



**Τεχνολογικό
Εκπαιδευτικό
Ίδρυμα Κρήτης**

**Τεχνολογικό Εκπαιδευτικό
Ίδρυμα Κρήτης
Παράρτημα Χανίων
Τμήμα Ηλεκτρονικών Μηχανικών**

Πτυχιακή Εργασία

**Εισαγωγή στην ανίχνευση σήματος και στην
εκτίμηση των παραμέτρων του.**

**Του φοιτητή
Θωμάδη Παναγιώτη
Α. Μ. 2352**

**Επιβλέπων Καθηγητής
Δρ. Κόκκινος Ευάγγελος**

Χανιά 2013

Περίληψη

Η παρούσα εργασία αναφέρεται στην ανίχνευση και εκτίμηση σήματος που έχει πολλές εφαρμογές, όπως στο δέκτη του ραντάρ. Από την τυχαία παρατήρηση του σήματος λήψης επιθυμούμε να εξάγουμε πληροφορίες για κάποιο φαινόμενο που συνδέεται με αυτή την παρατήρηση. Μελετώνται τα προβλήματα ανίχνευσης, στα οποία επιθυμούμε να αποφασίσουμε ανάμεσα σ' έναν πεπερασμένο αριθμό πιθανών καταστάσεων και τα προβλήματα εκτίμησης, στα οποία επιθυμούμε να εκτιμήσουμε την τιμή (τιμές) κάποιου μεγέθους ή μεγεθών που δεν παρατηρούνται άμεσα. Παρουσιάζονται τα βασικά στοιχεία του δυαδικού ελέγχου υπόθεσης με διακριτού χρόνου παρατηρήσεις, καθώς και η ανάλυση της απόδοσης τους. Τέλος γίνεται αναφορά σε προβλήματα εκτίμησης παραμέτρου του σήματος σε διακριτό χρόνο.

Summary

This project presents signal detection and estimation and its many applications, such as in the radar receiver. From the random observation of the reception signal we wish to extract information about a phenomenon associated with this observation. Detection problems are studied in order to decide between a finite number of possible states; and estimation problems are also studied in order to estimate those values of size that are not directly observable. We also present the basic elements of binary hypothesis-testing with discrete-time observations, as well as the analysis of their performance. Finally, we discuss problems of signal parameter estimation within discrete time.

Εισαγωγή

Μιλώντας γενικά, η ανίχνευση και εκτίμηση σήματος είναι το πεδίο μελέτης που αφορά στην επεξεργασία των φερόντων πληροφορίας σημάτων με σκοπό την εξαγωγή αυτών των πληροφοριών. Οι εφαρμογές της θεωρίας ανίχνευσης και εκτίμησης σήματος απαντούν σε πολλά πεδία, όπως οι επικοινωνίες και ο αυτόματος έλεγχος. Για παράδειγμα, σε εφαρμογές επικοινωνιών, όπως η μετάδοση δεδομένων ή τα ραντάρ, η ανίχνευση και η εκτίμηση παρέχουν τη θεωρητική και αναλυτική βάση για το σχεδιασμό αποτελεσματικών δεκτών επικοινωνίας. Εναλλακτικά, σε εφαρμογές αυτόματου ελέγχου, η θεωρία ανίχνευσης και εκτίμησης οδηγεί σε τεχνικές για την εξαγωγή ακριβών συμπερασμάτων σχετικά με τις συνθήκες που υπάρχουν σε μια διαδικασία ή ένα σύστημα που πρόκειται να ελεγχθεί. Ένα παράδειγμα εφαρμογής στην οποία οι τεχνικές ανίχνευσης και εκτίμησης είναι χρήσιμες, είναι το ραντάρ, το οποίο μεταδίδει ένα παλμικό ηλεκτρομαγνητικό σήμα και κατόπιν περιμένει ένα πιθανό επιστρεφόμενο σήμα να ανακλαστεί από έναν στόχο. Λόγω των θορύβων των ηλεκτρικών δεκτών, των ατμοσφαιρικών διαταράξεων, των ψευδών ανακλάσεων από το έδαφος και άλλα αντικείμενα, συνήθως δεν είναι δυνατό να οριστεί με απόλυτη βεβαιότητα αν ένας στόχος είναι παρών. Άρα, κανείς θα πρέπει να συμπεράνει από την (ατελή) παρατήρηση της κεραίας λήψης αν ο στόχος είναι παρών και η θεωρία ανίχνευσης παρέχει το μέσο για την επιλογή μιας καλής τεχνικής για τη διαδικασία αυτή. Επιπλέον, έχοντας ορίσει με κάποιο βαθμό βεβαιότητας ότι ένας στόχος είναι παρών, τότε θα θέλαμε να εκτιμήσουμε κάποια χαρακτηριστικά του στόχου, όπως η θέση και η ταχύτητά του, μια διαδικασία που εμπίπτει στο γενικό πλαίσιο της θεωρίας εκτίμησης. Τέτοιες εκτιμήσεις μπορούν να αποδειχθούν χρήσιμες στο έλεγχο της κεραίας του ραντάρ, για τον εντοπισμό του στόχου ή για τον εκ του μακρόθεν έλεγχο του ίδιου στόχου προκειμένου να διατηρήσει την επιθυμητή τροχιά. Άλλες συγκεκριμένες εφαρμογές στις οποίες οι τεχνικές ανίχνευσης και εκτίμησης είναι χρήσιμες είναι η σεισμολογία, η ραδιοαστρονομία, τα σόναρ, η επεξεργασία λόγου και εικόνας, η επεξεργασία ιατρικών σημάτων και οι οπτικές επικοινωνίες. Γενικά, οι τεχνικές ανίχνευσης και εκτίμησης αφορούν στην εξαγωγή συμπερασμάτων από παρατηρήσεις που παραμορφώνονται ή αλλοιώνονται κατά κάποιο άγνωστο τρόπο. Εξάλλου, οι πληροφορίες που επιθυμούμε να αποκομίσουμε από τέτοιες παρατηρήσεις είναι κατά μείζονα λόγο άγνωστες στον παρατηρητή. Άρα είναι πολύ χρήσιμο να θέσουμε τα προβλήματα ανίχνευσης και εκτίμησης σ' ένα πιθανολογικό πλαίσιο στο οποίο η άγνωστη συμπεριφορά θεωρείται ότι είναι τυχαία. Υπό αυτό το πρίσμα, η θεωρία ανίχνευσης και εκτίμησης αρμόζει κατάλληλα στο πεδίο των στατιστικών συμπερασμάτων και αυτή είναι η ερμηνεία που θα χρησιμοποιηθεί σε όλη αυτή τη μελέτη. Βασικό στοιχείο της μελέτης της θεωρίας ανίχνευσης και εκτίμησης σήματος είναι η έννοια μια τυχαίας παρατήρησης Y που λαμβάνει τιμές σ' ένα σύνολο παρατηρήσεων Γ , που μπορεί να είναι ένα σύνολο διανυσμάτων, κυματομορφών, πραγματικών αριθμών ή όποιο άλλο σύνολο. Από την παρατήρηση του Y επιθυμούμε να εξάγουμε πληροφορίες για κάποιο φαινόμενο που συνδέεται με το Y . Υπάρχουν δύο τύποι τέτοιων προβλημάτων που μας ενδιαφέρουν: προβλήματα ανίχνευσης, στα οποία επιθυμούμε να αποφασίσουμε ανάμεσα σ' έναν πεπερασμένο αριθμό πιθανών καταστάσεων ή «φυσικών καταστάσεων»

και προβλήματα εκτίμησης, στα οποία επιθυμούμε να εκτιμήσουμε την τιμή (τιμές) κάποιου μεγέθους ή μεγεθών που δεν παρατηρούνται άμεσα. Σε κάθε περίπτωση, η σύνδεση μεταξύ της παρατήρησης και της επιθυμούμενης πληροφορίας είναι μάλλον πιθανολογική παρά άμεση, υπό την έννοια ότι η στατιστική συμπεριφορά του Y επηρεάζεται από τη φυσική κατάσταση ή τις τιμές των μεγεθών που πρόκειται να εκτιμηθούν. Άρα ένα μοντέλο για αυτή την κατάσταση θα πρέπει να περιλαμβάνει μια οικογένεια κατανομών πιθανότητας στο Γ , τα μέλη της οποίας αντιστοιχούν σε στατιστικές συνθήκες παρούσες στις διάφορες φυσικές καταστάσεις ή στις διάφορες τιμές των μεγεθών που εκτιμώνται. Δοθέντος ενός τέτοιου μοντέλου, το πρόβλημα ανίχνευσης και εκτίμησης είναι να βρεθεί ο βέλτιστος τρόπος επεξεργασίας της παρατήρησης Y προκειμένου να εξαχθούν οι επιθυμητές πληροφορίες. Τα βασικά χαρακτηριστικά που διακρίνουν τέτοιου είδους προβλήματα από άλλα είναι η φύση των επιθυμητών πληροφοριών (διακριτές ή συνεχείς), το πόσες εκ των προτέρων γνώσεις έχουμε σχετικά με τα μεγέθη ή τη φυσική κατάσταση, και τα κριτήρια επίδοσης με τα οποία συγκρίνουμε διάφορες διεργασίες ανίχνευσης και εκτίμησης. Στόχος αυτής της πτυχιακής εργασίας είναι να εισάγει τον αναγνώστη στις βασικές αρχές της θεωρίας ανίχνευσης και εκτίμησης. Τα Κεφάλαια II, III και VI πραγματεύονται προβλήματα ανίχνευσης σήματος. Το Κεφάλαιο II παρουσιάζει τα βασικά στοιχεία του δυαδικού ελέγχου υπόθεσης, που αποτελεί τη βάση για τα περισσότερα προβλήματα ανίχνευσης σήματος. Στο Κεφάλαιο III αυτά τα βασικά στοιχεία εφαρμόζονται ώστε να δώσουν τις βέλτιστες διαδικασίες για μοντέλα συγκεκριμένα για προβλήματα ανίχνευσης σήματος. Στο κεφάλαιο αυτό επίσης μελετάται η ανάλυση απόδοσης αυτών των διεργασιών. Το Κεφάλαιο III επίσης ασχολείται με διάφορες ειδικές μεθόδους ανίχνευσης σήματος που μπορούν να εφαρμοστούν σε προβλήματα ανίχνευσης σημάτων για μη πάγιες καταστάσεις. Τα Κεφάλαια II και III πραγματεύονται κυρίως καταστάσεις στις οποίες οι παρατηρήσεις είναι διανύσματα. Η περίπτωση αυτή αναφέρεται στην ανίχνευση σήματος με διακριτού χρόνου (δηλ. με δειγματοληψία) παρατηρήσεις. Το πρόβλημα ανίχνευσης σήματος με παρατηρήσεις συνεχούς χρόνου αναλύεται στο Κεφάλαιο IV. Το πρόβλημα αυτό, παρόλο που φιλοσοφικά είναι παρόμοιο με την περίπτωση του διακριτού χρόνου, αναλύεται ξεχωριστά λόγω των πιο προηγμένων αναλυτικών μεθόδων που απαιτούνται για την ανάλυσή του. Το Κεφάλαιο IV αφορά σε προβλήματα εκτίμησης διακριτού χρόνου. Σε αυτό το Κεφάλαιο, ερευνώνται τα στοιχεία και η δομή των προβλημάτων εκτίμησης παραμέτρου.

Σημείωση για τη σημειογραφία (συμβολισμό).

Η συγκεκριμενοποίηση της κατανομής πιθανότητας στο σύνολο παρατήρησης r απαιτεί την ανάθεση των πιθανοτήτων σε υποσύνολα του Γ . Για κάποιους χώρους παρατήρησης που μας ενδιαφέρουν δεν είναι δυνατή η ανάθεση συνεπών πιθανοτήτων σε όλα τα υποσύνολα του Γ . Άρα θα συνδέουμε πάντα με το Γ μια τάξη G των υποσυνόλων του Γ , στο οποίο θέλουμε να αναθέσουμε πιθανότητα. Τα σύνολα στο G ονομάζονται γεγονότα παρατήρησης και το ζεύγος (Γ, G) καλείται χώρος παρατήρησης. Για λόγους ανάλυσης υποθέτουμε πάντα ότι η συλλογή G είναι μια σ -άλγεβρα. Δηλαδή, υποθέτουμε ότι το G περιλαμβάνει όλα τα συμπληρώματα (σχετικά με το Γ) και μετρήσιμες (δηλ. αριθμήσιμες) ενώσεις των μελών του.

Στο μεγαλύτερο μέρος αυτού του βιβλίου θα ασχοληθούμε με δύο περιπτώσεις για το (Γ, G) . Η πρώτη είναι η περίπτωση στην οποία $\Gamma = \mathbb{R}^n$, το σύνολο των n -διάστατων διανυσμάτων με πραγματικά στοιχεία και η δεύτερη είναι η περίπτωση στην οποία το Γ είναι ένα διακριτό (δηλ. μετρήσιμο) σύνολο, $\Gamma = \{\gamma_1, \gamma_2, \dots\}$. Στην πρώτη από τις περιπτώσεις αυτές, είναι φυσικό να θέλουμε να αναθέσουμε πιθανότητες σε σύνολα της μορφής $\{y = (y_1, \dots, y_n)^T \in \mathbb{R}^n \mid a_1 \leq y_1 \leq b_1, \dots, a_n \leq y_n \leq b_n\}$, όπου τα a_i και τα b_i είναι αυθαίρετα πραγματικοί αριθμοί. Άρα, για $\Gamma = \mathbb{R}^n$, θα θεωρήσουμε ότι το G είναι η μικρότερη σ -άλγεβρα που περιλαμβάνει όλα αυτά τα σύνολα με τα a_i και τα b_i να κυμαίνονται σε όλους τους πραγματικούς. Αυτή η σ -άλγεβρα συνήθως συμβολίζεται με \mathcal{B}^n και καλείται τάξη των συνόλων *Borel* στο \mathbb{R}^n . Για το δεύτερο από αυτά τα δύο σύνολα παρατήρησης είναι δυνατόν να επιτρέψουμε στο G να είναι το σύνολο όλων των υποσυνόλων του Γ . Αυτή η σ -άλγεβρα συνήθως συμβολίζεται με 2^Γ και καλείται σύνολο δύναμης του Γ . Αυτοί οι δύο χώροι παρατήρησης θα είναι επαρκείς για τη μελέτη των περισσότερων προβλημάτων ανίχνευσης και εκτίμησης διακριτού χρόνου που αναλύονται στα Κεφάλαια II ως V. Μέχρι να διατυπωθεί διαφορετικά, θα υποθέτουμε πάντα ότι (Γ, G) είναι μία από αυτές τις δύο περιπτώσεις. Για την ανάλυση προβλημάτων συνεχούς χρόνου απαιτούνται πιο αφηρημένοι χώροι παρατήρησης και αυτοί θα παρουσιαστούν στα Κεφάλαια VI και VII.

Για το διακριτό χώρο παρατήρησης $(\Gamma, 2^\Gamma)$, οι πιθανότητες μπορούν να ανατεθούν σε υποσύνολα του Γ με τη μορφή μιας συνάρτησης μάζας πιθανότητας, $p: \Gamma \rightarrow [0, 1]$,

μέσω του

$$P(A) = \sum_{\gamma_i \in A} p(\gamma_i), \quad A \in 2^\Gamma, \quad (I.1)$$

όπου η $P(A)$ συμβολίζει την πιθανότητα η παρατήρηση Y να βρίσκεται στο υποσύνολο A . Οποιαδήποτε συνάρτηση που αντιστοιχεί το Γ στο $[0, 1]$ είναι μια αποδεκτή συνάρτηση μάζας πιθανότητας με την προϋπόθεση ότι ικανοποιεί τον όρο $\sum_{i=1}^{\infty} p(\gamma_i) = 1$. Για το χώρο παρατήρησης $(\mathbb{R}^n, \mathcal{B}^n)$, μας ενδιαφέρουν πρωτίστως τα λεγόμενα *συνεχή τυχαία διανύσματα* για τα οποία οι πιθανότητες μπορούν να ανατεθούν σε μια συνάρτηση πυκνότητας πιθανότητας $p: \mathbb{R}^n \rightarrow [0, \infty)$ της μορφής

$$P(A) = \int_A p(y) dy, \quad A \in \mathcal{B}^n. \quad (I.2)$$

(Σημειώστε ότι το ολοκλήρωμα στο (I.2) είναι n -πλάσιο). Οποιαδήποτε ολοκληρώσιμη συνάρτηση που αντιστοιχεί το \mathbb{R}^n στο $[0, \infty)$ είναι μια αποδεκτή συνάρτηση πυκνότητας πιθανότητας, με την προϋπόθεση ότι ικανοποιεί τη συνθήκη $\int_{\mathbb{R}^n} p(y) dy = 1$. Για χάρη της ορολογίας και του συμβολισμού θα χρησιμοποιήσουμε τον όρο *πυκνότητα* τόσο για την μαζική συνάρτηση πιθανότητας όσο και για τη συνάρτηση πυκνότητας πιθανότητας και θα χρησιμοποιήσουμε το συμβολισμό

$$P(A) = \int_A p(y)\mu(dy) \quad (I.3)$$

για να συμβολίσουμε τόσο το άθροισμα του (I.1) όσο και το n -πλάσιο του (I.2). Όπου η μεταβλητή του ολοκληρώματος εννοείται, θα χρησιμοποιούμε ενίοτε τον εναλλακτικό συμβολισμό

$$P(A) = \int_A p d\mu. \quad (I.4)$$

Για μια πραγματικής τιμής συνάρτηση g της τυχαίας παρατήρησης Y , μας απασχολεί η αναμενόμενη τιμή της $g(Y)$, που συμβολίζεται με $E\{g(Y)\}$. Στην περίπτωση ενός διακριτού χώρου παρατήρησης $(\Gamma, 2^\Gamma)$ το μέγεθος αυτό δίνεται από

$$E\{g(Y)\} = \sum_{i=1}^{\infty} g(\gamma_i)p(\gamma_i), \quad (I.5)$$

και στην περίπτωση ενός συνεχούς τυχαίου διανύσματος στο $(\mathbb{R}^n, \mathcal{B}^n)$, έχουμε

$$E\{g(Y)\} = \int_{\mathbb{R}^n} g(y)p(y)dy, \quad (I.6)$$

όπου σε κάθε περίπτωση έχουμε υποθέσει την ύπαρξη του απαιτούμενου αθροίσματος του ολοκληρώματος. Και πάλι, για χάρη του συμβολισμού, θα χρησιμοποιήσουμε τους ακόλουθους συμβολισμούς για το (I.5) και για το (I.6):

$$E\{g(Y)\} = \int_{\Gamma} g(y)p(y)\mu(dy) = \int_{\Gamma} gp \, d\mu. \quad (I.7)$$

Ο συμβολισμός αυτός θα αποκτήσει περαιτέρω νόημα στο Κεφάλαιο VI. Σημειώστε ότι τα (I.3) και (I.4) είναι ειδικές περιπτώσεις του (I.7) με το g να δίνεται από

$$g(y) = \begin{cases} 1, & \text{if } y \in A \\ 0, & \text{if } y \in A^c. \end{cases} \quad (I.8)$$

Στο σύνολο της μελέτης αυτής θα χρησιμοποιηθούν κεφαλαία γράμματα για το συμβολισμό των τυχαίων μεγεθών και πεζά γράμματα για το συμβολισμό συγκεκριμένων τιμών των μεγεθών αυτών. Άρα η τυχαία παρατήρηση Y μπορεί να λάβει την τιμή y .

ΣΤΟΙΧΕΙΑ ΕΛΕΓΧΟΥ ΥΠΟΘΕΣΗΣ

II. A Εισαγωγή

Τα περισσότερα προβλήματα ανίχνευσης σήματος μπορούν να τεθούν στο πλαίσιο του ελέγχου υπόθεσης M -ary, στην οποία έχουμε μια παρατήρηση (πιθανόν ένα διάνυσμα ή μια συνάρτηση), με βάση την οποία θέλουμε να αποφασίσουμε ανάμεσα σε M πιθανές στατιστικές καταστάσεις που περιγράφουν τις παρατηρήσεις. Για παράδειγμα, σε έναν M -ary δέκτη επικοινωνιών παρατηρούμε μια ηλεκτρική κυματομορφή που συνίσταται σε ένα από τα M πιθανά σήματα που αλλοιώνονται από τυχαίο κανάλι ή θόρυβο του δέκτη, και προσπαθούμε να αποφασίσουμε ποιο από τα M πιθανά σήματα είναι παρόν. Προφανώς, για οποιοδήποτε δεδομένο πρόβλημα απόφασης, υπάρχουν διάφορες πιθανές στρατηγικές ή κανόνες που θα μπορούσαν να εφαρμοστούν. Ωστόσο, θα θέλαμε να επιλέξουμε έναν κανόνα που είναι βέλτιστος υπό κάποια έννοια. Υπάρχουν διάφοροι χρήσιμοι ορισμοί βελτιστότητας για τέτοια προβλήματα και στο κεφάλαιο αυτό εξετάζουμε τις τρεις πιο κοινές διατυπώσεις-την Μπαϊσειανή, την ελαχιστομέγιστη, και του Neyman-Pearson- και εξάγουμε τις αντίστοιχες βέλτιστες λύσεις. Γενικώς, εξετάζουμε το ειδικό πρόβλημα του δυαδικού ($M=2$) ελέγχου υπόθεσης, παρόλο που η επέκταση πολλών από τα αποτελέσματα του κεφαλαίου αυτού στη γενική υπόθεση M -ary είναι άμεση και θα αναπτυχθεί στις ασκήσεις. Η εφαρμογή της θεωρίας αυτής στα μοντέλα που αφορούν στην ανίχνευση σημάτων εξετάζονται λεπτομερώς στα Κεφάλαια III και VI.

II. B. Μπαϊσειανός Έλεγχος Υπόθεσης (ή κατά Bayes)

Το κύριο πρόβλημα που αντιμετωπίζουμε στο κεφάλαιο αυτό είναι το απλό πρόβλημα του ελέγχου υπόθεσης στο οποίο υποθέτουμε ότι υπάρχουν δύο πιθανές υποθέσεις ή «φυσικές καταστάσεις» H_0 και H_1 , που αντιστοιχούν σε δύο πιθανές κατανομές πιθανότητας P_0 και P_1 , αντίστοιχα, στο χώρο παρατήρησης (Γ, G) . Μπορούμε να γράψουμε το πρόβλημα ως

$$H_0: Y \sim P_0$$

ως προς

$$H_1: Y \sim P_1$$

Όπου ο συμβολισμός $Y \sim P$ δηλώνει τη συνθήκη «το Y έχει κατανομή P ». Οι υποθέσεις H_0 και H_1 αναφέρονται ορισμένες φορές ως μηδενικές και εναλλακτικές υποθέσεις, αντίστοιχα. Ένας κανόνας απόφασης (ή έλεγχος υπόθεσης) δ για H_0 ως προς H_1 είναι κάθε μερισμός του συνόλου παρατήρησης Γ σε σύνολα $\Gamma_1 \in \mathcal{G}$ και $\Gamma_0 = \Gamma_1^c$: τέτοιος ώστε επιλέγουμε H_j όταν $\mathbf{y} \in \Gamma_j$ για $j=0$ ή 1 . Το σύνολο Γ_1 είναι ενίοτε γνωστό ως περιοχή απόρριψης ή κρίσιμη περιοχή, ενώ το σύνολο Γ_0 είναι γνωστό ως περιοχή αποδοχής. Μπορούμε επίσης να σκεφτούμε τον κανόνα απόφασης δ ως συνάρτηση στο Γ που δίνεται από

$$\delta(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \in \Gamma_1 \\ 0 & \text{if } \mathbf{y} \in \Gamma_1^c, \end{cases} \quad (\text{II.B.2})$$

Ωστε η τιμή του δ για ένα δεδομένο $y \in \Gamma$ είναι ο δείκτης της υπόθεσης που γίνεται δεκτή από το δ .

Θα θέλαμε να επιλέξουμε το Γ_1 με κάποιο βέλτιστο τρόπο και, έχοντας αυτό στο μυαλό μας, θα θέλαμε να αποδώσουμε κόστη στις αποφάσεις μας. Ειδικότερα, θα υποθέσουμε για τώρα ότι έχουμε θετικούς αριθμούς για C_{ij} για $i=0,1$ και $j=0,1$, όπου C_{ij} είναι το κόστος που προκύπτει με την επιλογή της υπόθεσης H_i όταν η υπόθεση H_j είναι αληθής. Μπορούμε τότε να προσδιορίσουμε την υπό συνθήκη διακινδύνευση για κάθε υπόθεση ως το μέσο ή αναμενόμενο κόστος που προκύπτει από τον κανόνα απόφασης δ όταν η υπόθεση αυτή είναι αληθής. Δηλαδή:

$$R_j(\delta) = C_{1j}P_j(\Gamma_1) + C_{0j}P_j(\Gamma_0), \quad j = 0,1. \quad (\text{II.B.3})$$

Σημειώστε ότι $R_j(\delta)$ είναι το κόστος επιλογής H_j όταν H_j είναι αληθές επί την πιθανότητα να γίνει αυτό συν το κόστος επιλογής H_0 όταν το H_j είναι αληθές επί την πιθανότητα να γίνει αυτό

Τώρα ας υποθέσουμε περαιτέρω ότι μπορούμε να αποδώσουμε πιθανότητες π_0 και $\pi_1 = (1 - \pi_0)$ στην εμφάνιση των υποθέσεων H_0 και H_1 αντίστοιχα. Δηλαδή, π_j είναι η πιθανότητα η υπόθεση H_j να είναι πραγματικά μορφοποιημένο ως προς την αξία του Y . Οι πιθανότητες αυτές π_0 και π_1 είναι γνωστές ως οι εκ των προτέρων ή a priori πιθανότητες των δύο υποθέσεων. Για a priori δεδομένα μπορούμε να ορίσουμε μια μέση ή κατά Bayes διακινδύνευση ως το συνολικό μέσο κόστος που προκύπτει από τον κανόνα απόφασης δ .

Η ποσότητα αυτή δίνεται από :

$$r(\delta) = \pi_0 R_0(\delta) + \pi_1 R_1(\delta). \quad (\text{II.B.4})$$

Μπορούμε τώρα να ορίσουμε έναν βέλτιστο κανόνα απόφασης για H_0 ως προς H_1 , ως αυτόν που ελαχιστοποιεί, σε σχέση με όλους τους άλλους κανόνες, την κατά Bayes διακινδύνευση. Μια τέτοια απόφαση είναι γνωστή ως απόφαση κατά Bayes για H_0 ως προς H_1

Ας σημειωθεί ότι τα (II.B.3) και (II.B.4) μπορούν να συνδυαστούν και να δώσουν

$$\begin{aligned}
r(\delta) &= \sum_{j=0}^1 \pi_j [C_{0j}(1 - P_j(\Gamma_1)) + C_{1j}P_j(\Gamma_1)] \\
&= \sum_{j=0}^1 \pi_j C_{0j} + \sum_{j=0}^1 \pi_j (C_{1j} - C_{0j})P_j(\Gamma_1) \quad (\text{II.B.5})
\end{aligned}$$

όπου έχουμε χρησιμοποιήσει το δεδομένο ότι $P_j(\Gamma_1^c) = 1 - P_j(\Gamma_1)$. Αν υποθέσουμε ότι το P_j έχει πυκνότητα p_j για $j = 0, 1$, και χρησιμοποιώντας το συμβολισμό που είδαμε στο Κεφάλαιο I (II.B.5) συνάγεται ότι

$$\begin{aligned}
r(\delta) &= \sum_{j=0}^1 \pi_j C_{0j} \\
&\quad + \int_{\Gamma_1} \left[\sum_{j=0}^1 \pi_j (C_{1j} - C_{0j}) p_j(y) \right] \mu(dy), \quad (\text{II.B.6})
\end{aligned}$$

Και επομένως βλέπουμε ότι $r(\delta)$ είναι ένα ελάχιστο συνολικό Γ_1 αν επιλέξουμε

$$\begin{aligned}
\Gamma_1 &= \left\{ y \in \Gamma \left| \sum_{j=0}^1 \pi_j (C_{1j} - C_{0j}) p_j(y) \leq 0 \right. \right\} \\
&= \{ y \in \Gamma | \pi_1 (C_{11} - C_{01}) p_1(y) \leq \pi_0 (C_{00} - C_{10}) p_0(y) \}. \quad (\text{II.B.7})
\end{aligned}$$

Υποθέτοντας ότι $C_{11} < C_{01}$ (ότι δηλαδή το κόστος της ορθής επιλογής του H_1 είναι μικρότερο από το κόστος της εσφαλμένης απόρριψης του H_1), το (II.B.7) μπορεί να ξαναγραφεί ως :

$$\Gamma_1 = \{ y \in \Gamma | p_1(y) \geq \tau p_0(y) \} \quad (\text{II.B.8})$$

Όπου

$$\tau \triangleq \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})}. \quad (\text{II.B.9})$$

Ας σημειωθεί ότι το πεδίο $\{y \in \Gamma | p_1(y) = \tau p_0(y)\}$ δε συμβάλλει στο μέσο σφάλμα και επομένως μπορεί να παραλειφθεί εξ ολοκλήρου ή εν μέρει από το Γ_1 χωρίς να επηρεάζει τον κίνδυνο που προκύπτει.

Ο κανόνας απόφασης που περιγράφεται στην περιοχή απόρριψης του (II.B.8) είναι γνωστός ως έλεγχος λόγου πιθανοφάνειας (ή έλεγχος λόγου πιθανότητας). Ο έλεγχος αυτός παίζει κεντρικό ρόλο στη θεωρία ελέγχου υπόθεσης. Ας σημειωθεί ότι το Γ_1 του (II.B.8.) μπορεί να ξαναγραφεί ως

$$\Gamma_1 = \{y \in \Gamma | [p_1(y)/p_0(y)] \geq \tau\}, \quad (\text{II.B.10})$$

Όπου ερμηνεύουμε το $(k/0)$ ως ∞ για κάθε $k \geq 0$. Το μέγεθος

$$L(y) = \frac{p_1(y)}{p_0(y)}, \quad y \in \Gamma, \quad (\text{II.B.11})$$

είναι γνωστό ως *λόγος πιθανοφάνειας* (ή *στατιστική λόγου πιθανοφάνειας*) μεταξύ του H_0 και του H_1 . Έτσι ο κανόνας απόφασης κατά Bayes που αφορά στο (II.B.8) υπολογίζει το λόγο πιθανοφάνειας για την παρατηρούμενη τιμή του Y και μετά επιλέγει την απόφαση συγκρίνοντας το λόγο αυτό με το κατώφλι τ . Δηλαδή, ένας κανόνας κατά Bayes για το (II.B.1) είναι

$$\delta_B(y) = \begin{cases} 1 & \text{if } L(y) \geq \tau \\ 0 & \text{if } L(y) < \tau. \end{cases} \quad (\text{II.B.12})$$

Μια κοινώς χρησιμοποιούμενη απόδοση κόστους είναι η απόδοση ομοιόμορφου κόστους που δίνεται από

$$C_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases} \quad (\text{II.B.13})$$

Η διακινδύνευση κατά Bayes για έναν κανόνα απόφασης δ με κρίσιμο πεδίο Γ_1 δίνεται στην περίπτωση αυτή από

$$r(\delta) = \pi_0 P_0(\Gamma_1) + \pi_1 P_1(\Gamma_0). \quad (\text{II.B.14})$$

Ας σημειωθεί ότι $P_i(\Gamma_j)$ είναι η πιθανότητα επιλογής του H_j όταν το H_i είναι αληθές. Επομένως, $P_i(\Gamma_j)$ για $i \neq j$ είναι η υπό συνθήκη (ή δεσμευμένη) πιθανότητα πραγματοποίησης σφάλματος δεδομένου ότι το H_i είναι αληθές, και έτσι στην περίπτωση αυτή του $r(\delta)$ είναι η μέση πιθανότητα σφάλματος που προκύπτει από τον κανόνα απόφασης δ . Αφού η εξέταση λόγου πιθανοφάνειας με $\tau = \pi_0/\pi_1$ ελαχιστοποιεί το $r(\delta)$ για το κόστος δομής του (II.B.13), αποτελεί επομένως μία μέθοδο ελάχιστης πιθανότητας σφάλματος απόφασης.

Ο Μπαϊεσιανός τύπος (βλέπε, Thomas (1986) υποδεικνύει ότι η υπό συνθήκη πιθανότητα η υπόθεση H_j να είναι αληθής, δοθέντος ότι η τυχαία παρατήρηση Y παίρνει την τιμή y , δίνεται από

$$\pi_j(y) = P(H_j \text{ true} | Y = y) = \frac{p(y)\pi_j}{p(y)}, \quad (\text{II.B.15})$$

Όπου $p(y)$ είναι η μέση ή συνολική πυκνότητα του Y όπως δίνεται από $p(y) = \pi_0 p_0(y) + \pi_1 p_1(y)$. Οι πιθανότητες $\pi_0(y)$ και $\pi_1(y)$ ονομάζονται ύστερες ή εκ των υστέρων πιθανότητες των δύο υποθέσεων. Χρησιμοποιώντας το (II.B.15), το κρίσιμο πεδίο του Μπαϊεσιανού κανόνα (II.B.7) μπορεί να ξαναγραφεί ως

$$\Gamma_1 = \{y \in \Gamma | C_{10}\pi_0(y) + C_{11}\pi_1(y) \leq C_{00}\pi_0(y) + C_{01}\pi_1(y)\}. \quad (\text{II.B.16})$$

Επομένως, σ' αυτό το Μπαϊεσιανό πλαίσιο, οι βέλτιστες αποφάσεις βασίζονται σε ύστερες πιθανότητες και θα μπορούσαμε να θεωρήσουμε τη διαδικασία παρατήρησης ως ένα μηχανισμό ανανέωσης των πρότερων πιθανοτήτων των υποθέσεων με τις ύστερες πιθανότητες.

Ας σημειωθεί ότι το μέγεθος

$$C_{i0}\pi_0(y) + C_{i1}\pi_1(y) \quad (\text{II.B.17})$$

είναι το μέσο κόστος που προκύπτει από την επιλογή της υπόθεσης H_i δοθέντος ότι το Y ισούται με το y . Το μέγεθος αυτό καλείται το ύστερο κόστος επιλογής του H_i όταν η παρατήρηση είναι το y και επομένως ο Μπαϊεσιανός κανόνας παίρνει την απόφασή του επιλέγοντας

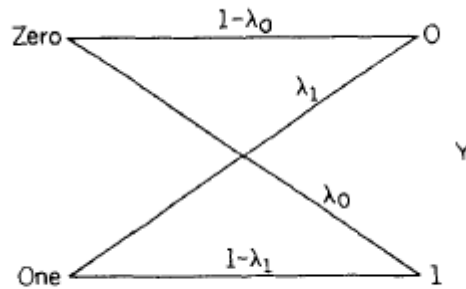


FIGURE II.B.1. The binary channel.

Σχήμα II.B.1 Το δυαδικό κανάλι

την υπόθεση που παράγει το ελάχιστο ύστερο κόστος. Για παράδειγμα για το κριτήριο ομοιόμορφου κόστους (II.B.13), ο Μπαϊεσιανός κανόνας μπορεί επομένως να γραφεί ως

$$\delta_B(y) = \begin{cases} 1 & \text{if } \pi_1(y) \geq \pi_0(y) \\ 0 & \text{if } \pi_1(y) < \pi_0(y). \end{cases} \quad (\text{II.B.18})$$

Επομένως, ο κανόνας ελάχιστης πιθανότητας σφάλματος απόφασης επιλέγει την υπόθεση που έχει τη μέγιστη εκ των υστέρων πιθανότητα να έχει προκύψει, δοθέντος ότι $Y = y$. Αυτός ο κανόνας απόφασης είναι ενίοτε γνωστός ως κανόνας απόφασης MAP για τον δυαδικό έλεγχο απόφασης (II.B.1). Τα ακόλουθα απλά παραδείγματα απεικονίζουν τον Μπαϊεσιανό (ή κατά Bayes) κανόνα απόφασης.

Παράδειγμα II.B.1. : Το Δυαδικό Κανάλι

Ας υποθέσουμε ότι ένα δυαδικό ψηφίο (δηλαδή, ένα «μηδέν» ή ένα «ένα») πρόκειται να μεταδοθεί από ένα κανάλι επικοινωνίας. Η παρατήρηση Y είναι η έξοδος του καναλιού που μπορεί να είναι είτε 0 είτε 1. Λόγω των θορύβων του καναλιού και της ατελούς διαμόρφωσης ή αποδιαμόρφωσης, ένα μεταδιδόμενο «μηδέν» λαμβάνεται ως 1 με πιθανότητα λ_0 και ως 0 με πιθανότητα $(1 - \lambda_0)$, όπου $0 < \lambda_0 < 1$. Παρομοίως, ένα μεταδιδόμενο «ένα» λαμβάνεται ως 0 με πιθανότητα λ_1 και ως 1 με πιθανότητα $(1 - \lambda_1)$. (Οι σχέσεις αυτές απεικονίζονται στο σχήμα II.B.1).

Επομένως, η παρατήρηση του Y δεν μας λέει ακριβώς αν το μεταδιδόμενο ψηφίο ήταν «μηδέν» ή «ένα» και θα θέλαμε να βρούμε έναν καλύτερο τρόπο για να αποφασίσουμε τι μεταδόθηκε.

Η κατάσταση μπορεί να μοντελοποιηθεί ως ένα πρόβλημα δυαδικού ελέγχου υπόθεσης στην οποία η υπόθεση H_j είναι ότι μεταδόθηκε ένα “ j ” ($j = 0,1$), το σύνολο παρατήρησης Γ είναι $\{0, 1\}$ και η παρατήρηση Y έχει πυκνότητες (δηλ. μαζικές λειτουργίες πιθανότητας)

$$p_j(y) = \begin{cases} \lambda_j & \text{if } y \neq j \\ (1 - \lambda_j) & \text{if } y = j \end{cases} \quad (\text{II.B.19})$$

Για $j = 0$ και 1 . Άρα ο λόγος πιθανοφάνειας δίνεται από

$$L(y) = \frac{p_1(y)}{p_0(y)} = \begin{cases} \frac{\lambda_1}{1 - \lambda_0} & \text{if } y = 0 \\ \frac{1 - \lambda_1}{\lambda_0} & \text{if } y = 1. \end{cases} \quad (\text{II.B.20})$$

Για έναν Μπαϊεσιανό έλεγχο, το κατώφλι ελέγχου τ ορίζεται από τα κόστη και τις πρότερες πιθανότητες από (II.B.9). Αν λ_1, λ_0 και τ είναι τέτοια ώστε $\lambda_1 \geq \tau(1 - \lambda_0)$, ο έλεγχος λόγου πιθανοφάνειας του (II.B.12) ερμηνεύει ένα λαμβανόμενο 0 ως ένα μεταδιδόμενο «ένα». Αλλιώς, ένα λαμβανόμενο 0 ερμηνεύεται ως ένα μεταδιδόμενο «μηδέν». Παρομοίως, αν $(1 - \lambda_1) \geq \tau \lambda_0$, ο έλεγχος λόγου πιθανοφάνειας ερμηνεύει ένα λαμβανόμενο 1 ως ένα μεταδιδόμενο «ένα», και αν $(1 - \lambda_1) < \tau \lambda_0$, ένα λαμβανόμενο 1 ερμηνεύεται ως ένα μεταδιδόμενο «μηδέν». Η ελάχιστη Μπαϊεσιανή διακινδύνευση $r(\delta_B)$ για το παράδειγμα αυτό μπορεί να υπολογιστεί ευθέως από το (II.B.5) (βλέπε Άσκηση 1).

Για παράδειγμα, με ομοιόμορφα κόστη και των προτέρων ίσα ($\pi_0 = \pi_1 = 1/2$), έχουμε ότι $\tau = 1$ και ο Μπαϊεσιανός κανόνας του (II.B.12) γίνεται

$$\delta_B(0) = \begin{cases} 1 & \text{if } \lambda_1 \geq (1 - \lambda_0) \\ 0 & \text{if } \lambda_1 < (1 - \lambda_0). \end{cases} \quad (\text{II.B.21a})$$

$$\delta_B(1) = \begin{cases} 1 & \text{if } (1 - \lambda_1) \geq \lambda_0 \\ 0 & \text{if } (1 - \lambda_1) < \lambda_0. \end{cases} \quad (\text{II.B.21b})$$

Αν επιπλέον υποθέσουμε ένα *συμμετρικό* κανάλι $\lambda_1 = \lambda_0 = \lambda$, τότε το (II.B.22) γίνεται

$$\delta_B(y) = \begin{cases} y & \text{if } \lambda \leq 1/2 \\ 1 - y & \text{if } \lambda > 1/2. \end{cases} \quad (\text{II.B.23})$$

Η ερμηνεία του (II.B.23) είναι άμεση – αν το κανάλι είναι πιο πιθανό να αντιστρέψει τα δυαδικά ψηφία από το να μην το κάνει (δηλ., $\lambda > 1/2$), παίρνουμε την απόφασή μας αντιστρέφοντας το ληφθέν δυαδικό ψηφίο. Διαφορετικά, δεχόμαστε το ληφθέν δυαδικό ψηφίο ως σωστό. Για την τελευταία περίπτωση, η ελάχιστη Μπαϊεσιανή διακινδύνευση αποδεικνύεται ότι είναι

$$r(\delta_B) = \min\{\lambda, 1 - \lambda\}. \quad (\text{II.B.24})$$

Άρα η απόδοση βελτιώνεται καθώς το κανάλι γίνεται πιο αξιόπιστο είτε στο να μεταδίδει το δυαδικό ψηφίο ευθέως είτε στο να αντιστρέφει το μεταδιδόμενο δυαδικό ψηφίο. Ας σημειωθεί ότι λόγω των ομοιόμορφων κοστών και εκ των προτέρων ισοδύναμων, η απλή εικασία για το μεταδιδόμενο δυαδικό ψηφίο χωρίς την παρατήρηση του y παράγει τον κίνδυνο του $1/2$. Άρα αν το $\lambda = 1/2$, η παρατήρηση είναι άνευ αξίας.

Παράδειγμα II.B.2: Έλεγχος θέσης με σφάλμα του Gauss

Μελετήστε τις ακόλουθες δύο υποθέσεις που αφορούν σε αληθινής – τιμής παρατήρηση Y :

$$\begin{array}{l} H_0 : Y = \epsilon + \mu_0 \\ \text{versus} \\ H_1 : Y = \epsilon + \mu_1 \end{array} \quad (\text{II.B.25})$$

Όπου ϵ είναι μια τυχαία μεταβλητή του Gauss με τιμή μηδέν και διακύμανση σ^2 , και όπου μ_0 και μ_1 είναι δύο σταθεροί αριθμοί με $\mu_1 > \mu_0$. Σημειώστε η πρόσθεση του μ_0 ή μ_1 στο ϵ αλλάζει μόνο τη μέση τιμή της παρατήρησης, έτσι ελέγχουμε σε ποια από τις δύο πιθανές τιμές ή «θέσεις» κατανέμεται η παρατήρηση. Οι εφαρμογές της πιο γενικής μορφής του απλού αυτού μοντέλου θα συζητηθεί αργότερα. Σε σχέση με τις κατανομές στο χώρο παρατήρησης, το ζεύγος υπόθεσης του (II.B.25) μπορεί να ξαναγραφεί ως

$$\begin{array}{l} H_0 : Y \sim \mathcal{N}(\mu_0, \sigma^2) \\ \text{versus} \\ H_1 : Y \sim \mathcal{N}(\mu_1, \sigma^2), \end{array} \quad (\text{II.B.26})$$

Όπου $N(\mu, \sigma^2)$ συμβολίζει την κατανομή Gauss ή (κανονική κατανομή) με μέσο μ και διακύμανση σ^2 . [Θυμηθείτε ότι μια $N(\mu, \sigma^2)$ τυχαία μεταβλητή είναι μία που έχει συνάρτηση πυκνότητας πιθανότητας

$$(1/\sqrt{2\pi}\sigma) \exp\{-(x - \mu)^2/2\sigma^2\}, x \in \mathbb{R}.$$

Ο λόγος πιθανοφάνειας για (II.B.26) δίνεται από

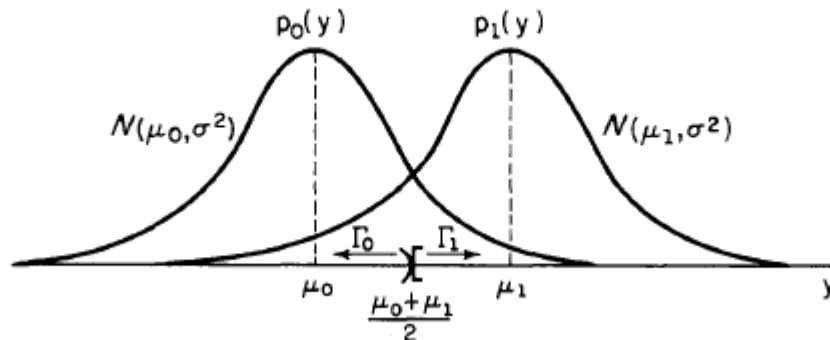
$$\begin{aligned} L(y) &= \frac{p_1(y)}{p_0(y)} = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu_1)^2/2\sigma^2}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu_0)^2/2\sigma^2}} \\ &= \exp\left\{\frac{\mu_1 - \mu_0}{\sigma^2} \left(y - \frac{\mu_0 + \mu_1}{2}\right)\right\}. \end{aligned} \quad (\text{II.B.27})$$

Άρα ο Μπαϊεσιανός έλεγχος για το (II.B.26) είναι

$$\delta_B(y) = \begin{cases} 1 & \text{if } \exp\left\{\frac{\mu_1 - \mu_0}{\sigma^2} \left(y - \frac{\mu_0 + \mu_1}{2}\right)\right\} \geq \tau \\ 0 & \text{otherwise} \end{cases} \quad (\text{II.B.28})$$

όπου το τ είναι το κατάλληλο κατώφλι. Αφού $\mu_1 > \mu_0$, ο λόγος πιθανοφάνειας του (II.B.27) είναι μια αυστηρά αυξανόμενη συνάρτηση της παρατήρησης y (δηλαδή, $dL(y)/dy = (\mu_1 - \mu_0)L(y)/\sigma^2 > 0$). Έτσι η σύγκριση του $L(y)$ με το κατώφλι τ είναι ισοδύναμη με τη σύγκριση του ίδιου του y με άλλο κατώφλι $\tau' = L^{-1}(\tau)$, όπου το L^{-1} είναι η αντίστροφη συνάρτηση του L . Συγκεκριμένα, η θέση λογαρίθμων στην ανισότητα του (II.B.28) και η αναδιάταξη των όρων δίνει

$$\delta_B(y) = \begin{cases} 1 & \text{if } y \geq \tau' \\ 0 & \text{if } y < \tau', \end{cases} \quad (\text{II.B.29})$$



Σχήμα II.B.2 Απεικόνιση του ελέγχου θέσης με σφάλματα Gauss, ομοιόμορφα κόστη και εκ των προτέρων ίσα

Όπου

$$\tau' = \frac{\sigma^2}{\mu_1 - \mu_0} \log(\tau) + \frac{\mu_0 + \mu_1}{2}. \quad (\text{II.B.30})$$

Για παράδειγμα, με ομοιόμορφα κόστη και εκ των προτέρων ίσα έχουμε $\tau = 1$ και $\tau' = (\mu_0 + \mu_1)/2$. Επομένως, σε αυτή την συγκεκριμένη περίπτωση, ο Μπαϊεσιανός κανόνας συγκρίνει την παρατήρηση με το μέσο του μ_0 και μ_1 . Αν το y είναι μεγαλύτερη ή ίσο από το μέσο, επιλέγουμε το H_1 . Αν το y είναι μικρότερο από αυτό το μέσο, επιλέγουμε το H_0 . Η εξέταση αυτή απεικονίζεται στο Σχ. II.B.2. Η ελάχιστη διακινδύνευση κατά Bayes, $r(\delta_B)$, για το πρόβλημα αυτό μπορεί να υπολογιστεί από (II.B.5) αν έχουμε $P_j(\Gamma_1)$ για $j = 0, 1$. Αφού $\Gamma_1 = \{y \in \mathbb{R} | y \geq \tau'\}$, έχουμε ότι

$$\begin{aligned} P_j(\Gamma_1) &= \int_{\tau'}^{\infty} p_j(y) dy = 1 - \Phi\left(\frac{\tau' - \mu_j}{\sigma}\right) \\ &= \begin{cases} 1 - \Phi\left(\frac{\log \tau}{d} + \frac{d}{2}\right), & j = 0 \\ 1 - \Phi\left(\frac{\log \tau}{d} - \frac{d}{2}\right), & j = 1, \end{cases} \end{aligned} \quad (\text{II.B.31})$$

όπου το Φ συμβολίζει την αθροιστική συνάρτηση κατανομής (cdf) πιθανότητας μιας $N(0,1)$ τυχαίας μεταβλητής και όπου $d \triangleq (\mu_1 - \mu_0)/\sigma$. Αν εξετάσουμε εκ νέου τη συγκεκριμένη περίπτωση των ομοιόμορφων κοστών και των εκ των προτέρων ίσων, έχουμε ευθέως ότι

$$r(\delta_B) = 1 - \Phi(d/2), \quad (\text{II.B.32})$$

που απεικονίζεται στο Σχ. II.B.3 ως συνάρτηση του d . Σημειώστε ότι η απόδοση βελτιώνεται μονοτονικά, καθώς ο διαχωρισμός των μέσων, $(\mu_1 - \mu_2)$, αυξάνεται σε σχέση με τη συνήθη απόκλιση του σφάλματος, σ . Αυτή η ποσότητα d είναι μια απλή εκδοχή του λόγου σήματος προς θόρυβο, και θα αποκτήσει περαιτέρω νόημα σε ακόλουθα κεφάλαια.

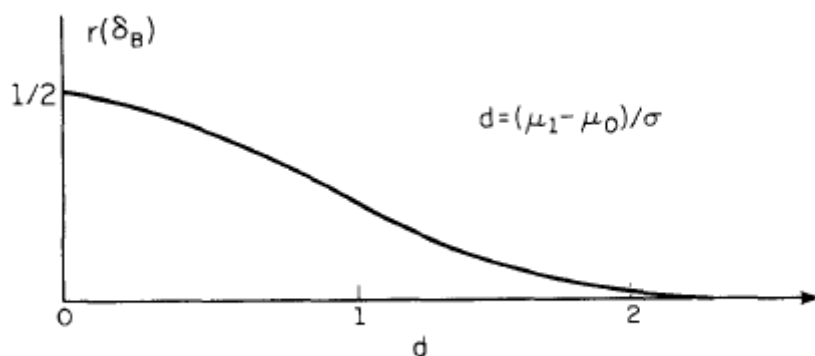


FIGURE II.B.3. Bayes risk in location testing with Gaussian error.

Σχήμα II.B.3. Μπαϊεσιανή διακινδύνευση σε έλεγχο θέσης με σφάλμα Gauss.

Τα παραδείγματα II.B.1. και II.B.2, αν και αρκετά απλά, αποδίδουν τις βασικές αρχές του Μπαϊεσιανού ελέγχου υπόθεσης. Περισσότερα παραδείγματα θα εξεταστούν στα Κεφάλαια III. και IV και μια σειρά σχετικών ασκήσεων περιλαμβάνονται στο τέλος αυτού του κεφαλαίου

Το κύριο συμπέρασμα αυτής της ενότητας είναι ότι ο Μπαϊεσιανός ορισμός της βελτιστότητας θεωρεί τον έλεγχο λόγου πιθανοφάνειας (II.B.12) ως τον βέλτιστο, όταν τα κόστη C_{ij} και τα πρότερα π_i είναι καθορισμένα. Στις ακόλουθες ενότητες εξετάζουμε άλλους ορισμούς βελτιστότητας για καταστάσεις στις οποίες τα πρότερα και/ή τα κόστη είναι άγνωστα.

II. C Ελαχιστομέγιστος έλεγχος υπόθεσης

Ας υποθέσουμε ότι στη διατύπωση της ενότητας II.B., οι πρότερες πιθανότητες π_0 και π_1 είναι άγνωστες στο σχεδιαστή. Τέτοιες καταστάσεις μπορούν να προκύψουν συχνά στην πράξη, καθώς ο σχεδιαστής ενός κανόνα απόφασης μπορεί να μην έχει έλεγχο ή πρόσβαση στο μηχανισμό που παράγει τη φυσική κατάσταση. Σε τέτοιες περιπτώσεις, η μέση ή Μπαϊεσιανή διακινδύνευση δεν είναι αποδεκτό κριτήριο αφού είναι μάλλον απίθανο ένας μόνο κανόνας απόφασης να ελαχιστοποιεί τη μέση διακινδύνευση για κάθε πιθανή πρότερη κατανομή. Επομένως στην περίπτωση αυτή είναι απαραίτητο να αναζητηθεί ένα εναλλακτικό κριτήριο

σχεδιασμού. Ένα τέτοιο κριτήριο είναι η αναζήτηση ενός κανόνα απόφασης που ελαχιστοποιεί για το δ το μέγιστο από τους υπό συνθήκη κινδύνους, $R_0(\delta)$ και $R_1(\delta)$. Ένα πιθανό κριτήριο σχεδιασμού είναι

$$\max\{R_0(\delta), R_1(\delta)\}. \quad (\text{II.C.1})$$

Ένας κανόνας απόφασης που ελαχιστοποιεί την ποσότητα στο (II.C.1) είναι γνωστός ως *ελαχιστομέγιστος κανόνας* και σ' αυτήν την υποενότητα θα εξετάσουμε τη δομή τέτοιων κανόνων.

Για να αναζητήσουμε έναν κανόνα απόφασης που ελαχιστοποιεί την ποσότητα στο (II.C.1) είναι χρήσιμο να λάβουμε υπόψη μας τη συνάρτηση $r(\pi_0, \delta)$, που ορίζεται για ένα δεδομένο πρότερο $\pi_0 \in [0,1]$ και τον κανόνα απόφασης δ ως μέσο κίνδυνο

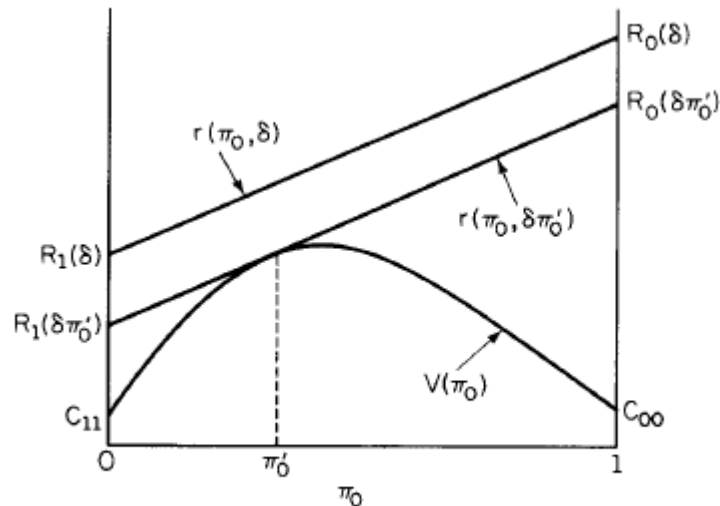


FIGURE II.C.1. Illustration of the functions $r(\pi_0, \delta)$ and $V(\pi_0)$.

Σχήμα II.C.1. Απεικόνιση των συναρτήσεων $r(\pi_0, \delta)$ και $V(\pi_0)$

$$r(\pi_0, \delta) = \pi_0 R_0(\delta) + (1 - \pi_0) R_1(\delta). \quad (\text{II.C.2})$$

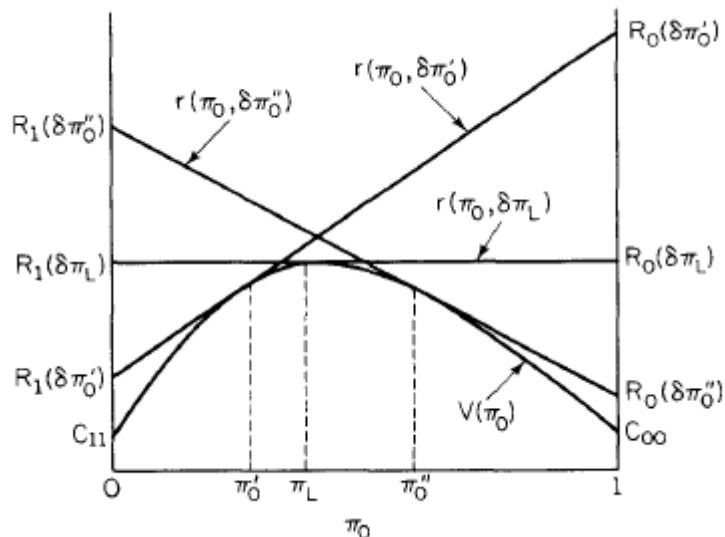
Ας σημειωθεί ότι ως συνάρτηση του π_0 , η $r(\pi_0, \delta)$ είναι μια ευθεία γραμμή από $r(0, \delta) = R_1(\delta)$ στο $r(1, \delta) = R_0(\delta)$, όπως απεικονίζεται στο Σχήμα II.C.1. Επομένως, για σταθερό δ , η μέγιστη τιμή του $r(\pi_0, \delta)$, καθώς το π_0 κυμαίνεται από 0 ως το 1, προκύπτει είτε στο $\pi_0 = 0$ ή $\pi_0 = 1$ και η μέγιστη τιμή είναι $\max\{R_0(\delta), R_1(\delta)\}$. Επίσης το πρόβλημα της ελαχιστοποίησης του (II.C.1) στο δ είναι το ίδιο με αυτό της ελαχιστοποίησης της ποσότητας στο δ . Η τελευταία ποσότητα είναι πιο βολική να ληφθεί υπόψη.

$$\max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta)$$

(II.C.3)

Για κάθε πρότερο $\pi_0 \in [0,1]$, έστω ότι δ_{π_0} συμβολίζει έναν Μπαϊεσιανό κανόνα που αντιστοιχεί σε αυτό το πρότερο και έστω ότι $V(\pi_0) = r(\pi_0, \delta_{\pi_0})$. Αυτό σημαίνει ότι, $V(\pi_0)$ είναι η ελάχιστη πιθανή Μπαϊεσιανή διακινδύνευση για το πρότερο π_0 . Θα πρέπει άμεσα να αποδειχθεί ότι το $V(\pi_0)$ είναι μια συνεχής κοίλα συνάρτηση του π_0 για $\pi_0 \in [0,1]$ με $V(0) = C_{11}$ και $V(1) = C_{00}$ (Βλέπε Άσκηση 8). Μια τυπική $V(\pi_0)$ καμπύλη απεικονίζεται στο Σχήμα II.C.1.

Για χάρη της συζήτησης ας υποθέσουμε ότι $V(\pi_0)$ και $r(\pi_0, \delta)$ είναι όπως απεικονίζονται στο Σχήμα II.C.1. Επίσης στο Σχήμα II.C.1. φαίνεται η γραμμή $r(\pi_0, \delta_{\pi_0})$ που είναι παράλληλη στο $r(\pi_0, \delta)$ και εφαπτόμενη στο $V(\pi_0)$. Ας σημειωθεί ότι, για την περίπτωση αυτή, το δ δεν μπορεί να είναι ελαχιστομέγιστος κανόνας, γιατί η γραμμή διακινδύνευσης που συμβολίζεται ως $r(\pi_0, \delta_{\pi_0})$ βρίσκεται τελείως κάτω από το $r(\pi_0, \delta)$ και επομένως έχει μικρότερη μέγιστη τιμή. Καθώς το $r(\pi_0, \delta_{\pi_0})$ αγγίζει το $V(\pi_0)$ στο $\pi_0 = \pi_0$, δ_{π_0} είναι Μπαϊεσιανός κανόνας για το πρότερο π_0 . Αφού μια παρόμοια εφαπτόμενη γραμμή (δηλ., μία που μειώνει και τους δύο υπό συνθήκη κινδύνους) μπορεί να χαρακτηί για οποιοδήποτε κανόνα απόφασης δ , είναι εύκολο να δει κανείς ότι μόνο οι Μπαϊεσιανοί κανόνες μπορούν να είναι ελαχιστομέγιστοι γι' αυτό το σχήμα.



Σχήμα II.C.2. Απεικόνιση του ελαχιστομέγιστου κανόνα όταν το V έχει ένα εσωτερικό μέγιστο.

Επιπλέον, εξετάζοντας το Σχ. II.C.2, βλέπουμε ότι ο ελαχιστομέγιστος κανόνας γι' αυτήν την περίπτωση είναι ένας Μπαϊεσιανός κανόνας για την πρότερη τιμή π_L που μεγιστοποιεί το

$V(\pi_0)$ σε σχέση με το $\pi_0 \in [0,1]$. Ας σημειωθεί ότι γι' αυτό το πρότερο έχουμε ότι το $r(\pi_0, \delta_{\pi_1})$, είναι σταθερά στο π_0 , άρα

$$\max\{R_0(\delta_{\pi_L}), R_1(\delta_{\pi_L})\} = R_0(\delta_{\pi_L}) = R_1(\delta_{\pi_L})$$

Ένας κανόνας απόφασης με ίσους υπό συνθήκη κινδύνους ονομάζεται εξισωτικός κανόνας. Το γεγονός ότι το δ_{π_L} είναι ελαχιστομέγιστο προκύπτει από το σχήμα, καθώς αν $\pi_0 < \pi_L$, έχουμε $\max\{R_0(\delta_{\pi'_0}), R_1(\delta_{\pi'_0})\} = R_0(\delta_{\pi'_0}) > R_0(\delta_{\pi_L})$, και αν $\pi'_0 > \pi_L$, έχουμε ότι $\max\{R_0(\delta_{\pi''_0}), R_1(\delta_{\pi''_0})\} = R_1(\delta_{\pi''_0}) > R_1(\delta_{\pi_L})$, όπως απεικονίζεται.

Καθώς το π_L στο Σχ. Π.С.2 μεγιστοποιεί τον ελάχιστο Μπαϊεσιανό κίνδυνο, ονομάζεται το λιγότερο ευνοϊκό πρότερο. Επομένως για την περίπτωση αυτή ένας ελαχιστομέγιστος κανόνας απόφασης είναι ένας Μπαϊεσιανός κανόνας για το λιγότερο ευνοϊκό πρότερο. Σε αμφισβήτηση των παραπάνω θα πρέπει να δούμε ότι δεν λάβαμε υπόψη μας την πιθανότητα ότι το $\max_{0 \leq \pi_0 \leq 1} V(\pi_0)$ μπορεί να προκύψει στο $\pi_0 = 0$ ή στο $\pi_0 = 1$, ή ότι το $V(\pi_0)$ μπορεί να μην έχει εφαπτόμενη γραμμή σε κάθε σημείο (δηλ, μπορεί να μην είναι διαφορίσιμο παντού). Ωστόσο, ακόμη και σ' αυτές τις περιπτώσεις είναι πάντα αληθές ότι ο ελαχιστομέγιστος κανόνας είναι ένας Μπαϊεσιανός κανόνας για το λιγότερο ευνοϊκό πρότερο. Τα ακόλουθα αποτελέσματα αναπτύσσουν επίσημα τη γενική λύση στο πρόβλημα του ελαχιστομέγιστου ελέγχου υπόθεσης. Αρχίζουμε με την ακόλουθη πρόταση που συνοψίζει σημαντικά τις περιπτώσεις στις οποίες το $V(\pi_0)$ είναι όπως απεικονίζεται στα Σχ. Π.С.1 και Π.С.2 ή σ' αυτές στις οποίες το $\pi_L = 0$ ή $\pi_L = 1$.

Πρόταση Π.С.1 : Ο ελαχιστομέγιστος έλεγχος. Ας υποθέσουμε ότι το π_L είναι λύση στο $V(\pi_L) = \max_{0 \leq \pi_0 \leq 1} V(\pi_0)$. Ας υποθέσουμε περαιτέρω ότι $\pi_L = 0$ ή $\pi_L = 1$ ή $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$. Τότε το δ_{π_L} είναι ένας ελαχιστομέγιστος κανόνας

Απόδειξη: Πρώτα, ας εξετάσουμε την περίπτωση $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$. Τότε για κάθε πρότερο π_0 έχουμε

$$\max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta) = r(\pi_L, \delta_{\pi_L}) = r(\pi_0, \delta_{\pi_L}), \quad (\text{Π.С.4})$$

όπου η πρώτη ισότητα είναι εξ ορισμού του V και του π_L , και όπου η δεύτερη ισότητα έπεται από το γεγονός ότι το $r(\pi_0, \delta_{\pi_L})$ είναι σταθερό στο π_0 . Άρα έχουμε ότι

$$\begin{aligned} \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta) &= \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta_{\pi_0}) \\ &\geq \min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta). \end{aligned} \quad (\text{II.C.5})$$

Αλλά για κάθε δ , έχουμε πάντα ότι

$$\max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) \geq \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta), \quad (\text{II.C.6})$$

που συνεπάγεται ότι

$$\min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) \geq \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta). \quad (\text{II.C.7})$$

Συνδυάζοντας το (II.C.5) και το (II.C.7) έχουμε ότι

$$\min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) = \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta) \quad (\text{II.C.8})$$

και η αριστερή πλευρά της ισότητας του (II.C.4) συνεπάγεται ότι

$$r(\pi_L, \delta_{\pi_L}) = \min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta), \quad (\text{II.C.9})$$

που έπρεπε να αποδειχθεί.

Τώρα ας υποθέσουμε ότι $\pi_L = 0$. Στην περίπτωση αυτή πρέπει να αποδειχθεί ότι $\max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta_{\pi_L}) = R_1(\delta_{\pi_L}) = r(\pi_L, \delta_{\pi_L})$.

Το Σχήμα II.C.3 απεικονίζει μια χαρακτηριστική τέτοια περίπτωση. Αυτό το δεδομένο και το όρισμα των ισοτήτων του (II.C.6) μέσω του (II.C.9) συνεπάγονται ότι το δ_{π_L} είναι ένας ελαχιστομέγιστος κανόνας. Παρόμοιο επιχείρημα ισχύει και για την περίπτωση του $\pi_L = 1$. Αυτό ολοκληρώνει την απόδειξη.

Ας σημειωθεί ότι για κάθε $\pi_0 \in [0,1]$, έχουμε πάντα $r(\pi_0, \delta_{\pi_0}) \geq V(\pi_0)$ για όλα τα

$\pi_0 \in [0,1]$ και $r(\pi_0, \delta_{\pi_0}) = V(\pi_0)$. Επομένως, αφού το $r(\pi_0, \delta_{\pi_0})$ ως μια συνάρτηση του π_0 περιγράφει μια ευθεία γραμμή, θα πρέπει να είναι εφαπτόμενη στο V στο $\pi_0 = \pi'_0$, όπως απεικονίζεται για παράδειγμα στο Σχήμα II.C.2.

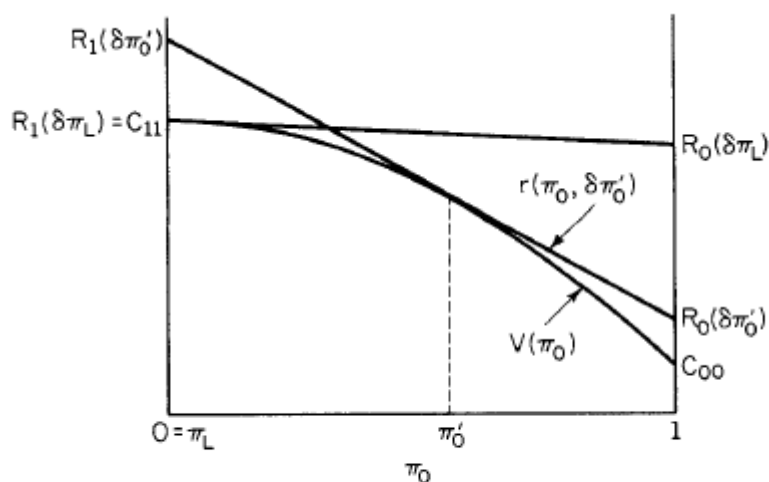


FIGURE II.C.3. Depiction of the minimax rule when V has an endpoint maximum.

Σχήμα II.C.3. Απεικόνιση του ελαχιστομέγιστου κανόνα όταν το V έχει ένα τελικό μέγιστο

Άρα, αν το V είναι διαφορίσιμο στο π_0 , θα πρέπει να έχουμε

$$V'(\pi'_0) = dr(\pi_0, \delta_{\pi'_0})/d\pi_0 = [R_0(\delta_{\pi'_0}) - R_1(\delta_{\pi'_0})],$$

όπου το V' συμβολίζει το παράγωγο του V . Αν το V έχει ένα εσωτερικό μέγιστο (δηλ., $\pi_L \in (0, 1)$), τότε το $V'(\pi_L)$ ισούται με μηδέν αν το V είναι διαφορίσιμο στο π_L . [Ας σημειωθεί ότι η κοιλότητα του V συνεπάγεται ότι το $V'(\pi_0)$ μπορεί να ισούται με μηδέν μόνο σε ένα μέγιστο. Άρα η κατάσταση $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$ ισχύει όποτε υπάρχει το $\pi_L \in (0, 1)$ και $V'(\pi_L)$ και επομένως η υπόθεση της πρότασης II.C.1 δεν είναι πολύ περιοριστική.

Έχουμε πλέον χαρακτηρίσει ελαχιστομέγιστους κανόνες για όλες τις περιπτώσεις, εκτός από αυτές στις οποίες το V έχει ένα εσωτερικό μέγιστο στο οποίο δεν είναι διαφορίσιμο. Για τέτοιες περιπτώσεις, ορίστε δύο κανόνες απόφασης από το $\delta_{\pi_L}^- = \lim_{\pi_0 \uparrow \pi_L} \delta_{\pi_0}$ και από το $\delta_{\pi_L}^+ = \lim_{\pi_0 \downarrow \pi_L} \delta_{\pi_0}$. Σημειώστε ότι το $\delta_{\pi_L}^-$ απαραίτητως έχει κρίσιμη περιοχή

$$\Gamma_1^- = \{y \in \Gamma | (1 - \pi_L)(C_{11} - C_{01})p_1(y) \leq \pi_L(C_{00} - C_{10})p_0(y)\}, \quad (\text{II.C.10})$$

Και το $\delta_{\pi_L}^+$ έχει κρίσιμη περιοχή (θυμηθείτε ότι υποθέτουμε $C_{11} < C_{01}$)

$$\Gamma_1^+ = \{y \in \Gamma | (1 - \pi_L)(C_{11} - C_{01})p_1(y) < \pi_L(C_{00} - C_{10})p_0(y)\}, \quad (\text{II.C.11})$$

ανεξάρτητα από το ποιο συγκεκριμένοι Μπαϊεσιανοί κανόνες δ_{π_0} χρησιμοποιούνται για τον προσδιορισμό τους. Για έναν αριθμό $q \in [0, 1]$, λάβετε υπόψη σας τον κανόνα δ_{π_L} που χρησιμοποιεί Γ_1^- με πιθανότητα q και χρησιμοποιεί Γ_1^+ με πιθανότητα $(1 - q)$. Δηλ. το $\bar{\delta}_{\pi_L}$ επιλέγει το H_1 αν $\bar{y} \in \Gamma_1^+$, επιλέγει H_0 αν $\bar{y} \in (\Gamma_1^-)^c$ και επιλέγει το H_1 με πιθανότητα q αν το y είναι στο όριο του Γ_1^- . Προσέξτε από το (II.B.6) ότι το όριο απόφασης είναι άσχετο από τη Μπαϊεσιανή διακινδύνευση, ώστε $r(\pi_L, \bar{\delta}_{\pi_L}) = V(\pi_L)$ και άρα το $\bar{\delta}_{\pi_L}$ είναι ένας Μπαϊεσιανός κανόνας για το π_L . Ωστόσο οι υπό συνθήκη κίνδυνοι όντως εξαρτώνται από το όριο και γίνονται

$$R_j(\bar{\delta}_{\pi_L}) = qR_j(\delta_{\pi_L}^-) + (1 - q)R_j(\delta_{\pi_L}^+), \quad (\text{II.C.12})$$

Έτσι η συνθήκη $R_0(\bar{\delta}_{\pi_L}) = R_1(\bar{\delta}_{\pi_L})$ επιτυγχάνεται επιλέγοντας

$$q = \frac{R_0(\delta_{\pi_L}^+) - R_1(\delta_{\pi_L}^+)}{R_0(\delta_{\pi_L}^+) - R_1(\delta_{\pi_L}^+) + R_1(\delta_{\pi_L}^-) - R_0(\delta_{\pi_L}^-)}. \quad (\text{II.C.13})$$

Επομένως, όπως στην Πρόταση II.C.1, το $\tilde{\delta}_{\pi_L}$ με το q να επιλέγεται από το (II.C.13) είναι ένας ελαχιστομέγιστος κανόνας.

Σημειώστε ότι, καθώς το V είναι κοίλο, θα πρέπει να έχει αριστερή και δεξιά παράγωγο στο π_L , τις οποίες συμβολίζουμε με $V'(\pi_L^-)$ και $V'(\pi_L^+)$ αντίστοιχα. Βλέπουμε ευθέως ότι

$$V'(\pi_L^+) = [R_0(\delta_{\pi_L}^+) - R_1(\delta_{\pi_L}^+)] \text{ και } V'(\pi_L^-) = [R_0(\tilde{\delta}_{\pi_L}^-) - R_1(\delta_{\pi_L}^-)], \text{ άρα το (II.C.13) γίνεται}$$

$$q = \frac{V'(\pi_L^+)}{V'(\pi_L^+) - V'(\pi_L^-)}. \quad (\text{II.C.14})$$

Ας σημειωθεί περαιτέρω ότι αν το πεδίο $\{y \in \Gamma | (1 - \pi_L)(C_{11} - C_{01})p_1(y) = \pi_L(C_{00} -$

$C_{10})p_0(y)\}$ προκύπτει με μηδέν πιθανότητα υπό το H_0 και H_1 , τότε η ανάγκη για τη χρήση του $\tilde{\delta}_{\pi_L}$ αντί οποιασδήποτε άλλης εκδοχής του δ_{π_L} , δεν θα προκύψει. Ο κανόνας απόφασης $\tilde{\delta}_{\pi_L}$ είναι ένα παράδειγμα τυχαιοποιημένου κανόνα απόφασης, μια έννοια που θα συζητηθεί περαιτέρω στην Ενότητα II.D.

Η δράση του κανόνα απόφασης $\tilde{\delta}_{\pi_L}$ είναι εμφανής στο παράδειγμα που απεικονίζεται στο Σχήμα II.C.4. . Ας σημειωθεί ότι οι γραμμές $r(\pi_0, \tilde{\delta}_{\pi_L}^-)$ and $r(\pi_0, \delta_{\pi_L}^+)$ διασταυρώνονται στο $\pi_0 = \pi_L$ και έχουν κλίσεις ίσες στο $V'(\pi_0^-)$ and $V'(\pi_0^+)$, αντίστοιχα. Μεταβάλλοντας την πιθανότητα q από 0 ως 1, μπορεί να ληφθεί οποιαδήποτε γραμμή μεταξύ αυτών των δύο γραμμών. Η συγκεκριμένη επιλογή του q που δίνεται από το (II.C.14) παράγει την οριζόντια γραμμή που βρίσκεται μεταξύ αυτών των δύο. Το ελαχιστομέγιστο του αντίστοιχου κανόνα απόφασης είναι προφανές από το σχήμα όπως εξετάστηκε σε συνδυασμό με το Σχ. II. C.2.

Για την απεικόνιση του σχεδίου των ελαχιστομέγιστων κανόνων απόφασης, μελετούμε τα δύο παραδείγματα που δίνονται στην Ενότητα II.B.

Παράδειγμα II.C.1 :Έλεγχος θέσης με Σφάλμα κατά Gauss.

Εξετάστε το πρόβλημα ελέγχου θέσης του Παραδείγματος II.B.2 με σφάλμα κατά Gauss και με ομοιόμορφα κόστη. Η συνάρτηση V έπεται από το II.B.5 και II.B.31 και δίνεται από

$$V(\pi_0) = \pi_0 \left(1 - \Phi \left(\frac{\tau' - \mu_0}{\sigma} \right) \right)$$

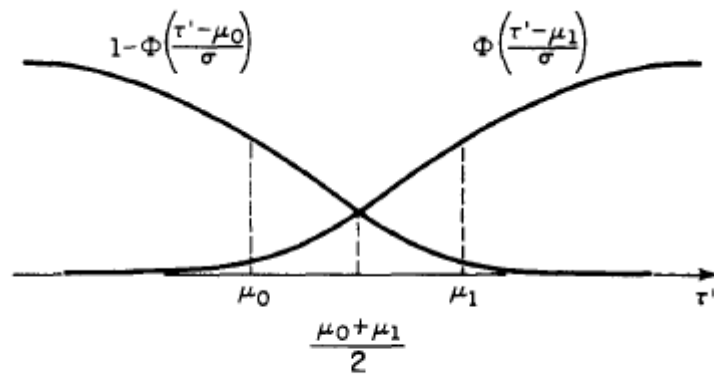


FIGURE II.C.5. Conditional risks for location testing with Gaussian error and uniform costs.

ΣΧΗΜΑ II.C.5. Υπό συνθήκη κίνδυνοι με σφάλμα κατά Gauss και ομοιόμορφα κόστη.

Από το (II.C. 17) έπεται ότι το λιγότερο ευνοϊκό πρότερο είναι $\pi_L = 1/2$, άρα ο ελάχιστος κίνδυνος είναι

$$V(1/2) = 1 - \Phi \left(\frac{\mu_1 - \mu_0}{2\sigma} \right).$$

Παράδειγμα II.C.2 : Το Δυαδικό Κανάλι

Ως ένα παράδειγμα, εξετάζουμε το δυαδικό κανάλι του Παραδείγματος II.B.1 με ομοιόμορφα κόστη. Η συνάρτηση ελάχιστου Μπαϊεσιανού κινδύνου για την περίπτωση αυτή δίνεται από

$$V(\pi_0) = \min\{(1 - \pi_0)\lambda_1, \pi_0(1 - \lambda_0)\} + \min\{(1 - \pi_0)(1 - \lambda_1), \pi_0\lambda_0\}. \quad (\text{II.C.19})$$

Η συνάρτηση αυτή μπορεί να ξαναγραφεί ως

$$V(\pi_0) = \begin{cases} \pi_0 & \text{if } 0 \leq \pi_0 \leq \underline{\pi} \\ \underline{\pi} + C\pi_0 & \text{if } \underline{\pi} < \pi_0 < \bar{\pi} \\ (1 - \pi_0) & \text{if } \bar{\pi} \leq \pi_0 \leq 1, \end{cases} \quad (\text{II.C.20})$$

όπου

$$\underline{\pi} = \min \left\{ \frac{\lambda_1}{1 - \lambda_0 + \lambda_1}, \frac{1 - \lambda_1}{1 - \lambda_1 + \lambda_0} \right\},$$

$$\bar{\pi} = \max \left\{ \frac{\lambda_1}{1 - \lambda_0 + \lambda_1}, \frac{1 - \lambda_1}{1 - \lambda_1 + \lambda_0} \right\},$$

και $C = (1 - \bar{\pi} - \underline{\pi})/(\bar{\pi} - \underline{\pi})$. Ας σημειωθεί ότι η $V(\pi_0)$ είναι τμηματικά γραμμική με αλλαγές κλίσης στο $\underline{\pi}$ και στο $\bar{\pi}$. Επομένως, αφού $V(0) = V(1) = 0$, το σύμβολο της κλίσης μεταξύ των $\underline{\pi}$ και $\bar{\pi}$ θα προσδιορίσει το π_L . Ειδικότερα, $\pi_L = \underline{\pi}$ αν

$C < 0$, $\pi_L = \bar{\pi}$ if $C > 0$, και π_L αν $C = 0$. Σε οποιαδήποτε περίπτωση, ο ελαχιστομέγιστος κίνδυνος είναι $V(\pi_L) = \max\{\underline{\pi}, 1 - \bar{\pi}\}$.

Καθώς το $V(\pi_0)$ δεν είναι διαφορίσιμο ούτε στο $\underline{\pi}$ ούτε στο $\bar{\pi}$, εδώ είναι απαραίτητο να λάβουμε υπόψη την περίπτωση $C > 0$ όπως απεικονίζεται στο Σχ. II. C.6. Τότε έχουμε $\pi_L = \bar{\pi}$. Ελέγχοντας έχουμε $V'(\pi_L^+) = -1$ και $V'(\pi_L^-) = C$; άρα η τυχαιοποιημένη σταθερά q δίνεται από

$$q = \frac{1}{1 + C} = \frac{\bar{\pi} - \underline{\pi}}{1 - 2\underline{\pi}}. \quad (\text{II.C.21})$$

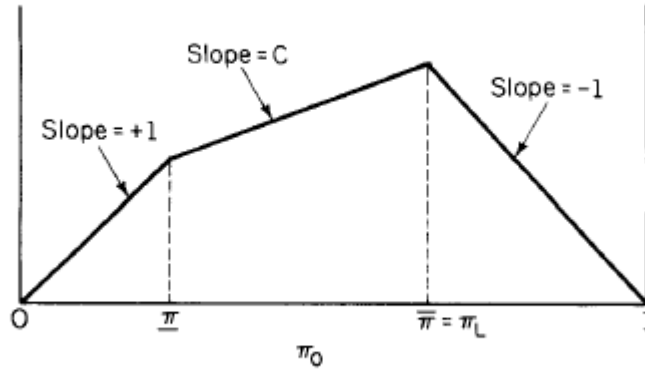


FIGURE II.C.6. $V(\pi_0)$ for the binary channel.

Σχήμα II.C. 6 $V(\pi_0)$ για το δυαδικό κανάλι.

Όπως φαίνεται από το Παράδειγμα II.B.1 αν $\pi_0 > \bar{\pi}$, τότε ο Μπαϊεσιανός κανόνας είναι $\delta_{\pi_0}(0) = \delta_{\pi_0}(1) = 0$. Επομένως το $\delta_{\pi_L}^+(0) = \delta_{\pi_L}^+(1) = 0$ και Γ_1^+ είναι το κενό σύνολο. Παρομοίως, για $\underline{\pi} < \pi_0 < \bar{\pi}$, έχουμε $\delta_{\pi_0}(y) = y$ if $\lambda_1 < 1 - \lambda_0$ και $\delta_{\pi_0}(y) = 1 - y$ if $\lambda_1 > 1 - \lambda_0$. Επομένως, για παράδειγμα, αν $\lambda_1 < 1 - \lambda_0$, έχουμε

$\delta_{\pi_L}^-(y) = y$ και ο ελαχιστομέγιστος κανόνας $\bar{\delta}_{\pi_L}$ επιλέγει το 0 αν $y = 0$ και επιλέγει το 1 με πιθανότητα q [και 0 με πιθανότητα $(1 - q)$] αν $y = 1$. Στην περίπτωση αυτή η ελαχιστομέγιστη διακινδύνευση είναι

$$V(\bar{\pi}_L) = 1 - \bar{\pi} = \frac{\lambda_0}{1 - \lambda_1 + \lambda_0}. \quad (\text{II.C.22})$$

Αν επιπλέον έχουμε ότι $\lambda_0 = \lambda_1 = \lambda$ (περίπτωση στην οποία $\lambda_1 < 1 - \lambda_0 < 1/2$), τότε $q = 1$ και η ελαχιστομέγιστη διακινδύνευση είναι

$$V(\pi_L) = \lambda,$$

που είναι ίδια με τη Μπαϊεσιανή διακινδύνευση για κάθε $\pi_0 \in [\lambda, 1 - \lambda]$. Επομένως, για την τελευταία αυτή περίπτωση, η αβεβαιότητα στα πρότερα δεν έχει κόστος σε απόδοση εφόσον $\lambda \leq \pi_0 \leq 1 - \lambda$.

Συνοψίζοντας την ενότητα αυτή, είδαμε ότι οι έλεγχοι βελτιστότητας μπορούν να σχεδιαστούν χωρίς την εικασία των γνωστών πρότερων, υιοθετώντας ένα ελαχιστομέγιστο κριτήριο σχεδιασμού. Η λύση είναι ένας Μπαϊεσιανός έλεγχος για το λιγότερο ευνοϊκό πρότερο, με την τυχαιοποίηση στο όριο της περιοχής απόφασης να είναι απαραίτητη για να δώσει το ελαχιστομέγιστο σε κάποια προβλήματα. Και πάλι, βλέπουμε ότι ο λόγος πιθανοφάνειας εμφανίζεται ως ο βέλτιστος κανόνας απόφασης. Στην ακόλουθη ενότητα, μελετούμε μία ακόμη διατύπωση του προβλήματος δυαδικού ελέγχου υπόθεσης.

II.D Έλεγχος Υπόθεσης Neyman – Pearson.

Στη Μπαϊεσιανή διατύπωση της Ενότητας II.B. η βελτιστότητα στον έλεγχο (II.B.1) προσδιορίστηκε με όρους ελαχιστοποίησης του συνολικού αναμενόμενου κόστους, οριζόμενη ως μέση διακινδύνευση. Παρομοίως, στην ελαχιστομέγιστη διατύπωση της Ενότητας II.C., τα πρότερα δεν θεωρούνταν γνωστά και η βελτιστότητα οριζόταν με όρους ελαχιστοποίησης του μέγιστου των υπό συνθήκη αναμενόμενων κοστών μεταξύ των δύο υποθέσεων. Σε πολλά προβλήματα πρακτικού ενδιαφέροντος, η επιβολή μιας συγκεκριμένης δομής κόστους στις αποφάσεις που γίνονται κατά τον έλεγχο (II.B.1) δεν είναι δυνατή ή επιθυμητή. Σε τέτοιες περιπτώσεις συχνά τίθεται ένα εναλλακτικό κριτήριο σχεδιασμού, γνωστό ως κριτήριο Neyman – Pearson. Στην ενότητα αυτή περιγράφουμε αυτή την εναλλακτική διατύπωση.

Κατά τον έλεγχο H_0 ως προς H_1 στο (II.B.1) υπάρχουν δύο τύποι σφάλματος που μπορούν να γίνουν. Το H_0 μπορεί εσφαλμένα να απορριφθεί ή το H_1 μπορεί εσφαλμένα να απορριφθεί. Ο πρώτος από αυτούς τους δύο τύπους λαθών ονομάζεται σφάλμα Τύπου I ή ψευδής συναγερμός. Ο δεύτερος τύπος ονομάζεται σφάλμα τύπου II ή αστοχία. Οι όροι «ψευδής συναγερμός» και «αστοχία» προέρχονται από προβλήματα ραντάρ στα οποία τα H_0 και H_1 συνήθως αντιπροσωπεύουν την απουσία και παρουσία ενός στόχου, αντίστοιχα. Η σωστή αποδοχή του H_1 παρομοίως καλείται ανίχνευση. Για έναν κανόνα απόφασης δ , η πιθανότητα σφάλματος Τύπου I είναι γνωστή ως πιθανότητα ψευδούς συναγερμού (ή ρυθμός ψευδούς συναγερμού) του δ , και συμβολίζεται με $P_F(\delta)$. Παρομοίως, η πιθανότητα σφάλματος Τύπου II καλείται πιθανότητα αστοχίας, $P_M(\delta)$. Ωστόσο, όταν μελετούμε την τελευταία ποσότητα, συνήθως μιλάμε για πιθανότητα ανίχνευσης, $P_D(\delta) = 1 - P_M(\delta)$ που καλείται επίσης η ισχύς του δ .

Προφανώς, ο σχεδιασμός ενός ελέγχου του H_0 ως προς το H_1 περιλαμβάνει μια ανταλλαγή μεταξύ των πιθανοτήτων των δύο τύπων σφαλμάτων, καθώς η μία μπορεί να γίνει αυθαίρετα μικρή εις βάρος της άλλης. Το μπαϊεσιανό ή το ελαχιστομέγιστο κριτήριο είναι δύο τρόποι ανταλλαγής αυτών. Το κριτήριο Neyman – Pearson για την πραγματοποίηση αυτής της ανταλλαγής είναι να τεθεί ένα φράγμα στην πιθανότητα ψευδούς συναγερμού και κατόπιν να ελαχιστοποιηθεί η πιθανότητα αστοχίας στο πλαίσιο αυτού του περιορισμού. Δηλ. το κριτήριο σχεδιασμού Neyman – Pearson είναι:

$$\max_{\delta} P_D(\delta) \text{ subject to } P_F(\delta) \leq \alpha, \quad (\text{II.D.1})$$

όπου το α είναι το προαναφερθέν φράγμα, το οποίο είναι γνωστό ως το επίπεδο ή επίπεδο σημαντικότητας του ελέγχου. Επομένως, ο στόχος του σχεδιασμού Neyman – Pearson είναι να βρει τον πιο ισχυρό έλεγχο α - επιπέδου του H_0 ως προς το H_1 . Ας σημειωθεί ότι, σε αντίθεση με το Μπαϊεσιανό και το ελαχιστομέγιστο κριτήριο, το κριτήριο Neyman – Pearson αναγνωρίζει τη βασική ασυμμετρία της σημασίας των δύο υποθέσεων.

Προκειμένου να δοθεί η γενική λύση στο πρόβλημα του Neyman – Pearson (II.D.1) είναι απαραίτητο να λάβουμε υπόψη τυχαιοποιημένους ελέγχους, παρόμοιους με τον έλεγχο $\tilde{\delta}_{\pi_L}$, όπως ορίζεται στη λύση του ελαχιστομέγιστου προβλήματος στην Ενότητα II.C.

Είναι βολικό να προσδιορισθεί ένα τυχαιοποιημένος κανόνας απόφασης $\tilde{\delta}$ για H_0 ως προς H_1 ως μια συνάρτηση που απεικονίζει το Γ στο διάστημα $[0,1]$ με την ερμηνεία ότι για $y \in \Gamma$, το $\tilde{\delta}(y)$ είναι η υπό συνθήκη πιθανότητα με την οποία δεχόμαστε το H_1 με δεδομένο ότι παρατηρούμε ότι $Y = y$.

Για παράδειγμα, ο τυχαιοποιημένος ελαχιστομέγιστος κανόνας $\tilde{\delta}_{\pi_L}$ που είδαμε στην Ενότητα II.C. μπορεί να γραφεί ως

$$\tilde{\delta}_{\pi_L}(y) = \begin{cases} 1 & \text{if } L(y) > \tau_L \\ q & \text{if } L(y) = \tau_L \\ 0 & \text{if } L(y) < \tau_L, \end{cases} \quad (\text{II.D.2})$$

όπου το τ_L είναι το κατώφλι που αντιστοιχεί στο λιγότερο ευνοϊκό πρότερο π_L . Με αυτόν τον ορισμό βλέπουμε ότι ένας μη-τυχαιοποιημένος κανόνας απόφασης είναι μια ειδική περίπτωση ενός τυχαιοποιημένου κανόνα απόφασης. Ειδικότερα, ένας μη τυχαιοποιημένος κανόνας δ αντιστοιχεί στον τυχαιοποιημένο κανόνα $\tilde{\delta}(y) = \delta(y)$. Η διαφορά μεταξύ των δύο είναι ότι η αξία του δ είναι ο δείκτης της αποδεχόμενης υπόθεσης και η αξία του $\tilde{\delta}$ είναι η πιθανότητα με την οποία αποδεχόμαστε το H_1 . Αυτά συμπίπτουν εφόσον το $\tilde{\delta}$ λαμβάνει μόνο τις δύο τιμές 0 και 1.

Η πιθανότητα ψευδούς συναγερού ενός κανόνα απόφασης είναι η πιθανότητα με την οποία δέχεται το H_1 δεδομένου ότι το H_0 είναι αληθές. Για ένα τυχαιοποιημένο κανόνα $\tilde{\delta}$ το μέγεθος αυτό δίνεται από

$$P_F(\tilde{\delta}) = E_0\{\tilde{\delta}(Y)\} = \int_{\Gamma} \tilde{\delta}(y)p_0(y)\mu(dy), \quad (\text{II.D.3})$$

όπου το $E_0\{\cdot\}$ συμβολίζει την προσδοκία στην υπόθεση H_0 . Η ισότητα (II.D.3) ακολουθεί γιατί η πιθανότητα αποδοχής του H_1 με δεδομένο ότι το H_0 είναι αληθές, είναι απλώς η πιθανότητα αποδοχής του H_1 δεδομένου του Y [δηλ. $\delta(Y)$] κατά μέσο όρο στην κατανομή του Y ως προς το H_0 . Παρομοίως, η πιθανότητα ανίχνευσης ενός τυχαιοποιημένου κανόνα $\tilde{\delta}$ δίνεται από

$$P_D(\tilde{\delta}) = E_1\{\tilde{\delta}(Y)\} = \int_{\Gamma} \tilde{\delta}(y)p_1(y)\mu(dy). \quad (\text{II.D.4})$$

Η γενική λύση στο πρόβλημα σχεδιασμού Neyman – Pearson μπορεί να συνοψιστεί στο ακόλουθο αποτέλεσμα

Πρόταση II.D.1.: Το Λήμμα Neyman – Pearson

Ας εξετάσουμε το ζεύγος υποθέσεων του (II.B.I) στο οποίο το P_j έχει πυκνότητα p_j για $j = 0$ και $j = 1$ και ας υποθέσουμε ότι $\alpha > 0$. Τότε οι ακόλουθες διατυπώσεις είναι αληθείς.

(i) (Βελτιστότητα). Έστω ότι $\tilde{\delta}$ είναι κάθε κανόνας απόφασης που ικανοποιεί το $P_F(\tilde{\delta}) \leq \alpha$ και έστω ότι $\tilde{\delta}'$ είναι κάθε κανόνας απόφασης της μορφής

$$\tilde{\delta}'(y) = \begin{cases} 1 & \text{if } p_1(y) > \eta p_0(y) \\ \gamma(y) & \text{if } p_1(y) = \eta p_0(y) \\ 0 & \text{if } p_1(y) < \eta p_0(y), \end{cases} \quad (\text{II.D.5})$$

όπου $\eta \geq 0$ και $0 \leq \gamma(y) \leq 1$ είναι τέτοια ώστε $P_F(\tilde{\delta}') = \alpha$. Τότε $P_D(\tilde{\delta}') \geq P_D(\tilde{\delta})$. Άρα, κάθε κανόνας απόφασης μεγέθους- α της μορφής (II.D.5) είναι κανόνας Neyman – Pearson.

(ii) (Υπαρξη) Για κάθε $\alpha \in (0, 1)$, υπάρχει ένας κανόνας απόφασης, $\tilde{\delta}_{NP}$ της μορφής του (II.D.5) με $\gamma(y) = \gamma_0$ (μια σταθερά) για τον οποίο $P_F(\tilde{\delta}_{NP}) = \alpha$

(iii) (Μοναδικότητα) Ας υποθέσουμε ότι $\tilde{\delta}''$ είναι οποιοσδήποτε α -επιπέδου κανόνας απόφασης Neyman Pearson για H_0 ως προς H_1 . Τότε το $\tilde{\delta}''$ θα πρέπει να είναι στη μορφή του (II.D.5) εκτός ενδεχομένως από ένα υποσύνολο του Γ που έχει πιθανότητα μηδέν ως προς το H_0 και H_1 .

Απόδειξη: (i) Υποθέτουμε ότι $\bar{\delta}$ και $\bar{\delta}'$ είναι όπως προσδιορίζονται παραπάνω. Σημειώστε ότι λόγω του τρόπου που είναι προσδιορισμένο το $\bar{\delta}'$, έχουμε πάντα $[\bar{\delta}'(y) - \bar{\delta}(y)][p_1(y) - \eta p_0(y)] \geq 0$ για κάθε $y \in \Gamma$. Επομένως

$$\int_{\Gamma} [\bar{\delta}'(y) - \bar{\delta}(y)][p_1(y) - \eta p_0(y)] \mu(dy) \geq 0. \quad (\text{II.D.6})$$

Αναπτύσσοντας τους όρους στο (II.D.6) και αναδιατάσσοντας έχουμε

$$\int_{\Gamma} \bar{\delta}' p_1 d\mu - \int_{\Gamma} \bar{\delta} p_1 d\mu \geq \eta \left[\int_{\Gamma} \bar{\delta}' p_0 d\mu - \int_{\Gamma} \bar{\delta} p_0 d\mu \right]. \quad (\text{II.D.7})$$

Εφαρμόζοντας (II.D.3) και (II.D.4), το (II.D.7) γίνεται

$$P_D(\bar{\delta}') - P_D(\bar{\delta}) \geq \eta [P_F(\bar{\delta}') - P_F(\bar{\delta})] = \eta [\alpha - P_F(\bar{\delta})] \geq 0, \quad (\text{II.D.8})$$

Όπου έχουμε χρησιμοποιήσει το δεδομένο ότι $P_F(\bar{\delta}) \leq \alpha$. Επομένως $P_D(\bar{\delta}') \geq P_D(\bar{\delta})$, το οποίο θα πρέπει να αποδειχθεί.

(ii) Έστω ότι η_0 είναι ο μικρότερος αριθμός τέτοιος ώστε

$$P_0(p_1(Y) > \eta_0 p_0(Y)) \leq \alpha. \quad (\text{II.D.9})$$

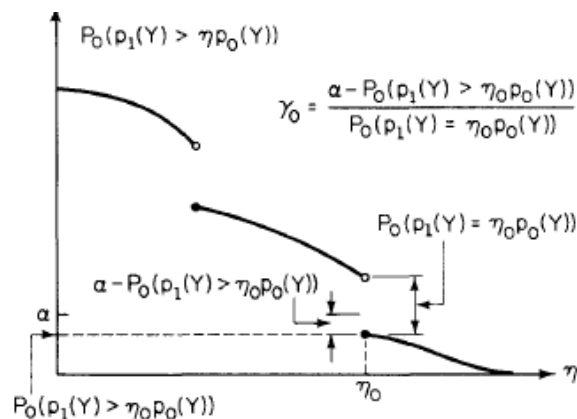
Τότε αν $P_0(p_1(Y) > \eta_0 p_0(Y)) < \alpha$, επιλέξτε

$$\gamma_0 = \frac{\alpha - P_0(p_1(Y) > \eta_0 p_0(Y))}{P_0(p_1(Y) = \eta_0 p_0(Y))}; \quad (\text{II.D.10})$$

Διαφορετικά, επιλέξτε γ_0 αυθαίρετα. Οι σχέσεις αυτές απεικονίζονται στο Σχ. II.D.1. Τότε, προσδιορίζοντας το $\bar{\delta}_{NP}$ ως κανόνα απόφασης του (II.D.5) με $\eta = \eta_0$ και $\gamma(y) = \gamma_0$, έχουμε

$$P_F(\bar{\delta}_{NP}) = E_0\{\bar{\delta}_{NP}(Y)\} = P_0(p_1(Y) > \eta_0 p_0(Y)) + \gamma_0 P_0(p_1(Y) = \eta_0 p_0(Y)) = \alpha. \quad (\text{II.D.11})$$

Επομένως έχουμε επιλέξει έναν κανόνα απόφασης της μορφής του (II.D.5) με $\gamma(y)$ σταθερά και πιθανότητα ψευδούς συναγερού α .



Σχήμα II.D.1 Κατώφλι και τυχαιοποίηση για έλεγχο Neyman-Pearson α -επιπέδου.

(iii) Ας υποθέσουμε ότι το $\bar{\delta}'$ είναι ένας κανόνας α -επιπέδου Neyman – Pearson της μορφής του (II.D.5) και έστω ότι $\bar{\delta}''$ είναι οποιοσδήποτε άλλος κανόνας α -επιπέδου Neyman – Pearson. Τότε το $P_D(\bar{\delta}'')$ πρέπει να είναι ίσο με το $P_D(\bar{\delta}')$, έτσι από το (II.D.8) έχουμε

$0 \geq \alpha - P_F(\bar{\delta}''') \geq 0$, or $\bar{P}_F(\bar{\delta}''') = \alpha$. Επομένως το $\bar{\delta}'''$ πρέπει να είναι μεγέθους α . Χρησιμοποιώντας τα δεδομένα ότι $P_D(\bar{\delta}''') = P_D(\bar{\delta}')$ και $P_F(\bar{\delta}''') = P_F(\bar{\delta}')$, και δουλεύοντας προς τα πίσω από το (II.D.8) στο (II.D.6) συμπεραίνουμε ότι

$$\int_{\Gamma} [\bar{\delta}'(y) - \bar{\delta}''(y)][p_1(y) - \eta p_0(y)] \mu(dy) = 0. \quad (\text{II.D.12})$$

Αφού η ολοκληρωτέα συνάρτηση είναι μη αρνητική (όπως ειπώθηκε παραπάνω), το (II.D.12) συνεπάγεται ότι είναι μηδέν εκτός πιθανόν σ' ένα σύνολο μηδενικής πιθανότητας ως προς το H_0 και H_1 . Επομένως τα $\bar{\delta}'$ και $\bar{\delta}''$ διαφέρουν μόνο στο σύνολο

$\{y \in \Gamma | p_1(y) = \eta p_0(y)\}$, που συνεπάγεται ότι $\bar{\delta}''$ είναι επίσης στη μορφή του (II.D.5), διαφέροντας πιθανόν από το $\bar{\delta}'$ μόνο στη συνάρτηση $\gamma(y)$. Αυτό ολοκληρώνει την απόδειξη της πρότασης.

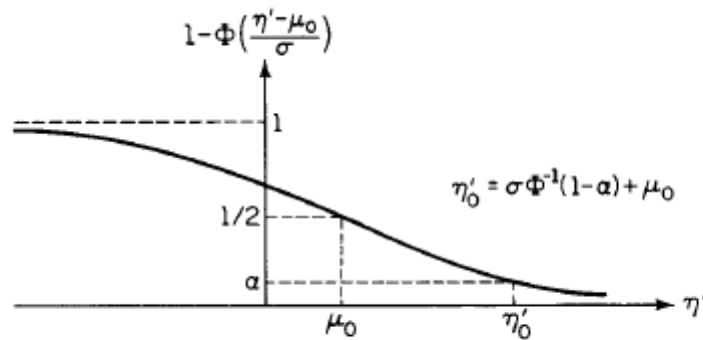


FIGURE II.D.2. Illustration of threshold (η'_0) for Neyman-Pearson testing of location with Gaussian error.

Σχήμα II.D.2 Απεικόνιση του κατωφλιού (η_0) για έλεγχο θέσης Neyman – Pearson με σφάλμα κατά Gauss.

Το παραπάνω αποτέλεσμα υποδεικνύει και πάλι τη βελτιστότητα του ελέγχου λόγου πιθανοφάνειας. Ο έλεγχος Neyman – Pearson για ένα δεδομένο ζεύγος υποθέσεων διαφέρει από τον Μπαϊεσιανό και τον ελαχιστομέγιστο έλεγχο μόνο στην επιλογή του κατωφλιού και

της τυχαιοποίησης (Ας σημειωθεί ότι, για $\alpha = 0$, ο έλεγχος Neyman-Pearson δίνεται άμεσα από το μη τυχαιοποιημένο έλεγχο με κρίσιμο χωρίο $\Gamma_1 = \{y \in \Gamma | p_0(y) = 0\}$.) Ο σχεδιασμός των ελέγχων Neyman – Pearson απεικονίζεται στα ακόλουθα δύο παραδείγματα.

Παράδειγμα II.D.1: Έλεγχος θέσης με σφάλμα κατά Gauss

Ας εξετάσουμε πρώτα το πρόβλημα ελέγχου θέσης με σφάλματα κατά Gauss όπως στο παράδειγμα II.B.2. Εδώ έχουμε

$$\begin{aligned} P_0(p_1(Y) > \eta p_0(Y)) &= P_0(L(Y) > \eta) = P_0(Y > \eta') \\ &= 1 - \Phi\left(\frac{\eta' - \mu_0}{\sigma}\right), \end{aligned} \quad (\text{II.D.13})$$

όπου $\eta' = \sigma^2 \log(\eta) / (\mu_1 - \mu_0) + (\mu_0 + \mu_1) / 2$. Η καμπύλη αυτή απεικονίζεται στο Σχ. II.D.2 ως συνάρτηση του η' . Ας σημειωθεί ότι οποιαδήποτε τιμή του α μπορεί να επιτευχθεί ακριβώς επιλέγοντας

$$\eta'_0 = \sigma \Phi^{-1}(1 - \alpha) + \mu_0 \quad (\text{II.D.14})$$

όπου Φ^{-1} είναι η αντίστροφη συνάρτηση του Φ . Αφού $P(Y = \eta'_0) = 0$, η τυχαιοποίηση μπορεί να επιλεγεί αυθαίρετα, ας πούμε $\gamma_0 = 1$. Ένας έλεγχος α -επιπέδου Neyman – Pearson για την περίπτωση αυτή δίνεται από

$$\bar{\delta}_{NP}(y) = \begin{cases} 1 & \text{if } y \geq \eta'_0 \\ 0 & \text{if } y < \eta'_0, \end{cases} \quad (\text{II.D.15})$$

όπου το η'_0 είναι από το (II.D.14).

Η ανίχνευση πιθανότητας του δ_{NP} δίνεται από

$$\begin{aligned} P_D(\bar{\delta}_{NP}) = E_1\{\bar{\delta}_{NP}(Y)\} &= P_1(Y \geq \eta'_0) \\ &= 1 - \Phi\left(\frac{\eta'_0 - \mu_1}{\sigma}\right) \\ &= 1 - \Phi(\Phi^{-1}(1 - \alpha) - d), \end{aligned} \quad (\text{II.D.16})$$

όπου $d = (\mu_1 - \mu_0)/\sigma$ είναι ο λόγος σήμα προς θόρυβο που ορίζεται στο Παράδειγμα II.B.2

Για ένα σταθερό α , το (II.D.16) δίνει την πιθανότητα ανίχνευσης ως συνάρτηση του d για τον έλεγχο του (II.D.16). Η σχέση αυτή είναι ενίοτε γνωστή ως *δυναμοσυνάρτηση* του ελέγχου. Ένα διάγραμμα της σχέσης αυτής απεικονίζεται στο Σχήμα II.D.3.

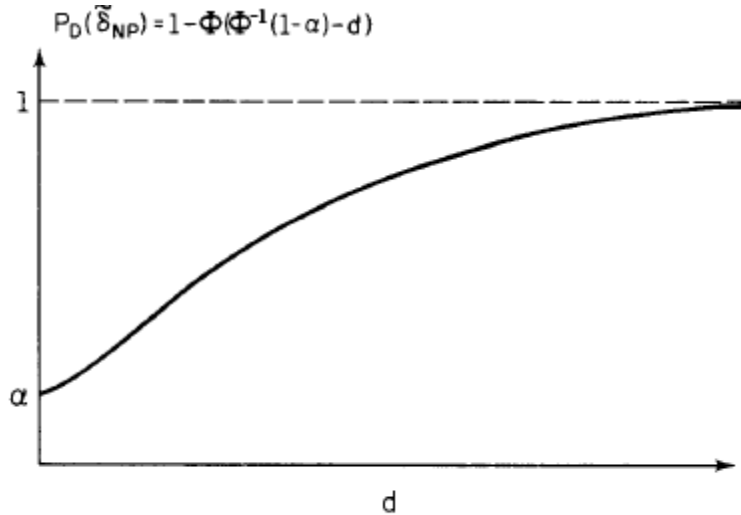


FIGURE II.D.3. Power function for Neyman-Pearson testing of location with Gaussian error [$d = (\mu_1 - \mu_0)/\sigma$].

Σχήμα II.D.3. Δυναμοσυνάρτηση για Neyman – Pearson έλεγχο θέσης με σφάλμα κατά Gauss [$d = (\mu_1 - \mu_0)/\sigma$].

Η ισότητα του (II.D.16) επίσης δίνει την πιθανότητα ανίχνευσης ως μια συνάρτηση της πιθανότητας ψευδούς συναγερμού για ένα σταθερό d . Και πάλι δανειζόμενο από την ορολογία των ραντάρ, ένα παραμετρικό διάγραμμα της σχέσης αυτής λέγεται *λειτουργικός χαρακτηριστικός δέκτης* (ROCs). Οι λειτουργικοί χαρακτηριστικοί δέκτες (ROCs) για τον έλεγχο του (II.D.15) φαίνονται στο Σχήμα II.D.4. Τα σχήματα II.D.2 ως II.D.4 εφαρμόζονται σε μεγάλη ομάδα προβλημάτων που αφορούν σε ανίχνευση σήματος σε θόρυβο κατά Gauss και θα εξεταστούν σε ακόλουθα κεφάλαια.

Παράδειγμα II.D.2 : Το Δυαδικό Κανάλι

Για να απεικονίσετε ένα πρόβλημα Neyman – Pearson στο οποίο η τυχαιοποίηση είναι απαραίτητη και πάλι λάβετε υπόψη σας το δυαδικό κανάλι του Παραδείγματος II.B.1. Ο λόγος πιθανοφάνειας για το πρόβλημα αυτό δίνεται από το (II.B.20). Για να βρούμε το κατώφλι για την επίτευξη ενός ελέγχου α -επιπέδου Neyman – Pearson, θα πρέπει να λάβουμε υπόψη το $P_0(L(Y) > \eta)$. Για λόγους απλοποίησης, υποθέτουμε ότι $\lambda_0 + \lambda_1 < 1$, περίπτωση στην οποία $\lambda_1/(1 - \lambda_0) < (1 - \lambda_1)/\lambda_0$. Τότε έχουμε

$$P_0(L(Y) > \eta) = \begin{cases} 1 & \text{if } \eta < \frac{\lambda_1}{(1-\lambda_0)} \\ \lambda_0 & \text{if } \frac{\lambda_1}{(1-\lambda_0)} \leq \eta < \frac{(1-\lambda_1)}{\lambda_0} \\ 0 & \text{if } \eta \geq \frac{(1-\lambda_1)}{\lambda_0}. \end{cases} \quad (\text{II.D.17})$$

Η συνάρτηση αυτή απεικονίζεται στο Σχ. II.D.5. Ελέγχοντας βλέπουμε ότι

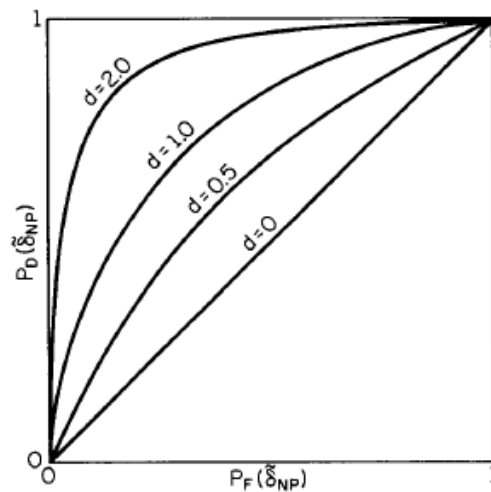


FIGURE II.D.4. Receiver operating characteristics (ROCs) for Neyman-Pearson location testing with Gaussian error [$d = (\mu_1 - \mu_0)/\sigma$].

Σχήμα II.D.4 Λειτουργικά Χαρακτηριστικά Δέκτη (ROCs) για έλεγχο θέσης Neyman – Pearson με σφάλμα κατά Gauss [$d = (\mu_1 - \mu_0)/\sigma$].

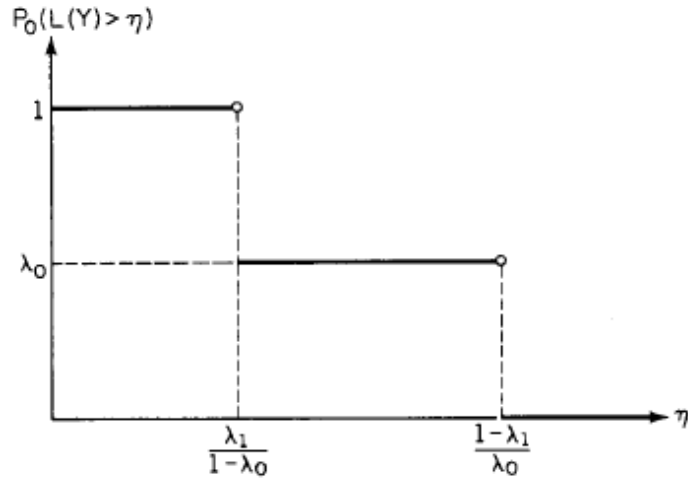


FIGURE II.D.5. Curve for threshold and randomization selection for a binary channel.

Σχήμα II.D.5. Καμπύλη για επιλογή κατώφλιού και τυχαιοποίησης για ένα δυαδικό κανάλι.

το επιθυμητό κατώφλι για έλεγχο α -επιπέδου Neyman – Pearson δίνεται από

$$\eta_0 = \begin{cases} \frac{(1-\lambda_1)}{\lambda_0} & \text{if } 0 \leq \alpha < \lambda_0 \\ \frac{\lambda_1}{(1-\lambda_0)} & \text{if } \lambda_0 \leq \alpha < 1 \\ 0 & \text{if } \alpha = 1. \end{cases} \quad (\text{II.D.18})$$

Επίσης, ελέγχοντας, βλέπουμε ότι η σταθερά τυχαιοποίησης πρέπει να δίνεται από

$$\gamma_0 = \begin{cases} \frac{\alpha}{\lambda_0} & \text{if } 0 \leq \alpha < \lambda_0 \\ \frac{(\alpha-\lambda_0)}{(1-\lambda_0)} & \text{if } \lambda_0 \leq \alpha < 1 \\ \text{arbitrary} & \text{if } \alpha = 1. \end{cases} \quad (\text{II.D.19})$$

(αυθαίρετα αν)

Χρησιμοποιώντας τα (II.B.20), (II.D.18) και (II.D.19), ο Neyman – Pearson έλεγχος που προκύπτει φαίνεται να είναι

$$\tilde{\delta}_{NP}(y) = \begin{cases} \frac{\alpha}{\lambda_0} & \text{if } y = 1 \\ 0 & \text{if } y = 0, \end{cases} \quad (\text{II.D.20})$$

για $0 \leq \alpha < \lambda_0$, και

$$\tilde{\delta}_{NP}(y) = \begin{cases} 1 & \text{if } y = 1 \\ \frac{(\alpha - \lambda_0)}{(1 - \lambda_0)} & \text{if } y = 0, \end{cases} \quad (\text{II.D.21})$$

για $\lambda_0 \leq \alpha \leq 1$.

Η πιθανότητα ανίχνευσης για τον έλεγχο Neyman – Pearson δίνεται από

$P_D(\tilde{\delta}_{NP}) = P_1(L(Y) > \eta_0) + \gamma_0 P_1(L(Y) = \eta_0)$, το οποίο φαίνεται εδώ άμεσα να είναι

$$P_D(\tilde{\delta}_{NP}) = \begin{cases} \alpha \frac{(1 - \lambda_1)}{\lambda_0} & \text{if } 0 \leq \alpha < \lambda_0 \\ (1 - \lambda_1) + \lambda_1 \frac{(\alpha - \lambda_0)}{(1 - \lambda_0)} & \text{if } \lambda_0 \leq \alpha \leq 1. \end{cases} \quad (\text{II.D.22})$$

Ας σημειωθεί ότι τα ROCs (P_D ως προς P_F) είναι τμηματικά γραμμικά με μια αλλαγή σε κλίση στο $P_F = \lambda_0$. Η συμπεριφορά αυτή απεικονίζεται στο Σχ. II.D.6 για την περίπτωση συμμετρικού καναλιού, $\lambda_0 = \lambda_1 = \lambda \leq 1/2$.

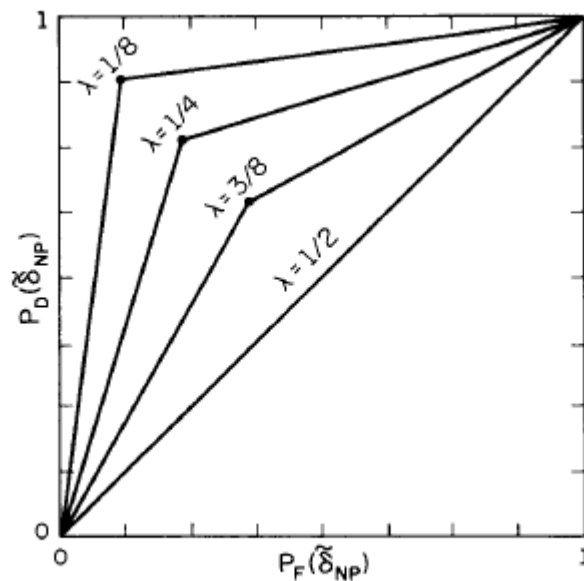


FIGURE II.D.6. ROCs for a binary symmetric channel.

Σχήμα II.D.6 ROCs για ένα δυαδικό συμμετρικό κανάλι

II.E Έλεγχος Σύνθετης Υπόθεσης

Τα προβλήματα ελέγχου υπόθεσης που εξετάστηκαν στις προηγούμενες ενότητες είναι γνωστά ως προβλήματα ελέγχου απλής υπόθεσης γιατί κάθε μία από τις δύο υποθέσεις στο (II.B.1) αντιστοιχεί μόνο σε μία κατανομή για την παρατήρηση. Σε πολλά προβλήματα ελέγχου υπόθεσης, ωστόσο, υπάρχουν πολλές πιθανές κατανομές που μπορεί να προκύψουν για κάθε μία από τις υποθέσεις. Οι υποθέσεις αυτού του τύπου είναι γνωστές ως σύνθετες

υποθέσεις. Ένα παράδειγμα του πού μπορεί να προκύψουν τέτοιου τύπου προβλήματα είναι στην ανίχνευση ραντάρ, στην οποία το επιστρεφόμενο σήμα, αν είναι παρόν, έχει άγνωστες παραμέτρους, όπως ο ακριβής χρόνος άφιξης του (που συνδέεται με τη θέση) και η μεταβολή Doppler (που συνδέεται με την ταχύτητα). Επομένως, η υπόθεση “παρόντος στόχου” είναι σύνθετη. Στην ενότητα αυτή θα εξετάσουμε σύντομα το σχεδιασμό ελέγχων υπόθεσης για σύνθετα προβλήματα.

Για να μοντελοποιήσουμε (ή προτυποποιήσουμε) τον πιο γενικό τύπο του προβλήματος ελέγχου σύνθετης υπόθεσης, μπορούμε να εξετάσουμε μια οικογένεια κατανομών πιθανότητας στο Γ με δείκτες από μια παράμετρο θ που λαμβάνει τιμές σ’ ένα σύνολο παραμέτρου Λ . Αυτό σημαίνει ότι, έχουμε μια οικογένεια $\{P_\theta; \theta \in \Lambda\}$, όπου το P_θ είναι η κατανομή πιθανότητας της παρατήρησης, δοθέντος ότι το θ είναι η πραγματική τιμή της παραμέτρου. Οι τιμές της παραμέτρου στο Λ αντιπροσωπεύουν το σύνολο όλων των πιθανών φυσικών καταστάσεων. Για το ζεύγος της απλής υπόθεσης του (II.B.1), έχουμε την ειδική περίπτωση, στην οποία $\Lambda = \{0,1\}$. Γενικότερα, μπορεί να έχουμε ένα παραμετρικό χώρο, ο οποίος είναι η ένωση δύο ασύνδετων συνόλων παραμέτρων Λ_0 και Λ_1 που αναπαριστούν τις εκτάσεις της παραμέτρου ως προς τις δύο υποθέσεις.

Σε μια Μπαϊεσιανή διατύπωση του προβλήματος ελέγχου σύνθετης υπόθεσης, υποθέτουμε ότι η παράμετρος είναι ένα τυχαίο μέγεθος, Θ , που λαμβάνει τιμές στο Λ . Στην περίπτωση αυτή το P_θ ερμηνεύεται ως η υπό συνθήκη κατανομή του Y , δοθέντος ότι $\Theta = \theta$. Θέλουμε να πάρουμε μία δυαδική απόφαση για το Θ και για λόγους απλοποίησης θα λάβουμε υπόψη μας μόνο μη τυχαιοποιημένους κανόνες απόφασης. Για να επιλέξουμε ένα βέλτιστο κανόνα απόφασης θα πρέπει πρώτα να αποδώσουμε κόστη στις αποφάσεις μας μέσω μιας συνάρτησης κόστους $C[i, \theta]$ όπου το $C[i, \theta]$ είναι το κόστος επιλογής της απόφασης i όταν

$Y \sim P_\theta$, για $i \in \{0,1\}$ και $\theta \in \Lambda$. Για παράδειγμα, στο πρόβλημα ανίχνευσης ραντάρ, στην οποία η παράμετρος θ μπορεί να συνδέεται με τη θέση του στόχου και την ταχύτητα, η δομή κόστους μπορεί να αποδώσει υψηλότερα κόστη στην αστοχία κοντινών ή γρήγορα κινούμενων στόχο από αυτά για την αστοχία βραδύτερων ή πιο μακρινών στόχων. Για λόγους απλοποίησης, υποθέτουμε ότι το C είναι μη αρνητικό και ανηγμένο.

Για ένα κανόνα απόφασης δ , μπορούμε να ορίσουμε υπό συνθήκη κινδύνους ανάλογους με αυτούς για την απλή περίπτωση μέσω

$$R_\theta(\delta) = E_\theta\{C[\delta(Y), \theta]\}, \quad \theta \in \Lambda, \quad (\text{II.E.1})$$

όπου το E_θ συμβολίζει προσδοκία, αν υποθέσουμε ότι $Y \sim P_\theta$. Επίσης μια μέση ή Μπαϊεσιανή διακινδύνευση μπορεί να οριστεί ως

$$r(\delta) = E\{R_\Theta(\delta)\}, \quad (\text{II.E.2})$$

Και ο μπαϊεσιανός κανόνας ορίζεται ως αυτός που ελαχιστοποιεί το $r(\delta)$.

Αφού $E_\theta\{C[\delta(Y), \theta]\} \triangleq E\{C[\delta(Y), \Theta]|\Theta = \theta\}$ το $r(\delta)$ μπορεί να ξαναγραφεί ως

$$r(\delta) = E\{E\{C[\delta(Y), \Theta]|\Theta\}\} = E\{C[\delta(Y), \Theta]\}, \quad (\text{II.E.3})$$

Όπου η δεύτερη ισότητα προκύπτει από τη χρήση των επαναλαμβανόμενων προσδοκιών $E\{X\} = E\{E\{X|Y\}\}$. Άρα το $r(\delta)$ είναι απλώς το κόστος χρήσης του δ κατά μέσο όρο στο Θ και το Y . Χρησιμοποιώντας και πάλι επαναλαμβανόμενες προσδοκίες μπορούμε να γράψουμε

$$r(\delta) = E\{E\{C[\delta(Y), \Theta]|Y\}\}. \quad (\text{II.E.4})$$

Παρατηρήστε στο (II.E.4) ότι το $r(\delta)$ ελαχιστοποιείται στο δ αν για κάθε $y \in \Gamma$, επιλέγουμε το $\delta(y)$ να είναι το ύστερο κόστος

$$E\{C[\delta(Y), \Theta]|Y = y\}. \quad (\text{II.E.5})$$

Αφού το $\delta(y)$ μπορεί να είναι μόνο 0 ή 1, βλέπουμε ότι ο Μπαϊεσιανός κανόνας γι' αυτό το πρόβλημα δίνεται από

$$\delta_B(y) = \begin{cases} 1 & \text{if } E\{C[1, \Theta]|Y = y\} < E\{C[0, \Theta]|Y = y\} \\ 0 \text{ or } 1 & \text{if } E\{C[1, \Theta]|Y = y\} = E\{C[0, \Theta]|Y = y\} \\ 0 & \text{if } E\{C[1, \Theta]|Y = y\} > E\{C[0, \Theta]|Y = y\}. \end{cases} \quad (\text{II.E.6})$$

Η ερμηνεία του (II.E.6) είναι απλή. Το δ_B επιλέγει την υπόθεση με το μικρότερο κόστος, κατά μέσο όρο, δοθείσης της παρατήρησής μας. Στην περίπτωση που $\Lambda = \{0,1\}$, το (II.E.6) ανάγεται φυσικά στον Μπαϊεσιανό κανόνα για τον έλεγχο απλής υπόθεσης, που όπως συζητήθηκε στην Ενότητα II.B. έχει επίσης την ερμηνεία της ελαχιστοποίησης του ύστερου κόστους

Για πολλά ενδιαφέροντα προβλήματα, ο παραμετρικός χώρος μπορεί να αναλυθεί σε δύο ασύνδετα σύνολα Λ_0 και Λ_1 , που αναπαριστούν τις υποθέσεις H_0 και H_1 , αντίστοιχα, με τα κόστη να είναι ομοιόμορφα στα σύνολα αυτά. Δηλ.

$$C[i, \theta] = C_{ij}, \quad \theta \in \Lambda_j. \quad (\text{II.E.7})$$

Σε αυτήν την περίπτωση είναι προφανές ότι υπό την υπόθεση ότι $C_{11} < C_{01}$, το (II.E.6) ανάγεται σε

$$\delta_B(y) = \begin{cases} 1 & > \\ 0 \text{ or } 1 & \text{if } \frac{P(\Theta \in \Lambda_1|Y=y)}{P(\Theta \in \Lambda_0|Y=y)} = \frac{C_{10} - C_{00}}{C_{01} - C_{11}}, \\ 0 & < \end{cases} \quad (\text{II.E.8})$$

όπου το $P(\Theta \in \Lambda_j|Y=y)$ συμβολίζει την υπό συνθήκη πιθανότητα το Θ να βρίσκεται στο Λ_j , δοθέντος ότι $Y=y$. Υποθέτοντας ότι το Y έχει υπό συνθήκη πυκνότητες $p(y|\Theta \in \Lambda_j)$ για $j = 0, 1$, ο Μπαϊεσιανός τύπος συνεπάγεται ότι

$$P(\Theta \in \Lambda_j|Y=y) = \frac{p(y|\Theta \in \Lambda_j)P(\Theta \in \Lambda_j)}{p(y)} \quad (\text{II.E.9})$$

για $j = 0, 1$ με $p(y) = \sum_{j=0}^1 p(y|\Theta \in \Lambda_j)P(\Theta \in \Lambda_j)$. Άρα το (II.E.8) ανάγεται σε

$$\delta_B(y) = \begin{cases} 1 & > \\ 0 \text{ or } 1 & \text{if } L(y) = \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} \\ 0 & < \end{cases} \quad (\text{II.E.10})$$

με $\pi_j = P(\Theta \in \Lambda_j)$ και $L(y) = p(y|\Theta \in \Lambda_1)/p(y|\Theta \in \Lambda_0)$. Από το (II.E.10), βλέπουμε ότι το πρόβλημα αυτό είναι ισοδύναμο με το πρόβλημα του Μπαϊεσιανού ελέγχου απλής υπόθεσης με $p_j(y) = p(y|\Theta \in \Lambda_j)$, ένα δεδομένο που θα έπρεπε να είναι προφανές εκ των υστέρων.

Υποθέτοντας ότι Θ έχει πυκνότητα $w(\theta)$ και το P_θ έχει πυκνότητα p_θ για κάθε $\theta \in \Lambda$, τότε, χρησιμοποιώντας το συμβολισμό που είδαμε στο Κεφάλαιο I, οι πυκνότητες $p(y|\Theta \in \Lambda_j)$ δίνονται από

$$p(y|\Theta \in \Lambda_j) = \int_{\Lambda} p_\theta(y)w_j(\theta)\mu(d\theta), \quad (\text{II.E.11})$$

όπου $w_j(\theta)$ είναι η υπό συνθήκη πυκνότητα του Θ δοθέντος ότι $\Theta \in \Lambda_j$. Δηλ.

$$w_j(\theta) = \begin{cases} 0 & \text{if } \theta \notin \Lambda_j \\ w(\theta)/\pi_j & \text{if } \theta \in \Lambda_j, \end{cases} \quad (\text{II.E.12})$$

με $\pi_j = \int_{\Lambda_j} w(\theta)\mu(d\theta)$.

Σημειώστε ότι το ζεύγος υπόθεσης στο πρόβλημα αυτό ορίζεται από τις πυκνότητες παρατήρησης του (Π.Ε.11) που εξαρτώνται μόνο από τις υπό συνθήκη πυκνότητες w_j . Επομένως σύνθετα ελαχιστομέγιστα και Neyman – Pearson προβλήματα μπορούν να τεθούν μόνο όταν τα w_j είναι γνωστά αλλά όχι τα p_j . Φυσικά, πέρα από τις πιθανές διαφορετικές φυσικές ερμηνείες του παραμετρικού μοντέλου, τα προβλήματα αυτά δεν διαφέρουν από τα προβλήματα ελέγχου απλής υπόθεσης στις Ενότητες Π.Β. ως Π.Δ., όπως σημειώθηκε παραπάνω. Το παρακάτω παράδειγμα απεικονίζει αυτόν τον τύπο προβλήματος.

Παράδειγμα Π.Ε.1: Έλεγχος της ακτίνας ενός σημείου στο επίπεδο

Ας υποθέσουμε ότι $\Gamma = \mathbb{R}^2$ [i.e., $Y = (Y_1, Y_2)^T$] και οι υποθέσεις είναι ως ακολούθως:

$$H_0 : \begin{aligned} Y_1 &= \epsilon_1 \\ Y_2 &= \epsilon_2 \end{aligned}$$

Ως προς

$$H_1 : \begin{aligned} Y_1 &= A \cos \Psi + \epsilon_1 \\ Y_2 &= A \sin \Psi + \epsilon_2, \end{aligned} \quad (\text{Π.Ε.13})$$

Όπου A είναι μια θετική σταθερά, Ψ είναι μια τυχαία μεταβλητή που κατανέμεται ομοιόμορφα στο $[0, 2\pi]$ και ϵ_1 και ϵ_2 είναι $N(0, \sigma^2)$ τυχαίες μεταβλητές που είναι ανεξάρτητες μεταξύ τους και με το Ψ . Η παρατήρηση εδώ μπορεί να θεωρηθεί ως μια θορυβώδης μέτρηση των συντεταγμένων ενός σημείου στο επίπεδο που βρίσκεται είτε στην αρχή είτε κατανέμεται ομοιόμορφα σ' ένα κύκλο της ακτίνας A . (Εφαρμογές του μοντέλου αυτού θα δούμε στο Κεφάλαιο ΙΙΙ.) Η παράμετρος σ' αυτή την περίπτωση μπορεί να

θεωρηθεί ότι είναι $\Theta = (\Theta_1, \Theta_2)$ με $\Theta_1 \in \{0, A\}$ και $\Theta_2 \in [0, 2\pi]$. Το παραμετρικό σύνολο Λ είναι άρα $\{0, A\} \times [0, 2\pi]$ με $\Lambda_0 = \{\theta \in \Lambda | \theta_1 = 0\}$ και

$\Lambda_1 = \{\theta \in \Lambda | \theta_1 = A\}$. Η πυκνότητα του Y , δοθέντος ότι $\Theta = \theta$, είναι η από κοινού πυκνότητα δύο ανεξάρτητων $\mathcal{N}(0, \sigma^2)$ τυχαίων μεταβλητών που μετατοπίζονται στο μέσον από το $\theta_1 \cos \theta_2$ και το $\theta_1 \sin \theta_2$, αντίστοιχα. Δηλ.,

$$\begin{aligned}
p_{\theta}(y) &= \frac{1}{2\pi\sigma^2} \exp\{-q(y, \theta)/2\sigma^2\}, y \in \mathbb{R}^2 \\
q(y, \theta) &\triangleq [(y - \theta_1 \cos \theta_2)^2 + (y - \theta_1 \sin \theta_2)^2]. \quad (\text{II.E.14})
\end{aligned}$$

Αυτό έπεται άμεσα ότι

$$p(y|\Theta \in \Lambda_0) = p_{\theta}(y)|_{\theta_1=0} \quad (\text{II.E.15})$$

Και

$$\begin{aligned}
p(y|\Theta \in \Lambda_1) &= \frac{1}{2\pi} \int_0^{2\pi} p_{\theta}(y)|_{\theta_1=A} d\theta_2 \\
&= \frac{1}{4\pi^2\sigma^2} \int_0^{2\pi} \exp\{-q(y, \theta)|_{\theta_1=A/2\sigma^2}\} d\theta_2. \quad (\text{II.E.16})
\end{aligned}$$

Από τα (II.E.15) και (II.E.16) ο λόγος πιθανοφάνειας δίνεται από

$$\begin{aligned}
L(y) &= \frac{p(y|\Theta \in \Lambda_1)}{p(y|\Theta \in \Lambda_0)} \\
&= \frac{e^{-A^2/2\sigma^2}}{2\pi} \int_0^{2\pi} \exp\left\{\frac{A}{\sigma^2}(y_1 \cos \theta_2 + y_2 \sin \theta_2)\right\} d\theta_2, \quad (\text{II.E.17})
\end{aligned}$$

όπου χρησιμοποιήσαμε την ταυτότητα $\sin^2(\theta_2) + \cos^2(\theta_2) = 1$. Για να απλοποιήσουμε το (II.E.17) εισάγουμε μεταβλητές $r = [y_1^2 + y_2^2]^{1/2}$ και $\phi = \text{Tan}^{-1}(y_2/y_1)$ έτσι ώστε $y_1 = r \cos \phi$ και $y_2 = r \sin \phi$. Τότε, χρησιμοποιώντας την ταυτότητα $\cos \phi \cos \theta_2 + \sin \phi \sin \theta_2 = \cos(\theta_2 - \phi)$, έχουμε

$$\begin{aligned}
L(y) &= \frac{e^{-A^2/2\sigma^2}}{2\pi} \int_0^{2\pi} \exp\left\{\frac{Ar}{\sigma^2} \cos(\theta_2 - \phi)\right\} d\theta_2 \quad (\text{II.E.18}) \\
&= e^{-A^2/2\sigma^2} I_0\left(\frac{Ar}{\sigma^2}\right),
\end{aligned}$$

Όπου I_0 είναι η μηδενικής τάξης τροποποιημένη συνάρτηση Bessel πρώτου είδους, οριζόμενη από το ολοκλήρωμα στο (II.E.18)

Η συνάρτηση $I_0(x)$ είναι μονότονα αυξάνουσα στα ορίσματά της. Έτσι ένας έλεγχος που συγκρίνει το $L(y)$ από το (II.E.16) μ' ένα κατώφλι τ είναι ισοδύναμο με αυτό που συγκρίνει το r με ένα άλλο κατώφλι τ' που δίνεται από το $\tau' = \sigma^2 I_0^{-1}(\tau e^{A^2/2\sigma^2})/A$.

Ο Μπαϊεσιανός, ο ελαχιστομέγιστος και ο Neyman – Pearson έλεγχος για το (II.E.13) έχουν επομένως τη μορφή

$$\bar{\delta}_0(y) = \begin{cases} 1 & \text{if } r > \tau' \\ \gamma & \text{if } r = \tau' \\ 0 & \text{if } r < \tau'. \end{cases} \quad (\text{II.E.19})$$

Έτσι οι βέλτιστοι έλεγχοι για το (II.E.13) λειτουργούν συγκρίνοντας το r , την απόσταση του σημείου (y_1, y_2) από την αρχή, με ένα κατώφλι. Αυτός ο τρόπος απόφασης απεικονίζεται στο Σχ. II. E.1 (Σημειώστε ότι η τυχαιοποίηση είναι άνευ σημασίας εδώ)

Περισσότερες πτυχές αυτού του θέματος και σχετικά μοντέλα θα εξετάσουμε στο Κεφάλαιο III.

Για προβλήματα ελέγχου σύνθετης υπόθεσης για τα οποία δεν έχουμε πρότερη κατανομή (ή υπό συνθήκη πρότερα) για την παράμετρο, η ανάπτυξη ελέγχων υπόθεσης που ικανοποιεί ακριβώς τους αναλυτικούς ορισμούς της βελτιστότητας είναι συχνά ουτοπική. Ένας τρόπος ορισμού της βελτιστότητας σε τέτοια προβλήματα είναι η γενίκευση του κριτηρίου Neyman – Pearson στην Ενότητα II.D. Ας υποθέσουμε ότι ο παραμετρικός χώρος αναλύεται σε δύο ασύνδετα σύνολα Λ_0 και Λ_1 , όπως πριν. Για έναν τυχαιοποιημένο κανόνα απόφασης $\bar{\delta}$, μπορούμε να ορίσουμε τις πιθανότητες ψευδούς συναγερμούς και ανίχνευσης ως ακολούθως

$$P_F(\bar{\delta}; \theta) = E_{\theta}\{\bar{\delta}(Y)\}, \quad \theta \in \Lambda_0 \quad (\text{II.E.20})$$

Και

$$P_D(\bar{\delta}; \theta) = E_{\theta}\{\bar{\delta}(Y)\}, \quad \theta \in \Lambda_1.$$

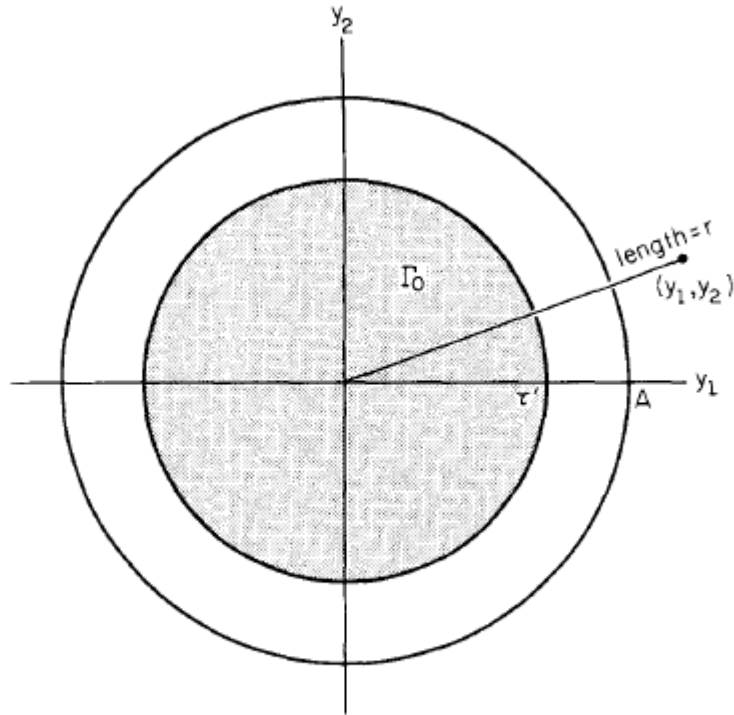


FIGURE II.E.1. Decision regions for Example II.E.1 ($\Gamma_1 = \Gamma_0^c$).

Σχήμα II.E.1 Τόποι απόφασης για το Παράδειγμα II.E.1 ($\Gamma_1 = \Gamma_0^c$).

Ας υποθέσουμε, όπως στη διατύπωση Neyman – Pearson, ότι επιθυμούμε να διασφαλίσουμε ότι η πιθανότητα ψευδούς συναγερωμούς δεν υπερβαίνει μια δεδομένη τιμή, ας πούμε, α . Τότε ιδανικός έλεγχος θα ήταν αυτός που μεγιστοποιεί το $P_D(\bar{\delta}; \theta)$ για κάθε $\theta \in \Lambda_1$, υποκείμενο σ' αυτόν τον περιορισμό ($P_F(\bar{\delta}; \theta) \leq \alpha, \theta \in \Lambda_0$). Αυτός ο έλεγχος είναι γνωστός ως ομοιόμορφα πιο ισχυρός (UMP) έλεγχος επιπέδου α .

Δυστυχώς, παρόλο που οι έλεγχοι UMP είναι επιθυμητοί, υπάρχουν μόνο υπό ειδικές συνθήκες. Για να το καταλάβετε, εξετάστε την κατάσταση στην οποία η μηδενική υπόθεση (H_0) είναι απλή, έτσι ώστε το Λ_0 συνίσταται σ' ένα μόνο στοιχείο θ_0 . Υποθέτοντας ότι το P_{θ_0} έχει πυκνότητα p_{θ_0} για κάθε $\theta \in \Lambda$, ο πιο ισχυρός έλεγχος α -επιπέδου για το H_0 ως προς την υπόθεση ότι $Y \sim P_{\theta}$, έχει κρίσιμη περιοχή $\Gamma_{\theta} = \{y \in \Gamma | p_{\theta}(y) > \tau p_{\theta_0}(y)\}$ με τ (και πιθανόν μια τυχαιοποίηση) που επιλέγεται για να δώσει μέγεθος α . Επίσης από το λήμμα Neyman – Pearson γνωρίζουμε ότι ο έλεγχος είναι εξαιρετικά μοναδικός και κάθε άλλος α -επιπέδου έλεγχος έχει μικρότερη ισχύ. Για παράδειγμα, αν επιλέξουμε δύο στοιχεία θ' και θ'' του Λ_1 , ο έλεγχος με κρίσιμη περιοχή $\Gamma_{\theta'}$ θα έχει μικρότερη ισχύ στον έλεγχο H_0 ως προς το $Y \sim P_{\theta''}$ απ' ό,τι έχει ο έλεγχος με κρίσιμη περιοχή $\Gamma_{\theta''}$ (και αντίστροφα), εκτός κι αν οι δύο

αυτές κρίσιμες περιοχές είναι κατ' ουσίαν όμοιες. Επομένως, έπεται ότι ένας έλεγχος UMP υπάρχει για το H_0 ως προς τη σύνθετη υπόθεση $H_1 : Y \sim P_\theta, \theta \in \Lambda_1$ αν και μόνο αν η κρίσιμη περιοχή Γ_θ είναι ίδια για όλα τα $\theta \in \Lambda_1$. Αυτό απεικονίζεται στο ακόλουθο παράδειγμα

Παράδειγμα II.E.2: Έλεγχος θέσης UMP

Εξετάστε την παραμετρική οικογένεια κατανομών $\{P_\theta; \theta \in \Lambda\}$, όπου P_θ είναι $N(\theta, \sigma^2)$ κατανομή και Λ είναι υποσύνολο του \mathbb{R} . Υποθέτουμε ότι έχουμε το ζεύγος υπόθεσης

$$H_0 : \theta = \mu_0$$

ως προς

$$H_1 : \theta > \mu_0 \quad (\text{II.E.21})$$

όπου μ_0 είναι ένας σταθερός αριθμός. Αυτό είναι ένα πρόβλημα με απλή μηδενική υπόθεση $\Lambda_0 = \{\mu_0\}$ και μια σύνθετη εναλλακτική $\Lambda_1 = (\mu_0, \infty)$. Από το Παράδειγμα II.D.1 γνωρίζουμε ότι για κάθε $\theta \in \Lambda$, ο πιο ισχυρός α -επιπέδου έλεγχος του H_0 ως προς $Y \sim N(\theta, \sigma^2)$ έχει κρίσιμη περιοχή $\Gamma_\theta = \{y \in \Gamma | y > \sigma\Phi^{-1}(1 - \alpha) + \mu_0\}$. Καθώς η περιοχή αυτή δεν εξαρτάται από το θ δίνει έναν έλεγχο UMP για το (II.E.21) το οποίο θα συμβολίσουμε με $\bar{\delta}_1$. Σημειώστε ότι το (II.D.16) συνεπάγεται ότι

$$P_D(\bar{\delta}_1; \theta) = 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{\theta - \mu_0}{\sigma}\right). \quad (\text{II.E.22})$$

Εναλλακτικά, για την ίδια οικογένεια κατανομών, υποθέστε ότι εξετάζουμε το ζεύγος υπόθεσης

$$H_0 : \theta = \mu_0$$

Ως προς

$$H_1 : \theta \neq \mu_0$$

(II.E.23)

Τώρα έχουμε τη σύνθετη εναλλακτική $\Lambda_1 = (-\infty, \mu_0) \cup (\mu_0, \infty)$. Για $\theta > \mu_0$, η πιο ισχυρή κρίσιμη περιοχή είναι όπως δίνεται στην προηγούμενη παράγραφο. Ωστόσο, για $\theta < \mu_0$ βλέπουμε προφανώς ότι ο πιο ισχυρός α -επιπέδου έλεγχος έχει κρίσιμη περιοχή

$$\Gamma_\theta = \{y \in \Gamma | y < \sigma\Phi^{-1}(\alpha) + \mu_0\}. \quad (\text{II.E.24})$$

Παρόλο που η περιοχή αυτή είναι ανεξάρτητη από το θ , είναι αρκετά διαφορετική από το Γ_θ για $\theta > \mu_0$. Άρα για το (II.E.23) δεν υπάρχει έλεγχος UMP

Αν συμβολίσουμε με $\bar{\delta}_2$ τον έλεγχο με κρίσιμη περιοχή το (II.E.24), τότε έχουμε ευθέως ότι

$$P_D(\bar{\delta}_2; \theta) = \Phi\left(\Phi^{-1}(\alpha) - \frac{\theta - \mu_0}{\sigma}\right). \quad (\text{II.E.25})$$

Το μέγεθος αυτό, μαζί με το $P_D(\bar{\delta}_1; \theta)$ από το (II.E.22) αναπαρίσταται ως προς το θ στο Σχ. II.E.2. Σημειώστε ότι κανένας από τους δύο ελέγχους δεν αποδίδει καλά όταν το θ είναι έξω από την περιοχή βελτιστότητάς του. (Ένας λογικότερος έλεγχος για το (II.E.23), είτε για το $\bar{\delta}_1$ είτε για το $\bar{\delta}_2$, είναι αυτός που συγκρίνει το $|y - \mu_0|$ με κάποιο κατώφλι. Ωστόσο, ο έλεγχος αυτός δεν μπορεί να είναι UMP για το (II.E.23).

Το παράδειγμα II.E.2 δείχνει ότι το κριτήριο UMP είναι πολύ ισχυρό για πολλές καταστάσεις. Κάποιες φορές αυτό μπορεί να ξεπεραστεί με την εφαρμογή άλλων περιορισμών για την εξάλειψη πραγματοποίησης αδικαιολόγητων ελέγχων. Μια τέτοια συνθήκη είναι η *αμεροληψία* που απαιτεί ότι $P_D(\bar{\delta}; \theta) \geq \alpha$ για όλα τα $\theta \in \Lambda_1$ μαζί με τον περιορισμό $P_F(\bar{\delta}; \theta) \leq \alpha$. Ας σημειωθεί ότι αυτή η προϋπόθεση θα εξαλείψει τόσο το

$\bar{\delta}_1$ όσο και το $\bar{\delta}_2$ στο παράδειγμα ως επιλογές ελέγχου για το (II.E.23). Το θέμα αυτό και άλλα σχετικά ζητήματα αναπτύσσονται περαιτέρω, για παράδειγμα, στο βιβλίο του Lehmann (1986).

Σε πολλές καταστάσεις, το σύνολο παραμέτρου Λ έχει τη μορφή $[\theta_0, \infty)$ με $\Lambda_0 = \{\theta_0\}$ και $\Lambda_1 = (\theta_0, \infty)$. Έτσι έχουμε το ζεύγος υπόθεσης

$$H_0 : \theta = \theta_0$$

ως προς

$$H_1 : \theta > \theta_0.$$

(II.E.26)

Αυτός ο τύπος κατάστασης προκύπτει, για παράδειγμα, σε πολλά προβλήματα ανίχνευσης σήματος στα οποία το $\theta_0 = 0$ και το θ είναι παράμετρος εύρους σήματος. Σε πολλές καταστάσεις αυτού του τύπου, μας ενδιαφέρει κυρίως η περίπτωση στην οποία, ως προς το H_1 , το θ είναι κοντά στο θ_0 . Αν για παράδειγμα το θ είναι μια παράμετρος εύρους σήματος, η τελευταία περίπτωση αναπαριστά την κατάσταση στην οποία η ισχύς του σήματος είναι μικρή.

Εξετάστε ένα κανόνα απόφασης $\bar{\delta}$. Εντός του πλαισίου της κανονικότητας μπορούμε να αναπτύξουμε σε σειρά Taylor το $P_D(\bar{\delta}; \theta)$ για το θ_0 . Δηλ.,

$$P_D(\bar{\delta}; \theta) = P_D(\bar{\delta}; \theta_0) + (\theta - \theta_0)P'_D(\bar{\delta}; \theta_0) + O((\theta - \theta_0)^2), \quad (\text{II.E.27})$$

όπου

$$P'_D(\bar{\delta}; \theta) = \partial P_D(\bar{\delta}; \theta) / \partial \theta.$$

Σημειώστε ότι $P_D(\bar{\delta}; \theta_0) = P_F(\bar{\delta})$

Έτσι για όλους τους ελέγχους μεγέθους- α , το $P_D(\bar{\delta}; \theta)$ δίνεται για θ κοντά στο θ_0 από

$$P_D(\bar{\delta}; \theta) \cong \alpha + (\theta - \theta_0)P'_D(\bar{\delta}; \theta_0). \quad (\text{II.E.28})$$

Άρα για θ κοντά στο θ_0 μπορούμε να πετύχουμε προσεγγιστική μέγιστη ισχύ με μέγεθος α επιλέγοντας $\bar{\delta}$ για τη μεγιστοποίηση του $P'_D(\bar{\delta}; \theta_0)$. Ο έλεγχος που μεγιστοποιεί το $P'_D(\bar{\delta}; \theta_0)$ που υπόκειται σε περιορισμό ψευδούς συναγερμού $P_F(\bar{\delta}) \leq \alpha$, ονομάζεται α -επιπέδου τοπικά πιο ισχυρός έλεγχος (LMP) ή απλώς τοπικά βέλτιστος έλεγχος.

Για να δούμε τη γενική δομή των ελέγχων LMP παρατηρούμε ότι, υποθέτοντας ότι το P_θ έχει πυκνότητα p_θ για κάθε $\theta \in \Lambda_1$, μπορούμε να γράψουμε

$$\begin{aligned}
P_D(\bar{\delta}; \theta) &= E_{\theta}\{\bar{\delta}(Y)\} \\
&= \int_{\Gamma} \bar{\delta}(y) p_{\theta}(y) \mu(dy).
\end{aligned}
\tag{II.E.29}$$

Αποδίδοντας επαρκή κανονικότητα στο $\{p_{\theta}(y); \theta \in \Lambda_1\}$, ώστε να μπορούμε να εναλλάσσουμε τη σειρά ολοκλήρωσης και διαφοροποίηση στο (II.E.29) έχουμε

$$P'_D(\bar{\delta}; \theta_0) = \int_{\Gamma} \bar{\delta}(y) \frac{\partial}{\partial \theta} p_{\theta}(y) |_{\theta=\theta_0} \mu(dy). \tag{II.E.30}$$

Η σύγκριση του (II.E.30) με το (II.D.4) υποδεικνύει ότι το πρόβλημα σχεδιασμού α -επιπέδου LMP είναι το ίδιο με το πρόβλημα σχεδιασμού α -επιπέδου Neyman – Pearson, αν αντικαταστήσουμε το $p_1(y)$ με $\frac{\partial p_{\theta}(y)}{\partial \theta} |_{\theta=\theta_0}$. Χρησιμοποιώντας αυτή την αναλογία, βλέπουμε ευθέως ότι, στο πλαίσιο της κανονικότητας, ένας έλεγχος α -επιπέδου LMP για το (II.E.26) δίνεται από

$$\bar{\delta}_{l\alpha}(y) = \begin{cases} 1 & > \\ \gamma & \text{if } \frac{\partial}{\partial \theta} p_{\theta}(y) |_{\theta=\theta_0} = \eta p_{\theta_0}(y), \\ 0 & < \end{cases} \tag{II.E.31}$$

όπου η και γ επιλέγονται ώστε $P_F(\bar{\delta}_{l\alpha}) = \alpha$. Λεπτομέρειες αυτού του αναπτύγματος βρίσκουμε στο βιβλίο του Ferguson (1968). Οι έλεγχοι LMP εξετάζονται περαιτέρω στο Κεφάλαιο III.

Εν απουσία εφαρμοσιμότητας οποιουδήποτε από τα παραπάνω αναφερθέντα κριτήρια βελτιστότητας, ένας έλεγχος που χρησιμοποιείται συχνά για σύνθετα προβλήματα στα οποία το θ είναι η ένωση των ασύνδετων Λ_0 και Λ_1 , είναι αυτός που βασίζεται στη σύγκριση του μεγέθους

$$\frac{\max_{\theta \in \Lambda_1} p_{\theta}(y)}{\max_{\theta \in \Lambda_0} p_{\theta}(y)} \tag{II.E.32}$$

με κάποιο κατώφλι. Ο έλεγχος αυτός είναι γνωστός ως γενικευμένος έλεγχος λόγου πιθανοφάνειας ή έλεγχος μέγιστης πιθανοφάνειας και περισσότερα κίνητρα για ελέγχους αυτού του τύπου απαντούν στο Κεφάλαιο IV.

ΑΝΙΧΝΕΥΣΗ ΣΗΜΑΤΟΣ ΣΕ ΔΙΑΚΡΙΤΟ ΧΡΟΝΟ

ΕΙΣΑΓΩΓΗ

Στο Κεφάλαιο II εξετάσαμε αρκετά βασικά κριτήρια βελτιστότητας και μεθόδους σχεδιασμού για δυαδικά προβλήματα ελέγχου υπόθεσης. Στο κεφάλαιο αυτό εφαρμόζουμε αυτές και άλλες σχετικές μεθόδους, ώστε να παράγουμε βέλτιστες διαδικασίες για την ανίχνευση σημάτων, ενσωματωμένων στο θόρυβο. Για να αποφύγουμε τις αναλυτικές επιπλοκές, εξετάζουμε αποκλειστικά την περίπτωση ανίχνευσης διακριτού χρόνου, αφήνοντας την περίπτωση του συνεχούς χρόνου για το Κεφάλαιο VI. Η περίπτωση διακριτού χρόνου έχει σημαντικό πρακτικό ενδιαφέρον, λόγω της επικράτησης των ψηφιακών εφαρμογών στις λειτουργίες επεξεργασίας σήματος.

Στην Ενότητα III.B εξετάζουμε διάφορα μοντέλα για προβλήματα ανίχνευσης σήματος και εξάγουμε τις βέλτιστες δομές ανιχνευτών που ανταποκρίνονται στα κριτήρια που τέθηκαν στο Κεφάλαιο II. Η Ενότητα III.C πραγματεύεται κάποιες μεθόδους ανάλυσης της επίδοσης των δομών αυτών για καταστάσεις στις οποίες η κλειστή μορφή υπολογισμού σχετικών σφαλμάτων πιθανοτήτων δεν είναι ανιχνεύσιμη. Υπάρχουν αρκετές χρήσιμες μέθοδοι σχεδιασμού για διαδικασίες ανίχνευσης, εκτός από εκείνες που μελετήθηκαν στο Κεφάλαιο II, ενώ στις Ενότητες III.D και III.E εισάγουμε τρεις τέτοιες μεθόδους, κυρίως, την ακολουθιακή, την ανθεκτική και τη μη-παραμετρική ανίχνευση.

III.B. Μοντέλα και Δομές Ανιχνευτών

Το βασικό φυσικό μοντέλο παρατήρησης που θέλουμε να εξετάσουμε είναι αυτό μιας παρατηρούμενης συνεχούς χρόνου κυματομορφής που συνίσταται σε ένα από τα δύο πιθανά σήματα που αλλοιώνονται από πρόσθετο θόρυβο. Σκοπός μας είναι να αποφασίσουμε ποιο από τα δύο πιθανά σήματα είναι παρόν, και επιθυμούμε να το κάνουμε επεξεργαζόμενοι έναν πεπερασμένο αριθμό (π.χ. n) σημάτων που λαμβάνονται από την παρατηρούμενη κυματομορφή.

Το πρόβλημα αυτό μπορεί να τυποποιηθεί στατιστικά από το ακόλουθο ζεύγος υπόθεσης για χώρο παρατήρησης $(\Gamma, \mathcal{G}) = (\mathbb{R}^n, \mathcal{B}^n)$:

$$\begin{array}{l} \text{versus} \\ H_0 : Y_k = N_k + S_{0k}, \quad k = 1, 2, \dots, n \\ H_1 : Y_k = N_k + S_{1k}, \quad k = 1, 2, \dots, n, \end{array} \quad (\text{III.B.1})$$

(ως προς)

όπου το $\underline{Y} = (Y_1, \dots, Y_n)^T$ είναι ένα διάνυσμα παρατήρησης που αποτελείται από τα δείγματα της υπό παρατήρησης κυματομορφής, το $\underline{N} = (N_1, \dots, N_n)^T$ είναι ένα διάνυσμα δειγμάτων θορύβου και τα $\underline{S}_0 = (S_{01}, \dots, S_{0n})^T$ και $\underline{S}_1 = (S_{11}, \dots, S_{1n})^T$ είναι διανύσματα δειγμάτων από τα δύο πιθανά σήματα¹. Ας

1

σημειωθεί ότι η ερμηνεία του \underline{Y} ως ένα διάνυσμα δειγμάτων χρόνου, δεν είναι η μόνη πιθανότητα για το (II.B.1), καθώς το ίδιο μοντέλο προκύπτει αν, για παράδειγμα, παίρνουμε παρόμοια (σε χρόνο) δείγματα από n χωρικά διαχωρισμένους ανιχνευτές σήματος ή από τις εξόδους μιας τράπεζας n παράλληλων φίλτρων. Σε κάθε περίπτωση, θα αναφέρουμε αυτόν τον κάτω δείκτη ως παράμετρο χρόνου, παρόλο που τα αποτελέσματα εφαρμόζονται εξίσου καλά και σε άλλες καταστάσεις που προτυποποιούνται από το (III.B.1).

Βέλτιστες διαδικασίες απόφασης μεταξύ του H_0 και του H_1 μπορούν να εξαχθούν με τη χρήση των αποτελεσμάτων του Κεφαλαίου II, αν έχουμε μοντέλα για τη στατιστική συμπεριφορά των σημάτων και του θορύβου. Για πρακτικούς σκοπούς, τα σήματα \underline{S}_0 και \underline{S}_1 μπορούν συνήθως να ταξινομηθούν ως ένας από τους τρεις βασικούς τύπους. Μπορούν να είναι απολύτως γνωστά (δηλ. ντετερμινιστικά), μπορούν να είναι γνωστά για ένα σύνολο αγνώστων (πιθανώς τυχαίων) παραμέτρων, ή μπορεί να είναι εντελώς τυχαία και επομένως να καθορίζονται μόνο από τις κατανομές των πιθανοτήτων τους. Κάποιες φορές (π.χ. σε προβλήματα ραντάρ/σόναρ) ένα από τα σήματα, συνήθως το \underline{S}_0 , είναι πανομοιότυπα μηδέν, ώστε στην ουσία προσπαθούμε να ανιχνεύσουμε ένα σήμα ενσωματωμένο στο θόρυβο. Για τους σκοπούς αυτής της μελέτης, θα υποθέσουμε ότι ο θόρυβος είναι ανεξάρτητος από τα σήματα ως προς κάθε υπόθεση και ότι η κατανομή πιθανότητάς του δεν εξαρτάται από το ποια υπόθεση είναι αληθής. Η παραδοχή αυτή είναι έγκυρη για τις περισσότερες εφαρμογές, παρόλο που σε κάποιες εφαρμογές ο θόρυβος μπορεί να εξαρτάται από το σήμα (ένα παράδειγμα αυτού είναι το πρόβλημα ραντάρ/σόναρ, στο οποίο ο θόρυβος κάποιες φορές συντίθεται μερικώς από πλασματικές ανακλάσεις σήματος από το έδαφος ή από αντικείμενα πέρα από το σκοπευόμενο στόχο). Επίσης θα υποθέσουμε γενικώς ότι αυτή η κατανομή θορύβου ορίζεται από μια (συνεχή ή διακριτή) πυκνότητα p_n ή \mathbb{R}^n .

Υπό τις παραπάνω παραδοχές, ο λόγος πιθανοφάνειας για το (III.B.1) μπορεί να υπολογιστεί, αν γνωρίζουμε τα στατιστικά του \underline{S}_j για $j = 0, 1$. Συγκεκριμένα, δοθέντος ότι

$$\underline{S}_j = \underline{s}_j \in \mathbb{R}^n, \text{ η παρατήρηση } \underline{Y} \text{ έχει υπό συνθήκη πυκνότητα (ως προς το } H_j)$$

¹ Εδώ και σε άλλα σημεία του βιβλίου, τα διανύσματα θεωρούνται κιονοειδή και ο άνω δείκτης T συμβολίζει μετάθεση.

$$p_{\underline{N}}(\underline{y} - \underline{s}_j), \quad \underline{y} \in \mathbb{R}^n. \quad (\text{III.B.2})$$

Από το (III.B.2) βλέπουμε ότι η πυκνότητα του \underline{Y} ως προς το H_j δίνεται από

$$p_j(\underline{y}) = E\{p_{\underline{N}}(\underline{y} - \underline{S}_j)\}, \quad \underline{y} \in \mathbb{R}^n, \quad (\text{III.B.3})$$

όπου θα πρέπει να σημειωθεί ότι η προσδοκία είναι ανάλογη με το σήμα \underline{S}_j . Τότε ο λόγος πιθανοφάνειας γίνεται

$$L(\underline{y}) = \frac{E\{p_{\underline{N}}(\underline{y} - \underline{S}_1)\}}{E\{p_{\underline{N}}(\underline{y} - \underline{S}_0)\}}, \quad \underline{y} \in \mathbb{R}^n. \quad (\text{III.B.4})$$

Επομένως, οι βέλτιστες διαδικασίες για το (III.B.1) εξάγονται υπολογίζοντας το (III.B.4) και αναλύοντας με την εύρεση των κατάλληλων πιθανοτήτων των προκυπτουσών κρίσιμων περιοχών. Τώρα, θα εξετάσουμε μια σειρά σημαντικών ειδικών περιπτώσεων του προβλήματος αυτού.

Περίπτωση III.B.1: Ανίχνευση ντετερμινιστικών σημάτων σε ανεξάρτητο θόρυβο

Σε πολλά από τα προβλήματα που μας αφορούν, τα δύο σήματα \underline{S}_0 και \underline{S}_1 είναι απολύτως ντετερμινιστικά. Συγκεκριμένα, έχουμε $\underline{S}_j = \underline{s}_j$, με το $\underline{s}_j \in \mathbb{R}^n$ να είναι γνωστό στο σχεδιαστή. Αυτό είναι ενίοτε γνωστό ως το *συνεκτικό* πρόβλημα ανίχνευσης.

Στην περίπτωση αυτή η $L(\underline{y})$ του (III.B.4) γίνεται

$$\begin{aligned} L(\underline{y}) &= \frac{p_{\underline{N}}(\underline{y} - \underline{s}_1)}{p_{\underline{N}}(\underline{y} - \underline{s}_0)} \\ &= \frac{p_{\underline{N}}(y_1 - s_{11}, \dots, y_n - s_{1n})}{p_{\underline{N}}(y_1 - s_{01}, \dots, y_n - s_{0n})}, \end{aligned} \quad (\text{III.B.5})$$

και άρα η βέλτιστη δομή ανιχνευτή προσδιορίζεται εύκολα αν το $p_{\underline{N}}$ είναι γνωστό. Δεν μπορούν να ειπωθούν πολλά για ελέγχους βασισμένους στο (III.B.5) χωρίς να γίνουν περαιτέρω απλουστευτικές υποθέσεις. Γενικά, το (III.B.5) θα μπορούσε να είναι μια πολύπλοκη συνάρτηση των παρατηρήσεων που θα ήταν πολύ δύσκολο να εφαρμοστούν και να αναλυθούν, ειδικότερα αν ο αριθμός των δειγμάτων (n) είναι μεγάλος. Για παράδειγμα, η θέση κατωφλιών στα προβλήματα ελαχιστομέγιστου ή Neyman – Pearson και ο υπολογισμός

της επίδοσης απαιτούν την αξιολόγηση των n-πλασίων ολοκληρωμάτων της μορφής $\int_{\{L(\underline{y}) > \tau\}} p_j(\underline{y}) \mu(d\underline{y})$, μια διαδικασία συχνά δύσκολη. Μια σημαντική ειδική περίπτωση του (III.B.1) με γνωστά σήματα είναι αυτή στην οποία τα δείγματα θορύβου N_1, \dots, N_n είναι στατιστικά ανεξάρτητα. Στην περίπτωση αυτή έχουμε

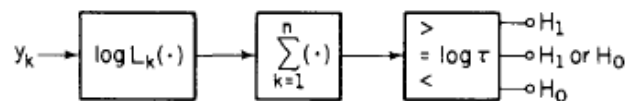
$$p_{\underline{N}}(\underline{y}) = \prod_{k=1}^n p_{N_k}(y_k), \quad (\text{III.B.6})$$

όπου το p_{N_k} είναι η οριακή πυκνότητα του N_k , ώστε η $L(\underline{y})$ γίνεται

$$L(\underline{y}) = \prod_{k=1}^n L_k(y_k) \quad (\text{III.B.7})$$

με $L_k(y_k) = p_{N_k}(y_k - s_{1k}) / p_{N_k}(y_k - s_{0k})$. Καθώς το $\log(x)$ είναι μια αυστηρά αυξανόμενη συνάρτηση του x , η σύγκριση της $L(\underline{y})$ με ένα κατώφλι τ είναι ισοδύναμη με τη σύγκριση της $\log L(\underline{y})$ με το κατώφλι $\log(\tau)$. Επομένως, οι βέλτιστοι έλεγχοι για την περίπτωση αυτή μπορούν να γραφεί ως

$$\bar{\delta}_o(\underline{y}) = \begin{cases} 1 \\ \gamma \\ 0 \end{cases} \quad \text{if } \sum_{k=1}^n \log L_k(y_k) \begin{matrix} > \\ = \\ < \end{matrix} \log \tau. \quad (\text{III.B.8})$$



Σχήμα III.B.2 Δομή ανιχνευτή για συνεκτικά σήματα σε ανεξάρτητο θόρυβο

Όπως απεικονίζεται στο Σχ. III.B.1 η δομή αυτή συνίσταται σε μια χρονικά μεταβαλλόμενη στιγμιαία μη γραμμικότητα, $\log L_k$, που ακολουθείται από έναν συσσωρευτή, που με τη σειρά του ακολουθείται από έναν συγκριτή κατωφλιού. Τα ακόλουθα δύο παραδείγματα απεικονίζουν δύο ειδικές περιπτώσεις του (III.B.8) που προκύπτουν πρακτικά.

Παράδειγμα III.B.1 Συνεκτική ανίχνευση με ανεξάρτητο και πανομοιότυπα καταναμημένο (i.i.d.) κατά Gauss θόρυβο.

Ας υποθέσουμε ότι τα δείγματα θορύβου N_1, \dots, N_n είναι ανεξάρτητα και πανομοιότυπα καταναμημένα (i.i.d.) με οριακή κατανομή $N(0, \sigma^2)$. Ένα τέτοιο μοντέλο θορύβου προκύπτει, για παράδειγμα, σε δέκτες επικοινωνίας, όταν η κύρια πηγή θορύβου είναι ο λεγόμενος *θερμικός θόρυβος*, που παράγεται από την κίνηση των ηλεκτρονίων στα ηλεκτρονικά του δέκτη. Για λόγους απλότητας, ας υποθέσουμε επίσης ότι $\underline{s}_0 = \underline{0}$, όπου το $\underline{0}$ συμβολίζει το n -διάνυσμα όλων των μηδέν και ας συμβολίσουμε το \underline{s}_1 με \underline{s} . (Ας σημειωθεί ότι αυτή η υπόθεση δεν επιφέρει οποιαδήποτε απώλεια σε γενικότητα, καθώς μπορούμε πάντα να επαναπροσδιορίσουμε τις παρατηρήσεις ως $\underline{y}' = \underline{y} - \underline{s}_0$ ώστε το σήμα να είναι $\underline{0}$ ως προς το H_0 και $\underline{s} = \underline{s}_1 - \underline{s}_0$ ως προς το H_1 . Τότε έχουμε (όπως στο Παράδειγμα II.B.2 ότι $\log L_k(\underline{y}_k) = s_k(\underline{y}_k - s_k/2)/\sigma^2$, οπότε ο βέλτιστος έλεγχος γίνεται

$$\tilde{\delta}_o(\underline{y}) = \begin{cases} 1 & \text{if } \sum_{k=1}^n s_k(\underline{y}_k - s_k/2) > \tau' \\ \gamma & \text{if } \sum_{k=1}^n s_k(\underline{y}_k - s_k/2) = \tau' \\ 0 & \text{if } \sum_{k=1}^n s_k(\underline{y}_k - s_k/2) < \tau' \end{cases} \quad (\text{III.B.9})$$

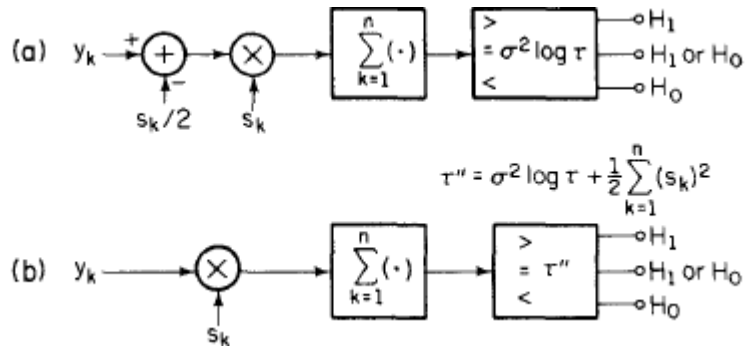
με $\tau' \triangleq \sigma^2 \log \tau$. Η δομή αυτή απεικονίζεται στο Σχ. III.B.2 (a). Ας σημειωθεί ότι ο όρος $-\frac{1}{2} \sum_{k=1}^n s_k^2$ μπορεί να ενσωματωθεί στο κατώφλι, οπότε ένας έλεγχος ισοδύναμος με το (III.B.9) είναι αυτός που συγκρίνει το $\sum_{k=1}^n s_k \underline{y}_k$ με το κατώφλι $\tau'' = \tau' + \frac{1}{2} \sum_{k=1}^n s_k^2$. Η δομή αυτή απεικονίζεται στο Σχ. III.B.2 (b) και είναι γνωστή ως *ανιχνευτής συσχέτισμού ή συσχέτιστής*.

Ένα σημαντικό χαρακτηριστικό αυτού του βέλτιστου ανιχνευτή για θόρυβο κατά Gauss είναι ότι λειτουργεί συγκρίνοντας την έξοδο ενός γραμμικού συστήματος με ένα κατώφλι.

Ειδικότερα, μπορούμε να γράψουμε $\sum_{k=1}^n s_k \underline{y}_k = \sum_{k=-\infty}^n h_{n-k} \underline{y}_k$, όπου

$$h_k = \begin{cases} s_{n-k} & \text{for } 0 \leq k \leq n-1 \\ 0 & \text{otherwise.} \end{cases} \quad (\text{III.B.10})$$

Άρα αυτός ο ανιχνευτής μπορεί να θεωρηθεί ως ένα σύστημα που εισάγει την ακολουθία παρατήρησης y_1, \dots, y_n σε ένα ψηφιακό γραμμικό φίλτρο και κατόπιν χρησιμοποιεί ως δείγμα την έξοδο σε χρόνο n για σύγκριση με κατώφλι. Η δομή αυτή είναι γνωστή ως *προσαρμοσμένο φίλτρο*.



Σχήμα III.B.2 Βέλτιστος ανιχνευτής για συνεκτικά σήματα i.i.d. θορύβου κατά Gauss

Παράδειγμα III.B.2.: Συνεκτική ανίχνευση σε i.i.d. Λαπλασιανού θορύβου.

Ας υποθέσουμε, όπως στο Παράδειγμα III.B.1 ότι τα δείγματα θορύβου N_1, \dots, N_n είναι i.i.d., αλλά με Λαπλασιανή οριακή πυκνότητα πιθανότητας

$$p_{N_k}(y_k) = \frac{\alpha}{2} e^{-\alpha|y_k|}, \quad y_k \in \mathbb{R}, \quad (\text{III.B.11})$$

όπου $\alpha > 0$ είναι μια παράμετρος κλίμακας της πυκνότητας. Το μοντέλο αυτό χρησιμοποιείται ενίοτε για να αναπαραστήσει τη συμπεριφορά των ωστικών θορύβων σε επικοινωνιακούς δέκτες. Σε σύγκριση με το μοντέλο κατά Gauss, χαρακτηρίζεται από μακρύτερες «ουρές», αναπαριστώντας υψηλότερες πιθανότητες μεγαλύτερων παρατηρήσεων.

$$\log L_k(y_k) = \alpha(|y_k| - |y_k - s_k|)$$

Η συνάρτηση $\log L_k(y_k)$ για το (III.B.11) δίνεται από

που μπορεί να γραφεί ως

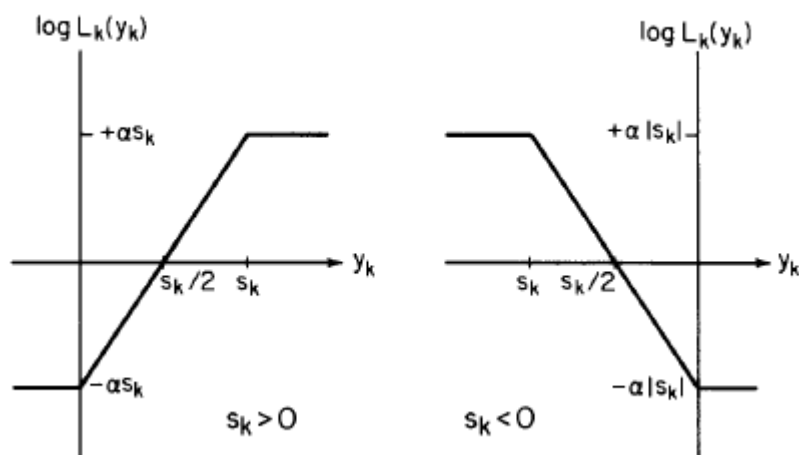
$$\log L_k(y_k) = \begin{cases} -\alpha|s_k| & \text{if } \text{sgn}(s_k)y_k \leq 0 \\ \alpha|2y_k - s_k| & \text{if } 0 < \text{sgn}(s_k)y_k < |s_k| \\ +\alpha|s_k| & \text{if } \text{sgn}(s_k)y_k \geq |s_k| \end{cases} \quad (\text{III.B.12})$$

Όπου το sgn συμβολίζει τη συνάρτηση signum

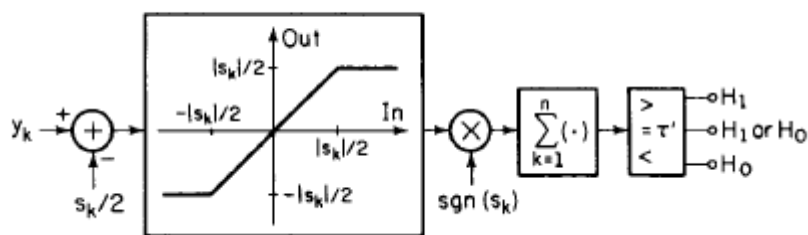
$$\text{sgn}(x) = \begin{cases} +1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases} \quad (\text{III.B.13})$$

Η συνάρτηση αυτή $\log L_k(y_k)$ απεικονίζεται στο Σχ. III.B.3 και για τις δύο περιπτώσεις $s_k < 0$ και $s_k > 0$. Μελετώντας τα σχήματα αυτά φαίνεται ότι οι βέλτιστοι ανιχνευτές βασισμένοι στο $\sum_{k=1}^n \log L_k(y_k)$ μπορούν να εφαρμοστούν ως

$$\tilde{\delta}_o(\underline{y}) = \begin{cases} 1 \\ \gamma \\ 0 \end{cases} \text{ if } \sum_{k=1}^n \text{sgn}(s_k) l_k(y_k) \begin{matrix} > \\ = \\ < \end{matrix} \tau', \quad (\text{III.B.14})$$



Σχήμα III.B.3 Ανά δείγμα λόγου λογαριθμικής πιθανοφάνειας για συνεκτική ανίχνευση σε Λαπλασιανό θόρυβο.



Σχήμα III.B.4. Βέλτιστος ανιχνευτής για συνεκτικά σήματα σε Λαπλασιανό θόρυβο.

όπου: $\tau' \triangleq \log \tau / 2\alpha$, και όπου οι συναρτήσεις l_k δίνονται από

$$l_k(x) = \begin{cases} -|s_k|/2 & \text{if } x \leq -|s_k|/2 \\ x & \text{if } -|s_k|/2 < x < |s_k|/2 \\ +|s_k|/2 & \text{if } x \geq +|s_k|/2. \end{cases} \quad (\text{III.B.15})$$

Ας σημειωθεί ότι η $l_k(x)$ είναι γραμμική στο x αν $x \in (-|s_k|/2, +|s_k|/2)$ και διαφορετικά η $l_k(x)$ ισούται με $|s_k| \text{sgn}(x)/2$. Η συνάρτηση αυτή είναι ενίοτε γνωστή ως απαλός περιοριστής ή περιοριστής ενισχυτή. Το σύστημα του (III.B.14) απεικονίζεται στο Σχ. III.B.4.

Έχει ενδιαφέρον να συγκρίνουμε τους ανιχνευτές των Σχημάτων (III.B.2(a) και III.B.4. Και τα δύο συστήματα «συγκεντρώνουν» τις παρατηρήσεις αφαιρώντας το $s_k/2$ από κάθε y_k . Το σύστημα του Σχ. III.B.2(a) τότε συσχετίζει τα συγκεντρωμένα δεδομένα με το γνωστό σήμα και συγκρίνει την έξοδο αυτού του συσχετισμού με ένα κατώφλι.

Εναλλακτικά, ο ανιχνευτής του Σχ. III.B.4 **περιορίζει μαλακά** τα συγκεντρωμένα δεδομένα και κατόπιν συσχετίζει αυτά τις **μαλακά περιορισμένες** παρατηρήσεις με την ακολουθία των σημείων των σημάτων. Η επίδραση αυτού του **μαλακού περιορισμού** είναι να μειώνει την ισχύ των μεγάλων παρατηρήσεων στο σύνολο, κάνοντας έτσι αυτό το σύστημα πιο ανεκτικό σε μεγάλες τιμές θορύβου απ' ό,τι το γραμμικό σύστημα του Σχ. III.B.2(a). Η διεργασία αυτή συνάδει με τη συμπεριφορά των Λαπλασιανών τυχαίων μεταβλητών κατά τη σύγκρισή της με την κατά Gauss, όπως αναφέρθηκε παραπάνω.

Παράδειγμα III.B.3: Τοπικά βέλτιστη ανίχνευση συνεκτικών σημάτων σε i.i.d. θόρυβο.

Σε πολλά προβλήματα ανίχνευσης, η μορφή του λαμβανόμενου σήματος είναι γνωστή, αλλά όχι το εύρος του. Για να προτυποποιήσουμε το πρόβλημα αυτό λαμβάνουμε υπόψη μας το πρόβλημα ελέγχου σύνθετης υπόθεσης που περιγράφεται από

$$\begin{aligned} & H_0 : Y_k = N_k, \quad k = 1, 2, \dots, n \\ \text{versus} & H_1 : Y_k = N_k + \theta s_k, \quad k = 1, 2, \dots, n, \quad \theta > 0, \end{aligned} \quad (\text{III.B.16})$$

όπου το $\underline{s} = (s_1, \dots, s_n)^T$ είναι ένα γνωστό σήμα, το $\underline{N} = (N_1, \dots, N_n)^T$ είναι ένα διάνυσμα συνεχούς τυχαίου θορύβου με στοιχεία i.i.d. και συναρτήσεις οριακής πιθανότητας πυκνότητας p_{N_k} , και το θ είναι παράμετρος σήματος – ισχύος. Δοθέντος του θ , ο λόγος πιθανοφάνειας μεταξύ του H_0 και H_1 δίνεται από

$$L_\theta(\underline{y}) = \prod_{k=1}^n \frac{p_{N_k}(y_k - \theta s_k)}{p_{N_k}(y_k)}. \quad (\text{III.B.17})$$

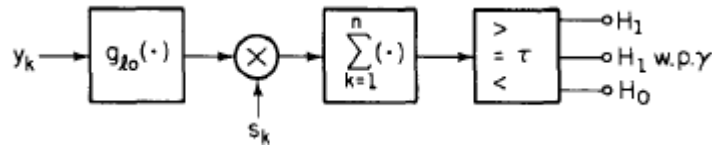
Ας σημειωθεί ότι η κρίσιμη περιοχή $\Gamma_\theta = \{y \in \mathbb{R}^n | L_\theta(y) > \tau\}$ θα εξαρτάται γενικά από το θ στην περίπτωση αυτή, εκτός από κάποιες ειδικές περιπτώσεις (Η περίπτωση θορύβου κατά Gauss του Παραδείγματος III.B.1 είναι μια τέτοια ειδική περίπτωση. Επομένως οι έλεγχοι UMP για το (III.B.16) υφίστανται μόνο για συγκεκριμένα μοντέλα θορύβου. Ωστόσο, οι έλεγχοι LMP για το (III.B.16) έχουν μια ιδιαίτερα απλή και διαισθητικά λογική δομή και επομένως έχει ενδιαφέρον να εξετάσουμε την τοπικά βέλτιστη ανίχνευση για την περίπτωση αυτή.

Υποθέτοντας ότι υπάρχει επαρκής κανονικότητα στο p_{N_k} , ο τοπικά βέλτιστος έλεγχος για H_0 ως προς H_1 δίνεται από [Βλέπε (II.E.31)]

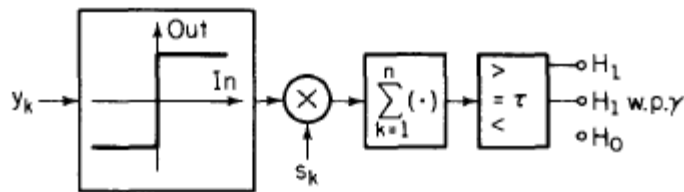
$$\bar{\delta}_{l_0}(\underline{y}) = \begin{cases} 1 & > \\ \gamma & \text{if } \frac{\partial}{\partial \theta} L_\theta(\underline{y}) |_{\theta=0} = \tau. \\ 0 & < \end{cases} \quad (\text{III.B.18})$$

Με διαφορισμό του (III.B.17) έχουμε

$$\frac{\partial}{\partial \theta} L_{\theta}(\underline{y})|_{\theta=0} = \sum_{k=1}^n s_k g_{l_0}(y_k), \quad (\text{III.B.19})$$



ΣΧΗΜΑ III.B.5 Δομή τοπικά βέλτιστου ανιχνευτή για συνεκτικά σήματα σε θόρυβο i.i.d.



ΣΧΗΜΑ III.B.6 Τοπικά βέλτιστος ανιχνευτής για Λαπλασιανό θόρυβο.

όπου $g_{l_0}(\mathbf{x}) \triangleq -p'_{N_1}(\mathbf{x})/p_{N_1}(\mathbf{x})$ και όπου $p'_{N_1}(\mathbf{x}) = dp_{N_1}(\mathbf{x})/d\mathbf{x}$. Η δομή αυτή απεικονίζεται στο Σχ. III.B.5. Αποτελείται από την αμνήμονα μη γραμμικότητα g_{l_0} που ακολουθείται από έναν συσχετιστή, ένας συνδυασμός γνωστός ως μη γραμμικός συσχετιστής

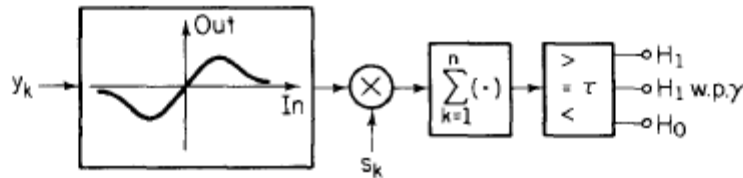
Όπως ο λόγος πιθανοφάνειας, η τοπικά βέλτιστη μη γραμμικότητα g_{l_0} διαμορφώνει τις παρατηρήσεις για να ελαττώσει τις επιζήμιες επιπτώσεις του θορύβου όσο το δυνατόν περισσότερο. Για παράδειγμα, με $N(0, \sigma^2)$ θόρυβο έχουμε $g_{l_0}(\mathbf{x}) = \mathbf{x}/\sigma^2$. Οπότε το Σχ. III.B.5 είναι απλώς ο ανιχνευτής συσχέτισης του Σχ. III.B.2 (b). (Έτσι πρέπει να είναι. Αφού αυτός ο ανιχνευτής είναι UMP είναι και LMP.). Για Λαπλασιανό θόρυβο με πυκνότητα του (III.B.11) έχουμε $g_{l_0}(\mathbf{x}) = \alpha \operatorname{sgn}(\mathbf{x})$, οπότε ο τοπικά βέλτιστος ανιχνευτής συσχετίζει το σήμα με την ακολουθία σημάτων των παρατηρήσεων, όπως απεικονίζεται στο Σχ. III.B.6. Η συνάρτηση $g_{l_0}(\mathbf{x})$, στην περίπτωση αυτή, είναι γνωστή ως σκληρός περιοριστής. Ένα μοντέλο θορύβου ακόμα «βαρύτερης ουράς» από το Λαπλασιανό περιγράφεται από την πυκνότητα του Cauchy που δίνεται από

$$p_{N_1}(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty. \quad (\text{III.B.20})$$

Για την περίπτωση αυτή $g_{10}(x) = 2x/(1+x^2)$, που είναι κατά προσέγγιση γραμμικό κοντά στο $x = 0$ και κατόπιν επανακατέρχεται (ασυμπτωματικά) στο μηδέν. Επομένως, αυτός ο ανιχνευτής αγνοεί τις παρατηρήσεις με πολύ μεγάλη έκταση. Η συμπεριφορά αυτή απεικονίζεται στο Σχ. III.B.7. (Μια προσέγγιση σ' αυτή τη μη γραμμικότητα είναι ο λεγόμενος *καταπιεστής θορύβου*,

$$g(x) = \begin{cases} x & \text{if } |x| \leq K \\ 0 & \text{if } |x| > K, \end{cases} \quad (\text{III.B.21})$$

όπου $K > 0$ είναι μια σταθερά, η οποία ενίοτε χρησιμοποιείται στην πράξη για να καταπολεμήσει τον εξαιρετικά ορμητικό θόρυβο).



ΣΧΗΜΑ III.B.7 Τοπικά βέλτιστος ανιχνευτής για θόρυβο του Cauchy

Περίπτωση III.B.2 : Ανίχνευση ντετερμινιστικών σημάτων σε θόρυβο κατά Gauss

Αν τα δείγματα θορύβου N_k στο (III.B.1) δεν είναι ανεξάρτητα μεταξύ τους, τότε οι βέλτιστες λύσεις του (III.B.4) δεν επιδεικνύουν κάποια ιδιαίτερα δομή (πέρα, ας πούμε, από το (III.B.5) ακόμη και για την περίπτωση των ντετερμινιστικών σημάτων. Μια σημαντική εξαίρεση σ' αυτή την έλλειψη γενικής δομής είναι η κατάσταση στην οποία τα σήματα είναι

ντετερμινιστικά ($\underline{S}_j = \underline{s}_j$) και το διάνυσμα θορύβου \underline{N} έχει μια πολυμεταβλητή κατά Gauss κατανομή. Στην περίπτωση αυτή, οι βέλτιστοι ανιχνευτές έχουν απλές, εύκολα εφαρμοζόμενες δομές και η επίδοση των βέλτιστων συστημάτων μπορούν να αναλυθούν διεξοδικά. Επιπλέον, η παραδοχή του κατά Gauss θορύβου δικαιολογείται συχνά στην πράξη,

και τα συστήματα που θεωρούνται ότι είναι βέλτιστα υπό την παροδοχή αυτή είναι διαισθητικά λογικά συστήματα για να χρησιμοποιηθούν ακόμη κι όταν ο θόρυβος δεν είναι κατά Gauss. Συνεπώς έχει ενδιαφέρον να μελετήσουμε αυτή την συγκεκριμένη περίπτωση λεπτομερέστερα.

Ας υποθέσουμε, λοιπόν, ότι τα σήματα \underline{s}_j λαμβάνουν γνωστές τιμές $\underline{s}_j \in \mathbb{R}^n$ και ότι το διάνυσμα θορύβου \underline{N} είναι ένα τυχαίο κατά Gauss διάνυσμα με μέσο διάνυσμα $\underline{0}$ και πίνακα συνδιακύμανσης $\underline{\Sigma}_N$. (Η παραδοχή του μηδενικού μέσου θορύβου δεν μειώνει τη γενικότητα αυτών των αποτελεσμάτων, καθώς μπορούμε πάντα να αφαιρέσουμε έναν μη μηδενικό μέσο θόρυβο από το \underline{y} για να παράγουμε μια νέα παρατήρηση με μηδενικό μέσο θόρυβο)

Ας θυμηθούμε ότι ένα τυχαίο κατά Gauss διάνυσμα στο \mathbb{R}^n με μέσο διάνυσμα $\underline{\mu} \triangleq E\{\underline{X}\} \in \mathbb{R}^n$ και $n \times n$ πίνακα συνδιακύμανσης $\underline{\Sigma} \triangleq E\{(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T\}$ είναι αυτό με συνάρτηση πυκνότητας πιθανότητας

$$p_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2} |\underline{\Sigma}|^{1/2}} \quad (\text{III.B.22})$$

όπου το $|\underline{\Sigma}|$ συμβολίζει την ορίζουσα του $\underline{\Sigma}$ και όπου το $\underline{\Sigma}^{-1}$ συμβολίζει το αντίστροφο του $\underline{\Sigma}$. Συμβολίζουμε μια τέτοια τυχαία μεταβλητή ως $N(\underline{\mu}, \underline{\Sigma})$ τυχαία μεταβλητή. Σημειώστε ότι ένας πίνακας συνδιακύμανσης είναι πάντα μη αρνητικό όρισμα (δηλ. $\underline{x}^T \underline{\Sigma} \underline{x} \geq 0$ με $\underline{\tilde{x}} = \underline{\Sigma}_N^{-1}(\underline{s}_1 - \underline{s}_0)$) και άρα η δομή του ανιχνευτή εδώ είναι όμοια με τον ανιχνευτή συσχέτισης του Σχ. III.B. 2(b), με το πραγματικό σήμα \underline{s} να αντικαθίσταται με το «ψευδοσήμα» $\underline{\tilde{x}}$. Άρα γι' αυτήν την κατά Gauss περίπτωση, η εφαρμογή του ανιχνευτή δεν είναι πιο δύσκολη για εξαρτημένο θόρυβο απ' ό,τι για τον ανεξάρτητο θόρυβο. Περισσότερες πλευρές της δομής του III.B.24 θα εξεταστούν παρακάτω. Ωστόσο, έχει ενδιαφέρον πρώτα να εξετάσουμε την απόδοση του (III.B.24). Ας σημειωθεί ότι η ποσότητα

$T(\underline{Y}) \triangleq (\underline{s}_1 - \underline{s}_0)^T \underline{\Sigma}_N^{-1} \underline{Y}$ είναι ένας γραμμικός μετασχηματισμός του κατά Gauss τυχαίου διανύσματος \underline{Y} . Μια βασική ιδιότητα της πολυμεταβλητής κατά Gauss κατανομής είναι ότι οι γραμμικοί μετασχηματισμοί των κατά Gauss διανυσμάτων είναι επίσης κατά Gauss. Στην περίπτωση αυτή, ο μετασχηματισμός είναι στο \mathbb{R} , ώστε το $T(\underline{Y})$ είναι μια κατά Gauss τυχαία μεταβλητή, και επομένως μπορούμε να χαρακτηρίσουμε τις κατανομές του ως προς H_0 και H_1 απολύτως, βρίσκοντας τα μέσα και τις μεταβλητές ως προς τις δύο υποθέσεις. Ως προς το H_j , το μέσο του $T(\underline{Y})$ δίνεται από [με $\underline{\tilde{x}} = \underline{\Sigma}_N^{-1}(\underline{s}_1 - \underline{s}_0)$]

$$\begin{aligned}
E\{T(\underline{Y})|H_j\} &= E\{\bar{\underline{s}}^T \underline{Y}|H_j\} = \bar{\underline{s}}^T E\{\underline{Y}|H_j\} \\
&= \bar{\underline{s}}^T E\{\underline{N}|H_j\} + \bar{\underline{s}}^T \underline{s}_j = \bar{\underline{s}}^T \underline{s}_j \stackrel{\Delta}{=} \bar{\mu}_j. \quad (\text{III.B.26})
\end{aligned}$$

Παρομοίως, η μεταβλητή του $T(\underline{Y})$ ως προς το H_j είναι

$$\begin{aligned}
\text{Var}(T(\underline{Y})|H_j) &= E\{(\bar{\underline{s}}^T \underline{Y} - \bar{\underline{s}}^T \underline{s}_j)^2|H_j\} = E\{(\bar{\underline{s}}^T \underline{N})^2\} \\
&= E\{\bar{\underline{s}}^T \underline{N} \underline{N}^T \bar{\underline{s}}\} = \bar{\underline{s}}^T E\{\underline{N} \underline{N}^T\} \bar{\underline{s}} \\
&= \bar{\underline{s}}^T \underline{\Sigma}_N \bar{\underline{s}} = (\underline{s}_1 - \underline{s}_0)^T \underline{\Sigma}_N^{-1} (\underline{s}_1 - \underline{s}_0) \stackrel{\Delta}{=} d^2. \quad (\text{III.B.27})
\end{aligned}$$

Ας σημειωθεί ότι η μεταβλητή του $T(\underline{Y})$ είναι ανεξάρτητη από την υπόθεση. Επίσης σημειώστε ότι η θετική οριστικότητα του $\underline{\Sigma}_N$ υπονοεί ότι το $\underline{\Sigma}_N^{-1}$ είναι επίσης θετικό όρισμα και ότι άρα $d^2 > 0$ εκτός κι αν $\underline{s}_1 = \underline{s}_0$.

Από την παραπάνω ανάλυση βλέπουμε ότι $T(\underline{Y}) \sim \mathcal{N}(\bar{\mu}_j, d^2)$ ως προς το H_j για $j = 0, 1$. Αυτό σημαίνει, μεταξύ άλλων, ότι η τυχαιοποίηση γ στο (III.B.24) είναι άσχετη. Η πιθανότητα επιλογής του H_1 ως προς το H_j δίνεται από

$$\begin{aligned}
P_j(\Gamma_1) &= \frac{1}{\sqrt{2\pi d}} \int_{\tau'}^{\infty} e^{-(x - \bar{\mu}_j)^2 / 2d^2} dx \\
&= 1 - \Phi\left(\frac{\tau' - \bar{\mu}_j}{d}\right) \quad (\text{III.B.28})
\end{aligned}$$

με τ' από το (III.B.24) και όπου d είναι η θετική τετραγωνική ρίζα του d^2 . Για το Μπαϊεσιανό πρόβλημα το (III.B.28) είναι βολικό να γραφεί με τους όρους του αρχικού καταωφλιού τ , περίπτωση στην οποία έχουμε

$$P_j(\Gamma_1) = \begin{cases} 1 - \Phi\left(\frac{\log \tau}{d} + \frac{d}{2}\right) & \text{for } j = 0 \\ 1 - \Phi\left(\frac{\log \tau}{d} - \frac{d}{2}\right) & \text{for } j = 1. \end{cases} \quad (\text{III.B.29})$$

Συγκρίνοντας τα (III.B.29) και (II.B.31) βλέπουμε ότι η Μπαϊεσιανή επίδοση στο υπό εξέταση πρόβλημα εδώ είναι πανομοιότυπη με αυτή στο απλό πρόβλημα ελέγχου θέσης του Παραδείγματος II.B.2 με το d που προσδιορίζεται στο (II.B.31) να είναι η περίπτωση $n = 1$ του d που προσδιορίζεται στο (III.B.27). Παρομοίως, η επιλογή επίδοσης και καταωφλιού για το ελαχιστομέγιστο και το κατά Neyman – Pearson πρόβλημα εδώ είναι το ίδιο με τις βαθμωτές περιπτώσεις των Παραδειγμάτων II.C.1 και II.D.1, αντίστοιχα, με την ταύτιση του

$\tilde{\mu}_j$ και του d του (III.B.28) με το μ_j και σ του (II.B.30). Για παράδειγμα, για έναν έλεγχο α -επιπέδου Neyman – Pearson θέτουμε $P_F(\delta_0) = P_0(\Gamma_1) = \alpha$ για να λάβουμε το κατώφλι του

$$\tau' = d\Phi^{-1}(1 - \alpha) + \tilde{\mu}_0, \quad (\text{III.B.30})$$

και η αντίστοιχη πιθανότητα ανίχνευσης γίνεται

$$\begin{aligned} P_D(\tilde{\delta}_{NP}) &= P_1(\Gamma_1) = 1 - \Phi\left(\frac{\tau' - \tilde{\mu}_1}{d}\right) \\ &= 1 - \Phi(\Phi^{-1}(1 - \alpha) - d). \end{aligned} \quad (\text{III.B.31})$$

Άρα, συγκρίνοντας τα (III.B.31) και (II.D.16), βλέπουμε ότι οι καμπύλες ισχύος και τα ROCs γι' αυτό το γενικό πρόβλημα είναι αυτά των Σχ. II.D.3 και II.D.4.

Στις παρακάτω παρατηρήσεις, εξετάζουμε διάφορα ενδιαφέροντα χαρακτηριστικά αυτού του γενικού προβλήματος της ανίχνευσης συνεκτικών σημάτων σε κατά Gauss θόρυβο.

Παρατήρηση III.B.1: Ερμηνεία του d^2 .

Από την παραπάνω συζήτηση βλέπουμε ότι η επίδοση της βέλτιστης ανίχνευσης των ντετερμινιστικών σημάτων σε κατά Gauss θόρυβο γενικά βελτιώνεται μονοτονικά με την αύξηση του d . Όπως είδαμε στο Παράδειγμα II.D.2 αυτή η ποσότητα (ή πιο σωστά το τετράγωνό της) μπορεί να ερμηνευθεί ως μια μέτρηση του λόγου σήματος προς θόρυβο. Για να το κατανοήσουμε αυτό, ας εξετάσουμε πρώτα την περίπτωση που αναφέρεται στο Παράδειγμα III.B.1, στο οποίο τα σήματα είναι $\underline{s}_0 = \underline{0}$ και $\underline{s}_1 = \underline{s}$ και ο θόρυβος είναι i.i.d. $N(0, \sigma^2)$ που αντιστοιχούν στην πολυμεταβλητή κατά Gauss περίπτωση με $\Sigma_N = \sigma^2 \mathbf{I}$ όπου το \mathbf{I} συμβολίζει τον $n \times n$ ταυτοτικού πίνακα. Δηλ. ,

$$\Sigma_N = \begin{pmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & \sigma^2 & 0 \\ 0 & \dots & 0 & 0 & \sigma^2 \end{pmatrix}. \quad (\text{III.B.32})$$

Στην περίπτωση αυτή $\Sigma_N^{-1} = \sigma^{-2} \mathbf{I}$, οπότε το d^2 γίνεται

$$\begin{aligned}
d^2 &= (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0) = \frac{\underline{s}^T \mathbf{I} \underline{s}}{\sigma^2} = \frac{\underline{s}^T \underline{s}}{\sigma^2} \\
&= \frac{1}{\sigma^2} \sum_{k=1}^n s_k^2 = n \frac{\overline{s^2}}{\sigma^2}, \tag{III.B.33}
\end{aligned}$$

όπου το $\overline{s^2} \triangleq (1/n) \sum_{k=1}^n s_k^2$ είναι η μέση ισχύς σήματος. Σημειώστε ότι το $\sigma^2 = (1/n) \sum_{k=1}^n E\{N_k^2\}$ είναι η μέση ισχύς θορύβου, οπότε το d^2 εδώ δίνεται από το μέσο λόγο ισχύος σήματος προς θόρυβο επί των αριθμό των δειγμάτων.

Επομένως η επίδοση ενισχύεται με την αύξηση οποιουδήποτε από αυτά τα μεγέθη και καθώς κάποιο από τα δύο αυξάνεται χωρίς φράγμα, μπορεί να επιτευχθεί η τέλεια επίδοση.

Μια παρόμοια ερμηνεία μπορεί να δοθεί στο d^2 στη μη i.i.d. περίπτωση με $\underline{s}_0 = \underline{0}$ και $\underline{s}_1 = \underline{s}$. Ειδικότερα, όπως στο (III.B.10), μπορούμε να γράψουμε το μέγεθος

$\sum_{k=1}^n \tilde{s}_k y_k$ ως την είσοδο σε χρόνο n ενός γραμμικού χρονικά αναλλοίωτου φίλτρου με κρουστική απόκριση

$$\tilde{h}_k = \begin{cases} \tilde{s}_{n-k}, & 0 \leq k \leq n-1 \\ 0, & \text{otherwise.} \end{cases} \tag{III.B.34}$$

Αν η είσοδος στο φίλτρο αυτό αποτελούταν από το σήμα s_1, \dots, s_n μόνο, τότε η έξοδος σε χρόνο n θα ήταν

$$\sum_{k=1}^n \tilde{s}_k s_k = \underline{s}^T \Sigma_N^{-1} \underline{s} = d^2.$$

Επομένως η ισχύς εξόδου στο χρόνο δειγματοληψίας λόγω του σήματος μόνο είναι d^4 . Από την άλλη, αν ο θόρυβος έμπαινε μόνο σ' αυτό το φίλτρο, η έξοδος σε χρόνο n θα ήταν $\sum_{k=1}^n \tilde{s}_k N_k$, ένα τυχαίο μέγεθος με ισχύ

$$E \left\{ \left(\sum_{k=1}^n \tilde{s}_k N_k \right)^2 \right\} = E \{ (\tilde{\underline{s}}^T \underline{N})^2 \} = d^2, \tag{III.B.35}$$

όπου χρησιμοποιήσαμε το (III.B.27). Άρα ο λόγος της ισχύος εξόδου του \tilde{h}_k λόγω του σήματος μόνο ως προς αυτό λόγω του θορύβου μόνο είναι

$$\frac{(\sum_{k=1}^n \tilde{s}_k s_k)^2}{E \left\{ (\sum_{k=1}^n \tilde{s}_k N_k)^2 \right\}} = \frac{d^4}{d^2} = d^2. \quad (\text{III.B.36})$$

Επομένως το μέγεθος d^2 στη γενική περίπτωση είναι ο λόγος ισχύος σήματος προς θόρυβο στην έξοδο του φίλτρου που χρησιμοποιείται για βέλτιστη ανίχνευση στο χρόνο δειγματοληψίας n . Είναι επομένως διαισθητικά λογικό ότι όσο μεγαλύτερη είναι αυτή η έξοδος SNR τόσο καλύτερα μπορεί να ανιχνευθεί το σήμα, συγκρίνοντας την έξοδο δειγματοληψίας μ' ένα κατώφλι και η διαίσθηση αυτή επιβεβαιώνεται από τη μονοτονία της απόδοσης ανίχνευσης ως μια συνάρτηση του d^2 που φαίνεται παραπάνω.

Είναι ενδιαφέρον να σημειωθεί ότι το φίλτρο \tilde{h}_k του (III.B.34) έχει μέγιστη αναλογία σήματος εξόδου προς θόρυβο σε χρόνο n ανάμεσα σε όλα τα γραμμικά φίλτρα με κρουστική απόκριση μήκους n (Βλέπε Άσκηση 1). Το αποτέλεσμα αυτό βασίζεται μόνο στο δεδομένο ότι \underline{N} είναι μηδενικό μέσο με συμμεταβλητή Σ_N , και άρα είναι αληθές ακόμη και για μη κατά Gauss θόρυβο. Ωστόσο, η βελτιστότητα του (III.B.34) για τον μπαϊεσιανό, ελαχιστομέγιστο και Neyman – Pearson έλεγχο εξαρτάται σε μεγάλο βαθμό από την παραδοχή του κατά Gauss θορύβου.

Το μέγεθος d^2 έχει μια άλλη ερμηνεία για την περίπτωση i.i.d. με γενικά σήματα. Στην περίπτωση αυτή μπορούμε να γράψουμε

$$d^2 = \frac{1}{\sigma^2} \|\underline{s}_1 - \underline{s}_0\|^2, \quad (\text{III.B.37})$$

Όπου το $\|\underline{s}_1 - \underline{s}_0\|$ συμβολίζει την Ευκλείδεια απόσταση μεταξύ των διανυσμάτων σήματος \underline{s}_0 και \underline{s}_1 που δίνεται από

$$\|\underline{s}_1 - \underline{s}_0\| = \left[\sum_{k=1}^n (s_{1k} - s_{0k})^2 \right]^{1/2}.$$

Άρα όσο μακρύτερα βρίσκονται μεταξύ τους τα διανύσματα σήματος, τόσο καλύτερη απόδοση μπορεί να επιτευχθεί. Μια παρόμοια ερμηνεία μπορεί να δοθεί για την περίπτωση μη i.i.d. θορύβου, όπως θα αναφερθεί παρακάτω.

Παρατήρηση III.B.2.: Αναγωγή στην περίπτωση i.i.d. θορύβου

Αφού το Σ_N είναι ένας $n \times n$ συμμετρικός θετικά ορισμένος πίνακας έχει αρκετές δομικές ιδιότητες που μπορούν να εξεταστούν για να δώσουν στοιχεία για τη δομή του βέλτιστου

συστήματος ανίχνευσης. Οι ιδιοτιμές $\lambda_1, \dots, \lambda_n$ και τα αντίστοιχα ιδιοδιανύσματα $\underline{v}_1, \dots, \underline{v}_n$ ενός $n \times n$ πίνακα Σ_N είναι οι λύσεις στην εξίσωση $\Sigma_N \underline{v}_k = \lambda_k \underline{v}_k$. (Το σύνολο των ιδιοτιμών ενός πίνακα είναι μοναδικό αλλά όχι και το σύνολο των ιδιοδιανυσμάτων). Καθώς το Σ_N στην περίπτωση μας είναι συμμετρικό και θετικό όρισμα, όλες οι ιδιοτιμές του είναι αληθείς και θετικές και τα ιδιοδιανύσματά του μπορούν να επιλεγούν για να είναι ορθοκανονικά (δηλ. $\underline{v}_k^T \underline{v}_l = 0$ αν $k \neq l$ και $\underline{v}_k^T \underline{v}_k = 1$, για όλα τα $l, k = 1, \dots, n$). Με αυτή την επιλογή ιδιοδιανυσμάτων μπορούμε να γράψουμε το Σ_N ως

$$\Sigma_N = \sum_{k=1}^n \lambda_k \underline{v}_k \underline{v}_k^T. \quad (\text{III.B.38})$$

Η εξίσωση του (III.B.38) καλείται φασματική παραγοντοποίηση του Σ_N και η ισχύς της έπεται εύκολα από το γεγονός ότι το ορθοκανονικό σύνολο $\underline{v}_1, \dots, \underline{v}_n$ δημιουργεί μια βάση για το \mathbf{R}^n (όπως κάθε σύνολο των n γραμμικά ανεξάρτητων διανυσμάτων στο \mathbf{R}^n).² Ας σημειωθεί ότι ο πίνακας $\underline{v}_k \underline{v}_k^T$ όταν πολλαπλασιαστεί με ένα διάνυσμα \underline{x} δίνει την προβολή

του x στο v_k . Χρησιμοποιώντας το (III.B.38) έπεται άμεσα ότι $\Sigma_N^{-1} = \sum_{k=1}^n \lambda_k^{-1} \underline{v}_k \underline{v}_k^T$, από το οποίο η στατιστική βέλτιστης ανίχνευσης $T(\underline{y})$ δίνεται από

$$T(\underline{y}) = (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} \underline{y} = \sum_{k=1}^n (\hat{s}_{1k} - \hat{s}_{0k}) \hat{y}_k, \quad (\text{III.B.39})$$

όπου

$$\hat{y}_k = \underline{v}_k^T \underline{y} / \sqrt{\lambda_k}, \quad k = 1, \dots, n \quad (\text{III.B.40})$$

και

$$\hat{s}_{jk} = \underline{v}_k^T \underline{s}_j / \sqrt{\lambda_k}, \quad k = 1, \dots, n \text{ and } j = 0, 1.$$

2

Ας σημειωθεί ότι το \underline{y} μπορεί να ληφθεί από το $\hat{\underline{y}}$ μέσω του $\underline{y} = \sum_{k=1}^n \sqrt{\lambda_k} \hat{y}_k \underline{v}_k$, οπότε το $\hat{\underline{y}}$ είναι μια ισοδύναμη παρατήρηση του \underline{y} . Σε ό,τι αφορά το $\hat{\underline{Y}}$ (το τυχαίο διάνυσμα που αντιστοιχεί στο $\hat{\underline{y}}$, το ζεύγος υπόθεσης του III.B.1) γίνεται

² Πράγματι, για κάθε $\underline{x} \in \mathbf{R}^n$, μπορούμε να γράψουμε $\underline{x} = \sum_{k=1}^n c_k \underline{v}_k$ με $c_k = \underline{v}_k^T \underline{x}$, οπότε έχουμε

$$\Sigma_N \underline{x} = \sum_{k=1}^n c_k \Sigma_N \underline{v}_k = \sum_{k=1}^n \lambda_k \underline{v}_k c_k = \sum_{k=1}^n \lambda_k \underline{v}_k \underline{v}_k^T \underline{x} = \left(\sum_{k=1}^n \lambda_k \underline{v}_k \underline{v}_k^T \right) \underline{x}.$$

$$\begin{aligned} H_0 : \hat{\mathbf{Y}} &= \hat{\mathbf{N}} + \hat{\mathbf{s}}_0 \\ \text{versus} \\ H_1 : \hat{\mathbf{Y}} &= \hat{\mathbf{N}} + \hat{\mathbf{s}}_1, \end{aligned} \quad (\text{III.B.41})$$

όπου $\hat{N}_k = \mathbf{v}_k^T \mathbf{N} / \sqrt{\lambda_k}$. Ας σημειωθεί ότι το $\hat{\mathbf{N}}$ είναι ένα κατά Gauss τυχαίο διάνυσμα, καθώς είναι ένας γραμμικός μετασχηματισμός του \mathbf{N} . Επίσης,

$$\begin{aligned} E\{\hat{N}_k \hat{N}_l\} &= E\{\mathbf{v}_k^T \mathbf{N} \mathbf{v}_l^T \mathbf{N}\} / \sqrt{\lambda_k \lambda_l} \\ &= E\{\mathbf{v}_k^T \mathbf{N} \mathbf{N}^T \mathbf{v}_l\} / \sqrt{\lambda_k \lambda_l} \\ &= \mathbf{v}_k^T E\{\mathbf{N} \mathbf{N}^T\} \mathbf{v}_l / \sqrt{\lambda_k \lambda_l} \\ &= \mathbf{v}_k^T \boldsymbol{\Sigma}_N \mathbf{v}_l / \sqrt{\lambda_k \lambda_l} = \mathbf{v}_k^T \mathbf{v}_l \sqrt{\lambda_l / \lambda_k}, \end{aligned} \quad (\text{III.B.42})$$

όπου έχουμε χρησιμοποιήσει το δεδομένο ότι το \mathbf{v}_l είναι ένα ιδιοδιάνυσμα του $\boldsymbol{\Sigma}_N$. Από την ορθοκανονικότητα του $\mathbf{v}_1, \dots, \mathbf{v}_n$, το (III.B.42) συνεπάγεται ότι

$$E\{\hat{N}_k \hat{N}_l\} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l, \end{cases} \quad (\text{III.B.43})$$

άρα τα $\hat{N}_1, \dots, \hat{N}_n$ είναι i.i.d. $N(0,1)$ τυχαίες μεταβλητές. Επομένως, με τον κατάλληλο γραμμικό μετασχηματισμό του $\hat{\mathbf{Y}}$ έχουμε μετατρέψει ένα πρόβλημα με εξαρτημένο κατά Gauss θόρυβο σ' ένα ισοδύναμο πρόβλημα με i.i.d. κατά Gauss θόρυβο. Φυσικά, το (III.B.39) δίνει τη στατιστική βέλτιστης ανίχνευσης γι' αυτό το τροποποιημένο πρόβλημα.

Βασικά, αυτό που κάναμε παραπάνω ήταν να αλλάξουμε από το αρχικό τυπικό σύστημα συντεταγμένων στο \mathbb{R}^n και να έχουμε ένα διαφορετικό σύστημα συντεταγμένων, στο οποίο οι συνήθεις άξονες ευθυγραμμίζονται με τα διανύσματα $\mathbf{v}_1, \dots, \mathbf{v}_n$. Στο σύστημα αυτό οι συντεταγμένες θορύβου $\hat{N}_1, \dots, \hat{N}_n$ είναι ανεξάρτητες. Ένας άλλος τρόπος θεώρησης αυτής της αλλαγής συντεταγμένων είναι να γράψουμε $\boldsymbol{\Sigma}_N = \mathbf{B}^2$, όπου \mathbf{B} είναι ο πίνακας

$\sum_{k=1}^n \lambda_k^{1/2} \mathbf{v}_k \mathbf{v}_k^T$. (Το \mathbf{B} καλείται τετραγωνική ρίζα του $\boldsymbol{\Sigma}_N$). Αυτός ο πίνακας \mathbf{B} έχει αντίστροφο $\mathbf{B}^{-1} = \sum_{k=1}^n \lambda_k^{-1/2} \mathbf{v}_k \mathbf{v}_k^T$, και $\boldsymbol{\Sigma}_N^{-1} = \mathbf{B}^{-2} \triangleq (\mathbf{B}^{-1})^2$. Αν ορίσουμε $\mathbf{s}_j^* = \mathbf{B}^{-1} \mathbf{s}_j$ και $\mathbf{y}^* = \mathbf{B}^{-1} \mathbf{y}$, έχουμε ότι

$$(\mathbf{s}_1 - \mathbf{s}_0)^T \boldsymbol{\Sigma}_N^{-1} \mathbf{y} = (\mathbf{s}_1^* - \mathbf{s}_0^*)^T \mathbf{y}^*. \quad (\text{III.B.44})$$

Επιπλέον, ως προς το H_j , μπορούμε να γράψουμε $\underline{Y}^* = \underline{N}^* + \underline{s}_j$ με $\underline{N}^* = \mathbf{B}^{-1}\underline{N}$, και

$$\begin{aligned} E\{\underline{N}^*(\underline{N}^*)^T\} &= E\{\mathbf{B}^{-1}\underline{N}\underline{N}^T\mathbf{B}^{-1}\} = \mathbf{B}^{-1}E\{\underline{N}\underline{N}^T\}\mathbf{B}^{-1} \\ &= \mathbf{B}^{-1}\Sigma_N\mathbf{B}^{-1} = \mathbf{B}^{-1}\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}. \end{aligned}$$

Άρα N_1^*, \dots, N_n^* είναι i.i.d. $N(0,1)$ μεταβλητές και έχουμε μια κατάσταση παρόμοια με αυτή των παρατηρήσιμων \hat{Y} που είδαμε παραπάνω. Στην ουσία, τα \underline{Y}^* και \hat{Y} είναι τα ίδια τυχαία διανύσματα σε δύο διαφορετικά συστήματα συντεταγμένων αφού $\underline{Y}^* = \sum_{k=1}^n \hat{Y}_k \underline{v}_k$ και $\hat{Y} = \sum_{k=1}^n \hat{Y}_k \underline{e}_k$, όπου τα $\underline{e}_1, \dots, \underline{e}_n$ είναι τα τυπικά διανύσματα βάσης για το \mathbf{R}^n , δηλ., το \underline{e}_k είναι όλα 0 εκτός από ένα 1 στο k th στοιχείο του.

Το διάνυσμα παρατήρησης \underline{Y} μπορεί να μετατραπεί και με έναν άλλον ενδιαφέροντα τρόπο για να δώσει μια ισοδύναμη παρατήρηση με θόρυβο i.i.d. Ειδικότερα, καθώς το Σ_N είναι θετικό όρισμα μπορεί να γραφεί ως

$$\Sigma_N = \mathbf{C}\mathbf{C}^T \quad (\text{III.B.45})$$

όπου \mathbf{C} είναι ένας $n \times n$ αντιστρέψιμος κατώτερος τριγωνικός πίνακας (δηλ, όλα τα παραπάνω διαγώνια στοιχεία του \mathbf{C} είναι μηδέν). Η εξίσωση του (III.B.45) ονομάζεται παραγοντοποίηση Cholesky του Σ_N , και υπάρχουν αρκετοί τυπικοί αλγόριθμοι για την εύρεση του \mathbf{C} από το Σ_N (βλέπε, π.χ., Bierman (1977)). Μπορούμε τότε να γράψουμε

$\Sigma_N^{-1} = (\mathbf{C}^T)^{-1}\mathbf{C}^{-1} = (\mathbf{C}^{-1})^T\mathbf{C}^{-1}$. Ορίζοντας νέα παρατηρούμενα $\tilde{Y} = \mathbf{C}^{-1}\underline{Y} = \mathbf{C}^{-1}\underline{N} + \mathbf{C}^{-1}\underline{s}_j \triangleq \tilde{N} + \tilde{s}_j$ έχουμε άμεσα ότι $\tilde{N} \sim \mathcal{N}(\underline{0}, \mathbf{I})$. Άρα μπορούμε και πάλι να έχουμε μια κατάσταση i.i.d. θορύβου και η στατιστική βέλτιστης ανίχνευσης είναι $(\tilde{s}_1 - \tilde{s}_0)^T \tilde{Y}$.

Το ενδιαφέρον σ' αυτή την ιδιαίτερη μετατροπή είναι ότι η κατώτερη τριγωνικότητα του \mathbf{C} συνεπάγεται ότι και το \mathbf{C}^{-1} είναι επίσης κατώτερα τριγωνικό. Αυτό με τη σειρά του συνεπάγεται ότι μπορούμε να γράψουμε

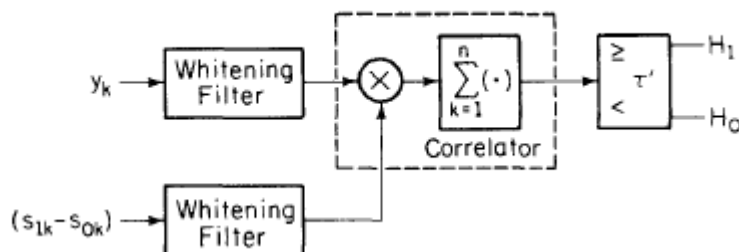
$$\bar{y}_k = \sum_{l=1}^k h_{k,l} y_l, \quad (\text{III.B.46})$$

Όπου το $h_{k,l}$ είναι το $k-l$ στοιχείο του C^{-1} . Σημειώστε ότι το (III.B.46) είναι ένας αιτιώδης μετασχηματισμός και στην ουσία το (III.B.46) δείχνει ότι το $\bar{y}_1, \dots, \bar{y}_n$ μπορεί να παραχθεί από ένα αιτιώδες, αλλά πιθανόν χρονικά μεταβαλλόμενο φιλτράρισμα του y_1, \dots, y_n . Καθώς ο θόρυβος στην έξοδο αυτού του φίλτρου είναι λευκός (δηλ., i.i.d.), το φίλτρο αυτό είναι γνωστό ενίοτε ως *λευκαντικό φίλτρο*. Άρα η βέλτιστη δομή ανιχνευτή του (III.B.24) μπορεί να απεικονισθεί ως το αιτιώδες γραμμικό φίλτρο με ωστική απόκριση $\{h_{k,l}\}$ οδηγούμενο από το y_1, \dots, y_n και ακολουθούμενο από έναν συσχετιστή, στον οποίο η έξοδος φίλτρου συσχετίζεται με το $(\bar{s}_{11} - \bar{s}_{01}), \dots, (\bar{s}_{1n} - \bar{s}_{0n})$, την έξοδο του ίδιου φίλτρου οδηγούμενο από το διαφορετικό σήμα $(s_{11} - s_{01}), \dots, (s_{1n} - s_{0n})$. Η δομή αυτή απεικονίζεται στο Σχ. III.B.8.

Ως τελικό σχόλιο σημειώνουμε ότι ο λόγος σήματος προς θόρυβο

$d^2 = (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0)$ μπορεί να γραφεί, σε ό,τι αφορά οποιοδήποτε από τα ζεύγη μετασχηματισμένων σημάτων, ως

$$d^2 = \|\hat{\underline{s}}_1 - \hat{\underline{s}}_0\|^2 = \|\underline{s}_1^* - \underline{s}_0^*\|^2 = \|\bar{\underline{s}}_1 - \bar{\underline{s}}_0\|^2. \quad (\text{III.B.47})$$



ΣΧΗΜΑ III.B.8. Βέλτιστος ανιχνευτής για συνεκτικά σήματα σε εξαρτημένο κατά Gauss θόρυβο. (Whitening filter = Φίλτρο λεύκανσης, Correlator = συσχετιστής)

Επομένως, η απόδοση της συνεκτικής ανίχνευσης σε εξαρτημένο θόρυβο εξαρτάται από το πόσο απέχουν τα σήματα όταν μετασχηματίζονται σ' ένα σύστημα συντεταγμένων στο οποίο τα στοιχεία θορύβου είναι i.i.d. [Συγκρίνετε με το (III.B.37)]. Θα πρέπει να σημειωθεί ότι και τα τρία ζεύγη σημάτων στο (III.B.47) έχουν την ίδια απόσταση γιατί είναι όλα απεικονίσεις του ίδιου ζεύγους διανυσμάτων σε διαφορετικά συστήματα συντεταγμένων που είναι απλές περιστροφές το ένα του άλλου.

Παρατήρηση III.B.3: Επιλογή σήματος

Η επίδοση της βέλτιστης συνεκτικής ανίχνευσης σε κατά Gauss θόρυβο βελτιώνεται με την αύξηση του μεγέθους $d^2 \triangleq (\underline{s}_1 - \underline{s}_0)^T \Sigma_N^{-1} (\underline{s}_1 - \underline{s}_0)$. Σε πολλές εφαρμογές στις οποίες προκύπτει συνεκτική ανίχνευση, υπάρχει συχνά κάποια ευελιξία στην επιλογή των σημάτων \underline{s}_0 και \underline{s}_1 . Σε τέτοιες καταστάσεις είναι λογικό να επιλεγούν τα σήματα αυτά για τη μεγιστοποίηση του d^2 .

Όπως σημειώθηκε στην προηγούμενη συζήτηση μπορούμε να γράψουμε

$\Sigma_N^{-1} = \sum_{k=1}^n \lambda_k^{-1} \underline{v}_k \underline{v}_k^T$, όπου τα $\lambda_1, \dots, \lambda_n$ και $\underline{v}_1, \dots, \underline{v}_n$ είναι οι ιδιοτιμές και τα αντίστοιχα ορθοκανονικά ιδιοδιανύσματα του Σ_N . Για κάθε διάνυσμα $\underline{x} \in \mathbb{R}^n$, έχουμε

$$\begin{aligned} \underline{x}^T \Sigma_N^{-1} \underline{x} &= \sum_{k=1}^n \lambda_k^{-1} \underline{x}^T \underline{v}_k \underline{v}_k^T \underline{x} \\ &\leq \lambda_{\min}^{-1} \sum_{k=1}^n \underline{x}^T \underline{v}_k \underline{v}_k^T \underline{x}, \end{aligned} \quad (\text{III.B.48})$$

όπου $\lambda_{\min} = \min\{\lambda_1, \dots, \lambda_n\}$. Αφού

$$\sum_{k=1}^n \underline{x}^T \underline{v}_k \underline{v}_k^T \underline{x} = \underline{x}^T \left(\sum_{k=1}^n \underline{v}_k \underline{v}_k^T \right) \underline{x} = \underline{x}^T \underline{x} = \|\underline{x}\|^2,$$

έχουμε ότι

$$\underline{x}^T \Sigma_N^{-1} \underline{x} \leq \lambda_{\min}^{-1} \|\underline{x}\|^2. \quad (\text{III.B.49})$$

Ας σημειωθεί ότι μπορούμε να έχουμε ισότητα στο (III.B.49) αν και μόνο αν το \underline{x} είναι ανάλογο με ένα ιδιοδιάνυσμα που να αντιστοιχεί στην ιδιοτιμή λ_{\min} . [Αν υπάρχουν περισσότερα από ένα ιδιοδιανύσματα που να αντιστοιχούν στο λ_{\min} , το \underline{x} μπορεί να είναι οποιοσδήποτε γραμμικός συνδυασμός αυτών και πάλι να έχουμε ισότητα στο (III.B.49). Οποιοσδήποτε τέτοιος γραμμικός συνδυασμός παραμένει ιδιοδιάνυσμα που αντιστοιχεί στο λ_{\min} .]

Από τα παραπάνω βλέπουμε ότι, για σταθερό $\|\underline{s}_1 - \underline{s}_0\|$, ο καλύτερος τρόπος επιλογής του διαφορετικού σήματος $\underline{s}_1 - \underline{s}_0$ είναι να βρίσκεται κατά μήκος ενός ιδιοδιανύσματος που

αντιστοιχεί στην ελάχιστη ιδιοτιμή του Σ_N . Οι ιδιοτιμές του Σ_N είναι μετρήσεις της ισχύος θορύβου στις κατευθύνσεις των αντίστοιχών τους ιδιοδιανυσμάτων. Επομένως, βάζοντας τη διαφορά σήματος κατά μήκος του ελάχιστης ιδιοτιμής ιδιοδιανύσματος είναι ισοδύναμο με τη σηματοδότηση προς τη λιγότερο θορυβώδη κατεύθυνση. Κάνοντας αυτό έχουμε μια τιμή του d^2 που δίνεται από

$$d^2 = \frac{1}{\lambda_{\min}} \|\underline{s}_1 - \underline{s}_0\|^2. \quad (\text{III.B.50})$$

Αφού έχουμε επιλέξει την κατεύθυνση της διαφοράς σήματος $\underline{s}_1 - \underline{s}_0$, μπορούμε να βελτιώσουμε περαιτέρω την απόδοση μεγιστοποιώντας το $\|\underline{s}_1 - \underline{s}_0\|^2$. Προφανώς, αυτό το μέγεθος μπορεί να είναι αυθαίρετα μεγάλο αν δεν θέσουμε περιορισμούς στα σήματα. Ωστόσο, τα σήματα συνήθως περιορίζονται από τη συνολική τους ισχύ και επομένως είναι ενδιαφέρον να μεγιστοποιούμε το (III.B.50) στο πλαίσιο ενός τέτοιου περιορισμού. Συγκεκριμένα, υποθέστε ότι περιορίζουμε $\|\underline{s}_1\|^2 \leq P$ και $\|\underline{s}_0\|^2 \leq P$, όπου $0 < P < \infty$.

Έχουμε

$$\begin{aligned} d^2 &= \frac{1}{\lambda_{\min}} (\underline{s}_1 - \underline{s}_0)^T (\underline{s}_1 - \underline{s}_0) \\ &= \frac{1}{\lambda_{\min}} (\|\underline{s}_1\|^2 - 2\underline{s}_1^T \underline{s}_0 + \|\underline{s}_0\|^2). \end{aligned} \quad (\text{III.B.51})$$

Σημειώστε ότι το $\underline{s}_1^T \underline{s}_0$ είναι το εσωτερικό προϊόν μεταξύ των \underline{s}_0 και \underline{s}_1 . Με σταθερά τα $\|\underline{s}_1\|$ και $\|\underline{s}_0\|$, το μέγεθος αυτό ελαχιστοποιείται (και άρα το d^2 μεγιστοποιείται) αν τα \underline{s}_1 και \underline{s}_0 βρίσκονται σε αντίθετες κατευθύνσεις. Δηλ. αν $\underline{s}_0 = \alpha \underline{s}_1$ με $\alpha < 0$. Στην περίπτωση αυτή έχουμε

$$\begin{aligned} d^2 &= \frac{1}{\lambda_{\min}} (\|\underline{s}_1\|^2 - 2\alpha \|\underline{s}_1\|^2 + \alpha^2 \|\underline{s}_1\|^2) \\ &= \frac{\|\underline{s}_1\|^2 (1 - \alpha)^2}{\lambda_{\min}}, \end{aligned} \quad (\text{III.B.52})$$

και το α πρέπει να προκύπτει από $\alpha = -\|\underline{s}_0\| / \|\underline{s}_1\|$. Άρα, για σταθερά $\|\underline{s}_0\|$ και $\|\underline{s}_1\|$, η μέγιστη τιμή του d^2 είναι

$$d^2 = \frac{(\|\underline{s}_1\| + \|\underline{s}_0\|)^2}{\lambda_{\min}}.$$

Βλέπουμε ότι το d^2 μεγιστοποιείται περαιτέρω επιλέγοντας $\|\underline{s}_0\|^2 = \|\underline{s}_1\|^2 = P$, περίπτωση στην οποία έχουμε $\alpha = -1$ και

$$\max_{\|\underline{s}_j\|^2 \leq P} d^2 = \frac{4P}{\lambda_{\min}}. \quad (\text{III.B.53})$$

Αφού έχουμε επιλέξει το $\underline{s}_1 - \underline{s}_0$ να βρίσκεται κατά μήκος ενός ελάχιστης ιδιοτιμής ιδιοδιανύσματος \underline{v}_{\min} , μπορούμε να πετύχουμε το (III.B.53) επιλέγοντας $\underline{s}_1 = c\underline{v}_{\min}$ και $\underline{s}_0 = -\underline{s}_1$, όπου το c επιλέγεται ώστε $\|\underline{s}_1\|^2 = \|\underline{s}_0\|^2 = P$. Το σωστό c είναι επομένως $\sqrt{P}/\|\underline{v}_{\min}\|$, οπότε τα βέλτιστα σήματα δίνονται από

$$\underline{s}_1 = \sqrt{P}\underline{v}_{\min}/\|\underline{v}_{\min}\| \quad \text{and} \quad \underline{s}_0 = -\underline{s}_1. \quad (\text{III.B.54})$$

Οι προηγούμενες έννοιες απεικονίζονται στο ακόλουθο απλό παράδειγμα

Παράδειγμα III.B.4: Βέλτιστα Σήματα για Ανίχνευση δύο δειγμάτων

Εξετάστε την περίπτωση $n = 2$ με

$$\Sigma_N = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad (\text{III.B.55})$$

όπου $|\rho| < 1$. Γι' αυτό το Σ_N , οι ιδιοτιμές και τα αντίστοιχα ορθοκανονικά ιδιοδιανύσματα φαίνεται εύκολα ότι δίνονται από

$$\begin{aligned} \lambda_1 &= \sigma^2(1 - \rho) \quad \text{and} \quad \lambda_2 = \sigma^2(1 + \rho) \\ \underline{v}_1 &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad \underline{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned} \quad (\text{III.B.56})$$

Άρα αν $\rho > 0$, $\lambda_{\min} = \lambda_1$ και τα βέλτιστα σήματα δίνονται από

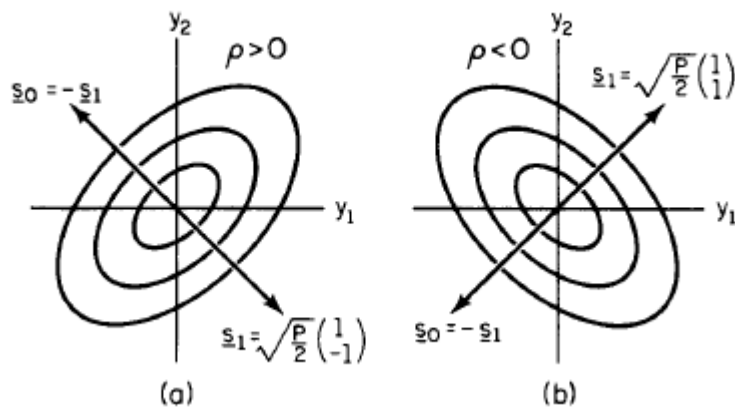
$$\underline{s}_1 = \sqrt{P/2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad \underline{s}_0 = \sqrt{P/2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad (\text{III.B.57})$$

Και αν $\rho < 0$, $\lambda_{\min} = \lambda_2$ και τα βέλτιστα σήματα δίνονται από

$$\underline{s}_1 = \sqrt{P/2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \underline{s}_0 = \sqrt{P/2} \begin{pmatrix} -1 \\ -1 \end{pmatrix}. \quad (\text{III.B.58})$$

Σε οποιαδήποτε από τις δύο περιπτώσεις, η μέγιστη τιμή του d^2 είναι

$$d^2 = \frac{4P}{\sigma^2(1 - |\rho|)}. \quad (\text{III.B.59})$$



ΣΧΗΜΑ III.B.9. Απεικόνιση των βέλτιστων σημάτων για κατά Gauss θόρυβο με

$$\Sigma_N = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Η βελτιστότητα των συνόλων σήματος για αυτές τις δύο περιπτώσεις αποδεικνύεται εύκολα από το Σχ. III.B.9, στο οποίο τα περιγράμματα ίσης πυκνότητας έχουν σχεδιαστεί για τις δύο περιπτώσεις. Ας σημειωθεί ότι και για τις δύο περιπτώσεις τα διανύσματα σήματος βρίσκονται στις κατευθύνσεις στις οποίες η πυκνότητα θορύβου μειώνεται ταχύτερα, δίνοντας έτσι έναν μέγιστο λόγο σήματος προς θόρυβο για σταθερά P , σ^2 και ρ . Είναι ενδιαφέρον να σημειωθεί ότι πρέπει μόνο να είναι γνωστό το αλγεβρικό σύμβολο του ρ , και όχι η πραγματική του τιμή, για να επιλεγθούν τα βέλτιστα σήματα στην περίπτωση αυτή.

Περίπτωση III.B.3: Ανίχνευση σημάτων με τυχαίες παραμέτρους

Στις περιπτώσεις III.B.1 και III.B.2 εξετάσαμε το πρόβλημα απόφασης μεταξύ δύο σημάτων που είναι απολύτως γνωστά στον παρατηρητή. Σε πολλές εφαρμογές, συχνά ερχόμαστε αντιμέτωποι με το σχετικό πρόβλημα απόφασης μεταξύ σημάτων που είναι γνωστά, με ένα σύνολο όμως αγνώστων παραμέτρων. Η κατάσταση αυτή προκύπτει, π.χ., σε ψηφιακά συστήματα επικοινωνίας στα οποία η μία από τις δύο κυματομορφές (που αναπαριστούν το «μηδέν» και το «ένα», αντίστοιχα) διαμορφώνεται πάνω σε έναν ημιτονοειδή φορέα και ο πομπός και ο δέκτης πρέπει να αποφασίσουν ποια από τις δύο στάλθηκε. Ακόμα κι αν οι δύο κυματομορφές και η συχνότητα του φορέα είναι γνωστά στο δέκτη, το πλάτος και η φάση του φορέα μπορεί να μην είναι. Και άρα αυτά τα δύο μεγέθη αποτελούν άγνωστες παραμέτρους

που πρέπει να ληφθούν υπόψη στη διαδικασία ανίχνευσης. Παρόμοιες καταστάσεις προκύπτουν στα ραντάρ, τα σόναρ και σε άλλες εφαρμογές, στις οποίες θα πρέπει να ανιχνευθούν σήματα άγνωστης συχνότητας, χρόνου άφιξης και πλάτους.

Γι' αυτή την κατάσταση είναι βολικό να γράψουμε το (III.B.1) ως

$$\begin{aligned} & H_0 : Y_k = N_k + s_{0k}(\theta), \quad k = 1, \dots, n \\ \text{versus} & H_1 : Y_k = N_k + s_{1k}(\theta), \quad k = 1, \dots, n \end{aligned} \quad (\text{III.B.60})$$

Όπου $s_0(\theta)$ και $s_1(\theta)$ είναι γνωστής τιμής διανύσματος συναρτήσεις του θ , που είναι μια άγνωστη παράμετρος που λαμβάνει τιμές σ' ένα σύνολο παραμέτρου Λ . Υποθέτοντας ότι το θ είναι τυχαίο (περίπτωση στην οποία το γράφουμε ως Θ) με πυκνότητα w_j ως προς την υπόθεση H_j , ο λόγος πιθανοφάνειας για το (III.B.60) είναι

$$\begin{aligned} L(\underline{y}) &= \frac{E_1\{p_N(\underline{y} - \underline{s}_1(\Theta))\}}{E_0\{p_N(\underline{y} - \underline{s}_0(\Theta))\}} \\ &= \frac{\int_{\Lambda} p_N(\underline{y} - \underline{s}_1(\theta)) w_1(\theta) \mu(d\theta)}{\int_{\Lambda} p_N(\underline{y} - \underline{s}_0(\theta)) w_0(\theta) \mu(d\theta)}. \end{aligned} \quad (\text{III.B.61})$$

Για τους σκοπούς της συζήτησης θα υποθέσουμε ότι $s_0(\theta) \equiv \underline{0}$ και $s_1(\theta) \triangleq \underline{s}(\theta)$, αφού μπορούμε να διαχειριστούμε παρόμοια άλλες περιπτώσεις. Στην περίπτωση αυτή έχουμε

$$\begin{aligned} L(\underline{y}) &= \int_{\Lambda} \frac{p_N(\underline{y} - \underline{s}(\theta))}{p_N(\underline{y})} w(\theta) \mu(d\theta) \\ &= \int_{\Lambda} L_{\theta}(\underline{y}) w(\theta) \mu(d\theta), \end{aligned} \quad (\text{III.B.62})$$

όπου $L_{\theta}(\underline{y})$ είναι ο λόγος πιθανοφάνειας με την υπόθεση ότι $\Theta = \theta$ και όπου έχουμε παραλείψει τον κάτω δείκτη από το w_1 . Από το (III.B.62) βλέπουμε ότι το $L(\underline{y})$ στην περίπτωση αυτή είναι απλώς ο μέσος (στο θ) λόγος πιθανοφάνειας για γνωστό θ . Με το θ γνωστό, το (III.B.60) είναι πρόβλημα ντετερμινιστικού σήματος και άρα η $L_{\theta}(\underline{y})$ βρίσκεται

άμεσα όπως στις περιπτώσεις III.B.1 και III.B.2. Για παράδειγμα, με δείγματα θορύβου i.i.d. $N(0, \sigma^2)$ η $L(\underline{y})$ γίνεται

$$L(\underline{y}) = \int_{\Lambda} \exp\{[\underline{s}^T(\theta)\underline{y} - \frac{1}{2} \|\underline{s}(\theta)\|^2]/\sigma^2\} w(\theta) \mu(d\theta). \quad (\text{III.B.63})$$

Παρομοίως, για μη i.i.d. κατά Gauss θόρυβο (III.B.63) ισχύει με $\sigma^2 = 1$ και με \underline{y} και $\underline{s}(\theta)$ να είναι μεγέθη που μετασχηματίστηκαν για να παράξουν ένα πρόβλημα i.i.d. θορύβου. Αυτός ο τύπος προβλήματος απεικονίζεται στο ακόλουθο παράδειγμα που προκύπτει σε σειρά εφαρμογών.

Παράδειγμα III.B.5: Μη συνεκτική ανίχνευση ενός διαμορφωμένου ημιτονοειδούς φορέα

Εξετάστε το ζεύγος σήματος $\underline{s}_0(\theta) = \underline{0}$ και $\underline{s}_1(\theta) = \underline{s}(\theta)$ με

$$s_k(\theta) = a_k \sin[(k-1)\omega_c T_S + \theta], \quad k = 1, \dots, n \quad (\text{III.B.64})$$

όπου a_1, a_2, \dots, a_n είναι μια γνωστή ακολουθία εύρους, θ είναι μια γωνία φάσης ανεξάρτητη από το θόρυβο και ομοιόμορφα κατανεμημένη στο $[0, 2\pi]$, και όπου ω_c και T_S είναι ένας γνωστός φορέας συχνότητας και διάστημα δειγματοληψίας με τη σχέση $n\omega_c T_S = m2\pi$ για κάποιο ακέραιο m (δηλ. υπάρχει ένας ακέραιος αριθμός περιόδων του ημιτονοειδούς στο χρονικό διάστημα $[0, nT_S]$). Επίσης υποθέτουμε ότι ο αριθμός των δειγμάτων που λήφθηκαν ανά κύκλο του ημιτονοειδούς (δηλ. n/m) είναι ένας ακέραιος μεγαλύτερος του 1. Τα σήματα αυτά παρέχουν ένα μοντέλο, για παράδειγμα, για ένα ψηφιακό σχήμα σηματοδότησης στο οποίο ένα «μηδέν» μεταδίδεται χωρίς να στέλνει τίποτα κατά τη διάρκεια του διαστήματος $[0, nT_S]$ και ένα «ένα» μεταδίδεται στέλνοντας ένα σήμα $a(t)$ διαμορφωμένο πάνω σ' ένα ημιτονοειδή φορέα συχνότητας ω_c . Αυτό το σχήμα σηματοδότησης είναι γνωστό ως on-off ψηφιοποίηση (OOK). Στην περίπτωση αυτή η ακολουθία a_1, \dots, a_n είναι η κυματομορφή δειγματοληψίας $a(t)$. [δηλ., τα $a_k = a((k-1)T_S)$ και θ αναπαριστούν τη γωνία φάσης του φορέα, που εδώ θεωρείται άγνωστη για το δέκτη]. Η ανίχνευση ενός διαμορφωμένου φορέα, στην οποία η φάση φορέα είναι άγνωστη στο δέκτη καλείται *μη συνεκτική ανίχνευση*. Η υπόθεση ότι η γωνία φάσης είναι ομοιόμορφη στο $[0, 2\pi]$ αφορά στην πεποίθηση ότι όλες οι φάσεις είναι εξίσου πιθανές να προκύψουν, μια υπόθεση που είναι λογική, αφού δεν υπάρχει κάποια άλλη πληροφορία για το αντίθετο.

Υποθέτοντας i.i.d. $N(0, \sigma^2)$ θόρυβο, ο λόγος πιθανοφάνειας για το πρόβλημα αυτό φαίνεται από το (III.B.63) να είναι

$$L(\underline{y}) = \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ \frac{1}{\sigma^2} \left(\sum_{k=1}^n y_k s_k(\theta) - \frac{1}{2} \sum_{k=1}^n s_k^2(\theta) \right) \right\} d\theta. \quad (\text{III.B.65})$$

Χρησιμοποιώντας την ταυτότητα $\sin(a+b) = \cos a \sin b + \sin a \cos b$, ο πρώτος όρος σε παρενθέσεις στον εκθέτη στο (III.B.65) μπορεί να γραφεί ως

$$\sum_{k=1}^n y_k s_k(\theta) = y_c \sin \theta + y_s \cos \theta \text{ with}$$

Και

$$y_s \triangleq \sum_{k=1}^n a_k y_k \sin[(k-1)\omega_c T_S].$$

Παρομοίως, με την ταυτότητα $\sin^2 a = \frac{1}{2} - \frac{1}{2} \cos 2a$, ο δεύτερος όρος σε παρένθεση στον εκθέτη γίνεται

$$-\frac{1}{2} \sum_{k=1}^n s_k^2(\theta) = -\frac{1}{4} \sum_{k=1}^n a_k^2 + \frac{1}{4} \sum_{k=1}^n a_k^2 \cos(2(k-1)\omega_c T_S + 2\theta). \quad (\text{III.B.67})$$

Για τις περισσότερες καταστάσεις που προκύπτουν πρακτικά, ο δεύτερος όρος στο δεξί μέλος του (III.B.67) είναι μηδέν ή κατά προσέγγιση μηδέν για όλες τις τιμές του θ . Για παράδειγμα, αν η ακολουθία σήματος a_1, \dots, a_n είναι μια σταθερά επί μια ακολουθία των ± 1 , ή αν a_1, \dots, a_n έχει σχήμα αυξημένου συνημίτονου της μορφής

$$a_k = A[1 - \cos((k-1)2\pi/(n-1))], k = 1, \dots, n, \text{ τότε αυτός ο δεύτερος όρος είναι}$$

όμοια μηδέν. Σε άλλες περιπτώσεις που μας αφορούν στην πράξη, η a_1^2, \dots, a_n^2 είναι συνήθως βραδώς μεταβαλλόμενη, συγκρινόμενη με τη διπλάσια συχνότητα του φορέα. Άρα αυτός ο δεύτερος όρος αφορά σ' ένα κατωπερατό φιλτράρισμα ενός υψηλής συχνότητας σήματος, μια διαδικασία που καταλήγει σε μια αμελητέα έξοδο. Σε κάθε περίπτωση, θα υποθέσουμε ότι η a_1, \dots, a_n είναι τέτοια ώστε αυτός ο δεύτερος όρος είναι μηδέν για όλα τα θ , και επομένως η $L(\underline{y})$ γίνεται

$$L(\underline{y}) = e^{-na^2/4\sigma^2} \times \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ \frac{1}{\sigma^2} (y_c \sin \theta + y_s \cos \theta) \right\} d\theta \quad (\text{III.B.68})$$

$$\text{με } \bar{a}^2 = (1/n) \sum_{k=1}^n a_k^2.$$

Η παράσταση (III.B.68) είναι παρόμοια με αυτή για το λόγο πιθανοφάνειας στο Παράδειγμα Π.Ε.1. Συγκεκριμένα, συγκρίνοντας τα (III.B.68) και (Π.Ε.17) βλέπουμε ότι

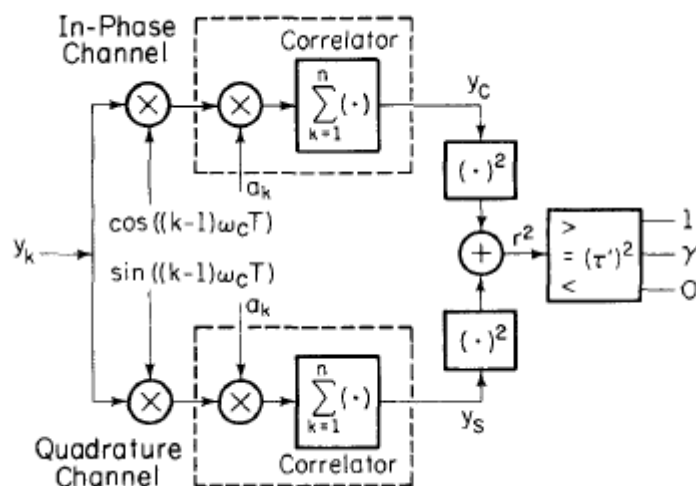
$$L(\underline{y}) = e^{-n\bar{a}^2/4\sigma^2} I_0(r/\sigma^2), \quad (\text{III.B.69})$$

όπου $r = [y_c^2 + y_s^2]^{1/2}$ και I_0 είναι η μηδενικής τάξης τροποποιημένη συνάρτηση Bessel πρώτου είδους. Λαμβάνοντας υπόψη τη μονοτονία του $I_0(\cdot)$, οι βέλτιστοι έλεγχοι στην περίπτωση αυτή δίνονται από

$$\bar{\delta}_o(\underline{y}) = \begin{cases} 1 & > \\ \gamma & \text{if } r = \tau' \triangleq \sigma^2 I_0^{-1}(\tau e^{n\bar{a}^2/4\sigma^2}). \\ 0 & < \end{cases} \quad (\text{III.B.70})$$

Η δομή αυτού του ανιχνευτή φαίνεται στο Σχ. III.B.10. Ας σημειωθεί ότι το παρατηρούμενο σήμα y_1, \dots, y_n διαιρείται σε δύο κανάλια, ένα από τα οποία πολλαπλασιάζει κάθε y_k με $\cos[(k-1)\omega_c T_s]$ και το άλλο πολλαπλασιάζει κάθε y_k

με $\sin[(k-1)\omega_c T_s]$. Αυτά ενίοτε ονομάζονται σε φάση και τετραγωνικά κανάλια αντίστοιχα.



ΣΧΗΜΑ III.B.10. Βέλτιστο σύστημα για μη συνεκτική ανίχνευση ενός διαμορφωμένου ημιτονοειδούς σε i.i.d. κατά Gauss θόρυβο.

Κάθε κανάλι συσχετίζει το τελικό προϊόν με την ακολουθία πλάτους a_1, \dots, a_n . Οι έξοδοι του καναλιού τότε συνδυάζονται για να δώσουν το r , το οποίο συγκρίνεται μ' ένα κατώφλι (Η δομή αυτή ενίοτε καλείται *ανιχνευτής φακέλων*). Διαισθητικά, η αρχή λειτουργίας αυτού του ανιχνευτή είναι ότι όταν ένα σήμα είναι παρόν, κάθε κανάλι διαλέγει μια ποσότητα της ενέργειας του σήματος ανάλογα με την πραγματική φάση γωνίας του φορέα. Ωστόσο, ανεξάρτητα από τη φάση του φορέα, ο συνδυασμός των δύο καναλιών συγκεντρώνει όλη την ενέργεια του σήματος.

Για να αναλύσουμε την επίδοση του ανιχνευτή του (III.B.70), πρέπει να βρούμε

$P_j(R > \tau') = P_j(R^2 > (\tau')^2)$ για $j = 0, 1$. Αφού $R^2 = Y_c^2 + Y_s^2$, όπου $Y_c \triangleq \sum_{k=1}^n a_k Y_k \cos[(k-1)\omega_c T_S]$ και $Y_s \triangleq \sum_{k=1}^n a_k Y_k \sin[(k-1)\omega_c T_S]$, οι επιθυμητές πιθανότητες μπορούν να βρεθούν από την κοινή συνάρτηση πυκνότητας πιθανότητας των Y_c και Y_s ως προς τις δύο υποθέσεις. Ως προς το H_0 , το \underline{Y} είναι $N(0, \sigma^2 \mathbf{I})$ και αφού το Y_c και το Y_s είναι γραμμικά στο \underline{Y} είναι από κοινού κατά Gauss ως προς το H_0 . Επομένως μπορούμε να καθορίσουμε την κοινότητα των (Y_c, Y_s) ως προς το H_0 βρίσκοντας τα μέσα και τις μεταβλητές των Y_c και Y_s και το συντελεστή συσχέτισης μεταξύ των Y_c και Y_s

Έχουμε ευθέως ότι

$$E\{Y_c|H_0\} = \sum_{k=1}^n a_k E\{N_k\} \cos[(k-1)\omega_c T_S] = 0$$

και

$$\begin{aligned} \text{Var}\{Y_c|H_0\} &= E\{Y_c^2|H_0\} = \sum_{k=1}^n \sum_{l=1}^n a_k a_l E\{N_k N_l\} \\ &\quad \times \cos[(k-1)\omega_c T_S] \cos[(l-1)\omega_c T_S] \\ &= \sigma^2 \sum_{k=1}^n a_k^2 \cos^2[(k-1)\omega_c T_S] = \frac{n\sigma^2 \bar{a}^2}{2} \end{aligned}$$

όπου έχουμε πάλι χρησιμοποιήσει την παραδοχή ότι ο δεύτερος όρος στο (III.B.67) είναι μηδέν για όλα τα θ . Παρομοίως, έχουμε $E\{Y_s|H_0\} = 0$ και $\text{Var}\{Y_s|H_0\} = \text{Var}\{Y_c|H_0\}$. Ο συντελεστής συσχέτισης μεταξύ του Y_c και του Y_s ως προς το H_0 δίνεται από $\text{Cov}(Y_c, Y_s|H_0)/[\text{Var}\{Y_c|H_0\} \text{Var}\{Y_s|H_0\}]^{1/2}$. Αφού τα Y_c και Y_s έχουν μηδενικά μέσα ως προς το H_0 , έχουμε

$$\begin{aligned}
\text{Cov}(Y_c, Y_s | H_0) &= E\{Y_c Y_s | H_0\} \\
&= \sum_{k=1}^n \sum_{l=1}^n a_k a_l E\{N_k N_l\} \cos[(k-1)\omega_c T_S] \sin[(l-1)\omega_c T_S] \\
&= \sigma^2 \sum_{k=1}^n a_k^2 \cos[(k-1)\omega_c T_S] \sin[(k-1)\omega_c T_S] \quad (\text{III.B.71}) \\
&= \frac{\sigma^2}{2} \sum_{k=1}^n a_k^2 \sin[2(k-1)\omega_c T_S] = 0.
\end{aligned}$$

Από το (III.B.71) βλέπουμε ότι ως προς το H_0 , τα Y_c και Y_s δε συσχετίζονται και αφού είναι από κοινού κατά Gauss είναι άρα ανεξάρτητα.

Συμπεραίνουμε ότι ως προς το H_0 , τα Y_c και Y_s είναι ανεξάρτητες $\mathcal{N}(0, n\sigma^2 \overline{a^2}/2)$ τυχαίες μεταβλητές. Λαμβάνοντας υπόψη ότι η τυχαιοποίηση δεν έχει σημασία εδώ, η πιθανότητα ψευδούς συναγερμού του (III.B.70) γίνεται επομένως

$$\begin{aligned}
P_0(\Gamma_1) &= \iint_{\{y_c^2 + y_s^2 \geq (\tau')^2\}} \frac{1}{\pi n \sigma^2 \overline{a^2}} e^{-(y_c^2 + y_s^2)/n\sigma^2 \overline{a^2}} dy_c dy_s \\
&= \frac{1}{\pi n \sigma^2 \overline{a^2}} \int_0^{2\pi} \int_{\tau'}^{\infty} r e^{-r^2/n\sigma^2 \overline{a^2}} dr d\psi \quad (\text{III.B.72}) \\
&= e^{-(\tau')^2/n\sigma^2 \overline{a^2}},
\end{aligned}$$

όπου για να έχουμε τη δεύτερη ισότητα εισαγάγαμε πολικές συντεταγμένες r και ψ .

Για να ορίσουμε την πιθανότητα ανίχνευσης του (III.B.70), πρέπει να βρούμε την κοινή πυκνότητα των Y_c και Y_s ως προς το H_1 . Ας σημειωθεί ότι δοθέντος ότι $\Theta = \theta$, το \underline{Y} έχει μια δεσμευμένη $\mathcal{N}(\underline{s}(\theta), \sigma^2 \mathbf{I})$ κατανομή ως προς το H_1 . Άρα, δόθεντος ότι $\Theta = \theta$, τα Y_c και Y_s είναι δεσμευμένα από κοινού Gauss. Έχουμε ευθέως ότι

$$\begin{aligned}
E\{Y_c | H_1, \Theta = \theta\} &= \sum_{k=1}^n a_k E\{Y_k | H_1, \Theta = \theta\} \cos[(k-1)\omega_c T_S] \\
&= \sum_{k=1}^n a_k^2 \sin[(k-1)\omega_c T_S + \theta] \cos[(k-1)\omega_c T_S] \\
&= \frac{n \overline{a^2}}{2} \sin \theta,
\end{aligned}$$

και ομοίως $E\{Y_s | H_1, \Theta = \theta\} = (n \overline{a^2}/2) \cos \theta$. Με σταθερό θ , οι μεταβλητές και οι συμμεταβλητές ως προς το H_1 των Y_c και Y_s δεν αλλάζουν από τις H_0 τιμές τους, καθώς η μόνη αλλαγή στο \underline{Y} είναι μια μετατόπιση σε μέσο.

Η μη εξαρτημένη πυκνότητα των Y_c και Y_s ως προς το H_1 βρίσκεται μέσω της εύρεσης του μέσου της δεσμευμένης πυκνότητας στο θ . Έχουμε

$$\begin{aligned} p_{Y_c, Y_s}(y_c, y_s | H_1) &= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{\pi n \sigma^2 a^2} \\ &\times \exp \left\{ -\frac{1}{n \sigma^2 a^2} q(y_c, y_s; \frac{n a^2}{2}, \theta) \right\} d\theta \quad (\text{III.B.73}) \\ &= p_{Y_c, Y_s}(y_c, y_s | H_0) e^{-n a^2 / 4 \sigma^2} I_0 \left(\frac{|y_c^2 + y_s^2|^{1/2}}{\sigma^2} \right), \end{aligned}$$

όπου $p_{Y_c, Y_s}(y_c, y_s | H_0)$ είναι η κοινή πυκνότητα των Y_c και Y_s ως προς το H_0 και όπου το q_1 ορίζεται στο (II.E.14). Η πιθανότητα ανίχνευσης του (III.B.70) άρα γίνεται

$$\begin{aligned} P_D(\delta_o) &= P_1(\Gamma_1) = \int \int_{\{y_c^2 + y_s^2 > (\tau')^2\}} p_{Y_c, Y_s}(y_c, y_s | H_1) dy_c dy_s \\ &= \frac{e^{-n a^2 / 4 \sigma^2}}{\pi n \sigma^2 a^2} \int_0^{2\pi} \int_{\tau'}^{\infty} r e^{-r^2 / n \sigma^2 a^2} I_0 \left(\frac{r}{\sigma^2} \right) dr d\psi \\ &= \int_{\tau_o}^{\infty} x e^{-(x^2 + b^2)/2} I_0(bx) dx \triangleq Q(b, \tau_o), \quad (\text{III.B.74}) \end{aligned}$$

όπου στην τρίτη ισότητα έχουμε ορίσει $b^2 = n a^2 / 2 \sigma^2$ και $\tau_o = \tau' / \sigma^2 b$ και έχουμε κάνει την αντικατάσταση $x = r / \sigma^2 b$. Η συνάρτηση Q που ορίζεται στο (III.B.74) είναι ενίοτε γνωστή ως Q συνάρτηση του Marcum. Σημειώστε στο (III.B.72) ότι

$Q(0, \tau_o) = P_F(\delta_o)$. Από το (III.B.72) μπορούμε εύκολα να θέσουμε το κατώφλι για ανίχνευση α-επιπέδου Neyman – Pearson. Έχουμε

$$\tau' = [n \sigma^2 a^2 \log(1/\alpha)]^{1/2}, \quad (\text{III.B.75})$$

από το οποίο, το (III.B.74) δίνει τα ROCs ως

$$P_D(\bar{\delta}_o) = Q[b, [2 \log(1/\alpha)]^{1/2}]. \quad (\text{III.B.76})$$

Αυτά τα χαρακτηριστικά λειτουργίας δέκτη μοιάζουν παρόμοια με αυτά του συνεκτικού προβλήματος (Σχ. II.D.4) και απεικονίζονται για παράδειγμα στο Σχ. V.D.3 του Helstrom (1968).

Από το (III.B.76) βλέπουμε ότι η επίδοση της ανίχνευσης Neyman – Pearson εδώ εξαρτάται μόνο από την παράμετρο b . Σημειώστε ότι η μέση ενέργεια σήματος είναι

$$E \left\{ \frac{1}{n} \sum_{k=1}^n s_k^2(\Theta) \right\} = \frac{1}{n} \frac{1}{2\pi} \int_0^{2\pi} \sum_{k=1}^n a_k^2 \sin^2[(k-1)\omega_c T_S + \theta] d\theta = \frac{\bar{a}^2}{2}.$$

Άρα το $b^2 = n\bar{a}^2/2\sigma^2$ έχει μια ερμηνεία λόγου σήματος προς θόρυβο, ανάλογη με το d^2 στο πρόβλημα συνεκτικής ανίχνευσης. Αν επρόκειτο να ανιχνεύσουμε το ίδιο σήμα συνεκτικά (δηλ. με γνωστό το θ), η αντίστοιχη τιμή του d^2 θα ήταν

$$d^2 = \frac{1}{\sigma^2} \sum_{k=1}^n s_k^2(\theta) = \frac{n\bar{a}^2}{2\sigma^2} = b^2. \quad (\text{III.B.77})$$

Επομένως, αυτές οι αναλογίες σήματος προς θόρυβο είναι στην πραγματικότητα οι ίδιες. Ωστόσο, η επίδοση για σταθερό a είναι διαφορετική για τα δύο συστήματα με $b = d$. Για τις κυμάνσεις του SNR και του a που προκύπτουν στα περισσότερα πρακτικά προβλήματα έχουμε

$$Q[b, [2 \log(1/\alpha)]^{1/2}] = 1 - \Phi[\Phi^{-1}(1 - \alpha) - d] \quad (\text{III.B.78})$$

Όταν $b \cong d + 0.4$ {βλέπε Helstrom (1968)}. Άρα αν επιθυμούμε να χρησιμοποιήσουμε μια μη συνεκτική τεχνική, απαιτείται ένα ελαφρώς υψηλότερο SNR για να επιτύχουμε την ίδια απόδοση με αυτή της αντίστοιχης συνεκτικής τεχνικής. Το μειονέκτημα ενός συνεκτικού συστήματος είναι ότι θα πρέπει να παρέχονται κάποια μέσα για την παραγωγή της φάσης φορέα.

Ολοκληρώνοντας αυτό το παράδειγμα, θα πρέπει να σημειωθεί ότι η παραγωγή λόγου πιθανοφάνειας (που αναφέρεται παραπάνω) μπορεί να εφαρμοστεί σε προβλήματα στα οποία έχουμε σήματα της μορφής του (III.B.64), αλλά με την παρουσία διαφορετικών ακολουθιών πλάτους ως προς και τις δύο υποθέσεις, στις οποίες έχουμε

$$s_{jk}(\theta) = a_{jk} \sin[(k-1)\omega_c T_S + \theta], \quad k = 1, \dots, n, \quad (\text{III.B.79})$$

για $j = 0, 1$. Εδώ, ο λόγος πιθανοφάνειας για i.i.d. θόρυβο μπορεί να βρεθεί από το (III.B.69) χρησιμοποιώντας μια τρίτη υπόθεση θορύβου μόνο ως «καταλύτη». Συγκεκριμένα,

$$\begin{aligned}
L(\underline{y}) &= \frac{E_1\{p_N(\underline{y} - \underline{s}_1(\Theta))\}}{E_0\{p_N(\underline{y} - \underline{s}_0(\Theta))\}} \\
&= \frac{E_1\{p_N(\underline{y} - \underline{s}_1(\Theta))/p_N(\underline{y})\}}{E_0\{p_N(\underline{y} - \underline{s}_0(\Theta))/p_N(\underline{y})\}} = \frac{L_1(\underline{y})}{L_0(\underline{y})},
\end{aligned}$$

όπου το $\overline{L_j(\underline{y})}$ είναι ο λόγος πιθανοφάνειας γι' αυτήν την τρίτη υπόθεση ως προς το H_j . Από το (III.B.69) συμπεραίνουμε ότι

$$L(\underline{y}) = \frac{e^{-n\overline{a_1^2}/4\sigma^2} I_0(r_1/\sigma^2)}{e^{-n\overline{a_0^2}/4\sigma^2} I_0(r_0/\sigma^2)}, \quad (\text{III.B.80})$$

όπου $\overline{a_j^2} = (1/n) \sum_{k=1}^n a_{jk}^2$ and $r_j = [y_{cj}^2 + y_{sj}^2]^{1/2}$ with

$$y_{cj} = \sum_{k=1}^n a_{jk} y_k \cos[(k-1)\omega_c T_S]$$

και

$$y_{sj} = \sum_{k=1}^n a_{jk} y_k \sin[(k-1)\omega_c T_S].$$

Η βέλτιστη ανίχνευση, επομένως, περιλαμβάνει το συνδυασμό των εξόδων των δύο συστημάτων, όπως αυτό στο Σχ. III.B.10, που να «ταιριάζει» σε κάθε μία από τις ακολουθίες πλάτους. Για παράδειγμα, αν τα σήματα έχουν ισορροπημένες ενέργειες ($\overline{a_0^2} = \overline{a_1^2}$) και υποθέτουμε ομοιόμορφα κόστη και εκ των προτέρων ισοδύναμα ($\tau = 1$), τότε ο Μπαϊεσιανός έλεγχος στην κατάσταση αυτή γίνεται

$$\delta_B(\underline{y}) = \begin{cases} 1 & > \\ 0 \text{ or } 1 & \text{if } r_1 = r_0, \\ 0 & < \end{cases} \quad (\text{III.B.81})$$

αφού το I_0 είναι μονότονα αυξανόμενο. Για την τελευταία αυτή περίπτωση, αν υποθέσουμε ότι οι ακολουθίες πλάτους είναι ορθογώνιες,

$$\sum_{k=1}^n a_{jk} a_{lk} = 0 \text{ if } j \neq l$$

και ότι

$$\sum_{k=1}^n a_{jk} a_{lk} \sin[(k-1)\omega_c T_S + \theta] = 0$$

για όλα τα θ (αυτή η υπόθεση αφορά σε αρκετά κοινά σύνολα σημάτων που χρησιμοποιούνται στην πρακτική), τότε η πιθανότητα σφάλματος μπορεί να αποδειχθεί ευθέως ότι είναι (βλ. Άσκηση 1)

$$P_e = \frac{1}{2} e^{-b^2/4}, \quad (\text{III.B.82})$$

όπου $b^2 = n I \bar{a}_1^2 / 2\sigma^2 (\equiv n \bar{a}_0^2 / 2\sigma^2)$.

Άλλες πλευρές αυτού του προβλήματος, όπως η ανίχνευση με άγνωστο πλάτος και φάση, αναπτύσσονται στις ασκήσεις.

Περίπτωση III.B.4.: Ανίχνευση Στοχαστικών Σημάτων

Σε κάποιες εφαρμογές, τα σήματα δείχνουν να είναι καλύτερα τυποποιημένα ως αμιγώς τυχαία ή στοχαστικά. Αυτός ο τύπος μοντέλου είναι χρήσιμος, για παράδειγμα, στη ραδιοαστρονομία, τα σήματα και άλλες εφαρμογές, στις οποίες τα σήματα διαταράσσονται από τη διάδοση μέσω τυρβωδών μέσων ή κατά μήκος πολλαπλών δρόμων. Για την περίπτωση αυτή έχουμε το γενικό μοντέλο του (III.B.1) και ο απλούστερος γενικός τύπος λόγου πιθανοφάνειας είναι αυτός του (III.B.4).

Μια σημαντική ειδική περίπτωση του προβλήματος αυτού είναι αυτό στο οποίο τα σήματα και ο θόρυβος είναι τυχαία κατά Gauss διανύσματα. Για να εξετάσουμε την περίπτωση αυτή, σημειώνουμε πρώτα ότι για το ακόλουθο πρόβλημα ελέγχου υπόθεσης στο \mathbb{R}^n :

$$\begin{aligned} & H_0 : \underline{Y} \sim \mathcal{N}(\underline{\mu}_0, \underline{\Sigma}_0) \\ \text{versus} & \\ & H_1 : \underline{Y} \sim \mathcal{N}(\underline{\mu}_1, \underline{\Sigma}_1) \end{aligned} \quad (\text{III.B.83})$$

ο λογάριθμος λόγου πιθανοφάνειας δίνεται από

$$\begin{aligned} \log L(\underline{y}) &= \frac{1}{2} \log[|\underline{\Sigma}_0|/|\underline{\Sigma}_1|] + \frac{1}{2}(\underline{y} - \underline{\mu}_0)^T \underline{\Sigma}_0^{-1}(\underline{y} - \underline{\mu}_0) \\ &\quad - \frac{1}{2}(\underline{y} - \underline{\mu}_1)^T \underline{\Sigma}_1^{-1}(\underline{y} - \underline{\mu}_1) \quad (\text{III.B.84}) \\ &= \frac{1}{2} \underline{y}^T [\underline{\Sigma}_0^{-1} - \underline{\Sigma}_1^{-1}] \underline{y} + [\underline{\mu}_1^T \underline{\Sigma}_1^{-1} - \underline{\mu}_0^T \underline{\Sigma}_0^{-1}] \underline{y} + C, \end{aligned}$$

όπου $C = \frac{1}{2}(\log(|\underline{\Sigma}_0|/|\underline{\Sigma}_1|) + \underline{\mu}_0^T \underline{\Sigma}_0^{-1} \underline{\mu}_0 - \underline{\mu}_1^T \underline{\Sigma}_1^{-1} \underline{\mu}_1)$. Ας σημειωθεί ότι η $\log L(\underline{y})$ εδώ συνίσταται σ' έναν τετραγωνικό (ή δευτεροβάθμιο) όρο στο \underline{y} , ένα γραμμικό όρο στο \underline{y} και μία σταθερά. Αν οι δύο συμμεταβλητές είναι ίδιες, ας πούμε $\underline{\Sigma}_0 = \underline{\Sigma}_1 = \underline{\Sigma}$, τότε ο τετραγωνικός όρος εξαφανίζεται, και έχουμε ένα γραμμικό στατιστικό έλεγχο

$(\underline{\mu}_1 - \underline{\mu}_0)^T \underline{\Sigma}^{-1} \underline{y}$, αφού το C μπορεί να ενσωματωθεί στο κατώφλι. Αυτή είναι η περίπτωση της συνεκτικής ανίχνευσης σε κατά Gauss θόρυβο που αναφέρεται στην Περίπτωση III.B.2. Αν, από την άλλη μεριά, τα διανύσματα μέσου είναι τα ίδια ως προς και τις δύο υποθέσεις ($\underline{\mu}_0 = \underline{\mu}_1$), μπορούμε (χωρίς απώλεια της γενικότητας) να θεωρήσουμε ότι είναι $\underline{0}$ και η δομή της $\log L(\underline{y})$ είναι επομένως τετραγωνική.

Η τελευταία περίπτωση είναι εφαρμόσιμη στο πρόβλημα ανίχνευσης μηδενικού μέσου στοχαστικών σημάτων σε κατά Gauss θόρυβο. Συγκεκριμένα, εξετάστε το ζεύγος υπόθεσης

$$\begin{aligned} & H_0 : \underline{y} = \underline{N} \\ \text{versus} & \\ & H_1 : \underline{y} = \underline{N} + \underline{S}, \end{aligned} \quad (\text{III.B.85})$$

όπου $\underline{N} \sim \mathcal{N}(\underline{0}, \sigma^2 \underline{I})$ και $\underline{S} \sim \mathcal{N}(\underline{0}, \underline{\Sigma}_S)$. Περιπτώσεις στις οποίες το \underline{N} είναι Gauss αλλά όχι i.i.d. μπορούν να ενταχθούν σ' αυτό το μοντέλο μέσω της προ-λευκάνσης, καθώς ένας γραμμικός μετασχηματισμός του Gauss σήματος \underline{S} θα συνεχίσει να είναι Gauss. Επίσης, περιπτώσεις στις οποίες υπάρχουν σήματα παρόντα ως προς κάθε μία από τις υποθέσεις μπορούμε να τις χειριστούμε, χρησιμοποιώντας το (III.B.85) ως έναν «καταλύτη», όπως στο προηγούμενο παράδειγμα.

Το ζεύγος υπόθεσης του (III.B.85) είναι μια ειδική περίπτωση του (III.B.83) με

$$\underline{\Sigma}_0 = \sigma^2 \underline{I} \text{ and } \underline{\Sigma}_1 = \sigma^2 \underline{I} + \underline{\Sigma}_S. \text{ (Υποθέτουμε, όπως πάντα, ότι σήμα και θόρυβος}$$

είναι ανεξάρτητα.) Επομένως, από το (III.B.84) βλέπουμε ότι οι βέλτιστοι έλεγχοι για το (III.B.85) έχουν τη μορφή

$$\tilde{\delta}_o(\underline{y}) = \begin{cases} 1 & \text{if } \underline{y}^T \mathbf{Q} \underline{y} > \tau' \\ \gamma & \text{if } \underline{y}^T \mathbf{Q} \underline{y} = \tau' \\ 0 & \text{if } \underline{y}^T \mathbf{Q} \underline{y} < \tau' \end{cases} \quad (\text{III.B.86})$$

με $\tau' \triangleq 2(\log \tau - C)$ and $\mathbf{Q} \triangleq \sigma^{-2} \mathbf{I} - (\sigma^2 \mathbf{I} + \Sigma_S)^{-1} \equiv \sigma^{-2} \Sigma_S (\sigma^2 \mathbf{I} + \Sigma_S)^{-1}$.

Από το (III.B.86) βλέπουμε ότι ο βέλτιστος ανιχνευτής υπολογίζει την τετραγωνική μορφή $\underline{y}^T \mathbf{Q} \underline{y}$ και τη συγκρίνει μ' ένα κατώφλι. Η δομή αυτή είναι γνωστή ως *τετραγωνικός ανιχνευτής*.

Για παράδειγμα, αν τα δείγματα σήματος είναι i.i.d. $\mathcal{N}(0, \sigma_S^2)$ τυχαίες μεταβλητές, τότε $\Sigma_S = \sigma_S^2 \mathbf{I}$ και

$$\underline{y}^T \mathbf{Q} \underline{y} = \frac{\sigma_s^2}{\sigma^2(\sigma^2 + \sigma_S^2)} \sum_{k=1}^n y_k^2. \quad (\text{III.B.87})$$

Αρα σ' αυτή την ειδική περίπτωση, ο βέλτιστος ανιχνευτής συγκρίνει το μέγεθος $\sum_{k=1}^n y_k^2$ μ' ένα κατώφλι. Αφού $(1/n) \sum_{k=1}^n y_k^2$ είναι η μέση ενέργεια της παρατηρούμενης κυματομορφής, η προκύπτουσα δομή ανιχνευτή είναι ενίοτε γνωστή ως *ανιχνευτής ενέργειας*. (Είναι γνωστή και ως *ραδιόμετρο*). Αυτός είναι ένας διαισθητικά λογικός τρόπος ανίχνευσης του σήματος στην περίπτωση αυτή, καθώς πρέπει εδώ να επιλέξουμε μεταξύ δύο καταστάσεων στις οποίες οι τυχαίες παρατηρήσεις Y_1, \dots, Y_n διαφέρουν μόνο ως προς τη (στατιστικά) μέση ενέργεια που περιέχουν. Αυτό μπορεί να αντιπαρατεθεί με το πρόβλημα επιλογής μεταξύ δύο σταθερών σημάτων $\underline{\mu}_0 = s_0 \underline{1}$ και $\underline{\mu}_1 = s_1 \underline{1}$, όπου $\underline{1} = (1, \dots, 1)^T$ και $s_1 > s_0$, σε λευκό θόρυβο. Στην περίπτωση αυτή, ο βέλτιστος ανιχνευτής συγκρίνει το $\sum_{k=1}^n y_k$ μ' ένα κατώφλι. Αφού οι τυχαίες παρατηρήσεις διαφέρουν εδώ ως προς τις δύο υποθέσεις μόνο σε ό,τι αφορά τα (στατιστικά) μέσα πλάτη τους, είναι διαισθητικά λογικό να αποφασίσουμε μεταξύ τους συγκρίνοντας τη μέση παρατήρηση $(1/n) \sum_{k=1}^n y_k$, μ' ένα κατώφλι. Η τελευταία αυτή δομή είναι ενίοτε γνωστή ως *γραμμικός ανιχνευτής*.

Προκειμένου να αναλύσουμε την επίδοση του ανιχνευτή στο (III.B.86), πρέπει να υπολογίσουμε τις πιθανότητες $P_j(\underline{Y}^T \mathbf{Q} \underline{Y} > \tau')$ για $j = 0, 1$. Το πρόβλημα αυτό μπορεί να

αντιμετωπιστεί ευκολότερα, αν πρώτα μετατρέψουμε τις παρατηρήσεις με τρόπο παρόμοιο μ' αυτό που εξετάσαμε στον πρόβλημα συνεκτικής ανίχνευσης. Συγκεκριμένα, υποθέστε ότι $\lambda_1, \dots, \lambda_n$ και $\underline{v}_1, \dots, \underline{v}_n$ είναι οι ιδιοτιμές και τα αντίστοιχα ορθοκανονικά ιδιοδιανύσματα του πίνακα διακύμανσης σήματος Σ_S . Τότε μπορούμε να γράψουμε $\Sigma_S = \sum_{k=1}^n \lambda_k \underline{v}_k \underline{v}_k^T$, και είναι προφανές ότι

$$\mathbf{I} = \sum_{k=1}^n \underline{v}_k \underline{v}_k^T.$$

Άρα

$$(\sigma^2 \mathbf{I} + \Sigma_S)^{-1} = \sum_{k=1}^n (\sigma^2 + \lambda_k)^{-1} \underline{v}_k \underline{v}_k^T,$$

και

$$\begin{aligned} \mathbf{Q} &= \sum_{k=1}^n \frac{1}{\sigma^2} \underline{v}_k \underline{v}_k^T - \sum_{k=1}^n \frac{1}{\sigma^2 + \lambda_k} \underline{v}_k \underline{v}_k^T \\ &= \sum_{k=1}^n \frac{\lambda_k}{\sigma^2(\sigma^2 + \lambda_k)} \underline{v}_k \underline{v}_k^T. \end{aligned} \quad (\text{III.B.88})$$

Βλέπουμε ότι η στατιστική ανίχνευσης μπορεί να γραφεί ως

$$\underline{y}^T \mathbf{Q} \underline{y} = \sum_{k=1}^n (\bar{y}_k)^2 \quad (\text{III.B.89})$$

με $\bar{y}_k \triangleq [\lambda_k / \sigma^2 (\sigma^2 + \lambda_k)]^{1/2} \underline{v}_k^T \underline{y}$.

Αφού το $\underline{v}_1, \dots, \underline{v}_n$ είναι ένα σύνολο ορθοκανονικών ιδιοδιανυσμάτων και για το $\sigma^2 \mathbf{I}$ και για το $(\sigma^2 \mathbf{I} + \Sigma_S)$, φαίνεται ευθέως ότι [παρόμοια με το (III.B.42)] ότι οι

$\bar{Y}_1, \dots, \bar{Y}_n$ είναι ανεξάρτητες μηδενικού μέσου Gauss τυχαίες μεταβλητές ως προς και τις δύο υποθέσεις με διακυμάνσεις

$$\sigma_{jk}^2 \triangleq \text{Var}(\bar{Y}_k | H_j) = \begin{cases} \frac{\lambda_k}{\sigma^2 + \lambda_k} & \text{if } j = 0 \\ \frac{\lambda_k}{\sigma^2} & \text{if } j = 1. \end{cases} \quad (\text{III.B.90})$$

Αυτό συνεπάγεται ότι ως προς το H_j , το $\underline{Y}^T \mathbf{Q} \underline{Y}$ είναι το σύνολο ανεξάρτητων μεταβλητών στο οποίο ο k th όρος \bar{Y}_k^2 έχει την κατανομή μιας $\mathcal{N}(0, \sigma_{jk}^2)$ τυχαίας μεταβλητή εις το τετράγωνο. Η συνάρτηση πυκνότητας πιθανότητας του $T_k \triangleq \bar{Y}_k^2$ ως προς το H_j μπορεί να είναι [βλέπε, π.χ., Papoulis (1986)]

$$p_{T_k}(t|H_j) = \begin{cases} \frac{1}{\sqrt{2\pi t \sigma_{jk}^2}} e^{-t/2\sigma_{jk}^2}, & t > 0 \\ 0, & t \leq 0 \end{cases} \quad (\text{III.B.91})$$

η οποία είναι μια γάμμα $(1/2, 1/2\sigma_{jk}^2)$ πυκνότητα. Η πυκνότητα πιθανότητας, p_T , του $T \triangleq \sum_{k=1}^n \bar{y}_k^2$ είναι η n -πλάσια συνέλιξη $p_{T_1} * p_{T_2} * \dots * p_{T_n}$, η οποία εκφράζεται ευκολότερα με τη χρήση μετασχηματισμών Fourier ως

$$p_T = \mathcal{F}^{-1} \left\{ \prod_{k=1}^n \phi_{T_k} \right\}, \quad (\text{III.B.92})$$

όπου η $\phi_{T_k}(u) \triangleq \mathcal{F}\{p_{T_k}\}(u) = E\{e^{iuT_k}\}$ είναι η χαρακτηριστική συνάρτηση του T_k

(εδώ $i = \sqrt{-1}$). Η χαρακτηριστική συνάρτηση της γάμμα $(1/2, 1/2\sigma_{jk}^2)$ δίνεται από (Lukacs, 1960)

$$\phi_{T_k}(u) = \frac{1}{[1 - 2iu\sigma_{jk}^2]^{1/2}}, \quad u \in \mathbf{R}. \quad (\text{III.B.93})$$

Άρα

$$p_T(t|H_j) = \int_{-\infty}^{\infty} e^{-iut} \prod_{k=1}^n [1 - 2iu\sigma_{jk}^2]^{-1/2} du. \quad (\text{III.B.94})$$

Δεν υπάρχει γνωστή γενική κλειστή μορφή για το (III.B.94). ωστόσο, στην ειδική περίπτωση

στην οποία $\sigma_{j1}^2 = \dots = \sigma_{jn}^2 \triangleq \sigma_j^2$, το (III.B.94) μπορεί να αντιστραφεί για να δώσει

$$p_T(t|H_j) = \begin{cases} \frac{1}{(2\sigma_j^2)^{n/2}\Gamma(n/2)} t^{(n/2-1)} e^{-t/2\sigma_j^2}, & t > 0 \\ 0, & t \leq 0, \end{cases} \quad (\text{III.B.95})$$

όπου η $\Gamma(x) = \int_0^\infty e^{-y} y^{x-1} dy$ είναι η γάμμα συνάρτηση. Η περίπτωση αυτή αντιστοιχεί στην κατάσταση, στην οποία $\lambda_1 = \lambda_2 = \dots = \lambda_n \triangleq \sigma_S^2$, από την οποία

$$\Sigma_S = \sigma_S^2 \sum_{k=1}^n \underline{u}_k \underline{u}_k^T = \sigma_S^2 \mathbf{I}.$$

Αυτό σημαίνει, στην περίπτωση αυτή ότι τα δείγματα σήματος είναι i.i.d. $\mathcal{N}(0, \sigma_S^2)$ και έχουμε $\sigma_0^2 = \sigma_S^2/(\sigma^2 + \sigma_S^2)$ and $\sigma_1^2 = \sigma_S^2/\sigma^2$. Η εξίσωση του (III.B.95) είναι η γάμμα $(n/2, 1/2\sigma_j^2)$ πυκνότητα και απ' αυτήν έχουμε

$$P_j(\underline{Y}^T \mathbf{Q} \underline{Y} > \tau') = 1 - \Gamma\left(\frac{n}{2}; \frac{\tau'}{2\sigma_j^2}\right), \quad (\text{III.B.96})$$

όπου

$$\Gamma(x; t) \triangleq \int_0^t e^{-y} y^{x-1} dy / \Gamma(x)$$

είναι η ατελής γάμμα συνάρτηση³ Για Neyman – Pearson ανίχνευση με πιθανότητα ψευδούς συναγεμού α , επιλέγουμε άρα

$$\tau' = 2\sigma_0^2 \Gamma^{-1}\left(\frac{n}{2}; 1 - \alpha\right)$$

όπου $\Gamma^{-1}(x; \cdot)$ είναι η αντίστροφη συνάρτηση του $\Gamma(x; \cdot)$ στη δεύτερή της μεταβλητή. Τα ROCs δίνονται από

$$P_D(\bar{\delta}_{NP}) = 1 - \Gamma\left[\frac{n}{2}; \frac{\sigma_0^2}{\sigma_1^2} \Gamma^{-1}\left(\frac{n}{2}; 1 - \alpha\right)\right]. \quad (\text{III.B.97})$$

Άρα η απόδοση εδώ παραμετρικοποιείται από δύο παραμέτρους n και $\sigma_0^2/\sigma_1^2 = 1/(1 + \sigma_S^2/\sigma^2)$. Ας σημειωθεί ότι σ_S^2/σ^2 είναι ο λόγος της μέσης ισχύος σήματος ως προς τη μέση ισχύ θορύβου στην περίπτωση αυτή και η απόδοση του (III.B.84) βελτιώνεται καθώς αυτό το μέγεθος και/ή το n αυξάνονται.

Για την περίπτωση στην οποία οι ιδιοτιμές του σήματος δεν είναι όμοιες (δηλ. για ένα μη i.i.d. θόρυβο), το (III.B.94) δεν μπορεί να βρεθεί σε κλειστή μορφή. Για την περίπτωση αυτή μπορούν να χρησιμοποιηθούν προσέγγιση ή όρια για να αναλυθεί η απόδοση ανίχνευσης. Τεχνικές για την πραγματοποίηση αυτής της μεθόδου εξετάζονται στην Ενότητα III.C.

Παρατήρηση III.B.4: Σχέση μεταξύ περιπτώσεων εξαρτημένου και ανεξάρτητου σήματος.

Εξετάστε το πρόβλημα του (III.B.85) στο οποίο $\underline{N} \sim \mathcal{N}(\underline{0}, \sigma^2 \mathbf{I})$ και $\underline{S} \sim \mathcal{N}(\underline{\mu}, \Sigma_S)$ με $\Sigma_S = \text{diag}\{\sigma_{S_1}^2, \dots, \sigma_{S_n}^2\}$. Αυτή είναι η περίπτωση στην οποία τα δείγματα θορύβου είναι i.i.d. $\mathcal{N}(0, \sigma^2)$ και τα δείγματα σήματος είναι ανεξάρτητα $\mathcal{N}(\mu_k, \sigma_{S_k}^2)$

3

Ο λόγος λογαριθμικής (log) πιθανοφάνειας για την περίπτωση αυτή δίνεται από

$$\begin{aligned} \log L(\underline{y}) &= \frac{1}{2} \sum_{k=1}^n y_k^2 / \sigma^2 - \frac{1}{2} \sum_{k=1}^n (y_k - \mu_k)^2 / (\sigma_{S_k}^2 + \sigma^2) \\ &\quad + \frac{1}{2} \sum_{k=1}^n \log[\sigma^2 / (\sigma_{S_k}^2 + \sigma^2)]. \end{aligned} \quad (\text{III.B.98})$$

Τώρα εξετάστε το ίδιο πρόβλημα στο οποίο το Σ_S δεν είναι διαγώνιο. Με το $p_j(\underline{y}_1, \dots, \underline{y}_l)$ να συμβολίζει την πυκνότητα του Y_1, \dots, Y_l ως προς το H_j , μπορούμε να γράψουμε

$$p_j(\underline{y}) = p_j(y_1) \prod_{k=2}^n p_j(y_k | y_1, \dots, y_{k-1}), \quad (\text{III.B.99})$$

³ Ας σημειωθεί ότι για ένα άρτιο n το (III.B.96) μπορεί να ολοκληρωθεί κατά μέλη για να παράγει

$$P_j(\underline{Y}^T \mathbf{Q} \underline{Y} > \tau') = e^{-\tau' / 2\sigma_j^2} \sum_{k=0}^{n/2-1} \frac{(\tau' / 2\sigma_j^2)^k}{k!}.$$

όπου $p_j(y_k|y_1, \dots, y_{k-1})$ είναι η δεσμευμένη (ή κατά συνθήκη) πυκνότητα του y_k δοθέντος ότι $Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}$. Η εξίσωση (III.B.99) ισχύει για κάθε πυκνότητα στο \mathbb{R}^n και συνεπάγεται εύκολα από το γεγονός ότι $p_j(y_k|y_1, \dots, y_{k-1}) = p_j(y_1, \dots, y_k) / p_j(y_1, \dots, y_{k-1})$. Ως προς το H_0 , το Y_k είναι ανεξάρτητο του $\bar{Y}_1, \dots, \bar{Y}_{k-1}$ αφού το \underline{Y} είναι i.i.d., άρα $p_0(\underline{y}) = \prod_{k=1}^n p_0(y_k)$. Ως προς το H_1 , το Y_k δεν είναι ανεξάρτητο του $\bar{Y}_1, \dots, \bar{Y}_{k-1}$. Ωστόσο, αφού το \underline{Y} είναι ένα Gauss τυχαίο διάλυμα ως προς το H_1 , το Y_k είναι κατά συνθήκη Gauss, δοθέντος ότι $Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}$. Το μέσο αυτής της κατά συνθήκη πυκνότητας δίνεται από

$$\begin{aligned} E_1\{Y_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}\} \\ &= E_1\{S_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}\} \\ &\quad + E_1\{N_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}\} \\ &= E_1\{S_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}\} \triangleq \hat{S}_k, \end{aligned} \quad (\text{III.B.100})$$

όπου χρησιμοποιήσαμε το δεδομένο ότι το N_k είναι ανεξάρτητο του $\bar{Y}_1, \dots, \bar{Y}_{k-1}$ και έχει μηδενικό μέσο. Παρομοίως, η διακύμανση της κατά συνθήκη πυκνότητας είναι

$$\begin{aligned} \text{Var}_1(Y_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}) \\ &= \text{Var}_1(S_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}) \\ &\quad + \text{Var}_1(N_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1}) \\ &= \hat{\sigma}_{\hat{S}_k}^2 + \sigma^2, \end{aligned} \quad (\text{III.B.101})$$

όπου $\hat{\sigma}_{\hat{S}_k}^2 \triangleq \text{Var}_1(S_k|Y_1 = y_1, \dots, Y_{k-1} = y_{k-1})$.

Μια ιδιότητα της πολυμεταβλητής κατά Gauss κατανομής είναι ότι το $\hat{\sigma}_{\hat{S}_k}^2$ δεν εξαρτάται από τις τιμές του y_1, \dots, y_{k-1} . (Αυτή και άλλες σχετικές ιδιότητες της πολυμεταβλητής κατά Gauss κατανομής αναπτύσσονται σε ακόλουθα κεφάλαια). Άρα με $\hat{S}_1 = E\{S_1\}$ και $\hat{\sigma}_{\hat{S}_1}^2 = \text{Var}(S_1)$, το $p_1(\underline{y})$ είναι το προϊόν των $\mathcal{N}(\hat{S}_k, \hat{\sigma}_{\hat{S}_k}^2 + \sigma^2)$ πυκνοτήτων και ο λόγος λογαριθμικής πιθανοφάνειας γίνεται

$$\log L(\underline{y}) = \frac{1}{2} \sum_{k=1}^n y_k^2 / \sigma^2 - \frac{1}{2} \sum_{k=1}^n (y_k - \hat{S}_k)^2 / (\hat{\sigma}_{\hat{S}_k}^2 + \sigma^2)$$

$$+ \frac{1}{2} \sum_{k=1}^n \log[\sigma^2 / (\hat{\sigma}_{S_k}^2 + \sigma^2)]. \quad (\text{III.B.102})$$

Συγκρίνοντας το (III.B.102) με το (III.B.98) βλέπουμε ότι η ανίχνευση ενός εξαρτημένου στοχαστικού σήματος είναι ανάλογη με την ανίχνευση ενός ανεξάρτητου στοχαστικού σήματος με μέσο \hat{S} και συμμεταβλητή $\text{diag}\{\hat{\sigma}_{S_1}^2, \dots, \hat{\sigma}_{S_k}^2\}$. Φυσικά η διαφορά είναι ότι το \hat{S}_k στο (III.B.102) εξαρτάται από το y_1, \dots, y_{k-1} , ενώ το μ_k στο (III.B.98) δεν εξαρτάται.

Άλλος ένας τρόπος για να το δούμε αυτό είναι να γράψουμε, ως προς το H_1 ,

$$Y_k = N_k + S_k = N_k + \epsilon_k + \hat{S}_k, \quad (\text{III.B.103})$$

όπου ερμηνεύουμε το \hat{S}_k ως το τυχαίο μέγεθος $E_1\{S_k | Y_1, \dots, Y_{k-1}\}$ και

$\epsilon_k = (S_k - \hat{S}_k)$. Όπως θα δούμε σε ακόλουθα κεφάλαια το μέγεθος \hat{S}_k είναι μια βέλτιστη προβλέπουσα (ως προς το H_1) του S_k από τις προηγούμενες παρατηρήσεις Y_1, \dots, Y_{k-1} . Άρα το ϵ_k μπορεί να ερμηνευθεί ως το σφάλμα στην πρόβλεψη αυτή ή, ισοδύναμα, ως τμήμα του S_k που δεν μπορεί να προβλεφθεί από τις προηγούμενες παρατηρήσεις. Έτσι, καθώς μελετάμε κάθε παρατήρηση μπορούμε να θεωρούμε ότι το σήμα

συνίσταται σ' ένα τμήμα, \hat{S}_k , γνωστό από το παρελθόν και σ' ένα νέο τμήμα ϵ_k που δεν μπορεί να προβλεφθεί από το παρελθόν⁴. Μπορεί να αποδειχθεί ότι ως προς το H_1 , το ϵ_k είναι στατιστικά ανεξάρτητο του Y_1, \dots, Y_{k-1} και ότι είναι μια $\mathcal{N}(0, \hat{\sigma}_{S_k}^2)$ τυχαία μεταβλητή. Συγκρίνοντας, στην περίπτωση στην οποία το S_1, \dots, S_n είναι μια ανεξάρτητη $\mathcal{N}(\mu_k, \sigma_{S_k}^2)$ ακολουθία, το \hat{S}_k ισούται με το μ_k και $\epsilon_k = S_k - \mu_k$, το οποίο είναι $\mathcal{N}(0, \sigma_{S_k}^2)$.

4

⁴ Μια παρόμοια ερμηνεία μπορεί να δοθεί στο \hat{S}_k και το $N_k + \epsilon_k$ σε ό,τι αφορά το Y_k . Δηλ. το \hat{S}_k είναι το μέρος του Y_k που είναι γνωστό από το παρελθόν και το $Y_k - \hat{S}_k = N_k + \epsilon_k$ είναι το μέρος του Y_k που δεν μπορεί να προσδιοριστεί από το παρελθόν. Αυτό σημαίνει ότι το $Y_k - \hat{S}_k$ περιλαμβάνει τις νέες πληροφορίες του $Y_k - \hat{S}_k$. Η ακολουθία $(Y_1 - \hat{S}_1), \dots, (Y_k - \hat{S}_k)$ είναι ενίοτε γνωστή ως καινοτόμος ακολουθία. Η ιδέα αυτή διαδραματίζει σημαντικό ρόλο στο φιλτράρισμα και θα αναπτυχθεί περαιτέρω.

Παρατήρηση III.B.5: Ερμηνεία εκτιμητή συσχέτισης του βέλτιστου ανιχνευτή για στοχαστικά σήματα σε i.i.d. κατά Gauss θόρυβο.

Ως περαιτέρω σχόλιο στη δομή του (III.B.102), σημειώστε ότι μπορούμε να γράψουμε την $L(\underline{y})$ ως

$$\log L(\underline{y}) = \frac{1}{2\sigma^2} \left[\sum_{k=1}^n y_k^2 - \sum_{k=1}^n (y_k - \hat{S}_k)^2 / (1 + \hat{\sigma}_{S_k}^2 / \sigma^2) \right] - \frac{1}{2} \sum_{k=1}^n \log(1 + \hat{\sigma}_{S_k}^2 / \sigma^2). \quad (\text{III.B.104})$$

Ας υποθέσουμε ότι η διακύμανση θορύβου σ^2 είναι ευρέως σχετική με τη μέγιστη διακύμανση σφάλματος πρόβλεψης $\max_{1 \leq k \leq n} \hat{\sigma}_{S_k}^2$, δηλ. ας υποθέσουμε ότι $\hat{\sigma}_{S_k}^2 \ll \sigma^2$ για όλα τα k .

Στην περίπτωση αυτή $(1 + \hat{\sigma}_{S_k}^2 / \sigma^2) \cong 1$ και μπορούμε να γράψουμε

$$\log L(\underline{y}) \cong \frac{1}{\sigma^2} \left[\sum_{k=1}^n y_k \hat{S}_k - \frac{1}{2} \sum_{k=1}^n (\hat{S}_k)^2 \right]. \quad (\text{III.B.105})$$

Συγκρίνοντας αυτό με το (III.B.9), βλέπουμε ότι (III.B.105) είναι η δομή ανίχνευση του $\hat{S}_1, \dots, \hat{S}_n$ σαν να ήταν ένα *συνεκτικό* σήμα. Αυτό σημαίνει ότι, μπορούμε να δούμε τον ανιχνευτή στοχαστικού σήματος στην περίπτωση αυτή, τουλάχιστον κατά προσέγγιση, ως ένα που εκτιμά το σήμα και κατόπιν το διαχειρίζεται ως ένα γνωστό σήμα.

Γενικότερα, ας υποθέσουμε ότι έχουμε ένα στοχαστικό σχήμα με πολυμεταβλητή συνάρτηση πυκνότητας πιθανότητας $p_{\underline{S}}$, όχι απαραίτητα Gauss, ενσωματωμένη σε $\mathcal{N}(\underline{0}, \sigma^2 \mathbf{I})$ θόρυβο. Ο λόγος πιθανοφάνειας για το (III.B.85) γίνεται

$$L(\underline{y}) = \int_{\mathbf{R}^n} \exp \left\{ \frac{1}{\sigma^2} \left(\underline{s}^T \underline{y} - \frac{1}{2} \|\underline{s}\|^2 \right) \right\} p_{\underline{S}}(\underline{s}) d\underline{s}. \quad (\text{III.B.106})$$

Εντός κανονικότητας στο $p_{\underline{S}}$, το θεώρημα μέσης τιμής [βλέπε Apostol (1974)] συνεπάγεται ότι

$$L(\underline{y}) = \exp \left\{ \frac{1}{\sigma^2} \left(\underline{\hat{S}}^T \underline{y} - \frac{1}{2} \|\underline{\hat{S}}\|^2 \right) \right\} \quad (\text{III.B.107})$$

για κάποιο $\underline{\hat{S}}$ στο \mathbb{R}^n (φυσικά, το $\underline{\hat{S}}$ εξαρτάται από το \underline{y}). Άρα βλέπουμε ότι, γενικά, ο λόγος πιθανοφάνειας για στοχαστικά σήματα σε i.i.d. κατά Gauss θόρυβο μπορεί να ερμηνευθεί ως ένας εκτιμητής, $\underline{\hat{S}}$, του σήματος που ακολουθείται από το βέλτιστο ανίχνευτή για το $\underline{\hat{S}}$ σαν να ήταν ένα συνεκτικό σήμα⁵. Η δομή αυτή είναι γνωστή ως *εκτιμητής συσχέτισης*. Η δομή αυτή δεν είναι ιδιαίτερα απλή, καθώς ο προσδιορισμός της συνάρτησης $\underline{\hat{S}}(\underline{y})$ μπορεί να είναι αρκετά δύσκολος. Επίσης, σε αντίθεση με το (III.B.105), το $\underline{\hat{S}}$ στο (III.B.107) εξαρτάται γενικά απ' όλα τα \underline{y} και δεν μπορεί να υπολογιστεί σε πραγματικό χρόνο. Ωστόσο, η δομή αυτή είναι ενδεικτική του πώς μπορεί κάποιος να σχεδιάσει ένα υποβέλτιστο σύστημα για την ανίχνευση στοχαστικών σημάτων σε i.i.d. κατά Gauss θόρυβο, δημιουργώντας πρώτα ένα σύστημα που θα μπορεί να εκτιμά το σήμα σαν να ήταν παρόν, και κατόπιν χρησιμοποιώντας αυτή την εκτίμηση σαν ένα γνωστό σήμα. Στο αναλογικό συνεχούς χρόνου του προβλήματος ανίχνευσης ενός στοχαστικού σήματος σε i.i.d. κατά Gauss θόρυβο, η δομή του εκτιμητή συσχέτισης προκύπτει μ' έναν πιο ευθύ τρόπο μ' έναν εύκολα χαρακτηριζόμενο εκτιμητή, όπως θα δούμε στο Κεφάλαιο VI.

Παρατήρηση III.B.6: Τοπικά βέλτιστη ανίχνευση στοχαστικών σημάτων

Ένα πρόβλημα με τον τετραγωνικό ανίχνευτή του (III.B.86) φαίνεται μέσω της εξέτασης του προβλήματος ελέγχου σύνθετης υπόθεσης

$$\begin{aligned} & H_0 : Y_k = N_k, \quad k = 1, 2, \dots, n \\ \text{versus} & H_1 : Y_k = N_k + \theta^{1/2} S_k, \quad k = 1, 2, \dots, n, \quad \theta > 0, \end{aligned} \quad (\text{III.B.108})$$

όπου τα \underline{N} και \underline{S} είναι όπως στο (III.B.85) με μοναδιαία διακύμανση θορύβου. Στην περίπτωση η συνδιακύμανση του $\theta^{1/2} \underline{S}$ είναι $\theta \underline{\Sigma}_S$ και έτσι η σχετική τετραγωνική στατιστική ανίχνευσης για έλεγχο Neyman – Pearson με σταθερό θ είναι

$$\theta \underline{y}^T \underline{\Sigma}_S (\mathbf{I} + \theta \underline{\Sigma}_S)^{-1} \underline{y}. \quad (\text{III.B.109})$$

Παρόλο που ο κύριος θ συντελεστής μπορεί να απορροφηθεί στο κατώφλι απόφασης, το θ που εμφανίζεται στο $(\mathbf{I} + \theta \Sigma_S)^{-1}$ δεν μπορεί να αποσυνδεθεί από τις παρατηρήσεις \underline{y} . Άρα, δεν υπάρχει έλεγχος UMP για το (III.B.108). Ωστόσο, μπορεί να βρεθεί ένας LMP στατιστικός έλεγχος με διαφορισμό του (III.B.109) σε ό,τι αφορά το θ και θέτοντας το θ στο μηδέν. Κάνοντας αυτό, έχουμε έναν LMP στατιστικό έλεγχο

$$2\underline{y}^T \Sigma_S \underline{y}. \quad (\text{III.B.110})$$

5

Η στατιστική του (III.B.110) έχει μια ενδιαφέρουσα ερμηνεία για την περίπτωση στην οποία το $k - l$ στοιχείο του Σ_S , ας πούμε $\rho_{k,l}$ εξαρτάται μόνο από τη διαφορά $(k - l)$. Στην περίπτωση αυτή μπορούμε να γράψουμε $\rho_{k,l} = \rho_{k-l,0} \stackrel{\Delta}{=} \rho_{k-l}$, όπου εξαλείφουμε το δεύτερο δείκτη για την ευκολία του συμβολισμού. Ένα σήμα με αυτή την ιδιότητα θεωρείται ότι είναι *στατικό με την ευρεία έννοια*, μια έννοια στην οποία θα επιστρέψουμε στο Κεφάλαιο V. Εξετάστε την ανηγμένη LMP στατιστική

$$T(\underline{y}) = \frac{1}{n} \underline{y}^T \Sigma_S \underline{y}, \quad (\text{III.B.111})$$

που είναι ισοδύναμη με το (III.B.110), αφού η αναγωγή ανακλιμακώνει μερικώς το κατώφλι. Χρησιμοποιώντας λίγη άλγεβρα (βλ. Άσκηση 18), μπορούμε να γράψουμε το $T(\underline{y})$ ως

$$T(\underline{y}) = \frac{1}{n} \sum_{k=1}^n \sum_{l=1}^n y_k y_l \rho_{k-l} = \rho_0 \hat{\rho}_0 + 2 \sum_{k=1}^{n-1} \rho_k \hat{\rho}_k, \quad (\text{III.B.112})$$

όπου το $\hat{\rho}_k$ προσδιορίζεται από

$$\hat{\rho}_k \stackrel{\Delta}{=} \frac{1}{n-k} \sum_{l=1}^{n-k} y_l y_{l+k}, \quad k = 0, \dots, n-1. \quad (\text{III.B.113})$$

Η αναπαράσταση του (III.B.112) οδηγεί στην ακόλουθη ερμηνεία της LMP στατιστικής (III.B.111). Ας σημειωθεί ότι για $n \gg k$, το $\hat{\rho}_k$ είναι μια εκτίμηση της συνδιακύμανσης $E\{Y_l Y_{l+k}\}$ για $l = 1, \dots, n-k$. Άρα, το $T(\underline{y})$ εκτιμά τη δομή συνδιακύμανσης των παρατηρήσεων και μετά τη συσχετίζει με την ακολουθία συνδιακύμανσης σήματος. Ως προς το H_0 έχουμε

⁵ Μια τέτοια ερμηνεία ισχύει για κάθε πυκνότητα θορύβου p_N που είναι επαρκώς κανονική, αφού για κάθε $\underline{y} \in \mathbb{R}^n$,

$$L(\underline{y}) = \int_{\mathbb{R}^n} \frac{p_N(\underline{y} - \underline{s})}{p_N(\underline{y})} p_{\underline{s}}(\underline{s}) d\underline{s} = \frac{p_N(\underline{y} - \underline{s})}{p_N(\underline{y})}$$

για κάποια $\underline{s} \in \mathbb{R}^n$.

$$E\{Y_l Y_{l+k}\} = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0, \end{cases} \quad (\text{III.B.114})$$

Ως προς το H_1 έχουμε

$$E\{Y_l Y_{l+k}\} = \begin{cases} 1 + \theta \rho_0 & \text{if } k = 0 \\ \theta \rho_k & \text{if } k \neq 0. \end{cases} \quad (\text{III.B.115})$$

Άρα, υποθέτοντας ότι οι εκτιμήσεις $\hat{\rho}_k$ ήταν σχετικά ακριβείς, θα είχαμε

$$T(\underline{y}) \cong \begin{cases} \rho_0 & \text{under } H_0 \\ \rho_0 + \theta \left(\rho_0^2 + 2 \sum_{k=1}^{n-1} \rho_k^2 \right) & \text{under } H_1. \end{cases} \quad (\text{III.B.116})$$

Από το (III.B.116), βλέπουμε ότι η στατιστική $T(\underline{y})$ είναι ένας διαισθητικά λογικός τρόπος ανίχνευσης του σήματος, ειδικά αν το σήμα είναι ιδιαίτερα συσχετιζόμενο (δηλ., το $\sum_{k=0}^n \rho_k^2$ είναι μεγάλο).

Η στατιστική $T(\underline{y})$ έχει επίσης μια ενδιαφέρουσα ερμηνεία στο πεδίο της συχνότητας. Ειδικότερα, ας σκεφτούμε το S_1, \dots, S_n ως ένα τμήμα μιας άπειρης τυχαίας ακολουθίας $\{S_k\}_{k=-\infty}^{\infty}$ με $E\{S_l S_{l+k}\} = \rho_k$, για όλους τους ακέραιους l και k . Τότε, ο διακριτού χρόνου μετασχηματισμός Fourier της ακολουθίας $\{\rho_k\}_{k=-\infty}^{\infty}$, κυρίως

$$\phi(\omega) \triangleq \sum_{k=-\infty}^{\infty} \rho_k e^{i\omega k}, \quad (\text{III.B.117})$$

είναι το φάσμα ισχύος του $\{S_k\}_{k=-\infty}^{\infty}$. Αφού το $\rho_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\omega) e^{-i\omega k} d\omega$, το $T(\underline{y})$ μπορεί να ξαναγραφεί ως

$$T(\underline{y}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(\omega) \hat{\phi}(\omega) d\omega, \quad (\text{III.B.118})$$

όπου

$$\hat{\phi}(\omega) \triangleq \frac{1}{n} \left| \sum_{k=1}^n y_k e^{i\omega k} \right|^2, \quad -\pi \leq \omega \leq \pi. \quad (\text{III.B.119})$$

Η συνάρτηση $\hat{\phi}$ είναι γνωστή ως *περιοδόγραμμα* των δεδομένων και είναι μια εκτίμηση του φάσματος της παρατήρησης. Άρα, στη μορφή του (III.B.118), το $T(\mathbf{y})$ εκτιμά το φάσμα παρατήρησης και συσχετίζει την εκτίμηση αυτή (στο πεδίο συχνότητας) με το φάσμα σήματος. Καθώς το φάσμα παρατήρησης ισούται με 1 για $\omega \in [-\pi, \pi]$ ως προς το H_0 και ισούται με $1 + \phi(\omega)$ για $\omega \in [-\pi, \pi]$ ως προς το H_1 , η πράξη του (III.B.118) έχει μια ερμηνεία παρόμοια με αυτή του (III.B.112).

III.C Αξιολόγηση Επίδοσης των Διαδικασιών Ανίχνευσης Σήματος

Στην Ενότητα III.B. αναλύθηκε ο σχεδιασμός και η ανάλυση των βέλτιστων διαδικασιών για ανίχνευση διακριτού χρόνου σήματος. Κατά μία έννοια, ο σχεδιασμός τέτοιων διαδικασιών είναι πιο άμεσος από την ανάλυση της επίδοσής τους, λόγω του δυσεπίλυτου αυτού του προβλήματος. Στάθηκε δυνατό να υπολογίσουμε την επίδοση στα περισσότερα από τα παραδείγματα που παρατέθηκαν στην Ενότητα III.B. μόνο επειδή τα συγκεκριμένα μοντέλα που εξετάστηκαν ανήκουν στα πιο ευεπίλυτα και άρα χρησιμοποιούνται για να περιγράψουν καλύτερα τη θεωρία. Κάποιες φορές, οι παραδοχές μπορεί να ποικίλουν ελαφρά από αυτές στα παραδείγματα αυτά και το ευεπίλυτο χάνεται. Για παράδειγμα, στο πρόβλημα ανίχνευσης ενός κατά Gauss σήματος σε i.i.d. κατά Gauss θόρυβο, ο υπολογισμός πιθανοτήτων σφάλματος είναι εφικτός αν και το σήμα είναι i.i.d., αλλά ανέφικτος αν η διακύμανση ενός μόνο δείγματος σήματος αλλάζει.

Τα βασικά μέτρα επίδοσης ενός δυαδικού συστήματος ανίχνευσης σήματος που χρησιμοποιεί ένα κανόνα απόφασης $\tilde{\delta}$ είναι οι δύο κατά συνθήκη πιθανότητες σφάλματος P_F και P_M που ορίζονται από

$$P_F(\tilde{\delta}) = P_0 \text{ (το } \tilde{\delta} \text{ επιλέγει } H_1)$$

και

$$P_M(\tilde{\delta}) = P_1 \text{ (το } \tilde{\delta} \text{ επιλέγει } H_0)$$

Οι έλεγχοι λόγου πιθανοφάνειας και οι περισσότεροι άλλοι κανόνες απόφασης που μας ενδιαφέρουν έχουν τη μορφή

$$\tilde{\delta}_T(\mathbf{y}) = \begin{cases} 1 & \text{if } T(\mathbf{y}) > \tau, \\ \gamma & \text{if } T(\mathbf{y}) = \tau, \\ 0 & \text{if } T(\mathbf{y}) < \tau, \end{cases} \quad (\text{III.C.1})$$

όπου T είναι μια απεικόνιση από το (Γ, G) στο $(\mathbb{R}, \mathcal{B})$ (δηλ. ο λόγος λογαριθμικής πιθανοφάνειας). Άρα, η αξιολόγηση επίδοσης για τα περισσότερα συστήματα περιλαμβάνει υπολογισμό των πιθανοτήτων των πεδίων $\{T(Y) > \tau\}$ (ή $\{T(Y) < \tau\}$) ως προς τις δύο υποθέσεις. Παρόλο που αυτό το πρόβλημα είναι εννοιολογικά απλό, ο πραγματικός υπολογισμός των απαιτούμενων πιθανοτήτων είναι συχνά αναλυτικά δύσκολος. Για παράδειγμα, αν το $Y = (Y_1, \dots, Y_n)$ έχει κοινό pdf p_0 ως προς το H_0 , τότε

$$P_F(\tilde{\delta}_T) = \int \dots \int_{\{T(\mathbf{y}) > \tau\}} p_0(y_1, \dots, y_n) dy_1 \dots dy_n + \gamma \int \dots \int_{\{T(\mathbf{y}) = \tau\}} p_0(y_1, \dots, y_n) dy_1 \dots dy_n,$$

που είναι δύσκολο να υπολογιστεί για μεγάλο n χωρίς περαιτέρω απλοποίηση. Στην ενότητα αυτή μελετούμε αρκετές κοινώς χρησιμοποιούμενες τεχνικές για τον υπολογισμό, την οριοθέτηση ή την προσέγγιση της επίδοσης των συστημάτων ανίχνευσης.

III.C.1 ΑΜΕΣΟΣ ΥΠΟΛΟΓΙΣΜΟΣ ΕΠΙΔΟΣΗΣ

Ας σημειωθεί ότι για ένα σύστημα της μορφής (III.C.1), έχουμε

$$\begin{aligned} P_F(\tilde{\delta}_T) &= P(T(Y) > \tau | H_0) + \gamma P(T(Y) = \tau | H_0) \\ &= [1 - F_{T,0}(\tau)] + \gamma [F_{T,0}(\tau) - \lim_{\sigma \rightarrow \tau^-} F_{T,0}(\sigma)] \end{aligned}$$

και

$$\begin{aligned} P_M(\tilde{\delta}_T) &= P(T(Y) < \tau | H_1) + (1 - \gamma) P(T(Y) = \tau | H_1) \\ &= P(T(Y) \leq \tau | H_1) - \gamma P(T(Y) = \tau | H_1) \\ &= F_{T,1}(\tau) - \gamma [F_{T,1}(\tau) - \lim_{\sigma \rightarrow \tau^-} F_{T,1}(\sigma)], \end{aligned}$$

όπου F_{T_j} είναι η αθροιστική συνάρτηση κατανομής (cdf) του $T(Y)$ ως προς την υπόθεση H_j . Έτσι, για διαδικασία ανίχνευσης της μορφής του (III.C.1) (όπως ο ανιχνευτής λόγου πιθανοφάνειας), η αξιολόγηση διευκολύνεται αν τα cdf $F_{T,j}$ μπορούν να προσδιοριστούν εύκολα σε μια περιοχή του κατωφλιού τ .

Μια περίπτωση στην οποία μπορεί να γραφεί μια άμεση παράσταση για το $F_{T,j}$ είναι αυτή στην οποία το $\underline{Y} = (Y_1, \dots, Y_n)^T$ είναι ένα διάνυσμα ανεξάρτητων (πραγματικών) τυχαίων μεταβλητών και το $T(\underline{y})$ έχει τη δομή

$$T(\underline{y}) = \sum_{k=1}^n g_k(y_k),$$

όπου το $\{g_k\}_{k=1}^n$ είναι μια ακολουθία από μη γραμμικότητες (δηλ. ο λόγος λογαριθμικής πιθανοφάνειας είναι αυτής της μορφής). Στην περίπτωση αυτή είναι συχνά απλό να υπολογίσουμε το $F_{T,j}$, χρησιμοποιώντας χαρακτηριστικές συναρτήσεις. Θυμηθείτε ότι η χαρακτηριστική συνάρτηση (χ.σ.) μιας τυχαίας μεταβλητής X ορίζεται από

$$\phi_X(u) = E\{e^{iuX}\}, \quad u \in \mathbb{R},$$

όπου το i συμβολίζει τη νοητή μονάδα $\sqrt{-1}$. Ας σημειωθεί ότι ένα cdf F και η χ.σ. ϕ αποτελούν ένα μοναδικό ζεύγος⁶. Συμβολίζοντας με $\phi_{T,j}$ και $\phi_{g_{k,j}}$, οι χ.σ. ως προς το H_j των $T(\underline{Y})$ και $g_k(Y_k)$, αντίστοιχα, έχουμε με την ανεξαρτησία του Y_k ότι

$$\begin{aligned} \phi_{T,j}(u) &= E \left\{ \exp \left\{ iu \sum_{k=1}^n g_k(Y_k) \right\} | H_j \right\} \\ &= \prod_{k=1}^n E \{ \exp \{ iu g_k(Y_k) \} | H_j \} = \prod_{k=1}^n \phi_{g_{k,j}}(u). \end{aligned}$$

6

⁶ Μια εκτενής θεώρηση των χαρακτηριστικών συναρτήσεων αναφέρεται στον Lucaks (1960)

Η χ.σ. $\phi_{T,j}$ μπορεί να αντιστραφεί για να δώσει την cdf $F_{T,j}$ μέσω του τύπου [βλέπε, π.χ. Billingsley (1979)]

$$F_{T,j}(b) - F_{T,j}(a) = \lim_{U \rightarrow \infty} \int_{-U}^U \frac{e^{-iua} - e^{-iub}}{iu} \phi_{T,j}(u) du, \quad (\text{III.C.2})$$

που ισχύει για όλα τα a και b που είναι σημεία συνέχειας του $F_{T,j}$. Η γνώση του $[F_{T,j}(b) - F_{T,j}(a)]$ σε όλα τα σημεία συνέχειας είναι αρκετή για να προσδιοριστεί το $F_{T,j}$ μοναδικά αφού τα $\check{F}_{T,j}(-\infty) = 0$ και $F_{T,j}$ πρέπει να είναι δεξιά συνεχή. Αν η $T(\check{Y})$ είναι μια συνεχής τυχαία μεταβλητή ως προς το H_j , η αντιστροφή απλοποιείται, καθώς το $F_{T,j}$ έχει μια συνάρτηση πυκνότητας πιθανότητας $p_{T,j}$ που δίνεται σ' αυτήν την περίπτωση από

$$p_{T,j}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_{T,j}(u) e^{-iut} du,$$

Και άρα τα $p_{T,j}$ και $\phi_{T,j}$ είναι ένα ζεύγος μετασχηματισμών Fourier⁷. Έτσι, για παράδειγμα, στην τελευταία αυτή περίπτωση έχουμε

$$P_F(\check{\delta}_T) = \frac{1}{2\pi} \int_{\tau}^{\infty} \int_{-\infty}^{\infty} e^{-iut} \left[\prod_{k=1}^n \phi_{g_{k,0}}(u) \right] du dt, \quad (\text{III.C.3})$$

και παρόμοια και για το $P_M(\check{\delta})$

Παράδειγμα III.C.1 : Συσχετισμός ανίχνευσης σε θόρυβο Cauchy.

Ως ένα παράδειγμα της προηγούμενης προσέγγισης, εξετάστε την επίδοση του ανιχνευτή συσχέτισης (που ορίζεται από $g_k(y_k) = s_k y_k, k = 1, \dots, n$) στην ανίχνευση ενός συνεκτικού σήματος σε πρόσθετο θόρυβο Cauchy. Συγκεκριμένα, εξετάστε το ζεύγος υπόθεσης.

$$H_0 : Y_k = N_k, \quad k = 1, \dots, n$$

versus

$$H_1 : Y_k = N_k + s_k, \quad k = 1, \dots, n,$$

όπου η N_1, \dots, N_n είναι μια ακολουθία i.i.d. τυχαίων μεταβλητών, κάθε μία εκ των οποίων έχει Cauchy pdf.

7

$$p_{N_k}(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbf{R},$$

και όπου η s_1, \dots, s_n είναι μια ακολουθία γνωστού σήματος. Σ' αυτή την περίπτωση έχουμε

$$\begin{aligned} \phi_{g_{k,0}}(u) &= E\{e^{i u s_k N_k}\} = \phi_{N_k}(u s_k) \\ &= \mathcal{F}\{p_{N_k}\}_{u s_k} = e^{-|s_k u|}, \quad u \in \mathbf{R}. \end{aligned}$$

Άρα

$$\phi_{T,0}(u) = \prod_{k=1}^n \phi_{g_{k,0}}(u) = e^{-n \bar{|s|} |u|}$$

όπου $\bar{|s|} \triangleq (1/n) \sum_{k=1}^n |s_k|$. Αφού η $\phi_{T,0}(u)$ είναι απολύτως ολοκληρώσιμη) έχουμε

$$\begin{aligned} P_F(\tilde{\delta}_T) &= \frac{1}{2\pi} \int_{\tau}^{\infty} \int_{-\infty}^{\infty} e^{-iut} e^{-n \bar{|s|} |u|} du dt \\ &= \frac{1}{n \bar{|s|} \pi} \int_{\tau}^{\infty} \frac{1}{1 + (t/n \bar{|s|})^2} dt \\ &= \frac{1}{2} - \frac{1}{\pi} \text{Tan}^{-1} \left(\frac{\tau}{n \bar{|s|}} \right). \end{aligned}$$

⁷ Μια επαρκής (άλλα όχι απαραίτητη συνθήκη) για να είναι το $T(\underline{Y})$ συνεχές ως προς H_j είναι $\int_{-\infty}^{\infty} |\phi_{T,j}(u)| du < \infty$ (βλ. Breimman (1968))

Παρομοίως έχουμε $\phi_{T,1}(u) = \phi_{T,0}(u)e^{iun\bar{s}^2}$, όπου $\bar{s}^2 \triangleq (1/n) \sum_{k=1}^n s_k^2$, που μας δίνει ευθέως

$$P_M(\bar{\delta}_T) = \frac{1}{2} + \frac{1}{\pi} \text{Tan}^{-1} \left(\frac{\tau}{n|\bar{s}|} - \sqrt{\bar{s}^2} \right).$$

{Παρεμπιπτόντως, είναι ενδιαφέρον να σημειωθεί εδώ ότι μπορούμε να πετύχουμε μια μεγέθους- α ανίχνευση επιλέγοντας $\tau = n|\bar{s}| \tan(\frac{1}{2} - \alpha)$, περίπτωση στην οποία έχουμε ROCs

$$\begin{aligned} P_D(\bar{\delta}_T) &= 1 - P_M(\bar{\delta}_T) \\ &= \frac{1}{2} - \frac{1}{\pi} \text{Tan}^{-1} \left(\tan \left(\frac{1}{2} - \alpha \right) - \sqrt{\bar{s}^2} \right). \end{aligned} \quad (\text{III.C.4})$$

Η εξίσωση του (III.C.4) συνεπάγεται το εκπληκτικό αποτέλεσμα ότι η επίδοση βελτιώνεται εδώ μόνο με την αύξηση της μέσης ισχύος σήματος και όχι με την αύξηση του αριθμού των δειγμάτων. Αυτή η περίεργη συμπεριφορά οφείλεται στο γεγονός ότι ο ανιχνευτής συσχέτισης είναι αρκετά διαφορετικός από τον βέλτιστο ανιχνευτή για συνεκτικά σήματα σε θόρυβο Cauchy, για τα οποία η επίδοση βελτιώνεται με αυξημένο αριθμό δειγμάτων. Ουσιαστικά, η βαρύτητα των ουρών της κατανομής Cauchy ακυρώνει την επίδραση της μείωσης θορύβου που επιτυγχάνει ο συσχετιστής στην κατά Gauss περίπτωση.

Στο παράδειγμα III.C.1, οι ακριβείς κλειστές μορφές παραστάσεις για P_F και P_M προέκυψαν με τη χρήση χαρακτηριστικών συναρτήσεων. Άλλο παράδειγμα στο οποίο αυτή η μέθοδος παράγει κλειστές μορφές παραστάσεις είναι αυτό της ανίχνευσης ενός i.i.d. κατά Gauss σήματος σε i.i.d. κατά Gauss θόρυβο, που αναλύθηκε στην Ενότητα III.B. Ωστόσο, η προσέγγιση της χαρακτηριστικής συνάρτησης δεν αποδίδει πάντα επιλύσιμες κλειστές μορφές παραστάσεις για πιθανότητες σφάλματος και είναι πιο σύνηθες το (III.C.3) και η αντίστοιχη παράσταση για P_M να χρησιμοποιούνται ως μια βάση προσέγγισης των πιθανοτήτων σφάλματος. Για μέτριες τιμές των πιθανοτήτων σφάλματος, η ευθεία αριθμητική ολοκλήρωση του (III.C.3) και το αντίστοιχο P_M ολοκλήρωμα, χρησιμοποιώντας για παράδειγμα το λογάριθμο ταχέος μετασχηματισμού Fourier (FFT), μπορεί να δώσει κλειστές προσεγγίσεις των πιθανοτήτων σφαλμάτων. Σε πολλά πρακτικά προβλήματα,

ωστόσο, μία ή δύο από τις πιθανότητες σφάλματος είναι αρκετά μικρή ($\leq 10^{-5}$) και άλλες αριθμητικές μέθοδοι είναι συνήθως πιο αποτελεσματικές. Σε τέτοιες περιπτώσεις, υπάρχει συνήθως μια παράμετρος (το κατώφλι, ο λόγος σήματος προς θόρυβο, ο αριθμός δειγμάτων, κτλ) η οποία είναι πολύ μεγάλη και άρα η επίδοση μπορεί να προσεγγισθεί χρησιμοποιώντας ασυμπτωτικά αναπτύγματα των πιθανοτήτων σφάλματος στην παράμετρο αυτή. Μια τεχνική αυτού του τύπου που εφαρμόζεται ευρέως στην ανάλυση των συστημάτων επικοινωνίας είναι η τεχνική της προσέγγισης σημείου σέλας [βλ. π.χ. DeBruijn (1961)]. Παραδείγματα εφαρμογών αυτής της μεθόδου απαντούν στο Lugannani and Rice (1980), στο οποίο βρίσκονται προσεγγίσεις για την περίπτωση στην οποία τα $g_k(Y_k)$ είναι πανομοιότυπα κατανομημένα, και στο Mazo and Salz (1965), στο οποίο αναπτύσσονται οι προσεγγίσεις σημείου σέλας για τον τετραγωνικό ανιχνευτή της Ενότητας III.B. για μη i.i.d. κατά Gauss σήματα.

III.C.2. Chernoff και Σχετικά φράγματα

Είναι συχνά μη πρακτικό (ή και αδύνατο) να υπολογίσουμε ακριβώς τις πιθανότητες σφάλματος P_F και P_M των ανιχνευτών της μορφής του (III.C.1). Ωστόσο, στην πράξη είναι συνήθως αρκετό να έχουμε καλά άνω φράγματα σε αυτά τα μεγέθη. Ένας τύπος του φράγματος επίδοσης που χρησιμοποιείται ευρέως στο πλαίσιο αυτό είναι το φράγμα Chernoff. Το φράγμα Chernoff είναι ένα φράγμα στην επίδοση του λόγου πιθανοφάνειας των ανιχνευτών και βασίζεται στην ακόλουθη απλή ανισότητα

Ανισότητα του Markov: Έστω ότι X είναι μια τυχαία μεταβλητή. Αν $P(X \geq 0) = 1$, τότε $P(X \geq a) \leq E\{X\}/a$ για όλα τα $a > 0$.

Απόδειξη: $P(X \geq a) \leq E\{X\}/a$, όπου $I_{[a, \infty)}$ είναι η δείκτρια συνάρτηση του συνόλου $[a, \infty)$, που ορίζεται από

$$I_{[a, \infty)}(x) = \begin{cases} 1 & \text{if } x \geq a \\ 0 & \text{if } x < a. \end{cases}$$

Αφού $X \geq 0$, παρατηρούμε ότι $I_{[a, \infty)}(X) \leq X/a$, και η ανισότητα του Markov έπεται.

Για να εφαρμόσουμε την ανισότητα Markov στο σύστημα του (III.C.1), παρατηρούμε ότι με την τυχαιοποίηση $\gamma \leq 1$, μπορούμε να γράψουμε

$$\begin{aligned}
P_F(\bar{\delta}_T) &\leq P_0(T(Y) \geq \tau) = P_0(e^{sT(Y)} \geq e^{s\tau}) \\
&\leq e^{-s\tau} E\{e^{sT(Y)} | H_0\} = \exp\{-s\tau + \mu_{T,0}(s)\} \quad (\text{III.C.5})
\end{aligned}$$

για όλα τα $s > 0$, όπου $\mu_{T,0}$ είναι η αθροιστική γεννήτρια συνάρτηση (cgf) του $T(Y)$ ως προς το H_0 που ορίζεται από

$$\mu_{T,0}(s) = \log(E\{e^{sT(Y)} | H_0\}).$$

Παρομοίως, αφού $\gamma \geq 0$, έχουμε

$$\begin{aligned}
P_M(\bar{\delta}_T) \leq P_1(T(Y) \leq \tau) &= P_1(e^{tT(Y)} \geq e^{t\tau}) \\
&\leq \exp\{-t\tau + \mu_{T,1}(t)\} \quad (\text{III.C.6})
\end{aligned}$$

για όλα τα $t < 0$, όπου $\mu_{T,1}$ είναι η cgf του $T(Y)$ ως προς το H_1 . Τα φράγματα των (III.C.5) και (III.C.6) μπορούν να ελαχιστοποιηθούν στο $s > 0$ και $t < 0$ για να βρουν τα ισχυρότερα τέτοια φράγματα, με την προϋπόθεση ότι οι cgf του $T(Y)$ είναι γνωστές. Τα φράγματα αυτά είναι ιδιαίτερα χρήσιμα για τον ανιχνευτή λόγου πιθανοφάνειας. Για να εξετάσουμε την περίπτωση αυτή, υποθέτουμε ότι το P_j έχει πυκνότητα p_j για $j = 0$ και 1 , και επιλέγουμε $T(y) = \log L(y)$, όπου $L = p_1/p_0$. Στην περίπτωση αυτή έχουμε

$$\mu_{T,0}(s) = \log \left(\int_{\Gamma} e^{s \log L} p_0 d\mu \right) = \log \left(\int_{\Gamma} L^s p_0 d\mu \right)$$

και

$$\begin{aligned}
\mu_{T,1}(t) &= \log \left(\int_{\Gamma} L^t p_1 d\mu \right) = \log \left(\int_{\Gamma} L^{t+1} p_0 d\mu \right) \\
&= \mu_{T,0}(t+1).
\end{aligned}$$

Άρα, μπορούμε να ξαναγράψουμε το φράγμα του (III.C.6) ως

$$P_M(\delta_T) \leq \exp\{(1-s)\tau + \mu_{T,0}(s)\}, \quad s < 1. \quad (\text{III.C.7})$$

Σημειώστε ότι και τα δύο φράγματα των (III.C.5) και (III.C.7) επιτυγχάνουν τα ελάχιστα τους στην ίδια τιμή του s αν

$$\arg\{\min_{s < 1} [\mu_{T,0}(s) - s\tau]\} > 0$$

και

$$\arg\{\min_{s > 0} [\mu_{T,0}(s) - s\tau]\} < 1.$$

Αποδεικνύεται ευθέως ότι η $[\mu_{T,0}(s) - s\tau]$ είναι μια κυρτή συνάρτηση του s στο πεδίο όπου $\mu_{T,0}(s) < \infty$ (που πρέπει να είναι διάστημα. Βλέπε Billingsley (1979)). Επομένως, η συνθήκη

$$\mu'_{T,0}(s) = \tau, \quad (\text{III.C.8})$$

όπου το $\mu'_{T,0}(s) = d\mu_{T,0}(s)/ds$ είναι επαρκές για ένα ελάχιστο του $[\mu_{T,0}(s) - s\tau]$. Θα πρέπει επίσης να αποδειχθεί ότι αν $E\{|\log L(Y)| | H_j\}$ για $j = 0$ ή 1 , τότε

$$\mu'_{T,0}(j) = E\{\log L(Y) | H_j\}. \quad (\text{III.C.9})$$

Η κυρτότητα του $[\mu_{T,0}(s) - s\tau]$ συνεπάγεται ότι $\mu'_{T,0}(s)$ είναι μια αυξανόμενη συνάρτηση του s , και άρα το (III.C.9) συνεπάγεται ότι το (III.C.8) έχει μια λύση με $s \in (0, 1)$ αν (και μόνο αν)

$$\mu_0 \triangleq E\{\log L(Y) | H_0\} < \tau < E\{\log L(Y) | H_1\} \triangleq \mu_1. \quad (\text{III.C.10})$$

Άρα, με την παραδοχή ότι το (III.C.10) είναι έγκυρο, τα (III.C.5) και (III.C.7) γίνονται

$$P_F(\bar{\delta}_T) \leq \exp\{\mu_{T,0}(s_0) - s_0\mu'_{T,0}(s_0)\} \quad (\text{III.C.11})$$

$$P_M(\bar{\delta}_T) \leq \exp\{\mu_{T,0}(s_0) + (1 - s_0)\mu'_{T,0}(s_0)\} \quad (\text{III.C.12})$$

με

$$\mu_0 < \mu'_{T,0}(s_0) = \tau < \mu_1. \quad (\text{III.C.13})$$

Οι ισότητες των (III.C.11) και (III.C.12) είναι γνωστές ως *φράγμα Chernoff*. Έπεται από την κυρτότητα του $[\mu_{0,T}(s) - s\tau]$ ότι αν $\tau \leq \mu_0$, τότε $\min_{s \geq 0} [\mu_{T,0}(s) - s\tau] = 0$ που συνεπάγεται ότι το φράγμα του (III.C.5) είναι τετριμμένο (δηλ. είναι ≥ 1) και παρομοίως ότι αν $\tau \geq \mu_1$, τότε το φράγμα του (III.C.7) είναι τετριμμένο, παρόλο που σε άλλη περίπτωση το άλλο φράγμα είναι μη τετριμμένο. Ας σημειωθεί ότι η ανισότητα του Jensen⁸ συνεπάγεται ότι $\mu_0 \leq 0$ και $\mu_1 \geq 0$ με $\mu_0 = 0$ και/ή $\mu_1 = 0$ αν και μόνο αν $P_0 = P_1$, οπότε ένα κατώφλι του $\tau = 0$ πάντα ικανοποιεί το (III.C.10).

Σημειώστε ότι αν τα πρότερα π_0 και π_1 είναι γνωστά, τότε (III.C.11) και (III.C.12) παράγουν ένα ανώτερο φράγμα στην μέση πιθανότητα σφάλματος. Συγκεκριμένα,

$$\begin{aligned} P_e &= \pi_0 P_F + \pi_1 P_M \\ &\leq [\pi_0 + \pi_1 e^{\mu'_{T,0}(s_0)\tau}] \exp\{\mu_{T,0}(s_0) - s_0\mu'_{T,0}(s_0)\} \end{aligned} \quad (\text{III.C.14})$$

Ωστόσο, ένα καλύτερο φράγμα στο P_e μπορεί να ληφθεί, παρατηρώντας ότι (βλ. Άσκηση 20)

$$P_e \leq \pi_0 e^{-s\tau} \int_{\Gamma_1} L^s p_0 d\mu + \pi_1 e^{(1-s)\tau} \int_{\Gamma_0} L^s p_0 d\mu \quad (\text{III.C.15})$$

για $0 \leq s \leq 1$, όπου $\Gamma_1 = \{L(Y) \geq \tau\}$ και $\Gamma_0 = \Gamma_1^c$. Η ισότητα του (III.C.15) συνεπάγεται ότι

$$P_e \leq \max\{\pi_0, \pi_1 e^\tau\} \exp\{\mu_{T,0}(s) - s\tau\}, \quad 0 \leq s \leq 1. \quad (\text{III.C.16})$$

Υπό τη συνθήκη του (III.C.13), το καλύτερο φράγμα του μορφής του (III.C.16) είναι

$$P_e \leq \max\{\pi_0, \pi_1 e^{\mu'_{T,0}(s_0)}\} \times \exp\{\mu_{T,0}(s_0) - s_0 \mu'_{T,0}(s_0)\}, \quad (\text{III.C.17})$$

που είναι ένα πιο αυστηρό φράγμα από ό,τι είναι το (III.C.14) εκτός κι αν π_0 είναι 0 ή 1. Ας σημειωθεί ότι ένας ανιχνευτής ελάχιστης πιθανότητας σφάλματος λόγω λογαριθμικής πιθανοφάνειας χρησιμοποιεί το κατώφλι $\tau = \log(\pi_0/\pi_1)$, περίπτωση στην οποία το (III.C.16) ανάγεται σε

$$P_e \leq \pi_0^{1-s} \pi_1^s e^{\mu_{T,0}(s)}, \quad 0 \leq s \leq 1. \quad (\text{III.C.18})$$

Μια ενδιαφέρουσα περίπτωση της προηγούμενης ανάλυσης είναι αυτή στην οποία το $\Gamma = \mathbb{R}^n$ και το $Y = (Y_1, \dots, Y_n)$ είναι μια ακολουθία ανεξάρτητα και ταυτόσημα κατανεμημένων παρατηρήσεων με οριακή πυκνότητα f_j ως προς την υπόθεση H_j . Στην περίπτωση αυτή έχουμε ευθέως ότι

$$\mu_{T,0}(s) = n \log \left(\int_{\mathbf{R}} f_1^s f_0^{1-s} \right),$$

και έπεται από την ανισότητα του Jensen ότι για $0 < s < 1$,

$$\log \left(\int_{\mathbf{R}} f_1^s f_0^{1-s} \right) < 0.$$

Άρα το (III.C.18) συνεπάγεται ότι η πιθανότητα σφάλματος σ' αυτή την περίπτωση μειώνεται τουλάχιστον σ' ένα εκθετικό ρυθμό καθώς ο αριθμός των δειγμάτων (n) αυξάνεται.

⁸ 8. Ανισότητα Jensen: Για κάθε τυχαία μεταβλητή X και κυρτή συνάρτηση C έχουμε $E\{C(X)\} \geq C(E\{X\})$ με ισότητα αν και μόνο αν $P(X = E\{X\}) = 1$ όταν το C είναι αυστηρά κυρτό.

Παράδειγμα III.C.2: Το φράγμα Chernoff για τετραγωνική ανίχνευση

Για να περιγράψουμε τη χρήση του φράγματος Chernoff, εξετάστε το πρόβλημα ανίχνευσης ενός $\mathcal{N}(\underline{0}, \Sigma_S)$ σήματος σε $\mathcal{N}(\underline{0}, \sigma^2 \mathbf{I})$ θόρυβο. Μετά το μετασχηματισμό του Y_1, \dots, Y_n στην ανεξάρτητη ακολουθία $\tilde{Y}_1, \dots, \tilde{Y}_n$ όπως στο (III.B.87) έχουμε ότι

$$L(\underline{y}) = \prod_{k=1}^n \frac{\sigma_{0k}}{\sigma_{1k}} e^{y_k^2}$$

με σ_{jk}^2 από το (III.B.90). Άρα με $T = \log L$ έχουμε

$$\begin{aligned} \mu_{T,0}(s) &= \log \left(E \left\{ \prod_{k=1}^n \left(\frac{\sigma_{0k}}{\sigma_{1k}} \right)^s e^{s \tilde{Y}_k^2} | H_0 \right\} \right) \\ &= \sum_{k=1}^n s \log \left(\frac{\sigma_{0k}}{\sigma_{1k}} \right) + \sum_{k=1}^n \log(E\{e^{s \tilde{Y}_k^2} | H_0\}). \quad (\text{III.C.19}) \end{aligned}$$

Η προσδοκία στο (III.C.19) δίνεται από

$$E\{e^{s \tilde{Y}_k^2} | H_0\} = \begin{cases} [1 - s\sigma_{0k}^2]^{-1/2} & \text{if } s < 1/\sigma_{0k}^2 \\ \infty & \text{if } s \geq 1/\sigma_{0k}^2. \end{cases}$$

Αφού $\overline{\sigma_{0k}^2} < 1$ για όλα τα k , το (III.C.19) είναι πεπερασμένο για όλα τα k όταν $s \leq 1$, και άρα μπορούμε άμεσα να ορίσουμε, ας πούμε, την ελάχιστη πιθανότητα σφάλματος με

$$P_e \leq \pi_0^{1-s} \pi_1^s \prod_{k=1}^n \frac{\sigma^s (\sigma^2 + \lambda_k)^{(1-s)/2}}{[\sigma^2 + (1-s)\lambda_k]^{1/2}}, \quad 0 \leq s \leq 1. \quad (\text{III.C.20})$$

Το φράγμα ελαχιστοποιείται με την τιμή s_0 λύνοντας

$$2 \log(\pi_0/\pi_1) + \sum_{k=1}^n \log(1 + \lambda_k/\sigma^2) = \sum_{k=1}^n \frac{\lambda_k}{\sigma^2 + (1 - s_0)\lambda_k}, \quad (\text{III.C.21})$$

που μπορεί να λυθεί ευθέως για την i.i.d. περίπτωση $(\lambda_1 = \dots = \lambda_n)$ και αριθμητικά για άλλες περιπτώσεις.

Υπάρχουν αρκετά άλλα φράγματα παρόμοια με το φράγμα Chernoff που είναι χρήσιμα σε εφαρμογές. Ένα τέτοιο φράγμα είναι η ειδική περίπτωση του (III.C.18) με $s = \frac{1}{2}$, για την οποία έχουμε

$$P_e \leq \sqrt{\pi_0\pi_1} \exp\{\mu_{T,0}(\frac{1}{2})\}. \quad (\text{III.C.22})$$

Το φράγμα αυτό καλείται, ενίοτε, φράγμα Bhattacharyya. Το μέγεθος

$$\rho \triangleq \exp\{\mu_{T,0}(\frac{1}{2})\} = \int_{\Gamma} [p_0 p_1]^{1/2} d\mu$$

λέγεται συχνά συντελεστής Bhattacharyya ή ολοκλήρωμα του Hellinger (επίσης και *συνάφεια*) και υπάρχουν επίσης διαθέσιμα κάτω φράγματα στο P_e σε ό,τι αφορά το ρ . Για παράδειγμα, έχουμε [βλέπε, Kobayashi and Thomas (1967)]

$$\pi_0\pi_1\rho^2 \leq P_e \leq (\pi_0\pi_1)^{1/2}\rho. \quad (\text{III.C.23})$$

Ένα σχετικό κάτω φράγμα βασίζεται σ' ένα μέγεθος γνωστό ως απόκλιση J που ορίζεται από

$$J = \int_{\Gamma} (L - 1) \log(L) p_0 d\mu,$$

και για το οποίο έχουμε το φράγμα [Kobayashi and Thomas (1967)]

$$P_e > \pi_0\pi_1 e^{-J/2}.$$

[Το μέγεθος J συνδέεται στενά με τη σχετική εντροπία ανάμεσα στο p_0 και p_1 που παίζει ένα ρόλο στη θεωρία της πληροφορίας, βλέπε Kullback (1959).] Καθώς τα μεγέθη ρ και J είναι

πιο ευάγωγα απ' ότι είναι το P_e , χρησιμοποιούνται κάποιες φορές ως κριτήρια για την επιλογή σήματος (δηλ. επιλογή των p_0 και p_1), όπως και κάποια άλλα σχετικά κριτήρια.

Η χρησιμότητα του φράγματος Chernoff και άλλων φραγμάτων που περιγράφησαν παραπάνω βασίζεται στο γεγονός ότι μεγέθη όπως τα $E\{L^s\}$ και $E\{(L-1)\log L\}$ είναι ευκολότερο να υπολογιστούν απ' ότι οι πιθανότητες σφάλματος και η εγκυρότητα των φραγμάτων αυτών βασίζεται στο γεγονός ότι οι συναρτήσεις του L που εμφανίζονται στα φράγματα δεν είναι σημειακά μικρότερες από τις δείκτριες συναρτήσεις που αντιστοιχούν στις πιθανότητες σφάλματος. Με άλλα λόγια, θα προτιμούσαμε να ορίσουμε το $E\{h(L)\}$ όπου h είναι η δείκτρια συνάρτηση ενός χώρου απόφασης και αυτό το κάνουμε βρίσκοντας μια συνάρτηση $g \geq h$ για την οποία το $E\{g(L)\}$ είναι εύκολο να υπολογιστεί⁹. Μια συναφής τάξη φραγμάτων πιθανότητας σφάλματος, γνωστών ως φράγματα χώρου – ροής, βασίζεται σε μια παρόμοια αρχή στην οποία προσπαθούμε να βρούμε μια συνάρτηση g τέτοια ώστε το $E\{g(L)\}$ να είναι εύκολο να υπολογιστεί και η καμπύλη που εντοπίζεται από $(h(L), g(L))$, καθώς το L μεταβάλλεται, να είναι σχεδόν μια ευθεία γραμμή. Τα φράγματα αυτά παρουσιάστηκαν από τους Yao και Tobin (1976) και είναι χρήσιμα στην αξιολόγηση των αποτελεσμάτων της επίδοσης διαφόρων τύπων θορύβου που προκύπτουν στα ψηφιακά συστήματα επικοινωνίας, όπως είναι η διασυμβολική παρεμβολή και ο θόρυβος πολλαπλής πρόσβασης

III.C.3 ΑΣΥΜΠΤΩΤΙΚΗ ΣΧΕΤΙΚΗ ΑΠΟΔΟΤΙΚΟΤΗΤΑ

Στις ενότητες III.C.1 και III.C.2, εξετάσαμε τεχνικές οριοθέτησης ή άμεσου υπολογισμού των πιθανοτήτων σφάλματος που συνδέονται με μια διαδικασία ανίχνευσης σήματος. Βλέπουμε ότι, γενικά, οι πιθανότητες σφάλματος μπορούν να είναι μεγέθη δύσκολα να ληφθούν ακριβώς. Υπάρχουν άλλα σημαντικά μέτρα επίδοσης που είναι ευκολότερο να υπολογιστούν από την πιθανότητα σφάλματος. Ένα τέτοιο κριτήριο, η ασυμπτωτική σχετική αποδοτικότητα (ARE) είναι ιδιαίτερα χρήσιμη στη σύγκριση συστημάτων ανίχνευσης διακριτού χρόνου σε συνθήκες μεγάλου μεγέθους δείγματος αδύναμου σήματος¹⁰. Η ARE μπορεί να τεθεί ως ακολούθως. Ας υποθέσουμε ότι έχουμε παρατηρήσεις Y_1, Y_2, \dots , που υπακούουν σε μία από τις δύο στατιστικές υποθέσεις, H_0 και H_1 . Ας υποθέσουμε περαιτέρω ότι $\bar{\delta}_1$ και $\bar{\delta}_2$ είναι δύο έλεγχοι για H_0 ως προς H_1 που έχουν πανομοιότυπες πιθανότητες σφάλματος, αλλά που χρησιμοποιούν δείγματα n_1 και n_2 , αντίστοιχα. Αν $n_1 < n_2$, θα μπορούσαμε να πούμε ότι το $\bar{\delta}_1$ είναι πιο αποδοτικό από το $\bar{\delta}_2$, αφού χρειάζεται λιγότερες πληροφορίες από το $\bar{\delta}_2$ για να πετύχει την ίδια ακριβώς επίδοση.

9

Παρομοίως, αν $n_1 > n_2$, θα μπορούσαμε να πούμε ότι το $\bar{\delta}_1$ είναι λιγότερο αποδοτικό από το $\bar{\delta}_2$. Άρα, ο λόγος n_2/n_1 είναι μια καλή μέτρηση της αποδοτικότητας του $\bar{\delta}_1$ σε σχέση με το

⁹ 9. Μια γενική κλάση άνω και κάτω φραγμάτων αυτού του τύπου έχουν αναπτυχθεί από τους Boeke και Ruitenberg (1981).

¹⁰ 10. Οι συνθήκες μεγάλου δείγματος και αδύναμου σήματος προκύπτουν σε εφαρμογές (π.χ. παθητικά σήματα, ραδιο-αστρονομία) στις οποίες τα σήματα είναι βαθιά ενσωματωμένα στο θόρυβο, απαιτώντας άρα μεγάλο χρόνο ολοκλήρωσης για να ανιχνευθούν. Υπάρχουν αρκετοί ορισμοί των ARE, αλλά ο πιο ευρέως χρησιμοποιούμενος είναι αυτός που παρουσίασε ο E.J.G Pitman [βλ. Noether (1955)] και που είναι ο ορισμός που αναλύεται εδώ.

$\bar{\delta}_2$. Μια τέτοια μέτρηση της σχετικής αποδοτικότητας θα ήταν ιδιαίτερα χρήσιμη για δείγματα μεγάλου μεγέθους (μεγάλα n_1 και n_2) γιατί, παρόλο που και τα δύο συστήματα θα επιτύγχαναν μια καλή επίδοση πιθανότητας σφάλματος με μεγάλο αριθμό δειγμάτων, ο λόγος $n_2/n_1 = 2$, π.χ., θα υποδείκνυε ότι το $\bar{\delta}_1$ είναι ¹⁰ σημαντικά πιο επιθυμητό από το $\bar{\delta}_2$ αν το n_1 είναι ένας μεγάλος αριθμός και αν τα $\bar{\delta}_1$ και $\bar{\delta}_2$ είναι παρόμοιας πολυπλοκότητας. Η ARE του Pitman είναι μια ασυμπτωτική ($n_1, n_2 \rightarrow \infty$) μέτρηση της απόδοσης ενός ανιχνευτή σχετική με κάποια άλλη, υπό την έννοια των σχετικών μεγεθών δείγματος που απαιτούνται για την επίτευξη ισοδύναμης επίδοσης.

Για να προσδιορίσουμε αυτή την έννοια της ARE ακριβέστερα, εξετάζουμε δύο ακολουθίες

$\{\bar{\delta}_1^{(n)}\}_{n=1}^{\infty}$ και $\{\bar{\delta}_2^{(n)}\}_{n=1}^{\infty}$ των ελέγχων του H_0 ως προς το H_1 , όπου το $\bar{\delta}_j^{(n)}$ λειτουργεί με δείγματα n . Υποθέτουμε ότι η πιθανότητα ψευδούς συναγερωμού του κάθε ελέγχου σε κάθε ακολουθία είναι σταθερή στο $\alpha \in (0, 1)$. Η σχετική αποδοτικότητα του $\{\bar{\delta}_1^{(n)}\}_{n=1}^{\infty}$ σε σχέση με το $\{\bar{\delta}_2^{(n)}\}_{n=1}^{\infty}$ για μέγεθος δείγματος n ορίζεται να είναι ο λόγος n_2/n , όπου n_2 είναι ο

μικρότερος αριθμός δειγμάτων, όπως το $P_D(\bar{\delta}_2^{(n_2)}) \geq P_D(\bar{\delta}_1^{(n)}) \triangleq \beta_n$. Τώρα θα θέλαμε να ορίσουμε την ARE ως το όριο της σχετικής αποδοτικότητας, καθώς το n προσεγγίζει το ∞ .

Ωστόσο, για τις περισσότερες λογικές ακολουθίες ελέγχου: $\lim_{n \rightarrow \infty} \beta_n = 1$,¹¹ και άρα για πολύ μεγάλο n , το β_n δεν είναι πλέον κατάλληλο κριτήριο επιλογής του ενός ελέγχου αντί του άλλου¹¹. Για να ξεπεράσουμε τη δυσκολία, εξετάζουμε μια ακολουθία εναλλακτικών υποθέσεων $\{H_1^{(n)}\}_{n=1}^{\infty}$ που συγκλίνουν με κάποιο τρόπο στο H_0 έτσι ώστε

$\lim_{n \rightarrow \infty} \beta_n \triangleq \beta \in (\alpha, 1)$. Υπολογίζουμε κατόπιν τη σχετική αποδοτικότητα, υποθέτοντας ότι και το $\bar{\delta}_1^{(n)}$ και το $\bar{\delta}_2^{(n)}$ είναι έλεγχοι του H_0 ως προς το $H_1^{(n)}$ και ορίζουμε την ασυμπτωτική αποδοτικότητα του $\{\bar{\delta}_1^{(n)}\}_{n=1}^{\infty}$ από

$$ARE_{1,2} = \lim_{n \rightarrow \infty} (n_2/n).$$

Σημειώστε ότι το γεγονός ότι το $H_1^{(n)}$ βρίσκεται «κοντά» στο H_0 (για μεγάλο n) αντιστοιχεί στο πρόβλημα τοπικού ελέγχου που αναλύθηκε στην Ενότητα II.E. (δηλ. στην περίπτωση ενός αδύναμου σήματος σε μοντέλο ανίχνευσης σήματος).

Γενικά, το μέγεθος n_2/n είναι μια συνάρτηση του α και του $\{\beta_n\}_{n=1}^{\infty}$ μεταξύ άλλων. Αποδεικνύεται, ωστόσο, ότι υπό ήπιες παραδοχές η ARE δεν εξαρτάται από τα μεγέθη αυτά. Συγκεκριμένα, υποθέστε ότι οι έλεγχοι tests $\bar{\delta}_j^{(n)}$ είναι της μορφής

¹⁰ 11. Μια ακολουθία ελέγχων με αυτή την ιδιότητα θεωρείται ότι είναι *συνεπής*

$$\bar{\delta}_j^{(n)}(\underline{y}) = \begin{cases} 1 & \text{if } T_j^{(n)}(y_1, \dots, y_n) > \tau_j^{(n)}, \\ \gamma_j & \text{if } T_j^{(n)}(y_1, \dots, y_n) = \tau_j^{(n)}, \\ 0 & \text{if } T_j^{(n)}(y_1, \dots, y_n) < \tau_j^{(n)}, \end{cases}$$

και ότι οι υποθέσεις είναι της μορφής

$$\begin{aligned} H_0 &: Y \sim P_{\theta_0} \\ H_1^{(n)} &: Y \sim P_{\theta_n}, \end{aligned}$$

όπου $\theta_n > \theta_0$ και όπου $\{P_\theta; \theta \geq \theta_0\}$ είναι μια οικογένεια κατανομών για το Y .

Άρα η έννοια ότι το $H_1^{(n)}$ συγκλίνει στο H_0 μπορεί εδώ να αναπαρασταθεί με

$\lim_{n \rightarrow \infty} \theta_n = \theta_0$. Ορίστε για $j = 0, 1$, για $n = 1, 2, \dots$, και για $\theta \geq \theta_0$, τα δύο μεγέθη

$$\psi_j^{(n)}(\theta) = E\{T_j^{(n)}(Y_1, \dots, Y_n) | Y \sim P_\theta\}$$

και

$$\sigma_j^{(n)}(\theta) = [\text{Var}(T_j^{(n)}(Y_1, \dots, Y_n) | Y \sim P_\theta)]^{1/2}$$

δηλ. τα $\psi_j(\theta)$ και $\sigma_j(\theta)$ είναι η μέση και τυπική απόκλιση του στατιστικού ελέγχου $T_j(Y)$ όταν $Y \sim P_\theta$.

Εξετάστε τις ακόλουθες προϋποθέσεις κανονικότητας :

1. Υπάρχει ένας θετικός ακέραιος m τέτοιος ώστε το πρώτο μέχρι το $(m-1)$ ο παράγωγο του $\psi_j^{(n)}(\theta)$ είναι μηδέν στο $\theta = \theta_0$ και

$$\frac{d^m}{d\theta^m} \psi_j^{(n)}(\theta) |_{\theta=\theta_0} > 0 \quad \text{για } j = 0, 1.$$

2. Υπάρχει $\delta > 0$ τέτοιο ώστε, για $j = 0, 1$,

$$\lim_{n \rightarrow \infty} \left[n^{-m\delta} \frac{d^m}{d\theta^m} \psi_j^{(n)}(\theta) |_{\theta=\theta_0} \sigma_j^{(n)}(\theta_0) \right] \stackrel{\Delta}{=} c_j > 0. \quad (\text{III.C.24})$$

3. Ορίστε $\theta_n = \theta_0 + K n^{-\delta}$ για $n = 1, 2, \dots$. Τότε,

$$\lim_{n \rightarrow \infty} \left[\frac{d^m}{d\theta^m} \psi_j^{(n)}(\theta) |_{\theta=\theta_n} / \frac{d^m}{d\theta^m} \psi_j^{(n)}(\theta) |_{\theta=\theta_0} \right] = 1$$

και

$$\lim_{n \rightarrow \infty} \left[\sigma_j^{(n)}(\theta_n) / \sigma_j^{(n)}(\theta_0) \right] = 1.$$

4. Ορίστε

$$W_j^{(n)}(\underline{Y}) = [T_j^{(n)}(Y_1, \dots, Y_n) - \psi_j^{(n)}(\theta)] / \sigma_j^{(n)}(\theta).$$

τότε

$$\lim_{n \rightarrow \infty} P_\theta(W_j^{(n)}(Y) \leq w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-x^2/2} dx \triangleq \Phi(w)$$

για όλα τα $w \in \mathbb{R}$ ομοιόμορφα στο θ για $\theta_0 \leq \theta \leq \theta_0 + d$ για κάποιο $d > 0$. Μπορούμε τώρα να υποστηρίξουμε τα ακόλουθα

Πρόταση III.C.1: Το θεώρημα Pitman – Noether

Ας υποθέσουμε ότι τα $\{\tilde{\delta}_1^{(n)}\}_{n=1}^\infty$ και $\{\tilde{\delta}_2^{(n)}\}_{n=1}^\infty$ καλύπτουν τις προϋποθέσεις 1 ως 4. Τότε για την ακολουθία εναλλακτικών $\theta_n = \theta_0 + Kn^{-\delta}$, έχουμε

$$ARE_{1,2} = \eta_1 / \eta_2,$$

όπου το η_j ορίζεται από

$$\eta_j = (c_j)^{1/m\delta}, \quad j = 0, 1$$

και το c_j ορίζεται από το (III.C.24).

Περίληψη της απόδειξης : Μια πλήρης απόδειξη αυτού του αποτελέσματος απάντα στον Noether (1955). Εδώ έχουμε μια περίληψη (περίγραμμα) της βασικής ιδέας της απόδειξης.

Πρώτα, η προϋπόθεση 4 προϋποθέτει ότι το $T_j^{(n)}(Y_1, \dots, Y_n)$ είναι κατά προσέγγιση $\mathcal{N}(\psi_j(\theta_0), [\sigma_j^{(n)}(\theta_0)]^2)$ ως προς το H_0 . Άρα για $P_F(\delta_j) = \alpha$,

$$\tau_j^{(n)} \sim \sigma_j^{(n)}(\theta_0) \Phi^{-1}(1 - \alpha) + \psi_j^{(n)}(\theta_0).$$

Παρομοίως, η πιθανότητα ανίχνευσης για $\tilde{\delta}_j$ με επίπεδο α είναι (ασυμπτωματικά)

$$\begin{aligned}
P_D(\tilde{\delta}_j) &\sim 1 - \Phi\left(\frac{\tau_j^{(n)} - \psi_j^{(n)}(\theta_n)}{\sigma_j^{(n)}(\theta_n)}\right) \\
&= 1 - \Phi\left(\frac{\sigma_j^{(n)}(\theta_0)}{\sigma_j^{(n)}(\theta_n)}\Phi^{-1}(1 - \alpha) - \frac{\psi_j^{(n)}(\theta_n) - \psi_j^{(n)}(\theta_0)}{\sigma_j^{(n)}(\theta_n)}\right).
\end{aligned}$$

Με την προϋπόθεση 1 έχουμε

$$\psi_j^{(n)}(\theta_0) - \psi_j^{(n)}(\theta_n) \sim \left(\frac{\theta_n - \theta_0}{m!}\right)^m \frac{d^m}{d\theta^m} \psi_j^{(n)}(\theta) |_{\theta=\theta_0}$$

και με την προϋπόθεση 3, $\sigma_j^{(n)}(\theta_0) \sim \sigma_j^{(n)}(\theta_n)$. Άρα, αφού $\theta_n = \theta_0 + Kn^{-\delta}$,

$$P_D(\delta_j) \sim 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{K^m}{m!n^{m\delta}}c_j\right). \quad (\text{III.C.25})$$

Αν εξισώσουμε τα $P_D(\tilde{\delta}_1)$ και $P_D(\tilde{\delta}_2)$, τότε, αφήνοντας τα n_1 και n_2 να συμβολίζουν τα αντίστοιχα μεγέθη δειγμάτων, έχουμε από το (III. C.25)

$$\frac{c_1}{(n_1)^{m\delta}} \sim \frac{c_2}{(n_2)^{m\delta}},$$

ή ισοδύναμα

$$\frac{n_2}{n_1} \sim \left(\frac{c_2}{c_1}\right)^{1/m\delta} = \frac{\eta_2}{\eta_1}, \quad (\text{III.C.26})$$

που είναι το επιθυμητό αποτέλεσμα

Παρατηρήσεις :

1. Το μέγεθος η_j είναι γνωστό ως (οριακή) *αποτελεσματικότητα* της ακολουθίας ελέγχου $\{\tilde{\delta}_j^{(n)}\}_{n=1}^{\infty}$. Άρα το θεώρημα Pitman – Noether επιβεβαιώνει ότι η ακολουθία ελέγχου με την μεγαλύτερη αποτελεσματικότητα είναι η πιο αποδοτική ασυμπτωματικά.
2. Οι προϋποθέσεις κανονικότητας 1 ως 4 ικανοποιούνται εύκολα από πολλά μοντέλα ανίχνευσης σήματος. Για παράδειγμα, εξετάστε την περίπτωση στην οποία οι παρατηρήσεις Y_1, Y_2, \dots , είναι ανεξάρτητες και ταυτόσημα καταναμημένες με οριακή πυκνότητα $f_{\theta}(y)$. Εξετάστε στατιστικές ανίχνευσης της μορφής

$$T_j^{(n)}(y_1, \dots, y_n) = \sum_{k=1}^n g_j(y_k), \quad j = 0, 1. \quad (\text{III.C.27})$$

Τότε έχουμε

$$\psi_j^{(n)}(\theta) = n \int g_j f_\theta d\mu$$

και

$$\sigma_j^{(n)}(\theta) = \left[n \left\{ \int (g_j)^2 f_\theta d\mu - (\psi_j^{(1)}(\theta))^2 \right\} \right]^{1/2}.$$

Όπως θα δούμε παρακάτω, δεν είναι παράλογο να υποθέσουμε ότι $\int g_j f_{\theta_0} d\mu = 0$ [το μέσο του $T_j^{(n)}(Y_1, \dots, Y_n)$ ως προς το H_0 είναι άσχετο σε κάθε περίπτωση καθώς κάθε σταθερά που προστίθεται στο $T_j^{(n)}$ καταλήγει στην ίδια σταθερά που προστίθεται στο $\tau_j^{(n)}$] και $\partial \int g_j f_\theta d\mu / \partial \theta |_{\theta=\theta_0} > 0$, που δίνει τιμές του $m = 1$ και $\delta = \frac{1}{2}$

και

$$\eta_j = \left[\frac{\partial}{\partial \theta} \int g_j f_\theta d\mu \right]_{\theta=\theta_0}^2 / \int g_j^2 f_{\theta_0} d\mu. \quad (\text{III.C.28})$$

Άρα, αν υποθέσουμε ότι οι προϋποθέσεις 3 και 4 ισχύουν, είναι αρκετά απλό να υπολογίσουμε τις ARE των ανιχνευτών της μορφής του (III.C.27). Η προϋπόθεση 3 είναι απλώς μια προϋπόθεση ομαλότητας στην πυκνότητα f_θ ως μια συνάρτηση του θ και στη μη γραμμικότητα g_j , και αν $\int g_j^2 f_\theta d\mu < \infty$, έχουμε πάντα

$$P_\theta(W_j^{(n)}(Y) \leq w) \rightarrow \Phi(w),$$

από το κεντρικό οριακό θεώρημα [βλέπε Breiman (1968)]. Η προϋπόθεση 4 απαιτεί την ομοιομορφία αυτής της σύγκλισης στο $[\theta_0, \theta_0 + d]$. Ωστόσο, έχουμε για i.i.d. δείγματα το ακόλουθο αποτέλεσμα [γνωστό ως το φράγμα Berry – Eseen. Breiman (1968)]

$$\sup_{w \in \mathbf{R}} |P_\theta(W_j^{(n)}(Y) \leq w) - \Phi(w)| \leq \frac{4 \int |g_j|^3 f_\theta d\mu}{\sqrt{n} [\int g_j^2 f_\theta d\mu]^{3/2}} \quad (\text{III.C.29})$$

Έτσι, αν υπάρχουν θετικές σταθερές a , b και d με

$$\int |g_j|^3 f_{\theta} d\mu \leq a \quad \text{and} \quad \int g_j^2 f_{\theta} d\mu \geq b \quad (\text{III.C.30})$$

για όλα τα $\theta \in [\theta_0, \theta_0 + d]$, η προϋπόθεση 4 ικανοποιείται. Ας σημειωθεί ότι η (III.C.30) είναι μια πολύ ήπια προϋπόθεση.

3. Σχετικά με το μοντέλο που αναλύεται στην Παρατήρηση 2, βλέπουμε ότι η

αποδοτικότητα του $\{\tilde{\delta}_j^{(n)}\}_{n=1}^{\infty}$ εξαρτάται από το μέγεθος $\frac{\partial \int g_j f_{\theta} d\mu}{\partial \theta} \Big|_{\theta=\theta_0}$.

Μπορούμε να ξαναγράψουμε το μέγεθος αυτό ως $\frac{\partial \int g_j l_{\theta} f_{\theta_0} d\mu}{\partial \theta} \Big|_{\theta=\theta_0}$, όπου $f_{\theta} / f_{\theta_0}$.

Υποθέτοντας την επαρκή ομαλότητα του l_{θ} , μπορούμε να ανταλλάξουμε τη σειρά της ολοκλήρωσης και του διαφορισμού για να έχουμε

$$\begin{aligned} \frac{\partial}{\partial \theta} \int g_j f_{\theta} d\mu \Big|_{\theta=\theta_0} &= \int g_j \left(\frac{\partial}{\partial \theta} l_{\theta} \right) \Big|_{\theta=\theta_0} f_{\theta_0} d\mu \\ &= \int g_j T_{i_0} f_{\theta_0} d\mu, \end{aligned} \quad (\text{III.C.31})$$

όπου

$$T_{i_0} = \frac{\partial}{\partial \theta} l_{\theta} \Big|_{\theta=\theta_0}. \quad (\text{III.C.32})$$

Η εξίσωση (III.C.33) και η ανισότητα του Schwarz¹² συνεπάγονται ότι το πιο αποδοτικό σύστημα της μορφής του (III.C.27) επιτυγχάνεται με τη χρήση της μη γραμμικότητας

$g_j = c T_{i_0}$ για κάθε θετική σταθερά c . Η τιμή της σταθεράς c δεν έχει σχέση αφού μια αλλαγή στο c έχει ως αποτέλεσμα μόνο μια αλλαγή στο κατώφλι. Ωστόσο, το c θα πρέπει να είναι θετικό, αφού

$$\frac{\partial}{\partial \theta} \int [c T_{i_0}] f_{\theta} d\mu \Big|_{\theta=\theta_0} = c \int (T_{i_0})^2 f_{\theta_0} d\mu > 0.$$

Η μέγιστη δυνατή τιμή του η_j δίνεται με την αντικατάσταση $g_j = T_{i_0}$ στο (III.C.33) για να λάβουμε

$$\max \eta_j = \int (T_{i_0})^2 f_{\theta_0} d\mu, \quad (\text{III.C.34})$$

ένα μέγεθος που θα προκύψει ξανά στο Κεφάλαιο IV στο πλαίσιο της εκτίμησης παραμέτρων

Αξίζει να σημειωθεί ότι ο ανιχνευτής που βασίζεται στη μη γραμμικότητα T_{i_0} είναι και τοπικά βέλτιστος για $\theta = \theta_0$ ως προς $\theta > \theta_0$ με την έννοια που αναλύθηκε στο Κεφάλαιο II. Ξέρουμε ότι για σταθερό n , ο ανιχνευτής λόγου πιθανοφάνειας έχει την καλύτερη επίδοση.

Άρα για κάθε πεπερασμένο n , η αποδοτικότητα κάθε ανιχνευτή που συνδέεται με τον ανιχνευτή λόγου πιθανοφάνειας δεν μπορεί να είναι μεγαλύτερη από τη μονάδα.

Ωστόσο, βλέπουμε από τα παραπάνω ότι ο ανιχνευτής που βασίζεται στο T_{I_0} είναι πιο αποδοτικός ασυμπτωτικά και άρα, υπό αυτή την έννοια, ο ανιχνευτής που βασίζεται στο T_{I_0} είναι ασυμπτωτικά ισοδύναμος με τον ανιχνευτή λόγου πιθανοφάνειας.

11

III.D. Ακολουθιακή Ανίχνευση

Όλες οι διαδικασίες ανίχνευσης που αναλύθηκαν στην Ενότητα III.B. είναι ανιχνευτές παγίου μεγέθους δείγματος. Δηλαδή, σε κάθε περίπτωση είχαμε έναν σταθερό αριθμό παρατηρήσεων και επιθυμούσαμε να παράγουμε το βέλτιστο ανιχνευτή με βάση αυτά τα δείγματα. Μια εναλλακτική προσέγγιση στο πρόβλημα αυτό μπορεί να είναι να καθορίσουμε την επιθυμητή επίδοση και να επιτρέψουμε στον αριθμό των δειγμάτων να μεταβάλλεται προκειμένου να πετύχουμε την επίδοση αυτή. Δηλ. για κάποιες υλοποιήσεις της ακολουθίας παρατήρησης μπορεί να επιθυμούμε να συνεχίσουμε τη δειγματοληψία για να πάρουμε μια καλύτερη απόφαση. Ένας ανιχνευτής που χρησιμοποιεί ένα τυχαίο αριθμό δειγμάτων που εξαρτάται από την ακολουθία παρατήρησης είναι γενικά γνωστό ως *ακολουθιακός ανιχνευτής*. Για να περιγράψουμε τέτοιους ανιχνευτές μπορούμε να χρησιμοποιήσουμε το ακόλουθο μοντέλο.

Ας υποθέσουμε ότι το σύνολο παρατήρησής μας $\Gamma = \mathbb{R}^\infty$, το σύνολο όλων (των μονομερών) πραγματικών ακολουθιών και ότι οι παρατηρήσεις $\{Y_k; k = 1, 2, \dots\}$ είναι ανεξάρτητες και πανομοιότυπα κατανομημένες σύμφωνα με ¹³

$$^{12} H_0 : Y_k \sim P_0, \quad k = 1, 2, \dots$$

ως προς

$$H_1 : Y_k \sim P_1, \quad k = 1, 2, \dots, \quad (\text{III.D.1})$$

όπου P_0 και P_1 είναι δύο πιθανές κατανομές στο $(\mathbb{R}, \mathcal{B})$, με το \mathcal{B} να συμβολίζει την σ -άλγεβρα Borel στο \mathbb{R} . Ένας *ακολουθιακός κανόνας απόφασης* είναι ένα ζεύγος ακολουθιών $(\underline{\varphi}, \underline{\delta})$, όπου το $\underline{\varphi} = \{\varphi_j; j = 0, 1, 2, \dots\}$ καλείται κανόνας τερματισμού ή διακοπής $(\varphi_j : \mathbb{R}^j \rightarrow \{0, 1\})$ και το $\underline{\delta} = \{\delta_j; j = 0, 1, 2, \dots\}$ καλείται τερματικός κανόνας απόφασης, με το δ_j να είναι ένας κανόνας απόφαση στο $(\mathbb{R}^j, \mathcal{B}^j)$ για κάθε $j \geq 0$.

Ο ακολουθιακός κανόνας απόφασης $(\underline{\varphi}, \underline{\delta})$ λειτουργεί ως ακολούθως: Για μια ακολουθία παρατήρησης $\{y_k; k = 1, 2, \dots\}$, ο κανόνας $(\underline{\varphi}, \underline{\delta})$ παίρνει την απόφαση

¹¹12. Ανισότητα του Schwarz: $[\int |fg|d\mu]^2 \leq \int f^2 d\mu \int g^2 d\mu$ με ισότητα αν και μόνον αν $J = cg$ για κάποια σταθερά $c \in \mathbb{R}$.

¹²13. Δηλ., $\mathbb{R}^\infty = \{y | y = \{y_k\}_{k=1}^\infty \text{ with } y_k \in \mathbb{R}, k \geq 1\}$.

$\delta_N(y_1, y_2, \dots, y_N)$, όπου N είναι ο χρόνος τερματισμού που ορίζεται από το $N = \min\{n | \phi_n(y_1, y_2, \dots, y_n) = 1\}$. Αυτό σημαίνει ότι το $\underline{\varphi}$ μας λέει πότε να σταματήσουμε να παίρνουμε δείγματα με τον μηχανισμό ότι όταν $\phi_n(y_1, \dots, y_n) = 0$,

παίρνουμε ένα άλλο δείγμα [το $(n+1)$ st] και όταν $\phi_n(y_1, \dots, y_n) = 1$, σταματάμε τη δειγματοληψία και παίρνουμε μια απόφαση. Με τον τρόπο αυτό, ο αριθμός των δειγμάτων, N , είναι τυχαίος, αφού εξαρτάται από την ακολουθία δεδομένων. Ο τερματικός κανόνας απόφασης $\underline{\delta}$ μας λέει ποια απόφαση να πάρουμε όταν σταματάμε τη δειγματοληψία. Για παράδειγμα, ένας συνηθισμένος σταθερού μεγέθους δείγματος κανόνας απόφασης δ που λειτουργεί με n δείγματα δίνεται από τον ακολουθιακό κανόνα απόφασης $(\underline{\varphi}, \underline{\delta})$ που ορίζεται από

$$\begin{aligned} \phi_j(y_1, \dots, y_j) &= \begin{cases} 0 & \text{if } j \neq n \\ 1 & \text{if } j = n \end{cases} \\ \delta_j(y_1, \dots, y_j) &= \begin{cases} \delta(y_1, \dots, y_n) & \text{if } j = n \\ \text{arbitrary} & \text{if } j \neq n. \end{cases} \end{aligned} \quad (\text{III.D.2})$$

Για να εξάγουμε βέλτιστους ακολουθιακούς κανόνες απόφασης για το (III.D.1) πρώτα εξετάζουμε τη Μπαϊεσιανή εκδοχή του προβλήματος αυτού, στην οποία τα πρότερα π_1 και $\pi_0 = (1 - \pi_1)$ αποδίδονται στις υποθέσεις H_1 και H_0 , αντίστοιχα, και τα κόστη C_{ij} αποδίδονται στις αποφάσεις. Προς χάριν της απλότητας, θα υποθέσουμε ομοιόμορφα κόστη, παρόλο που μπορούμε να χειριστούμε εύκολα και άλλες αποδόσεις κόστους. Αφού θεωρητικά έχουμε ένα άπειρο αριθμό i.i.d. παρατηρήσεων στη διάθεσή μας, για να κάνουμε το πρόβλημα ρεαλιστικό θα πρέπει να αποδώσουμε ένα κόστος στην παρατήρηση. Επομένως θα αποδώσουμε ένα κόστος $C > 0$ σε κάθε δείγμα που λαμβάνουμε, ώστε το κόστος λήψης n δειγμάτων να είναι nC .

Με τις προηγούμενες αποδόσεις κόστους, οι υπό συνθήκη διακινδυνεύσεις για ένα δεδομένο ακολουθιακό κανόνα απόφασης είναι

$$R_0(\underline{\phi}, \underline{\delta}) = E_0\{\delta_N(Y_1, \dots, Y_N)\} + CE_0\{N\}$$

και

$$R_1(\underline{\phi}, \underline{\delta}) = 1 - E_1\{\delta_N(Y_1, \dots, Y_N)\} + CE_1\{N\}, \quad (\text{III.D.3})$$

όπου οι κάτω δείκτες συμβολίζουν την υπόθεση με την οποία υπολογίζεται η προσδοκία και N είναι ο χρόνος τερματισμού που ορίζεται παραπάνω ότι λειτουργεί με την τυχαία ακολουθία $\{Y_k\}$. Η Μπαϊεσιανή διακινδύνευση, επομένως δίνεται από

$$r(\underline{\phi}, \underline{\delta}) = (1 - \pi_1)R_0(\underline{\phi}, \underline{\delta}) + \pi_1R_1(\underline{\phi}, \underline{\delta}), \quad (\text{III.D.4})$$

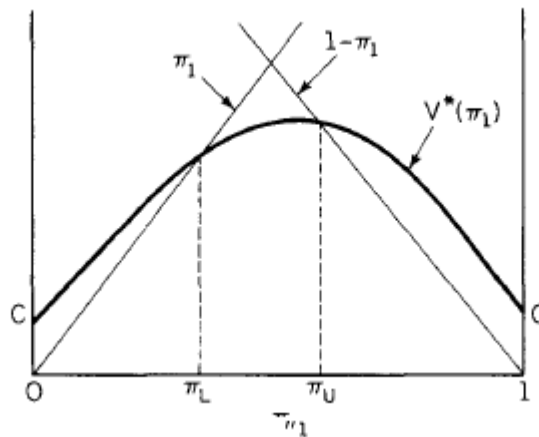
και ένας Μπαϊεσιανός ακολουθιακός κανόνας είναι κάποιος που ελαχιστοποιεί το $r(\underline{\phi}, \underline{\delta})$.

Για να καταλάβουμε τη δομή του βέλτιστου κανόνα απόφασης σ' αυτό το Μπαϊεσιανό πλαίσιο, είναι χρήσιμο να εξετάσουμε τη συνάρτηση

$$V^*(\pi_1) \triangleq \min_{\substack{\underline{\phi}, \underline{\delta} \\ \phi_0=0}} r(\underline{\phi}, \underline{\delta}), \quad 0 \leq \pi_1 \leq 1. \quad (\text{III.D.5})$$

Αφού $\phi_0 = 0$ σημαίνει ότι ο έλεγχος δεν σταματά με μηδενικές παρατηρήσεις, το (III.D.5) περιγράφει την ελάχιστη Μπαϊεσιανή διακινδύνευση σε όλους τους ακολουθιακούς ελέγχους που παίρνουν τουλάχιστον ένα δείγμα. Αποδεικνύεται ευθέως ότι η συνάρτηση

$V^*(\pi_1)$ είναι μια κοίλα, συνεχής συνάρτηση του π_1 με $V^*(0) = V^*(1) = C$, όπως απεικονίζεται στο Σχ. III.D.1 [βλέπε Ferguson (1967)]. Στο Σχ. III.D.1 φαίνονται επίσης διαγράμματα της Μπαϊεσιανής διακινδύνευσης ως προς το π_1 για δύο άλλους ακολουθιακούς κανόνες απόφασης. Αυτόν που δεν παίρνει δείγματα και αποφασίζει H_1 (δηλ. $\phi_0 = \delta_0 = 1$) και αυτόν που δεν παίρνει δείγματα και αποφασίζει H_0 (δηλ. $\phi_0 = 1 - \delta_0 = 1$).



ΣΧΗΜΑ III.D.1 Σχέσεις που αποδίδουν τον Μπαϊεσιανό ακολουθιακό νόμο για ομοιόμορφα κόστη σφαλμάτων και κόστος C ανά δείγμα

Σημειώστε ότι

$$r(\underline{\phi}, \underline{\delta}) |_{\phi_0=\delta_0=1} = (1 - \pi_1) \text{ and } r(\underline{\phi}, \underline{\delta}) |_{\phi_0=1-\delta_0=1} = \pi_1 .$$

Οι τελευταίοι αυτοί δύο κανόνες απόφασης αποτελούν τους μόνους πιθανούς Μπαϊεσιανούς κανόνες που δεν περιλαμβάνονται στην ελαχιστοποίηση του (III.D.5)¹⁴ Ας σημειωθεί ότι στο Σχ. III.D.1 συμβολίσαμε με π_U την τετμημένη της τομής του $r(\underline{\phi}, \underline{\delta}) |_{\phi_0=1-\delta_0=1}$ και του $V^*(\pi_1)$.

Μελετώντας το Σχ. III.D.1, βλέπουμε ότι ο Μπαϊεσιανός κανόνας για ένα σταθερό πρότερο π_1 είναι $\phi_0 = 1 - \delta_0 = 1$ αν $\pi_1 \leq \pi_L$, είναι $\phi_0 = \delta_0 = 1$ αν $\pi_1 \geq \pi_U$, και είναι ο κανόνας απόφασης με ελάχιστη Μπαϊεσιανή διακινδύνευση μεταξύ όλων των (φ, δ) με $\phi_0 = 0$ αν $\pi_L < \pi_1 < \pi_U$. Άρα αν $\pi_1 \leq \pi_L$, δεν παίρνουμε δείγματα και επιλέγουμε H_0 , αν $\pi_1 \geq \pi_U$ δεν παίρνουμε δείγματα και επιλέγουμε H_1 και σε άλλη περίπτωση παίρνουμε τουλάχιστον ένα δείγμα.

Τώρα υποθέστε ότι έχουμε τη συνθήκη $\pi_L < \pi_1 < \pi_U$. Εδώ γνωρίζουμε ότι ο βέλτιστος έλεγχος λαμβάνει τουλάχιστον ένα δείγμα, αλλά ο έλεγχος δεν μπορεί να προσδιοριστεί με άλλο τρόπο. Ωστόσο, ας σημειωθεί ότι αφού έχουμε λάβει ένα δείγμα, το πρόβλημα βελτιστοποίησης του ελέγχου είναι υπό συνθήκη το ίδιο με αυτό χωρίς δείγματα, υπό την έννοια ότι και πάλι έχουμε απείρως πολλά i.i.d. δείγματα στη διάθεσή μας και τα κόστη είναι τα ίδια. Η μία διαφορά είναι ότι τώρα έχουμε πάρει ένα δείγμα και έτσι έχουμε περισσότερες πληροφορίες για το ποια υπόθεση είναι αληθής. Συγκεκριμένα, αντί να έχουμε μια πρότερη πιθανότητα π_1 , τώρα έχουμε πρότερο $\pi_1(y_1)$ που είναι στην πραγματικότητα η ύστερη πιθανότητα του H_1 , με δεδομένη την παρατήρησή μας στο Y_1 . Δηλ., $\pi_1(y_1) = P(H_1 \text{ είναι αληθές} | Y_1 = y_1)$. Άρα η εικόνα μετά τη λήψη ενός δείγματος είναι ακριβώς ίδια με αυτή του Σχ. III.D.1 εκτός από το ότι η τετμημένη μεταβλητή π_1 αντικαθίσταται από το $\pi_1(y_1)$. Καθώς τα δείγματα είναι ανεξάρτητα, η γνώση του Y_1 δεν επηρεάζει το σχήμα το V^* (που τώρα συμβολίζει την ελάχιστη διακινδύνευση σε όλους τους ελέγχους που λαμβάνουν τουλάχιστον δύο δείγματα). Άρα συμπεραίνουμε ότι αφού λάβουμε ένα δείγμα, ο βέλτιστος έλεγχος σταματά και επιλέγει H_0 , αν $\pi_1(y_1) \leq \pi_L$, σταματά και επιλέγει H_1 , αν $\pi_1(y_1) \geq \pi_U$ και λαμβάνει άλλο δείγμα, αν $\pi_L < \pi_1(y_1) < \pi_U$.

Αν όσο το $\pi_L < \pi_1 < \pi_U$ όσο και το $\pi_L < \pi_1(y_1) < \pi_U$, τότε από τα παραπάνω βλέπουμε ότι ο βέλτιστος έλεγχος λαμβάνει τουλάχιστον δύο δείγματα. Στην περίπτωση αυτή, ξεκινάμε εκ νέου με το νέο πρότερο $\pi_1(y_1, y_2) = P(H_1 \text{ είναι αληθές} | Y_1 = y_1, Y_2 = y_2)$ και κάνουμε την ίδια σύγκριση ξανά. Συνεχίζοντας με αυτή τη λογική για ένα αυθαίρετο αριθμό ληφθέντων δειγμάτων, βλέπουμε ότι ο Μπαϊεσιανός ακολουθιακός έλεγχος συνεχίζει τη δειγματοληψία μέχρι το μέγεθος $\pi_1(y_1, \dots, y_n) \stackrel{\Delta}{=} P(H_1 \text{ είναι αληθές} | Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n)$ να μην εμπίπτει στο διάστημα (π_L, π_U) και τότε επιλέγει H_0 αν $\pi_1(y_1, \dots, y_n) \leq \pi_L$ και H_1 αν $\pi_1(y_1, \dots, y_n) \geq \pi_U$. [Για $n = 0$, $\pi_1(y_1, \dots, y_n) \leq \pi_L$ και H_1 αν $\pi_1(y_1, \dots, y_n) \geq \pi_U$. [Για $n = 0$, το $\pi_1(y_1, \dots, y_n)$ συμβολίζει π_1]. Ο έλεγχος αυτός περιγράφεται από τον κανόνα τερματισμού

13

14. Βλέπουμε ευθέως ότι αυτή η τυχαιοποίηση δεν μπορεί να βοηθήσει με το $\phi_0 = 1$, αφού αν επιλέξουμε $\delta_0 = 1$ με πιθανότητα γ και $\delta_0 = 0$ με πιθανότητα $(1 - \gamma)$ έχουμε ένα Μπαϊεσιανό κίνδυνο $\gamma(1 - \pi_1) + (1 - \gamma)\pi_1$, που είναι πάντα μεγαλύτερος από το $\min\{\pi_1, 1 - \pi_1\}$.

$$\phi_n(y_1, \dots, y_n) = \begin{cases} 0 \\ 1 \end{cases} \text{ αν } \pi_L < \pi_1(y_1, \dots, y_n) < \pi_U$$

και ο κανόνας τερματισμού απόφασης

$$\delta_n(y_1, \dots, y_n) = \begin{cases} 1 & \text{if } \pi_1(y_1, \dots, y_n) \geq \pi_U \\ 0 & \text{if } \pi_1(y_1, \dots, y_n) \leq \pi_L. \end{cases} \quad (\text{III.D.7})$$

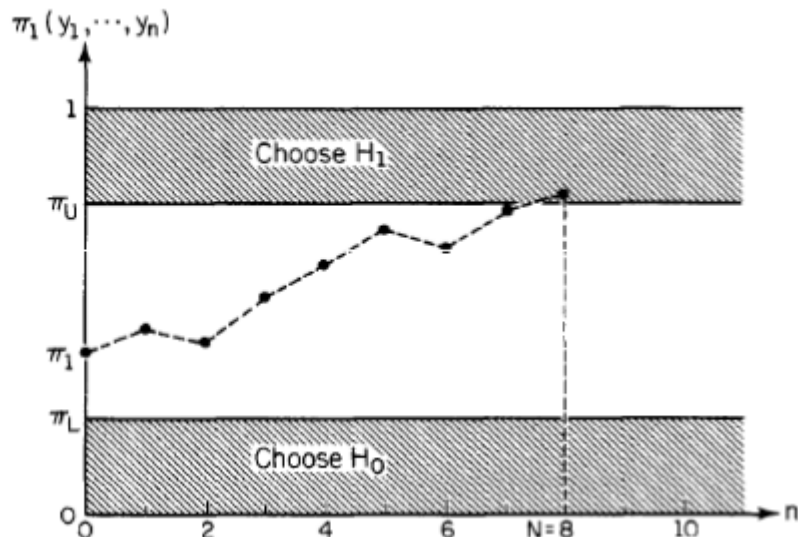
Ο Μπαϊεσιανός έλεγχος απεικονίζεται στο Σχ. III.D.2 Υπό ομαλές συνθήκες, το

$\pi_1(Y_1, \dots, Y_n)$ συγκλίνει σχεδόν βέβαια στο 1 ως προς το H_1 και στο 0 ως προς το H_0 . Επομένως, ο έλεγχος τερματίζει με πιθανότητα 1. Όλα όσα είναι απαραίτητα για τον προσδιορισμό του βέλτιστου ελέγχου είναι οι δύο πιθανότητες π_L και π_U κι ένα σύστημα υπολογισμού $\pi_1(y_1, \dots, y_n)$. Δυστυχώς, είναι δύσκολο να λάβουμε τις π_L και π_U ακριβώς, εκτός από ορισμένες ειδικές περιπτώσεις (π.χ. όταν τα Y_n λαμβάνουν μόνο διακριτές τιμές). Από την άλλη, ο υπολογισμός της ύστερης πιθανότητας $\pi_1(y_1, \dots, y_n)$ είναι αρκετά εύκολος. Συγκεκριμένα, υποθέτοντας ότι τα P_0 και P_1 έχουν πυκνότητες p_0 και p_1 , ο Μπαϊεσιανός τύπος συνεπάγεται ότι το $\pi_1(Y_1, \dots, Y_n)$ μπορεί να γραφεί ως

$$\begin{aligned} \pi_1(y_1, \dots, y_n) &= \frac{\pi_1 \prod_{k=1}^n p_1(y_k)}{\pi_0 \prod_{k=1}^n p_0(y_k) + \pi_1 \prod_{k=1}^n p_1(y_k)} \\ &= \frac{\pi_1 \lambda_n(y_1, \dots, y_n)}{\pi_0 + \pi_1 \lambda_n(y_1, \dots, y_n)}, \end{aligned} \quad (\text{III.D.8})$$

όπου λ_n είναι ο λόγος πιθανοφάνειας βασισμένος σε n δείγματα που δίνεται από

$$\lambda_n(y_1, \dots, y_n) = \prod_{k=1}^n [p_1(y_k)/p_0(y_k)]. \quad (\text{III.D.9})$$



ΣΧΗΜΑ III.D.2. Απεικόνιση υλοποίησης ενός Μπαϊεσιανού ακολουθιακού ελέγχου

(Για λόγους συμβατότητας ορίζουμε $\lambda_0 = 1$.) Καθώς το (III.D.8) είναι μονότονα αυξανόμενο στο λ_n , ο έλεγχος του (III.D.6) και του (III.D.7) μπορεί να ξαναγραφεί ως

$$\phi_n(y_1, \dots, y_n) = \begin{cases} 0 & \text{if } \underline{\pi} < \lambda_n(y_1, \dots, y_n) < \bar{\pi} \\ 1 & \text{otherwise} \end{cases} \quad (\text{III.D.10})$$

και

$$\delta_n(y_1, \dots, y_n) = \begin{cases} 1 & \text{if } \lambda_n(y_1, \dots, y_n) \geq \bar{\pi} \\ 0 & \text{if } \lambda_n(y_1, \dots, y_n) \leq \underline{\pi}, \end{cases} \quad (\text{III.D.11})$$

όπου $\underline{\pi} \triangleq \pi_0 \pi_L / \pi_1 (1 - \pi_L)$ και $\bar{\pi} \triangleq \pi_0 \pi_U / \pi_1 (1 - \pi_U)$. Άρα ο Μπαϊεσιανός ακολουθιακός έλεγχος λαμβάνει δείγματα μέχρι ο λόγος πιθανοφάνειας να βρίσκεται εκτός του διαστήματος $(\underline{\pi}, \bar{\pi})$ και κατόπιν αποφασίζει για το H_0 ή το H_1 , ανάλογα από το αν το λ_n εμπίπτει κάτω από το $\underline{\pi}$ ή πάνω από το $\bar{\pi}$.

Ο έλεγχος του (III.D.10) και (III.D.11) είναι ένα παράδειγμα ακολουθιακού ελέγχου λόγου πιθανοφάνειας (SPRT). Συγκεκριμένα, για όλους πραγματικούς αριθμούς A και B που ικανοποιούν το $0 < A \leq 1 \leq B < \infty$, ο SPRT με *σύνορα* A και B [που συμβολίζονται στο SPRT (A, B)] ορίζεται όπως στο (III.D.10) και το (III.D.11) με το $\underline{\pi}$ να αντικαθίσταται από το A και το $\bar{\pi}$ να αντικαθίσταται από το B , και με τον κανόνα απόφασης να μένει αυθαίρετος αν $A = B$. Άρα ο SPRT (A, B) συνεχίζει τη δειγματοληψία μέχρι ο λόγος πιθανοφάνειας λ_n να βρεθεί εκτός των «συνόρων» A και B και κατόπιν επιλέγει H_1 αν $\lambda_n \geq B$ και H_0 αν $\lambda_n \leq A$. Ας σημειωθεί ότι αν $A = 1 < B$, δεν παίρνουμε δείγματα και επιλέγουμε H_0 . Και αν $A = B = 1$, δεν παίρνουμε δείγματα και κάνουμε μια αυθαίρετη επιλογή.

Παράδειγμα III.D.1: Ακολουθιακή Ανίχνευση ενός Διαρκούς Σήματος

Εξετάστε το πρόβλημα ανίχνευσης ενός διαρκούς σήματος σε προσθετικό i.i.d. θόρυβο.

$$\begin{aligned} H_0 : Y_k &= N_k, \quad k = 1, 2, \dots, \\ H_1 : Y_k &= N_k + \theta, \quad k = 1, 2, \dots, \end{aligned} \quad (\text{III.D.12})$$

όπου $\theta > 0$ και $\{N_k\}_{k=1}^{\infty}$ είναι μια i.i.d. ακολουθία των $\mathcal{N}(0, \sigma^2)$ δειγμάτων θορύβου. Για το πρόβλημα αυτό, ο λόγος πιθανοφάνειας που βασίζεται σε n δείγματα δίνεται από

$$\lambda_n(y_1, \dots, y_n) = \exp \left\{ \frac{\theta}{\sigma^2} \sum_{k=1}^n (y_k - \theta/2) \right\}, \quad (\text{III.D.13})$$

οπότε ο SPRT (A, B) συνεχίζει τη δειγματοληψία για όσο το μέγεθος αυτό βρίσκεται μεταξύ των A και B . Ισοδύναμα, με τη λήψη λογαρίθμων, βλέπουμε ότι ο SPRT (A, B) υπολογίζει το $(\theta/\sigma^2) \sum_{k=1}^n (y_k - \theta/2)$ σε κάθε στάδιο και το συγκρίνει με το λογάριθμο A και το

λογάριθμο B , σταματώντας και επιλέγοντας H_1 όταν το $(\tilde{\theta}/\sigma^2) \sum_{k=1}^n (y_k - \theta/2)$ υπερβαίνει το λογάριθμο B και σταματώντας και επιλέγοντας H_0 όταν το $(\theta/\sigma^2) \sum_{k=1}^n (y_k - \theta/2)$ πέφτει κάτω από το λογάριθμο A .

Εκτός της βελτιστότητας του SPRT $(\underline{\pi}, \bar{\pi})$ στο Μπαϊεσιανό πρόβλημα, ο SPRT (A, B) έχει μια ακόμη ιδιότητα βελτιστότητας η οποία, από την άποψη της ανίχνευσης σήματος, είναι ενδεχομένως πιο σημαντική. Η ιδιότητα αυτή συνοψίζεται στο θεώρημα Wald – Wolfowitz (που αναφέρεται παρακάτω), το οποίο είναι ανάλογο με το λήμμα Neyman – Pearson.

Για ένα ακολουθιακό κανόνα απόφασης $(\underline{\phi}, \underline{\delta})$ έστω ότι $P_F(\underline{\phi}, \underline{\delta})$ συμβολίζει την πιθανότητα ψευδούς συναγερμού και έστω ότι $P_M(\underline{\phi}, \underline{\delta})$ συμβολίζει την πιθανότητα μιας αστοχίας.

Δηλ.,

$$P_F(\underline{\phi}, \underline{\delta}) = P(\delta_N(Y_1, \dots, Y_N) = 1 | H_0)$$

και

$$P_M(\underline{\phi}, \underline{\delta}) = P(\delta_N(Y_1, \dots, Y_N) = 0 | H_1).$$

Επίσης, έστω ότι $N(\underline{\phi})$ συμβολίζει τον τυχαίο χρόνο τερματισμού που συνδέεται με το $\underline{\phi}$. Δηλ.,

$$N(\underline{\phi}) = \min\{n | \phi_n(Y_1, \dots, Y_n) = 1\}.$$

Το $N(\underline{\phi})$ είναι επίσης γνωστό ως ο αριθμός δείγματος του $\underline{\phi}$. Τότε έχουμε το ακόλουθο αποτέλεσμα

Πρόταση III.D.1: Το Θεώρημα Wald – Wolfowitz

Ας υποθέσουμε ότι $(\underline{\phi}_0, \underline{\delta}_0)$ είναι ο SPRT (A, B) και ότι $(\underline{\phi}, \underline{\delta})$ είναι κάθε άλλος ακολουθιακός κανόνας απόφασης για τον οποίο

$$P_F(\underline{\phi}, \underline{\delta}) \leq P_F(\underline{\phi}_0, \underline{\delta}_0)$$

και

$$P_M(\underline{\phi}, \underline{\delta}) \leq P_M(\underline{\phi}_0, \underline{\delta}_0).$$

Τότε

$$P_M(\underline{\phi}, \underline{\delta}) \leq P_M(\underline{\phi}_0, \underline{\delta}_0).$$

Αρα βλέπουμε ότι για ένα δεδομένο επίπεδο επίδοσης, κανένας ακολουθιακός κανόνας απόφασης δεν έχει μικρότερο αναμενόμενο μέγεθος δείγματος απ' ό,τι ο SPRT με αυτή την επίδοση. Ας σημειωθεί επίσης ότι καθώς ένας ανιχνευτής σταθερού μεγέθους δείγματος μπορεί να χαρακτηριστεί ως ένας συγκεκριμένος τύπος ακολουθιακού κανόνας απόφασης, αυτό το θεώρημα συνεπάγεται ότι το μέσο μέγεθος δείγματος ενός SPRT δεν είναι μεγαλύτερο από το μέγεθος δείγματος ενός ελέγχου σταθερού μεγέθους δείγματος με την ίδια επίδοση. Είναι επίσης αληθές ότι για δεδομένα αναμενόμενα μεγέθη δείγματος, κανένας ακολουθιακός κανόνας απόφασης δεν έχει μικρότερες πιθανότητες σφάλματος από τον SPRT [βλ. π.χ., Ferguson (1967)]. Η ισχύς του θεωρήματος Wald – Wolfowitz ουσιαστικά απορρέει από τη Μπαϊσιανή βελτιστότητα του SPRT. Εδώ η απόδειξη παραλείπεται.

Το θεώρημα Wald – Wolfowitz μας λέει ότι για δεδομένα P_F και P_M , ο SPRT είναι ο ανιχνευτής με το μικρότερο μέσο μέγεθος δείγματος. Αποδεικνύεται ότι ο SPRT (A, B) μας επιτρέπει να επιλέξουμε τα P_F και P_M αυθαίρετα εις βάρος μεγαλύτερων αναμενόμενων μεγεθών δείγματος, όπως θα δούμε παρακάτω. Το επόμενο ερώτημα που προκύπτει: Πώς επιλέγει κάποιος το A και B για να επιτύχει ένα επιθυμητό επίπεδο επίδοσης; Η απάντηση στο ερώτημα αυτό δίνεται εν μέρει από την παρακάτω ανάλυση.

Ας υποθέσουμε ότι το $(\underline{\phi}, \underline{\delta})$ είναι ο SPRT (A, B) με $A < 1 < B$, και έστω ότι $\alpha = P_F(\underline{\phi}, \underline{\delta})$, $\gamma = (1 - \beta) = P_M(\underline{\phi}, \underline{\delta})$ και $N = N(\underline{\phi})$. Σημειώστε ότι η περιοχή απόρριψης του $(\underline{\phi}, \underline{\delta})$ μπορεί να γραφεί ως

$$\Gamma_1 = \{y \in \mathbb{R}^\infty | \lambda_N(y_1, \dots, y_N) \geq B\} = \cup_{n=1}^{\infty} Q_n,$$

όπου $Q_n = \{y \in \mathbb{R}^\infty | N = n \text{ και } \lambda_n(y_1, y_2, \dots, y_n) \geq B\}$

Αφού Q_n και Q_m είναι ξένα (ή αμοιβαίως αποκλειόμενα) σύνολα για $m \neq n$, μπορούμε να γράψουμε

$$\alpha = P(\lambda_N(Y_1, \dots, Y_N) \geq B | H_0) = \sum_{n=1}^{\infty} \int_{Q_n} \prod_{k=1}^n [p_0(y_k) \mu(dy_k)].$$

Στο Q_n έχουμε $\prod_{k=1}^n p_0(y_k) \leq B^{-1} \prod_{k=1}^n p_1(y_k)$, ώστε

$$\begin{aligned} \alpha &\leq B^{-1} \sum_{k=1}^{\infty} \int_{Q_n} \prod_{k=1}^n [p_1(y_k) \mu(dy_k)] \\ &= B^{-1} P(\lambda_N(Y_1, \dots, Y_N) \geq B | H_1) \\ &= B^{-1} (1 - \gamma). \end{aligned}$$

Παρομοίως,

$$\begin{aligned}
\gamma &= P(\lambda_N(Y_1, \dots, Y_N) \leq A | H_1) \\
&\leq AP(\lambda_N(Y_1, \dots, Y_N) \leq A | H_0) \\
&= A(1 - \alpha).
\end{aligned}$$

Άρα έχουμε

$$B \leq (1 - \gamma)/\alpha \quad \text{and} \quad A \geq \gamma/(1 - \alpha). \quad (\text{III.D.14})$$

Μπορούμε να χρησιμοποιήσουμε τις ανισότητες του (III.D.14) για να λάβουμε προσεγγιστικές τιμές ώστε τα σύνορα A και B να δώσουν τα επιθυμητά α και γ , υποθέτοντας ότι όταν ο λόγος πιθανοφάνειας λ_n ξεπερνά ένα σύνορο, η υπέρβαση του συνόρου [δηλ., $(\lambda_N(Y_1, \dots, Y_N) - B)$ ή $(A - \lambda_N(Y_1, \dots, Y_N))$] είναι αμελητέα. Η προσέγγιση αυτή θα είναι ακριβής, αν το N είναι σχετικά μεγάλο στο μέσο. Άρα, υποθέτουμε ότι είτε $\lambda_N(Y_1, \dots, Y_N) \cong A$ είτε $\lambda_N(Y_1, \dots, Y_N) \cong B$ και η ανισότητα του (III.D.14) γίνονται προσεγγιστικές ισότητες. Δηλ.,

$$B \cong (1 - \gamma)/\alpha \quad \text{and} \quad A \cong \gamma/(1 - \alpha). \quad (\text{III.D.15})$$

Οι προσεγγίσεις αυτές είναι γνωστές ως *προσεγγίσεις Wald*

Έστω ότι α_d και γ_d είναι επιθυμητές πιθανότητες σφάλματος και έστω ότι χρησιμοποιούμε την προσέγγιση του (III.D.15) για να επιλέξουμε τα πραγματικά σύνορα, δηλ.,

$$A_a = \gamma_d/(1 - \alpha_d) \quad \text{and} \quad B_a = (1 - \gamma_d)/\alpha_d. \quad (\text{III.D.16})$$

Τότε οι πραγματικές πιθανότητες σφάλματος α_a και γ_a θα ικανοποιούν τις ανισότητες του (III.D.13), έτσι ώστε

$$\alpha_a/(1 - \gamma_a) \leq B_a^{-1} = \alpha_d/(1 - \gamma_d)$$

Και

$$\gamma_a/(1 - \alpha_a) \leq A_a = \gamma_d/(1 - \alpha_d),$$

από το οποίο έχουμε

$$\alpha_a \leq \alpha_d(1 - \gamma_a)/(1 - \gamma_d) \leq \alpha_d/(1 - \gamma_d)$$

και

$$\gamma_a \leq \gamma_d(1 - \alpha_a)/(1 - \alpha_d) \leq \gamma_d/(1 - \alpha_d). \quad (\text{III.D.17})$$

Έτσι, για παράδειγμα, αν $\gamma_d = \alpha_d$ έχουμε

$$\alpha_a \leq \alpha_d + O(\alpha_d^2)$$

και

$$\gamma_a \leq \gamma_d + O(\gamma_d^2).$$

Ας σημειωθεί ότι το (III.D.17) εγγυάται ότι μπορούμε να λάβουμε αυθαίρετα καλή επίδοση πιθανότητας σφάλματος, με την ορθή επιλογή ορίων A και B . Σημειώστε επίσης ότι οι ανισότητες (III.D.17) είναι ακριβείς και δεν εξαρτώνται (εκτός από το κίνητρο για την επιλογή A_a και B_a από το (III.D.15)] από την προσέγγιση που βασίζεται στην άγνοια της υπέρβασης των συνόρων. Ενδεχομένως το πιο εκπληκτικό απ' όλα τα παραπάνω είναι ότι η πραγματική κατανομή του Y_k δεν σχετίζεται με την ανάλυση. Το γεγονός αυτό αποτελεί σημαντικό πλεονέκτημα της χρήσης ενός SPRT αντί ενός ελέγχου σταθερού μεγέθους δείγματος.

Παρόλο που μπορούμε αυθαίρετα να μειώσουμε τις πιθανότητες σφάλματος μ' έναν SPRT, αυτό γίνεται εις βάρος της αύξησης του αναμενόμενου μεγέθους δείγματος. Για να το κατανοήσουμε αυτό, ας υποθέσουμε ότι a και γ είναι μικρά ώστε $A \cong \gamma$ και $B \cong 1/\alpha$. Τότε, καθώς το γ και το a μειώνονται, το A μικραίνει και το B μεγαλώνει. Αυτό διευρύνει το διάστημα (A, B) , οπότε για μια δεδομένη υλοποίηση, ο SPRT (A, B) πρέπει να λάβει περισσότερα δείγματα. Η ανάλυση των προηγούμενων παραγράφων υποδεικνύει ότι οι πιθανότητες σφάλματος του SPRT μπορούν να ελεγχθούν χωρίς να ληφθεί υπόψη η κατανομή των παρατηρήσεων. Ωστόσο, το αναμενόμενο μέγεθος εξαρτάται από την κατανομή των δεδομένων και χρησιμοποιείται ο βασικός δείκτης απόδοσης για τη σύγκριση διάφορων ακολουθιακών ανιχνευτών. Το επόμενο ερώτημα που προκύπτει τότε είναι: Πώς μπορεί κάποιος να αξιολογήσει το αναμενόμενο μέγεθος δείγματος ενός ακολουθιακού ανιχνευτή; Για να απαντήσουμε στο ερώτημα αυτό θα πρέπει να εξετάσουμε πρώτα έναν ελαφρώς πιο γενικό ακολουθιακό έλεγχο για H_0 ως προς H_1 , που ορίζεται ως ακολούθως

Για κάθε $a < 0 < b$ και κάθε συνάρτηση $g : \mathbb{R} \rightarrow \mathbb{R}$, ορίστε τον ακολουθιακό κανόνα απόφασης $ST(a, b; g)$ από το ζεύγος (ϕ, δ) που δίνεται από

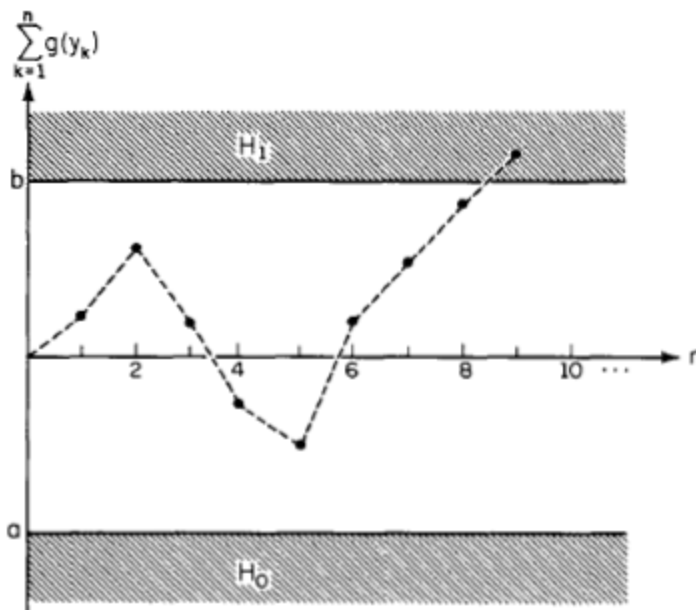
$$\phi_j(y_1, \dots, y_j) = \begin{cases} 0 \\ 1 \end{cases} \text{ αν } a < \sum_{k=1}^j g(y_k) < b$$

Και

$$\delta_j(y_1, \dots, y_j) = \begin{cases} 1 \\ 0 \end{cases} \text{ αν } \sum_{k=1}^j g(y_k) \geq b \\ \text{αν } \sum_{k=1}^j g(y_k) \leq a.$$

Ας σημειωθεί ότι για $0 < A < 1 < B < \infty$ και $p_1/p_0 < \infty$, ο SPRT (A, B) είναι $ST(a, b; g)$ με $a = \log A, b = \log B$, και $g = \log(p_1/p_0)$. Για την περίπτωση του παραδείγματος III.D.1, ο SPRT (A, B) είναι $ST(a, b; \bar{g})$ με $g(x) = \theta(x - \theta/2)/\sigma^2$. Ο έλεγχος $ST(a, b, g)$ απεικονίζεται στο Σχ. III.D.3.

Στο πλαίσιο του $ST(a, b; g)$ έχουμε το ακόλουθο αποτέλεσμα που οφείλεται στο Wald (βλέπε Ferguson 1967).



ΣΧΗΜΑ III.D.1 Απεικόνιση του $ST(a, b; g)$.

Πρόταση III.D.2.: Η θεμελιώδης ταυτότητα της ακολουθιακής ανάλυσης

Έστω $(\phi, \bar{g}) = ST(a, b; g)$. Ορίστε $N = \min\{n | \phi_n(Y_1, \dots, Y_n) = 1\}$ και $S_n = \sum_{k=1}^n \bar{g}(Y_k)$ και έστω ότι M_j συμβολίζει τη στιγμιαία συνάρτηση παραγωγής της τυχαίας μεταβλητής $g(Y_1)$ ως προς την υπόθεση H_j . Δηλ.,

$$M_j(t) = E\{\exp\{tg(Y_1)\} | H_j\}, \quad j = 0, 1.$$

Ας υποθέσουμε ότι $j = 0$ ή 1 . Τότε, αν $P(g(Y_1) = 0 | H_j) \neq 1$ και $P(|g(Y_1)| < \infty | H_j) = 1$, έχουμε

$$E\{\exp\{tS_N\}[M_j(t)]^{-N}|H_j\} = 1$$

για όλα τα αληθή t για τα οποία $M_j(t) < \infty$.

Το αποτέλεσμα αυτό είναι επίσης γνωστό ως *ταυτότητα του Wald*. Είναι άμεση συνέπεια του θεωρήματος βέλτιστης δειγματοληψίας για τα μαρτινγκέλ [Breiman (1968)] και δεν θα αποδειχθεί εδώ. Ένα αποτέλεσμα που έπεται από την Πρόταση III.D.2 είναι το ακόλουθο:

Πρόταση III.D.3: Συνακόλουθο της ταυτότητας του Wald.

Ως προς την Πρόταση III.D.2, υποθέστε ότι $M_j(t) < \infty$ σε μια περιοχή του $t = 0$. Ορίστε $\mu_j = E\{g(Y_1)|H_j\}$ και $\sigma_j^2 = \text{Var}\{g(Y_1)|H_j\}$.

Τότε

$$(a) E\{S_N|H_j\} = \mu_j E\{N|H_j\}$$

και

$$(b) E\{(S_N - N\mu_j)^2|H_j\} = \sigma_j^2 E\{N|H_j\}.$$

Χρησιμοποιώντας την ταυτότητα του Wald και το συνακόλουθό της μπορούμε να επεκτείνουμε τις προσεγγίσεις Wald (III.D.15) μέχρι τον έλεγχο $ST(a, b; g)$, ως ακολούθως

Ας υποθέσουμε ότι μπορούμε να βρούμε δύο μη μηδενικούς αριθμούς t_0 και t_1 τέτοιους ώστε

$M_j(t_j) = 1$ για $j = 0$ και 1 (η ύπαρξη ενός τέτοιου t_j εξασφαλίζεται αν $P_j(g(Y_1) < 0) > 0, P_j(g(Y_1) > 0) > 0$, και $\mu_j \neq 0$ [Βλ. Ferguson (1967)]). Τότε η ταυτότητα του Wald συνεπάγεται ότι

$$E\{\exp\{t_j S_N\}|H_j\} = 1 \text{ for } j = 0, 1. \quad (\text{III.D.18})$$

Αν αγνοήσουμε την «υπέρβαση των ορίων», τότε ως προς το H_0 , η S_N είναι μια διακριτή τυχαία μεταβλητή που λαμβάνει τιμές b και a με πιθανότητες $P_F(\underline{\phi}, \underline{\delta}) \triangleq \alpha$ και $(1 - \alpha)$ αντίστοιχα, και ως προς το H_1 , η S_N λαμβάνει τιμές a και b με πιθανότητες $\gamma = P_M(\underline{\phi}, \underline{\delta})$ και $(1 - \gamma)$, αντίστοιχα. Άρα, το (III.D.18) συνεπάγεται ότι

$$(1 - \alpha)e^{t_0 a} + \alpha e^{t_0 b} \cong 1$$

και

$$\gamma e^{t_1 a} + (1 - \gamma)e^{t_1 b} \cong 1,$$

από το οποίο έχουμε

$$\alpha \cong (1 - e^{t_0 a}) / (e^{t_0 b} - e^{t_0 a}) \quad (\text{III.D.19})$$

και

$$\gamma \cong (1 - e^{t_1 b}) / (e^{t_1 a} - e^{t_1 b}). \quad (\text{III.D.20})$$

Ως παράδειγμα, ας υποθέσουμε ότι παίρνουμε $g = \log(p_1/p_0)$, ώστε

$$ST(a, b; g) = \text{SPRT}(e^a, e^b).$$

Τότε έχουμε

$$\begin{aligned} M_0(t) &= \int_{\mathbf{R}} \exp\{t \log[p_1(y)/p_0(y)]\} p_0(y) \mu(dy) \\ &= \int_{\mathbf{R}} [p_1(y)]^t [p_0(y)]^{1-t} \mu(dy). \end{aligned}$$

Σημειώνοντας ότι

$$M_0(1) = \int_{\mathbf{R}} p_1(y) \mu(dy) = 1,$$

έχουμε $t_0 = 1$. Παρομοίως για την περίπτωση αυτή

$$M_1(t) = \int_{\mathbf{R}} [p_1(y)]^{t+1} [p_0(y)]^{-t} \mu(dy),$$

που παράγει $t_1 = -1$. Ορίζοντας $A = e^a$ και $B = e^b$, τα (III.D.19) και (III.D.20) συνεπάγονται ότι

$$\alpha \cong (1 - A)/(1 - B) \quad \text{και} \quad \gamma \cong (1 - B^{-1})/(A^{-1} - B^{-1}),$$

που είναι ισοδύναμα με τις προσεγγίσεις Wald του (III.D.15).

Αν $\mu = 0$, τότε δεν υπάρχει μη μηδενικό t_0 τέτοιο ώστε $M_0(t_0) = 1$. Αλλά σ' αυτή την περίπτωση, το συμπλήρωμα στην ταυτότητα του Wald συνεπάγεται ότι $E\{S_N | H_0\} = 0$, το οποίο παράγει

$$a(1 - \alpha) + b\alpha \cong 0$$

ή

$$\alpha \cong -a/(b - a). \quad (\text{III.D.21})$$

Παρομοίως, αν $\mu_1 = 0$, έχουμε

$$\gamma \cong -b(a - b). \quad (\text{III.D.22})$$

Με παρόμοιο τρόπο, μπορούμε να προσεγγίσουμε τα αναμενόμενα μεγέθη δείγματος του $ST(g; a, b)$ χρησιμοποιώντας την Πρόταση III.D.3. Πρώτα, υποθέτουμε ότι $\mu_0 \neq 0$. Τότε, αγνοώντας την υπέρβαση των ορίων, το συμπλήρωμα συνεπάγεται ότι

$$E\{N|H_0\} = \frac{1}{\mu_0} E\{S_N|H_0\} \cong \frac{1}{\mu_0} (a(1 - \alpha) + b\alpha).$$

Αν $\mu_0 = 0$, μπορούμε να χρησιμοποιήσουμε

$$E\{N|H_0\} = \frac{1}{\sigma_0^2} E\{S^2|H_0\} \cong \frac{1}{\sigma_0^2} (a^2(1 - \alpha) + B^2\alpha).$$

Παρόμοιες παραστάσεις ισχύουν και για το $E\{N|H_1\}$. Ας σημειωθεί ότι είτε το (a, b) ή το (α, γ) μπορούν να απαλειφθούν από τις παραστάσεις αυτές μέσω των προσεγγίσεων των (III.D.19) και (III.D.20) ή των (III.D.21) και (III.D.22). Για παράδειγμα, αν $\mu_j = 0$, έχουμε

$$E\{N|H_j\} \cong -ab/\sigma_j^2.$$

Για τη SPRT (e^a, e^b) έχουμε $\mu_j \neq 0$ για $j = 0$ και 1 και οι προσεγγίσεις Wald παράγουν

$$E\{N|H_0\} \cong \frac{1}{\mu_0} \left[(1 - \alpha) \log \frac{\gamma}{1 - \alpha} + \alpha \log \frac{1 - \gamma}{\alpha} \right]$$

και (III.D.23)

$$E\{N|H_1\} \cong \frac{1}{\mu_1} \left[\gamma \log \frac{\gamma}{1 - \alpha} + (1 - \gamma) \log \frac{1 - \gamma}{\alpha} \right].$$

Τώρα, ας εξετάσουμε το ακόλουθο παράδειγμα

Παράδειγμα III.D.2: Σύγκριση μιας ακολουθιακής και μιας FSS ανίχνευσης

Για να απεικονίσετε τις εξοικονομήσεις σε αναμενόμενο μέγεθος δείγματος που προκαλείται από τη χρήση ενός ελέγχου SPRT αντί της χρήσης ενός ελέγχου FSS (δείγματος παγίου μεγέθους), εξετάστε ξανά το πρόβλημα ανίχνευσης διαρκούς σήματος στο Παράδειγμα III.D.1.

Φυσικά, ο καλύτερος FSS ανιχνευτής είναι ο ανιχνευτής λόγου πιθανοφάνειας, ο οποίος για έναν δεδομένο δείγμα μεγέθους n , έχει πιθανότητες σφάλματος α και γ που συνδέονται μέσω της παράστασης (III.B.31).

$$(1 - \gamma) = 1 - \Phi(\Phi^{-1}(1 - \alpha) - n^{1/2}\theta/\sigma). \quad (\text{III.D.24})$$

Αρα για δεδομένα α και γ , το μέγεθος δείγματος που απαιτείται ώστε ο FSS έλεγχος λόγου πιθανοφάνειας να παράγει αυτές τις πιθανότητες σφάλματος δίνεται με την αντιστροφή του (III.D.24), δηλαδή

$$n_{FSS} = \lceil \sigma^2 [\Phi^{-1}(1 - \alpha) - \Phi^{-1}(\gamma)]^2 / \theta^2 \rceil, \quad (\text{III.D.25})$$

όπου το $\lceil x \rceil$ συμβολίζει τον μικρότερο ακέραιο που δεν είναι μικρότερος από το x .

Για τον SPRT με όρια επιλεγμένα για δεδομένες πιθανότητες σφάλματος, τα αναμενόμενα μεγέθη δείγματος δίνονται από το (III.D.23). Για να υπολογίσουμε την παράσταση πρέπει να υπολογίσουμε τα μ_0 και μ_1 που δίνονται από

$$\begin{aligned} \mu_j &= E\{\log[p_1(Y_1)/p_0(Y_1)]|H_j\} \\ &= \frac{\theta}{\sigma^2} E\{(Y_1 - \theta/2)|H_j\} \\ &= \begin{cases} -\theta^2/2\sigma^2 & \text{if } j = 0 \\ +\theta^2/2\sigma^2 & \text{if } j = 1. \end{cases} \end{aligned} \quad (\text{III.D.26})$$

Για λόγους απλότητας υποθέτουμε ότι $\alpha = \gamma$, περίπτωση στην οποία

$$\begin{aligned} E\{N|H_1\} &= E\{N|H_0\} \\ &\cong 2\sigma^2 \left[\gamma \log \frac{\gamma}{1 - \alpha} + (1 - \gamma) \log \frac{1 - \gamma}{\alpha} \right]. \end{aligned} \quad (\text{III.D.27})$$

Για $\alpha = \gamma = 0.1$ και $\theta^2/\sigma^2 = 1$, ο υπολογισμός των (III.D.25) και (III.D.27) δίνει $n_{FSS} \cong 22$ και $E\{N|H_j\} \cong 9$. Αρα, κατά μέσο όρο, ο SPRT χρησιμοποιεί λιγότερα από τα μισά δείγματα του FSS ελέγχου στην περίπτωση αυτή. Αυτό που είναι ενδιαφέρον είναι ότι, χρησιμοποιώντας τα (III.D.25) και (III.D.27) μπορεί να αποδειχθεί ότι

$$\lim_{\alpha=\gamma \rightarrow 0} \frac{E\{N|H_j\}}{n_{FSS}} = 1/4. \quad (\text{III.D.28})$$

Αρα, για εξαφανιζόμενα μικρές πιθανότητες σφάλματος (με $\alpha = \gamma$) ο SPRT απαιτεί μόνο το ένα τέταρτο δειγμάτων, κατά μέσο όρο, από αυτά που απαιτεί ο FSS έλεγχος.

Το παράδειγμα III.D.2 απεικονίζει ότι ο SPRT μπορεί, κατά μέσο όρο, να προσφέρει σημαντικές εξοικονομήσεις σε σχέση με τον καλύτερο FSS έλεγχο σε ό,τι αφορά τον αριθμό

των δειγμάτων που απαιτούνται για την πραγματοποίηση ενός ελέγχου με δεδομένο επίπεδο επίδοσης. Αυτό είναι ιδιαίτερα ευνοϊκό σε εφαρμογές στις οποίες πρόκειται να διεξαχθεί ένας μεγάλος αριθμός πανομοιότυπων ελέγχων. Ένα παράδειγμα μιας τέτοιας εφαρμογής είναι το ερευνητικό ραντάρ, στο οποίο το ραντάρ πραγματοποιεί έναν έλεγχο (παρόν στόχος ως προς απόντα στόχο) σε κάθε ένα από τα πολλά κελιά σ' ένα πεδίο έρευνας.

Από τα οφέλη σε επίδοση που φαίνονται στο Παράδειγμα III.D.2 θα μπορούσε να δημιουργηθεί το ερώτημα: Γιατί να μην εγκαταλείψουμε τη χρήση των FSS ελέγχων λόγω πιθανοφάνειας προς χάρη των SPRT για όλες τις περιπτώσεις. Δυστυχώς, οι SPRT έχουν αρκετά πρακτικά μειονεκτήματα που δεν είναι άμεσα προφανή από την παραπάνω ανάλυση. Ένα τέτοιο μειονέκτημα έγκειται στο γεγονός ότι παρόλο που το μέγεθος δείγματος ενός SPRT είναι πεπερασμένο με πιθανότητα 1, δεν είναι φραγμένο. Ο SPRT εξοικονομεί δείγματα παίρνοντας γρήγορες αποφάσεις, όταν η υπόθεση είναι σαφής ως προς τα παρατηρούμενα δεδομένα. Από την άλλη, αν τα παρατηρούμενα δεδομένα είναι ασαφή, ο SPRT μπορεί να λειτουργήσει για ένα μεγάλο αριθμό δειγμάτων. Όπως υποδεικνύει το θεώρημα Wald – Wolfowitz, ο μέσος όρος αυτών των δύο επιδράσεων είναι ευεργετικός. Ωστόσο, η περιστασιακή μακρά λειτουργία μπορεί να μην είναι πρακτική για πολλές εφαρμογές. Ευτυχώς, αυτή η δυσκολία μπορεί να ξεπεραστεί αρκετά εύκολα, μετασχηματίζοντας τον SPRT ώστε να σταματήσει τη δειγματοληψία και να πάρει μια δύσκολη (μονού καταφλιού) απόφαση μετά από κάποιο μέγιστο αριθμό δειγμάτων. Αυτός ο τύπος ελέγχου είναι γνωστός ως αποκομμένος SPRT, και οι αποκομμένοι SPRT διατηρούν τις θετικές ιδιότητες των SPRT με την προϋπόθεση ότι το σημείο αποκοπής δεν είναι πάρα πολύ μικρό.

Ένα άλλο πρακτικό μειονέκτημα των SPRT είναι ότι η πραγματοποίησή τους απαιτεί ακριβή γνώση και του p_0 και του p_1 . Για παράδειγμα, στο πρόβλημα ανίχνευσης διαρκούς σήματος του Παραδείγματος III.D.1 είναι απαραίτητο να γνωρίζουμε τη μόνη τιμή θ , προκειμένου να πραγματοποιήσουμε τον έλεγχο. Αυτό έρχεται σε αντίθεση με τον FSS Neyman – Pearson έλεγχο για τον ίδιο έλεγχο που είναι ομοιόμορφα πιο ισχυρός για $\theta > 0$. Μια λανθασμένη εικασία ως προς την πραγματική μόνη τιμή μπορεί να έχει ως αποτέλεσμα κάποια απώλεια στην επίδοση του SPRT στην περίπτωση αυτή. Για παράδειγμα, αν η παράμετρος πραγματικής θέσης ήταν μόνο η μισή από την εικαζόμενη, η στατιστική ελέγχου $\theta \sum_{k=1}^n (y_k - \theta/2) / \sigma^2$ θα κυμαινόταν γύρω στο μηδέν ως προς το H_1 , έχοντας πιθανόν ως αποτέλεσμα πολύ μακρείς ελέγχους. Αυτό μπορεί να αμβλυνθεί σε κάποιο βαθμό με την αποκοπή. Ωστόσο, αυτός ο τύπος προβλήματος αποτελεί έναν από τους βασικούς περιορισμούς των ακολουθιακών ελέγχων.

Ένα τρίτο μειονέκτημα των ακολουθιακών ελέγχων είναι ότι η θεωρία αυτών των ελέγχων περιορίζεται όταν δεν μπορεί να γίνει επίκληση της i.i.d. παραδοχής. Σημειώστε ότι η αρχική μας εξαγωγή του SPRT ως βέλτιστου Μπαϊεσιανού ελέγχου δεν θα λειτουργούσε αν δεν είχαμε ανεξαρτησία από το παρελθόν και το μέλλον και αν το μέλλον δεν ήταν όμοιο σε κάθε στάδιο.

Παρά αυτά τα τρία μειονεκτήματα, τα πλεονεκτήματα των SPRT τους καθιστούν πολύ ελκυστικούς για πολλές εφαρμογές. Μια σειρά πρόσθετων ιδιοτήτων του SPRT και των πρακτικών τροποποιήσεων του αναλύονται σ' ένα ερευνητικό άρθρο από τον Tantarata (1986)

III. Ε Μη παραμετρική και ανθεκτική ανίχνευση

Στο κεφάλαιο II και στην Ενότητα III.B εξετάσαμε προβλήματα ελέγχου υπόθεσης και ανίχνευσης σήματος ως προς μια ποικιλία παραδοχών σχετικά με τη στατιστική συμπεριφορά των διαθέσιμων παρατηρήσεων. Μια παραδοχή κοινή σε όλα αυτά τα προβλήματα είναι ότι η κατανομή πιθανότητας των δεδομένων είναι γνωστή (ίσως μόνο μέχρι ένα σύνολο άγνωστων παραμέτρων) ως προς κάθε υπόθεση.

Σε πρακτικές καταστάσεις, συχνά είναι μη ρεαλιστικό να εικάσουμε ότι οι κατανομές αυτές είναι γνωστές με ακρίβεια και ενίοτε δεν μπορεί καν να εικασθεί ότι είναι γνωστές κατά προσέγγιση. Χωρίς αυτή τη γνώση, οι τεχνικές των προηγούμενων ενοτήτων δεν μπορούν να εφαρμοστούν ευθέως και άρα πρέπει να τεθούν εναλλακτικά κριτήρια σχεδιασμού. Δύο φιλοσοφίες σχεδιασμού που μπορούν να εφαρμοστούν στο πλαίσιο αυτό είναι η μη παραμετρική και η ανθεκτική ανίχνευση. Βασικά, οι μη παραμετρικές τεχνικές χειρίζονται το πρόβλημα της ανίχνευσης σημάτων μόνο με πολύ «τραχιές» (γενικές και αόριστες) πληροφορίες σχετικά με τη στατιστική συμπεριφορά του θορύβου, ενώ οι ανθεκτικές τεχνικές είναι εφαρμόσιμες σε καταστάσεις στις οποίες οι στατιστικές θορύβου είναι γνωστές κατά προσέγγιση αλλά όχι ακριβώς. Στην ενότητα αυτή, παρουσιάζουμε μια σύντομη επισκόπηση αυτών των δύο τύπων στρατηγικών ανίχνευσης.

Εξετάστε το ακόλουθο γενικό πρόβλημα σύνθετου ελέγχου δυαδικής υπόθεσης που βασίζεται σε μια ανεξάρτητη και ταυτόσημα κατανομημένη (i.i.d.) ακολουθία παρατήρησης.

$$H_0 : Y_k \sim P \in \mathcal{P}_0, \quad k = 1, 2, \dots, n$$

ως προς

(III.E.1)

$$H_1 : Y_k \sim P \in \mathcal{P}_1, \quad k = 1, 2, \dots, n,$$

όπου οι \mathcal{P}_0 και \mathcal{P}_1 είναι δύο μη επικαλυπτόμενες κλάσεις πιθανών οριακών κατανομών για τις παρατηρήσεις. Αυτό το πρόβλημα θεωρείται ότι είναι ένα πρόβλημα *παραμετρικού*

ελέγχου υπόθεσης αν οι κλάσεις \mathcal{P}_0 και \mathcal{P}_1 μπορούν να παραμετροποιηθούν από μια πραγματική ή διανυσματική παράμετρο. Για παράδειγμα, τα προβλήματα ελέγχου σύνθετης

υπόθεσης του αναλύθηκαν στην Ενότητα II.E είναι παραμετρικά προβλήματα. Αν τα \mathcal{P}_0 και

\mathcal{P}_1 δεν μπορούν να παραμετροποιηθούν με αυτόν τον τρόπο, το (III.E.1) θεωρείται ότι είναι ένα πρόβλημα *μη παραμετρικού* ελέγχου υπόθεσης. Η γενική ιδέα στα μη παραμετρικά

προβλήματα είναι ότι τα \mathcal{P}_0 και \mathcal{P}_1 είναι πολύ ευρέα για να παραμετροποιηθούν από μια παράμετρο πεπερασμένης διάστασης.

Ένα παράδειγμα μη παραμετρικού προβλήματος ελέγχου υπόθεσης είναι το πρόβλημα ελέγχου θέσης

$$H_0 : Y_k = N_k, \quad k = 1, 2, \dots, n$$

ως προς

(III.E.2)

$$H_1 : Y_k = N_k + \theta, \quad k = 1, 2, \dots, n,$$

στο οποίο η $\{N_k\}_{k=1}^n$ είναι μια i.i.d. ακολουθία για την οποία γνωρίζουμε μόνο ότι η οριακή της κατανομή είναι συμμετρική περίπου στο μηδέν. Ένα τέτοιο μοντέλο μπορεί να προκύψει, για παράδειγμα, στο πρόβλημα ανίχνευσης ενός διαρκούς σήματος σ' ένα περιβάλλον θορύβου που είναι εντελώς άγνωστο στατιστικά εκτός από τις ιδιότητες του πανομοιότυπου, της ανεξαρτησίας και της συμμετρίας. Το πρόβλημα είναι μη παραμετρικό, αφού η τάξη όλων των συμμετρικών κατανομών είναι σίγουρα μη πεπερασμένη διάστασης.

Οι ανθεκτικοί και οι μη παραμετρικοί έλεγχοι υπόθεσης είναι σχεδιασμένοι μέσα στο πλαίσιο των μη παραμετρικών υποθέσεων. Ξεκινάμε τη συζήτησή μας με τους μη παραμετρικούς ελέγχους. Οι ανθεκτικοί έλεγχοι θα συζητηθούν στη συνέχεια.

III.E.1 ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΑΝΙΧΝΕΥΣΗ

Μιλώντας γενικά, μη παραμετρικός έλεγχος είναι αυτός που σχεδιάζεται για να λειτουργήσει

σε ευρείες κλάσεις \mathcal{P}_0 και \mathcal{P}_1 με κάποια χαρακτηριστικά επίδοσης να παραμένουν αμετάβλητα σε όλες τις κλάσεις. Οι έλεγχοι αυτοί είναι συνήθως απλοί, που χρησιμοποιούν πρόχειρες πληροφορίες για τα δεδομένα (π.χ., πρόσημα, διατάξεις κτλ.), παρά τις ακριβείς τιμές των δεδομένων. Σχεδόν πάντα, το χαρακτηριστικό επίδοσης που πρέπει να διατηρείται αμετάβλητο σε μη παραμετρικά προβλήματα είναι η πιθανότητα ψευδούς συναγεμού.

Επομένως, ο τυπικός ορισμός ενός μη παραμετρικού ελέγχου (ή ανιχνευτή) για το (III.E.1)

είναι αυτός στον οποίο η πιθανότητα ψευδούς συναγεμού είναι διαρκής ως προς το \mathcal{P}_0 . Για καταστάσεις στις οποίες το Y αποτελεί μια ακολουθία παρατηρήσεων Y_1, \dots, Y_n , μας

ενδιαφέρουν και για ακολουθίες ελέγχων, $\{\delta_n(y_1, \dots, y_n)\}_{n=1}^{\infty}$ που είναι *ασυμπτωματικά μη παραμετρικές* για το (III.E.1), δηλ. ακολουθίες ελέγχων για τις οποίες το

$$\lim_{n \rightarrow \infty} P_F(\delta_n) \text{ είναι σταθερό για όλα τα } P \in \mathcal{P}_0.$$

Οι μη παραμετρικοί έλεγχοι και ανιχνευτές έχουν βρει πολλές εφαρμογές σε πεδία, όπως είναι τα ραντάρ και τα σόναρ. Σε αυτές τις εφαρμογές, οι μη παραμετρικοί ανιχνευτές καλούνται ενίοτε *ανιχνευτές σταθερού ρυθμού ψευδούς συναγεμού (CFAR)*. Στις ακόλουθες παραγράφους, περιγράφουμε κάποιες από τις πιο ευρέως χρησιμοποιούμενες μη παραμετρικούς μεθόδους.

Ο προσημικός έλεγχος

Ας υποθέσουμε ότι έχουμε μια ακολουθία Y_1, \dots, Y_n ανεξάρτητων και .. κατανομημένων (i.i.d.)

αληθούς τιμής παρατηρήσεων. Ορίστε την παράμετρο p ως $p = P(Y_1 > 0)$ και εξετάστε το ζεύγος υπόθεσης

$$H_0 : p = \frac{1}{2}$$

Ως προς (III.E.3)

$$H_1 : \frac{1}{2} < p < 1.$$

Με λέξεις, το (III.E.3) είναι η υπόθεση ότι τα Y_k έχουν μηδενικό διάμεσο ως προς την υπόθεση ότι ο διάμεσος των Y_k είναι μεγαλύτερος από μηδέν. Με διάφορες ερμηνείες, το μοντέλο αυτό προκύπτει σε διάφορες εφαρμογές. Για παράδειγμα, το μοντέλο του (III.E.2) με $\theta > 0$ είναι ένα υποσύνολο το μοντέλου αυτού.

Οι υποθέσεις του (III.E.3) είναι μη παραμετρικές αφού, σε ό,τι αφορά το (III.E.1), έχουμε

$$\mathcal{P}_0 = \{P \in \mathcal{M} | P((0, \infty)) = \frac{1}{2}\} \quad (\text{III.E.4})$$

και

$$\mathcal{P}_1 = \{P \in \mathcal{M} | 1 > P((0, \infty)) > \frac{1}{2}\}, \quad (\text{III.E.5})$$

όπου το \mathcal{M} συμβολίζει την κλάση όλων των κατανομών στο $(\mathbb{R}, \mathcal{B})$. Καμία από αυτές τις κλάσεις δεν μπορεί να παραμετροποιηθεί (δηλ. να τεθεί σε μία προς μία αντιστοιχία) με μια παράμετρο πεπερασμένων διαστάσεων.

Για να παράγουμε ένα βέλτιστο έλεγχο για το (III.E.3) ας επιλέξουμε πρώτα μια αυθαίρετη κατανομή Q_1 στο \mathcal{P}_1 . Για λόγους απεικόνισης, θα υποθέσουμε ότι το Q_1 έχει πυκνότητα q_1 , παρόλο που το ακόλουθο ανάπτυγμα μπορεί να πραγματοποιηθεί χωρίς αυτή την υπόθεση. Ορίστε δύο συναρτήσεις

$$q_1^+(x) = \begin{cases} q_1(x) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

και

$$q_1^-(x) = \begin{cases} 0 & \text{if } x > 0 \\ q_1(x) & \text{if } x \leq 0 \end{cases} \quad (\text{III.E.6})$$

και ορίστε μια πυκνότητα q_0 στο $(\mathbb{R}, \mathcal{B})$ με

$$q_0(x) = \frac{q_1^+(x)}{2 \int_0^\infty q_1(t) dt} + \frac{q_1^-(x)}{2 \int_{-\infty}^0 q_1(t) dt}. \quad (\text{III.E.7})$$

Σημειώστε ότι η κατανομή Q_0 που αντιστοιχεί στην πυκνότητα q_0 είναι στοιχείο του \mathcal{P}_0 αφού

$$Q_0((0, \infty)) = \int_0^\infty q_0(x) dx = \frac{\int_0^\infty q_1^+(x) dx}{2 \int_0^\infty q_1(x) dx} = \frac{1}{2}. \quad (\text{III.E.8})$$

Εξετάστε το απλό ζεύγος υπόθεσης

$$H'_0 : Y_k \sim Q_0, \quad k = 1, \dots, n$$

Ως προς (III.E.9)

$$H'_1 : Y_k \sim Q_1, \quad k = 1, \dots, n.$$

Σύμφωνα με το λήμμα Neyman – Pearson, ένας πιο ισχυρός α - επιπέδου έλεγχος του H'_0 ως προς το H'_1 είναι ο έλεγχος λόγου πιθανοφάνειας που βασίζεται στη σύγκριση της στατιστικής

$$L(\underline{y}) = \prod_{k=1}^n \frac{q_1(y_k)}{q_0(y_k)} \quad (\text{III.E.10})$$

μ' ένα κατώφλι. Σημειώστε ότι

$$\frac{q_1(y_k)}{q_0(y_k)} = \begin{cases} 2Q_1^+ & \text{if } y_k > 0 \\ 2(1 - Q_1^+) & \text{if } y_k \leq 0 \end{cases} \quad (\text{III.E.11})$$

όπου $Q_1^+ \triangleq Q_1((0, \infty))$, έτσι ώστε η $L(\underline{y})$ μπορεί να γραφεί ως

$$L(\underline{y}) = 2^n [Q_1^+]^n [Q_1^+ / (1 - Q_1^+)]^{t(\underline{y})}, \quad (\text{III.E.12})$$

όπου

$$t(\underline{y}) \triangleq \sum_{k=1}^n u(y_k), \quad (\text{III.E.13})$$

με το u να συμβολίζει τη συνάρτηση μοναδιαίου βήματος που ορίζεται από, $u(x) = 1$ αν $x > 0$ και $u(x) = 0$ αν $x \leq 0$. Σημειώστε ότι $t(\underline{y})$ είναι ο αριθμός των παρατηρούμενων y_k που είναι θετικά

Με την υπόθεση $\frac{1}{2} < Q_1^+ < 1$, τότε $Q_1^+/(1 - Q_1^+) > 1$. Αυτό συνεπάγεται ότι η $L(\underline{y})$ είναι μια μονότονα αυξανόμενη συνάρτηση του $t(\underline{Y})$ και έτσι ένας πιο ισχυρός α -επιπέδου έλεγχος του H_0' ως προς το H_1' δίνεται από

$$\bar{\delta}_s(\underline{y}) = \begin{cases} 1 & \text{if } t(\underline{y}) > \tau, \\ \gamma & \text{if } t(\underline{y}) = \tau, \\ 0 & \text{if } t(\underline{y}) < \tau, \end{cases} \quad (\text{III.E.14})$$

όπου τα γ και τ είναι η τυχαιοποίηση και το κατώφλι για μια πιθανότητα ψευδούς συναγερμού α . Είναι εύκολο να δούμε ότι η $t(\underline{Y})$ είναι μια διωνυμική τυχαία μεταβλητή με παραμέτρους (n, Q_1^+) ως προς το H_1' και $(n, \frac{1}{2})$ ως προς το H_0' . Άρα για ένα μέγεθος α , το κατώφλι τ είναι ο μικρότερος ακέραιος, τέτοιος ώστε

$$2^{-n} \sum_{k=\tau}^n \frac{n!}{(n-k)!k!} \leq \alpha \quad (\text{III.E.15})$$

και η σταθερά τυχαιοποίησης είναι

$$\gamma = \frac{\alpha - 2^{-n} \sum_{k=\tau+1}^n \frac{n!}{(n-k)!k!}}{2^{-n} \frac{n!}{(n-\tau)!\tau!}}. \quad (\text{III.E.16})$$

Η κατανομή του $t(\underline{Y})$ είναι διωνυμική με παραμέτρους $(n, \frac{1}{2})$ για κάθε $Q_0 \in \mathcal{P}_0$; άρα το (III.E.14) έχει μέγεθος α για ολόκληρη την κλάση \mathcal{P}_0 . Επομένως, ο έλεγχος του (III.E.14) είναι μη παραμετρικός για το πρόβλημα αυτό. Αυτό συνεπάγεται επίσης ότι το (III.E.14) είναι ο πιο ισχυρός α -επιπέδου έλεγχος του H_0 ως προς το H_1' . Επιπλέον, σημειώνοντας ότι ο έλεγχος του (III.E.14) δεν εξαρτάται από την επιλογή του Q_1 , βλέπουμε ότι είναι ένας ομοιόμορφα πιο ισχυρός α -επιπέδου έλεγχος του H_0 ως προς το H_1 .

Καθώς η $t(\underline{Y})$ είναι διωνυμική (n, p) ως προς το H_1 , η ανίχνευση πιθανότητας του (III.E.14) δίνεται από

$$P_D = \sum_{k=\tau+1}^n \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} + \gamma \frac{n!}{(n-\tau)!\tau!} p^\tau (1-p)^{n-\tau}. \quad (\text{III.E.17})$$

Από το (III.E.17) βλέπουμε ότι παρόλο που η πιθανότητα ψευδούς συναγερμού του δ_s είναι σταθερά ως προς το H_0 , η ανίχνευση πιθανότητάς της εξαρτάται από το p και άρα δεν εξαρτάται από την επιλογή $P \in \mathcal{P}_1$. Μπορεί να αποδειχθεί ότι η P_D αυξάνεται μονοτονικά από το α στο 1 καθώς το p αυξάνεται από το $1/2$ στο 1.

Ο έλεγχος του (III.E.14) χρησιμοποιεί μόνο τα (αλγεβρικά) πρόσημα των παρατηρήσεων y_1, \dots, y_n και έτσι είναι γνωστός ως προσημικός έλεγχος. Παρόλο που ο προσημικός έλεγχος είναι α -επιπέδου UMP για \mathcal{P}_0 ως προς \mathcal{P}_1 , θα είχαμε καλύτερο αποτέλεσμα απ' ό,τι με τον προσημικό έλεγχο, αν γνωρίζαμε την ακριβή κατανομή των παρατηρήσεων ως προς τις δύο υποθέσεις, χρησιμοποιώντας τον έλεγχο λόγου πιθανοφάνειας μεταξύ αυτών των δύο κατανομών. Αυτό σημαίνει ότι ο προσημικός έλεγχος δεν είναι ένας UMP α -επιπέδου έλεγχος για ένα συγκεκριμένο $P \in \mathcal{P}_0$ ως προς την κλάση \mathcal{P}_1 . Ένα ενδιαφέρον ερώτημα είναι: Πόση απόδοση χάνουμε, αν υποθέσουμε για την κατανομή μόνο τις πολύ «τραχιές» παραδοχές που γίνονται στην υπόθεση του (III.E.3);

Για να απαντήσουμε μερικώς στο ερώτημα αυτό, εξετάζουμε μια ασυμπτωτική ($n \rightarrow \infty$) ανάλυση βασισμένη στην ασυμπτωτική σχετική αποδοτικότητα (ARE) του Pitman που παρουσιάστηκε στην Ενότητα III.C. Θυμηθείτε ότι η ασυμπτωτική αποδοτικότητα ενός ανιχνευτή που συνδέεται με κάποιον άλλον είναι μια μέτρηση του σχετικού αριθμού δειγμάτων που χρειάζεται ο ένας για να πετύχει την ίδια απόδοση με τον άλλον στο όριο, καθώς ο αριθμός των δειγμάτων αυξάνεται χωρίς φράγμα.

Για να αναλύσουμε τον προσημικό έλεγχο μέσω της ARE, χρειαζόμαστε συγκεκριμένο μοντέλο και έλεγχο για να χρησιμοποιήσουμε ως βάση για σύγκριση. Για να το κάνουμε αυτό, εξετάζουμε εκ νέου το μοντέλο

$$H_0 : Y_k = N_k, \quad k = 1, 2, \dots, n$$

ως προς (III.18.E)

$$H_1 : Y_k = N_k + \theta, \quad k = 1, 2, \dots, n$$

όπου η N_1, N_2, \dots, N_n είναι μια i.i.d. ακολουθία με μηδενικό μέσο. Όπως αναφέρθηκε παραπάνω, το (III.E.18) είναι, φυσικά, το πρόβλημα ανίχνευσης ενός σταθερού σήματος σε i.i.d. προσθετικό θόρυβο.

Αν υποθέσουμε ότι η κατανομή θορύβου στο (III.E.18) είναι $\mathcal{N}(0, \sigma^2)$ με το σ^2 άγνωστο, τότε μπορεί να αποδειχθεί [βλέπε, π.χ., Lehmann (1986)] ότι ένας UMP (μεταξύ αμερόληπτων ελέγχων) α -επιπέδου ελέγχος του H_0 ως προς το H_1 , δίνεται από

$$\tilde{\delta}_t(\mathbf{y}) = \begin{cases} 1 & \text{if } \frac{\bar{y}}{[\bar{s}^2]^{1/2}} > \tau \\ \gamma & \text{if } \frac{\bar{y}}{[\bar{s}^2]^{1/2}} = \tau \\ 0 & \text{if } \frac{\bar{y}}{[\bar{s}^2]^{1/2}} < \tau \end{cases} \quad (\text{III.E.19})$$

όπου \bar{y} είναι η μέση τιμή δείγματος $[\bar{y} \triangleq (1/n) \sum_{k=1}^n y_k]$ και \bar{s}^2 είναι η διακύμανση δείγματος $[\bar{s}^2 = (1/n) \sum_{k=1}^n (y_k - \bar{y})^2]$.

Ο έλεγχος του (III.E.19) είναι γνωστός ως t -έλεγχος. Όχι μόνο αυτός ο έλεγχος είναι UMP για την κατά Gauss περίπτωση του (III.E.18), αλλά επίσης, επιλέγοντας το κατώφλι

$\tau = \Phi^{-1}(1 - \alpha)/\sqrt{n}$ και την τυχαιοποίηση γ αυθαίρετα, ο t -έλεγχος γίνεται ασυμπτωτικά μη παραμετρικός στο $P_F = \alpha$ για το (III.E.18) με κάθε κατανομή θορύβου που έχει μηδενικό μέσο και πεπερασμένη διακύμανση. Για να γίνει κατανοητό αυτό, σημειώνουμε ότι η πιθανότητα ψευδούς συναγερού του ελέγχου αυτού δίνεται από

$$\begin{aligned} P_F(\tilde{\delta}_t) &= P_0(\bar{Y}/(\bar{S}^2)^{1/2} > \tau) + \gamma P_0(\bar{Y}/(\bar{S}^2)^{1/2} = \tau) \\ &= P_0\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k/(\bar{S}^2)^{1/2} > \Phi^{-1}(1 - \alpha)\right) \\ &\quad + \gamma P_0\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k/(\bar{S}^2)^{1/2} = \Phi^{-1}(1 - \alpha)\right). \end{aligned} \quad (\text{III.E.20})$$

Με τον αδύναμο νόμο των μεγάλων αριθμών, το \bar{S}^2 συγκλίνει σε πιθανότητα στο $\text{Var}(N_1)$, και με το θεώρημα κεντρικού ορίου, το $n^{-1/2} \sum_{k=1}^n Y_k/(\bar{S}^2)^{1/2}$ συγκλίνει σε κατανομή σε μια $\mathcal{N}(0, 1)$ τυχαία μεταβλητή ως προς το H_0 , άρα

$$\lim_{n \rightarrow \infty} P_F(\tilde{\delta}_t) = \frac{1}{\sqrt{2\pi}} \int_{\Phi^{-1}(1-\alpha)}^{\infty} e^{-x^2/2} dx = \alpha. \quad (\text{III.E.21})$$

Η ισότητα του (III.E.21) συνεπάγεται ότι ο t -έλεγχος είναι ασυμπτωτικά μη παραμετρικός. (Σημειώστε ότι το γ είναι άσχετο καθώς η οριακή κατανομή είναι συνεχής).

Από τα παραπάνω, βλέπουμε ότι ο t -έλεγχος είναι βέλτιστος για το (III.E.18) με κατά Gauss θόρυβο και είναι ασυμπτωτικά μη παραμετρικός για το (III.E.18) με θόρυβο πεπερασμένης διακύμανσης. Ας σημειωθεί ότι το πρώτο από αυτά τα προβλήματα αντιστοιχεί στον έλεγχο ενός υποσυνόλου των κατανομών από το (III.E.3) και, αν θέσουμε τον πρόσθετο περιορισμό ότι ο θόρυβος έχει μηδενικό διάμεσο επιπρόσθετα στο μηδενικό μέσο, το δεύτερο πρόβλημα επίσης αντιστοιχεί στον έλεγχο ενός υποσυνόλου των κατανομών για το (III.E.3). Είναι ενδιαφέρον να συγκρίνουμε τον προσημικό έλεγχο και τον t -έλεγχο υπό αυτές τις τελευταίες συνθήκες.

Αν υποθέσουμε ότι ο θόρυβος του (III.E.18) έχει μια pdf f που έχει μηδενικό μέσο, διακύμανση $\sigma^2 < \infty$ και είναι συνεχής στο μηδέν, τότε έπεται ευθέως από το θεώρημα Pitman – Noether (βλέπε Ενότητα III.C) ότι η ασυμπτωτική αποδοτικότητα του προσημικού ελέγχου σε σχέση με τον t -έλεγχο ως προς το (III.E.18) δίνεται από

$$\text{ARE}_{s,t} = 4\sigma^2 f^2(0). \quad (\text{III.E.22})$$

Για τη συγκεκριμένη περίπτωση του κατά Gauss θορύβου, στην οποία η f είναι η $\mathcal{N}(0, \sigma^2)$ πυκνότητα, το (III.E.22) γίνεται

$$\text{ARE}_{s,t} = 4\sigma^2 \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^2 = \frac{2}{\pi} \cong 0.64,$$

οπότε ο t -έλεγχος απαιτεί το 64% των δειγμάτων που απαιτούνται από έναν ισοδύναμο προσημικό έλεγχο. Εναλλακτικά, για την περίπτωση του Λαπλασιανού θορύβου ($f(x) = \frac{\alpha}{2} e^{-\alpha|x|}$), έχουμε $\sigma^2 = 2/\alpha^2$ και

$$\text{ARE}_{s,t} = \frac{8}{\alpha^2} \left(\frac{\alpha}{2} \right)^2 = 2.$$

Άρα για την περίπτωση αυτή, ο t -έλεγχος απαιτεί διπλάσια δείγματα από τον ισοδύναμο προσημικό έλεγχο. Θα πρέπει να σημειωθεί ότι ο προσημικός έλεγχος είναι βέλτιστος σε ό,τι αφορά την ασυμπτωτική αποδοτικότητα για την περίπτωση του Λαπλασιανού θορύβου.

Μπορεί να αποδειχθεί ότι [βλ. Kendall and Stuart (1961)] για κάθε συμμετρική μονοκόρυφη πυκνότητα [δηλ., $f(x) = f(-x)$ και $f(|x_1|) > f(|x_2|)$ αν $|x_1| < |x_2|$], η $\text{ARE}_{s,t}$ ικανοποιεί

την ανισότητα

$$\text{ARE}_{s,t} \geq 1/3. \quad (\text{III.E.23})$$

Άρα, ο t -έλεγχος απαιτεί τουλάχιστον το ένα τρίτο του αριθμού των δειγμάτων που απαιτούνται από έναν ισοδύναμο προσημικό έλεγχο υπό αυτές τις συνθήκες. Καθώς δεν υπάρχει αντίστοιχο άνω φράγμα στην $\text{ARE}_{s,t}$, ο προσημικός έλεγχος είναι προτιμότερος από τον t -έλεγχο όταν η κλάση των πιθανών κατανομών θορύβου είναι αρκετά ευρεία. Επιπλέον, ο προσημικός έλεγχος είναι ακριβώς μη παραμετρικός σε μια πολύ ευρεία κλάση κατανομών, ενώ ο t -έλεγχος είναι μόνο ασυμπτωτικά μη παραμετρικός σε μια κάπως στενότερη κλάση. Οι παράγοντες αυτοί, μαζί με την υπολογιστική απλότητα του προσημικού ελέγχου, καθιστούν τον προσημικό έλεγχο μία πολύ χρήσιμη εναλλακτική επιλογή σε σχέση με τους βέλτιστους ελέγχους των προηγούμενων εννοιών για προβλήματα ανίχνευσης σήματος. Τόσο ο προσημικός έλεγχος και ο t -έλεγχος χρησιμοποιούνται αρκετά συχνά σε εφαρμογές όπως η CFAR ανίχνευση ραντάρ (στη συγκεκριμένη αυτή εφαρμογή, ο προσημικός έλεγχος ορίζεται ενίοτε ως *δυναδικός ολοκληρωτής*).

Έλεγχος βαθμίδων

Από τα παραπάνω βλέπουμε ότι, παρόλο που χρησιμοποιεί πολύ «...| πληροφορίες σχετικά με τα παρατηρούμενα δεδομένα, ο προσημικός έλεγχος είναι αρκετά αποδοτικός ακόμη και

στη χειρότερη του περίπτωση συγκρινόμενος μ' ένα ανταγωνιστικό έλεγχο που χρησιμοποιεί πολύ περισσότερες πληροφορίες σχετικά με τα δεδομένα. Ωστόσο, χρησιμοποιώντας περισσότερες πληροφορίες για τα δεδομένα, ο μη παραμετρικός χαρακτήρας του προσημικού ελέγχου μπορεί να διατηρηθεί ενόσω βελτιώνεται η αποδοτικότητα της χειρότερης περίπτωσης σε σχέση με τον t -έλεγχο.

Για παράδειγμα, ας υποθέσουμε ότι αντικαθιστούμε τη στατιστική του προσημικού ελέγχου $t(\underline{y}) = \sum_{k=1}^n u(y_k)$ του (III.E.13) με μια σταθμισμένη εκδοχή $\sum_{k=1}^n \lambda_k u(y_k)$ όπου το λ_k είναι η βαθμίδα του y_k στο δείγμα y_1, \dots, y_n όταν αναδιατάσσεται σε αυξανόμενη σειρά απόλυτης τιμής. Δηλαδή, υποθέτουμε ότι κατατάσσουμε το y_1, \dots, y_n ως y_{k_1}, \dots, y_{k_n} όπου $|y_{k_1}| \leq |y_{k_2}| \leq \dots \leq |y_{k_n}|$, και πραγματοποιούμε έναν έλεγχο κατωφλιού βασισμένο στη στατιστική

$$t_W(\underline{y}) = \sum_{i=1}^n i u(y_{k_i}). \quad (\text{III.E.24})$$

Ο απορρέων έλεγχος είναι γνωστός ως έλεγχος Wilcoxon και είναι παράδειγμα ενός ελέγχου βαθμίδας, καθώς βασίζεται στις βαθμίδες των ατομικών παρατηρήσεων στο πλαίσιο ολόκληρου του δείγματος παρατήρησης.

Η στατιστική του ελέγχου Wilcoxon του (III.E.24) μπορεί να ξαναγραφεί ως

$$t_W(\underline{y}) = \sum_{k=1}^n \sum_{j=1}^k u(y_k + y_j), \quad (\text{III.E.25})$$

το αποτέλεσμα της οποίας δίνεται ως άσκηση. Μπορεί να αποδειχθεί από το (III.E.25) ότι ο έλεγχος Wilcoxon είναι μη παραμετρικός για την υπόθεση ότι τα Y_1, \dots, Y_n είναι i.i.d. με μια συμμετρική οριακή κατανομή [δηλ., $F_{Y_k}(b) = 1 - F_{Y_k}(-b)$ για όλα τα αληθή b]. Ας σημειωθεί ότι αυτή είναι μια μικρότερη κλάση μοντέλων από την κλάση όλων των κατανομών με μηδενικό διάμεσο (για τις οποίες ο προσημικός έλεγχος είναι μη παραμετρικός). Η ασυμπτωτική αποδοτικότητα του ελέγχου Wilcoxon σε σχέση με τον t -έλεγχο στο ζεύγος υπόθεσης του (III.E.18) δίνεται από το θεώρημα Pitman – Noether ως

$$\text{ARE}_{w,t} = 12\sigma^2 \left[\int_{-\infty}^{\infty} f^2(x) dx \right]^2, \quad (\text{III.E.26})$$

όπου έχουμε υποθέσει ότι οι μεταβλητές θορύβου έχουν συμμετρική πυκνότητα f . Για την περίπτωση του υπολογισμού του κατά Gauss θορύβου $[N_k \sim \mathcal{N}(0, \sigma^2)]$ του (III.E.26) δίνει

$\text{ARE}_{w,t} = 3/\pi = 0.955$. Άρα ο έλεγχος Wilcoxon είναι σχεδόν βέλτιστος για την κατά Gauss περίπτωση. Για τη Λαπλασιανή περίπτωση η $\text{ARE}_{w,t} = 1.5$ που υποδεικνύει απώλεια αποδοτικότητας της τάξης του 25% σε σχέση με τον προσημικό έλεγχο στην περίπτωση αυτή. Ωστόσο, μπορεί να αποδειχθεί, ελαχιστοποιώντας το

$\int_{-\infty}^{\infty} f^2(x)dx$ υποκείμενο στον περιορισμό $\int_{-\infty}^{\infty} x^2 f(x)dx = \sigma^2$ [για λεπτομέρειες, βλ. Kendall and Stuart (1961)] ότι

$$ARE_{w,t} \geq 0.864 \quad (\text{III.E.27})$$

για κάθε συμμετρική πυκνότητα θορύβου. Άρα, ο έλεγχος Wilcoxon είναι πάντα τουλάχιστον 86,4% αποδοτικός όπως ο t -έλεγχος και, καθώς δεν υπάρχει αντίστοιχο άνω φράγμα στην $ARE_{w,t}$ (η διακύμανση σ^2 δεν είναι φραγμένη), ο έλεγχος Wilcoxon προσφέρει σημαντικά πλεονεκτήματα σε σχέση με τον t -έλεγχο. Ωστόσο, ένα μειονέκτημα του ελέγχου Wilcoxon είναι ότι όλα τα δείγματα πρέπει να είναι αποθηκευμένα για να υπολογιστεί ο στατιστικός έλεγχός τους. Αυτό δεν ισχύει ούτε για τον προσημικό ούτε για τον t -έλεγχο.

Καλύτερη ακόμη επίδοση σε σχέση με τον t -έλεγχο μπορεί να επιτευχθεί με τη χρήση ελέγχων βαθμίδων που είναι πιο περίπλοκοι από τον έλεγχο Wilcoxon. Ένας τέτοιος έλεγχος είναι ο *Fisher-Yates* ή *έλεγχος κανονικών αποτελεσμάτων*, που χρησιμοποιεί το στατιστικό έλεγχο

$$t_{FY}(\underline{y}) = \sum_{i=1}^n h_n(iu(y_{k_i})), \quad (\text{III.E.28})$$

όπου το y_{k_1}, \dots, y_{k_n} είναι το διατεταγμένο δείγμα όπως στον έλεγχο Wilcoxon και όπου h_n είναι μια συνάρτηση που ορίζεται από

$$h_n(i) = \begin{cases} 0 & \text{if } i = 0 \\ E\{X_{(i)}\} & \text{if } i = 1, \dots, n, \end{cases} \quad (\text{III.E.29})$$

όπου οι $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ είναι διατεταγμένες τιμές των i.i.d. $N(0,1)$ τυχαίων μεταβλητών X_1, X_2, \dots, X_n . Ο έλεγχος Fisher – Yates έχει αποδοτικότητα σε σχέση με τον t -έλεγχο που ικανοποιεί

$$ARE_{FY,t} \geq 1$$

στο μοντέλο του (III.E.18) με συμμετρικό θόρυβο. Άρα, ο έλεγχος Fisher – Yates είναι πάντα, τουλάχιστον το ίδιο αποδοτικό με τον t -έλεγχο στο μοντέλο αυτό. Και πάλι, η αποδοτικότητα αυτή επιτυγχάνεται εις βάρος της πολυπλοκότητας.

Για περαιτέρω συζήτηση των ελέγχων βαθμίδων, ο αναγνώστης παραπέμπεται στα βιβλία των Hajek and Sidak (1967) και Kendall (1948).

Έλεγχος δύο καναλιών

Ένας αριθμός εφαρμογών περιλαμβάνει σύνολα παρατήρησης που αποτελούνται από δείγματα που λαμβάνονται από δύο ή περισσότερους αισθητήρες ή κανάλια. Τέτοιες εφαρμογές προκύπτουν σε προβλήματα σόναρ, σεισμολογίας και ραδιαστρονομίας στα οποία σχηματισμοί αισθητήρων χρησιμοποιούνται συχνά στην ανίχνευση σημάτων. Έχουν αναπτυχθεί διάφοροι σημαντικοί μη παραμετρικοί έλεγχοι γι' αυτού του τύπου προβλήματα, και κάποιιοι από αυτούς θα αναλυθούν εδώ για τη συγκεκριμένη περίπτωση στην οποία είναι διαθέσιμα δύο κανάλια παρατήρησης.

Εξετάζουμε μια ακολουθία παρατήρησης που αποτελείται από n ανεξάρτητα ζεύγη τυχαίων μεταβλητών. Δηλ., $\underline{Y} = [(U_1, V_1), (U_2, V_2), \dots, (U_n, V_n)]$, όπου (U_k, V_k) , $k = 1, \dots, n$, είναι αμοιβαία ανεξάρτητα.

Ένας τύπος προβλήματος σ' αυτό το πλαίσιο είναι αυτός της ανίχνευσης της παρουσίας ή απουσίας ενός κοινού τυχαίου σήματος σε δύο αισθητήρες. Το πρόβλημα αυτό μπορεί να μοντελοποιηθεί από το ακόλουθο ζεύγος υπόθεσης.

$$H_0 : \begin{array}{l} U_k = N_k \\ V_k = W_k \end{array}, k = 1, 2, \dots, n$$

ως προς

(III.E.30)

$$H_1 : \begin{array}{l} U_k = N_k + S_k \\ V_k = W_k + S_k \end{array}, k = 1, 2, \dots, n$$

όπου οι $\{N_k\}_{k=1}^n$, $\{W_k\}_{k=1}^n$, και $\{S_k\}_{k=1}^n$ είναι ανεξάρτητες ακολουθίες των i.i.d. τυχαίων μεταβλητών με συναρτήσεις οριακής κατανομής F_N, F_W , και F_S , αντίστοιχα.

Με τις παραδοχές ανεξαρτησίας, αποδεικνύεται ότι, ως προς το H_0 , κάθε ζεύγος (U_k, V_k) έχει κοινή συνάρτηση κατανομής

$$Q_0(u, v) = F_N(u)F_W(v) \quad (III.E.31)$$

και, ως προς το H_1 , το (U_k, V_k) έχει κοινή κατανομή

$$Q_1(u, v) = \int_{-\infty}^{\infty} F_N(u-s)F_W(v-s)dF_S(s). \quad (III.E.32)$$

Από αυτές τις κατανομές ο λόγος πιθανοφάνειας για βέλτιστη ανίχνευση στο (III.E.30) μπορεί να επιτευχθεί αν οι F_N, F_W , και F_S είναι γνωστές.

Ας υποθέσουμε, για παράδειγμα, ότι οι F_N, F_W , και F_S είναι όλες κατά Gauss κατανομές με μηδενικά μέσα, $\text{Var}(N_k) = \text{Var}(W_k) = \sigma^2$ και $\text{Var}(S_k) = \sigma_S^2$. Τότε, η Q_0 είναι η διδιάστατη κατά Gauss πυκνότητα και με τα δύο μηδενικά μέσα, με τις δύο διακυμάνσεις σ^2 και με μηδενικό συντελεστή συσχέτισης. Και η Q_1 είναι η διδιάστατη κατά Gauss πυκνότητα

και με τα δύο μηδενικά μέσα, με τις δύο διακυμάνσεις $\sigma^2/(1-\rho)$ και συντελεστή συσχέτισης ρ , όπου $\rho \triangleq \sigma_S^2/(\sigma^2 + \sigma_S^2)$. Για γνωστό σ^2 , ένας α -επιπέδου UMP έλεγχος γι' αυτό το κατά Gauss πρόβλημα δίνεται από έναν ανιχνευτή ενέργειας της μορφής

$$\delta_{ED}(\underline{y}) = \begin{cases} 1 & \text{if } \sum_{k=1}^n (u_k + v_k)^2 > \tau, \\ \gamma & \text{if } \sum_{k=1}^n (u_k + v_k)^2 = \tau, \\ 0 & \text{if } \sum_{k=1}^n (u_k + v_k)^2 < \tau, \end{cases} \quad (\text{III.E.33})$$

όπου το τ επιλέγεται για ένα μέγεθος α . Η ανάλυση αυτού του ελέγχου είναι ουσιαστικά όμοια με αυτή του ραδιόμετρου μονού καναλιού της Ενότητας III.B.

Αν, όπως συνήθως συμβαίνει στην πράξη, οι κατανομές F_N , F_W , και F_S είναι όλες άγνωστες, τότε θα πρέπει να αναζητηθεί μια εναλλακτική του βέλτιστου ανιχνευτή, όπως αυτή του (III.E.33). Ένας τέτοιος ανιχνευτής που χρησιμοποιείται ευρέως στην πράξη είναι ο *συσχετιστής πολικότητας – σύμπτωσης (ή ταύτισης) (PCC)*, που δίνεται από

$$\delta_{PCC}(\underline{y}) = \begin{cases} 1 & \text{if } \sum_{k=1}^n u(u_k v_k) > \tau, \\ \gamma & \text{if } \sum_{k=1}^n u(u_k v_k) = \tau, \\ 0 & \text{if } \sum_{k=1}^n u(u_k v_k) < \tau, \end{cases} \quad (\text{III.E.34})$$

όπου το u και πάλι συμβολίζει τη συνάρτηση μοναδιαίου βήματος

$$u(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases} \quad (\text{III.E.35})$$

Ας σημειωθεί ότι αυτός ο ανιχνευτής παίρνει την απόφασή του με βάση τον αριθμό των «συμπτώσεων πολικότητας» στις παρατηρήσεις. Δηλ., η PCC απόφαση βασίζεται στον αριθμό των φορών που οι έξοδοι και των δύο καναλιών έχουν το ίδιο σήμα. Καθώς το σήμα είναι κοινό και στα δύο κανάλια ως προς το H_1 και απόν και από τα δύο κανάλια ως προς το H_0 στο μοντέλο του (III.E.30), θα μπορούσε κανείς να περιμένει περισσότερες συμπτώσεις πολικότητας ως προς το H_1 απ' ό,τι ως προς το H_0 , και άρα αυτός είναι ένας λογικός τρόπος ανίχνευσης του σήματος εδώ.

Ως προς οποιαδήποτε από τις υποθέσεις του (III.E.30), η στατιστική PCC,

$t_{PCC}(\underline{Y}) \triangleq \sum_{k=1}^n u(U_k V_k)$ είναι το σύνολο των n i.i.d. Bernoulli τυχαίων μεταβλητών.

Αρα, ως προς το H_j , η $t_{PCC}(\underline{Y})$ είναι μια διωνυμική (n, λ_j) τυχαία μεταβλητή, όπου $\lambda_j = P_j(U_1 V_1 \geq 0)$, $j = 0, 1$. Μπορούμε να γράψουμε

$$\lambda_j = P_j(U_1 \geq 0, V_1 \geq 0) + P_j(U_1 \leq 0, V_1 \leq 0),$$

που ευθέως ανάγεται σε

$$\lambda_j = 1 - Q_j(0, \infty) - Q_j(\infty, 0) + 2Q_j(0, 0). \quad (\text{III.E.36})$$

Χρησιμοποιώντας τα (III.E.31) και (III.E.32) έχουμε ότι

$$\begin{aligned}\lambda_0 &= 1 - F_N(0) - F_W(0) + 2F_N(0) \\ &= \frac{1}{2} + 2[F_N(0) - \frac{1}{2}][F_W(0) - \frac{1}{2}]\end{aligned}\quad (\text{III.E.37})$$

και

$$\begin{aligned}\lambda_1 &= 1 - \int_{-\infty}^{\infty} F_N(-s)dF_S(s) - \int_{-\infty}^{\infty} F_W(-s)dF_S(s) \\ &\quad + 2 \int_{-\infty}^{\infty} F_N(-s)F_W(-s)F_S(s) \\ &= \frac{1}{2} + 2 \int_{-\infty}^{\infty} [F_N(s) - \frac{1}{2}][F_W(s) - \frac{1}{2}]dF_S(s).\end{aligned}\quad (\text{III.E.38})$$

Από το (III.E.36) βλέπουμε ότι αν η διεργασία θορύβου έχει μηδενικό διάμεσο (δηλ., αν $F_N(0) = \frac{1}{2}$ ή $F_W(0) = \frac{1}{2}$) τότε $\lambda_0 = \frac{1}{2}$. Άρα ο PCC είναι μη παραμετρικός ως προς την κλάση των μοντέλων θορύβου στα οποία τουλάχιστον μία από τις διεργασίες θορύβου έχει μηδενικό διάμεσο, αφού η πιθανότητα ψευδούς συναγερμού μπορεί να οριστεί στο α για την κλάση αυτή επιλέγοντας τ και γ για να δώσουν μέγεθος α , για t_{PCC} που είναι διωνυμικά κατανομημένο με παραμέτρους $(n, \frac{1}{2})$.

Είναι ενδιαφέρον να συγκρίνουμε τον PCC με τον βέλτιστο ανιχνευτή για κατά Gauss κανάλια που δίνονται από το (III.E.33). Υποθέτοντας ότι οι F_N και F_W έχουν συναρτήσεις πυκνότητας πιθανότητας f_N και f_W , αντίστοιχα και ότι οι f_N και f_W είναι συνεχείς στο μηδέν και έχουν πεπερασμένες δεύτερες και τέταρτες ροπές, μπορεί να αποδειχθεί ότι η ασυμπτωτική αποδοτικότητα του δ_{PCC} σε σχέση με το δ_{ED} για το ζεύγος υπόθεσης του (III.E.30) μπορεί να δοθεί από

$$\text{ARE}_{PCC,ED} = [\gamma_N^4 + \gamma_W^4 + 4\sigma_N^2\sigma_W^2 - \sigma_N^2 - \sigma_W^4]f_N^2(0)f_W^2(0),$$

όπου

$$\sigma_N^2 \triangleq E\{N_1^2\}, \sigma_W^2 \triangleq E\{W_1^2\}, \gamma_N^2 \triangleq E\{N_1^4\}, \text{ and } \gamma_W^4 = E\{W_1^4\}.$$

Για παράδειγμα, για όμοια κατά Gauss κανάλια έχουμε

$$N_1 \sim \mathcal{N}(0, \sigma^2) \quad \text{and} \quad W_1 \sim \mathcal{N}(0, \sigma^2),$$

έχουμε

$$\sigma_N^2 = \sigma_W^2 = \sigma^2,$$

το οποίο δίνει,

$$\text{ARE}_{PCC,ED} = \frac{2}{\pi^2} = 0.202.$$

Εναλλακτικά, για πανομοιότυπα Λαπλασιανά κανάλια με

$$f_{N_1}(x) = f_{W_1}(x) = \frac{\alpha}{2} e^{-\alpha|x|}$$

έχουμε $\sigma_N^2 = \sigma_W^2 = 2/\alpha^2$ και $\gamma_N^4 = \gamma_W^4 = 24/\alpha^2$, εκ των οποίων

$$\text{ARE}_{PCC,ED} = 3.5.$$

Άρα, ο PCC αποδίδει σχετικά ανεπαρκώς σε σύγκριση με ανιχνευτές ισχύος για κατά Gauss κανάλια, αλλά ο PCC αποδίδει αρκετά καλά για τα Λαπλασιανά κανάλια. Η ανεπαρκής αποδοτικότητα για τα κατά Gauss κανάλια είναι το τίμημα που πληρώνεται εδώ προς χάριν της απλότητας πραγματοποίησης και της μη παραμετρικής απόδοσης που επιδεικνύει ο PCC. Όπως στην περίπτωση του μονού καναλιού, η βελτιωμένη αποδοτικότητα μπορεί να επιτευχθεί για τα προβλήματα διπλού ή πολλαπλού δείγματος με παράλληλη διατήρηση της μη παραμετρικής απόδοσης με την εξέταση πιο πολύπλοκων ανιχνευτών που βασίζονται στις βαθμίδες [βλ. για παράδειγμα, Carlyle (1968)].

Οι υποθέσεις του (III.E.30) περιγράφουν ένα συνήθη τύπο προβλήματος δύο καναλιών. Ένας άλλος τύπος προβλήματος δύο καναλιών που προκύπτει συχνά είναι το πρόβλημα ανίχνευσης ενός σήματος με αναφορά σε πηγή θορύβου. Αυτό το πρόβλημα περιγράφεται από το ζεύγος υπόθεσης

$$H_0 : \begin{array}{l} U_k = N_k \\ V_k = W_k \end{array}, k = 1, 2, \dots, n$$

ως προς

(III.E.39)

$$H_1 : \begin{array}{l} U_k = N_k \\ V_k = W_k + S_k \end{array}, k = 1, 2, \dots, n$$

όπου οι $\{N_k\}_{k=1}^n$ και $\{W_k\}_{k=1}^n$ είναι ανεξάρτητες i.i.d. ακολουθίες θορύβου με το ίδιο όριο και η $\{S_k\}_{k=1}^n$ είναι μια ακολουθία σήματος, ανεξάρτητη από το θόρυβο. Άρα το πρόβλημα αυτό αντιστοιχεί στο σύνθετο πρόβλημα μονού δείγματος στο κανάλι V, με την πρόσθετη παρατήρηση ενός δείγματος θορύβου μόνο στο κανάλι U. Ας σημειωθεί ότι και τα δύο σύνολα δειγμάτων μπορεί να ληφθούν από ένα μονό κανάλι με τα δείγματα μόνο θορύβου να λαμβάνονται σε στιγμή κατά την οποία είναι γνωστό ότι δεν υπάρχει κανένα

σήμα παρόν και να αποθηκεύονται για ύστερη διεργασία με τα πρόσθετα δείγματα. [Για λόγους απλότητας, υποθέσαμε στο (III.E.39) ότι λαμβάνεται ο ίδιος αριθμός δειγμάτων από κάθε κανάλι, παρόλο που αυτό δεν συμβαίνει απαραίτητα στην πράξη.]

Υπάρχει ένας αριθμός ελέγχων που μπορούν να εφαρμοστούν για την ανίχνευση σημάτων με παρατηρήσεις της μορφής του (III.E.39). Πρώτα βλέπουμε ότι αν οι στατιστικές του θορύβου είναι γνωστές με ακρίβεια, τότε οι πληροφορίες που παρέχονται από το μόνο θορύβου κανάλι είναι άχρηστες υπό τις δεδομένες παραδοχές ανεξαρτησίας. Αυτό είναι εμφανές από το λόγο πιθανοφάνειας, από τον οποίο όλοι οι όροι που αφορούν στα U_k εξαφανίζονται. Άρα, το κανάλι χωρίς σήμα μπορεί είναι χρήσιμο μόνο όταν υπάρχει κάτι άγνωστο σχετικά με την κατανομή θορύβου.

Ένας έλεγχος που είναι χρήσιμος για το (III.E.39) όταν το σήμα είναι μια θετική σταθερά, είναι η διπλού δείγματος εκδοχή του ελέγχου Wilcoxon, που βασίζεται στη στατιστική

$$\sum_{k=1}^n r_k, \quad (\text{III.E.40})$$

όπου η r_k είναι η βαθμίδα του v_k όταν όλες οι παρατηρήσεις u_1, \dots, u_n και v_1, \dots, v_n συγκεντρώνονται και τίθενται σε αύξουσα σειρά. Ο έλεγχος αυτός είναι γνωστός και ως έλεγχος Mann – Whitney. Μια στατιστική ισοδύναμη με αυτή του (III.E.40) είναι

$$\sum_{k=1}^n \sum_{j=1}^n u(v_k - u_j),$$

από την οποία είναι προφανές ότι ο ανιχνευτής Mann – Whitney είναι μη παραμετρικός για συμμετρικά κατανομημένο θόρυβο. Όπως με τον μονού καναλιού ανιχνευτή Wilcoxon, αυτή η εκδοχή δύο καναλιών αποδίδει αρκετά καλά σε σύγκριση με άλλους ανιχνευτές για το (III.E.39).

Ο έλεγχος Mann – Whitney είναι ένας από τους πολλούς χρήσιμους ελέγχους για το (III.E.39) που εμπίπτουν στην κατηγορία των ελέγχων Kolmogorof – Smirnov, που βασίζεται σε συναρτησιακά που υπολογίζονται στη συνάρτηση

$$\hat{F}_V(x) - \hat{F}_U(x), \quad -\infty < x < \infty,$$

όπου οι \hat{F}_V και \hat{F}_U είναι οι λεγόμενες εμπειρικές συναρτήσεις κατανομής των δειγμάτων $\{u_k\}_{k=1}^n$ και $\{v_k\}_{k=1}^n$, αντίστοιχα. Δηλ., η \hat{F}_V ορίζεται ως

$$\hat{F}_V(x) = \frac{1}{n} \sum_{k=1}^n u(x - v_k), \quad -\infty < x < \infty, \quad (\text{III.E.41})$$

και η \hat{F}_U ορίζεται ανάλογα.¹⁵ Καθώς $n \rightarrow \infty$, οι \hat{F}_U και \hat{F}_V συγκλίνουν στις αντίστοιχες οριακές κατανομές των U_k και V_k αντίστοιχα και άρα οι συναρτήσεις της διαφοράς

$(\hat{F}_V - \hat{F}_U)$ μπορεί να είναι χρήσιμες στην απόφαση αν οι $\{U_k\}_{k=1}^n$ και $\{V_k\}_{k=1}^n$ έχουν την ίδια κατανομή ή όχι [δηλ. η $(\hat{F}_V - \hat{F}_U)$ είναι χρήσιμη για τον έλεγχο της ομοιογένειας]. Εκτός από τον έλεγχο Mann – Whitney, άλλοι χρήσιμοι έλεγχοι αυτού του τύπου είναι αυτοί που βασίζονται στη σύγκριση της στατιστικής

$$\sup_{-\infty < x < \infty} [\hat{F}_V(x) - \hat{F}_U(x)],$$

ή

$$\sup_{-\infty < x < \infty} |\hat{F}_V(x) - \hat{F}_U(x)|$$

μ' ένα κατώφλι. Οι έλεγχοι αυτοί είναι μη παραμετρικοί για την υπόθεση ότι οι $\{U_k\}_{k=1}^n$ και $\{V_k\}_{k=1}^n$ έχουν την ίδια οριακή κατανομή [δηλ. για H_0 στο (III.E.39)] και μπορεί να ξεπεράσει σε επιδόσεις του ελέγχου Mann – Whitney για κάποια πεδία της πιθανότητας ψευδούς συναγερωμού.

Για περαιτέρω συζήτηση και λεπτομέρειες της θεωρίας και των εφαρμογών της μη παραμετρικής ανίχνευσης, ο αναγνώστης παραπέμπεται στο βιβλίο των Kassam and Thomas (1980).

14

III.E.2 ΑΝΘΕΚΤΙΚΗ ΑΝΙΧΝΕΥΣΗ

¹⁴ 15. Σημειώστε ότι για κάθε x , το $\hat{F}_V(x)$ είναι απλώς ο αριθμός των u_k που είναι μικρότερα ή ίσα με το x διαιρεμένα με το συνολικό αριθμό των u_k .

Στην Ενότητα III.B. πραγματευτήκαμε το σχεδιασμό των συστημάτων ανίχνευσης υπό την παραδοχή ότι είναι διαθέσιμη μια πλήρης στατιστική περιγραφή των δεδομένων παρατήρησης. Εναλλακτικά, στις παραπάνω παραγράφους, εξετάσαμε συστήματα ανίχνευσης για καταστάσεις στις οποίες πολύ λίγα είναι γνωστά για τις στατιστικές παρατήρησης. Μεταξύ αυτών των δύο άκρων υπάρχει η κατάσταση στην οποία είναι διαθέσιμο ένα σχετικά ακριβές ονομαστικό μοντέλο για τα στατιστικά δεδομένα, αλλά στην οποία μπορεί να προκύψουν κάποιες μικρές αποκλίσεις από το μοντέλο αυτό.

Εξετάστε, για παράδειγμα, το πρόβλημα ελέγχου μεταξύ δύο πιθανών οριακών κατανομών, P_0 και P_1 , για μια i.i.d. ακολουθία Y_1, \dots, Y_n . Υποθέτοντας πυκνότητες p_0 και p_1 , οι βέλτιστοι έλεγχοι βασίζονται στο λόγο πιθανοφάνειας

$$L(\underline{y}) = \prod_{k=1}^n \frac{p_1(y_k)}{p_0(y_k)}. \quad (\text{III.E.42})$$

Ας σημειωθεί ότι ο λόγος πιθανοφάνειας είναι πολύ ευαίσθητος σε παρατηρήσεις για τις οποίες είτε $p_1(y_k) \gg p_0(y_k)$ είτε $p_0(y_k) \gg p_1(y_k)$. Καθώς η συνθήκη $p_1(y_k) \gg p_0(y_k)$ είναι πολύ πιο πιθανό να προκύψει ως προς το H_1 παρά ως προς το H_0 (και αντίστροφα για τη συνθήκη $p_0(y_k) \gg p_1(y_k)$), αυτή η ευαισθησία είναι απλώς τμήμα της προτιθέμενης δράσης του ελέγχου ως προς το εικαζόμενο μοντέλο.

Ας υποθέσουμε, ωστόσο, ότι η πραγματική οριακή κατανομή των δεδομένων δεν είναι ακριβώς P_0 ή P_1 , αλλά είναι μάλλον μόνο κατά προσέγγιση P_0 ή P_1 . Για παράδειγμα, ας υποθέσουμε ότι η πραγματική κατανομή είναι της μορφής

$$(1 - \epsilon)P_j + \epsilon M_j, \quad j = 0, 1, \quad (\text{III.E.43})$$

όπου P_0 και P_1 είναι οι ονομαστικές κατανομές, οι M_0 και M_1 είναι άγνωστες και αυθαίρετα «μολύνουσες» κατανομές και ϵ είναι ένας αριθμός μεταξύ του 0 και του 1 που αναπαριστά το βαθμό αβεβαιότητας που πρέπει να τεθεί στο μοντέλο. Ένα τέτοιο μοντέλο μπορεί να προκύψει, για παράδειγμα, σε ένα κανάλι επικοινωνιών ή ραντάρ στο οποίο είναι παρών ένας εξωτερικός παρεμβολέας για ένα κλάσμα ϵ του χρόνου ή στο οποίο ο παλμικός θόρυβος (κεραυνός κλπ.) προκύπτει με πιθανότητα ϵ . Επίσης, τα διαλείποντα σφάλματα αισθητήρα και άλλες μετρήσεις σφαλμάτων καταγραφής δεδομένων μπορούν να μοντελοποιηθούν με τον τρόπο αυτό. Η βασική ιδέα εδώ είναι ότι τα M_j αναπαριστούν μια πλευρά του καθορισμένου μοντέλου που είναι απολύτως άγνωστη στο σχεδιαστή.

Ας υποθέσουμε ότι η $p_1(y_k)/p_0(y_k)$ είναι μια μη φραγμένη συνάρτηση του y_k . Αφού το M_0 είναι αυθαίρετο, θα μπορούσε να θέσει όλες τις πιθανότητες του σε περιοχές όπου

$p_1(y_k) \gg p_0(y_k)$. Αυτό θα μπορούσε να προκαλέσει τον έλεγχο που βασίζεται στον ονομαστικό λόγο πιθανοφάνειας (III.E.42) για να κάνει ψευδείς συναγερούς (δηλ., σφάλματα ως προς το H_0) πιο συχνά απ' ό,τι θα έπρεπε. Όπως θα δούμε παρακάτω, αυτό θα έκανε την πιθανότητα ψευδούς συναγερού (και συνολικά την πιθανότητα σφάλματος) σχεδόν στην τάξη του $1 - (1 - \epsilon)^n$, που αυξάνει με n και ισούται με ϵ για $n = 1$. Καθώς πολλά συστήματα ανίχνευσης είναι σχεδιασμένα να λειτουργούν σε πιθανότητες ψευδούς συναγερού σε μια έκταση από το 10^{-4} ως το 10^{-6} , ακόμη και 1% αβεβαιότητας σ' αυτό το μοντέλο (δηλ., $\epsilon = 0,01$) θα μπορούσε εν δυνάμει να καταστρέψει την επίδοση ψευδούς συναγερού του ανιχνευτή. Ένα παρόμοιο φαινόμενο θα μπορούσε να προκύψει ως προς το H_1 , αν η $p_1(y_k)/p_0(y_k)$ δεν είναι φραγμένη μακριά από το μηδέν, καθώς το M_1 μπορεί να θέσει την πιθανότητα σε περιοχές παρατήρησης όπου $p_1(y_k) \ll p_0(y_k)$, τείνοντας έτσι να οδηγήσει το λόγο πιθανοφάνειας κάτω από το κατώφλι.

Η παραπάνω συζήτηση, αν και ευρητική, επισημαίνει μια σχετική έλλειψη ανθεκτικότητας στην επίδοση των ελέγχων λόγου πιθανοφάνειας σε καταστάσεις όπου η $p_1(y_k)/p_0(y_k)$ δεν είναι φραγμένη από πάνω και (μακριά από το μηδέν) από κάτω. Συγκεκριμένα, βλέπουμε ότι ακόμη και σχετικά μικρές αποκλίσεις από το μοντέλο μπορεί να έχουν ως αποτέλεσμα σημαντική απώλεια επίδοσης στην κατάσταση αυτή. Ένα ερώτημα που προκύπτει είναι αν μπορεί να γίνει οτιδήποτε λογικό για να αμβλυνθεί αυτή η απώλεια ανθεκτικότητας. Ένας σχετικά προφανής πιθανός τρόπος σταθεροποίησης της επίδοσης του ελέγχου λόγου πιθανοφάνειας είναι να αντικατασταθεί ο λόγος πιθανοφάνειας p_1/p_0 με μια εκδοχή που είναι περιορισμένη από πάνω και από κάτω. Δηλ., έστω ότι αντικαθιστούμε το $l \triangleq p_1/p_0$ στο γινόμενο $\prod_{k=1}^n p_1(y_k)/p_0(y_k)$ με τη συνάρτηση

$$[l]_a^b(y) = \begin{cases} b & \text{if } l(y) > b \\ l(y) & \text{if } a \leq l(y) \leq b \\ a & \text{if } l(y) < a, \end{cases} \quad (\text{III.E.44})$$

όπου $0 < a < b < \infty$. Τότε για κατάλληλα επιλεγμένα a και b , ένας έλεγχος βασισμένος στο $\prod_{k=1}^n [l]_a^b(y_k)$ δε θα εμφάνιζε τις δυσκολίες που παρατηρήθηκαν παραπάνω για μη φραγμένο l . Ωστόσο, εισάγοντας τον περιορισμό του (III.E.44), χάνεται κάποια επίδοση (και σίγουρα βελτιστότητα) στο ονομαστικό, καθώς αυτές οι παρατηρήσεις για τις οποίες η $l(y)$ είναι πολύ μεγάλη ή πολύ κοντά στο μηδέν είναι, κατά μία έννοια, αυτές που δίνουν τις περισσότερες πληροφορίες.

Έχει ενδιαφέρον να βρεθεί ένα κριτήριο μέσω του οποίου οι έλεγχοι μπορούν να βελτιστοποιηθούν στην επίδοση υπό την αβεβαιότητα μοντελοποίησης. Η παραπάνω συζήτηση που αφορά στην έλλειψη ανθεκτικότητας του ελέγχου λόγου πιθανοφάνειας με μη φραγμένο λόγο πιθανοφάνειας οδηγεί σ' ένα τέτοιο κριτήριο. Συγκεκριμένα, αυτή η συζήτηση επικεντρώθηκε στην χειρότερη περίπτωση επίδοσης του ελέγχου στην κλάση των στατιστικών που είναι πιθανόν να υπάρξουν υπό το μοντέλο αβεβαιότητας. Άρα ένα λογικό

κριτήριο σχεδιασμού θα ήταν να αντικατασταθούν οι συνήθεις πιθανότητες σφάλματος με τις τιμές σε κάποια λογική γειτονιά του ονομαστική μοντέλου [όπως αυτό του (III.E.43)] και να επιλεγεί ένας έλεγχος για τη βελτιστοποίηση του αντίστοιχου κριτηρίου

Θυμηθείτε ότι αν P_0 είναι η αληθής οριακή κατανομή του Y_k , η πιθανότητα ψευδούς συναγερμού ενός ελέγχου δ είναι

$$P_F(\delta, P_0) = \int_{\Gamma} \delta(\underline{y}) \left[\prod_{k=1}^n p_0(y_k) \right] \mu(d\underline{y}), \quad (\text{III.E.45})$$

όπου p_0 είναι η οριακή πυκνότητα που αντιστοιχεί στο P_0 . Παρομοίως, η χαμένη πιθανότητα, όταν η P_1 είναι η αληθής οριακή, είναι

$$P_M(\delta, P_1) = \int_{\Gamma} [1 - \delta(\underline{y})] \left[\prod_{k=1}^n p_1(y_k) \right] \mu(d\underline{y}). \quad (\text{III.E.46})$$

Υποθέτοντας για λόγους απλότητας ότι τα κόστη είναι ομοιόμορφα, τα τρία συνήθη κριτήρια για απλό έλεγχο δυαδικής υπόθεσης είναι:

- (i) $\min_{\delta} [\pi_0 P_F(\delta, P_0) + \pi_1 P_M(\delta, P_1)]$ (Bayes).
- (ii) $\min_{\delta} [\max\{P_F(\delta, P_0), P_M(\delta, P_1)\}]$ (Minimax).
- (iii) $\min_{\delta} P_M(\delta, P_1)$ subject to $P_F(\delta, P_0) \leq \alpha$ (Neyman-Pearson).

- (i) Μπαϊεσιανό
- (ii) (Ελαχιστομέγιστο)
- (iii) (Neyman – Pearson).

Αν αντί να υποθέσουμε ότι η οριακή κατανομή του Y_k είναι ακριβώς P_0 ή P_1 , υποθέσουμε ότι το όριο βρίσκεται είτε σε μια περιοχή \mathcal{P}_0 του P_0 είτε σε μια περιοχή \mathcal{P}_1 του P_1 (όπως π.χ. στο (III.E.43), τότε αντικαθιστώντας το $P_F(\delta, P_0)$ και το $P_M(\delta, P_1)$ στο (i) – (iii) με τις τιμές των χειρότερων περιπτώσεών τους

$$P_F(\delta, \mathcal{P}_0) \triangleq \sup_{P \in \mathcal{P}_0} P_F(\delta, P) \quad (\text{III.E.47})$$

και

$$P_M(\delta, \mathcal{P}_1) \triangleq \sup_{P \in \mathcal{P}_1} P_M(\delta, P), \quad (\text{III.E.48})$$

καταλήγουμε στα προβλήματα εναλλακτικού σχεδιασμού

- (i') $\min_{\delta} [\pi_0 P_F(\delta, \mathcal{P}_0) + \pi_1 P_M(\delta, \mathcal{P}_1)].$
- (ii') $\min_{\delta} [\max\{P_F(\delta, \mathcal{P}_0), P_M(\delta, \mathcal{P}_1)\}].$
- (iii') $\min_{\delta} P_M(\delta, \mathcal{P}_1)$ subject to $P_F(\delta, \mathcal{P}_0) \leq \alpha.$

Οι λύσεις στο (i') – (iii') θα έχουν την καλύτερη απόδοση χειρότερης περίπτωσης (στις περιοχές \mathcal{P}_0 και \mathcal{P}_1) όλων των πιθανών ελέγχων. Φυσικά, υπάρχει ο κίνδυνος ότι τέτοιοι έλεγχοι μπορεί να είναι υπερβολικά συντηρητικοί και αυτό σίγουρα θα ήταν αληθές αν οι \mathcal{P}_0 και \mathcal{P}_1 ήταν πολύ μεγάλες. Ωστόσο, η ιδέα εδώ είναι ότι οι \mathcal{P}_0 και \mathcal{P}_1 είναι μικρές περιοχές ενός ονομαστικού μοντέλου και ο στόχος είναι να αποφευχθεί η πιθανή αβεβαιότητα απόδοσης για τέτοιες περιοχές που αναφέρθηκαν παραπάνω. Αποδεικνύεται ότι οι λύσεις των (i') – (iii') επιτυγχάνουν αυτόν το στόχο για μοντέλα αβεβαιότητας όπως αυτό του (III.E.43).

Υπάρχει μια γενική προσέγγιση στην επίλυση προβλημάτων (i') – (iii') των Huber και Strassen (1973) η οποία έχει κάποια σχέση που δεν θα συζητήσουμε εδώ. Αντ' αυτού, θα εστιάσουμε στις λύσεις των (i') – (iii') για τις περιοχές ιδιαίτερης αβεβαιότητας που περιγράφονται στο (III.E.43), που είναι γνωστές ως ϵ -μολυσμένα μίγματα. Η λύση αυτή απαντά στο Huber (1965), στον οποίο η διατύπωση (i') – (iii') προτάθηκε για πρώτη φορά ως μια τεχνική σχεδιασμού για ανθεκτικούς ελέγχους.

Αποδεικνύεται ότι οι λύσεις των (i')-(iii') για ϵ -μολυσμένα μίγματα είναι οι αντίστοιχοι βέλτιστοι έλεγχοι για τα (i)-(iii), όταν τα \mathcal{P}_0 και \mathcal{P}_1 στα (i)-(iii) αντικαθίστανται από ένα ζεύγος $Q_0 \in \mathcal{P}_0$ και $Q_1 \in \mathcal{P}_1$ των λιγότερο ευνοϊκών κατανομών. Τα Q_0 και Q_1 δίνονται σε σχέση με τις πυκνότητές τους από

$$q_0(y_k) = \begin{cases} (1 - \epsilon)p_0(y_k) & \text{if } p_1(y_k) < c''p_0(y_k) \\ \frac{1-\epsilon}{c''}p_1(y_k) & \text{if } p_1(y_k) \geq c''p_0(y_k) \end{cases} \quad (\text{III.E.49a})$$

Και

$$q_1(y_k) = \begin{cases} (1 - \epsilon)p_1(y_k) & \text{if } p_1(y_k) > c'p_0(y_k) \\ c'(1 - \epsilon)p_0(y_k) & \text{if } p_1(y_k) \leq c'p_0(y_k), \end{cases} \quad (\text{III.E.49b})$$

όπου $0 < c' < 1 < c'' < \infty$ είναι δύο σταθερές που επιλέγονται ώστε τα Q_0 και Q_1 να είναι πιθανές κατανομές, ώστε, δηλαδή η συνολική τους πιθανότητα να ισούται με 1. Αυτή η συνθήκη δίνεται από

$$(1 - \epsilon)[P_0(l(Y_k) < c'') + P_1(l(Y_k) \geq c'')/c''] = 1 \quad (\text{III.E.50a})$$

Και

$$(1 - \epsilon)[P_1(l(Y_k) > c') + c'P_0(l(Y_k) \leq c')] = 1. \quad (\text{III.E.50b})$$

Καθώς κάθε ένα από τα (i) – (iii) λύνεται από έναν έλεγχο λόγου πιθανοφάνειας, οι λύσεις στα (i') – (iii') είναι έλεγχοι λόγου πιθανοφάνειας μεταξύ των Q_0 και Q_1 . Κυρίως, βασίζονται στο λόγο πιθανοφάνειας $\prod_{k=1}^n q_1(y_k)/q_0(y_k)$. Χρησιμοποιώντας το (III.E.49a) έχουμε ότι

$$\frac{q_1(y_k)}{q_0(y_k)} = \begin{cases} c' & \text{if } l(y_k) < c' \\ l(y_k) & \text{if } c' \leq l(y_k) \leq c'' \\ c'' & \text{if } l(y_k) > c''. \end{cases} \quad (\text{III.E.51})$$

Άρα οι λύσεις των (i') – (iii') είναι έλεγχοι κατωφλιού που βασίζονται στο

$$\prod_{k=1}^n [l]_{c'}^{c''}(y_k), \quad (\text{III.E.52})$$

και έχουμε φτάσει αναλυτικά σ' έναν έλεγχο του τύπου που προτάθηκε νωρίτερα ως ένας ανθεκτικός έλεγχος. Ο έλεγχος αυτός, ωστόσο, έχει μια συγκεκριμένη ιδιότητα βελτιστότητας στην επίλυση των (i') – (iii'). Επιπλέον, εκτός από την επίλυση των (i') – (iii'), μπορεί να αποδειχθεί ότι

$$P_F(\delta_R, P_0) = P_F(\delta_R, Q_0) \quad (\text{III.E.53a})$$

και

$$P_M(\delta_R, \mathcal{P}_1) = P_M(\delta_R, Q_1) \quad (\text{III.E.53b})$$

για κάθε έλεγχο κατώφλιού δ_R που βασίζεται στο λόγο πιθανοφάνειας (III.E.52). Αυτό συνεπάγεται ότι η απόδοση χειρότερης περίπτωσης του δ_R είναι στην ουσία η απόδοσή της στο ζεύγος κατανομών (Q_0, Q_1) για το οποίο είναι βέλτιστο. Αυτό μας επιτρέπει να υπολογίσουμε τα άνω φράγματα στις πιθανότητες σφάλματος του δ_R για ολόκληρες τις τάξεις \mathcal{P}_0 και \mathcal{P}_1 , υπολογίζοντας απλώς τις πιθανότητες σφάλματος του δ_R για οριακά Q_0 και Q_1 .

Παρατηρήσεις

- Κατά μία έννοια, οι λιγότερο ευνοϊκές πυκνότητες Q_0 και Q_1 είναι τόσο κοντά σε σχήμα μεταξύ τους όσο είναι δυνατόν στο πλαίσιο των περιορισμών $Q_0 \in \mathcal{P}_0$ και $Q_1 \in \mathcal{P}_1$. Μπορεί να αποδειχθεί ότι αν $\mathcal{P}_0 \neq \mathcal{P}_1$ και το ϵ είναι αρκετά μικρό (και θετικό), τότε οι ισότητες του (III.E.50a) έχουν λύσεις που ικανοποιούν το $0 < c' < 1 < c'' < \infty$. Αυτό συνεπάγεται ότι $Q_0 \neq Q_1$ και ότι ο δ_R δεν είναι ένας τετριμμένος έλεγχος. Από την άλλη, αν το ϵ είναι πολύ μεγάλο, τότε οι \mathcal{P}_0 και \mathcal{P}_1 θα επικαλύπτονται, $c' = c'' = 1$ και το Q_0 θα ισούται με το Q_1 . Στην περίπτωση αυτή, $q_1(y_k)/q_0(y_k) \equiv 1$ και οι έλεγχοι που επιλύουν τα (i')-(iii') απλώς αγνοούν τις παρατηρήσεις και εικάζουν στην υπόθεση. Άρα, αν οι περιοχές \mathcal{P}_0 και \mathcal{P}_1 είναι πολύ μεγάλες για να επικαλυφθούν, τα (i')-(iii') δεν είναι καλά κριτήρια σχεδιασμού. Ωστόσο, όπως σημειώθηκε παραπάνω, χρησιμοποιούνται στην ουσία για μικρό ϵ .
- Ακόμη και για πολύ μικρό ϵ , η διαφορά απόδοσης μεταξύ των ελέγχων που βασίζονται στο l και $|l|_{c''}$ μπορεί να είναι αρκετά σημαντική. Εξετάστε, για παράδειγμα, τη Μπαϊεσιανή διατύπωση (i) με $\pi_0 = \pi_1 = 1/2$. Τότε, για οποιονδήποτε από τους δύο ελέγχους, το κατώφλι είναι μονάδα και η τυχαιοποίηση είναι αυθαίρετη. Έστω ότι $0 < l(y_k) < \infty$ και $\sup_{y_k \in \mathbf{R}} l(y_k) = \infty$. Τότε, όπως σημειώθηκε παραπάνω, αφού το M_0 είναι αυθαίρετο, μπορεί να επιλεγθεί να τεθεί όλη του η πιθανότητα σε μια τιμή του y_k για το οποίο $l(y_k)$ είναι αυθαίρετα μεγάλη. Με τον τρόπο αυτό, οποιαδήποτε από τις παρατηρήσεις μπορεί να προκαλέσει το δ_0 να πραγματοποιήσει έναν ψευδή συναγερμό με πιθανότητα ϵ , όπου το δ_0 είναι ο έλεγχος που βασίζεται στο l . Καθώς υπάρχουν n παρατηρήσεις, αυτό συνεπάγεται ότι

$$P_F(\delta_0, \mathcal{P}_0) \geq 1 - (1 - \epsilon)^n. \quad (\text{III.E.54})$$

Παρομοίως, αν $\inf_{y_k \in \mathbf{R}} l(y_k) = 0$, η πιθανότητα απώλειας του δ_0 στο \mathcal{P}_1 ικανοποιεί

$$P_M(\delta_0, \mathcal{P}_1) \geq 1 - (1 - \epsilon)^n, \quad (\text{III.E.55})$$

όπου το supremum (ελάχιστο άνω φράγμα) αναλαμβάνει τις \mathcal{P}_0 και \mathcal{P}_1 . Για κάθε $\epsilon > 0$, $\lim_{n \rightarrow \infty} (1 - \epsilon)^n = 0$. Άρα,

$$\lim_{n \rightarrow \infty} [\sup P_\epsilon(\delta_0)] = 1; \quad (\text{III.E.56})$$

άρα η απόδοση του δ_0 μπορεί να είναι αυθαίρετα κακή, και στην ουσία μπορεί να είναι χειρότερη από την εικασία της υπόθεσης, καθώς η εικασία θα προκαλούσε μια πιθανότητα σφάλματος του 1/2.

Αντίθετα, εξετάστε τον έλεγχο δ_R που βασίζεται στο q_1/q_0 . Από την ιδιότητα του (III.E.50a), έπεται ότι

$$\sup P_\epsilon(\delta_R) = \frac{1}{2}P_F(\delta_R, Q_0) + \frac{1}{2}P_M(\delta_R, Q_1). \quad (\text{III.E.57})$$

Εφαρμόζοντας το φράγμα Chernoff (III.C.18) στο δεξιό μέλος του (III.E.57), έχουμε ότι

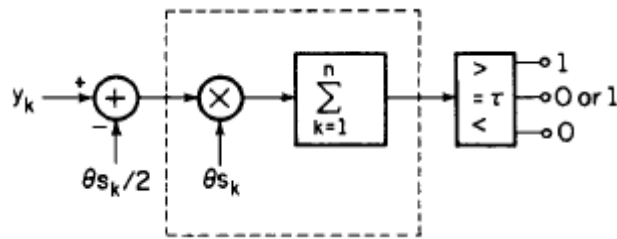
$$\sup P_\epsilon(\delta_R) \leq \frac{1}{2} \left[\int [q_0 q_1]^{1/2} \right]^n. \quad (\text{III.E.58})$$

Όπως σημειώθηκε στην Ενότητα III.C, αν $Q_0 \neq Q_1$, θα έχουμε $\int [q_0 q_1]^{1/2} < 1$, ώστε

$$\lim_{n \rightarrow \infty} [\sup P_\epsilon(\delta_R)] = 0. \quad (\text{III.E.59})$$

Από τα (III.E.56) και (III.E.59), βλέπουμε ότι σε ό,τι αφορά την επίδοση χειρότερης περίπτωσης, η $P_\epsilon(\delta_0)$ συγκλίνει εκθετικά στη μονάδα, ενώ η $P_\epsilon(\delta_R)$ συγκλίνει εκθετικά στο μηδέν. Αν και η συμπεριφορά που οδηγεί το $\sup P_\epsilon(\delta_0)$ στη μονάδα είναι κάπως ακραία, ωστόσο καταλήγει σε έναν περιορισμό τέτοιων ελέγχων. Θα πρέπει να σημειωθεί ότι υπό ονομαστικές συνθήκες, η $P_\epsilon(\delta_0)$ επίσης συγκλίνει εκθετικά στο μηδέν μέσω του φράγματος Chernoff. Άρα, η δ_R επιτυγχάνει, στην χειρότερή της περίπτωση, συμπεριφορά παρόμοια με την ονομαστική συμπεριφορά της δ_0 , ενώ η δ_0 στην χειρότερή της περίπτωση επιτυγχάνει συμπεριφορά εντελώς διαφορετική.

3. Οι λύσεις των (i')-(iii') είναι γνωστές για μια σειρά μοντέλων αβεβαιότητας πέρα από το ϵ μολυσμένο μίγμα. (Το ίδιο το ϵ μολυσμένο μίγμα μπορεί να γενικευθεί ελαφρώς για να επιτρέψει διαφορετικές τιμές του ϵ ως προς κάθε υπόθεση). Για λόγους απλότητας, επιλέγουμε ότι είναι ίσες). Για παράδειγμα, διάφοροι ενδιαφέροντες τύποι περιοχών που μπορούν να αντιμετωπιστούν σε αυτό το πλαίσιο είναι της μορφής $\mathcal{P}_0 = \{P|\rho(P, P_0) \leq \epsilon_0\}$ και $\mathcal{P}_1 = \{P|\rho(P, P_1) \leq \epsilon_1\}$, όπου ρ είναι κάποια μέτρηση της απόστασης μεταξύ των κατανομών πιθανότητας. Στις περισσότερες περιπτώσεις, άλλες γνωστές λύσεις των (i')-(iii') είναι παρόμοιες με αυτές που αναλύθηκαν εδώ για την ϵ μολυσμένη περίπτωση στο ότι είναι λύσεις των (i)-(iii) για λιγότερο ευνοϊκά ζεύγη και στο ότι συνήθως περιλαμβάνουν κάποια μορφή περιορισμού εύρους του λόγου πιθανοφάνειας. Το ϵ μολυσμένο μοντέλο μπορεί επίσης να γενικευθεί για να καταστούν δυνατά τα χρονικά μεταβαλλόμενα ονομαστικά και τα ϵ . Στην περίπτωση αυτή το $\prod_{k=1}^n [l_k]_{c'_k}^{c''_k}(y_k)$ απλώς αντικαθίσταται από $\prod_{k=1}^n [l_k]_{c'_k}^{c''_k}(y_k)$ όπου l_k είναι ονομαστικός λόγος πιθανοφάνειας για το k th δείγμα και τα c'_k και c''_k επιλύουν το (III.E.50a) για τις k th ονομαστικές κατανομές.



Συσχετιστής

ΣΧΗΜΑ III.E.1. Ονομαστικά βέλτιστος ανιχνευτής για συνεκτικά σήματα σε κατά Gauss θόρυβο

Περιγράφουμε την εφαρμογή των αποτελεσμάτων που συζητήθηκαν στην παραπάνω παράγραφο με το ακόλουθο παράδειγμα.

Παράδειγμα III.E.1: Ο περιοριστής συσχετιστή

Εξετάστε το πρόβλημα ανίχνευσης συνεκτικού σήματος.

$$H_0 : Y_k = N_k, \quad k = 1, \dots, n$$

ως προς

$$(III.E.60)$$

$$H_1 : Y_k = N_k + \theta s_k, \quad k = 1, \dots, n,$$

όπου η s_1, \dots, s_n είναι μια γνωστή ακολουθία σήματος. Η N_1, \dots, N_n είναι μια i.i.d. ακολουθία των $N(0,1)$ δειγμάτων θορύβων και θ είναι ένα γνωστό θετικό εύρος. Ο λόγος πιθανοφάνειας του k th - δείγματος για το πρόβλημα αυτό δίνεται [βλέπε (III.B.9)] από $l_k(y_k) = \exp[\theta s_k(y_k - \theta s_k/2)]$, και άρα ο λόγος λογαριθμικής πιθανοφάνειας είναι

$$\log L(\underline{y}) = \theta \sum_{k=1}^n s_k(y_k - \theta s_k/2), \quad (\text{III.E.61})$$

που οδηγεί στον ανιχνευτή συσχέτισης που απεικονίζεται στο Σχήμα III.E.1.

Στην πράξη, το μοντέλο του (III.E.60) μπορεί μόνο να υποθεθεί ότι είναι κατά προσέγγιση σωστό. Συγκεκριμένα, η κατανομή είναι απίθανο να είναι ακριβώς $N(0,1)$ και το μοντέλο μέτρησης στο οποίο το σήμα και ο θόρυβος θεωρούνται ότι είναι προσθετικά δεν είναι απολύτως ακριβές λόγω των μη γραμμικοτήτων στο μηχανισμό παρατήρησης. Άρα, για να έχουμε ένα πιο ρεαλιστικό μοντέλο, θα μπορούσαμε να μετασχηματίσουμε το (III.E.60) σε

$$H_0 : Y_k \sim (1 - \epsilon)P_0 + \epsilon M_0, \quad k = 1, \dots, n$$

ως προς

$$(\text{III.E.62})$$

$$H_1 : Y_k \sim (1 - \epsilon)P_1^{(k)} + \epsilon M_1^{(k)}, \quad k = 1, \dots, n,$$

όπου P_0 είναι η $N(0,1)$ κατανομή, $P_1^{(k)}$ είναι η $\mathcal{N}(\theta s_k, 1)$ κατανομή, M_0 και $M_1^{(k)}, k = 1, \dots, n$ είναι αυθαίρετα (σημείωση: το M_0 θα μπορούσε επίσης να αλλάξει με το k . Αυτό δεν θα άλλαζε τη λύση που δίνεται παρακάτω) και το ϵ βρίσκεται μεταξύ του 0 και του 1.

Ας σημειωθεί ότι με $l_k(y_k) = \exp[\theta s_k(y_k - \theta s_k/2)]$, έχουμε

$0 < l_k(y_k) < \infty$, $\sup_{y_k \in \mathbf{R}} l_k(y_k) = \infty$, και $\inf_{y_k \in \mathbf{R}} l_k(y_k) = 0$, οπότε ο ανιχνευτής συσχέτισης θα υποστεί τη μείωση απόδοσης που αναφέραμε παραπάνω με την παρουσία αβεβαιότητας, μοντελοποιημένης όπως στο (III.E.62). Άρα, ο ανθεκτικός έλεγχος λόγου

πιθανοφάνειας του Huber δ_R είναι προτιμητέος στην κατάσταση αυτή. Για το μοντέλο του (III.E.62) ο λογάριθμος του ανθεκτικού λόγου πιθανοφάνειας

$$L_R(\underline{y}) = \prod_{k=1}^n [l_k]_{c'_k}^{c''_k}(y_k),$$

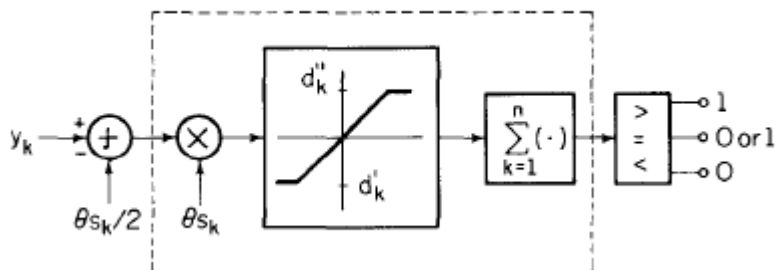
μπορεί να γραφεί ως

$$\log L_R(\underline{y}) = \sum_{k=1}^n [\theta s_k (y_k - \theta s_k / 2)]_{d'_k}^{d''_k}, \quad (\text{III.E.63})$$

όπου $d'_k \triangleq \log c'_k$ και $d''_k \triangleq \log c''_k$. Χρησιμοποιώντας τις ιδιότητες συμμετρίας της $N(0,1)$ κατανομής και του λόγου πιθανοφάνειας μπορεί να αποδειχθεί ότι $d'_k = -d''_k$ και ότι d''_k είναι η λύση στο

$$\Phi \left(\frac{d''_k}{\theta |s_k|} + \frac{\theta |s_k|}{2} \right) + e^{-d''_k} \left[1 + \Phi \left(\frac{d''_k}{\theta |s_k|} - \frac{\theta |s_k|}{2} \right) \right] = (1 - \epsilon)^{-1}, \quad (\text{III.E.64})$$

όπου Φ είναι η $N(0,1)$ αθροιστική συνάρτηση κατανομής. Το d''_k μειώνεται μονότονα από το ∞ στο μηδέν καθώς το ϵ αυξάνεται από το μηδέν.



Συσχετιστής

ΣΧΗΜΑ III.E.2. Ανθεκτικός ανιχνευτής για συνεκτικά σήματα σε ονομαστικό κατά Gauss θόρυβο.

Ο ανιχνευτής του (III.E.63) απεικονίζεται στο Σχ. III.E.2. Ας σημειωθεί ότι είναι πανομοιότυπος με το δέκτη συσχέτισμού στο Σχ. III.E.1, εκτός του ότι υπάρχει ένας (χρονικά μεταβαλλόμενος) περιοριστής με όρια $\pm d_k''$ τοποθετημένος μεταξύ του πολλαπλασιαστή και του συσσωρευτή στο συσχετιστή. Αυτή η δομή ανιχνευτή ονομάζεται περιοριστής συσχετιστή και αρχικά δημιουργήθηκε και αναλύθηκε από τους Martin και Schwartz (1971). Θα είναι ανθεκτικός με την προϋπόθεση ότι το ϵ είναι αρκετά μικρό ώστε να αποφευχθεί η επικάλυψη των κλάσεων ως προς τα H_0 και H_1 . Για σταθερό ϵ , αυτό αντιστοιχεί στο ότι η ισχύ σήματος θ είναι αρκετά μεγάλη ώστε να αποφευχθεί η επικάλυψη. Συγκεκριμένα, αν $|s_k| > 0$ για $k = 1, \dots, n$ οι κλάσεις αβεβαιότητας ως προς τα H_0 και H_1 δεν επικαλύπτονται για κανένα από τα k αν $\theta > \theta_\epsilon / \min |s_k|$, όπου θ_ϵ είναι η λύση στο $(1 - \epsilon)\Phi(\theta_\epsilon/2) = 1/2$. Για παράδειγμα, αν $\epsilon = 0.138$, τότε $\theta_\epsilon = 0.03$, αν $\epsilon = 0.055$, τότε $\theta_\epsilon = 0.15$ και αν $\epsilon = 0.138$ τότε $\theta_\epsilon = 0.4$. Άρα η ανθεκτικότητα διατηρείται για σχετικά μικρούς λόγους σήματος προς θόρυβο. Θα πρέπει να σημειωθεί ότι σε αντίθεση με το γραμμικό συσχετιστή, ο περιοριστής συσχετιστή δεν είναι ομοιόμορφα βέλτιστος στο θ ακόμη και για $\theta > \theta_\epsilon$ αφού το d_k'' εξαρτάται από το θ . Αυτό είναι ένα μειονέκτημα για πολλές εφαρμογές, αφού το θ δεν είναι απαραίτητα γνωστό. Έχουν αναπτυχθεί διάφορες μέθοδοι χειρισμού αυτής της δυσκολίας και των προβλημάτων ανθεκτικότητα μικρών θ . Αυτές και άλλες πλευρές της ανθεκτικής ανίχνευσης αναλύονται σε ένα άρθρο ανασκόπησης των Kassam και Poor (1985)

ΚΕΦΑΛΑΙΟ IV

ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΩΝ

IV. ΕΙΣΑΓΩΓΗ

Στα κεφάλαια II και III εξετάσαμε το σχεδιασμό βέλτιστων διεργασιών για την απόφαση μεταξύ δύο πιθανών στατιστικών καταστάσεων με βάση μια τυχαία παρατήρηση Y . Σε πολλές καταστάσεις που προκύπτουν στην πράξη, ενδιαφερόμαστε όχι τόσο να επιλέξουμε μεταξύ δύο (ή περισσότερων) διακριτών καταστάσεων, αλλά μάλλον να επιλέξουμε μεταξύ ενός συνεχούς πιθανών φυσικών καταστάσεων. Συγκεκριμένα, όπως στο πρόβλημα ελέγχου σύνθετης υπόθεσης που αναλύσαμε στο Κεφάλαιο II, μπορούμε να σκεφτούμε μια οικογένεια κατανομών στο χώρο παρατήρησης με δείκτη μία παράμετρο ή ένα σύνολο παραμέτρων. Αλλά σε αντίθεση με την περίπτωση ελέγχου σύνθετης υπόθεσης, στην οποία επιθυμούμε να λάβουμε μια δυαδική απόφαση για την παράμετρο, εδώ επιθυμούμε να προσδιορίσουμε όσο το δυνατόν ακριβέστερα την πραγματική τιμή της παραμέτρου από την παρατήρηση.

Αυτού του είδους τα προβλήματα είναι γνωστά ως *προβλήματα εκτίμησης παραμέτρου* και στο κεφάλαιο αυτό αναλύουμε τις βασικές ιδέες που συνδέονται με το σχεδιασμό των βέλτιστων διεργασιών για την εκτίμηση παραμέτρων. Όπως με το πρόβλημα ελέγχου υπόθεσης (το οποίο παρεμπιπτόντως μπορεί να θεωρηθεί ως μια ειδική περίπτωση του προβλήματος εκτίμησης παραμέτρου), μπορούν να χρησιμοποιηθούν διάφορες φιλοσοφίες σχεδιασμού εκτίμησης, που διαφοροποιούνται κυρίως στην ποσότητα πρότερων γνωστών πληροφοριών σχετικά με την παράμετρο και τα κριτήρια απόδοσης που εφαρμόστηκαν.

Στο παρόν κεφάλαιο, αναλύουμε δύο βασικές προσεγγίσεις της εκτίμησης παραμέτρου – μία τη Μπαϊεσιανή, στην οποία η παράμετρος θεωρείται ότι είναι ένα τυχαίο μέγεθος που συνδέεται στατιστικά με την παράμετρο, και μία δεύτερη στην οποία η παράμετρος θεωρείται ότι είναι άγνωστη αλλά χωρίς να συνδέεται με κάποια πιθανοθεωρητική δομή. Από τις δύο αυτές προσεγγίσεις, η Μπαϊεσιανή είναι η πιο άμεση και έτσι εξετάζεται πρώτη, στην Ενότητα IV.B, με την εκτίμηση μη τυχαίας παραμέτρου να εξετάζεται στο υπόλοιπο του κεφαλαίου.

Θα πρέπει να σημειωθεί ότι σ' αυτή τη μελέτη εξετάζουμε μόνο την εκτίμηση παραμέτρων που είναι στατικές δηλ. διαρκείς στο χρόνο. Η εκτίμηση δυναμικών παραμέτρων (δηλ. των σημάτων) εξετάζεται στο Κεφάλαιο V.

IV.B Μπαϊεσιανή εκτίμηση παραμέτρου

Σε όλο αυτό το κεφάλαιο, θεωρούμε ως μοντέλο μια οικογένεια κατανομών για την τυχαία παρατήρηση Y , με δείκτη μια παράμετρο θ που λαμβάνει τιμές σ' ένα σύνολο παραμέτρου Λ . Δηλ. έχουμε την οικογένεια $\{P_\theta; \theta \in \Lambda\}$, όπου το P_θ συμβολίζει μια κατανομή στο χώρο παρατήρησης (Γ, \mathcal{G}) . Επίσης υποθέτουμε ότι το σύνολο παραμέτρου Λ είναι ένα υποσύνολο του \mathbf{R}^m για κάποιο m . Στο πλαίσιο του μοντέλου αυτού, ο στόχος του προβλήματος εκτίμησης παραμέτρου είναι να βρει μια συνάρτηση $\hat{\theta}: \Gamma \rightarrow \Lambda$ τέτοια ώστε η $\hat{\theta}(y)$ να είναι η «καλύτερη» εικασία της αληθινής τιμής του θ (δηλ. η τιμή του θ για την οποία $Y \sim P_\theta$) με βάση την παρατήρηση $Y = y$.

Φυσικά, η λύση σ' αυτό το πρόβλημα εξαρτάται από το κριτήριο ποιότητας με το οποίο μετράμε την απόδοση εκτίμηση. Έτσι, όπως στο πρόβλημα ελέγχου υπόθεσης, ξεκινάμε αποδίδοντας κόστη στις αποφάσεις μας σχετικά με την παράμετρο. Συγκεκριμένα, υποθέτουμε ότι υπάρχει μια συνάρτηση $C: \Lambda \times \Lambda \rightarrow \mathbb{R}$ τέτοια που το $C[a, \theta]$ είναι το κόστος εκτίμησης μιας αληθινής τιμής του θ ως a , για a και θ στο Λ . Με δεδομένη μια τέτοια συνάρτηση C μπορούμε κατόπιν να συσχετίσουμε, μ' έναν εκτιμητή $\hat{\theta}$ την κατά συνθήκη διακινδύνευσης ή το μέσο κόστος στο Y για κάθε $\theta \in \Lambda$. Δηλ. έχουμε

$$R_{\theta}(\hat{\theta}) = E_{\theta}\{C[\hat{\theta}(Y), \theta]\}. \quad (\text{IV.B.1})$$

Αν τώρα υιοθετήσουμε την ερμηνεία ότι η πραγματική αξία παραμέτρου θ είναι η πραγματοποίηση μιας τυχαίας μεταβλητής Θ , μπορούμε να προσδιορίσουμε μια μέση ή Μπαϊεσιανή διακινδύνευση ως

$$r(\hat{\theta}) \triangleq E\{R_{\Theta}(\hat{\theta})\}, \quad (\text{IV.B.2})$$

και ο κατάλληλος στόχος σχεδιασμού είναι να βρεθεί μια εκτιμητρια συνάρτηση που ελαχιστοποιεί το $r(\hat{\theta})$. Μια τέτοια εκτιμητρια συνάρτηση είναι γνωστή ως Μπαϊεσιανή εκτίμηση του θ .

Παρατηρώντας ότι $R_{\theta}(\theta) = E\{C[\theta(Y), \Theta] | \Theta = \theta\}$, έχουμε

$$r(\hat{\theta}) = E\{C[\hat{\theta}(Y), \Theta]\} = E\{E\{C[\hat{\theta}(Y), \Theta] | Y\}\}. \quad (\text{IV.B.3})$$

Ανατρέχοντας στο (IV.B.3) βλέπουμε ότι η Μπαϊεσιανή εκτίμηση του θ μπορεί να βρεθεί (εάν υπάρχει) ελαχιστοποιώντας, για κάθε $y \in \Gamma$, το ύστερο κόστος δοθέντος ότι $Y = y$:

$$E\{C[\hat{\theta}(y), \Theta] | Y = y\}. \quad (\text{IV.B.4})$$

Αυτή είναι η ίδια διαδικασία όπως αυτή που ακολουθήθηκε στο πρόβλημα Μπαϊεσιανού ελέγχου υπόθεσης (βλέπε Ενότητα Π.Ε.) Ας σημειωθεί ότι αν υποθέσουμε ότι το Θ έχει μια κατά συνθήκη πυκνότητα $w(\theta|y)$ δοθέντος ότι $Y = y$ για κάθε $y \in \Gamma$, τότε η Μπαϊεσιανή εκτίμηση $\hat{\theta}(y)$ που αντιστοιχεί στο $y \in \Gamma$ μπορεί να αναζητηθεί ελαχιστοποιώντας

$$\int_{\Lambda} C[\hat{\theta}(y), \theta] w(\theta|y) \mu(d\theta). \quad (\text{IV.B.5})$$

Οι παρακάτω περιπτώσεις απεικονίζουν την εφαρμογή αυτού του κριτηρίου

Περίπτωση IV. B.1: Εκτίμηση Ελάχιστου Μέσου Τετραγωνικού Σφάλματος (MMSE).

Για καταστάσεις στις οποίες $\Lambda = \mathbb{R}$ and $E\{\Theta^2\} < \infty$, μια κοινώς χρησιμοποιούμενη συνάρτηση κόστους είναι αυτή που δίνεται από

$$C[a, \theta] = (a - \theta)^2, \quad (a, \theta) \in \mathbb{R}^2. \quad (\text{IV.B.6})$$

Αυτή η συνάρτηση κόστους είναι φυσική για πολλές καταστάσεις, καθώς μετρά την απόδοση ενός εκτιμητή ως προς το τετράγωνο του σφάλματος εκτίμησης $\hat{\theta}(y) - \theta$. Ο Μπαϊεσιανός κίνδυνος εδώ είναι $E\{(\hat{\theta}(Y) - \Theta)^2\}$, ένα μέγεθος γνωστό ως μέσο τετραγωνικό σφάλμα (MSE). Άρα η Μπαϊεσιανή εκτίμηση στην περίπτωση αυτή είναι ένας εκτιμητής ελάχιστου μέσου τετραγωνικού σφάλματος (MMSE).

Το ύστερο κόστος δοθέντος ότι $Y = y$ δίνεται, στην περίπτωση αυτή, από

$$\begin{aligned} E\{(\hat{\theta}(y) - \Theta)^2 | Y = y\} &= E\{[\hat{\theta}(y)]^2 | Y = y\} \\ &\quad - 2E\{\hat{\theta}(y)\Theta | Y = y\} \\ &\quad + E\{\Theta^2 | Y = y\} \quad (\text{IV.B.7}) \\ &= [\hat{\theta}(y)]^2 - 2\hat{\theta}(y)E\{\Theta | Y = y\} \\ &\quad + E\{\Theta^2 | Y = y\}. \end{aligned}$$

Η παράσταση στο (IV.B.7) είναι μια τετραγωνική συνάρτηση του $\hat{\theta}(y)$, άρα επιτυγχάνει το μοναδικό της ελάχιστο στο σημείο όπου η παραγωγός της ως προς το $\hat{\theta}(y)$ είναι μηδέν. Διαφορίζοντας το (IV.B.7) έχουμε ότι η Μπαϊεσιανή εκτίμηση, που συμβολίζεται με $\hat{\theta}_{MMSE}$ δίνεται από

$$\hat{\theta}_{MMSE}(y) = E\{\Theta|Y = y\}. \quad (IV.B.8)$$

Άρα η MMSE εκτίμηση του Θ δοθέντος ότι $Y = y$ είναι το κατά συνθήκη μέσο του Θ δοθέντος ότι $Y = y$. Αυτό είναι πολύ βασικό αποτέλεσμα στο οποίο θα επιστρέψουμε σε ακόλουθα κεφάλαια. Αυτή η εκτίμηση καλείται ενίοτε *δεσμευμένη μέση εκτίμηση* (CME).

Περίπτωση IV. B. 2: Εκτίμηση ελάχιστου μέσου απόλυτου σφάλματος (MMAE)

Μια άλλη συνάρτηση κόστους που ενίοτε εφαρμόζεται στην περίπτωση του $\Lambda = \mathbb{R}$ είναι αυτή του απόλυτου σφάλματος που δίνεται από

$$C[a, \theta] = |a - \theta|, \quad (a, \theta) \in \mathbb{R}^2. \quad (IV.B.9)$$

Η Μπαϊεσιανή διακινδύνευση εδώ είναι $E\{|\hat{\theta}(Y) - \Theta|\}$, ένα μέγεθος γνωστό ως *μέσο απόλυτο σφάλμα*, άρα η αντίστοιχη Μπαϊεσιανή εκτίμηση είναι γνωστή ως *εκτίμηση ελάχιστου μέσου απόλυτου σφάλματος* (MMAE).

Για να εξάγουμε την MMAE εκτίμηση κάνουμε χρήση του δεδομένου ότι αν X είναι μια τυχαία μεταβλητή με $P(X \geq 0) = 1$, τότε $E\{X\} = \int_0^{\infty} P(X > x) dx$. Αυτό το αποτέλεσμα έπεται κυρίως με την ολοκλήρωση κατά μέρη [βλέπε, π.χ. Breiman (1968)]

Αφού $|\hat{\theta}(y) - \Theta| \geq 0$, από το παραπάνω αποτέλεσμα έχουμε ότι

$$\begin{aligned}
E\{|\hat{\theta}(y) - \Theta| | Y = y\} &= \int_0^{\infty} P(|\hat{\theta}(y) - \Theta| > x | Y = y) dx \\
&= \int_0^{\infty} P(\Theta > x + \hat{\theta}(y) | Y = y) dx \\
&\quad + \int_0^{\infty} P(\Theta < -x + \hat{\theta}(y) | Y = y) dx.
\end{aligned}
\tag{IV.B.10}$$

Αντικαθιστώντας το $t = x + \hat{\theta}(y)$ στο πρώτο ολοκλήρωμα και το $t = -x + \hat{\theta}(y)$ στο δεύτερο ολοκλήρωμα στα δεξιά του (IV.B.10), έχουμε

$$\begin{aligned}
E\{|\hat{\theta}(y) - \Theta| | Y = y\} &= \int_{\hat{\theta}(y)}^{\infty} P(\Theta > t | Y = y) dt \\
&\quad + \int_{-\infty}^{\hat{\theta}(y)} P(\Theta < t | Y = y) dt.
\end{aligned}
\tag{IV.B.11}$$

Με $E\{|\hat{\theta}(y) - \Theta| | Y = y\}$ στη μορφή του (IV.B.11) βλέπουμε ότι είναι μια διαφορίσιμη συνάρτηση του $\hat{\theta}(y)$. Με διαφορισμό έχουμε

$$\begin{aligned}
\frac{\partial}{\partial \hat{\theta}(y)} E\{|\hat{\theta}(y) - \Theta| | Y = y\} &= P(\Theta < \hat{\theta}(y) | Y = y) \\
&\quad - P(\Theta > \hat{\theta}(y) | Y = y).
\end{aligned}
\tag{IV.B.12}$$

Από το (IV.B.12) παρατηρούμε ότι αυτή η παράγωγος είναι μη φθίνουσα συνάρτηση του $\hat{\theta}(y)$ που προσεγγίζει το -1 καθώς $\hat{\theta}(y) \rightarrow -\infty$ και το +1 καθώς $\hat{\theta}(y) \rightarrow +\infty$. Άρα $E\{|\hat{\theta}(y) - \Theta| | Y = y\}$ επιτυγχάνει το ελάχιστό του ως προς το $\hat{\theta}(y)$ στο σημείο (ή στο σύνολο σημείων) όπου η παράγωγός του αλλάζει πρόσημο. Αυτό σημαίνει ότι η Μπαϊεσιανή εκτίμηση σ' αυτή την περίπτωση, που συμβολίζεται με $\hat{\theta}_{ABS}(y)$, είναι κάθε σημείο τέτοιο που

$$P(\Theta < t | Y = y) \leq P(\Theta > t | Y = y), \quad t < \hat{\theta}_{ABS}(y)$$

και

$$P(\Theta < t|Y = y) \geq P(\Theta > t|Y = y), \quad t > \hat{\theta}_{ABS}(y). \quad (\text{IV.B.13})$$

Ας σημειωθεί ότι ένα σημείο $\hat{\theta}_{ABS}(y)$ που ικανοποιεί το (IV.B.13) είναι μια διάμεσος της δεσμευμένης κατανομής του Θ , δοθέντος ότι $Y = y$. Άρα η MMAE εκτίμηση είναι μια *εκτίμηση δεσμευμένης διαμέσου*. Η εκτίμηση αυτή συμπίπτει με τη MMSE εκτίμηση μόνο όταν η κατανομή του Θ δοθέντος ότι $Y = y$ έχει την ίδια τιμή για το μέσο και τη διάμεσό της.

Περίπτωση IV.B.3. Εκτίμηση Μέγιστης εκ των υστέρων (A posteriori) πιθανότητας (MAP)

Μια άλλη μέθοδος εκτίμησης που, αν και δεν είναι απολύτως μπαϊεσιανή, ταιριάζει στο Μπαϊεσιανό πλαίσιο είναι η εκτίμησης μέγιστης εκ των υστέρων πιθανότητας (MAP).

Για να εφαρμόσουμε τη μέθοδο αυτή, υποθέσουμε την περίπτωση $\Lambda = \mathbb{R}$ και εξετάζουμε τη λεγόμενη *συνάρτηση ομοιόμορφου κόστους*,

$$C[a, \theta] = \begin{cases} 0 & \text{if } |a - \theta| \leq \Delta \\ 1 & \text{if } |a - \theta| > \Delta, \end{cases} \quad (\text{IV.B.14})$$

όπου $\Delta > 0$. Για έναν εκτιμητή $\hat{\theta}$ το μέσο ύστερο κόστος, δοθέντος ότι $Y = y$ στην περίπτωση αυτή δίνεται από

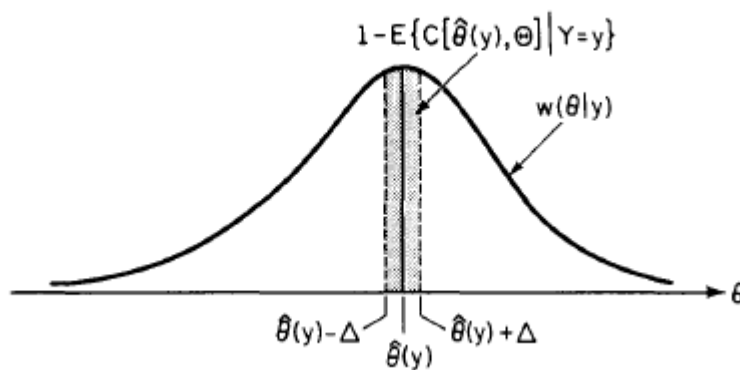
$$\begin{aligned} E\{C[\hat{\theta}(y), \Theta]|Y = y\} &= P(|\hat{\theta}(y) - \Theta| > \Delta|Y = y) \\ &= 1 - P(|\hat{\theta}(y) - \Theta| \leq \Delta|Y = y). \end{aligned} \quad (\text{IV.B.15})$$

Για να εξετάσουμε την ελαχιστοποίηση του (IV.B.15) υποθέτουμε πρώτα ότι Θ είναι μια διακριτή τυχαία μεταβλητή που λαμβάνει τιμές σ' ένα πεπερασμένο σύνολο

$\Lambda = \{\theta_0, \dots, \theta_{M-1}\}$ με $|\theta_i - \theta_j| > \Delta$ για $i \neq j$. Τότε έχουμε

$$\begin{aligned}
E\{C[\hat{\theta}(y), \Theta] | Y = y\} &= 1 - P(\Theta = \hat{\theta}(y) | Y = y) \\
&= 1 - w(\hat{\theta}(y) | y) \quad \text{for } \hat{\theta}(y) \in \Lambda,
\end{aligned}
\tag{IV.B.16}$$

όπου $w(\theta|y)$ είναι η συνάρτηση δεσμευμένης πυκνότητας πιθανότητας του Θ δοθέντος ότι $Y = y$. Από το (IV.B.16) βλέπουμε ότι η Μπαϊεσιανή εκτίμηση στην περίπτωση αυτή δίνεται για κάθε $y \in \Gamma$ από κάθε τιμή του θ που μεγιστοποιεί το $w(\theta|y)$ ως προς το $\theta \in \Lambda$. Αυτό σημαίνει ότι η Μπαϊεσιανή εκτίμηση είναι η τιμή του Θ που έχει την μέγιστη εκ των υστέρων πιθανότητα να προκύψει δοθέντος ότι $Y = y$.¹



ΣΧΗΜΑ IV.B.1 Απεικόνιση της MAP εκτίμησης

Τώρα ας υποθέσουμε ότι $\Lambda = \mathbb{R}$ και ότι Θ είναι μια συνεχής τυχαία μεταβλητή με συνάρτηση δεσμευμένη πυκνότητας $w(\theta|y)$ δοθέντος ότι $Y = y$. Στην περίπτωση αυτή, η ύστερη διακινδύνευση γίνεται

$$E\{C[\hat{\theta}(y), \Theta] | Y = y\} = 1 - \int_{\hat{\theta}(y) - \Delta}^{\hat{\theta}(y) + \Delta} w(\theta|y) d\theta. \tag{IV.B.17}$$

Το μέγεθος το (IV. B.17) ελαχιστοποιείται στο $\hat{\theta}(y)$ με τη μεγιστοποίηση της περιοχής ως προς η $w(\theta|y)$ στο διάστημα $(\hat{\theta}(y) - \Delta, \hat{\theta}(y) + \Delta)$. Ανατρέχοντας στο Σχ. IV.B.1 βλέπουμε ότι αν η $w(\theta|y)$ είναι μια λεία συνάρτηση του θ και αν το Δ είναι επαρκώς μικρό, αυτή η περιοχή θα είναι κατά προσέγγιση μεγιστοποιημένη επιλέγοντας το $\hat{\theta}(y)$ να είναι ένα μέγιστο σημείο της $w(\theta|y)$. Αυτό σημαίνει ότι για μικρό Δ και λεία $w(\theta|y)$, έχουμε

$$\int_{\hat{\theta}(\mathbf{y})-\Delta}^{\hat{\theta}(\mathbf{y})+\Delta} w(\theta|\mathbf{y})d\theta \cong 2\Delta w(\theta|\mathbf{y})|_{\theta=\hat{\theta}(\mathbf{y})}, \quad (\text{IV.B.18})$$

και η πλευρά του δεξιού μέλους μεγιστοποιείται επιλέγοντας το $\hat{\theta}(\mathbf{y})$ ώστε να είναι η τιμή του θ που μεγιστοποιεί τη $w(\theta|\mathbf{y})$ στο Λ .

Σε όποια από τις δύο παραπάνω περιπτώσεις το κριτήριο ομοιόμορφου κόστους οδηγεί στη διαδικασία εκτίμησης του Θ ως την τιμή που μεγιστοποιεί την εκ των υστέρων (διακριτή ή συνεχή) πυκνότητα $w(\theta|\mathbf{y})$. [Παρομοίως, με θ διακριτό αλλά λαμβάνοντας απείρως πολλές τιμές, μπορεί να υποστηριχθεί ότι το (IV.B.15) ελαχιστοποιείται κατά προσέγγιση,

επιλέγοντας το $\hat{\theta}(\mathbf{y})$ για τη μεγιστοποίηση της δεσμευμένης πυκνότητας πιθανότητας $w(\theta|\mathbf{y})$. Η εκτίμηση αυτή είναι γνωστή ως *εκτίμηση μέγιστης εκ των υστέρων πιθανότητας* (MAP) και

συμβολίζεται με $\hat{\theta}_{MAP}$. Παρόλο που η εκτίμηση αυτή συχνά προσεγγίζει την Μπαϊεσιανή εκτίμηση μόνο για ομοιόμορφο κόστος με μικρό Δ , το MAP κριτήριο χρησιμοποιείται ευρέως για το σχεδιασμό εκτιμήσεων. Ένας βασικός λόγος γι' αυτό είναι ότι οι MAP εκτιμήσεις είναι ευκολότερο να υπολογιστούν από τις MMSE, MMAE ή άλλες εκτιμήσεις.

Σημειώστε ότι ένα σημείο στο οποίο μια πυκνότητα επιτυγχάνει τη μέγιστή της τιμή καλείται *ταλάντωση* της αντίστοιχης κατανομής πιθανότητας. Άρα, αφού η $\hat{\theta}_{MAP}$ εκτιμά το Θ με την ταλάντωση της δεσμευμένης του κατανομής, είναι μια *εκτίμηση δεσμευμένης ταλάντωσης*.

Από τις περιπτώσεις IV.B.1 ως IV.B.3 [και από το (IV.B.5)] βλέπουμε ότι οι Μπαϊεσιανές εκτιμήσεις για μια δεδομένη κατάσταση ορίζονται από τη δεσμευμένη κατανομή της παραμέτρου δεδομένων των παρατηρήσεων. Συγκεκριμένα, οι MMSE, MMAE και MAP εκτιμήσεις είναι το μέσο, η διάμεσος και η ταλάντωση αυτής της κατανομής, αντίστοιχα. Όπως στην περίπτωση του ελέγχου υπόθεσης, μπορούμε να θεωρήσουμε την παρατήρηση ως ένα μέσο για τη μετατροπή της πρότερης κατανομής της παραμέτρου σε μια ύστερη κατανομή. Γενικά, οι Μπαϊεσιανοί εκτιμητές είναι χαρακτηριστικά αυτής της ύστερης κατανομής.

Στη μοντελοποίηση μιας δεδομένης στατιστικής κατάστασης, συνήθως ξεκινάμε με την οικογένεια $\{P_{\theta}; \theta \in \Lambda\}$ των δεσμευμένων κατανομών του Y δοθέντος ότι $\Theta = \theta$ και για τη Μπαϊεσιανή διατύπωση έχουμε επίσης μια πρότερη κατανομή για το Θ . Για να βρούμε τη δεσμευμένη κατανομή του Θ δοθέντος του Y από την πρότερη και την δεσμευμένη κατανομή του Y δοθέντος του Θ , χρειάζεται μόνο να εφαρμόσουμε το Μπαϊεσιανό τύπο. Συγκεκριμένα, υποθέτοντας ότι το P_{θ} έχει πυκνότητα p_{θ} για κάθε $\theta \in \Lambda$ και ότι η πρότερη κατανομή του Θ

1. Φυσικά, η περίπτωση αυτή στην οποία το Λ είναι ένα πεπερασμένο σύνολο είναι απλώς ένα πρόβλημα ελέγχου M -ary υπόθεσης και το κριτήριο κόστους του (IV.B.4) ανάγεται εδώ σε $C[\mathbf{a}, \theta] = 1$ αν $\mathbf{a} \neq \theta$ και $C[\mathbf{a}, \theta] = 0$ αν $\mathbf{a} = \theta$, αφού $|\theta_i - \theta_j| > \Delta$ για $i \neq j$. Άρα η Μπαϊεσιανή εκτίμηση στην περίπτωση αυτή είναι ο M -ary Μπαϊεσιανός κανόνας απόφασης για ομοιόμορφο κόστος (όπως στην Άσκηση 16 του Κεφαλαίου II.)

έχει πυκνότητα $w(\theta)$, έχουμε ότι η δεσμευμένη κατανομή του Θ δοθέντος ότι $Y = y$ έχει πυκνότητα

$$w(\theta|y) = \frac{p_\theta(y)w(\theta)}{\int_{\Lambda} p_\theta(y)w(\theta)\mu(d\theta)}. \quad (\text{IV.B.19})$$

Ας σημειωθεί ότι ο παρονομαστής του (IV.B.19) είναι $p(y)$, η αδέσμευτη πυκνότητα του Y .

Οι Μπαϊεσιανές εκτιμήσεις για τις τρεις παραπάνω περιπτώσεις μπορούν να βρεθούν ευθέως από το (IV.B.19). Σημειώστε ότι η MAP εκτίμηση μπορεί να βρεθεί χωρίς τον υπολογισμό του $p(y)$ καθώς ο όρος αυτός δε θα επηρεάσει τη μεγιστοποίηση στο θ . Δηλ. η $\hat{\theta}_{MAP}(y)$ βρίσκεται με τη μεγιστοποίηση του $p_\theta(y)w(\theta)$ στο $\theta \in \Lambda$. Καθώς ο λογάριθμος είναι μια αυξανόμενη συνάρτηση, η $\hat{\theta}_{MAP}(y)$ επίσης μεγιστοποιεί το $[\log p_\theta(y) + \log w(\theta)]$ στο $\theta \in \Lambda$. Αν το Θ είναι μια συνεχής τυχαία μεταβλητή δοθέντος ότι $Y = y$, για επαρκώς λεία p_θ και w , μια αναγκαία συνθήκη γι' αυτή τη μεγιστοποίηση είναι

$$\frac{\partial}{\partial \theta} \log p_\theta(y)|_{\theta=\hat{\theta}_{MAP}(y)} = -\frac{\partial}{\partial \theta} \log w(\theta)|_{\theta=\hat{\theta}_{MAP}(y)}. \quad (\text{IV.B.20})$$

Η εξίσωση του (IV.B.20) είναι γνωστή ως *MAP εξίσωση*. Τα ακόλουθα δύο παραδείγματα απεικονίζουν τον υπολογισμό των MMSE, MMAE και MAP εκτιμήσεων.

Παράδειγμα IV.B.1: Εκτίμηση της παραμέτρου μιας εκθετικής κατανομής

Εξετάστε την κατάσταση $\Lambda = (0, \infty)$ και $\Gamma = \mathbb{R}$, στην οποία οι παρατηρήσεις έχουν την ακόλουθη συνάρτηση δεσμευμένης πυκνότητας πιθανότητας, δοθέντος ότι $\Theta = \theta$:

$$p_\theta(y) = \begin{cases} \theta e^{-\theta y} & \text{if } y \geq 0 \\ 0 & \text{if } y < 0. \end{cases} \quad (\text{IV.B.21})$$

Αυτή είναι η *εκθετική πυκνότητα* με παράμετρο θ . Η εκθετική πυκνότητα μοντελοποιεί πολλά φυσικά φαινόμενα. Είναι ιδιαίτερα χρήσιμη στη μοντελοποίηση των χρονικών διαστημάτων μεταξύ διαδοχικών γεγονότων που προκύπτουν τυχαία στο χρόνο, όπως μηνύματα ή πακέτα δεδομένων που φτάνουν σ' ένα σταθμό μεταγωγής επικοινωνιών, οχήματα που φτάνουν σε μια διασταύρωση δρόμων, φωτόνια που εκπέμπουν από μια συνεκτική πηγή φωτός ή συσκευές που καταρρέουν σ' ένα λογικό κύκλωμα. Η παράμετρος θ σ' αυτό το μοντέλο μπορεί να ερμηνευθεί ως ο ρυθμός τέτοιων συμβάντων και άρα μπορούμε να θεωρήσουμε το

πρόβλημα εκτίμησης εδώ ως αυτό της εκτίμησης του ρυθμού εμφάνισης τέτοιων συμβάντων από μια παρατήρηση του χρόνου μεταξύ των διαδοχικών εμφανίσεών τους.

Ας υποθέσουμε ότι οι πρότερες πληροφορίες μας για το Θ είναι ότι επίσης έχει μια εκθετική κατανομή με πυκνότητα

$$w(\theta) = \begin{cases} \alpha e^{-\alpha\theta} & \text{if } \theta \geq 0 \\ 0 & \text{if } \theta < 0, \end{cases} \quad (\text{IV.B.22})$$

όπου $\alpha > 0$ είναι γνωστό. Μπορούμε τότε να βρούμε την ύστερη κατανομή του Θ δοθέντος ότι $Y = y$ από το (IV.B.19). Έχουμε

$$\begin{aligned} w(\theta|y) &= \frac{\alpha\theta e^{-(\alpha+y)\theta}}{\int_0^\infty \alpha\theta e^{-(\alpha+y)\theta} d\theta} \\ &= (\alpha+y)^2 \theta e^{-\theta(\alpha+y)}, \end{aligned} \quad (\text{IV.B.23})$$

για $\theta > 0$ και $y \geq 0$, και $w(\theta|y) = 0$

Η MMSE εκτίμηση είναι το μέσο του (IV.B.23) και άρα δίνεται από

$$\begin{aligned} \hat{\theta}_{\text{MMSE}}(y) &= \int_0^\infty \theta w(\theta|y) d\theta = (\alpha+y)^2 \int_0^\infty \theta^2 e^{-\theta(\alpha+y)} d\theta \\ &= \frac{2}{\alpha+y}. \end{aligned} \quad (\text{IV.B.24})$$

Ας σημειωθεί ότι για σταθερό α , αυτή η εκτίμηση του Θ μεταβάλλεται αντιστρόφως με το y . Αυτό είναι διαισθητικά λογικό από την προηγούμενη ερμηνεία του εκθετικού μοντέλου, καθώς ένας μεγάλος χρόνος μεταξύ αφίξεων (μεγάλο y) θα ήταν απόδειξη ενός χαμηλού ρυθμού (μικρό θ). Η συμπεριφορά αυτή μετριάζεται ως ένα βαθμό ανάλογα από την τιμή του α , καθώς η εκτίμηση δεν είναι ποτέ μεγαλύτερη από $2/\alpha$. Σημειώστε ότι μια μικρή τιμή του α συνεπάγεται ότι το Θ κατανέμεται διάχυτα [δηλ. το $w(\theta)$ είναι σχετικά εξαπλωμένο] και η αντίστοιχη εκτίμηση επιτρέπει μεγαλύτερες τιμές Θ αν συνεπάγεται από την παρατήρηση. Εναλλακτικά, μια μεγάλη τιμή του α συνεπάγεται ότι το Θ είναι κοντά στο μηδέν με υψηλή πιθανότητα, άρα η εκτίμηση δεν είναι ποτέ μεγάλη σ' αυτή την περίπτωση.

Η ελάχιστη τιμή της MSE μπορεί να υπολογιστεί ευθέως στην περίπτωση αυτή. Πρώτα, παρατηρούμε από το (IV.B.3) ότι η Μπαϊεσιανή διακινδύνευση είναι ο μέσος όρος του ύστερου κόστους έτσι ώστε

$$\begin{aligned} \text{MMSE} = r(\hat{\theta}_{\text{MMSE}}) &= E\{E\{(\hat{\theta}_{\text{MMSE}}(Y) - \Theta)^2|Y\}\} \\ &= E\{E\{(\Theta - E\{\Theta|Y\})^2|Y\}\} \\ &= E\{\text{Var}(\Theta|Y)\}. \end{aligned} \quad (\text{IV.B.25})$$

Άρα η ελάχιστη MSE είναι ο μέσος όρος της δεσμευμένης μεταβλητής του Θ , δοθέντος του Y . Αφού

$$\text{Var}(\Theta|Y = y) = E\{\Theta^2|Y = y\} - E^2\{\Theta|Y = y\},$$

έχουμε

$$\begin{aligned} \text{Var}(\Theta|Y = y) &= \int_0^\infty \theta^2 w(\theta|y) d\theta - [\hat{\theta}_{\text{MMSE}}(y)]^2 \\ &= (\alpha + y)^2 \int_0^\infty \theta^3 e^{-\theta(\alpha+y)} d\theta - \frac{4}{(\alpha + y)^2} \\ &= \frac{2}{(\alpha + y)^2}. \end{aligned}$$

Άρα

$$\begin{aligned} \text{MMSE} = E\left\{\frac{2}{(\alpha + Y)^2}\right\} &= \int_0^\infty \frac{2}{(\alpha + y)^2} p(y) dy \\ &= \int_0^\infty \frac{2\alpha}{(\alpha + y)^4} dy \\ &= \frac{2}{3\alpha^2}, \end{aligned} \quad (\text{IV.B.26})$$

όπου χρησιμοποιήσαμε το $p(y) = \int_0^\infty \alpha \theta e^{-(\alpha+y)\theta} d\theta = \alpha/(\alpha+y)^2$ όπως στο (IV.B.23).

Η MMAE εκτίμηση $\hat{\theta}_{ABS}(y)$ είναι η διάμεσος του $w(\theta|y)$. Αφού το Θ είναι συνεχές δοθέντος ότι $Y = y$, μπορούμε να βρούμε τη $\hat{\theta}_{ABS}(y)$ λύνοντας την εξίσωση

$$\int_{\hat{\theta}_{ABS}(y)}^{\infty} w(\theta|y) d\theta = \frac{1}{2}. \quad (\text{IV.B.27})$$

Ενθέτουμε το (IV.B.23) και ολοκληρώνουμε τα παράγωγα

$$[1 + (\alpha + y)\hat{\theta}_{ABS}(y)]e^{-(\alpha+y)\hat{\theta}_{ABS}(y)} = \frac{1}{2}, \quad (\text{IV.B.28})$$

έτσι ώστε να έχουμε

$$\hat{\theta}_{ABS}(y) = \frac{T_o}{\alpha + y}, \quad (\text{IV.B.29})$$

με το T_o να είναι η λύση στο $(1 + T_o)e^{-T_o} = 1/2$, που δίνεται από $T_o \cong 1.68$. Συγκρίνοντας το (IV.B.29) με το (IV.B. 24) βλέπουμε την ίδια συμπεριφορά με αυτής της MMSE εκτίμησης, με μόνη διαφοροποίηση τη σταθερά στον αριθμητή. Η ελάχιστη Μπαϊεσιανή διακινδύνευση για την κατάσταση αυτή μπορεί να υπολογιστεί όπως αυτή για την MMSE εκτίμηση και ο υπολογισμός αυτός δίνεται ως άσκηση.

Και η MAP εκτίμηση του Θ μπορεί επίσης να ληφθεί εύκολα στην περίπτωση αυτή. Σημειώνοντας ότι

$$\begin{aligned} \frac{\partial}{\partial \theta} [\log p_{\theta}(y) + \log w(\theta)] &= \frac{\partial}{\partial \theta} (\log \theta - \theta y + \log \alpha - \alpha \theta) \\ &= \theta^{-1} - (\alpha + y) \end{aligned}$$

Και

$$\frac{\partial^2}{\partial \theta^2} [\log p_{\theta}(y) + \log w(\theta)] = -\theta^{-2} < 0,$$

βλέπουμε ότι η $w(\theta|\underline{y})$ έχει το μοναδικό μέγιστό της στο

$$\hat{\theta}_{MAP}(\underline{y}) = \frac{1}{\alpha + \underline{y}}. \quad (\text{IV.B.30})$$

Άρα και πάλι έχουμε μια εκτίμηση που διαφέρει από την MMSE εκτίμηση μόνο σε ένα συντελεστή κλίμακας.

Στο Παράδειγμα IV.B.1 τα τρία κριτήρια εκτίμησης που εξετάζονται οδηγούν σε τρεις διαφορετικούς εκτιμητές (ή εκτιμήτριες συναρτήσεις) για το Θ . Για να αποφασίσουμε ποιο από αυτά θα χρησιμοποιήσουμε, θα πρέπει να αποφασίσουμε ποια από τις τρεις αντίστοιχες συναρτήσεις κόστους «τιμωρεί» την εκτίμηση σφάλματος με τον πιο κατάλληλο τρόπο για την εφαρμογή που μας ενδιαφέρει. Σε πολλά προβλήματα δεν χρειάζεται να γίνει μια τέτοια επιλογή γιατί οι τρεις εκτιμήσεις συμπίπτουν. Ακολουθεί ένα παράδειγμα μιας τέτοιας κατάστασης.

Παράδειγμα IV.B.2.: Εκτίμηση Εύρους σήματος

Εξετάστε την περίπτωση $\Gamma = \mathbb{R}^n$ και $\Lambda = \mathbb{R}^-$ με

$$Y_k = N_k + \Theta s_k, \quad k = 1, \dots, n, \quad (\text{IV.B.31})$$

όπου $\underline{N} \sim \mathcal{N}(\underline{0}, \Sigma)$, \underline{s} είναι γνωστό, $\Theta \sim \mathcal{N}(\mu, v^2)$ και \underline{N} και Θ είναι ανεξάρτητα. Ας σημειωθεί ότι αυτό το πρόβλημα αντιστοιχεί στην εκτίμηση του άγνωστου εύρους ενός ειδήλλως γνωστού παρατηρούμενου σήματος με την παρουσία προσθετικού θορύβου.

Δοθέντος ότι $\Theta = \theta$, έχουμε ότι $\underline{Y} \sim \mathcal{N}(\theta \underline{s}, \Sigma)$. Άρα η ύστερη πυκνότητα για το Θ είναι

$$\begin{aligned} w(\theta|\underline{y}) &= \frac{\frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-(1/2)(\underline{y} - \theta \underline{s})^T \Sigma^{-1} (\underline{y} - \theta \underline{s})} \frac{1}{\sqrt{2\pi v}} e^{-(\theta - \mu)^2 / 2v^2}}{\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-(1/2)(\underline{y} - \theta \underline{s})^T \Sigma^{-1} (\underline{y} - \theta \underline{s})} \frac{1}{\sqrt{2\pi v}} e^{-(\theta - \mu)^2 / 2v^2} d\theta} \\ &= K(\underline{y}) \exp \left\{ -\frac{\theta^2}{2} (d^2 + 1/v^2) + \theta (\underline{s}^T \Sigma^{-1} \underline{y} + \frac{\mu}{v^2}) \right\}, \quad (\text{IV.B.32}) \end{aligned}$$

όπου όπως στο Κεφάλαιο III, ορίζουμε $d^2 = \underline{s}^T \Sigma^{-1} \underline{s}$ και όπου $K(\underline{y})$ είναι μια συνάρτηση που εξαρτάται από το \underline{y} αλλά όχι από το θ . Σημειώστε από το (IV.B.32) ότι $w(\theta|\underline{y})$ είναι ο εκθέτης ενός τετραγωνικού όρου στο θ , άρα πρέπει να πρόκειται για κατά Gauss πυκνότητα.

Αν $w(\theta|\underline{y})$ ήταν $\mathcal{N}(m, q^2)$, θα είχαμε

$$\begin{aligned} w(\theta|\underline{y}) &= \frac{1}{\sqrt{2\pi}q} e^{-(\theta-m)^2/2q^2} \\ &= \frac{e^{-m^2/2q^2}}{\sqrt{2\pi}q} e^{-\theta^2/2q^2 + \theta m/q^2}. \end{aligned} \quad (\text{IV.B.33})$$

Συγκρίνοντας το (IV.B.32) με (IV.B.33), βλέπουμε ότι δοθέντος ότι

$$\underline{Y} = \underline{y}, \Theta \sim \mathcal{N}(m, q^2) \text{ με}$$

$$q^2 = (d^2 + 1/v^2)^{-1}$$

και

$$m = (d^2 + 1/v^2)^{-1} (\underline{s}^T \Sigma^{-1} \underline{y} + \mu/v^2),$$

και η $K(\underline{y})$ γίνεται $e^{-m^2/2q^2} / \sqrt{2\pi}q$.

Καθώς η πρώτη παράμετρος της κατά Gauss πυκνότητας είναι το μέσο της, αμέσως έχουμε ότι η δεσμευμένη μέση εκτίμηση του Θ είναι

$$\begin{aligned} \hat{\theta}_{MMSE}(\underline{y}) &= \frac{\underline{s}^T \Sigma^{-1} \underline{y} + \mu/v^2}{d^2 + 1/v^2} \\ &= \frac{v^2 d^2 \hat{\theta}_1(\underline{y}) + \mu}{v^2 d^2 + 1}, \end{aligned} \quad (\text{IV.B.34})$$

όπου $\hat{\theta}_1(\underline{y}) \triangleq \underline{s}^T \Sigma^{-1} \underline{y} / d^2$. Επιπλέον, το ελάχιστο μέσο τετραγωνικό σφάλμα είναι

$$\text{MMSE} = E\{\text{Var}(\Theta|\underline{Y})\} = \frac{1}{d^2 + 1/v^2} = \frac{v^2}{v^2 d^2 + 1}, \quad (\text{IV.B.35})$$

αφού $\text{Var}(\Theta|\underline{Y}) = (d^2 + 1/v^2)^{-1}$, που δεν εξαρτάται από το \underline{Y} . Επίσης, αφού η κατά Gauss πυκνότητα είναι συμμετρική ως προς το μέσο της και επιτυγχάνει το μέγιστό της στο μέσο της, η δεσμευμένη διάμεσος και η δεσμευμένη ταλάντωση ισούνται με το δεσμευμένο μέσο.

Δηλ., έχουμε $\hat{\theta}_{\text{ABS}} = \hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{MMSE}}$.

Η συμπεριφορά της εκτίμησης αυτής απεικονίζει ικανοποιητικά τη φύση της Μπαϊεσιανής εκτίμησης. Σημειώστε ότι το v^2 απεικονίζει την ακρίβεια της πρότερης γνώσης μας για το Θ . Δηλαδή, όσο μικρότερο είναι το v^2 τόσο ακριβέστερα γνωρίζουμε το θ ελλείψει παρατηρήσεων. Από την άλλη, λαμβάνοντας υπόψη το θέμα της συνεκτικής ανίχνευσης σε κατά Gauss θόρυβο στο Κεφάλαιο III, το μέγεθος d^2 είναι μια μέτρηση της ποιότητας στην οποία το \underline{y} μπορεί να διακριθεί από τον $\mathcal{N}(0, \Sigma)$ θόρυβο. Αυτό σημαίνει ότι το d^2 είναι μια μέτρηση της ακρίβειας των παρατηρήσεών μας ως προς την παραγωγή πληροφοριών σχετικά με το σήμα – το μεγάλο d^2 αντιστοιχεί σε υψηλής ποιότητας παρατηρήσεις και το μικρό d^2 αντιστοιχεί σε χαμηλής ποιότητας παρατηρήσεις κατ' αυτήν την έννοια.

Έχοντας αυτό στο μυαλό σας, εξετάστε την εκτίμηση $\hat{\theta}_{\text{MMSE}}$ του (IV.B.32). Αν το $v^2 d^2$ είναι πολύ μικρό σε σχέση με άλλα μεγέθη αυτής της εκτίμησης, έχουμε $\hat{\theta}_{\text{MMSE}}(\underline{y}) \cong \mu$. Αυτό προκύπτει όταν η πρότερη γνώση είναι πολύ ακριβής ως προς τις παρατηρήσεις (δηλ. το v^2 είναι μικρό σε σχέση με το $1/d^2$), οπότε ο εκτιμητής αγνοεί τις παρατηρήσεις και επιλέγει το μέσο της πρότερης κατανομής ως εκτίμησή του. Σημειώστε ότι η MMSE στην περίπτωση αυτή είναι κατά προσέγγιση v^2 , η πρότερη μεταβλητή. Από την άλλη αν το $v^2 d^2$ είναι μεγάλο, τότε $\hat{\theta}_{\text{MMSE}}(\underline{y}) \cong \hat{\theta}_1(\underline{y})$, μια εκτίμηση που εξαρτάται μόνο από τις παρατηρήσεις και δεν ενσωματώνει καθόλου τις πρότερες πληροφορίες. Η τελευταία αυτή περίπτωση είναι επίσης λογική καθώς με μεγάλο v^2 σε σχέση με το $1/d^2$ είναι προτιμητέο να εμπιστευτούμε τις παρατηρήσεις παρά τις πρότερες πληροφορίες. Η MMSE στην τελευταία περίπτωση είναι κατά προσέγγιση $1/d^2$. Μεταξύ αυτών των δύο άκρων, ο βέλτιστος εκτιμητής ισορροπεί την πρότερη γνώση και τις παρατηρήσεις και η αντίστοιχη MMSE αντανακλά αυτή την ισορροπία.

Έχει ενδιαφέρον να εξετάσουμε την ειδική περίπτωση $\Sigma = \sigma^2 \mathbf{I}$ και $\underline{\mathbf{s}} = \underline{\mathbf{1}} \triangleq (1, 1, \dots, 1)^T$ στο πλαίσιο της παραπάνω συζήτησης. Στην περίπτωση αυτή οι παρατηρήσεις μας είναι

$$Y_k = N_k + \Theta, \quad k = 1, \dots, n,$$

με N_1, \dots, N_n i.i.d. $\mathcal{N}(0, \sigma^2)$. Το μέγεθος $v^2 d^2 = nv^2/\sigma^2$ και

$\hat{\theta}_1(\underline{\mathbf{y}}) = \bar{y} \triangleq (1/n) \sum_{k=1}^n y_k$ το μέσο δείγμα. Αν δεν έχουμε παρατηρήσεις ($n = 0$), απλώς εκτιμούμε το Θ ως το πρότερο μέσο του μ , αλλά καθώς λαμβάνουμε περισσότερες παρατηρήσεις (αυξάνουμε το n) το μέσο δείγμα \bar{y} γίνεται πιο αξιόπιστο και βασιζόμαστε περισσότερο σ' αυτό.

Στο όριο όπως $n \rightarrow \infty$ παραβλέπουμε τελείως το πρότερο μέσο και υιοθετούμε το μέσο δείγμα ως εκτίμησή μας.

Η κλίμακα αυτής της συμπεριφοράς ελέγχεται από το λόγο v^2/σ^2 (σημειώστε ότι το σ^2 ορίζει την ακρίβεια κάθε παρατήρησης).

Στην παραπάνω συζήτηση επικεντρωθήκαμε στην εκτίμηση μίας μονής πραγματικής παραμέτρου. Ωστόσο, σε πολλά πρακτικά προβλήματα επιθυμούμε να εκτιμήσουμε ταυτόχρονα διάφορες παραμέτρους. Η Μπαϊεσιανή διατύπωση, φυσικά, εφαρμόζεται εξίσου καλά και στην κατάσταση διανύσματος παραμέτρου, και παρακάτω μελετούμε την περίπτωση αυτή.

Περίπτωση IV.B.4: Εκτίμηση διανύσματος παραμέτρων

Εξετάζουμε τώρα την περίπτωση στην οποία $\Lambda = \mathbb{R}^m$. Για να ακολουθήσουμε τη Μπαϊεσιανή διαδικασία σχεδιασμού μιας εκτίμησης του $\underline{\Theta}$ πρέπει να ορίσουμε μια συνάρτηση κόστους $C : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$. Ενίοτε είναι χρήσιμο να χρησιμοποιήσουμε μια συνάρτηση κόστους της μορφής

$$C[\underline{\mathbf{a}}, \underline{\boldsymbol{\theta}}] = \sum_{i=1}^m C_i[\mathbf{a}_i, \boldsymbol{\theta}_i], \quad (\text{IV.B.36})$$

2. Σημειώστε ότι, ελλείψει άλλων παρατηρήσεων, η MMSE εκτίμηση είναι μ και η MMSE είναι v^2 που αντιστοιχούν στις κατά προσέγγιση συνθήκες για μικρά $v^2 d^2$. Αποδεικνύεται, όπως θα δούμε παρακάτω, ότι η $\hat{\theta}_1(\underline{\mathbf{y}})$ και η ακρίβεια $1/d^2$ είναι βέλτιστες, ελλείψει άλλων πρότερων πληροφοριών. Άρα αυτά τα δύο άκρα είναι αρκετά λογικά.

όπου C_i είναι μια συνάρτηση κόστους που συνδέεται με την εκτίμηση της i -οστής συνιστώσας της παραμέτρου. Αν έχουμε μια συνάρτηση κόστους αυτής της μορφής, το δεσμευμένο πρότερο κόστος για μια εκτίμηση $\hat{\underline{\theta}}$ δίνεται από

$$E\{C[\hat{\underline{\theta}}(\underline{y}), \underline{\Theta}]|Y = \underline{y}\} = \sum_{i=1}^m E\{C_i[\hat{\theta}_i(\underline{y}), \Theta_i]|Y = \underline{y}\}, \quad (\text{IV.B.37})$$

οπότε ουσιαστικά έχουμε m προβλήματα βαθμωτής εκτίμησης. Δηλαδή, το $\hat{\theta}_i(\underline{y})$ [η i -οστή συνιστώσα της $\hat{\underline{\theta}}(\underline{y})$] επιλέγεται για να ελαχιστοποιήσει το $E\{C_i[\hat{\theta}_i(\underline{y}), \Theta_i]|Y = \underline{y}\}$.

Ένα παράδειγμα μιας χρήσιμης συνάρτησης κόστους που αποσυντίθεται όπως στο (IV.B.36) είναι το τετράγωνο του Ευκλείδειου κανόνα του σφάλματος:

$$C[\underline{a}, \underline{\theta}] = \|\underline{a} - \underline{\theta}\|^2 = \sum_{i=1}^m (a_i - \theta_i)^2. \quad (\text{IV.B.38})$$

Έπεται από την περίπτωση IV.B.1 ότι γι' αυτή την συνάρτηση κόστους η i -οστή συνιστώσα της Μπαϊεσιανής εκτίμησης είναι $E\{\Theta_i|Y = \underline{y}\}$. Δηλ., η Μπαϊεσιανή εκτίμηση είναι

$$\hat{\underline{\theta}}_B(\underline{y}) = E\{\underline{\Theta}|Y = \underline{y}\}, \quad (\text{IV.B.39})$$

το δεσμευμένο μέσο του $\underline{\Theta}$ δοθέντος ότι $Y = \underline{y}$.

Ένα άλλο παράδειγμα συνάρτησης κόστους που ικανοποιεί το (IV.B.36) είναι το ακόλουθο

$$C[\underline{a}, \underline{\theta}] = \sum_{i=1}^m |a_i - \theta_i|. \quad (\text{IV.B.40})$$

Αυτή η συνάρτηση παρέχει μια εναλλακτική στο $\| \underline{a} - \underline{\theta} \|$ ως μια μέτρηση της απόστασης μεταξύ \underline{a} και $\underline{\theta}$. Από την Περίπτωση IV.B.2 βλέπουμε ότι αυτή η συνάρτηση κόστους οδηγεί στην εκτίμησης της i -οστής συνιστώσας της οποίας είναι η δεσμευμένη διάμεσος του Θ_i δοθέντος ότι $Y = y$.

Για να επεκτείνουμε την έννοια της MAP εκτίμησης στο διάνυσμα παραμέτρων, θα πρέπει να εξετάσουμε μια συνάρτηση κόστους της μορφής (IV.B.36), στην οποία η $C_i[\underline{a}_i, \theta_i]$ είναι η συνάρτηση ομοιόμορφου κόστους του (IV.B.14). Αυτό οδηγεί στον εκτιμητή διανύσματος που έχει την i -οστή του συνιστώσα στη δεσμευμένη ταλάντωση του Θ_i δοθέντος ότι $Y = y$. Ωστόσο, αυτή η αποσύνθεση συνάρτησης κόστους δεν είναι η πιο σημαντική επέκταση της συνάρτησης ομοιόμορφου κόστους στην περίπτωση του διανύσματος. Πιο σημαντική είναι

$$C[\underline{a}, \underline{\theta}] = \begin{cases} 1 & \text{if } \max_{1 \leq i \leq m} |a_i - \theta_i| > \Delta \\ 0 & \text{if } \max_{1 \leq i \leq m} |a_i - \theta_i| \leq \Delta, \end{cases} \quad (\text{IV.B.41})$$

για την οποία έχουμε

$$E\{C[\hat{\underline{\theta}}(y), \underline{\Theta}] | Y = y\} = 1 - P(|\hat{\theta}_1(Y) - \Theta_1| \leq \Delta, \dots, |\hat{\theta}_m(Y) - \Theta_m| \leq \Delta | Y = y). \quad (\text{IV.B.42})$$

Από το (IV.B.42) μπορούμε να θεωρήσουμε την κατά προσέγγιση βελτιστότητα της εκτίμησης του $\underline{\Theta}$ ως τη δεσμευμένη του ταλάντωση, δοθέντος ότι $Y = y$, ένα μέγεθος που διαφοροποιείται γενικά από το διάνυσμα, η i -οστή συνιστώσα του οποίου είναι η δεσμευμένη ταλάντωση του Θ_i δοθέντος ότι $Y = y$ που λαμβάνεται από την αποσύνθεση του κόστους. Η εκτίμηση που επιλέγει τη δεσμευμένη ταλάντωση του $\underline{\Theta}$ δοθέντος ότι $Y = y$ είναι η MAP εκτίμηση για την περίπτωση διανύσματος παραμέτρου. Σημειώστε ότι η περιοχή όπου το $\max_{1 \leq i \leq m} |a_i - \theta_i| \leq \Delta$ είναι ένα m -διάστατος κύβος, κεντρικός στο $\underline{\theta}$ με μήκος πλευράς 2Δ . Θα μπορούσαμε να ορίσουμε παρόμοιες συναρτήσεις κόστους αντικαθιστώντας αυτόν τον κύβο με άλλα σχήματα (π.χ. μια m -διάστατη σφαίρα, $\| \underline{a} - \underline{\theta} \| \leq \Delta$). Ωστόσο, η κατά προσέγγιση βελτιστότητα της MAP εκτίμησης θα ήταν συνεπαγόμενη στο πλαίσιο των κατάλληλων συνθηκών λειότητας.

Μια ακόμη χρήσιμη συνάρτηση κόστους στην εκτίμηση διανύσματος παραμέτρων είναι μια γενίκευση του κανόνα τετραγωνικού σφάλματος. Συγκεκριμένα έχει ενδιαφέρον να εξετάσουμε συναρτήσεις κόστους της μορφής

$$C[\underline{a}, \underline{\theta}] = (\underline{a} - \underline{\theta})^T \mathbf{A} (\underline{a} - \underline{\theta}), \quad (\text{IV.B.43})$$

όπου \mathbf{A} είναι ένας συμμετρικός, θετικά ορισμένος πίνακας. Σημειώστε ότι αυτή η συνάρτηση κόστους επιτρέπει τις από κοινού σταθμίσεις σφαλμάτων σε διαφορετικές παραμέτρους, ένα επιθυμητό χαρακτηριστικό για κάποιες εφαρμογές καθώς η ακρίβεια της εκτίμησης μας για μια από τις πραγματικές παραμέτρους που σχηματίζουν το $\underline{\Theta}$ μπορεί να επηρεάσει το πόσο καλά πρέπει να γνωρίζουμε άλλες παραμέτρους.

Για να εξάγουμε την Μπαϊεσιανή εκτίμηση για το (IV.B.43), γράφουμε

$$\begin{aligned} E\{(\hat{\theta}(y) - \underline{\Theta})^T \mathbf{A}(\hat{\theta}(y) - \underline{\Theta}) | Y = y\} \\ = [\hat{\theta}(y)]^T \mathbf{A} \hat{\theta}(y) - 2[\hat{\theta}(y)]^T \mathbf{A} E\{\underline{\Theta} | Y = y\} \\ + E\{\underline{\Theta}^T \mathbf{A} \underline{\Theta} | Y = y\}. \end{aligned} \quad (\text{IV.B.44})$$

Καθώς η συνάρτηση του (IV.B.44) είναι τετραγωνική στο $\hat{\theta}(y)$, φτάνει στο ελάχιστό της στο σημείο στο οποίο η βαθμίδα της ως προς το $\hat{\theta}(y)$ εξαφανίζεται. Έχουμε ευθέως ότι

$$\nabla_{\hat{\theta}(y)} E\{C[\hat{\theta}(y), \underline{\Theta}] | Y = y\} = 2\mathbf{A}\hat{\theta}(y) - 2\mathbf{A}E\{\underline{\Theta} | Y = y\}. \quad (\text{IV.B.45})$$

Άρα η Μπαϊεσιανή εκτίμηση, $\hat{\theta}_B$ για το (IV.B.43) ικανοποιεί

$$2\mathbf{A}\hat{\theta}_B(y) = 2\mathbf{A}E\{\underline{\Theta} | Y = y\}. \quad (\text{IV.B.46})$$

Πολλαπλασιάζοντας από αριστερά το (IV.B.46) με $(1/2)\mathbf{A}^{-1}$ έχουμε ότι

$$\hat{\theta}_B(y) = E\{\underline{\Theta} | Y = y\}.$$

Έτσι βλέπουμε ότι το τετραγωνικό κριτήριο κόστους του (IV.B.43) αποδίδει το διάνυσμα δεσμευμένου μέσου ως μια Μπαϊεσιανή εκτίμηση ανεξάρτητα από την επιλογή του \mathbf{A} . Η απορρέουσα Μπαϊεσιανή διακινδύνευση φυσικά εξαρτάται από το \mathbf{A} και θα πρέπει ευθέως να αποδειχθεί αυτό (βλέπε Άσκηση 10) για την περίπτωση αυτή.

$$r(\hat{\theta}_B) = \text{tr}\{\mathbf{A}E\{\text{Cov}(\underline{\Theta} | Y)\}\}, \quad (\text{IV.B.47})$$

όπου το $\text{tr}\{\cdot\}$ συμβολίζει τον τελεστή ίχνους (δηλ. άθροιση των διαγώνιων όρων) και όπου το $\text{Cov}(\underline{\Theta}|Y)$ είναι ο πίνακας δεσμευμένης διακύμανσης του $\underline{\Theta}$ δοθέντος ότι $Y = y$. Ας σημειωθεί ότι ο κανόνας τετραγωνικού σφάλματος είναι η ειδική περίπτωση του (IV.B.47) με $A = I$, άρα η τελευταία περίπτωση $r(\hat{\underline{\theta}}_B)$ είναι απλώς το ίχνος του $E\{\text{Cov}(\underline{\Theta}|Y)\}$

Παράδειγμα IV.B.3: Εκτίμηση ενός κατά Gauss διανύσματος από μια από κοινού κατά Gauss παρατήρηση.

Εξετάστε την περίπτωση στην οποία $\Gamma = \mathbb{R}^n$, $\Lambda = \mathbb{R}^m$, και Y και $\underline{\Theta}$ είναι από κοινού κατά Gauss με μέσα διανύσματα $\underline{\mu}_Y$ και $\underline{\mu}_\Theta$, πίνακες διακύμανσης Σ_Y και Σ_Θ και πίνακα διασταυρούμενης διακύμανσης $\Sigma_{Y\Theta} \triangleq E\{(\underline{Y} - \underline{\mu}_Y)(\underline{\Theta} - \underline{\mu}_\Theta)^T\}$. Δηλαδή, υποθέτουμε ότι

$$\begin{pmatrix} \underline{Y} \\ \underline{\Theta} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \underline{\mu}_Y \\ \underline{\mu}_\Theta \end{pmatrix}, \begin{pmatrix} \Sigma_Y & \Sigma_{Y\Theta} \\ \Sigma_{\Theta Y} & \Sigma_\Theta \end{pmatrix}\right) \quad (\text{IV.B.48})$$

με $\Sigma_{\Theta Y} = \Sigma_{Y\Theta}^T$.

Στο πλαίσιο αυτού του μοντέλου πρέπει να αποδειχθεί ευθέως ότι η δεσμευμένη κατανομή του $\underline{\Theta}$ δοθέντος ότι $\underline{Y} = \underline{y}$ είναι επίσης κατά Gauss, με δεσμευμένο μέσο $\hat{\underline{\mu}}(\underline{y})$ που δίνεται από

$$\hat{\underline{\mu}}(\underline{y}) = \underline{\mu}_\Theta + \Sigma_{\Theta Y} \Sigma_Y^{-1} (\underline{y} - \underline{\mu}_Y) \quad (\text{IV.B.49a})$$

και με πίνακα δεσμευμένης διακύμανσης $\hat{\Sigma}$ που δίνεται από

$$\hat{\Sigma} = \Sigma_\Theta - \Sigma_{\Theta Y} \Sigma_Y^{-1} \Sigma_{Y\Theta}. \quad (\text{IV.B.49b})$$

Από την ιδιότητα αυτή μπορούμε να βρούμε όλες τις βέλτιστες εκτιμήσεις που αναλύθηκαν στην Περίπτωση IV.B.4. Συγκεκριμένα, παρατηρούμε αμέσως ότι η εκτίμηση δεσμευμένου

μέσου ισούται με το $\hat{\underline{\mu}}(\underline{y})$ του (IV.B.49a). Επίσης, αφού η πολυμεταβλητή κατά Gauss πυκνότητα έχει τον τρόπο της στο μέσο της, η MAP εκτίμηση δίνεται επίσης από το $\hat{\underline{\mu}}(\underline{y})$. Επιπλέον, αφού το $\underline{\Theta}$ είναι κατά Gauss δοθέντος ότι $\underline{Y} = \underline{y}$, η περιθώρια ταλάντωση και η διάμεσος του Θ_i είναι οριακά κατά Gauss δοθέντος ότι $\underline{Y} = \underline{y}$ προκύπτει στο $\hat{\mu}_i(\underline{y})$, στην i -οστή συνιστώσα του $\hat{\underline{\mu}}(\underline{y})$. Άρα το $\hat{\underline{\mu}}(\underline{y})$ παρέχει τη βέλτιστη εκτίμηση υπό όλες τις έννοιες που αναλύθηκαν στην Περίπτωση IV.B.4. Θα πρέπει να σημειωθεί ότι αυτή η εκτίμηση είναι γραμμική (ή ορθότερα, *αφροινική*) στο \underline{y} , οπότε υπολογίζεται εύκολα, αν το Σ_Y^{-1} μπορεί να προσδιοριστεί επαρκώς. Το ζήτημα αυτό θα σχολιαστεί εκτενέστερα αργότερα.

Η ελάχιστη Μπαϊεσιανή διακινδύνευση μπορεί επίσης εύκολα να υπολογιστεί για τη συνάρτηση τετραγωνικού κόστους του (IV.B.43) μέσω του (IV.B.47). Συγκεκριμένα, παρατηρούμε ότι $\text{Cov}(\underline{\Theta}|\underline{Y}) = \hat{\Sigma}$, που δεν εξαρτάται από το \underline{Y} . Άρα το $E\{\text{Cov}(\underline{\Theta}|\underline{Y})\} = \hat{\Sigma}$ και η ελάχιστη Μπαϊεσιανή διακινδύνευση γίνεται

$$r(\hat{\underline{\theta}}_B) = \text{tr}\{\mathbf{A}\hat{\Sigma}\} = \text{tr}\{\mathbf{A}\Sigma_{\theta}\} - \text{tr}\{\mathbf{A}\Sigma_{\theta Y}\Sigma_Y^{-1}\Sigma_{Y\theta}\}. \quad (\text{IV.B.50})$$

Παρατηρήστε επίσης ότι $\hat{\Sigma} = E\{(\underline{\Theta} - \hat{\underline{\theta}}_B(\underline{Y}))(\underline{\Theta} - \hat{\underline{\theta}}_B(\underline{Y}))^T\}$, άρα το $\hat{\Sigma}$ είναι ο πίνακας συνδιακύμανσης της εκτίμησης σφάλματος $\underline{\Theta} - \hat{\underline{\theta}}_B(\underline{Y})$.

Μια ειδική ενδιαφέρουσα περίπτωση αυτού του γενικού κατά Gauss προβλήματος προκύπτει από το λεγόμενο *γραμμικό μοντέλο παρατήρησης*:

$$\underline{Y} = \mathbf{H}\underline{\Theta} + \underline{N}, \quad (\text{IV.B.51})$$

όπου $\underline{\Theta} \sim \mathcal{N}(\underline{\mu}_{\theta}, \Sigma_{\theta})$, $\underline{N} \sim \mathcal{N}(\underline{0}, \Sigma)$, \mathbf{H} είναι ένας σταθερός $n \times m$ πίνακας και τα $\underline{\Theta}$ και \underline{N} είναι ανεξάρτητα. Τέτοια μοντέλα προκύπτουν σε πολλές εφαρμογές. Για παράδειγμα, το μοντέλο του Παραδείγματος IV.B.2 στο οποίο επιθυμούμε να εκτιμήσουμε το εύρος σήματος είναι αυτό της μορφής με $m = 1$ και $\mathbf{H} = \underline{g}$. Επιπλέον, αν θεωρήσουμε τα $\Theta_1, \dots, \Theta_m$ ως δείγματα ενός στοχαστικού σήματος, τότε

$$Y_k = \sum_{j=1}^m h_{k,j}\Theta_j + N_k, \quad k = 1, \dots, n \quad (\text{IV.B.52})$$

είναι μια ακολουθία παρατήρησης αποτελούμενη από γραμμικά φιλτραρισμένο σήμα συν πρόσθετο θόρυβο – μια κατάσταση που προκύπτει, για παράδειγμα, όταν ένα σήμα παρατηρείται μέσω ενός καναλιού με πεπερασμένο εύρος ζώνης ή άλλο γραμμικά

παραμορφωτικό χαρακτηριστικό. Στην περίπτωση αυτή η εκτίμηση του $\underline{\Theta}$ είναι γνωστή ως το πρόβλημα *εξίσωσης καναλιού*. Περισσότερες εφαρμογές του μοντέλου εξετάζονται στο Κεφάλαιο V στο ζήτημα του φίλτρου Kalman – Bucy.

Στο μοντέλο αυτό είναι εύκολο να δείξουμε ότι τα \underline{Y} και $\underline{\Theta}$ είναι από κοινού κατά Gauss, με τα $\underline{\mu}_\theta$ και Σ_θ δεδομένα, $\underline{\mu}_Y = \mathbf{H}\underline{\mu}_\theta$, $\Sigma_Y = \mathbf{H}\Sigma_\theta\mathbf{H}^T + \Sigma$, και $\Sigma_{\theta Y} = \Sigma_\theta\mathbf{H}^T$. Άρα έχουμε την Μπαϊεσιανή εκτίμηση

$$\hat{\underline{\mu}}(\underline{y}) = \underline{\mu}_\theta + \Sigma_\theta\mathbf{H}^T(\mathbf{H}\Sigma_\theta\mathbf{H}^T + \Sigma)^{-1}(\underline{y} - \mathbf{H}\underline{\mu}_\theta) \quad (\text{IV.B.53})$$

και τον πίνακα συνδιακύμανσης σφάλματος

$$\hat{\Sigma} = \Sigma_\theta - \Sigma_\theta\mathbf{H}^T(\mathbf{H}\Sigma_\theta\mathbf{H}^T + \Sigma)^{-1}\mathbf{H}\Sigma_\theta. \quad (\text{IV.B.54})$$

Σε ό,τι αφορά τον υπολογισμό του (IV.B.53) παρατηρούμε ότι περιλαμβάνει την αντιστροφή ενός $n \times n$ πίνακα, ένας υπολογισμός η πολυπλοκότητα του οποίου είναι της τάξης του n^3 , εκτός κι αν ο πίνακας έχει κάποια ειδική δομή. Αυτή η υπολογιστική πολυπλοκότητα μπορεί ενίοτε να μειωθεί κάνοντας χρήση της ακόλουθης απλής ταυτότητας πίνακα

$$\begin{aligned} \Sigma_\theta\mathbf{H}^T(\mathbf{H}\Sigma_\theta\mathbf{H}^T + \Sigma)^{-1} \\ = (\mathbf{H}^T\Sigma^{-1}\mathbf{H} + \Sigma_\theta^{-1})^{-1}\mathbf{H}^T\Sigma^{-1}. \end{aligned} \quad (\text{IV.B.55})$$

Αν το Σ^{-1} είναι γνωστό (δηλ., αν $\Sigma = \sigma^2\mathbf{I}$) και $m < n$, ο πίνακας στο δεξιό μέλος του (IV.B.55) είναι ευκολότερο να υπολογιστεί από αυτόν στο αριστερό.

Στο Κεφάλαιο V, τα (IV.B.53) και (IV.B.54) θα χρησιμοποιηθούν για να παράγουν το φίλτρο Kalman – Bucy. Είναι επίσης ενδιαφέρον αν επανεξετάσουμε το Παράδειγμα IV.B.2 σ' αυτό το γενικό πλαίσιο. Στην περίπτωση αυτή έχουμε $m = 1$, $\mathbf{H} = \underline{s}$, $\underline{\mu}_\theta = \mu$, και $\Sigma_\theta = v^2$. Εισάγοντας τα μεγέθη αυτά στα (IV.B.53) και (IV.B.54) και εφαρμόζοντας το (IV.B.55), έχουμε

$$\begin{aligned} \hat{\underline{\mu}}(\underline{y}) &= \mu + (\underline{s}^T\Sigma^{-1}\underline{s} + 1/v^2)^{-1}\underline{s}^T\Sigma^{-1}(\underline{y} - \underline{s}\mu) \\ &= \frac{v^2 d^2 \hat{\theta}_1(\underline{y}) + \mu}{v^2 d^2 + 1} \end{aligned}$$

και

$$\begin{aligned} r(\hat{\mu}) &= \hat{\Sigma} = v^2 - (\underline{s}^T \Sigma^{-1} \underline{s} + v^{-2})^{-1} \underline{s}^T \Sigma^{-1} \underline{s} v^2 \\ &= \frac{v^2}{v^2 d^2 + 1} \end{aligned}$$

όπως στα (IV.B.34) και (IV.B.35).

IV.C Εκτίμηση μη τυχαίας παραμέτρου: Γενική Δομή

Στην ενότητα IV.B. εξετάσαμε το πρόβλημα εκτίμησης μιας τυχαίας παραμέτρου που δεικνύει μια κλάση κατανομών στο χώρο παρατήρησης. Ένα συναφές πρόβλημα είναι αυτό στο οποίο έχουμε μια παράμετρο (που δεικνύει την κλάση των στατιστικών παρατήρησης) που δεν μοντελοποιείται ως τυχαία μεταβλητή, αλλά ωστόσο, είναι άγνωστη. Συγκεκριμένα, μπορεί να μην έχουμε αρκετές πρότερες πληροφορίες σχετικά με την παράμετρο ώστε να της αποδώσουμε μια πρότερη κατανομή πιθανότητας, αλλά ωστόσο θέλουμε να χειριστούμε την εκτίμηση τέτοιων παραμέτρων με οργανωμένο τρόπο.

Έστω, λοιπόν, ότι έχουμε μια παρατήρηση $Y \in \Gamma$ και ότι η κατανομή του Y είναι μέλος μιας κλάσης κατανομών στο (Γ, \mathcal{G}) που δεικνύεται από μια παράμετρο θ που βρίσκεται σε κάποιο σύνολο Λ . Όπως προηγουμένως, συμβολίζουμε αυτό το σύνολο κατανομών με $\{P_\theta; \theta \in \Lambda\}$. Ας υποθέσουμε προς το παρόν ότι η παράμετρος θ είναι πραγματικής τιμής.

Δεν ξέρουμε τίποτα για την πραγματική τιμή του θ πέρα από το γεγονός ότι βρίσκεται στο Λ , και με απλά λόγια, το πρόβλημα που θέλουμε να λύσουμε είναι: Δοθείσης της παρατήρησης $Y=y$, ποια είναι η καλύτερη εκτίμηση του θ ; Έχοντας υπόψη τις διαδικασίες που αναπτύχθηκαν στην Ενότητα IV.B., μπορούμε να αρχίσουμε να απαντούμε αυτό το ερώτημα αναζητώντας μια εκτίμηση $\hat{\theta}(y)$ που ελαχιστοποιεί κάποιο κριτήριο μέσης απόδοσης. Σε όλο το υπόλοιπο κεφάλαιο, εξετάζουμε αποκλειστικά το κόστος τετραγωνικού σφάλματος, παρόλο που κάποια αποτελέσματα που θα μελετηθούν εδώ εφαρμόζονται ευθέως και σε άλλες αποδόσεις κόστους. Εν απουσία ενός πρότερου στο Λ , ο μόνος τρόπος εύρεσης μέσου όρου του κόστους σχετίζεται με την κατανομή του Y δοθέντος του θ . Δηλ., μπορούμε να χρησιμοποιήσουμε μόνο τη δεσμευμένη συνάρτηση διακινδύνευσης

$$R_\theta(\hat{\theta}) \triangleq E_\theta\{(\hat{\theta}(Y) - \theta)^2\}, \theta \in \Lambda.$$

Όπως είδαμε στην περίπτωση ελέγχου υπόθεσης στο Κεφάλαιο II, δεν μπορούμε γενικά να περιμένουμε την ομοιόμορφη ελαχιστοποίηση του $R_\theta(\hat{\theta})$ για $\theta \in \Lambda$. Αυτό είναι προφανές για το κόστος τετραγωνικού σφάλματος, καθώς για κάθε συγκεκριμένη τιμή του θ , π.χ. θ_0 το δεσμευμένο μέσο τετραγωνικό σφάλμα μπορεί να γίνει μηδέν επιλέγοντας το $\hat{\theta}(y)$ να είναι πανομοιότυπα θ_0 για όλες τις παρατηρήσεις $y \in \Gamma$. Αλλά μια τέτοια εκτίμηση θα είχε χαμηλή

επίδοση αν τα $\hat{\theta}_o$ δεν ήταν κοντά στην πραγματική τιμή του θ . Άρα είναι προφανές ότι το δεσμευμένο μέσο τετραγωνικό σφάλμα δεν είναι από μόνο του ένα κατάλληλο κριτήριο σχεδιασμού για έναν εκτιμητή μιας μη τυχαίας παραμέτρου, εκτός και αν η κλάση των εκτιμητών περιορίζεται κατά κάποιον τρόπο ώστε να περιλαμβάνει μόνο λογικούς εκτιμητές [δηλ., να αποκλείει εκτιμητές όπως $\hat{\theta}(y) \equiv \theta_o$]

Ένας λογικός περιορισμός που μπορεί να τεθεί σε έναν εκτιμητή του θ είναι ότι η αναμενόμενη τιμή του ισούται με την πραγματική τιμή της παραμέτρου. Δηλ. ότι

$$E_{\theta}\{\hat{\theta}(Y)\} = \theta, \quad \theta \in \Lambda. \quad (\text{IV.C.1})$$

Μια τέτοια εκτίμηση ορίζεται ως *αμερόληπτη*. Στο πλαίσιο αυτού του περιορισμού, το δεσμευμένο μέσο τετραγωνικό σφάλμα γίνεται η διακύμανση του εκτιμητή ως προς το P_{θ} , και μια αμερόληπτη εκτίμηση που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα για κάθε $\theta \in \Lambda$ ονομάζεται *ελάχιστη διακύμανση αμερόληπτου εκτιμητή (MVUE)*.

Στην παρούσα ενότητα εξετάζουμε τη γενική δομή προβλημάτων εκτίμησης μη τυχαίας παραμέτρου με στόχο το χαρακτηρισμό των MVUE.

Ξεκινάμε με την έννοια της επάρκειας, όπως ορίζεται ακολούθως (μέχρι να ισχύσει κάτι διαφορετικό, υποθέτουμε ότι το Λ είναι γενικό, δηλ. όχι απαραίτητα ένα υποσύνολο του \mathbb{R})

Ορισμός IV.C.1 : Επάρκεια

Ας υποθέσουμε ότι Δ είναι ένα αυθαίρετο σύνολο και \mathcal{D} είναι μία κλάση ενδεχομένων στο Δ . Μια συνάρτηση $T : (\Gamma, \mathcal{G}) \rightarrow (\Delta, \mathcal{D})$ λέγεται ότι είναι *επαρκής στατιστική* για $\{P_{\theta}; \theta \in \Lambda\}$ αν η κατανομή του Y δεσμευμένου στο $T(Y)$ όταν $Y \sim P_{\theta}$ δεν εξαρτάται από το θ για $\theta \in \Lambda$ (Όταν η $\{P_{\theta}; \theta \in \Lambda\}$ εννοείται, μπορούμε απλώς να πούμε ότι η T είναι επαρκής για το θ .)

Σημειώστε ότι το θ επηρεάζει τις παρατηρήσεις μόνο μέσω της κατανομής της P_{θ} . Άρα μπορούμε να μάθουμε για το θ βλέποντας τη στατιστική συμπεριφορά του Y . Άρα αν η γνώση της $T(Y)$ απομακρύνει όποια επιπλέον εξάρτηση της κατανομής του Y από το θ , μπορούμε να συμπεράνουμε ότι η $T(Y)$ περιλαμβάνει όλες τις πληροφορίες στο Y που είναι χρήσιμες για την εκτίμηση του θ – εξ ου και η προέλευση του όρου «επαρκής».

Σημειώστε ότι κάθε μία από τις μία προς μία απεικονίσεις των παρατηρήσεων είναι τετριμμένα επαρκής για θ , άρα υπάρχουν πάντα πολλές επαρκείς στατιστικές για κάθε δεδομένο μοντέλο εκτίμησης. Ωστόσο, είναι επιθυμητό να βρεθεί μια επαρκής στατιστική

που να μειώνει όσο τον δυνατόν περισσότερο τις περισσότερες. Στο πλαίσιο αυτό, έχουμε τον ακόλουθο ορισμό.

Πρόταση IV.C.1: Το Θεώρημα της Παραγοντοποίησης

Έστω ότι $\{P_\theta; \theta \in \Lambda\}$ έχει μια αντίστοιχη οικογένεια πυκνοτήτων $\{p_\theta; \theta \in \Lambda\}$. Μια στατιστική T είναι επαρκής για το θ , αν και μόνο αν υπάρχουν συναρτήσεις g_θ και h τέτοιες

$$p_\theta(y) = g_\theta[T(y)]h(y) \quad (\text{IV.C.2})$$

ώστε

για όλα τα $y \in \Gamma$ και $\theta \in \Lambda$

Απόδειξη: Αποδεικνύουμε αυτό το αποτέλεσμα μόνο για την περίπτωση στην οποία το Γ είναι διακριτό. Η περίπτωση αυτή απεικονίζει τη γενική ιδέα αυτής της πρότασης χωρίς να χρησιμοποιεί τις τεχνικές λεπτομέρειες που απαιτούνται για τη γενική περίπτωση. Μια απόδειξη της γενικής περίπτωσης απαντά στον Lehmann (1986).

Έστω ότι το Γ είναι διακριτό και το $\{p_\theta; \theta \in \Lambda\}$ ικανοποιεί το (IV.C.2) για μια συνάρτηση T .

Έστω ότι το $p_\theta(y|t)$ συμβολίζει την πυκνότητα του Y δοθέντος ότι $T(Y) = t$ όταν $Y \sim P_\theta$. Από τον Μπαϊεσιανό τύπο έχουμε

$$\begin{aligned} p_\theta(y|t) &\stackrel{\Delta}{=} \frac{P_\theta(Y = y | T(Y) = t)}{P_\theta(T(Y) = t)} \\ &= \frac{P_\theta(T(Y) = t | Y = y) P_\theta(Y = y)}{P_\theta(T(Y) = t)}. \end{aligned} \quad (\text{IV.C.3})$$

Αφού $P_\theta(T(Y) = t | Y = y)$ ισούται με 1 αν $T(y) = t$ και 0 αν $T(y) \neq t$ και αφού $P_\theta(Y = y) = p_\theta(y)$, το (IV.C.3) γίνεται

$$p_\theta(y|t) = \begin{cases} p_\theta(y)/P_\theta(T(Y) = t) & \text{if } T(y) = t \\ 0 & \text{if } T(y) \neq t. \end{cases} \quad (\text{IV.C.4})$$

Τώρα $P_\theta(T(Y) = t) = \sum_{\{y|T(y)=t\}} p_\theta(y)$. Άρα από το (IV.C.2) έχουμε

$$\begin{aligned}
P_{\theta}(T(Y) = t) &= \sum_{\{y|T(y)=t\}} g_{\theta}[T(y)]h(y) \\
&= g_{\theta}(t) \sum_{\{y|T(y)=t\}} h(y),
\end{aligned}$$

και επίσης έχουμε $p_{\theta}(y) = g_{\theta}[T(y)]h(y) = g_{\theta}(t)h(y)$. Από το (IV.C.4) έχουμε τότε

$$p_{\theta}(y|t) = \begin{cases} h(y) / \sum_{\{y|T(y)=t\}} h(y), & \text{if } T(y) = t \\ 0, & \text{if } T(y) \neq t. \end{cases}$$

Καθώς αυτή η παράσταση δεν εξαρτάται από το θ , η T είναι μια επαρκής στατιστική για το $\{P_{\theta}; \theta \in \Lambda\}$. Αυτό αποδεικνύει ότι η T είναι επαρκής αν το (IV.C.2) ισχύει.

Για να αποδείξουμε ότι η T είναι επαρκής μόνο αν το (IV.C.2) ισχύει, έστω ότι T είναι κάθε επαρκής στατιστική για το θ . Από το (IV.C.4) μπορούμε να γράψουμε

$$p_{\theta}(y) = p_{\theta}[y|T(y)]P_{\theta}[T(Y) = T(y)]. \quad (\text{IV.C.5})$$

Καθώς η T είναι επαρκής για το θ , το $p_{\theta}[y|T(y)]$ εξαρτάται μόνο από το y και όχι από το θ . Επίσης, η $P_{\theta}[T(Y) = T(y)]$ είναι συνάρτηση μόνο του $T(y)$ και του θ . Ορίζοντας $h(y) \triangleq p_{\theta}[y|T(y)]$ και $g_{\theta}[T(y)] \triangleq P_{\theta}[T(Y) = T(y)]$, βλέπουμε ότι το (IV.C.5) συνεπάγεται την παραγοντοποίηση του (IV.C.2). Αυτό ολοκληρώνει την απόδειξη της πρότασης αυτής για την περίπτωση του διακριτού Γ .

Για να απεικονίσουμε την Πρόταση IV.C.1, εξετάζουμε το ακόλουθο απλό παράδειγμα.

Παράδειγμα IV.C.1: Μια επαρκής στατιστική για έλεγχο υπόθεσης

Εξετάστε το πρόβλημα ελέγχου υπόθεσης $\Lambda = \{0,1\}$ με πυκνότητες p_0 και p_1 . Παρατηρώντας ότι

$$p_{\theta}(y) = \begin{cases} p_0(y) & \text{if } \theta = 0 \\ \frac{p_1(y)}{p_0(y)} p_0(y) & \text{if } \theta = 1, \end{cases}$$

$$\begin{aligned} & (\text{αν } \theta = 0 \\ & \text{αν } \theta = 1) \end{aligned}$$

βλέπουμε την παραγοντοποίηση $p_{\theta}(y) = g_{\theta}[T(y)]h(y)$ με $h(y) = p_0(y)$, $T(y) = p_1(y)/p_0(y) \triangleq L(y)$ και $g_{\theta}(t)$ να ορίζεται από

$$g_{\theta}(t) = \begin{cases} 1 & \text{if } \theta = 0 \\ t & \text{if } \theta = 1. \end{cases}$$

$$\begin{aligned} & (\text{αν } \theta = 0 \\ & \text{αν } \theta = 1) \end{aligned}$$

Αρα βλέπουμε ότι ο λόγος πιθανοφάνειας $L(y)$ είναι μια επαρκής στατιστική για το πρόβλημα ελέγχου δυαδικής υπόθεσης. Είναι μια πολύ χρήσιμη επαρκής στατιστική γιατί είναι μονοδιάστατη, ανεξάρτητα από τη φύση του Γ . Φυσικά, έχουμε ήδη δει ότι όλοι οι βέλτιστοι έλεγχοι για $\Lambda = \{0,1\}$ που ορίζονται στο Κεφάλαιο II εξαρτώνται από την παρατήρηση y μόνο μέσω αυτής της επαρκούς στατιστικής $L(y)$

Η χρησιμότητα των επαρκών στατιστικών στην αναζήτηση αμερόληπτων εκτιμητών των πραγματικών παραμέτρων είναι εν μέρει εμφανής από το ακόλουθο αποτέλεσμα. Εδώ επιτρέπουμε το Λ να είναι αυθαίρετο, αλλά υποθέτουμε ότι επιθυμούμε να εκτιμήσουμε κάποια πραγματικής τιμής συνάρτηση g του θ .

Πρόταση IV.C.2: Το Θεώρημα Rao – Blackwell

Έστω ότι η $\hat{g}(y)$ είναι μια αμερόληπτη εκτίμηση της $g(\theta)$ και ότι η T είναι επαρκής για το θ . Ορίστε τη $\bar{g}[T(y)]$ με

$$\bar{g}[T(y)] = E_{\theta}\{\hat{g}(Y)|T(Y) = T(y)\}.$$

Τότε η $\bar{g}[T(Y)]$ είναι επίσης μια αμερόληπτη εκτίμηση της $g(\theta)$. Επιπλέον,

$$\text{Var}_\theta(\tilde{g}[T(Y)]) \leq \text{Var}_\theta[\hat{g}(Y)],$$

$$P_\theta(\hat{g}(Y) = \tilde{g}[T(Y)]) = 1.$$

με ισότητα αν και μόνο αν

Απόδειξη: Παρατηρούμε πρώτα ότι η προσδοκία που ορίζει το \tilde{g} , δεν εξαρτάται από το θ δυνάμει της επάρκειας της T [δοθέντος ότι $T(Y)$, η κατανομή του Y και ως εκ τούτου το μέσω της $\hat{g}(Y)$ δεν εξαρτώνται από το θ]. Για να δούμε ότι το \tilde{g} είναι αμερόληπτο, παρατηρούμε ότι

$$\begin{aligned} E_\theta\{\tilde{g}[T(Y)]\} &= E_\theta\{E_\theta\{\hat{g}(Y)|T(Y)\}\} \\ &= E_\theta\{\hat{g}(Y)\} = g(\theta), \end{aligned}$$

όπου έχουμε χρησιμοποιήσει το δεδομένο ότι $E\{E\{X|Z\}\} = E\{X\}$ για να έχουμε τη δεύτερη ισότητα και την αμεροληψία του \tilde{g} για να έχουμε την τρίτη ισότητα.

Για να δούμε ότι $\text{Var}_\theta(\tilde{g}[T(Y)]) \leq \text{Var}_\theta[\hat{g}(Y)]$, σημειώνουμε πρώτα ότι

$$\text{Var}_\theta(\tilde{g}[T(Y)]) = E_\theta\{[\tilde{g}[T(Y)]]^2\} - g^2(\theta)$$

και

$$\text{Var}_\theta(\hat{g}(Y)) = E_\theta\{[\hat{g}(Y)]^2\} - g^2(\theta).$$

Πρέπει λοιπόν μόνο να αποδείξουμε ότι $E_\theta\{[\tilde{g}[T(Y)]]^2\} \leq E_\theta\{[\hat{g}(Y)]^2\}$. Έχουμε

$$\begin{aligned} E_\theta\{(\tilde{g}[T(Y)])^2\} &= E_\theta\{[E_\theta\{\hat{g}(Y)|T(Y)\}]^2\} \\ &\leq E_\theta\{E_\theta\{[\hat{g}(Y)]^2|T(Y)\}\} = E_\theta\{[\hat{g}(Y)]^2\}, \end{aligned} \tag{IV.C.6}$$

όπου η ανισότητα έπεται από την εφαρμογή της ανισότητας του Jensen ώστε να έχουμε

$$[E_\theta\{\hat{g}(Y)|T(Y)\}]^2 \leq E_\theta\{[\hat{g}(Y)]^2|T(Y)\}$$

και η τελική ισότητα έπεται από

επαναλαμβανόμενες προσδοκίες. Ας σημειωθεί ότι έχουμε ισότητα στην ανισότητα του Jensen αν και μόνο αν $P_\theta[\hat{g}(Y) = E_\theta\{\hat{g}(Y)|T(Y)\}|T(Y)] = 1$. Αφού

$\bar{g}[T(Y)] \triangleq E_{\theta}\{\hat{g}(Y)|T(Y)\}$. αυτό είναι ισοδύναμο με τη συνθήκη $P_{\theta}[\hat{g}(Y) = \bar{g}[T(Y)]] = 1$. Αυτό ολοκληρώνει την απόδειξη της Πρότασης IV.C.2.

Από το θεώρημα Rao – Blackwell βλέπουμε ότι με μια επαρκή στατιστική T μπορούμε να βελτιώσουμε οποιονδήποτε αμερόληπτο εκτιμητή που να μην είναι ήδη μια συνάρτηση της T , δεσμευόντάς τον στην $T(Y)$. Επιπλέον, το θεώρημα αυτό συνεπάγεται ότι αν η T είναι επαρκής για το θ και αν υπάρχει μόνο μία συνάρτηση της T που είναι αμερόληπτη εκτίμηση της $g(\theta)$, αυτή η συνάρτηση είναι MVUE για την $g(\theta)$. Για να το κατανοήσουμε αυτό, έστω ότι $g^*[T(y)]$ είναι η μόνη συνάρτηση της $T(y)$ για την οποία $E_{\theta}\{g^*[T(Y)]\} = g(\theta)$. Έστω ότι $\hat{g}(y)$ είναι οποιαδήποτε αμερόληπτη εκτιμήτρια της $g(\theta)$. Τότε, από το θεώρημα Rao – Blackwell, η $\bar{g}[T(y)] \triangleq E_{\theta}\{\hat{g}(Y)|T(Y) = T(y)\}$ είναι αμερόληπτη για την $g(\theta)$ και είναι μια συνάρτηση της $T(y)$. Άρα από τη μοναδικότητα του g^* , θα πρέπει να έχουμε $g^* = \bar{g}$. Το θεώρημα Rao – Blackwell επίσης επιβεβαιώνει ότι $\text{Var}_{\theta}(\bar{g}[T(Y)]) \leq \text{Var}_{\theta}[\hat{g}(Y)]$. Αφού το \hat{g} είναι αυθαίρετο, βλέπουμε ότι $\text{Var}_{\theta}(g^*[T(Y)]) \leq \text{Var}_{\theta}[\hat{g}(Y)]$ για κάθε αμερόληπτη εκτίμηση της $g(\theta)$. Με άλλα λόγια, η $g^*[T(y)]$ είναι μια MVUE της $g(\theta)$.

Ορισμός IV.C.3: Πληρότητα

Η οικογένεια $\{P_{\theta}; \theta \in \Lambda\}$ θεωρείται ότι είναι *πλήρης* αν η συνθήκη $E_{\theta}\{f(Y)\} = 0$ για όλα τα $\theta \in \Lambda$ συνεπάγεται ότι $P_{\theta}[f(Y) = 0] = 1$ για όλα τα $\theta \in \Lambda$.

Αυτή η έννοια της πληρότητας είναι πολύ παρόμοια με την έννοια της πληρότητας ενός συνόλου διανυσμάτων στο \mathbb{R}^n . Για να γίνει αυτό κατανοητό, εξετάστε την κατάσταση στην οποία το Γ είναι ένα πεπερασμένο σύνολο $\{\gamma_1, \dots, \gamma_n\}$. Στην περίπτωση αυτή, για κάθε συνάρτηση f στο Γ μπορούμε να γράψουμε

$$E_{\theta}\{f(Y)\} = \underline{f}^T \underline{p}_{\theta},$$

όπου $\underline{f} = [f(\gamma_1), f(\gamma_2), \dots, f(\gamma_n)]^T$ και $\underline{p}_{\theta} = [p_{\theta}(\gamma_1), p_{\theta}(\gamma_2), \dots, p_{\theta}(\gamma_n)]^T$. Υποθέτοντας ότι $p_{\theta}(\gamma_i) > 0$ για όλα τα $\theta \in \Lambda$ και $i = 1, \dots, n$, η πληρότητα του $\{p_{\theta}; \theta \in \Lambda\}$ ορίζεται από τη συνθήκη ότι η $\underline{f}^T \underline{p}_{\theta} = 0$ για όλα τα $\theta \in \Lambda$ συνεπάγεται ότι το \underline{f} είναι το n -διάνυσμα όλων των μηδέν. Δηλαδή, η $\{\underline{p}_{\theta}; \theta \in \Lambda\}$ είναι πλήρης αν το $\underline{0}$ είναι το μόνο διάνυσμα που είναι ορθογώνιο σε όλα τα διανύσματα $\{\underline{p}_{\theta}; \theta \in \Lambda\}$. Αυτή είναι φυσικά η συνηθισμένη έννοια της πληρότητας του συνόλου διανυσμάτων $\{\underline{p}_{\theta}; \theta \in \Lambda\}$ στο \mathbb{R}^n . (Θυμηθείτε ότι ένα πλήρες σύνολο διανυσμάτων στο \mathbb{R}^n θεωρείται ότι *αναπτύσσει* το \mathbb{R}^n). Παρόμοιες αναλογίες ισχύουν για περισσότερους γενικούς χώρους παρατήρησης.

Για να κατανοήσετε περαιτέρω την έννοια της πληρότητας, εξετάστε το ακόλουθο παράδειγμα.

Παράδειγμα IV.C.2: Πληρότητα της Διωνυμικής Κατανομής

Υποθέστε ότι $\Gamma = \{0, 1, \dots, n\}$, $\Lambda = (0, 1)$, και

$$p_{\theta}(y) = \frac{n!}{y!(n-y)!} \theta^y (1-\theta)^{n-y}, \quad y = 0, \dots, n, \quad 0 < \theta < 1.$$

Για κάθε συνάρτηση f στο Γ έχουμε

$$\begin{aligned} E_{\theta}\{f(Y)\} &= \sum_{y=0}^n \frac{n!}{y!(n-y)!} f(y) \theta^y (1-\theta)^{n-y}, \\ &= (1-\theta)^n \sum_{y=0}^n a_y x^y, \end{aligned}$$

όπου

$$a_y \triangleq \frac{n!}{y!(n-y)!} f(y), \quad \text{for } y = 0, \dots, n,$$

και

$$x \triangleq \theta / (1-\theta).$$

Η συνθήκη $E_{\theta}\{f(Y)\} = 0$ για όλα τα $\theta \in \Lambda$ είναι ισοδύναμη με τη συνθήκη

$$\sum_{y=0}^n a_y x^y = 0, \quad \text{for all } x > 0. \quad (\text{IV.C.7})$$

Η συνάρτηση $\sum_{y=0}^n a_y x^y$ είναι n -οστής σειρά πολυωνυμική και άρα έχει τα περισσότερα n μηδέν της εκτός και αν όλοι οι συντελεστές της είναι μηδέν. Έπεται ότι το (IV.C.7) μπορεί να ικανοποιηθεί μόνο με $f(y) = 0, y = 0, \dots, n$. Άρα η $\{p_{\theta}; \theta \in \Lambda\}$ είναι πλήρης. Σημειώστε ότι η πληρότητα διατηρείται εδώ για κάθε Λ που περιλαμβάνει τουλάχιστον $(n+1)$ μη μηδενικές τιμές παραμέτρου.

Οι έννοιες της πληρότητας και της επάρκειας συνδέονται στενά. Για να το κατανοήσετε αυτό, υποθέστε ότι η T είναι επαρκής για την πλήρη οικογένεια $\{P_\theta; \theta \in \Lambda\}$ και για λόγους ευκολίας υποθέστε ότι $E_\theta\{|Y|\} < \infty$ για κάθε $\theta \in \Lambda$. Ορίστε μια συνάρτηση $f(y)$ με

$$f(y) = y - E_\theta\{Y|T(Y) = T(y)\}.$$

Παρατηρείτε ότι το f δεν εξαρτάται από το θ αφού η T είναι επαρκής. Για κάθε $\theta \in \Lambda$, έχουμε

$$\begin{aligned} E_\theta\{f(Y)\} &= E_\theta\{Y\} - E_\theta\{E_\theta\{Y|T(Y)\}\} \\ &= E_\theta\{Y\} - E_\theta\{Y\} = 0. \end{aligned}$$

Άρα η πληρότητα του $\{P_\theta; \theta \in \Lambda\}$ συνεπάγεται ότι $P_\theta\{Y = E_\theta\{Y|T(Y)\}\} = 1$ για όλα τα $\theta \in \Lambda$, ή ουσιαστικά, ότι $y = E_\theta\{Y|T(Y) = T(y)\}$. Αφού

η $E_\theta\{Y|T(Y) = T(y)\}$ είναι μια συνάρτηση του $T(y)$, η τελευταία συνθήκη συνεπάγεται ότι το ίδιο το y είναι μια συνάρτηση του $T(y)$. Καθώς το $T(y)$ είναι προφανώς μια συνάρτηση του y , βλέπουμε ότι το $T(y)$ πρέπει να είναι μια μία προς μία συνάρτηση του y . Αυτό σημαίνει ότι το T είναι μια τετριμμένη επαρκής στατιστική. Συμπεραίνουμε τότε ότι αν η $\{P_\theta; \theta \in \Lambda\}$ είναι πλήρης, τότε δεν υπάρχει μη τετριμμένη επαρκής στατιστική για το θ . Δηλ. η παρατήρηση Y δεν μπορεί να αναχθεί χωρίς καταστροφή των πληροφοριών σχετικά με το θ .

Η πληρότητα είναι μια χρήσιμη έννοια στο χαρακτηρισμό των MVUE. Για να γίνει αυτό κατανοητό, υποθέστε ότι η T είναι επαρκής για το θ και έστω ότι το Q_θ συμβολίζει την κατανομή του $T(Y)$ όταν $Y \sim P_\theta$. Αν η $\{Q_\theta; \theta \in \Lambda\}$ είναι πλήρης, τότε η T θεωρείται ότι είναι μια *πλήρης επαρκής στατιστική*³. Υποθέστε ότι η T είναι πλήρης και έστω ότι οι $\tilde{g}[T(y)]$ και $g^*[T(y)]$ είναι οποιεσδήποτε συναρτήσεις του $T(y)$ που είναι αμερόληπτες εκτιμήτριες συναρτήσεις του $g(\theta)$.

Έχουμε

$$\begin{aligned} E_\theta\{\tilde{g}[T(Y)] - g^*[T(Y)]\} &= E_\theta\{\tilde{g}[T(Y)]\} - E_\theta\{g^*[T(Y)]\} \\ &= g(\theta) - g(\theta) = 0 \end{aligned}$$

για όλα τα $\theta \in \Lambda$. Άρα, μέσω της πληρότητας της T , βλέπουμε ότι

$$P_\theta\{\tilde{g}[T(Y)] = g^*[T(Y)]\} = 1 \text{ για όλα τα } \theta \in \Lambda, \text{ δηλ. ότι οι } \tilde{g}[T(y)] \text{ και } g^*[T(y)] \text{ είναι}$$

η ίδια εκτιμήτρια συνάρτηση. Άρα, αφού τα \tilde{g} και g^* επιλέχθηκαν αυθαίρετα βλέπουμε ότι οποιοσδήποτε αμερόληπτος εκτιμητής που είναι συνάρτηση μιας πλήρους επαρκούς στατιστικής είναι μοναδικός από την άποψη αυτή και άρα είναι MVUE.

Βλέπουμε λοιπόν μια διαδικασία για την αναζήτηση των MVUE.

1. Βρίσκουμε μια πλήρη επαρκή στατιστική T για $\{P_\theta; \theta \in \Lambda\}$.
2. Βρίσκουμε οποιαδήποτε αμερόληπτη εκτιμήτρια συνάρτηση $\hat{g}(y)$ του $g(\theta)$
3. Τότε η $\bar{g}[T(y)] \triangleq E_\theta\{\hat{g}(Y)|T(Y) = T(y)\}$ είναι μια MVUE του $g(\theta)$

Από τα στάδια αυτά το πρώτο μοιάζει να είναι το λιγότερο ξεκάθαρο, καθώς το δεύτερο βήμα είναι συχνά αρκετά εύκολο και το τρίτο επιτυγχάνεται ευθέως μέσω λογισμού πιθανότητας. Ωστόσο, για πολλά μοντέλα που μας ενδιαφέρουν στην πράξη, το πρώτο βήμα αποδεικνύεται αρκετά εύκολο. Για να αναπτύξουμε το θέμα αυτό, παρουσιάζουμε πρώτα τον ακόλουθο ορισμό.

Ορισμός IV.C.4: Εκθετικές Οικογένειες.

Μια κλάση κατανομών $\{P_\theta; \theta \in \Lambda\}$ θεωρείται ότι είναι μια *εκθετική οικογένεια* αν υπάρχουν πραγματικής τιμής συναρτήσεις $C, Q_1, \dots, Q_m, T_1, \dots, T_m$,

και h τέτοιες ώστε το P_θ έχει πυκνότητα

$$p_\theta(y) = C(\theta) \exp \left\{ \sum_{l=1}^m Q_l(\theta) T_l(y) \right\} h(y), \quad (\text{IV.C.8})$$

για όλα τα $\theta \in \Lambda$ και $y \in \Gamma$.

Πολλές κατανομές που αντιμετωπίζονται στην πράξη μπορούν να τεθούν με τη μορφή των εκθετικών οικογενειών, όπως οι κατά Gauss, Poisson, Λαπλασιανή, διωνυμική, γεωμετρική και κάποιες πολυμεταβλητές μορφές αυτών. Οι εκθετικές οικογένειες παίζουν σημαντικό ρόλο στη θεωρία της αμερόληπτης εκτίμησης ελάχιστης διακύμανσης, δυνάμει του ακόλουθου αποτελέσματος.

3. Με βάση τα παραπάνω, βλέπουμε ότι η παρατήρηση δεν μπορεί να αναχθεί πέρα από την $T(Y)$ χωρίς την καταστροφή των πληροφοριών για το θ . Στην ουσία, η T πρέπει να είναι μια ελάχιστη επαρκής στατιστική για $\{P_\theta; \theta \in \Lambda\}$. Αυτό έπεται αφού, για κάθε επαρκή στατιστική T' πρέπει να έχουμε $T = E_\theta\{T|T'\}$ μέσω της πληρότητας. Άρα η T είναι μια συνάρτηση της T'

Πρόταση IV.C.3: Το θεώρημα πληρότητας για τις εκθετικές οικογένειες

Έστω ότι $\Gamma = \mathbb{R}^n$, $\Lambda \subset \mathbb{R}^m$ και ότι κάθε P_θ έχει πυκνότητα p_θ που δίνεται από

$$p_\theta(y) = C(\theta) \exp \left\{ \sum_{l=1}^m \theta_l T_l(y) \right\} h(y), \quad (\text{IV.C.9})$$

όπου οι C, T_1, \dots, T_m , και h είναι πραγματικής τιμής συναρτήσεις⁴. Τότε η $T(y) = [T_1(y), \dots, T_m(y)]$ είναι μια πλήρης επαρκής στατιστική για $\{P_\theta; \theta \in \Lambda\}$ αν το Λ περιέχει ένα m -διάστατο ορθογώνιο.

Περίληψη της απόδειξης: Μια πλήρης απόδειξη της Πρότασης IV.C.3 απαντά στον Lehmann (1986). Τα βήματα της απόδειξης αυτής περιγράφονται ως ακολούθως.

Πρώτα παρατηρούμε ότι η T είναι επαρκής για το θ με το θεώρημα της παραγοντοποίησης (Πρόταση IV.C.1), άρα πρέπει μόνο να αποδείξουμε την πληρότητα της T . Με το Y να κατανέμεται σύμφωνα με το (IV.C.9) βλέπουμε ευθέως ότι η $T(Y)$ θα έχει πυκνότητα στο \mathbb{R}^m της μορφής

$$g_\theta(t) = C(\theta) \exp \left\{ \sum_{l=1}^m \theta_l t_l \right\} h_T(t), \quad (\text{IV.C.10})$$

όπου η h_T είναι μια πραγματικής τιμής συνάρτηση του t . Υποθέστε ότι f είναι μια πραγματικής τιμής συνάρτηση στο \mathbb{R}^m τέτοια ώστε $E_\theta\{f[T(Y)]\} = 0$. Έχουμε

$$E_\theta\{f[T(Y)]\} = C(\theta) \int_{\mathbb{R}^m} f(t) \exp \left\{ \sum_{l=1}^m \theta_l t_l \right\} h_T(t) \mu(dt). \quad (\text{IV.C.11})$$

Έστω ότι η Λ περιλαμβάνει ένα m -διάστατο ορθογώνιο $J = \{\theta | a_1 \leq \theta_1 \leq b_1, a_2 \leq \theta_2 \leq b_2, \dots, a_m \leq \theta_m \leq b_m\}$. Με απλή μετάφραση των παραμέτρων, μπορούμε πάντα να επιλέγουμε το ορθογώνιο αυτό να είναι της μορφής $J' = \{\theta | -1 \leq \theta_1 \leq 1, -1 \leq \theta_2 \leq 1, \dots, -1 \leq \theta_m \leq 1\}$. Εξετάστε το (IV.C.11) ως μια συνάρτηση μιας περίπλοκης μεταβλητής αντικαθιστώντας το θ_l με $\theta_l + iu_l, l = 1, \dots, m$. Μπορεί να αποδειχθεί ότι αυτή η συνάρτηση είναι αναλυτική στην περιοχή $C = \{\theta + iu | -1 \leq \theta_l \leq 1, -\infty < u_l < \infty, l = 1, \dots, m\}$ και άρα η συνθήκη

4. Ας σημειωθεί ότι το (IV.C.8) μπορεί να επαναπαραμετροποιηθεί για να τεθεί στη μορφή του (IV.C.9).

ότι είναι μηδέν για όλα τα πραγματικά ορίσματα στα J' συνεπάγεται ότι είναι μηδέν σε όλη τη λωρίδα C . Συγκεκριμένα, η συνάρτηση αυτή είναι μηδέν στην περιοχή $C' = \{\theta + i\mathbf{u} | \theta_l = 0, -\infty < u_l < \infty, l = 1, \dots, m\}$, δηλ. έχουμε,

$$C(\theta) \int_{\mathbf{R}^m} f(t) \exp \left\{ i \sum_{l=1}^m u_l t_l \right\} h_T(t) \mu(dt) = 0, \quad (\text{IV.C.12})$$

για όλα τα $\mathbf{u} \in \mathbf{R}^m$. Ας σημειωθεί ότι η συνάρτηση στα αριστερά του (IV.C.12) είναι ένας πολυδιάστατος Fourier μετασχηματισμός. Καθώς αυτό είναι πανομοιότυπα μηδέν για όλα τα $\theta \in \Lambda$ συνεπάγεται ότι η συνάρτηση που μετασχηματίζεται είναι μηδέν για όλα τα $\theta \in \Lambda$ ή ισοδύναμα ότι $P_\theta\{f(Y) = 0\} = 1$ για όλα τα $\theta \in \Lambda$. Αυτό συνεπάγεται με τη σειρά του ότι η T είναι πλήρης και έτσι ολοκληρώνεται η απόδειξη της πρότασης.

Για να περιγράψουμε τη χρήση της Πρότασης IV.C.3 εξετάζουμε το ακόλουθο παράδειγμα.

Παράδειγμα IV.C.3: Αμερόληπτη Εκτίμηση Ελάχιστης Διακύμανσης Εύρους σήματος

Εξετάστε το μοντέλο

$$Y_k = N_k + \mu s_k, \quad k = 1, \dots, n$$

όπου N_1, \dots, N_n είναι i.i.d. $\mathcal{N}(0, \sigma^2)$ δείγματα θορύβου, $\mathbf{s} = (s_1, \dots, s_n)^T$ είναι γνωστό σήμα, και το μ είναι παράμετρος εύρους σήματος. Υποθέστε για τώρα ότι το σ^2 είναι γνωστό και ότι επιθυμούμε να εκτιμήσουμε την παράμετρο εύρους μ . Η πυκνότητα του \underline{Y} δίνεται από

$$\begin{aligned} & \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n (y_k - \mu s_k)^2 \right\} \\ & = C(\theta_1) \exp\{\theta_1 T_1(\underline{y})\} h(\underline{y}), \end{aligned} \quad (\text{IV.C.13})$$

όπου έχουμε ορίσει

$$\begin{aligned} \theta_1 & = \mu/\sigma^2, \\ T_1(\underline{y}) & = \sum_{k=1}^n s_k y_k, \\ C(\theta_1) & = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{\theta_1^2 \sigma^2}{2} \sum_{k=1}^n s_k^2 \right\}, \end{aligned}$$

και

$$h(\underline{y}) = \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n y_k^2 \right\}.$$

Με την παραδοχή ότι μ είναι ένας αυθαίρετος πραγματικός αριθμός, το σύνολο παραμέτρου είναι $\Lambda = \{\theta_1 | -\infty < \theta_1 < \infty\} = \mathbb{R}$. Ένα μονοδιάστατο ορθογώνιο είναι διάστημα και το Λ προφανώς περιλαμβάνει ένα διάστημα, άρα από την Πρόταση IV.C.3 και το (IV.C.13) βλέπουμε ότι η $T_1(\underline{y})$ είναι μια πλήρης επαρκής στατιστική για το θ_1 .

Θέλουμε να εκτιμήσουμε το $\mu = g(\theta) = \sigma^2 \theta_1$. Ας σημειωθεί ότι $E_\theta\{Y_1\} = \mu s_1$. Άρα, υποθέτοντας ότι $s_1 \neq 0$, η εκτίμηση $\hat{g}(\underline{y}) = y_1/s_1$ είναι ένας αμερόληπτος εκτιμητής του $g(\theta)$. Άρα, αφού η T_1 είναι πλήρης, η εκτίμηση

$$\tilde{g}[T_1(\underline{y})] = E_\theta\{\hat{g}(\underline{Y}) | T_1(\underline{Y}) = T_1(\underline{y})\} \quad (\text{IV.C.14})$$

είναι MVUE. Για να υπολογίσουμε το (IV.C.14) βλέπουμε ότι οι $\hat{g}(\underline{Y})$ και $T_1(\underline{Y})$ είναι γραμμικές συναρτήσεις του \underline{Y} το οποίο είναι κατά Gauss. Άρα οι $\hat{g}(\underline{Y})$ και $T_1(\underline{Y})$ είναι από κοινού κατά Gauss. Είναι εύκολα κατανοητό ότι

$$\begin{aligned} E_\theta\{\hat{g}(\underline{Y})\} &= \mu, \\ E_\theta\{T_1(\underline{Y})\} &= n\mu\bar{s}^2 \\ \text{Var}_\theta\{\hat{g}(\underline{Y})\} &= \sigma^2/s_1^2, \\ \text{Var}_\theta\{T_1(\underline{Y})\} &= n\sigma^2\bar{s}^2, \end{aligned}$$

και

$$\text{Cov}_\theta[\hat{g}(\underline{Y}), T_1(\underline{Y})] = \sigma^2,$$

όπου έχουμε ορίσει $\bar{s}^2 \triangleq (1/n) \sum_{k=1}^n s_k^2$. Άρα, εφαρμόζοντας τα αποτελέσματα της Ενότητας IV.B μπορούμε να γράψουμε αυτό το δεσμευμένο μέσο του (IV.C.14) ως

$$\begin{aligned} \tilde{g}[T_1(\underline{y})] &= E_\theta\{\hat{g}(\underline{Y})\} + \text{Cov}_\theta[\hat{g}(\underline{Y}), T_1(\underline{Y})] \\ &\quad \times [\text{Var}_\theta\{T_1(\underline{Y})\}]^{-1} [T_1(\underline{y}) - E_\theta\{T_1(\underline{Y})\}] \\ &= \mu + \sigma^2 (n\sigma^2\bar{s}^2)^{-1} [T_1(\underline{y}) - n\mu\bar{s}^2] \\ &= T_1(\underline{y})/n\bar{s}^2 = \left(\sum_{k=1}^n s_k y_k \right) / n\bar{s}^2. \quad (\text{IV.C.15}) \end{aligned}$$

Άρα έχουμε δημιουργήσει μια MVUE για το εύρος σήματος μ . Η διακύμανση αυτού του εκτιμητή είναι

$$\text{Var}_\theta(\tilde{g}[T_1(\underline{Y})]) = \sigma^2/n\bar{s}^2. \quad (\text{IV.C.16})$$

Ας υποθέσουμε τώρα ότι τα μ και σ^2 είναι άγνωστα, με το μ να κυμαίνεται πάνω από το \mathbb{R} και το σ^2 να κυμαίνεται πάνω από το $(0, \infty)$ και ότι θα θέλαμε να εκτιμήσουμε και τις δύο αυτές παραμέτρους. Από το (IV.C.16) βλέπουμε ότι η εκτίμηση του σ^2 μας δίνει μια εκτίμηση της ακρίβειας της εκτίμησης μας για το εύρος. Ας σημειωθεί ότι η $h(\underline{y})$, όπως ορίζεται στο (IV.C.13) είναι μια συνάρτηση του σ^2 , έτσι το (IV.C.13) όπως είναι γραμμένο δεν είναι μια ορθή εκθετική οικογένεια αν το σ^2 είναι άγνωστο. Ωστόσο, μπορούμε να ξαναγράψουμε την πυκνότητα ως

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n (y_k - \mu s_k)^2 \right\} \\ = C(\theta) \exp \{ \theta_1 T_1(\underline{y}) + \theta_2 T_2(\underline{y}) \} h(\underline{y}), \quad (\text{IV.C.17})$$

όπου το θ_1 και T_1 είναι όπως στο (IV.C.13), αλλά, τώρα ορίζουμε $\theta = (\theta_1, \theta_2)$,

$$\theta_2 = -\frac{1}{2\sigma^2}, \\ T_2(\underline{y}) = \sum_{k=1}^n y_k^2, \\ C(\theta) = \left(-\frac{\theta_2}{\pi} \right)^{n/2} \exp \left\{ \frac{\theta_1^2}{4\theta_2} \sum_{k=1}^n s_k^2 \right\},$$

και

$$h(\underline{y}) \equiv 1.$$

Η κύμανση $\{(\mu, \sigma^2) | \mu \in \mathbb{R}, \sigma^2 > 0\}$ αντιστοιχεί στο

$\Lambda = \{(\theta_1, \theta_2) | \theta_1 \in \mathbb{R}, \theta_2 < 0\}$, που σίγουρα περιλαμβάνει ένα ορθογώνιο. Άρα η $T = (T_1, T_2)$ είναι μια πλήρης επαρκής στατιστική για το θ .

Θέλουμε να εκτιμήσουμε τα $\sigma^2 = g_2(\theta) \triangleq -1/2\theta_2$, και $\sigma^2 = g_2(\theta) \triangleq -1/2\theta_2$. Σημειώστε ότι η εκτίμηση στο (IV.C.15) έχει υπολογιστεί χωρίς γνώση του σ^2 , είναι αμερόληπτη και είναι μια συνάρτηση του $T_1(\underline{y})$ [και ως εκ τούτου του $T(\underline{y})$]. Άρα είναι μια MVUE του μ , ακόμα κι όταν το σ^2 είναι άγνωστο.

Για να βρούμε μια MVUE για το σ^2 θα πρέπει πρώτα να αναζητήσουμε έναν αμερόληπτο εκτιμητή του σ^2 και κατόπιν να τον δεσμεύσουμε στο $T(\underline{y})$. Είναι απλούστερο σ' αυτήν την περίπτωση, ωστόσο, να ψάξουμε ευθέως για μια αμερόληπτη συνάρτηση του T .

Συγκεκριμένα, παρατηρούμε ότι αφού $T_1(\underline{Y}) \sim \mathcal{N}(n\mu\bar{s}^2, n\sigma^2\bar{s}^2)$, έχουμε

$$\begin{aligned} E_{\theta}\{T_1^2(\underline{Y})\} &\equiv \text{Var}_{\theta}[T_1(\underline{Y})] + (E_{\theta}\{T_1(\underline{Y})\})^2 \\ &= n\sigma^2\bar{s}^2 + n^2\mu^2(\bar{s}^2)^2. \end{aligned}$$

Επίσης, έχουμε ότι

$$\begin{aligned} E_{\theta}\{T_2(\underline{Y})\} &= \sum_{k=1}^n E_{\theta}\{Y_k^2\} = \sum_{k=1}^n (\sigma^2 + \mu^2 s_k^2) \\ &= n\sigma^2 + n\mu^2\bar{s}^2. \end{aligned}$$

Από αυτά τα δύο αποτελέσματα βλέπουμε ότι το μέγεθος $[T_2(\underline{Y}) - T_1^2(\underline{Y})/ns^2]$ έχει μέσο

$$E_{\theta}\{T_2(\underline{Y})\} - E_{\theta}\{T_1^2(\underline{Y})/ns^2\} = (n-1)\sigma^2. \quad (\text{IV.C.18})$$

Άρα η συνάρτηση $\tilde{g}_2[T(\underline{y})] = [T_2(\underline{y}) - T_1^2(\underline{y})/ns^2]/(n-1)$ είναι μια αμερόληπτη εκτιμήτρια του σ^2 και με την πληρότητα της T είναι μια MVUE. Μπορούμε να ξαναγράψουμε το \tilde{g}_2 ως

$$\tilde{g}_2[T(\underline{y})] = \frac{1}{n-1} \sum_{k=1}^n (y_k - \hat{\mu}s_k)^2 \triangleq \hat{\sigma}^2, \quad (\text{IV.C.19})$$

όπου το $\hat{\mu}$ είναι η MVUE του μ από το (IV.C.15). Σημειώστε ότι η $\hat{\eta}_k \triangleq y_k - \hat{\mu}s_k$ είναι μια εκτίμηση του θορύβου το k -οστού δείγματος, άρα το $\hat{\sigma}^2$ εκτιμά τη διακύμανση (που ισούται με τη ροπή δεύτερης τάξης) του θορύβου με $[1/(n-1)] \sum_{k=1}^n (\hat{\eta}_k)^2$. Ας σημειωθεί ότι ένας πιο φυσικός εκτιμητής για τη ροπή δεύτερης τάξης θα ήταν $(1/n) \sum_{k=1}^n (\hat{\eta}_k)^2$; αλλά όπως βλέπουμε από την παραπάνω ανάλυση, η τελευταία εκτίμηση είναι μεροληπτική. Περαιτέρω ανάλυση του θέματος αυτού γίνεται στην ενότητα IV.D.

Η θεωρία που περιγράφεται στις παραπάνω παραγράφους παρέχει ένα μέσο αναζήτησης των αμερόληπτων εκτιμητών ελάχιστης διακύμανσης. Για πολλά μοντέλα που μας ενδιαφέρουν, ωστόσο, η δομή που απαιτείται για την εφαρμογή αποτελεσμάτων όπως αυτών στην Πρόταση IV.C.3 δεν είναι παρούσα. Άρα, συχνά αντιμετωπίζουμε το πρόβλημα να προτείνουμε έναν εκτιμητή και να αξιολογούμε την απόδοση του (δηλ., τη μεροληψία και τη διακύμανσή του) εν απουσία οποιασδήποτε γνώσης σχετικά με τη βελτιστότητα του εκτιμητή. Σε τέτοιες περιπτώσεις είναι χρήσιμο να έχουμε ένα πρότυπο με το οποίο μπορούν να συγκριθούν οι εκτιμητές. Δηλ. θα ήταν χρήσιμο να γνωρίζουμε τους βασικούς περιορισμούς της απόδοσης εκτίμησης που τίθενται από ένα δεδομένο μοντέλο. Ένα τέτοιο πρότυπο παρέχεται εν μέρει από το ακόλουθο αποτέλεσμα.

Πρόταση IV.C.4: Η Ανισότητα της Πληροφορίας

Έστω ότι $\hat{\theta}$ είναι μια εκτίμηση της παραμέτρου θ σε μια οικογένεια $\{P_\theta; \theta \in \Lambda\}$ και ότι ισχύουν οι ακόλουθες συνθήκες:

- (1) Το Λ είναι ένα ανοιχτό διάστημα
- (2) Η οικογένεια $\{P_\theta; \theta \in \Lambda\}$ έχει μια αντίστοιχη οικογένεια πυκνοτήτων $\{p_\theta; \theta \in \Lambda\}$, όλα τα μέλη της οποίας έχουν την ίδια υποστήριξη⁵
- (3) Η $\partial p_\theta(y)/\partial \theta$ υπάρχει και είναι πεπερασμένη για όλα τα $\theta \in \Lambda$ και όλα τα y στην στήριξη του p_θ .
- (4) Η $\partial \int_\Gamma h(y)p_\theta(y)\mu(dy)/\partial \theta$ υπάρχει και ισούται με $\int_\Gamma h(y)[\partial p_\theta(y)/\partial \theta]\mu(dy)$, για όλα τα $\theta \in \Lambda$ για $h(y) = \hat{\theta}(y)$ και $h(y) = 1$

Τότε

$$\text{Var}_\theta[\hat{\theta}(Y)] \geq \frac{\left[\frac{\partial}{\partial \theta} E_\theta\{\hat{\theta}(Y)\} \right]^2}{I_\theta} \quad (\text{IV.C.20})$$

όπου

$$I_\theta \triangleq E_\theta \left\{ \left(\frac{\partial}{\partial \theta} \log p_\theta(Y) \right)^2 \right\}. \quad (\text{IV.C.21})$$

Επιπλέον αν η ακόλουθη συνθήκη ισχύει

- (5) Το $\partial^2 p_\theta(y)/\partial \theta^2$ ισχύει για όλα τα $\theta \in \Lambda$ και y στην υποστήριξη του p_θ και

$$\int \frac{\partial^2}{\partial \theta^2} p_\theta(y)\mu(dy) = \frac{\partial^2}{\partial \theta^2} \int p_\theta(y)\mu(dy),$$

Τότε το I_θ μπορεί να υπολογιστεί μέσω

$$I_\theta = -E_\theta \left\{ \frac{\partial^2}{\partial \theta^2} \log p_\theta(Y) \right\}. \quad (\text{IV.C.22})$$

Απόδειξη: Η απόδειξη του αποτελέσματος αυτού έπεται ευθέως από την ανισότητα Schwarz. Συγκεκριμένα, έχουμε ότι

$$E_{\theta}\{\hat{\theta}(Y)\} = \int_{\Gamma} \hat{\theta}(y) p_{\theta}(y) \mu(dy). \quad (\text{IV.C.23})$$

Διαφορίζοντας το (IV.C.23) και εφαρμόζοντας τη συνθήκη (4), έχουμε

$$\frac{\partial}{\partial \theta} E_{\theta}\{\hat{\theta}(Y)\} = \int_{\Gamma} \hat{\theta}(y) \frac{\partial}{\partial \theta} p_{\theta}(y) \mu(dy).$$

Η συνθήκη (4) επίσης συνεπάγεται ότι

$$\int_{\Gamma} \frac{\partial}{\partial \theta} p_{\theta}(y) \mu(dy) = \frac{\partial}{\partial \theta} \int_{\Gamma} p_{\theta}(y) \mu(dy) = \frac{\partial}{\partial \theta} (1) = 0,$$

ώστε να έχουμε

$$\begin{aligned} \frac{\partial}{\partial \theta} E_{\theta}\{\hat{\theta}(Y)\} &= \int_{\Gamma} (\hat{\theta}(y) - E_{\theta}\{\hat{\theta}(Y)\}) \frac{\partial}{\partial \theta} p_{\theta}(y) \mu(dy) \\ &= \int_{\Gamma} (\hat{\theta}(y) - E_{\theta}\{\hat{\theta}(Y)\}) \left[\frac{\partial}{\partial \theta} \log p_{\theta}(y) \right] p_{\theta}(y) \mu(dy) \\ &= E_{\theta} \left\{ [\hat{\theta}(Y) - E_{\theta}\{\hat{\theta}(Y)\}] \left[\frac{\partial}{\partial \theta} \log p_{\theta}(Y) \right] \right\}, \end{aligned} \quad (\text{IV.C.24})$$

όπου η δεύτερη ισότητα έπεται από το γεγονός ότι

$\frac{\partial \log p_{\theta}(y)}{\partial \theta} = [\partial p_{\theta}(y) / \partial \theta] / p_{\theta}(y)$. Εφαρμόζοντας την ανισότητα Schwarz στο (IV.C.24), έχουμε

$$\left(\frac{\partial}{\partial \theta} E_{\theta}\{\hat{\theta}(Y)\} \right)^2 \leq E_{\theta}\{[\hat{\theta}(Y) - E_{\theta}\{\hat{\theta}(Y)\}]^2\} I_{\theta}, \quad (\text{IV.C.25})$$

όπου το I_{θ} είναι από το (IV.C.21). Παρατηρώντας ότι

$$E_{\theta}\{[\hat{\theta}(Y) - E_{\theta}\{\hat{\theta}(Y)\}]^2\} = \text{Var}_{\theta}[\hat{\theta}(Y)], \text{ το (IV.C.20) έπεται.}$$

5. Δηλ. το σύνολο $\{y | p_{\theta}(y) > 0\}$ είναι ίδιο για όλα τα $\theta \in \Lambda$

Για να δούμε το (IV.C.22) παρατηρούμε ότι

$$\frac{\partial^2}{\partial \theta^2} \log p_{\theta}(Y) = \left(\frac{\partial^2}{\partial \theta^2} p_{\theta}(Y) / p_{\theta}(Y) \right) - \left(\frac{\partial}{\partial \theta} \log p_{\theta}(Y) \right)^2. \quad (\text{IV.C.26})$$

Λαμβάνοντας $E_{\theta} \{ \cdot \}$ και στις δύο πλευρές του (IV.C.26) και αναδιατάσσοντας αποδόσεις

$$I_{\theta} = -E_{\theta} \left(\frac{\partial^2}{\partial \theta^2} \log p_{\theta}(Y) \right) - \int_{\Gamma} \frac{\partial^2}{\partial \theta^2} p_{\theta}(y) \mu(dy).$$

Χρησιμοποιώντας τη συνθήκη (5) έχουμε

$$\int_{\Gamma} \frac{\partial^2}{\partial \theta^2} p_{\theta}(y) \mu(dy) = \frac{\partial^2}{\partial \theta^2} \int_{\Gamma} p_{\theta}(y) \mu(dy) = \frac{\partial^2}{\partial \theta^2} (1) = 0,$$

και το (IV.C.22) έπεται.

Το μέγεθος I_{θ} που ορίζεται το (IV.C.21) είναι γνωστό ως *πληροφορία Fisher* για την εκτίμηση του θ από το Y , και το (IV.C.20) ονομάζεται *ανισότητα της πληροφορίας*. Όσο μεγαλύτερη είναι αυτή η μέτρηση πληροφορίας για ένα δεδομένο μοντέλο, τόσο καλύτερο είναι το κατώτερο φράγμα στην ακρίβεια εκτίμησης που δίνεται από την ανισότητα της πληροφορίας. Η ύπαρξη μιας εκτίμησης που επιτυγχάνει ισότητα στην ανισότητα πληροφορίας είναι δυνατή μόνο υπό ειδικές συνθήκες [βλέπε, π.χ. Lehmann (1983) και την παρακάτω ανάλυση]. Για τη συγκεκριμένη περίπτωση στην οποία το $\hat{\theta}$ είναι αμερόληπτο $[E_{\theta}\{\hat{\theta}(Y)\} = \theta]$, η ανισότητα πληροφορίας ανάγεται σε

$$\text{Var}_{\theta}[\hat{\theta}(Y)] \geq \frac{1}{I_{\theta}}, \quad (\text{IV.C.27})$$

ένα αποτέλεσμα γνωστό ως *κατώτατο φράγμα Cramer – Rao (CRLB)*.

Παραδείγματα που περιγράφουν την ανισότητα της πληροφορίας σε συγκεκριμένα προβλήματα εκτίμησης θα συζητηθούν στην ακόλουθη ενότητα. Το ακόλουθο γενικό παράδειγμα περιγράφει αναλυτικότερα το ρόλο των εκθετικών οικογενειών στην εκτίμηση παραμέτρου.

Παράδειγμα IV.C.4: Η Ανισότητα πληροφορίας για τις εκθετικές οικογένειες

Έστω ότι Λ είναι ένα ανοιχτό διάστημα και το $p_{\theta}(y)$ δίνεται από

$$p_{\theta}(y) = C(\theta) e^{g(\theta)T(y)} h(y), \quad (\text{IV.C.28})$$

όπου οι C, g, T και h είναι πραγματικής τιμής συναρτήσεις των ορισμάτων τους και όπου η $g(\theta)$ έχει παραγωγό τη $g'(\theta)$. Υποθέτοντας ότι $E_{\theta}\{|T(Y)|\} < \infty$ και

$$\frac{\partial}{\partial \theta} \int_{\Gamma} e^{g(\theta)T(y)} h(y) \mu(dy) = \int_{\Gamma} \frac{\partial}{\partial \theta} e^{g(\theta)T(y)} h(y) \mu(dy),$$

οι συνθήκες (1) – (4) της Πρότασης IV.C.4 ισχύουν. Αφού η $p_{\theta}(y)$ πρέπει να ολοκληρωθεί στη μονάδα, μπορούμε να γράψουμε $C(\theta) = \left[\int_{\Gamma} e^{g(\theta)T(y)} h(y) \mu(dy) \right]^{-1}$. Για να υπολογίσουμε το I_{θ} γι' αυτή την οικογένεια πυκνοτήτων, γράφουμε

$$\begin{aligned} \log p_{\theta}(y) &= g(\theta)T(y) + \log h(y) \\ &\quad - \log \left[\int_{\Gamma} e^{g(\theta)T(y)} h(y) \mu(dy) \right]. \end{aligned}$$

Με διαφορισμό έχουμε

$$\begin{aligned} \frac{\partial}{\partial \theta} \log p_{\theta}(y) &= g'(\theta)T(y) - \frac{g'(\theta) \int_{\Gamma} T(y) e^{g(\theta)T(y)} h(y) \mu(dy)}{\int_{\Gamma} e^{g(\theta)T(y)} h(y) \mu(dy)} \\ &= g'(\theta)[T(y) - E_{\theta}\{T(Y)\}]. \end{aligned}$$

Άρα

$$\begin{aligned} I_{\theta} &\triangleq E_{\theta} \left\{ \left(\frac{\partial}{\partial \theta} \log p_{\theta}(Y) \right)^2 \right\} = [g'(\theta)]^2 E_{\theta} \{ [T(Y) - E_{\theta}\{T(Y)\}]^2 \} \\ &= [g'(\theta)]^2 \text{Var}_{\theta}[T(Y)], \end{aligned}$$

Και η ανισότητα πληροφορίας στην περίπτωση αυτή είναι

$$\text{Var}_{\theta}[\hat{\theta}(Y)] \geq \frac{\left[\frac{\partial}{\partial \theta} E_{\theta}\{\hat{\theta}(Y)\} \right]^2}{[g'(\theta)]^2 \text{Var}_{\theta}[T(Y)]}. \quad (\text{IV.C.29})$$

Εστω ότι θεωρούμε την $T(y)$ ως εκτιμήτρια του θ . Τότε έχουμε

$$E_{\theta}\{T(Y)\} = \frac{\int_{\Gamma} T(y) e^{g(\theta)T(y)} h(y) \mu(dy)}{\int_{\Gamma} e^{g(\theta)T(y)} h(y) \mu(dy)}. \quad (\text{IV.C.30})$$

Διαφορίζοντας το (IV.C.30) έχουμε ευθέως ότι

$$\frac{\partial}{\partial \theta} E_{\theta}\{T(Y)\} = g'(\theta) \text{Var}_{\theta}[T(Y)],$$

και άρα το (IV.C.29) συνεπάγεται ότι το κατώτερο φράγμα στην ανισότητα πληροφορίας ισούται με

$$\frac{[\frac{\partial}{\partial \theta} E_{\theta}\{T(Y)\}]^2}{[g'(\theta)]^2 \text{Var}_{\theta}[T(Y)]} = \text{Var}_{\theta}[T(Y)]. \quad (\text{IV.C.31})$$

Από το (IV.C.31) βλέπουμε ότι η $T(Y)$ επιτυγχάνει το κατώτερο φράγμα πληροφορίας, άρα έχει την ελάχιστη διακύμανση μεταξύ όλων των εκτιμητών $\hat{\theta}$ που ικανοποιούν την $\partial E_{\theta}\{\hat{\theta}(Y)\}/\partial \theta = \partial E_{\theta}\{T(Y)\}/\partial \theta$. Συγκεκριμένα, αν η T είναι αμερόληπτη για το θ , τότε είναι μια MVUE, ένα δεδομένο που γνωρίζουμε ήδη από το γεγονός ότι η T είναι μια πλήρης επαρκής στατιστική για το θ στην περίπτωση αυτή.

Βλέπουμε ότι η εκθετική μορφή του (IV.C.28) είναι επαρκής για τη διακύμανση του T ώστε να επιτύχει το κατώτερο φράγμα πληροφορίας στο πλαίσιο της κανονικότητας που υποτέθηκε παραπάνω. Αποδεικνύεται ότι αυτή η μορφή είναι επίσης απαραίτητη για την επίτευξη του κατώτερου φράγματος για όλα τα $\theta \in \Lambda$, και πάλι σε συνθήκες κανονικότητας.

Συγκεκριμένα, παρατηρούμε ότι ένας εκτιμητής $\hat{\theta}$ έχει διακύμανση ίση με το κατώτερο φράγμα πληροφορίας για όλα τα $\theta \in \Lambda$ αν και μόνο αν έχουμε ισότητα στην ανισότητα Schwarz που εφαρμόζεται στο (IV.C.25). Αυτό, με τη σειρά του, θα συμβεί, αν και μόνο αν

$$\frac{\partial}{\partial \theta} \log p_{\theta}(Y) = k(\theta)[\hat{\theta}(Y) - E_{\theta}\{\hat{\theta}(Y)\}]$$

με πιθανότητα 1 ως προς το P_{θ} , για κάποια $k(\theta)$. Αν (a, b) συμβολίζει το Λ και $f(\theta)$ συμβολίζει το $E_{\theta}\{\hat{\theta}(Y)\}$, συμπεραίνουμε ότι το $\hat{\theta}$ επιτυγχάνει το φράγμα πληροφορίας αν και μόνο αν

$$p_{\theta}(y) = h(y) \exp \left\{ \int_a^{\theta} k(\sigma)[\hat{\theta}(y) - f(\sigma)] d\sigma \right\}, \quad y \in \Gamma, \quad (\text{IV.C.32})$$

όπου το $h(y)$ δεν εξαρτάται από το θ . Η ισότητα του (IV.C.32) θα αναγνωριστεί ως η εκθετική μορφή του (IV.C.28) με το h ως δεδομένο,

$$C(\theta) = \exp \left\{ - \int_a^{\theta} k(\sigma) f(\sigma) d\sigma \right\},$$

$$g(\theta) = \int_a^{\theta} k(\sigma) d\sigma,$$

και

$$T(y) = \hat{\theta}(y).$$

[Ας σημειωθεί ότι το $k(\theta)$ πρέπει να ισούται με $I_{\theta}/[\partial E_{\theta}\{\hat{\theta}(Y)\}/\partial \theta]$ στην κατάσταση αυτή, όπως είναι προφανές και από την αντικατάσταση του (IV.C.32) με (IV.C.34)]. Άρα

συμπεραίνουμε ότι, στο πλαίσιο της κανονικότητας, το κατώτερο φράγμα πληροφορίας επιτυγχάνεται με το $\hat{\theta}$, αν και μόνο αν $\hat{\theta}(y) = T(y)$ σε μια εκθετική οικογένειας μίας παραμέτρου.

IV.D Εκτίμηση Μέγιστης Πιθανοφάνειας

Για πολλά μοντέλα παρατήρησης που προκύπτουν στην πράξη, δεν είναι δυνατή η εφαρμογή των αποτελεσμάτων της Ενότητας IV.C. για την εύρεση των MVUE, είτε λόγω μη επιλυσιμότητας της απαιτούμενης ανάλυσης, είτε λόγω έλλειψης μιας χρήσιμης πλήρους επαρκούς στατιστικής. Για τέτοιου είδους μοντέλα, απαιτείται μια εναλλακτική μέθοδος αναζήτησης καλών εκτιμητών. Μια ευρέως χρησιμοποιούμενη μέθοδος σχεδιασμού εκτιμητών είναι η μέθοδος μέγιστης πιθανοφάνειας, που αποτελεί το αντικείμενο της ενότητας αυτής.

Για να εκκινήσουμε την εκτίμηση μέγιστης πιθανοφάνειας, εξετάζουμε αρχικά την εκτίμηση MAP, στην οποία αναζητούμε το $\hat{\theta}_{MAP}(y)$ που δίνεται από

$$\hat{\theta}_{MAP}(y) = \arg\{\max_{\theta \in \Lambda} p_{\theta}(y)w(\theta)\}. \quad (IV.D.1)$$

Ελλείψει οποιασδήποτε πρότερης πληροφορίας για την παράμετρο, μπορούμε να υποθέσουμε ότι κατανέμεται ομοιόμορφα στην κύμανσή της [δηλ., η $w(\theta)$ είναι σταθερά στο Λ] καθώς αυτό αναπαριστά, λίγο ή πολύ, ένα πρότερο χειρότερης περίπτωσης. Στην περίπτωση αυτή η MAP εκτίμηση για δεδομένο $y \in \Gamma$ είναι οποιαδήποτε τιμή του θ που μεγιστοποιεί την $p_{\theta}(y)$ στο Λ . Αφού η $p_{\theta}(y)$ ως συνάρτηση του θ καλείται ενίοτε *συνάρτηση πιθανοφάνειας* [άρα το $p_1(y)/p_0(y)$ είναι ο λόγος πιθανοφάνειας], η εκτίμηση αυτή ονομάζεται *εκτίμηση μέγιστης πιθανοφάνειας* (MLE). Συμβολίζοντας αυτή την εκτίμηση με $\hat{\theta}_{ML}$, έχουμε

$$\hat{\theta}_{ML}(y) = \arg\{\max_{\theta \in \Lambda} p_{\theta}(y)\}. \quad (IV.D.2)$$

Δύο πράγματα δεν είναι σωστά με το παραπάνω ζήτημα. Πρώτον, δεν είναι πάντα δυνατό να δημιουργήσουμε μια ομοιόμορφη κατανομή στο Λ , αφού το Λ μπορεί να μην είναι φραγμένο σύνολο. Δεύτερον και σημαντικότερο, η υπόθεση για ένα ομοιόμορφο πρότερο για την παράμετρο διαφέρει από την υπόθεση ότι το πρότερο είναι άγνωστο ή ότι η παράμετρος δεν είναι μια τυχαία μεταβλητή. Ωστόσο, η εκτίμηση μέγιστης πιθανοφάνειας αποδεικνύεται πολύ χρήσιμη σε πολλές καταστάσεις, και όπως θα δούμε στην ενότητα αυτή, η χρήση της μπορεί να ενεργοποιηθεί με άλλους, πιο ευθείς, τρόπους. Επιπλέον, η εύρεση της τιμής του θ που καθιστά τις παρατηρήσεις πιο πιθανές είναι ένα θεμιτό κριτήριο από μόνο του.

Η μεγιστοποίηση της $p_{\theta}(y)$ είναι ισοδύναμη με τη μεγιστοποίηση του $\log p_{\theta}(y)$ και υποθέτοντας επαρκή λειότητα της συνάρτησης αυτής, μια απαραίτητη συνθήκη για την εκτίμηση μέγιστης πιθανοφάνειας είναι

$$\left. \frac{\partial}{\partial \theta} \log p_{\theta}(y) \right|_{\theta = \hat{\theta}_{ML}(y)} = 0. \quad (IV.D.3)$$

Η εξίσωση του (IV.D.3) είναι γνωστή ως *εξίσωση πιθανοφάνειας* και θα δούμε ότι οι λύσεις της έχουν χρήσιμες ιδιότητες ακόμη κι όταν δεν είναι μέγιστες της $p_{\theta}(y)$. Για παράδειγμα, έστω ότι έχουμε ισότητα στο κατώτερο φράγμα Cramer – Rao (IV.C.27). Δηλ. υποθέστε ότι $\hat{\theta}$ είναι μια αμερόληπτη εκτίμηση του θ με $\text{Var}_{\theta}[\hat{\theta}(Y)] = 1/I_{\theta}$. (Σημειώστε ότι ένα τέτοιο $\hat{\theta}$ είναι MVUE του θ .) Τότε από το (IV.C.32) βλέπουμε ότι το $\log p_{\theta}(y)$ πρέπει να είναι της μορφής

$$\log p_{\theta}(y) = \int_a^{\theta} I_{\sigma}[\hat{\theta}(y) - \sigma] d\sigma + \log h(y), \quad (\text{IV.D.4})$$

όπου έχουμε χρησιμοποιήσει τα δεδομένα ότι $f(\theta) = \theta$ και $k(\theta) = I_{\theta}/f'(\theta)$. Από το (IV.D.4), η εξίσωση πιθανοφάνειας γίνεται

$$\frac{\partial}{\partial \theta} \log p_{\theta}(y) \Big|_{\theta=\hat{\theta}_{ML}(y)} = I_{\theta}[\hat{\theta}(y) - \theta] \Big|_{\theta=\hat{\theta}_{ML}(y)} = 0, \quad (\text{IV.D.5})$$

που έχει τη λύση $\hat{\theta}_{ML}(y) = \hat{\theta}(y)$. Άρα, συμπεραίνουμε ότι αν το θ επιτυγχάνει την CRLB, είναι η λύση στην εξίσωση πιθανοφάνειας. Με άλλα λόγια, μόνο λύσεις της εξίσωσης πιθανοφάνειας μπορούν να πετύχουν την CRLB. Δυστυχώς, δεν είναι πάντα αληθές ότι οι λύσεις της εξίσωσης πιθανοφάνειας θα επιτύχουν την CRLB ούτε ότι είναι αμερόληπτες. [Ωστόσο, όταν ο $\log p_{\theta}$ έχει τη μορφή του (IV.D.4), αυτό θα συμβεί]. Επίσης, όταν η λύση της εξίσωσης πιθανοφάνειας δεν ικανοποιεί την CRLB, μπορεί να υπάρχουν άλλοι εκτιμητές με την ίδια μεροληψία που έχουν μικρότερη διακύμανση από το $\hat{\theta}_{ML}$.

Από την παραπάνω συζήτηση, βλέπουμε ότι η λύση της εξίσωσης πιθανοφάνειας μπορεί ενίοτε να είναι μια MVUE. Για την περίπτωση στην οποία ο χώρος παρατήρησης είναι \mathbb{R}^n με το Y να αποτελείται από i.i.d. στοιχεία, σημαίνει ότι στο πλαίσιο της κανονικότητας, η λύση στην εξίσωση πιθανοφάνειας είναι αμερόληπτη και επιτυγχάνει την CRLB ασυμπτωτικά καθώς $n \rightarrow \infty$. Πριν τη μελέτη των ασυμπτωτικών αυτών ιδιοτήτων, δίνουμε τα ακόλουθα δύο παραδείγματα για να περιγράψουμε την προσέγγιση μέγιστης πιθανοφάνειας.

Παράδειγμα IV.D.1: Εκτίμηση Μέγιστης Πιθανοφάνειας της Παραμέτρου Εκθετικής Κατανομής

Έστω ότι $\Gamma = \mathbb{R}^n$, $\Lambda = (0, \infty)$, και Y_1, \dots, Y_n είναι i.i.d. εκθετικές τυχαίες μεταβλητές με παράμετρο θ , δηλ. $p_{\theta}(\underline{y}) = \prod_{k=1}^n f_{\theta}(y_k)$ με

$$f_{\theta}(y_k) = \begin{cases} \theta e^{-\theta y_k} & \text{if } y_k \geq 0 \\ 0 & \text{if } y_k < 0. \end{cases} \quad (\text{IV.D.6})$$

Έχουμε $p_{\theta}(\underline{y}) = \theta^n \exp\{-\theta n\bar{y}\}$ με $\bar{y} \triangleq (1/n) \sum_{k=1}^n y_k$, οπότε η εξίσωση πιθανοφάνειας είναι

$$\frac{\partial}{\partial \theta} \log p_{\theta}(\underline{y}) \Big|_{\theta=\hat{\theta}_{ML}(\underline{y})} = \frac{n}{\theta} - n\bar{y} \Big|_{\theta=\hat{\theta}_{ML}(\underline{y})} = 0, \quad (\text{IV.D.7})$$

που έχει τη μοναδική λύση $\hat{\theta}_{ML}(\underline{y}) = 1/\bar{y}$. Αφού $\partial^2 \log p_{\theta}(\underline{y})/\partial \theta^2 = -n/\theta^2 < 0$, η λύση αυτή δίνει το μοναδικό μέγιστο του $p_{\theta}(\underline{y})$. Ας σημειωθεί ότι $E_{\theta}\{Y_k\} = 1/\theta$, και άρα είναι λογικό να εκτιμήσουμε το θ ως $1/\bar{y}$. Στην ουσία, ο αδύναμος νόμος των μεγάλων αριθμών συνεπάγεται ότι $\bar{Y} \rightarrow 1/\theta$ στην πιθανότητα ως προς το P_{θ} . Δηλ. η MLE συγκλίνει σε πιθανότητα στην πραγματική τιμή παραμέτρου, μια ιδιότητα γνωστή ως *συνέπεια*. Η ιδιότητα αυτή της MLE δεν είναι συγκεκριμένη στο παράδειγμα αυτό αλλά είναι αληθής σ' ένα γενικό πλαίσιο όπως θα δούμε παρακάτω.

Η πληροφορία του Fisher για την περίπτωση αυτή μπορεί να υπολογιστεί μέσω

$$I_{\theta} = -E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \log p_{\theta}(\underline{Y}) \right] = -E_{\theta} \{-n/\theta^2\} = n/\theta^2,$$

και άρα η CRLB είναι θ^2/n . Αφού το $\partial \log p_{\theta}(\underline{y})/\partial \theta$ δεν είναι της μορφής $k(\theta)[\hat{\theta}_{ML}(\underline{y}) - f(\theta)]$, γνωρίζουμε ότι η ανισότητα πληροφορίας δεν επιτυγχάνεται στο πρόβλημα αυτό. Ωστόσο, μπορούμε να υπολογίσουμε το μέσο και τη διακύμανση του $\hat{\theta}_{ML}$ ευθέως. Συγκεκριμένα, χρησιμοποιώντας χαρακτηριστικές συναρτήσεις αποδεικνύεται ευθέως ότι το μέσο δείγμα \bar{Y} έχει pdf

$$p_{\bar{Y}}(\bar{y}) = \begin{cases} \frac{(n\theta)^n}{n!} \bar{y}^{n-1} e^{-n\theta\bar{y}} & \text{if } \bar{y} \geq 0 \\ 0 & \text{if } \bar{y} < 0, \end{cases}$$

από το οποίο μπορούμε να υπολογίσουμε (για $n > 1$)

$$E_{\theta}\{\hat{\theta}_{ML}(\underline{Y})\} = E_{\theta} \left\{ \frac{1}{\bar{Y}} \right\} = \frac{n\theta}{n-1} \quad (\text{IV.D.8})$$

και (για $n > 2$)

$$\text{Var}_{\theta}[\hat{\theta}_{ML}(\underline{Y})] = \frac{\theta^2 n^2}{(n-1)^2(n-2)}. \quad (\text{IV.D.9})$$

Βλέπουμε από το (IV.D.8) ότι παρόλο που η $\hat{\theta}_{ML}(\underline{Y})$ είναι μη αμερόληπτη, δεν έχει την ιδιότητα ότι $\lim_{n \rightarrow \infty} E_{\theta}\{\hat{\theta}_{ML}(\underline{Y})\} = \theta$. Αυτό σημαίνει ότι είναι *ασυμπτωτικά αμερόληπτη*. Επίσης, παρατηρούμε ότι

$$\text{Var}_{\theta}[\hat{\theta}_{ML}(\underline{Y})]I_{\theta} = \frac{n^3}{(n-1)^2(n-2)} \rightarrow 1$$

καθώς $n \rightarrow \infty$; Και άρα η $\hat{\theta}_{ML}$ έχει διακύμανση ασυμπτωτικά ίση με την CRLB, μια ιδιότητα γνωστή ως *ασυμπτωτική επάρκεια*. Όπως θα δούμε, αυτές οι δύο ιδιότητες της ασυμπτωτικής αμεροληψίας και της επάρκειας είναι χαρακτηριστικές των MLE υπό γενικές συνθήκες για i.i.d. παρατηρήσεις.

Ως τελευταίο σχόλιο στο παράδειγμα αυτό, παρατηρούμε από την Πρόταση IV.C.3 ότι \bar{Y} είναι μια πλήρης επαρκής στατιστική για το θ στο μοντέλο αυτό. Επίσης από το (IV.D.8) βλέπουμε ότι

$$\frac{n-1}{n}\hat{\theta}_{ML}(\underline{y}) \equiv \left(\frac{1}{n-1} \sum_{k=1}^n y_k \right)^{-1}$$

είναι ένας αμερόληπτος εκτιμητής του θ που εξαρτάται από το \bar{Y} . Άρα η

$$\frac{n-1}{n}\hat{\theta}_{ML}(\underline{Y}) \triangleq \hat{\theta}_{MV}(\underline{y})$$

είναι μια MVUE του θ στο πρόβλημα αυτό. Από το (IV.D.9) η διακύμανσή του φαίνεται να δίνεται από

$$\text{Var}_{\theta}[\hat{\theta}_{MV}(\underline{Y})] = \frac{\theta^2}{n-2}, \quad (\text{IV.D.10})$$

ένα μέγεθος που είναι μεγαλύτερο από την CRLB (όπως πρέπει να είναι αφού γνωρίζουμε ότι η CRLB δεν μπορεί εδώ να επιτευχθεί), αλλά που προσεγγίζει την CRLB καθώς το n γίνεται μεγαλύτερο

Η διακύμανση του (IV.D.10) ισούται με την MSE του $\hat{\theta}_{MV}$ αφού είναι αμερόληπτο

Για την MLE, η MSE είναι

$$E_{\theta}\{[\hat{\theta}_{ML}(\underline{Y}) - \theta]^2\} = \text{Var}_{\theta}[\hat{\theta}_{ML}(\underline{Y})] + b^2(\theta), \quad (\text{IV.D.11})$$

όπου η $b(\theta) \triangleq E_{\theta}\{\hat{\theta}_{ML}(\underline{Y})\} - \theta$ είναι η μεροληψία του $\hat{\theta}_{ML}$. Χρησιμοποιώντας τα (IV.D.8) και (IV.D.9), έχουμε

$$E_{\theta}\{[\hat{\theta}_{ML}(\underline{Y}) - \theta]^2\} = \frac{\theta^2(n+2)}{(n-1)(n-2)},$$

ένα μέγεθος που είναι αυστηρά μεγαλύτερο από το $\frac{\theta^2}{(n-2)}$, η MSE του $\hat{\theta}_{MV}$. Άρα στην περίπτωση αυτή, η MVUE είναι προτιμότερη από την MLE, παρόλο που είναι ασυμπτωτικά ισοδύναμες.

Παράδειγμα IV.D.2: Εκτίμηση μέγιστης πιθανοφάνειας εύρους σήματος

Εξετάστε το μοντέλο που αναλύεται στο Παράδειγμα IV.C.3:

$$Y_k = N_k + \mu s_k, \quad k = 1, \dots, n$$

με τα N_1, \dots, N_n i.i.d. $\mathcal{N}(0, \sigma^2)$ και $\underline{s} = (s_1, \dots, s_n)^T$ γνωστά. Η εξίσωση πιθανοφάνειας για την εκτίμηση του μ με το σ^2 γνωστό δίνεται από

$$\begin{aligned} & - \frac{\partial}{\partial \mu} \left(\frac{1}{2} \sum_{k=1}^n \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{k=1}^n (y_k - \mu s_k)^2 \right) \Big|_{\mu = \hat{\mu}_{ML}(\underline{y})} \\ & = \frac{1}{\sigma^2} \sum_{k=1}^n s_k [y_k - \hat{\mu}_{ML}(\underline{y}) s_k] = 0, \end{aligned} \quad (\text{IV.D.12})$$

που συνεπάγεται ότι

$$\hat{\mu}_{ML}(\underline{y}) = \frac{1}{n} \sum_{k=1}^n s_k y_k / \bar{s}^2, \quad (\text{IV.D.13})$$

όπου, όπως πριν, $\bar{s}^2 \triangleq (1/n) \sum_{k=1}^n s_k^2$. Αφού

$$- \frac{\partial^2}{\partial \mu^2} \sum_{k=1}^n \left(\frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} (y_k - \mu s_k)^2 \right) = -n\bar{s}^2/\sigma^2 < 0, \quad (\text{IV.D.14})$$

βλέπουμε ότι το $\log p_{\theta}(\underline{y})$ είναι κοίλο στο μ , άρα η λύση της εξίσωσης πιθανοφάνειας πράγματι δίνει ένα σφαιρικό μέγιστο εδώ.

Ας σημειωθεί ότι το $\hat{\mu}_{ML}$ είναι το ίδιο με την MVUE του μ (βλ. Παράδειγμα IV.C.3) ώστε $E_{\theta}\{\hat{\mu}_{ML}(\underline{Y})\} = \mu$ και $\text{Var}_{\theta}\{\hat{\mu}_{ML}(\underline{Y})\} = \sigma^2/n\bar{s}^2$. Από το (IV.D.14) βλέπουμε ότι $I_{\theta} = n\bar{s}^2/\sigma^2$, άρα

$$\text{CRLB} = \frac{\sigma^2}{n\bar{s}^2} = \text{Var}_{\theta}\{\hat{\mu}_{ML}(\underline{Y})\}. \quad (\text{IV.D.15})$$

Σημειώστε ότι με $\theta = \mu$, μπορούμε να γράψουμε

$$\frac{\partial}{\partial \theta} \log p_{\theta}(\underline{y}) = k(\theta)[\hat{\theta}_{ML}(\underline{y}) - \theta]$$

με $k(\theta) = I_{\theta} = n\overline{s^2}/\sigma^2$, όπως απαιτείται για την επίτευξη του CRLB.

Έστω ότι τώρα το μ είναι γνωστό, αλλά θέλουμε να εκτιμήσουμε το σ^2 . Η εξίσωση πιθανοφάνειας γίνεται

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} \log p_{\theta}(\underline{y}) \Big|_{\sigma^2 = \hat{\sigma}_{ML}^2(\underline{y})} \\ = \frac{1}{2\hat{\sigma}_{ML}^2(\underline{y})} - \frac{1}{2[\hat{\sigma}_{ML}^2(\underline{y})]^2} \sum_{k=1}^n (y_k - \mu s_k)^2 = 0, \end{aligned} \quad (\text{IV.D.16})$$

που έχει τη μοναδική λύση

$$\hat{\sigma}_{ML}^2(\underline{y}) = \frac{1}{n} \sum_{k=1}^n (y_k - \mu s_k)^2. \quad (\text{IV.D.17})$$

Αφού

$$\frac{\partial}{\partial \sigma^2} \log p_{\theta}(\underline{y}) = \frac{n}{2\sigma^4} [\hat{\sigma}_{ML}^2(\underline{y}) - \sigma^2], \quad (\text{IV.D.18})$$

βλέπουμε ότι το $\log p_{\theta}(\underline{y})$ αυξάνεται στο σ^2 για $\sigma^2 < \hat{\sigma}_{ML}^2(\underline{y})$ και μειώνεται στο σ^2 για $\sigma^2 > \hat{\sigma}_{ML}^2(\underline{y})$. Άρα το $\log p_{\theta}(\underline{y})$ επιτυγχάνει το απόλυτο μέγιστό του στο $\hat{\sigma}_{ML}^2(\underline{y})$. Επίσης βλέπουμε από το (IV.D.18) ότι με $\theta = \sigma^2$,

$$\frac{\partial}{\partial \theta} \log p_{\theta}(\underline{y}) = \frac{n}{2\theta^2} [\hat{\theta}_{ML}(\underline{y}) - \theta], \quad (\text{IV.D.19})$$

το οποίο από το Παράδειγμα IV.C.4 συνεπάγεται ότι το $\hat{\sigma}_{ML}^2(\underline{y})$ είναι αμερόληπτο και επιτυγχάνει την CRLB και άρα ότι $\hat{\sigma}_{ML}^2$ είναι μια MVUE του σ^2 . Αναθεωρώντας το (IV.D.19) έχουμε $I_{\theta} = n/2\theta^2 \equiv n/2\sigma^4$, άρα

$$\text{CRLB} = \frac{2\sigma^4}{n} = \text{Var}_{\theta}[\hat{\sigma}_{ML}^2(\underline{Y})]. \quad (\text{IV.D.20})$$

Τώρα ας υποθέσουμε ότι και το μ και το σ^2 είναι άγνωστα. Θέτοντας $\theta = (\mu, \sigma^2)$, η MLE του θ βρίσκεται μεγιστοποιώντας το $p_{\theta}(\underline{y})$ στο μ και το σ^2 . Καθώς το μέγιστο $\hat{\mu}_{ML}(\underline{y})$ από το (IV.D.13) δεν εξαρτάται από το σ^2 , έχουμε ότι

$$\begin{aligned} \max_{(\mu, \sigma^2)} \log p_{\theta}(\underline{y}) &= \max_{\sigma^2} \{ \max_{\mu} \log p_{\theta}(\underline{y}) \} \\ &= \max_{\sigma^2} \left\{ -\frac{1}{2} \sum_{k=1}^n \log(2\pi\sigma^2) \right. \\ &\quad \left. - \frac{1}{2\sigma^2} \sum_{k=1}^n [y_k - \hat{\mu}_{ML}(\underline{y})s_k]^2 \right\}. \end{aligned}$$

Αλλά το δεξί μέλος της ισότητας αυτής είναι το ίδιο πρόβλημα μεγιστοποίησης όπως αυτό για την εκτίμηση του σ^2 με γνωστό μ , με το μ σύνολο ίσο με το $\hat{\mu}_{ML}(\underline{y})$. Άρα το μέγιστο επιτυγχάνεται από το (IV.D.17) με το $\hat{\mu}_{ML}$ που αντικαθίσταται για το μ και το MLE για το $\theta = (\mu, \sigma^2)$ να είναι $\hat{\theta}_{ML}(\underline{y}) = [\hat{\mu}_{ML}(\underline{y}), \hat{\sigma}_{ML}^2(\underline{y})]$, όπου

$$\hat{\mu}_{ML}(\underline{y}) = \frac{1}{n} \sum_{k=1}^n s_k y_k / s^2 \quad (\text{IV.D.21a})$$

και

$$\hat{\sigma}_{ML}^2(\underline{y}) = \frac{1}{n} \sum_{k=1}^n [y_k - \hat{\mu}_{ML}(\underline{y})s_k]^2. \quad (\text{IV.D.21b})$$

Η εκτίμηση $\hat{\mu}_{ML}(\underline{y})$ παραμένει μια MVUE του μ στην περίπτωση αυτή. Ωστόσο, από το (IV.C.19) βλέπουμε ότι $\hat{\sigma}_{ML}^2(\underline{y})$ is $[(n-1)/n]\hat{\sigma}_{MV}^2(\underline{y})$. Άρα,

$$E_{\theta}\{\hat{\sigma}_{ML}^2(\underline{Y})\} = \frac{n-1}{n}\sigma^2,$$

και η MLE του σ^2 είναι μεροληπτική εδώ (παρόλο που είναι ασυμπτωτικά αμερόληπτη).

Ας σημειωθεί ότι

$$\text{Var}_{\theta}[\hat{\sigma}_{ML}^2(\underline{Y})] = [(n-1)^2/n^2]\text{Var}_{\theta}[\hat{\sigma}_{MV}^2(\underline{Y})],$$

οπότε η $\hat{\sigma}_{ML}^2(\underline{y})$ έχει χαμηλότερη διακύμανση από την MVUE. Μπορεί να αποδειχθεί ότι (βλέπε Άσκηση 14)⁶.

$$\text{Var}_{\theta}(\hat{\sigma}_{MV}^2(\underline{Y})) = \frac{2\sigma^4}{n-1}. \quad (\text{IV.D.22})$$

Άρα για την MVUE του σ^2 , η MSE, $E_{\theta}\{[\hat{\sigma}_{MV}^2(\underline{Y}) - \sigma^2]^2\}$, είναι $2\sigma^4/(n-1)$.

Εναλλακτικά, για την MLE του σ^2 , η MSE δίνεται από

$$\begin{aligned}
E_{\theta}\{[\sigma_{ML}^2(\underline{Y}) - \sigma^2]^2\} &= \text{Var}_{\theta}[\sigma_{ML}^2(\underline{Y})] + [E_{\theta}\{\hat{\sigma}_{ML}^2(\underline{Y})\} - \sigma^2]^2 \\
&= \frac{(n-1)^2}{n^2} \frac{2\sigma^4}{n-1} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\
&= \sigma^4 \frac{(2n-1)}{n^2}. \quad (\text{IV.D.23})
\end{aligned}$$

Ο λόγος αυτών των δύο μεγεθών είναι

$$\frac{E_{\theta}\{[\hat{\sigma}_{MV}^2(\underline{Y}) - \sigma^2]^2\}}{E_{\theta}\{[\hat{\sigma}_{ML}^2(\underline{Y}) - \sigma^2]^2\}} = \left(\frac{n}{n-1}\right) \left(\frac{2n}{2n-1}\right) > 1. \quad (\text{IV.D.24})$$

Από το (IV.D.24) βλέπουμε ότι η MLE στην περίπτωση αυτή έχει μια ομοιόμορφα χαμηλότερη MSE από την MVUE. Αυτό συμβαίνει γιατί η αύξηση στην MSE λόγω της μεροληψίας της MLE είναι περισσότερο αντισταθμισμένη, με την αύξηση στη διακύμανση της MVUE που απαιτείται για την επίτευξη της αμεροληψίας. Άρα, η επίτευξη του στόχου ελάχιστης διακύμανσης αμερόληπτης εκτίμησης δεν οδηγεί πάντα σε μια βέλτιστη εκτίμηση σε ό,τι αφορά το μέσο τετραγωνικό σφάλμα.

Ένα από τα βασικά κίνητρα για τη χρήση της εκτίμησης μέγιστης πιθανοφάνειας απεικονίζεται από τα δύο παραπάνω παραδείγματα. Κυρίως, εκτιμήσεις βασισμένες σε ανεξάρτητα δείγματα έχουν καλές ασυμπτωτικές ιδιότητες, καθώς ο αριθμός των δειγμάτων αυξάνεται χωρίς φράγμα. Ο λόγος για αυτή την ασυμπτωτική συμπεριφορά είναι προφανής από τα ορίσματα στις ακόλουθες παραγράφους.

Έστω ότι έχουμε μια ακολουθία των i.i.d. παρατηρήσεων Y_1, Y_2, \dots, Y_n , κάθε μία με οριακή πυκνότητα f_{θ} που να προέρχεται από την οικογένεια $\{f_{\theta}; \theta \in \Lambda\}$. Έστω ότι το $\hat{\theta}_n$ συμβολίζει μια λύση της εξίσωσης πιθανοφάνειας για δείγμα μεγέθους n , δηλ.,

$$\frac{\partial}{\partial \theta} \log p_{\theta}(\underline{y}) \Big|_{\theta=\hat{\theta}_n(\underline{y})} = \sum_{k=1}^n \psi[y_k; \hat{\theta}_n(\underline{y})] = 0,$$

όπου $\psi(y_k; \theta) \triangleq \partial \log f_{\theta}(y_k) / \partial \theta$. Παρομοίως, μπορούμε να γράψουμε

$$\frac{1}{n} \sum_{k=1}^n \psi[y_k; \hat{\theta}_n(\underline{y})] = 0. \quad (\text{IV.D.25})$$

Αξίζει να σημειώσουμε ότι η MVUE του σ^2 με γνωστό μ έχει διακύμανση $2\sigma^4/n$ (από το (IV.D.20)) και η MVUE του σ^2 με άγνωστο μ έχει διακύμανση $2\sigma^4/(n-1)$ (από το (IV.D.22)). Άρα για αμερόληπτη εκτίμηση του σ^2 , υπάρχει μια «ποινή» μιας παρατήρησης όταν το μ είναι άγνωστο.

Για μια πάγια τιμή παραμέτρου $\theta' \in \Lambda$ εξετάστε το μέγεθος $\sum_{k=1}^n \psi(Y_k; \theta') / n$.

Υποθέτοντας ότι η θ είναι η πραγματική τιμή παραμέτρου (δηλ. $Y_k \sim f_\theta$), ο αδύναμος νόμος μεγάλων αριθμών συνεπάγεται ότι

$$\frac{1}{n} \sum_{k=1}^n \psi(Y_k; \theta') \xrightarrow{i.p.} E_\theta \{\psi(Y_1; \theta')\}.$$

Έχουμε

$$\begin{aligned} E_\theta \{\psi(Y_1; \theta')\} &= \int_{\mathbf{R}} \frac{\partial}{\partial \theta} \log f_\theta(y_1) |_{\theta=\theta'} f_\theta(y_1) \mu(dy_1) \\ &\triangleq J(\theta; \theta'). \end{aligned} \quad (\text{IV.D.26})$$

Υποθέτοντας ότι η σειρά ενσωμάτωσης και διαφορισμού μπορεί να εναλλαχθεί στο (IV.D.26), το $J(\theta; \theta)$ μπορεί να γραφεί ως

$$\begin{aligned} J(\theta; \theta) &= \int \left[\frac{\partial}{\partial \theta} \log f_\theta(y_1) \right] f_\theta(y_1) \mu(dy_1) \\ &= \int \frac{\partial}{\partial \theta} f_\theta(y_1) \mu(dy_1) \\ &= \frac{\partial}{\partial \theta} \int f_\theta(y_1) \mu(dy_1) = \frac{\partial}{\partial \theta} (1) = 0. \end{aligned}$$

Άρα η ισότητα $J(\theta; \theta') = 0$ έχει μια λύση $\theta' = \theta$. Υποθέστε ότι αυτή είναι η μοναδική ρίζα του $J(\theta; \theta')$ και υποθέστε ότι οι $J(\theta; \theta')$ και $\sum_{k=1}^n \psi(Y_k; \theta') / n$ είναι και οι δύο λείες συναρτήσεις του θ' . Τότε, αφού η $\sum_{k=1}^n \psi(Y_k; \theta') / n$ είναι κοντά στο $J(\theta; \theta')$ για μεγάλο n , θα περιμέναμε οι ρίζες αυτών των δύο συναρτήσεων να είναι κοντά όταν το n είναι μεγάλο.

Δηλαδή, το $\hat{\theta}_n(\underline{Y})$ θα πρέπει να είναι κοντά στην πραγματική τιμή παραμέτρου θ όταν το n είναι μεγάλο. Και καθώς $n \rightarrow \infty$, θα περιμέναμε ότι $\hat{\theta}_n(\underline{Y}) \rightarrow \theta$ υπό κάποια στατιστική έννοια. Ουσιαστικά, υπό τις κατάλληλες συνθήκες λειότητας και μοναδικότητας, οι λύσεις στην εξίσωση πιθανοφάνειας είναι *συνεπείς*. Δηλ. συγκλίνουν σε πιθανότητα στην πραγματική τιμή παραμέτρου:

$$\lim_{n \rightarrow \infty} P_\theta(|\hat{\theta}_n(\underline{Y}) - \theta| > \epsilon) = 0 \text{ for all } \epsilon > 0.$$

Ένα σύνολο συνθηκών κάτω από τις οποίες οι λύσεις της εξίσωσης πιθανοφάνειας είναι συνεπείς συνοψίζεται παρακάτω.

Πρόταση IV.D.1: Συνέπεια των MLE.

Έστω ότι η $\{Y_k\}_{k=1}^{\infty}$ είναι μια i.i.d. ακολουθία τυχαίων μεταβλητών, η κάθε μία με πυκνότητα f_{θ} , και έστω ότι τα J και ψ ορίζονται ορθά όπως παραπάνω. Έστω επίσης ότι ισχύουν οι ακόλουθες συνθήκες.

- (1) Η $J(\theta; \theta')$ είναι μια συνεχής συνάρτηση του θ' και έχει μια μοναδική ρίζα στο $\theta' = \theta$, σημείο στο οποίο αλλάζει πρόσημο.
- (2) Η $\psi(Y_k; \theta')$ είναι μια συνεχής συνάρτηση του θ' (με πιθανότητα 1).
- (3) Για κάθε n , η $\sum_{k=1}^n \psi(Y_k; \theta')/n$ έχει μια μοναδική ρίζα $\hat{\theta}_r$ (με πιθανότητα 1).

Τότε $\hat{\theta}_n \rightarrow \theta$ (i.p.).

Απόδειξη: Επιλέξτε $\epsilon > 0$. Με τη συνθήκη (1), οι $J(\theta; \theta + \epsilon)$ και $J(\theta; \theta - \epsilon)$ πρέπει να έχουν αντίθετα πρόσημα. Ορίστε $\delta = \min\{|J(\theta; \theta + \epsilon)|, |J(\theta; \theta - \epsilon)|\}$ και για κάθε n , ορίστε τις περιπτώσεις

$$A_n^- = \{ |J(\theta; \theta - \epsilon) - \frac{1}{n} \sum_{k=1}^n \psi(Y_k; \theta - \epsilon)| \leq \delta \}, \quad (\text{IV.D.27})$$

και $A_n = A_n^+ \cap A_n^-$.

Τώρα στο A_n^+ , το $\sum_{k=1}^n \psi(Y_k; \theta + \epsilon)/n$ πρέπει να έχει το ίδιο πρόσημο με το $J(\theta; \theta + \epsilon)$ και στο A_n^- , το $\sum_{k=1}^n \psi(Y_k; \theta - \epsilon)/n$ πρέπει να έχει το ίδιο πρόσημο με το $J(\theta; \theta - \epsilon)$. Άρα στο A_n , τα $\sum_{k=1}^n \psi(Y_k; \theta + \epsilon)/n$ και $\sum_{k=1}^n \psi(Y_k; \theta - \epsilon)/n$ έχουν αντίθετα πρόσημα. Με την παραδοχή της συνέχειας (2), το $\sum_{k=1}^n \psi(Y_k; \theta')/n$ μπορεί να αλλάξει πρόσημο μόνο περνώντας από το μηδέν. Άρα στο A_n , η ρίζα $\hat{\theta}_n$ βρίσκεται ανάμεσα στο $\theta - \epsilon$ και $\theta + \epsilon$. Αυτό συνεπάγεται ότι το A_n είναι ένα υποσύνολο του $\{|\hat{\theta}_n - \theta| \leq \epsilon\}$, έτσι ώστε $P(|\hat{\theta}_n - \theta| \leq \epsilon) \geq P(A_n)$.

Με τον αδύναμο νόμο των μεγάλων αριθμών

$$\frac{1}{n} \sum_{k=1}^n \psi(Y_k; \theta + \epsilon) \rightarrow J(\theta; \theta + \epsilon) \text{ (i.p.)}$$

και

$$\frac{1}{n} \sum_{k=1}^n \psi(Y_k; \theta - \epsilon) \rightarrow J(\theta; \theta - \epsilon) \text{ (i.p.)}, \quad (\text{IV.D.28})$$

Άρα $P(A_n^+) \rightarrow 1$ και $P(A_n^-) \rightarrow 1$ καθώς $n \rightarrow \infty$. Έχουμε

$$\begin{aligned} 1 &\geq P(|\hat{\theta}_n - \theta| \leq \epsilon) \\ &\geq P(A_n) = P(A_n^+) + P(A_n^-) - P(A_n^+ \cup A_n^-) \\ &\geq P(A_n^+) + P(A_n^-) - 1 \rightarrow 1. \end{aligned} \quad (\text{IV.D.29})$$

Άρα $P(|\hat{\theta}_n - \theta| \leq \epsilon) \rightarrow 1$, και αφού το ϵ επιλέχθηκε αυθαίρετα έχουμε το επιθυμητό αποτέλεσμα

Παρατηρήσεις: Οι συνθήκες στην πρόταση αυτή μπορούν να χαλαρώσουν με διάφορους τρόπους. Πρώτον, η συνέχεια των συναρτήσεων $J(\theta; \theta')$ και $\psi(Y_k; \theta')$ μπορεί να χαλαρώσουν σε μια γειτονιά του $\theta' = \theta$. Επίσης, δεν είναι απαραίτητο να υποθέσουμε την ύπαρξη των ριζών $\hat{\theta}_n$, αφού η παραπάνω ανάπτυξη δείχνει ότι θα πρέπει να υπάρχει μια ρίζα στην εξίσωση πιθανοφάνειας στο A_n , που έχει πιθανότητα που τείνει το 1. Στην ουσία, με μόνο την παραδοχή της τοπικής συνέχειας, η παραπάνω απόδειξη μπορεί να χρησιμοποιηθεί για να αποδειχθεί ότι με πιθανότητα που τείνει στο 1, υπάρχει μια ακολουθία ριζών στην εξίσωση πιθανοφάνειας που κλίνει σε κάθε απομονωμένη ρίζα του $J(\theta; \theta')$. Άρα αν το $J(\theta; \theta')$ έχει πολλαπλές ρίζες, μπορούν να προκύψουν ασυνεπείς ακολουθίες λύνοντας την εξίσωση πιθανοφάνειας.

Πέρα από τη συνέπεια, βλέπουμε στα παραπάνω παραδείγματα ότι οι λύσεις της εξίσωσης πιθανοφάνειας μπορεί επίσης να είναι ασυμπτωτικά αμερόληπτες και επαρκείς. Ξέρουμε ότι υπό τις προϋποθέσεις της Πρότασης IV.D.1, το $\hat{\theta}_n$ συγκλίνει στο θ στην πιθανότητα. Άρα, αν γράφαμε

$$\lim_{n \rightarrow \infty} E_{\theta}\{\hat{\theta}_n\} = E_{\theta}\{\lim_{n \rightarrow \infty} \hat{\theta}_n\} \quad (\text{IV.D.30})$$

γι' αυτόν τον τύπο σύγκλισης, τότε η ασυμπτωτική αμεροληψία θα έποταν. Η εναλλαγή ορίων και προσδοκιών στο (IV.D.30) δεν ισχύει πάντα για τη σύγκλιση στην πιθανότητα. Ωστόσο, υπό διάφορες συνθήκες στο ψ , η εναλλαγή αυτή μπορεί να αποδειχθεί έγκυρη. (Μια επαρκής συνθήκη για την εγκυρότητα αυτής της εναλλαγής είναι η ύπαρξη μιας τυχαίας μεταβλητής X τέτοια ώστε $|\hat{\theta}_n| \leq X$ για κάθε n και $E_{\theta}\{X\} < \infty$. Αυτό είναι γνωστό ως *θεώρημα κυριαρχημένης σύγκλισης*). Άρα δεν είναι παράλογο να αναμένεται ως ιδιότητα η ασυμπτωτική αμεροληψία σε ό,τι αφορά τη συνέπεια του $\hat{\theta}_n$.

Είναι λιγότερο σαφές γιατί οι MLE μπορεί να είναι ασυμπτωτικά επαρκείς. Για να δούμε γιατί μπορεί να συμβαίνει αυτό, εξετάζουμε το σχετικό ζήτημα της εύρεσης της ασυμπτωτικής κατανομής σφάλματος, $\hat{\theta}_n - \theta$. Συγκεκριμένα, αποδεικνύουμε την ακόλουθη πρόταση.

Πρόταση IV.D.2: Ασυμπτωτική Κανονικότητα των MLE.

Έστω ότι η $\{Y_k\}_{k=1}^{\infty}$ είναι μια ακολουθία των i.i.d. τυχαίων μεταβλητών, κάθε μία από τις οποίες έχει πυκνότητα f_{θ} και ότι η $\{\hat{\theta}_n\}_{n=1}^{\infty}$ είναι μια συνεπής ακολουθία ριζών της εξίσωσης πιθανοφάνειας. Υποθέστε επίσης ότι το ψ ικανοποιεί τις παρακάτω συνθήκες κανονικότητας.

- (1) $0 < i_{\theta} \triangleq E_{\theta}\{[\psi(Y_1; \theta)]^2\} < \infty$.
- (2) Οι παράγωγοι $\psi'(Y_1; \theta') \triangleq \partial\psi(Y_1; \theta')/\partial\theta'$ και $\psi''(Y_k; \theta') \triangleq \partial^2\psi(Y_k; \theta')/(\partial\theta')^2$ υπάρχουν (με πιθανότητα 1).
- (3) Υπάρχει μια συνάρτηση $M(Y_1)$ τέτοια που $|\psi''(Y_1; \theta')| \leq M(Y_1)$ για όλα τα $\theta' \in \Lambda$ και $E_{\theta}\{M(Y_1)\} < \infty$.
- (4) $J(\theta; \theta) = 0$, όπου η $J(\theta; \theta')$ ορίζεται όπως στο (IV.D.26)
- (5) Ισχύει η συνθήκη (5) της Πρότασης IV.C.24

Τότε

$$P_{\theta}(\sqrt{ni_{\theta}}(\hat{\theta}_n - \theta) \leq x) \rightarrow \Phi(x) \text{ for all } x \in \mathbb{R},$$

όπου Φ είναι η πάγια κατά Gauss συνάρτηση κατανομή. Δηλαδή, το $\sqrt{n}(\hat{\theta}_n - \theta)$ συγκλίνει σε κατανομή σε μια $\mathcal{N}(0, 1/i_{\theta})$ τυχαία μεταβλητή.

Απόδειξη: Χρησιμοποιώντας το θεώρημα του μπορούμε να επεκτείνουμε το δεξιό μέλος της εξίσωσης πιθανοφάνειας, $(1/n) \sum_{k=1}^n \psi(Y_k; \hat{\theta}_n) = 0$, προς το θ για να λάβουμε

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n \psi(Y_k; \theta) + (\hat{\theta}_n - \theta) \frac{1}{n} \sum_{k=1}^n \psi'(Y_k; \theta) \\ + \frac{1}{2} (\hat{\theta}_n - \theta)^2 \frac{1}{n} \sum_{k=1}^n \psi''(Y_k; \bar{\theta}_n) = 0, \quad (\text{IV.D.31}) \end{aligned}$$

όπου το $\bar{\theta}_n$ βρίσκεται μεταξύ του θ και του $\hat{\theta}_n$. Αναδιατάσσοντας το (IV.D.31) έχουμε μια παράσταση για το μέγεθος $\sqrt{n}(\hat{\theta}_n - \theta)$:

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{-\frac{1}{\sqrt{n}} \sum_{k=1}^n \psi(Y_k; \theta)}{\frac{1}{n} \sum_{k=1}^n \psi'(Y_k; \theta) + (\hat{\theta}_n - \theta) \frac{1}{2n} \sum_{k=1}^n \psi''(Y_k; \bar{\theta}_n)}. \quad (\text{IV.D.32})$$

Εξετάστε τον παρονομαστή στο δεξί μέλος του (IV.D.32). Με τον αδύναμο νόμο των μεγάλων αριθμών, ο πρώτος όρος, $\sum_{k=1}^n \psi'(Y_k; \theta)/n$, συγκλίνει στο $E_{\theta}\{\psi'(Y_1; \theta)\}$ σε πιθανότητα. Με τη συνθήκη (3), ο δεύτερος όρος $(\hat{\theta}_n - \theta) \sum_{k=1}^n \psi''(Y_k; \bar{\theta}_n)/2n$ φράσσεται ως

$$\begin{aligned} & \left| \frac{1}{2}(\hat{\theta}_n - \theta) \frac{1}{n} \sum_{k=1}^n \psi''(Y_k; \hat{\theta}_n) \right| \\ & \leq \frac{1}{2} |\hat{\theta}_n - \theta| \frac{1}{n} \sum_{k=1}^n M(Y_k). \end{aligned} \quad (\text{IV.D.33})$$

Τώρα, $|\hat{\theta}_n - \theta| \rightarrow 0$ (i.p.) και ο αδύναμος νόμος των μεγάλων αριθμών συνεπάγεται ότι $(1/n) \sum_{k=1}^n M(Y_k) \rightarrow E_{\theta}\{M(Y_1)\} < \infty$. Άρα, ο δεύτερος όρος συγκλίνει σε πιθανότητα στο μηδέν και ο παρονομαστής τότε συγκλίνει σε πιθανότητα στο $E_{\theta}\{\psi'(Y_1; \theta)\}$.

Το άθροισμα αριθμητή $\sum_{k=1}^n \psi(Y_k; \theta)$ στο (IV.D.32) είναι το σύνολο των n i.i.d. τυχαίων μεταβλητών, η κάθε μία με μέσο $E_{\theta}\{\psi(Y_1; \theta)\} = J(\theta; \theta) = 0$ και διακύμανση $E_{\theta}\{\psi^2(Y_1; \theta)\} = i_{\theta} < \infty$. Άρα με το κεντρικό οριακό θεώρημα,

$-(1/\sqrt{n}) \sum_{k=1}^n \psi(Y_k; \theta)$ συγκλίνει σε κατανομή σε μια $\mathcal{N}(0, i_{\theta})$ τυχαία μεταβλητή.

Τα δύο παραπάνω αποτελέσματα συνεπάγονται ότι το $\sqrt{n}(\hat{\theta}_n - \theta)$ συγκλίνει σε κατανομή σε μια $\mathcal{N}(0, v^2)$ τυχαία μεταβλητή με

$$v^2 = i_{\theta} / E_{\theta}^2\{\psi'(Y_1; \theta)\}. \quad (\text{IV.D.34})$$

Αλλά χρησιμοποιώντας το όρισμα που χρησιμοποιήθηκε για την εξαγωγή του (IV.C.22), $E_{\theta}\{\psi'(Y_k; \theta)\} = -E_{\theta}\{\psi^2(Y_1; \theta)\} = -i_{\theta}$, άρα $v^2 = 1/i_{\theta}$. Αυτό ολοκληρώνει την απόδειξη.

Παρατηρήσεις: Είναι εύκολα κατανοητό ότι η πληροφορία Fisher δίνεται από το

$I_{\theta} = ni_{\theta}$ γι' αυτή την i.i.d. περίπτωση. Ευρετικά, μπορούμε να θεωρήσουμε το

συμπέρασμα αυτής της πρότασης ως τη συνθήκη ότι το $\hat{\theta}_n$ είναι ασυμπτωτικά $\mathcal{N}(\theta, 1/ni_{\theta})$.

Δηλαδή το $\hat{\theta}_n$ ασυμπτωτικά έχει μέσο θ και διακύμανση ίση με $1/ni_{\theta}$, το κατώτερο φράγμα Cramer – Rao. Στην ουσία, αυτό που αποδείξαμε είναι ότι η ασυμπτωτική κατανομή του

$\sqrt{n}(\hat{\theta}_n - \theta)$ έχει μηδενικό μέσο και διακύμανση $1/i_{\theta}$ που δεν είναι ίδιο με

$E_{\theta}\{\sqrt{n}(\hat{\theta}_n - \theta)\} \rightarrow 0$ και $\text{Var}_{\theta}[\sqrt{n}(\hat{\theta}_n - \theta)] \rightarrow 1/i_{\theta}$. Οι δύο αυτές τελευταίες συνθήκες (η δεύτερη των οποίων είναι η ασυμπτωτική επάρκεια) μπορεί να ισχύει. Ωστόσο, απαιτούνται πρόσθετες συνθήκες για να υποτεθεί αυτό. [Οι ιδιότητες αυτές μπορούν να εξεταστούν μέσω του (IV.D.32). Παρόλ' αυτά, το συμπέρασμα της IV.D.2 είναι επαρκώς πρακτική αιτιολογία για να εξετάσουμε την MLE ως έναν ασυμπτωτικά βέλτιστο (MVUE) εκτιμητή. Και, στην ουσία, η ασυμπτωτική αμεροληψία και επάρκεια ορίζονται ενίοτε εναλλακτικά σε σχέση με το μέσο και τη διακύμανση της ασυμπτωτικής κατανομής σφάλματος.

IV. Περαιτέρω πλευρές και επεκτάσεις της Εκτίμησης μέγιστης πιθανοφάνειας

IV.E.1 ΕΚΤΙΜΗΣΗ ΤΩΝ ΔΙΑΝΥΣΜΑΤΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ

Θα πρέπει να σημειωθεί ότι όλη η ανάλυση της προηγούμενης ενότητας μπορεί να γενικευθεί για την περίπτωση στην οποία η παράμετρος είναι διάνυσμα, ας πούμε της διάστασης m . Στην περίπτωση αυτή η ισότητα είναι μια διανυσματική ισότητα

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \log p_{\underline{\theta}}(y) \Big|_{\underline{\theta}=\hat{\underline{\theta}}} &= 0 \\ &\vdots \\ &\vdots \\ \frac{\partial}{\partial \theta_m} \log p_{\underline{\theta}}(y) \Big|_{\underline{\theta}=\hat{\underline{\theta}}} &= 0, \end{aligned} \quad (\text{IV.E.1})$$

το οποίο για i.i.d. μοντέλα γίνεται

$$\begin{aligned} \sum_{k=1}^n \psi_1(y_k; \hat{\underline{\theta}}_n) &= 0 \\ &\vdots \\ &\vdots \\ \sum_{k=1}^n \psi_m(y_k; \hat{\underline{\theta}}_n) &= 0, \end{aligned} \quad (\text{IV.E.2})$$

όπου $\psi_j(y_k; \underline{\theta}) = \partial \log f_{\underline{\theta}}(y_k) / \partial \theta_j$ και όπου η $f_{\underline{\theta}}$ είναι η οριακή πυκνότητα του Y_k .

Η ανισότητα της πληροφορίας (Πρόταση IV.C.4) μπορεί, στο πλαίσιο της κανονικότητας, να επεκταθεί στη διανυσματική περίπτωση. Για παράδειγμα, το κατώτερο φράγμα Cramer – Rao στη διακύμανση των αμερόληπτων εκτιμήσεων γίνεται

$$\text{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq \mathbf{I}_{\underline{\theta}}^{-1}, \quad (\text{IV.E.3})$$

όπου $\text{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \triangleq E_{\underline{\theta}}\{(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})^T\}$, και $\mathbf{I}_{\underline{\theta}}$ είναι ο $m \times m$ πίνακας πληροφορίας Fisher με το j -ιστό στοιχείο

$$(\mathbf{I}_{\underline{\theta}})_{j,l} = E_{\underline{\theta}} \left(\left[\frac{\partial}{\partial \theta_j} \log p_{\underline{\theta}}(Y) \right] \left[\frac{\partial}{\partial \theta_l} \log p_{\underline{\theta}}(Y) \right] \right). \quad (\text{IV.E.4})$$

Σημειώστε ότι $\mathbf{I}_{\underline{\theta}}$ είναι ο πίνακας συνδιακύμανσης του μηδενικού μέσου διανύσματος

$$\left(\frac{\partial}{\partial \theta_1} \log p_{\underline{\theta}}(Y), \frac{\partial}{\partial \theta_2} \log p_{\underline{\theta}}(Y), \dots, \frac{\partial}{\partial \theta_m} \log p_{\underline{\theta}}(Y) \right)^T,$$

και άρα είναι τουλάχιστον μη αρνητικά ορισμένο. Η ισότητα του (IV.E.3) υποθέτει ότι είναι θετικά ορισμένο. Η ανισότητα $\mathbf{A} \geq \mathbf{B}$ για πίνακες σημαίνει ότι $(\mathbf{A} - \mathbf{B})$ είναι μια αρνητικά ορισμένο. Για την i.i.d. περίπτωση, το (IV.E.3) γίνεται

$$\text{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq \frac{1}{n} \mathbf{i}_{\underline{\theta}}^{-1} \quad (\text{IV.E.5})$$

όπου

$$(\mathbf{i}_{\underline{\theta}})_{j,l} = E_{\underline{\theta}}\{\psi_j(Y_1; \underline{\theta})\psi_l(Y_1; \underline{\theta})\}. \quad (\text{IV.E.6})$$

και σε συνθήκες παρόμοιες με αυτές της Πρότασης IV.D.2

$$\sqrt{n}(\hat{\underline{\theta}}_n - \underline{\theta}) \rightarrow \mathcal{N}(\underline{0}, \mathbf{i}_{\underline{\theta}}^{-1}) \quad (\text{IV.E.8})$$

σε κατανομή. Άρα, η περίπτωση διανυσματικής παραμέτρου είναι παρόμοια με την βαθμωτή.

Λεπτομέρειες γι' αυτήν και άλλες πλευρές της συμπεριφοράς των MLE για i.i.d. μοντέλα, απαντούν στο βιβλίο του Lehmann (1983).

IV.E.2 ΕΚΤΙΜΗΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΣΗΜΑΤΟΣ

Οι ασυμπτωτικές ιδιότητες των MLE μπορούν επίσης να επεκταθούν σε κάποια προβλήματα μεταβαλλόμενου χρόνου. Ιδιαίτερο ενδιαφέρον έχει η κατάσταση στην οποία έχουμε πραγματικής τιμής παρατηρήσεις της μορφής

$$Y_k = s_k(\theta) + N_k, \quad k = 1, \dots, n, \quad (\text{IV.E.9})$$

όπου η $\{s_k(\theta)\}_{k=1}^n$ είναι μια ακολουθία σήματος, η οποία είναι μια γνωστή συνάρτηση της άγνωστης παραμέτρου θ , και όπου η $\{N_k\}_{k=1}^n$ είναι μια i.i.d. ακολουθία θορύβου με οριακή πυκνότητα πιθανότητας f . Υποθέτουμε, για λόγους απλότητας, ότι το θ είναι μια βαθμωτή παράμετρος που βρίσκεται σ' ένα διάστημα Λ .

Η εκτίμηση μέγιστης πιθανοφάνειας του θ στο (IV.E.9) λύνει την εξίσωση

$$\hat{\theta}_n = \arg \max_{\theta \in \Lambda} \left[\sum_{k=1}^n \log f[Y_k - s_k(\theta)] \right],$$

ή ισοδύναμα

$$\hat{\theta}_n = \arg \min_{\theta \in \Lambda} \left[- \sum_{k=1}^n \log f[Y_k - s_k(\theta)] \right], \quad (\text{IV.E.10})$$

Και άρα η εξίσωση πιθανοφάνειας είναι

$$\sum_{k=1}^n s'_k(\hat{\theta}_n) \psi[Y_k - s_k(\hat{\theta}_n)] = 0, \quad (\text{IV.E.11})$$

όπου $\psi \triangleq -f'/f$, $f'(x) \triangleq df(x)/dx$, και $s'_k(\theta) \triangleq \partial s_k(\theta)/\partial \theta$. Για παράδειγμα, όταν f είναι μια $\mathcal{N}(0, \sigma^2)$ πυκνότητα, τα (IV.E.10) και (IV.E.11) είναι ισοδύναμα με

$$\hat{\theta}_n = \arg \left[\min_{\theta \in \Lambda} \sum_{k=1}^n [Y_k - s_k(\theta)]^2 \right] \quad (\text{IV.E.12})$$

Και

$$\sum_{k=1}^n s'_k(\hat{\theta}_n) [Y_k - s_k(\hat{\theta}_n)] = 0, \quad (\text{IV.E.13})$$

αντίστοιχα. Ο ειδικός εκτιμητής (IV.E.12) είναι ενίοτε γνωστός ως *εκτίμηση ελάχιστων τετραγώνων του θ* , καθώς επιλέγει αυτή την τιμή του θ για την οποία $\{s_k(\theta)\}_{k=1}^n$ είναι τα ελάχιστα τετράγωνα, κατάλληλα για τα δεδομένα. Δηλαδή, επιλέγει το θ ώστε να ελαχιστοποιήσει το σύνολο των τετραγωνικών σφαλμάτων μεταξύ των δεδομένων και του σήματος που προκύπτουν από αυτή την επιλογή του θ . Τα ελάχιστα τετράγωνα είναι μια κλασική τεχνική εκτίμησης και χρησιμοποιείται συχνά σε μοντέλα, όπως στο (IV.E.9) ακόμη κι όταν τα σφάλματα δεν μπορούν να εικασθούν ότι είναι κατά Gauss.

Οι λύσεις της εξίσωσης πιθανοφάνειας του (IV.E.11) μπορεί να έχουν ασυμπτωτικές ιδιότητες παρόμοιες με αυτές για τις MLE στα i.i.d. μοντέλα. Ωστόσο, η μεταβολή χρόνου του σήματος προσθέτει διαφορετικές θεωρήσεις στην ασυμπτωτική ανάλυση. Για παράδειγμα, αν το σήμα γίνει πανομοιότυπα μηδέν (ή ειδιάλλως ανεξάρτητο του θ) μετά από κάποιο πεπερασμένο αριθμό δειγμάτων, θα ήταν μη ρεαλιστικό να περιμένουμε συνέπεια στο μοντέλο αυτό. Για να περιγράψουμε τους τύπους συνθηκών που απαιτούνται από το σήμα ώστε οι λύσεις της εξίσωσης πιθανοφάνειας να απολαμβάνουν τις ιδιότητες των i.i.d. αντίστοιχών τους, θα αναλύσουμε την ειδική περίπτωση της εκτίμησης ελάχιστων τετραγώνων του (IV.E.13). Παρόμοια αποτελέσματα θα ισχύουν για τη γενική περίπτωση του (IV.E.11) στο πλαίσιο επαρκούς κανονικότητας στο ψ .

Η εξίσωση (IV.E.13) που ικανοποιείται από την εκτίμηση ελάχιστων τετραγώνων μπορεί να γραφεί χρησιμοποιώντας το μοντέλο παρατήρησης (IV.E.9) ως

$$\sum_{k=1}^n s'_k(\hat{\theta}_n) N_k + \sum_{k=1}^n s'_k(\hat{\theta}_n) [s_k(\theta) - s_k(\hat{\theta}_n)] = 0. \quad (\text{IV.E.14})$$

Για να αναλύσουμε τη συμπεριφορά του $\hat{\theta}_n$, ας εξετάσουμε για κάθε $\theta' \in \Lambda$ την ακολουθία των τυχαίων μεταβλητών

$$J_n(\theta; \theta') \triangleq \sum_{k=1}^n s'_k(\theta') N_k + \sum_{k=1}^n s'_k(\theta') [s_k(\theta) - s_k(\theta')]. \quad (\text{IV.E.15})$$

Σημειώστε ότι εν απουσία του θορύβου ($N_k \equiv 0$), το $\hat{\theta}_n = \theta$ είναι μια λύση στην εξίσωση πιθανοφάνειας (IV.E.14). Ωστόσο, εκτός κι αν το $\theta' = \theta$ είναι η μόνη ρίζα του

$$K_n(\theta; \theta') \triangleq \sum_{k=1}^n s'_k(\theta') [s_k(\theta) - s_k(\theta')], \quad (\text{IV.E.16})$$

η εξίσωση (IV.E.14) μπορεί να μην καταλήγει σε μια τέλεια εκτίμηση ακόμη και στην περίπτωση χωρίς θόρυβο. Άρα για συνέπεια στο (IV.E.14) θα πιστεύαμε ότι απαιτείται ο όρος θορύβου $\sum_{k=1}^n s'_k(\theta') N_k$, στο (IV.E.15) να είναι ασυμπτωτικά αμελητέος ως προς τον όρο $K_n(\theta; \theta')$ και ότι αυτός ο τελευταίος όρος να έχει μια μοναδική ρίζα ασυμπτωτικά. Καθώς η λύση στο (IV.E.14) είναι αμετάβλητη, αν διαιρέσουμε κάθε μέλος με κάποιο $d_n > 0$, μπορούμε να μετασχηματίσουμε τις παραπάνω διατυπώσεις ώστε να εφαρμόζονται στους αντίστοιχους όρους στο $J_n(\theta; \theta')/d_n$. Δηλ., αν μπορούμε να βρούμε μια ακολουθία $\{d_n\}_{n=1}^{\infty}$ τέτοια ώστε η $\sum_{k=1}^n s'_k(\theta') N_k/d_n$ να είναι ασυμπτωτικά αμελητέα και το $K_n(\theta; \theta')/d_n$ να έχει μια μοναδική ρίζα ασυμπτωτικά, τότε μπορούμε να περιμένουμε οι ρίζες του (IV.E.14) να είναι συνεπείς σε αναλογία με ό,τι συμβαίνει στην i.i.d. περίπτωση.

Σημειώστε ότι, υποθέτοντας $\mathcal{N}(0, \sigma^2)$ θόρυβο, έχουμε

Από αυτό είναι εύκολα κατανοητό ότι για δεδομένο $\theta, \theta' \in \Lambda$, η $J_n(\theta; \theta')/d_n$ συγκλίνει σε πιθανότητα σε μια σταθερά, αν και μόνο αν

$$\lim_{n \rightarrow \infty} \frac{1}{d_n^2} \sum_{k=1}^n [s'_k(\theta')]^2 = 0 \quad (\text{IV.E.18})$$

Και

$$\lim_{n \rightarrow \infty} \frac{1}{d_n} K_n(\theta; \theta') \text{ exists.} \quad (\text{IV.E.19})$$

Από το αποτέλεσμα αυτό μπορούμε να αποδείξουμε την ακόλουθη πρόταση, που είναι ανάλογη με την Πρόταση IV.D.1.

Πρόταση IV.E.1: Συνέπεια των ελάχιστων τετραγώνων

Έστω ότι έχουμε το μοντέλο του (IV.E.9) με $\mathcal{N}(0, \sigma^2)$ θόρυβο και ότι υπάρχει μια ακολουθία βαθμωτών $\{d_n\}_{n=1}^{\infty}$, τέτοια ώστε τα (IV.E.18) και (IV.E.19) ισχύουν για όλα τα $\theta' \in \Lambda$. Υποθέστε περαιτέρω ότι $s_k(\theta')$, $s'_k(\theta')$, και

$$J(\theta; \theta') \triangleq \lim_{n \rightarrow \infty} \frac{1}{d_n} K_n(\theta; \theta') \quad (\text{IV.E.20})$$

είναι όλες συνεχείς συναρτήσεις του θ' και ότι $J(\theta; \theta')$ έχει μια μοναδική ρίζα στο $\theta' = \theta$. Τότε με πιθανότητα που τείνει στο 1, η εξίσωση πιθανοφάνειας του (IV.E.13) έχει μια ακολουθία ριζών που συγκλίνει σε πιθανότητα στο θ . Συγκεκριμένα, αν το (IV.E.13) έχει μια μοναδική ρίζα $\hat{\theta}_n$ για κάθε n , τότε $\hat{\theta}_n \rightarrow \theta$ (i.p.).

Η απόδειξη αυτού του αποτελέσματος είναι σχεδόν ταυτόσημη με αυτή της Πρότασης IV.D.1 και θα τη δούμε στις ασκήσεις. Ως ένα παράδειγμα, εξετάστε το πρόβλημα της εκτίμησης εύρους σήματος (βλ. Παράδειγμα IV.D.2), στο οποίο

$$s_k(\theta) = \theta s_k, \quad k = 1, 2, \dots, n, \quad (\text{IV.E.21})$$

για μια γνωστή ακολουθία $\{s_k\}_{k=1}^{\infty}$. Στην περίπτωση αυτή, έχουμε $s'_k(\theta) = s_k$, και άρα

$\sum_{k=1}^n [s'_k(\theta)]^2 = \sum_{k=1}^n s_k^2$ και $K_n(\theta; \theta') = (\theta - \theta') \sum_{k=1}^n s_k^2$. Άρα μια επαρκής συνθήκη για συνέπεια που έπεται από την πρόταση είναι η ύπαρξη μιας αποκλίνουσας ακολουθίας $\{d_n\}_{n=1}^{\infty}$ τέτοια ώστε

$$0 < \lim_{n \rightarrow \infty} \frac{1}{d_n} \sum_{k=1}^n s_k^2 < \infty. \quad (\text{IV.E.22})$$

Η ασυμπτωτική κανονικότητα μπορεί επίσης να εξασφαλιστεί για την εκτίμηση ελάχιστων τετραγώνων στο (IV.E.9) υπό συνθήκες κανονικότητας στην ακολουθία σήματος. Σημειώστε ότι αν η $s_k(\theta)$ έχει τρίτες παραγώγους, η εξίσωση πιθανοφάνειας μπορεί να αναπτυχθεί σε μια σειρά Taylor για το θ , για να δώσει

$$\begin{aligned} & \sum_{k=1}^n s'_k(\theta) [Y_k - s_k(\theta)] \\ & + (\hat{\theta}_n - \theta) \sum_{k=1}^n [s''_k(\theta) [Y_k - s_k(\theta)] - [s'_k(\theta)]^2] \\ & + \frac{1}{2} (\hat{\theta}_n - \theta)^2 \sum_{k=1}^n [s'''_k(\hat{\theta}_n) [Y_k - s_k(\hat{\theta}_n)] - 3s''_k(\hat{\theta}_n) s'_k(\hat{\theta}_n)] = 0. \end{aligned} \quad (\text{IV.E.23})$$

με το $\bar{\theta}_n$ μεταξύ του θ και του $\hat{\theta}_n$. Αναδιατάσσοντας, έχουμε

$$\hat{\theta}_n - \theta = \frac{-\sum_{k=1}^n s'_k(\theta) N_k}{\sum_{k=1}^n s''_k(\theta) N_k - \sum_{k=1}^n [s'_k(\theta)]^2 + \frac{1}{2}(\hat{\theta}_n - \theta) \sum_{k=1}^n Z_k(\bar{\theta}_n)} \quad (\text{IV.E.24})$$

όπου

$$Z_k(\theta') \triangleq [s'''_k(\theta') \{N_k + s_k(\theta) - s_k(\theta')\} - 3s''_k(\theta') s'_k(\theta')].$$

Από την παράσταση αυτή για το σφάλμα, μπορεί να αποδειχθεί το ακόλουθο αποτέλεσμα.

Πρόταση IV.E.2: Ασυμπτωτική κανονικότητα των ελάχιστων τετραγώνων

Έστω ότι έχουμε το μοντέλο του (IV.E.9) με $\mathcal{N}(0, \sigma^2)$ θόρυβο και $\{\bar{\theta}_n\}_{n=1}^{\infty}$ είναι μια συνεπής ακολουθία εκτιμήσεων ελάχιστων τετραγώνων του θ . Έστω, περαιτέρω, ότι ισχύουν οι ακόλουθες συνθήκες κανονικότητας:

- (1) Υπάρχει μια συνάρτηση M τέτοια ώστε $|Z_k(\theta')| \leq M(N_k)$ ομοιόμορφα στο θ' και $E_{\theta}\{M(N_k)\} < \infty$ [Υποτίθεται επίσης η ύπαρξη των σχετικών παραγώγων του $s_k(\theta)$]
- (2) $\lim_{n \rightarrow \infty} (1/n) \sum_{k=1}^n [s'_k(\theta)]^2 > 0$.
- (3) $\lim_{n \rightarrow \infty} \sum_{k=1}^n [s''_k(\theta)]^2 / [\sum_{k=1}^n [s'_k(\theta)]^2]^2 = 0$.

Τότε

$$\left(\sum_{k=1}^n [s'_k(\theta)]^2 \right)^{1/2} (\hat{\theta}_n - \theta) \rightarrow \mathcal{N}(0, \sigma^2) \quad (\text{IV.E.25})$$

σε κατανομή.

Η απόδειξη αυτού του αποτελέσματος είναι παρόμοια με αυτή για την ανάλογη i.i.d. περίπτωση και θα τη δούμε στις ασκήσεις. Σημειώστε ότι η πληροφορία Fisher δίνεται εδώ από

$$I_{\theta} = \sum_{k=1}^n [s'_k(\theta)]^2 / \sigma^2. \quad (\text{IV.E.26})$$

Άρα με την ίδια λογική όπως στην i.i.d. περίπτωση, η εκτίμηση ελάχιστων τετραγώνων είναι ασυμπτωτικά επαρκής για το (IV.E.9) με $\mathcal{N}(0, \sigma^2)$ σφάλματα.

Το πρόβλημα εκτίμησης εύρους σήματος, $s_k(\theta) = \theta s_k$, αποτελεί και πάλι ένα απλό παράδειγμα. Στην περίπτωση αυτή, οι συνθήκες διαφορισιμότητας είναι τετριμμένες, $Z_k(\theta') \equiv 0$, και $s_k''(\theta) = 0$. Άρα η μόνη συνθήκη απαραίτητη για ασυμπτωτική κανονικότητα είναι ότι $\lim_{n \rightarrow \infty} \sum_{k=1}^n s_k^2/n > 0$. Θυμηθείτε, ωστόσο, ότι οι επιθυμητές ιδιότητες της MLE σ' αυτή τη συγκεκριμένη περίπτωση έπονται από την ευθεία ανάλυση (ακόμη και για πεπερασμένο n), όπως είδαμε στο Παράδειγμα IV.D.2.

Ένα λιγότερο προφανές παράδειγμα δίνεται από τα παρακάτω.

Παράδειγμα IV.E.1: Αναγνώριση ενός γραμμικού συστήματος πρώτης τάξης

Μια σημαντική κλάση εφαρμογών των προβλημάτων εκτίμησης παραμέτρου εμπίπτει στο πλαίσιο της αναγνώρισης συστήματος, στο οποίο επιθυμούμε να συνάγουμε τη δομή κάποιου συστήματος εισόδου/εξόδου, εισάγοντας μια είσοδο και παρατηρώντας την έξοδο. Ένα από τα απλούστερα πιθανά προβλήματα αναγνώρισης είναι αυτό της αναγνώρισης ενός σταθερού πρώτης τάξης χρονικά αμετάβλητου γραμμικού συστήματος. Αυτός ο τύπος συστήματος μπορεί να περιγραφεί από το μοντέλο σήματος

$$s_k(\theta) = \theta s_{k-1}(\theta) + u_k, \quad k = 1, 2, \dots, n, \quad (\text{IV.E.27})$$

όπου $|\theta| < 1$ και $\{u_k\}_{k=1}^n$ είναι η γνωστή ακολουθία εισόδου. Σημειώστε ότι το θ εδώ είναι ο συντελεστής της ομογενούς εξίσωσης $s_k(\theta) = \theta s_{k-1}(\theta)$ και άρα αυτή η παράμετρος προσδιορίζει πλήρως το σύστημα, εφόσον έχουν γίνει οι υποθέσεις της γραμμικότητας, του αμετάβλητου χρόνου και της μοναδιαίας τάξης. Η παρατήρηση του συστήματος εξόδου αλλοιώνεται συνήθως από το θόρυβο μέτρησης, άρα υποθέτοντας ότι αυτός ο θόρυβος είναι i.i.d., η εκτίμηση του θ είναι ένα πρόβλημα της μορφής του IV.E.9. Εξετάζουμε την περίπτωση των $\mathcal{N}(0, \sigma^2)$ σφαλμάτων και της εκτίμησης ελάχιστων τετραγώνων του θ .

Ας υποθέσουμε ότι το σύστημα (IV.E.27) είναι αρχικά σε ηρεμία [$s_0(\theta) = 0$], περίπτωση στην οποία η λύση στο (IV.E.27) δίνεται από

$$s_k(\theta) = \sum_{l=1}^k \theta^{k-l} u_l. \quad (\text{IV.E.28})$$

Αν το θ μπορεί ή όχι να αναγνωριστεί (καθώς $n \rightarrow \infty$) εξαρτάται από την ακολουθία εισόδου $\{u_k\}_{k=1}^n$. Εξετάστε, για παράδειγμα, ένα συνεχές σήμα εισόδου $u_k = 1$ για όλα τα $k \geq 1$. Η έξοδος τότε είναι

$$s_k(\theta) = \sum_{l=1}^k \theta^{l-1} = \sum_{m=0}^{k-1} \theta^m = \frac{1 - \theta^k}{1 - \theta},$$

και

$$s'_k(\theta) = \frac{(1 - \theta^k) - k\theta^{k-1}(1 - \theta)}{(1 - \theta)^2}.$$

Αυτό συνεπάγεται ότι

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n [s'_k(\theta)]^2 = \frac{(2 - \theta)^2}{(1 - \theta)^4} \quad (\text{IV.E.29})$$

Και

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{k=1}^n [s'_k(\theta') | s_k(\theta) - s_k(\theta')] \right] = \frac{(2 - \theta')(\theta - \theta')}{(1 - \theta')^2(1 - \theta)}. \quad (\text{IV.E.30})$$

Καθώς το (IV.E.30) έχει μια μοναδική ρίζα στο $\theta' = \theta$ και τα σχετικά μεγέθη είναι συνεχή για $|\theta'| < 1$, τα (IV.E.29) και (IV.E.30) συνεπάγονται ότι η υπόθεση της Πρότασης IV.E.1 ικανοποιείται με $d_n = n$. Άρα έχουμε μια συνεπή ακολουθία ριζών στην εξίσωση πιθανοφάνειας. [Στην ουσία, αφού το $J(\theta; \theta')$ είναι δεσμευμένο μακριά από το μηδέν σε μια γειτονιά του $\theta' = \theta$, μπορεί να αποδειχθεί ότι οποιαδήποτε ακολουθία ριζών είναι συνεπής.]

Δεν είναι δύσκολο να κατανοηθεί γιατί η συνεπής εκτίμηση του θ είναι δυνατή στην περίπτωση αυτή. Σημειώστε ότι η ασυμπτωτική τιμή του $s_k(\theta)$ είναι $1/(1 - \theta)$. Άρα, το σύστημα επιτυγχάνει μια μοναδική σταθερής κατάστασης τιμή για κάθε τιμή της παραμέτρου θ . Από αυτό θα περιμέναμε να είναι δυνατόν να προσδιοριστεί η τιμή της παραμέτρου τέλεια με την παρατήρηση της θορυβώδους εξόδου για $k = 1, 2, \dots, \infty$, αφού ο θόρυβος μπορεί να γίνει μέσος σε άπειρο χρόνο. Από την άλλη, έστω ότι χρησιμοποιούμε μια είσοδο με πεπερασμένη διάρκεια. Τότε, αφού το σύστημα είναι σταθερό, η σταθερής κατάστασης έξοδος του συστήματος είναι μηδέν για κάθε τιμή παραμέτρου.

Είναι εύκολο να δούμε ότι η υπόθεση της Πρότασης IV.E.1 αποτυγχάνει να ισχύσει στην περίπτωση αυτή. Αν ο θόρυβος μέτρησης δεν ήταν παρών, μπορεί να ήταν δυνατός ο προσδιορισμός της παραμέτρου στην περίπτωση αυτή από τη μεταβατική συμπεριφορά. Ωστόσο, η παρουσία του θορύβου απαιτεί η παράμετρος να είναι αναγνωρίσιμη και στη σταθερή κατάσταση. Η ιδιότητα μιας εισόδου που παράγει αυτό το αποτέλεσμα είναι ενίοτε γνωστή ως *διατήρηση της διέγερσης*. (Μια σχετική ιδιότητα που απαιτείται ενίοτε από μια είσοδο στα προβλήματα αναγνώρισης εισόδου είναι ο *επαρκής πλούτος*. (Βασικά, η ιδιότητα αυτή σημαίνει ότι το περιεχόμενο συχνότητα του σήματος εισόδου είναι επαρκώς πλούσιο ώστε να διεγείρει όλες τις ταλαντευόμενες λειτουργίες του συστήματος.)

Για το σταθερό σήμα εισόδου, η Πρόταση IV.E.2 δεν μπορεί να εφαρμοστεί ευθέως σε αυτό το μοντέλο με $\Lambda = (-1, 1)$ γιατί η $Z_k(\theta')$ δεν μπορεί να είναι φραγμένη ομοιόμορφα στο σύνολο αυτό. Ωστόσο, αν υποθέσουμε ότι το θ είναι φραγμένο μακριά από τη μονάδα (δηλ., αν έχουμε $\Lambda = (-1, \theta_u)$ με $\theta_u < 1$) τότε οι συνθήκες κανονικότητας της Πρότασης IV.D.4 πράγματι ισχύουν και η ασυμπτωτική κανονικότητα και επάρκεια των συνεπών ριζών της εξίσωσης πιθανοφάνειας έπονται. Σημειώστε ότι η ασυμπτωτική διακύμανση του $\sqrt{n}(\theta_n - \theta)$ στην περίπτωση αυτή είναι $\sigma^2(1 - \theta)^4 / (2 - \theta)^2$.

Κάποιες πρόσθετες πλευρές των εκτιμήσεων μέγιστης πιθανοφάνειας και ελάχιστων τετραγώνων των μονών παραμέτρων αναλύονται παρακάτω και στο Κεφάλαιο VII. Ωστόσο, πριν αφήσουμε το ζήτημα αυτό προς το παρόν, παρατηρούμε ότι οι ιδιότητες των ελάχιστων τετραγώνων που συνοψίζονται στις Προτάσεις IV.E.1 και IV.E.2 ισχύουν πιο γενικά.

Συγκεκριμένα έχουμε το ακόλουθο.

Πρόταση IV.E.3: Συνέπεια και Ασυμπτωτική Κανονικότητα των Ελάχιστων Τετραγώνων με μη κατά Gauss θόρυβο.

Οι προτάσεις IV.E.1 και IV.E.2 παραμένουν σε ισχύ αν η υπόθεση

$$N_k \sim \mathcal{N}(0, \sigma^2) \text{ αντικατασταθεί από την υπόθεση } E\{N_k\} = 0 \text{ και } E\{N_k^2\} = \sigma^2 < \infty$$

Σημειώστε, ωστόσο, ότι αυτό το αποτέλεσμα δεν συνεπάγεται ότι τα ελάχιστα τετράγωνα είναι ασυμπτωτικά επαρκή όταν ο θόρυβος δεν είναι κατά Gauss, καθώς η πληροφορία Fisher δεν δίνεται πλέον από το (IV.E.26) στη μη κατά Gauss περίπτωση.

IV.E.3 ΑΝΘΕΚΤΙΚΗ ΕΚΤΙΜΗΣΗ ΤΩΝ ΜΟΝΩΝ ΠΑΡΑΜΕΤΡΩΝ

Εξετάστε ξανά το μοντέλο του (IV.E.9), στο οποίο έχουμε παρατηρήσει ότι οι MLE είναι ασυμπτωτικά βέλτιστες υπό την έννοια της ελάχιστης ασυμπτωτικής διακύμανσης.

Όπως αναλύθηκε στην Ενότητα III.E, στατιστικά μοντέλα όπως αυτό είναι μόνο κατά προσέγγιση έγκυρα στην πράξη και ένα σημαντικό ζήτημα που προκύπτει σε τέτοιες καταστάσεις είναι το αν οι διαδικασίες που έχουν σχεδιαστεί για ένα συγκεκριμένο μοντέλο είναι *ανθεκτικές*. Δηλ. το αν η απόδοσή τους δεν είναι ευαίσθητη σε μικρές αλλαγές στο μοντέλο.

Εξετάστε, για παράδειγμα, ένα ονομαστικό μοντέλο στο οποίο τα δείγματα θορύβου έχουν την $N(0,1)$ κατανομή. Τότε, στο πλαίσιο της κανονικότητας και υποθέτοντας ότι το

$$\epsilon_\theta \triangleq \lim_{n \rightarrow \infty} \sum_{k=1}^n [s'_k(\theta)]^2 / n$$

υφίσταται και είναι θετικό, η εκτίμηση ελάχιστων τετραγώνων είναι ασυμπτωτικά $\mathcal{N}(\theta, 1/n\epsilon_\theta)$. Έστω, ωστόσο, ότι η πραγματική στατιστική συμπεριφορά του θορύβου περιγράφεται από μία pdf που είναι μόνο κατά προσέγγιση $N(0,1)$. Για παράδειγμα, υποθέστε ότι η πυκνότητα θορύβου f είναι της μορφής

$$f(x) = (1 - \epsilon) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + \epsilon h(x), \quad x \in \mathbb{R}, \quad (\text{IV.E.31})$$

όπου το $h(x)$ είναι μια αυθαίρετη πυκνότητα, συμμετρική γύρω από το μηδέν και με διακύμανση

$$\sigma_h^2 \triangleq \int_{-\infty}^{\infty} x^2 h(x) dx$$

πεπερασμένη αλλά όχι φραγμένη. Τότε, με την Πρόταση IV.E.3 η εκτίμηση ελάχιστων τετραγώνων θα έχει ασυμπτωτική διακύμανση

$$v_h^2 \sim \frac{(1 - \epsilon) + \epsilon \sigma_h^2}{n e_\theta}, \quad (\text{IV.E.32})$$

Σημειώστε ότι το v_h^2 μπορεί να είναι αυθαίρετα μεγάλο για κάθε $\epsilon > 0$ αφού το σ_h^2 δεν είναι φραγμένο. Συγκεκριμένα, η χειρότερης περίπτωσης ασυμπτωτική διακύμανση στην κλάση των πυκνοτήτων (IV.E.31) είναι

$$\sup_h [(1 - \epsilon) + \epsilon \sigma_h^2] = \infty \quad (\text{IV.E.33})$$

για κάθε $\epsilon > 0$

Αυτό υποδεικνύει μια έλλειψη ανθεκτικότητας της εκτίμησης ελάχιστων τετραγώνων για καταστάσεις στις οποίες ένα μικρό κλάσμα των δειγμάτων θορύβου μπορεί να προέρχεται από μια κατανομή υψηλής διακύμανσης. (Αυτό μπορεί να συμβεί, για παράδειγμα, σε μετρήσεις ραντάρ, στις οποίες μπορεί να είναι παρούσα υψηλής διακύμανσης ωστική παρεμβολή σ' ένα μικρό κλάσμα ϵ των μετρήσεων. Παρατηρήσεις που είναι βέβαια μεγάλες για ένα δεδομένο ονομαστικό μοντέλο ενίοτε καλούνται *ακραίες τιμές*). Όπως στα προβλήματα ανίχνευσης σήματος που αναλύθηκαν στην Ενότητα III.E, απαιτείται μια εναλλακτική στην ασυμπτωτική διακύμανση σ' ένα ονομαστικό μοντέλο, ως ένα σχεδιαστικό κριτήριο για τέτοιες καταστάσεις.

Έστω ότι η πυκνότητα θορύβου f στο (IV.E.9) είναι μια άρτια συμμετρική συνάρτηση. Εξετάστε εκτιμήσεις του θ της μορφής

$$\sum_{k=1}^n s'_k(\hat{\theta}_n) \psi[Y_k - s_k(\hat{\theta}_n)] = 0, \quad (\text{IV.E.34})$$

όπου ψ είναι μια γενική περιττή συμμετρική συνάρτηση. Με $\psi(x) = x$, το (IV.E.34) δίνει την εκτίμηση ελάχιστων τετραγώνων και με $\psi(x) = -f'(x)/f(x)$, το (IV.E.34) δίνει την MLE. Οι εκτιμήσεις αυτής της μορφής είναι γνωστές ως *M-εκτιμήσεις*. Υποθέτοντας ότι $0 < e_\theta < \infty$ και με κανονικότητα στα ψ, f , και $\{s_k(\theta)\}_{k=1}^\infty$, μπορεί να αποδειχθεί, με τη χρήση των παραπάνω τεχνικών, ότι οι *M-εκτιμήσεις* είναι συνεπείς και ασυμπτωτικά $N[\theta, V(\psi, f)/n e_\theta]$, όπου

$$V(\psi, f) \triangleq \frac{\int \psi^2 f}{(\int \psi' f)^2} \quad (\text{IV.E.35})$$

με $\psi'(x) = d\psi(x)/dx$.

Με βάση αυτές τις ιδιότητες, ένας πιθανός τρόπος σχεδιασμού ενός ανθεκτικού εκτιμητή για μια κλάση αβεβαιότητας \mathcal{F} των πυκνοτήτων θορύβου είναι να αναζητηθεί μια συνάρτηση ψ που ελαχιστοποιεί την χειρότερης περίπτωσης διακύμανση M -εκτίμησης,

$\sup_{f \in \mathcal{F}} V(\psi, f)$. Δηλαδή, μια πιθανή μέθοδος σχεδιασμού είναι να εστιαστεί η προσοχή στις M -εκτιμήσεις και να λυθεί

$$\min_{\psi} \sup_{f \in \mathcal{F}} V(\psi; f). \quad (\text{IV.E.36})$$

Το πρόβλημα (IV.E.36) έχει μελετηθεί από τον Huber (1981) για γενικά σύνολα \mathcal{F} . Σε κατάλληλες συνθήκες, η λύση του είναι βασικά ως ακολούθως.

Εξετάστε το συναρτησιακό

$$I(f) \triangleq \int (f')^2 / f, \quad (\text{IV.E.37})$$

και έστω ότι το f_L είναι μια πυκνότητα \mathcal{F} που ελαχιστοποιεί το $I(f)$ στο \mathcal{F} . Δηλ.

$$I(f_L) = \min_{f \in \mathcal{F}} I(f). \quad (\text{IV.E.38})$$

Τότε η M -εκτίμηση με ψ -συνάρτηση $\psi_R(x) = -f'_L(x)/f_L(x)$, λύνει το (IV.E.36). Σημειώστε ότι για κάθε f ,

$$V(\psi, f) |_{\psi = -f'/f} = 1/I(f), \quad (\text{IV.E.39})$$

οπότε το $[ne_{\theta} I(f)]^{-1}$ είναι η ασυμπτωτική διακύμανση της MLE στο μοντέλο μας με δεδομένο f . [Η πληροφορία Fisher εδώ είναι $ne_{\theta} I(f)$]. Άρα f_L είναι το μέλος του \mathcal{F} , η αντίστοιχη βέλτιστη εκτίμηση του οποίου (η MLE) έχει τη χειρότερη βέλτιστη απόδοση. Για το λόγο αυτό η f_L μπορεί να θεωρηθεί η *λιγότερο ευνοϊκή πυκνότητα* και η ανθεκτική M -εκτίμηση είναι η καλύτερη εκτίμηση γι' αυτό το λιγότερο ευνοϊκό μοντέλο.

Το πρόβλημα $\min_{f \in \mathcal{F}} I(f)$ έχει λυθεί για έναν αριθμό μοντέλων αβεβαιότητας \mathcal{F} [βλέπε Huber (1981)]. Για παράδειγμα, για το ϵ -μολυσμένο $N(0,1)$ μοντέλο του (IV.E.31), η λιγότερο ευνοϊκή πυκνότητα δίνεται από

$$f_L(x) = \begin{cases} (1 - \epsilon) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} & \text{if } |x| \leq k' \\ (1 - \epsilon) e^{-k'(|x|-k')} \frac{1}{\sqrt{2\pi}} e^{-(k')^2/2} & \text{if } |x| > k', \end{cases} \quad (\text{IV.E.40})$$

όπου k' είναι μια σταθερά που δίνεται από τη λύση στο

$$(1 - \epsilon)^{-1} = 2\Phi(k') - 1 + \frac{1}{k'} \left(\frac{2}{\pi}\right)^{1/2} e^{-(k')^2/2}. \quad (\text{IV.E.41})$$

Η αντίστοιχη ψ ανθεκτική συνάρτηση είναι

$$\psi_k(x) = \begin{cases} x & \text{if } |x| \leq k' \\ k' \operatorname{sgn}(x) & \text{if } |x| > k'. \end{cases} \quad (\text{IV.E.42})$$

Άρα, όπως στο ανάλογο πρόβλημα ελέγχου υπόθεσης, η ανθεκτικότητα επιτυγχάνεται με τον περιορισμό των επιδράσεων των ακραίων τιμών.

Για περαιτέρω ανάλυση αυτής και άλλων προσεγγίσεων της ανθεκτικής εκτίμησης, ο αναγνώστης παραπέμπεται στο ερευνητικό άρθρο των Kassam και Poor (1985) και τα βιβλία των Huber (1981) και Hampel, *et al.* (1986).

IV.E.4 ΑΝΑΔΡΟΜΙΚΗ ΕΚΤΙΜΗΣΗ ΠΑΡΑΜΕΤΡΟΥ

Από τις προηγούμενες αναλύσεις βλέπουμε ότι οι εκτιμήσεις μέγιστης πιθανοφάνειας συχνά έχουν καλές ιδιότητες, ειδικότερα όταν το μέγεθος δείγματος είναι μεγάλο. Ωστόσο, ενίοτε έχουν το μειονέκτημα να είναι πολύ δύσκολο να υπολογιστούν. Για παράδειγμα, με n i.i.d. δείγματα που λαμβάνονται από την πυκνότητα f_θ , ο υπολογισμός της MLE απαιτεί τη μεγιστοποίηση της συνάρτησης $\sum_{k=1}^n \log f_\theta(y_k)$. Αν η μεγιστοποίηση του θ δεν μπορεί να βρεθεί ως μια κλειστής μορφής συνάρτηση του \underline{y} , θα πρέπει να χρησιμοποιηθεί μια επαναληπτική τεχνική για να βρεθεί η $\hat{\theta}_{ML}(\underline{y})$. Αυτό απαιτεί την αποθήκευση και τον ταυτόχρονο χειρισμό όλων των n δειγμάτων (εκτός κι αν είναι διαθέσιμη μια πιο ολιγοδιάστατη επαρκής στατιστική), μια διαδικασία που δεν είναι επιθυμητή αν το n είναι πολύ μεγάλο. Είναι άρα ενίοτε επιθυμητό να εξετάσουμε εναλλακτικές της μέγιστης πιθανοφάνειας που μπορούν να εφαρμοστούν με αναδρομικό ή ακολουθιακό τρόπο, ώστε η συμβολή κάθε δείγματος στην εκτίμηση να υπολογίζεται με τη λήψη του δείγματος.

Μια τέτοια τεχνική εκτίμησης συνιστάται από την MLE. Συγκεκριμένα, εξετάστε μια συνεπή ακολουθία $\{\hat{\theta}_n\}_{n=1}^\infty$ που λύνει την εξίσωση πιθανοφάνειας

$$\sum_{k=1}^n \psi(Y_k; \hat{\theta}_n) = 0 \quad (\text{IV.E.43})$$

με $\psi(Y_k; \theta) = \partial \log f_\theta(Y_k) / \partial \theta$, όπως προηγουμένως. Αφού η $\{\hat{\theta}_n\}_{n=1}^\infty$ είναι συνεπής, η διαφορά, $\hat{\theta}_n - \hat{\theta}_{n-1}$ συγκλίνει στο μηδέν καθώς $n \rightarrow \infty$. Άρα το (IV.E.43) μπορεί να προσεγγισθεί, αναπτύσσοντας .. $\hat{\theta}_{n-1}$ για να δώσει

$$\sum_{k=1}^n \psi(Y_k; \hat{\theta}_{n-1}) + (\hat{\theta}_n - \hat{\theta}_{n-1}) \sum_{k=1}^n \psi'(Y_k; \hat{\theta}_{n-1}) \sim 0, \quad (\text{IV.E.44})$$

με $\psi'(Y_k; \theta) = \partial \psi(Y_k; \theta) / \partial \theta$. Αναδιατάσσοντας το (IV.E.44) έχουμε

$$\hat{\theta}_n \sim \hat{\theta}_{n-1} - \frac{\sum_{k=1}^n \psi(Y_k; \hat{\theta}_{n-1})}{\sum_{k=1}^n \psi'(Y_k; \hat{\theta}_{n-1})}. \quad (\text{IV.E.45})$$

Αφού το $\hat{\theta}_{n-1}$ λύνει το $\sum_{k=1}^{n-1} \psi(Y_k; \hat{\theta}_{n-1}) = 0$, το άθροισμα αριθμητή στο δεξί μέλος του (IV.E.45) έχει μόνο έναν όρο, $\psi(Y_n; \hat{\theta}_{n-1})$. Έστω ότι γράφουμε το άθροισμα παρονομαστή ως

$$n \left[\frac{1}{n} \sum_{k=1}^n \psi'(Y_k; \hat{\theta}_{n-1}) \right]. \quad (\text{IV.E.46})$$

Τώρα, ο αδύναμος νόμος των μεγάλων αριθμών συνεπάγεται ότι

$$-\frac{1}{n} \sum_{k=1}^n \psi'(Y_k; \theta) \rightarrow i_\theta \quad (i.p.),$$

όπου $i_\theta = -E_\theta\{\psi'(Y_k; \theta)\} = E\{\psi^2(Y_k; \theta)\}$ είναι η πληροφορία Fisher ανά δείγμα. Καθώς $\hat{\theta}_{n-1} \rightarrow \theta$, μπορούμε να προσεγγίσουμε

$$\frac{1}{n} \sum_{k=1}^n \psi'(Y_k; \hat{\theta}_{n-1}) \sim i_{\hat{\theta}_{n-1}}. \quad (\text{IV.E.47})$$

Συνδυάζοντας τα (IV.E.45) και (IV.E.47) έχουμε ότι, ασυμπτωτικά, μια συνεπής ακολουθία λύσεων στην εξίσωση πιθανοφάνειας θα ικανοποιεί

$$\hat{\theta}_n \sim \hat{\theta}_{n-1} + \frac{\psi(Y_n; \hat{\theta}_{n-1})}{ni_{\hat{\theta}_{n-1}}}. \quad (\text{IV.E.48})$$

Αυτή είναι μια ασυμπτωτική αναδρομική εξίσωση για το $\hat{\theta}_n$, αφού το $\hat{\theta}_n$ υπολογίζεται από το $\hat{\theta}_{n-1}$ και το Y_n μόνο.

Αποδεικνύεται ότι η (μη-ασυμπτωτική) αναδρομή

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \frac{\psi(Y_n; \hat{\theta}_{n-1})}{n \dot{\ell}_{\hat{\theta}_{n-1}}}, \quad n = 1, \dots, \quad (\text{IV.E.49})$$

(με το $\hat{\theta}_0$ αυθαίρετο) που προτείνεται από το (IV.E.48) έχει τις ίδιες επιθυμητές ασυμπτωτικές ιδιότητες (δηλ., συνέπεια και επάρκεια) όπως η MLE στο πλαίσιο κανονικότητας του μοντέλου. Η αναδρομή αυτή είναι ένα παράδειγμα μιας πιο γενικής κλάσης αλγορίθμων εκτίμησης αναδρομικής παραμέτρου, που είναι γνωστοί ως *αλγόριθμοι στοχαστικής προσέγγισης*. Λόγω της αναδρομικής τους φύσης, αυτοί οι αλγόριθμοι έχουν σημαντικό ενδιαφέρον σε εφαρμογές στις οποίες η απευθείας σύνδεση και πραγματικού χρόνου εκτίμηση παραμέτρου είναι απαραίτητη. Σε μετασχηματισμένη μορφή είναι επίσης χρήσιμοι στον πραγματικού χρόνου εντοπισμό αργά μεταβαλλόμενων παραμέτρων. Ο αναγνώστης που ενδιαφέρεται για περαιτέρω χαρακτηριστικά τέτοιων αλγορίθμων παραπέμπεται στο βιβλίο των Nev'ison και Has'minskii (1973). Παρόμοιοι αναδρομικοί μετασχηματισμοί της MLE και της εκτίμησης ελάχιστων τετραγώνων για προβλήματα μεταβαλλόμενου χρόνου όπως αυτό του (IV.E.9) έχουν επίσης αναπτυχθεί. Ο αναγνώστης παραπέμπεται στους Ljung και Soderstrom (1982) και στους Goodwin και Sin (1984) για την ανάπτυξη αυτών των ιδεών.

Αναφορές

1. Anderson, B. D. O., and J. B. Moore (1979), *Optimal Filtering* (Prentice- Hall: Englewood Cliffs, NJ).
2. Apostol, A. (1974), *Mathematical Analysis*, 2nd Ed. (Addison-Wesley: Reading, MA).
3. Ash, R. B. (and M. F. Gardner (1975), *Topics in Stochastic Processes* (Academic: New York).
4. Barton, R. J., and H. V. Poor (1992), "On Generalized Signal-to-Noise Ratios in Quadratic Detection," *Mathematics of Control, Signals and Systems*, vol, 5, no. 1, pp. 81-91.
5. Benes, V. E. (1981), "Exact Finite-Dimensional Filters for Certain Diffusions with Nonlinear Drift," *Stochastics*, vol. 5, pp. 65-92.
6. Benes, V. E. (1987), "Nonlinear Filtering: Problems, Examples, Applications," Chapter 1 in *Advances in Statistical Signal Processing, Volume 1: Estimation*, H. V. Poor, Ed. (JAI Press: Greenwich, CT).
7. Bierman, G. J. (1977), *Factorization Methods for Discrete Sequential Estimation* (Academic: New York).
8. Billingsley, P. (1979), *Probability and Measure* (Wiley: New York).
9. Boekee. D. E., and J. C. Ruilenbeck (1981), "A Class of Lower Bounds on the Bayesian Probability of Error," *Inform. Sciences*, vol. 25, pp. 21-25.
10. Breiman, L. (1968), *Probability* (Addison-Wesley: Reading, MA).
11. Brockett, R., and J. M. C. Clark (1980), "The Geometry of the Conditional Density Equation," in *Analysis and Optimization of Stochastic Systems*, O. L. R. Jacobs, Ed, (Academic: New York).
12. Carlyle, J. W. (1968), "Nonparametric Methods in Detection Theory," Chapter 8 in *Communication Theory*, A. V. Balakrishnan, Ed. (McGraw- Hill: New York).
13. Kailath, T. (1981), *Lectures on Wiener and Kalman Filtering* (Springer- Verlag; New York).
14. Kassam, S. A., and H. V. Poor (1985), "Robust Techniques for Signal Processing: A Survey," *Proc. IEEE*, vol. 73, pp. 433-481.
15. Kassam, S. A., and J. B. Thomas (1980), *Nonparametric Detection: Theory and Applications* (Dowden, Hutchinson & Ross: Stroudsburg, PA).
16. Kendall, M. G. (1948), *Rank Correlation Methods* (Griffin: London).
17. Kendall, M. G., and A. Stuart (1961), *The Advanced Theory of Statistics*, Vol. 1 (Hafner: New York).
18. Kobayashi, I., and J. B. Thomas (1967), "Distance Measures and Related Criteria," *Proc. 5 th Ann. Allerton Conf. Circuit and System Theory*, Monticello, IL, pp. 491-500.
19. Kullback, S. (1959), *Information Theory and Statistics* (Wiley: New York).
20. Lehmann, E. L. (1983), *Theory of Point Estimation* (Wiley: New York).
21. Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (Wiley: New York).
22. Lipster, R. S., and A. N. Shiriyayev (1977), *Statistics of Random Processes I: General Theory* (Springer-Verlag: New York).

23. Ljung, L., and T. Soderstrom (1982), Theory and Practice of Recursive Identification (MIT Press: Cambridge, MA).
24. Lovitt, W. V. (1950), Linear Integral Equations (Dover: New York).
25. Thomas, J. B. (1971), An Introduction to Applied Probability and Random Processes (Wiley: New York).
26. Thomas, J. B. (1986), An Introduction to Applied Probability (Springer- Verlag: New York),
27. Trench, W. F. (1964), "An Algorithm for the Inversion of Finite Toeplitz Matrices," J SIAM, vol. 12, pp. 515-522.
28. Van Trees, H. L. (1968), Detection, Estimation and Modulation Theory, Part I (Wiley: New York).
29. Viterbi, A. (1968), Principles of Coherent Communications (Wiley: New York).
30. Wong, E. (1983), Introduction to Random Processes (Springer- verlag: New York).
31. Wong, E., and B. Hajek (1985), Stochastic Processes in Engineering Systems (Springer-Verlag: New York).
32. Yao, K., and R. M. Tobin (1976), "Moment Space Upper and Lower Bounds for Digital Systems with Intersymbol Interference," IEEE Trans. Inform. Theory, vol. IT-22, pp. 65-74.