

Κεφάλαιο 6

Εκτίμηση Παραμέτρων.

6.1 Ιδιότητες Εκτιμητών

Στην καθημερινή πρακτική, εκτίμηση είναι μία κατά προσέγγιση διαδικασία που οδηγεί σε μη-ιδιαιτέρως ακριβή αποτελέσματα. Συνήθως αντιπαραβάλλουμε τις έννοιες «εκτίμησης» και «προσδιορισμού» ή «υπολογισμού». Όσον αφορά στη στατιστική και την ανάλυση των πειραματικών δεδομένων, η εκτίμηση είναι ιδιαίτερα ακριβής μέθοδος που έχει ως αποτέλεσμα τον προσδιορισμό ποσοτήτων. Η ακρίβεια της εκτίμησης εξαρτάται από «στατιστικούς» παράγοντες, επί παραδείγματι: από το πλήθος των παρατηρήσεων. Επιπλέον, η ακρίβεια κάθε εκτίμησης μπορεί και αυτή να εκτιμηθεί.

Εκτίμηση είναι η διαδικασία προσδιορισμού της τιμής μίας ή περισσότερων παραμέτρων χρησιμοποιώντας έναν περιορισμένο αριθμό, N , παρατηρήσεων (μετρήσεων). Οι παράμετροι, επί παραδείγματι: φυσικές σταθερές που καθορίζουν την εξέλιξη φυσικών συστημάτων, έχουν σταθερές (αλλά άγνωστες) τιμές, τις οποίες θα καλούμε εφεξής, «πραγματικές τιμές» των παραμέτρων.

Το αποτέλεσμα της εκτίμησης της τιμής παραμέτρου (*εκτίμηση σημείου*) συμπεριφέρεται ως τυχαία μεταβλητή. Εάν δηλαδή επαναλάβουμε την εκτίμηση της τιμής της ίδιας παραμέτρου χρησιμοποιώντας ένα άλλο σύνολο, N τω πλήθος, παρατηρήσεων που συλλέχθηκαν από την ίδια πειραματική συσκευή υπό τις ίδιες ακριβώς συνθήκες, θα καταλήξουμε σε διαφορετική τιμή.

Λόγω του «στατιστικού χαρακτήρα» της εκτίμησης, πολλές φορές ενδιαφερόμαστε για τον προσδιορισμό μίας περιοχής τιμών (*εκτίμηση περιοχής*) η οποία είναι πολύ πιθανόν να περιέχει την πραγματική τιμή της παραμέτρου.

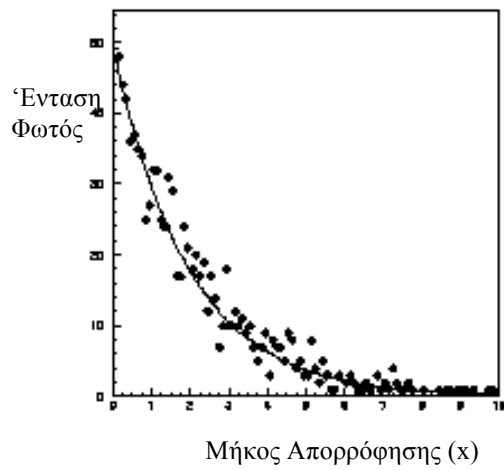
Παράδειγμα 6.1.1.

Στο σχήμα 6.1.1, φαίνονται οι μετρήσεις της έντασης φωτεινής δέσμης που διαδίδεται μέσα σε ημιδιαφανές υγρό σαν συνάρτηση του μήκους διείσδυσης του φωτός στο υγρό. Η εξασθένηση της έντασης της δέσμης, I , οφείλεται στην απορρόφηση του φωτός από το υλικό του μέσου και ακολουθεί τον εκθετικό νόμο $I = I_0 \cdot e^{-\lambda x}$, όπου x είναι το μήκος διείσδυσης.

Ο συντελεστής λ είναι χαρακτηριστικό μέγεθος εξαρτούμενο από τις ατομικές (μοριακές) ιδιότητες του υλικού. Οι μετρήσεις, που παρίστανται γραφικά στο σχήμα 5.1.1, μπορούν να χρησιμοποιηθούν στη εκτίμηση της τιμής της παραμέτρου λ .

Επειδή το αποτέλεσμα της εκτίμησης (εν προκειμένω εκτίμηση σημείου) είναι προϊόν στατιστικής επεξεργασίας των μετρήσεων και συμπεριφέρεται ως τυχαία μεταβλητή, επιβάλλεται να προσδιορισθεί και το σφάλμα της εκτίμησης.

Εναλλακτικά, θα μπορούσαμε να εκτιμήσουμε μία περιοχή τιμών, ώστε η πραγματική τιμή του λ να έχει 90% (σαν παράδειγμα) πιθανότητα να ανήκει σ' αυτή την περιοχή. Σ' αυτή την περίπτωση αναφερόμαστε σε εκτίμηση περιοχής.



Σχήμα 6.1.1: Μετρήσεις της έντασης δέσμης φωτός κατά την διάδοση του μέσα σε απορροφητικό μέσο, σαν συνάρτηση του μήκους διείσδυσης (απορρόφησης). Όλα τα μεγέθη είναι σε αυθαίρετες μονάδες .

Για να εκτιμήσουμε την τιμή μίας παραμέτρου από ένα πλήθος παρατηρήσεων, κάποιος θα πρέπει πρώτα να επιλέξει μία κατάλληλη συνάρτηση των παρατηρήσεων, την οποία καλούμε **Εκτιμητή**. Είναι φανερό ότι ο εκτιμητής είναι στατιστική συνάρτηση και όλα όσα έχουμε μάθει μέχρι τώρα σχετικά με την πληροφορία που μεταφέρει η στατιστική συνάρτηση ισχύουν και για την περίπτωση του εκτιμητή.

Έχοντας διαλέξει έναν εκτιμητή, μπορούμε να συζητήσουμε για τις ιδιότητες του και να αξιολογήσουμε τα αποτελέσματα της εκτίμησης χρησιμοποιώντας τέσσερα πολύ σημαντικά κριτήρια.

- *Consistency* **Συνέπεια**
- *Unbiasedness* **Έλλειψη Προκατάληψης**
- *Efficiency* **Αποδοτικότητα**
- *Robustness* **Σταθερότητα**

6.1.1 Συνέπεια:

Ένας εκτιμητής λέγεται ότι είναι συνεπής (consistent) εάν η εκτίμηση της τιμής της παραμέτρου, $\hat{\theta}_n$, συγκλίνει προς την πραγματική της τιμή, θ_0 , καθώς ο αριθμός των παρατηρήσεων, n , αυξάνει.

Εδώ επιλέγουμε να μιλήσουμε για σύγκλιση πιθανότητας.

Έστω ότι $\hat{\theta}_n$ είναι η εκτίμηση της παραμέτρου θ , η οποία βασίζεται σε n παρατηρήσεις, $\{x_1, x_2, \dots, x_n\}$:

$$\hat{\theta}_n = \Theta(x_1, x_2, \dots, x_n) \quad 6.1.1$$

όπου η στατιστική συνάρτηση $\Theta(x_1, x_2, \dots, x_n)$ είναι ο εκτιμητής.

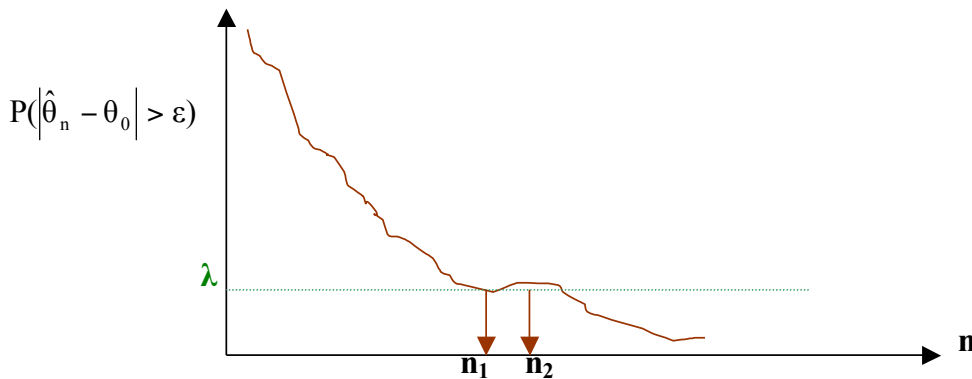
Δοθέντος οποιουδήποτε πραγματικού αριθμού $\varepsilon > 0$, η έκφραση:

$$P(|\hat{\theta}_n - \theta_0| > \varepsilon) \quad 6.1.2$$

ορίζεται ως η πιθανότητα ώστε: η απόλυτη απόκλιση της εκτίμησης της μεταβλητής από την πραγματική τιμή της, θ_0 , να είναι μεγαλύτερη από ε .

Ο εκτιμητής Θ , είναι συνεπής εάν για οποιοδήποτε δοθέντα πραγματικό αριθμό $\lambda > 0$, υπάρχει ένας αριθμός παρατηρήσεων, N , τέτοιος ώστε:

$$P(|\hat{\theta}_n - \theta_0| > \varepsilon) < \lambda \quad \text{για } n > N \quad 6.1.3$$



Σχήμα 6.1.2: Η ιδιότητα της συνέπειας για πεπερασμένο αριθμό μετρήσεων.

Επισημαίνεται ότι η ιδιότητα της συνέπειας είναι ασυμπτωτική ιδιότητα, υπό την έννοια ότι η αύξηση του αριθμού των παρατηρήσεων από n σε $n+\Delta n$ (όπου το n είναι πεπερασμένος αριθμός) δεν σημαίνει απαραίτητα ότι θα μειώνεται η απόλυτη απόκλιση της εκτίμησης της παραμέτρου θ από την πραγματική της τιμή.

Σο σχήμα 6.1.2, περιγράφεται η μεταβολή της πιθανότητας, $P(|\hat{\theta}_n - \theta_0| > \varepsilon)$, σε ένα υποθετικό παράδειγμα εκτίμησης κάποιας παραμέτρου θ , καθώς μεταβάλλεται ο αριθμός των παρατηρήσεων, n . Για $n = n_1$ παρατηρήσεις (και για κάποια επιλογή της τιμής του ε) η πιθανότητα $P(|\hat{\theta}_n - \theta_0| > \varepsilon)$ είναι μικρότερη του λ . Ωστόσο, για $n = n_2$ ($n_2 > n_1$) παρατηρήσεις η πιθανότητα, τοπικά, υπερβαίνει το λ . Παρ' όλα αυτά το παράδειγμα αναφέρεται σε συνεπή εκτίμηση.

Παράδειγμα 6.1.2

Η εκτίμηση της μέσης τιμής, μ , της τυχαίας μεταβλητής X , με συνάρτηση πυκνότητας πιθανότητας $f(x)$, ως ο μέσος όρος n τιμών της μεταβλητής, είναι συνεπής εκτίμηση.

Πράγματι, εάν η διασπορά, V , της τυχαίας μεταβλητής είναι πεπερασμένη και κατά την επιλογή τιμών δεν παραβιάζονται οι άλλες προϋποθέσεις του θεωρήματος του κεντρικού ορίου, ο μέσος όρος έχει συνάρτηση πυκνότητα πιθανότητας που εκφράζεται από την

$$\text{κανονική συνάρτηση: } \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(\mu - \bar{x})^2}{2 \cdot \sigma^2}} \quad \text{όπου } \sigma = \sqrt{\frac{V}{n}} \text{ και } \hat{x} = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Καθώς ο αριθμός των διαθέσιμων μετρήσεων, n , τείνει προς το άπειρο, το εύρος (σ) της κανονικής συνάρτησης τείνει στο μηδέν και η σχέση 6.1.3 ικανοποιείται για κάθε ε και λ .

6.1.2 Προκατάληψη:

Ορίζουμε ως **προκατάληψη (bias)**, b_N , του εκτιμητή της παραμέτρου θ , την απόκλιση της αναμενόμενης τιμής της εκτίμησης $\hat{\theta}$, η οποία βασίζεται σε N παρατηρήσεις, από την πραγματική τιμή θ_0 . Δηλαδή:

$$\begin{aligned} b_N(\hat{\theta}) &= E(\hat{\theta}) - \theta_0 \\ &= E(\hat{\theta} - \theta_0) \\ &= \int (\hat{\theta} - \theta_0) \cdot f(x_1, x_2, \dots, x_N; \theta) dx_1 dx_2 \dots dx_N \\ &= \int (\hat{\theta} - \theta_0) \cdot f(\underline{x}; \theta) d\underline{x} \end{aligned} \quad 6.1.4$$

όπου $\underline{x} = \{x_1, x_2, \dots, x_N\}$ και η συνάρτηση $f(\underline{x}; \theta)$, η οποία εκφράζει την κοινή συνάρτηση πυκνότητας πιθανότητας των N μετρήσεων, εξαρτάται από την παράμετρο θ .

Επομένως ένας εκτιμητής δεν είναι προκατειλημμένος εάν, για κάθε πλήθος μετρήσεων N , θα ισχύει :

$$b_N(\hat{\theta}) = 0 \quad \text{ή} \quad E(\hat{\theta}) = \theta_0 \quad 6.1.5$$

Παράδειγμα 6.1.1.

Ας υποθέσουμε ότι n ανεξάρτητοι παρατηρητές πραγματοποιούν το ίδιο πείραμα, μέτρησης της τιμής κάποιας φυσικής ποσότητας, x . Έστω ότι κάθε παρατηρητής συλλέγει N ανεξάρτητες μετρήσεις, με κοινή συνάρτηση πυκνότητας πιθανότητας:

$f(\underline{x}) = f(x_1, x_2, \dots, x_N) = \prod_{i=1}^N g(x_i)$, όπου $g(x)$ παριστά την συνάρτηση πυκνότητας πιθανότητας της κάθε μέτρησης.

Στην συνέχεια κάθε παρατηρητής εξάγει τον μέσο όρο, $\langle x \rangle$, των μετρήσεων που συνέλεξε. Τελικά, τα αποτελέσματα των παρατηρήσεων, συνοψίζονται στα ακόλουθα:

$$\begin{aligned} \underline{x}^1 &\equiv \{x_1^1, x_2^1, x_3^1, \dots, x_N^1\} \rightarrow \langle x \rangle^1 \\ \underline{x}^2 &\equiv \{x_1^2, x_2^2, x_3^2, \dots, x_N^2\} \rightarrow \langle x \rangle^2 \end{aligned} \quad 6.1.6$$

$\underline{x}^3 \equiv \{x_1^3, x_2^3, x_3^3, \dots, x_N^3\}$

$$\underline{x}^n \equiv \{x_1^n, x_2^n, x_3^n, \dots, x_N^n\} \rightarrow \langle x \rangle^n$$

όπου ο άνω δείκτης δηλώνει από ποιά παρατηρητή προήλθαν τα δεδομένα. Προφανώς, η 6.1.6 περιέχει n τιμές της στατιστικής συνάρτησης, $\langle x \rangle$.

Επειδή οι μετρήσεις, κάθε παρατηρητή, είναι ανεξάρτητες μεταξύ τους, θα ισχύει ότι:

$$E(< x >) = E(x) = \int_{\Omega} x \cdot g(x) dx = \mu \quad 6.1.7$$

Στη συνέχεια, ως χρησιμοποιήσουμε την στατιστική συνάρτηση $< x > = \frac{1}{N} \cdot \sum_{i=1}^N x_i$ ως εκτιμητή της μέσης τιμής, μ . Εύκολα μπορείτε να δείξετε ότι η προκατάληψη του $< x >$ είναι μηδέν (Σύγκρινε 6.1.5 με 6.1.7)

Στο ίδιο συμπέρασμα καταλήγει κανείς εάν χρησιμοποιήσει τον ισχυρό νόμο των μεγάλων αριθμών.

«Για ένα σύνολο N τυχαίων μεταβλητών, $y_1, y_2, \dots, y_N$, που παίρνουν τιμές σύμφωνα με διαφορετικές συναρτήσεις πυκνότητας πιθανότητας αλλά με την ίδια μέση τιμή, μ , και τετραγωνικές διασπορές σ_i^2 (όπου $i=1,2,3,\dots,N$), εάν η ποσότητα $\lim_{N \rightarrow \infty} \sum_{i=1}^N \left(\frac{\sigma_i}{i} \right)^2$ είναι

πεπερασμένη, τότε η έκφραση $< y > = \frac{1}{N} \cdot \sum_{i=1}^N y_i$ συγκλίνει κατά πιθανότητα στο μ .»

Στην περίπτωση αυτού του παραδείγματος

$$\sigma_i^2 = \sigma^2 = \int_{\Omega} (x - \mu)^2 \cdot g(x) dx$$

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N \left(\frac{\sigma_i}{i} \right)^2 = \lim_{N \rightarrow \infty} \left(\sigma^2 \cdot \sum_{i=1}^N \frac{1}{i^2} \right) < \sigma^2 \cdot \int_1^{\infty} \frac{1}{x^2} dx < \sigma^2 \cdot \frac{1}{3}$$

και επομένως η ποσότητα $\lim_{N \rightarrow \infty} \sum_{i=1}^N \left(\frac{\sigma_i}{i} \right)^2$ είναι πεπερασμένη.

Σύμφωνα λοιπόν με τον νόμο των μεγάλων αριθμών, η στατιστική συνάρτηση $< x > = \frac{1}{N} \cdot \sum_{i=1}^N x_i$ συγκλίνει κατά πιθανότητα στο μ , και κατ' επέκταση εκφράζει συνεπή εκτιμητή του μ .

Γενικά, ένας εκτιμητής της παραμέτρου θ , ο οποίος κατασκευάζεται από τα αποτελέσματα N παρατηρήσεων, όπως: $\hat{\theta}_N = \Theta(x_1, x_2, \dots, x_N)$, μπορεί να πάρει διαφορετικές τιμές για n διαφορετικά σύνολα (δείγματα) παρατηρήσεων $\{x_1, x_2, \dots, x_N\}$.

$$\hat{\theta}_N^i = \Theta\{x_1^i, x_2^i, x_3^i, \dots, x_N^i\} \quad i = 1, \dots, n \quad 6.1.8$$

Την συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $\hat{\theta}_N$, $q(\hat{\theta}_N; \theta)$, θα καλούμε, για ευνόητους λόγους, στα επόμενα δειγματική συνάρτηση.

Ερώτηση: Δείξετε ότι η δειγματική κατανομή εξαρτάται από την παράμετρο θ .

Σύμφωνα με τον ορισμό της συνέπειας, ένας συνεπής εκτιμητής χαρακτηρίζεται από δειγματική κατανομή με μέση τιμή, απειροστά κοντά στην πραγματική τιμή της παραμέτρου

θ, για μεγάλο αριθμό παρατηρήσεων. Όταν το πλήθος των παρατηρήσεων γίνει άπειρο, τότε η μέση τιμή της δειγματικής κατανομής και η πραγματική τιμή της παραμέτρου συμπίπτουν.

Επισημαίνεται ότι ένας εκτιμητής χωρίς προκατάληψη, για κάθε αριθμό παρατηρήσεων (N), χαρακτηρίζεται από δειγματική κατανομή που έχει μέση τιμή ίση με την πραγματική τιμή της υπό εκτίμηση παραμέτρου.

Η σχέση μεταξύ της μέσης τιμής της δειγματικής κατανομής εκτιμητή και της πραγματικής τιμής της παραμέτρου παρίσταται γραφικά στο Σχήμα 6.1.3 για τις περιπτώσεις συνεπούς και μη προκατειλημμένης εκτίμησης.

Παράδειγμα 6.1.3:

Εκτίμηση της τετραγωνικής διασποράς συνάρτησης πυκνότητας πιθανότητας.

Έστω N μετρήσεις ενός μεγέθους x, $\{x_1, x_2, x_3, \dots, x_N\}$, με συνάρτηση πυκνότητας πιθανότητας g(x). Ας υποθέσουμε ότι γνωρίζουμε την μέση τιμή (μ) και ας προσπαθήσουμε να βρούμε ένα «καλό» εκτιμητή της διασποράς της των μετρήσεων, V(x).

Ας δοκιμάσουμε την στατιστική συνάρτηση $\hat{V}(x) = \frac{1}{N} \sum (x_i - \mu)^2$

Η αναμενόμενη τιμή του εκτιμητή $\hat{V}(x)$ υπολογίζεται εύκολα ως:

$$E(\hat{V}(x)) = \frac{1}{N} \sum_{i=1}^N E[(x_i - \mu)^2] = \frac{1}{N} \sum_{i=1}^N V(x) = V(x) \quad 6.1.9$$

Συνεπώς ο εκτιμητής $\hat{V}(x)$ είναι μη προκατειλημμένος.

Ερώτηση: Αποδείξτε ότι ο εκτιμητής $\hat{V}(x)$ είναι επίσης συνεπής.

Υπόδειξη: Κάνετε χρήση του θεωρήματος του κεντρικού ορίου και δείξτε ότι η διασπορά της εκτίμησης μηδενίζεται όταν ο αριθμός των παρατηρήσεων τείνει στο άπειρο.

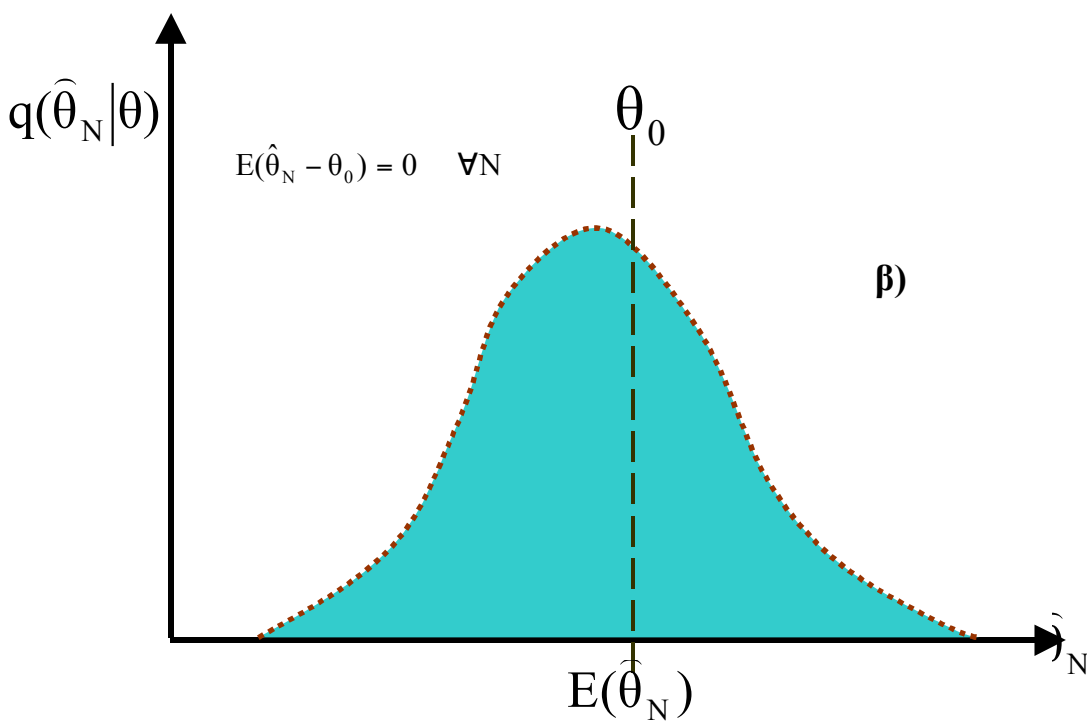
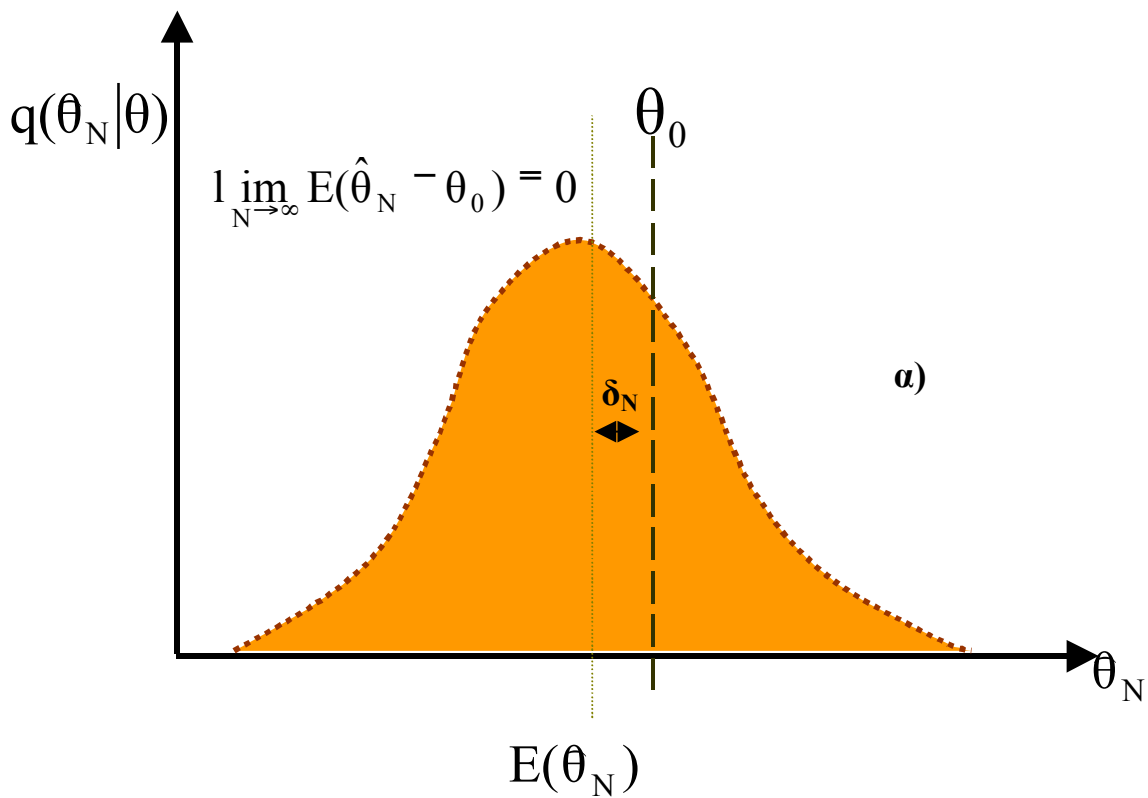
Τώρα ας υποθέσουμε ότι και η μέση τιμή είναι άγνωστη. Θα δοκιμάσουμε τον εκτιμητή:

$$\hat{V}(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2 = \frac{1}{N} \sum_{i=1}^N (x_i^2 - \langle x \rangle^2) \quad 6.1.10$$

$$\text{με } \langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i$$

και θα υπολογίσουμε την αναμενόμενη τιμή του εκτιμητή, $\hat{V}(x)$ ως:

$$\begin{aligned} E(\hat{V}(x)) &= \frac{1}{N} \sum_{i=1}^N [E(x_i^2) - E(\langle x \rangle^2)] = E(x_i^2) - E(\langle x \rangle^2) \\ &= E(x_i^2) - E(\langle x \rangle)^2 - [E(\langle x \rangle^2) - E(\langle x \rangle)^2] = V(x) - V(\langle x \rangle) \\ &= V(x) - \frac{V(x)}{N} = \frac{N-1}{N} \cdot V(x) \neq V(x) \end{aligned}$$



Σχήμα 6.1.3: Σχηματική παράσταση των ιδιοτήτων της συνέπειας (α) και της μη προκατάληψης (β).

Συνεπώς στην περίπτωση όπου η μέση τιμή δεν είναι γνωστή και εκτιμάται ως:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i$$

η στατιστική συνάρτηση:

$$\hat{V}(x) = \frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2$$

είναι προκατειλημμένος εκτιμητής της διασποράς έστω και εάν είναι συνεπής εκτιμητής.

Ερώτηση: Εξετάσετε τις ιδιότητες της στατιστικής συνάρτησης

$$\hat{V}(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \langle x \rangle)^2 \quad 6.1.11$$

ως εκτιμητή της τετραγωνικής διασποράς, όταν η μέση τιμή της συνάρτησης πυκνότητας πιθανότητας δεν είναι γνωστή.

Υπόδειξη: Επαναλάβετε τον έλεγχο που εφαρμόσαμε στο παράδειγμα 6.1.3 και δείξτε ότι η 6.1.10 είναι μη προκατειλημμένος εκτιμητής.

Παράδειγμα 6.1.4:

Μία σειρά δοκιμών ενός νέου τύπου αερόσακου καταλήγει σε N_S επιτυχίες και σε N_F αποτυχίες.

Η στατιστική συνάρτηση: $\hat{p} = \frac{N_S}{N_S + N_F}$

είναι ένας συνεπής εκτιμητής της πιθανότητας επιτυχίας p .

Η συνέπεια αναφέρεται στην περίπτωση όπου ο αριθμός των συνολικών δοκιμών $n = N_S + N_F$ τείνει στο άπειρο. Βεβαίως υπό αυτές τις συνθήκες ο εκτιμητής $\hat{p}$

$$\hat{p} = \lim_{n \rightarrow \infty} \frac{N_S}{n}$$

ισούται, εξ ορισμού, με την πιθανότητα επιτυχίας και ο εκτιμητής $\hat{p}$ είναι συνεπής.

Στην περίπτωση που θα χρησιμοποιήσουμε τον ίδιο εκτιμητή για πεπερασμένο πλήθος δοκιμών θα πρέπει να εξασφαλίσουμε ότι η εκτίμηση δεν είναι προκαταλειμμένη.

Η αναμενόμενη τιμή του εκτιμητή $\hat{p}$ είναι:

$$E(\hat{p}) = \sum_{N_s=0}^n \frac{N_s}{n} p^{N_s} (1-p)^{n-N_s} \frac{n!}{N_s!(n-N_s)!}$$

$$= p \sum_{N_s=1}^n p^{N_s-1} (1-p)^{n-N_s} \frac{(n-1)!}{(N_s-1)!(n-N_s)!}$$

αντικαθιστώντας $n' = n - 1$ και $N'_s = N_s - 1$

$$E(\hat{p}) = p \sum_{N'_s=0}^{n'} p^{N'_s} (1-p)^{n'-N'_s} \frac{n'!}{N'_s!(n'-N'_s)!} = p$$

συνθήκη κανονικοποίησης

Συνεπώς ο εκτιμητής $\hat{p}$ είναι μη προκατειλημμένος εκτιμητής της πιθανότητας επιτυχίας.

6.1.3 Αποδοτικότητα:

Στις προηγούμενες υποενότητες επισημάναμε ότι το αποτέλεσμα της εκτίμησης παραμέτρου συμπεριφέρεται ως τυχαία μεταβλητή και επιπλέον δείξαμε ότι η μέση τιμή της δειγματικής κατανομής, συνεπών και μη προκατειλημμένων εκτιμήσεων, συνδέεται με την πραγματική τιμή της παραμέτρου.

Είναι προφανές ότι «εύρος» της δειγματικής κατανομής εκφράζει την ακρίβεια της εκτίμησης. Επί παραδείγματι, η εκτίμηση της τιμής κάποιας φυσικής σταθεράς, από τα δεδομένα κατάλληλου πειράματος, καταλήγει σε διαφορετικό αποτέλεσμα κάθε φορά που θα επαναληφθεί το πείραμα. Η ακρίβεια της εκτίμησης είναι βέλτιστη όταν τα αποτελέσματα αυτά κατανέμονται, γύρω από την πραγματική τιμή της φυσικής σταθεράς, με αμελητέο εύρος. Εν άλλους λόγους, η τετραγωνική διασπορά του αποτελέσματος της εκτίμησης αποτελεί μέτρο της ακρίβειας.

Η ακρίβεια της εκτίμησης εξαρτάται προφανώς από την στατιστική συνάρτηση που επιλέγεται ως εκτιμητής αλλά εξαρτάται επίσης και από την συνάρτηση πυκνότητας πιθανότητας των πειραματικών δεδομένων.

Παράδειγμα 6.1.5

Έστω ότι x είναι το αποτέλεσμα της μέτρησης φυσικού μεγέθους (π.χ. της περιόδου ενός εκκρεμούς). Θα υποθέσουμε επίσης ότι κατά την μετρητική διαδικασία πληρούνται όλες οι προϋποθέσεις του θεωρήματος του κεντρικού ορίου. Η τυχαία μεταβλητή, x , χαρακτηρίζεται από κανονική συνάρτηση πυκνότητας πιθανότητας, με μέση τιμή, μ , ίση με την πραγματική τιμή του φυσικού μεγέθους και τετραγωνική απόκλιση, $V = \sigma^2$, που εκφράζει την διακριτική ικανότητα του οργάνου.

Έστω ότι συλλέγονται N μετρήσεις και η πραγματική τιμή εκτιμάται από τον μέσο όρο των μετρήσεων, $\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i$. Στο παράδειγμα 6.1.1, δείξαμε ότι πρόκειται περί συνεπούς εκτιμητή ο οποίος δεν πάσχει από προκατάληψη.

Σύμφωνα με τα προηγούμενα, η ακρίβεια της εκτίμησης εκφράζεται από την διασπορά, $V(<x>)$, η οποία εύκολα υπολογίζεται ως:

$$V(<x>) = V\left(\frac{1}{N} \sum_{i=1}^N x_i\right) = \frac{1}{N^2} \cdot \sum_{i=1}^N V(x) = \frac{1}{N^2} \cdot N \cdot V(x) = \frac{\sigma^2}{N} \quad 6.1.12$$

Είναι προφανές ότι η ακρίβεια της εκτίμησης εξαρτάται από την διασπορά της συνάρτησης πυκνότητας πιθανότητας που χαρακτηρίζει τις μετρήσεις.

Παράδειγμα 6.1.6

Ας υποθέσουμε ότι, στο Παράδειγμα 6.1.5, η διακριτική ικανότητα της μετρητικής συσκευής δεν είναι γνωστή και για την εκτίμησή της χρησιμοποιείται ο συνεπής και μη-προκατειλημμένος εκτιμητής 6.1.11

$$\hat{V}(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - <x>)^2$$

Η διασπορά της εκτίμησης ευρίσκεται ως:

$$V[\hat{V}(x)] = V\left[\frac{\sigma^2}{N-1} \cdot \sum_{i=1}^N \frac{(x_i - <x>)^2}{\sigma^2}\right] = \left(\frac{\sigma^2}{N-1}\right)^2 V\left[\sum_{i=1}^N \frac{(x_i - <x>)^2}{\sigma^2}\right] \quad 6.1.13$$

Η τυχαία μεταβλητή $\chi^2 = \left[\sum_{i=1}^N \frac{(x_i - <x>)^2}{\sigma^2}\right]$, όπου κάθε ένα από τα x_i ακολουθεί

κανονική συνάρτηση πυκνότητας πιθανότητας με διασπορά σ^2 , χαρακτηρίζεται από την λεγόμενη $\chi^2(N-1)$ συνάρτηση πυκνότητας πιθανότητας (βλέπε υποενότητα), για $N-1$ βαθμούς ελευθερίας. Η διασπορά $V(\chi^2)$ ισούται με: $V[\chi^2(N-1)] = 2(N-1)$, ώστε η σχέση 6.1.13 καταλήγει σε:

$$V[\hat{V}(x)] = \left(\frac{2 \cdot \sigma^4}{N-1}\right) \quad 6.1.14$$

όπου είναι εμφανής η εξάρτηση της ακρίβειας της εκτίμησης από την διασπορά των μετρήσεων.

Ένας εκτιμητής είναι **αποδοτικός (efficient)** εάν προσφέρει την μεγαλύτερη δυνατή ακρίβεια εκτίμησης της πραγματικής τιμής της παραμέτρου. Επιπλέον, η αποδοτικότητα ενός εκτιμητή θα πρέπει να εξαρτάται από το «μέγεθος» της πληροφορίας που μεταφέρει. Κατά συνέπεια η τετραγωνική διασπορά ενός εκτιμητή και η πληροφορία που μεταφέρει θα πρέπει να συνδέονται. Η σχέση της αποδοτικότητας και της πληροφορίας που μεταφέρεται από ένα εκτιμητή θα μας απασχολήσει στην επόμενη υποενότητα.

6.2 Μέγιστη Αποδοτικότητα - Η Ανισότητα των Cramer-Rao

Έστω $\underline{x} = \{x^1, x^2, \dots, x^p\}$ είναι ένα σύνολο p μεταβλητών που χαρακτηρίζουν ένα γεγονός (π.χ. οι τρεις συνιστώσες της ταχύτητας ενός μορίου ιδανικού αερίου). Οι τιμές των μεταβλητών αυτών κατανέμονται σύμφωνα με την συνάρτηση πυκνότητας πιθανότητας $f(\underline{x}; \theta)$, η οποία με την σειρά της εξαρτάται από την παράμετρο θ (π.χ. την θερμοκρασία του αερίου για κάποια σταθερή και γνωστή πίεση).

Δεδομένων N μετρήσεων των χαρακτηριστικών αυτών φυσικών μεγεθών, τυχαίων μεταβλητών: $\underline{x}_i, i=1, 2, \dots, N$ (π.χ. μετρώντας τις τρεις συνιστώσες της ταχύτητας N μορίων ενός αερίου σε θερμοδυναμική ισορροπία), η τιμή της παραμέτρου θ εκτιμάται από κάποιον εκτιμητή, έστω τον εκτιμητή $\hat{\theta} = \Theta(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N)$.

Η παράμετρος θ έχει καλά καθορισμένη τιμή, θ_0 (π.χ. το αέριο έχει καθορισμένη θερμοκρασία) αλλά η εκτίμηση, $\hat{\theta} = \Theta(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N)$, θα καταλήγει σε διαφορετικό αποτέλεσμα κάθε φορά που θα επαναλαμβάνεται η πειραματική διαδικασία και θα συλλέγονται N μετρήσεις. Η εκτίμηση $\hat{\theta}$ θα συμπεριφέρεται ως τυχαία μεταβλητή με δειγματική κατανομή $q(\hat{\theta}; \theta_0)$. Επισημαίνεται ότι η δειγματική κατανομή εξαρτάται από την πραγματική τιμή της παραμέτρου, θ_0 .

Σύμφωνα με τον ορισμό της προκατάληψης της εκτίμησης, η προκατάληψη, b , του εκτιμητή $\hat{\theta} = \Theta(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N)$ εκφράζεται ως:

$$b = E(\hat{\theta}) - \theta_0$$

$$= \left[\int_{\Omega_{\theta}} \int_{\Omega_{\underline{x}}} \Theta(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N) \cdot f(\underline{x}_1; \theta_0) \cdot f(\underline{x}_2; \theta_0) \cdot \dots \cdot f(\underline{x}_N; \theta_0) \cdot d\underline{x}_1 d\underline{x}_2 \dots d\underline{x}_N - \theta_0 \right] \quad 6.2.1$$

πιθανοφάνεια $L_{\underline{x}}$

$$= \left[\int \hat{\theta} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} - \theta_0 \right]$$

Προφανώς η προκατάληψη εξαρτάται από την πραγματική τιμή της παραμέτρου. Στην ανάλυση που ακολουθεί θα θεωρήσουμε την πραγματική τιμή ως παράμετρο της δειγματικής συνάρτησης.

Στην περίπτωση όπου:

1. το πεδίο ορισμού των τυχαίων μεταβλητών, $\underline{x}$, δεν εξαρτάται από την τιμή της παραμέτρου, $\Omega_{\theta} \equiv \Omega$, και

2. η πιθανοφάνεια είναι ομαλή συνάρτηση ώστε οι τελεστές $\partial^2/\partial\theta_0^2$ και $\int d\underline{x}$ να μετατίθενται

η διασπορά της δειγματικής κατανομής

$$V(\hat{\theta}) = \int (\hat{\theta} - E(\hat{\theta}))^2 \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \quad 6.2.2$$

σχετίζεται με την πληροφορία μέσω της ανισότητας των Cramer-Rao

$$V(\hat{\theta}) \geq \frac{\left(1 + \frac{db}{d\theta_0}\right)^2}{I_{\hat{\theta}}} \geq \frac{\left(1 + \frac{db}{d\theta_0}\right)^2}{I_x} \quad 6.2.3$$

όπου I_x και $I_{\hat{\theta}}$ είναι οι ποσότητες πληροφορίας, αναφορικά με την πραγματική τιμή της παραμέτρου - θ_0 , που εμπεριέχονται στο σύνολο των N παρατηρήσεων $\{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N\}$ και στην στατιστική συνάρτηση $\hat{\theta} = \Theta(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N)$, αντίστοιχα.

Επειδή, προφανώς, $I_x \geq I_{\theta}$ η απόδειξη της δεύτερης ανισότητας, $\frac{\left(1 + \frac{db}{d\theta_0}\right)^2}{I_{\hat{\theta}}} \geq \frac{\left(1 + \frac{db}{d\theta_0}\right)^2}{I_x}$,

στην έκφραση 6.2.3 είναι τετριμμένη. Στα επόμενα θα αποδειχθεί η πρώτη ανισότητα:

$$V(\hat{\theta}) \geq \frac{\left(1 + \frac{db}{d\theta_0}\right)^2}{I_{\hat{\theta}}}.$$

Η πληροφορία που μεταφέρει η στατιστική συνάρτηση $\hat{\theta} = \Theta(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N)$ εκφράζεται ως:

$$\begin{aligned} I_{\hat{\theta}} &= \int \left(\frac{\partial \ln L_{\hat{\theta}}}{\partial \theta_0} \right)^2 \cdot L_{\hat{\theta}} d\hat{\theta} \\ &= \int \left(\frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \right)^2 \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \end{aligned}$$

Όπου η πιθανοφάνεια που αντιστοιχεί σε μία τυχαία μεταβλητή, $\hat{\theta}$, δεν είναι τίποτα άλλο παρά η (δειγματική) συνάρτηση πυκνότητας πιθανότητας, $q(\hat{\theta}; \theta_0)$.

Εξ' ορισμού ισχύει ότι: $E(\hat{\theta}) = \int \hat{\theta} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta}$

Συνεπώς:

$$\frac{\partial E(\hat{\theta})}{\partial \theta_0} = \int \hat{\theta} \cdot \frac{\partial q(\hat{\theta}; \theta_0)}{\partial \theta_0} d\hat{\theta} = \int \hat{\theta} \cdot \frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \quad 6.2.4$$

Επειδή όμως, σύμφωνα με τον ορισμό 6.2.1, ισχύει ότι: $E(\hat{\theta}) = b(\theta_0) - \theta_0$, συνεπάγεται ότι:

$$\frac{\partial E(\hat{\theta})}{\partial \theta_0} = \frac{db(\theta_0)}{d\theta_0} + 1 \quad 6.2.5$$

Από τις σχέσεις 6.2.4 και 6.2.5 προκύπτει ότι:

$$\begin{aligned}
1 + \frac{db(\theta_0)}{d\theta_0} &= \int \hat{\theta} \cdot \frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \\
&= \int [\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta})] \cdot A \cdot d\hat{\theta} \\
&= \int [\hat{\theta} - E(\hat{\theta})] \cdot A \cdot d\hat{\theta} + \int E(\hat{\theta}) \cdot A d\hat{\theta}
\end{aligned} \tag{6.2.6}$$

$$\text{όπου, } A \equiv \frac{\partial \ln q(\hat{\theta}; \theta)}{\partial \theta} q(\hat{\theta}; \theta)$$

Επισημαίνεται ότι η αναμενόμενη τιμή $E(\hat{\theta}) \left(\equiv \int_{-\infty}^{+\infty} \hat{\theta} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \right)$ είναι ανεξάρτητη του $\hat{\theta}$,
ώστε ο δεύτερος όρος του δεξιού σκέλους της 6.2.6 επιδέχεται την ακόλουθη απλοποίηση:

$$\begin{aligned}
\int E(\hat{\theta}) \cdot A d\hat{\theta} &= E(\hat{\theta}) \cdot \int \frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \\
&= E(\hat{\theta}) \cdot \int \frac{\partial q(\hat{\theta}; \theta_0)}{\partial \theta_0} d\hat{\theta} \\
&= E(\hat{\theta}) \cdot \frac{\partial}{\partial \theta_0} \int q(\hat{\theta}; \theta_0) d\hat{\theta} \\
&= 0
\end{aligned} \tag{6.2.7}$$

Αντικαθιστώντας της σχέση 6.2.7 στην σχέση 6.2.6 καταλήγουμε ότι:

$$1 + \frac{db(\theta_0)}{d\theta_0} = \int [\hat{\theta} - E(\hat{\theta})] \cdot \frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \tag{6.2.8}$$

Στα επόμενα θα χρησιμοποιηθεί η ανισότητα του Schwarz, η οποία, για τις συναρτήσεις $u(y)$ και $v(y)$, εκφράζεται ως:

$$\int u^2 dy \cdot \int v^2 dy \geq \left[\int u \cdot v dy \right]^2 \tag{6.2.9}$$

Αντικαθιστώντας, στην 6.2.9, με:

$$u(\hat{\theta}) \equiv (\hat{\theta} - E(\hat{\theta})) \sqrt{q(\hat{\theta}; \theta_0)} \quad , \quad v(\hat{\theta}) \equiv \frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \cdot \sqrt{q(\hat{\theta}; \theta_0)} \tag{6.2.10}$$

η ανισότητα εκφράζεται ως:

$$\left[\int (\hat{\theta} - E(\hat{\theta})) \frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \right]^2 \leq \left[\int (\hat{\theta} - E(\hat{\theta}))^2 q(\hat{\theta}; \theta_0) d\hat{\theta} \right] \cdot \left[\int \left(\frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \right)^2 \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \right]$$

Αντικαθιστώντας το αριστερό σκέλος της ανισότητας με την σχέση 6.2.8, καταλήγουμε ότι:

$$\left(1 + \frac{db(\theta_0)}{d\theta_0}\right)^2 \leq \left[\int (\hat{\theta} - E(\hat{\theta}))^2 \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \right] \cdot \left[\int \left(\frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \right)^2 \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \right] \quad 6.2.11$$

Παρατηρήσετε ότι το ολοκλήρωμα, $\left[\int (\hat{\theta} - E(\hat{\theta}))^2 \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \right]$, εκφράζει την διασπορά της εκτίμησης $\hat{\theta}$, $V(\hat{\theta})$, ενώ το ολοκλήρωμα, $\left[\int \left(\frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} \right)^2 \cdot q(\hat{\theta}; \theta_0) d\hat{\theta} \right]$, εκφράζει την πληροφορία που εμπεριέχει η εκτίμηση $\hat{\theta}$, $I_{\hat{\theta}}$, αναφορικά με την πραγματική τιμή της παραμέτρου.

Συνεπώς, η 6.2.11 μπορεί να αναδιατυπωθεί ως:

$$V(\hat{\theta}) \cdot I_{\hat{\theta}} \geq \left[1 + \frac{db(\theta_0)}{d\theta_0} \right]^2 \quad \text{ή} \quad V(\hat{\theta}) \geq \frac{\left[1 + \frac{db(\theta_0)}{d\theta_0} \right]^2}{I_{\hat{\theta}}} \quad 6.2.12$$

Στην περίπτωση όπου η εκτίμηση δεν πάσχει από προκατάληψη, $b(\theta_0)=0$, η ανισότητα των Cramer-Rao εκφράζεται ως:

$$V(\hat{\theta}) \geq \frac{1}{I_{\hat{\theta}}} \quad 6.2.13$$

Η περίπτωση της ισότητας στην ανισότητα του Schwarz, 6.2.9, ισχύει όταν $v(y) = \lambda \cdot u(y)$ και ο παράγων λ είναι ανεξάρτητος του y . Στην περίπτωση των συναρτήσεων 6.2.10, ο παράγων λ θα πρέπει να είναι ανεξάρτητος της εκτίμησης, $\hat{\theta}$, αλλά μπορεί να εξαρτάται από την πραγματική τιμή της παραμέτρου, θ_0 . Εν άλλους λόγους, η συνθήκη της ισότητας εκφράζεται ως:

$$\frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} = R(\theta_0) \cdot (\hat{\theta} - E(\hat{\theta})) \quad 6.2.14$$

όπου $R(\theta_0)$ εκφράζει το πολλαπλασιαστικό παράγοντα. Η σχέση 6.2.14 γράφεται συναρτήσει της προκατάληψης ως:

$$\frac{\partial \ln q(\hat{\theta}; \theta_0)}{\partial \theta_0} = R(\theta_0) \cdot (\hat{\theta} - (b(\theta_0) + \theta_0)) \quad 6.2.15$$

Ερώτηση: Δείξτε ότι η συνθήκη ισότητας 6.2.15 ισχύει εάν η δειγματική συνάρτηση είναι της εκθετικής μορφής:

$$q(\hat{\theta}; \theta_0) = \exp(a(\theta_0) \cdot \hat{\theta} + \beta(\hat{\theta}) + c(\theta_0)) \quad 6.2.16$$

Υπόδειξη:

$$\int R(\theta_0) d\theta_0 = a(\theta_0) + \beta_1(\hat{\theta})$$

$$\int R(\theta_0) \cdot (b(\theta_0) + \theta_0) d\theta_0 = c(\theta_0) + \beta_2(\hat{\theta})$$

$$\beta(\hat{\theta}) = \beta_1(\hat{\theta}) + \beta_2(\hat{\theta})$$

Ανακεφαλαιώνοντας, αποδείξαμε ότι:

$$V(\hat{\theta}) \geq \frac{\left[1 + \frac{db(\theta_0)}{d\theta_0}\right]^2}{I_{\hat{\theta}}} \geq \frac{\left[1 + \frac{db(\theta_0)}{d\theta_0}\right]^2}{I_x} \quad 6.2.17$$

όπου η περίπτωση της ισότητας ισχύει όταν η δειγματική συνάρτηση είναι της εκθετικής μορφής 6.2.16.

Ειδικά, στην περίπτωση όπου $I_x = I_{\theta}$, η εκτίμηση έχει διασπορά ίση με τον ελάχιστο δυνατό όριο. Η σχέση αυτή, $I_x = I_{\theta}$, αποτελεί την αναγκαία και ικανή συνθήκη ώστε ο εκτιμητής $\hat{\theta} = \Theta(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N)$ να είναι επαρκής στατιστική συνάρτηση όσον αφορά στην παράμετρο θ .

Όταν όμως υπάρχει επαρκής στατιστική συνάρτηση για την παράμετρο θ , η συνάρτηση πυκνότητας πιθανότητας των τυχαίων μεταβλητών, $f(\underline{x}; \theta)$, (π.χ. των προβολών της ταχύτητας των μορίων) είναι και αυτή εκθετικής μορφής, όπως επιβάλλει το θεώρημα του Darmois.

Η ανισότητα των Cramer-Rao γενικεύεται εύκολα για την περίπτωση όπου η συνάρτηση πυκνότητας πιθανότητας των δεδομένων εξαρτάται από πολλές παραμέτρους.

Έστω, η συνάρτηση πυκνότητας πιθανότητας, $f(\underline{x}; \underline{\theta})$, όπου $\underline{\theta} = \{\theta_1, \theta_2, K, \theta_r\}$ παριστά τις r παραμέτρους από τις οποίες εξαρτάται η κατανομή των πειραματικών δεδομένων. Έστω, οι εκτιμητές των παραμέτρων, $\hat{\underline{\theta}} = \{\hat{\theta}_1, \hat{\theta}_2, K, \hat{\theta}_r\}$, όπου $\hat{\theta}_i = \Theta_i(\underline{x}_1, \underline{x}_2, K, \underline{x}_N)$, $i = 1, 2, K, r$ είναι στατιστικές συναρτήσεις των N μετρήσεων.

Η πληροφορία, σύμφωνα με την 5.2.5, εκφράζεται από πίνακα με στοιχεία:

$$\left[I_{\underline{\theta}}^x(\underline{\theta}) \right]_{ij} = E \left[\frac{\partial \ln L(\underline{x}_1, \underline{x}_2, K, \underline{x}_N; \underline{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln L(\underline{x}_1, \underline{x}_2, K, \underline{x}_N; \underline{\theta})}{\partial \theta_j} \right] \quad 6.2.18$$

ενώ η πληροφορία που μεταφέρει το σύνολο των εκτιμητών εκφράζεται από τον πίνακα:

$$\left[I_{\underline{\theta}}^{\hat{\underline{\theta}}}(\underline{\theta}) \right]_{ij} = E \left[\frac{\partial \ln Q(\hat{\underline{\theta}}; \underline{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln Q(\hat{\underline{\theta}}; \underline{\theta})}{\partial \theta_j} \right] \quad 6.2.19$$

όπου $L(\underline{x}_1, \underline{x}_2, K, \underline{x}_N; \underline{\theta}) = \prod_{i=1}^N f(\underline{x}_i; \underline{\theta})$ είναι η πιθανοφάνεια, $Q(\hat{\underline{\theta}}; \underline{\theta})$ είναι η κοινή δειγματική

κατανομή των εκτιμήσεων $\hat{\underline{\theta}} = \{\hat{\theta}_1, \hat{\theta}_2, K, \hat{\theta}_r\}$ και οι παράμετροι $\underline{\theta}$ αναφέρονται στις φυσικές (πραγματικές) τιμές των σταθερών από τις οποίες εξαρτάται η κατανομή των δεδομένων.

Η ανισότητα των Cramer-Rao, 6.2.17, γενικεύεται, για την περίπτωση εκτιμήσεων χωρίς προκατάληψη, ως:

$$V(\hat{\theta}_i) \geq \frac{1}{\left[I_{\underline{\theta}}^x \right]_{ii}} \geq \frac{1}{\left[I_{\underline{\theta}}^{\hat{\underline{\theta}}} \right]_{ii}} \quad 6.2.20$$

Γενικότερα, για κάθε θετική τιμή c , το ελλειψοειδές

$$(\hat{\underline{\theta}} - \underline{\theta})^T \cdot \mathbf{I}_{\hat{\underline{\theta}}} \cdot (\hat{\underline{\theta}} - \underline{\theta}) = c$$

κείται στο εσωτερικό του ελλειψοειδούς

$$(\hat{\underline{\theta}} - \underline{\theta})^T \cdot \mathbf{V}_{\hat{\underline{\theta}}}(\hat{\underline{\theta}}) \cdot (\hat{\underline{\theta}} - \underline{\theta}) = c,$$

όπου $\left[\mathbf{V}_{\hat{\underline{\theta}}}(\hat{\underline{\theta}}) \right]_{ij} = E(\hat{\theta}_i \cdot \hat{\theta}_j) - E(\hat{\theta}_i) \cdot E(\hat{\theta}_j)$ είναι ο πίνακας συνέλιξης των εκτιμήσεων.

Τα δύο ελλειψοειδή ταυτίζονται και ισχύει η ισότητα στην σχέση 6.2.20, όταν οι εκτιμητές, $\hat{\theta}_i = \Theta_i(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N)$, $i = 1, 2, \dots, K$, είναι από κοινού επαρκείς στατιστικές συναρτήσεις των παραμέτρων $\underline{\theta} = \{\theta_1, \theta_2, \dots, \theta_K, \theta_r\}$

6.3 Μέθοδοι ορισμού συνεπών εκτιμητών – Μέθοδος των ορμών .

Στις προηγούμενες υποενότητες αυτού του κεφαλαίου συζητήσαμε τις ιδιότητες των εκτιμητών και τις συνθήκες που διασφαλίζουν την επάρκεια των εκτιμητών. Σ' αυτή την υποενότητα θα συζητήσουμε μεθόδους «κατασκευής» συνεπών εκτιμητών. Για απλότητα, θα θεωρήσουμε ότι κάθε γεγονός των πειραματικών δεδομένων χαρακτηρίζεται από τη μέτρηση μίας μόνο φυσικής ποσότητας (π.χ. μας απασχολεί μόνο η συνολική ορμή του μορίου ενός ιδανικού αερίου)

Όπως και στις προηγούμενες υποενότητες, θεωρείστε ότι διαθέτουμε N γεγονότα (μετρήσεις) $x_1, x_2, \dots, x_N$. Έστω ότι $a(x)$ είναι μία συνάρτηση του αποτελέσματος της μέτρησης (τυχαία μεταβλητή), τέτοια ώστε η τετραγωνική της διασπορά, $V(a(x))$, να είναι πεπερασμένη. Σύμφωνα με τον νόμο των μεγάλων αριθμών, ο μέσος όρος των τιμών αυτής της συνάρτησης, συγκλίνει προς την μέση τιμή της τυχαίας μεταβλητής, $E[a(x)]$, καθώς το πλήθος των διαθέσιμων παρατηρήσεων τείνει στο άπειρο:

$$N^{-1} \cdot \sum_{i=1}^N a(x_i) \xrightarrow{N \rightarrow \infty} E[a(x)] = \int a(x) \cdot f(x; \theta) dx \quad 6.3.1$$

όπου $f(x; \theta)$ παριστά την συνάρτηση πυκνότητας πιθανότητας που χαρακτηρίζει κάθε μία των πειραματικών μετρήσεων.

Το δεξιό μέλος της σχέσης 6.3.1 είναι συνάρτηση μόνο της παραμέτρου θ . Επιπλέον στην περίπτωση όπου η αναμενόμενη τιμή, $E[a(x)]$, μπορεί να εκφρασθεί ως αναλυτική συνάρτηση του θ , $\int a(x) \cdot f(x; \theta) dx = h(\theta)$, η πραγματική τιμή της παραμέτρου εκτιμάται ως:

$$\hat{\theta} = h^{-1} \left(\lim_{N \rightarrow \infty} \sum_{i=1}^N a(x_i) \right); \quad h^{-1} \left(\sum_{i=1}^N a(x_i) \right) \quad 6.3.2$$

για N μεγάλο

όπου το όρισμα, $\sum_{i=1}^N a(x_i)$, της αντιστρόφου της συνάρτησης $h(\theta)$ υπολογίζεται

χρησιμοποιώντας το πειραματικά δεδομένα, $\underline{x} = \{x_1, x_2, \dots, x_N\}$.

Γενικά, η αντιστροφή της συνάρτησης $h(\theta)$, σχέση 6.3.2, μπορεί να καταλήγει σε περισσότερες από μία εκτιμήσεις της παραμέτρου θ . Στην περίπτωση αυτή μόνο μία από αυτές θα αντιστοιχεί σε συνεπή εκτίμηση. Επισημαίνεται επίσης ότι η εκτίμηση 6.3.2 είναι συνεπής, λόγω της 6.3.1, αλλά δεν οδηγεί πάντα σε απροκατάληπτα αποτελέσματα, όπως θα δειχθεί στα επόμενα.

Η απλούστερη συνάρτηση $a(x)$ αντιστοιχεί στην επιλογή: $a(x) = x$, με αναμενόμενη τιμή που είναι ίση με την μέση τιμή της μεταβλητής x . Ο εκτιμητής της παραμέτρου θ , σ' αυτή την περίπτωση, ορίζεται ως

$$\mu_x = E(x) = \int x \cdot f(x; \theta) dx = h(\theta) \quad 6.3.3$$

$$\hat{\theta} = h^{-1} \left[\frac{1}{N} \sum_{i=1}^N x_i \right] = h^{-1} [\langle x \rangle]$$

Στην γενικότερη περίπτωση όπου η συνάρτηση πυκνότητας πιθανότητας, $f(x;\underline{\theta})$, εξαρτάται από περισσότερες της μίας μεταβλητές, $\underline{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$, χρειάζονται k συναρτήσεις ($a_i(x)$, $i=1,2,\dots,k$) για να εκτιμηθούν οι τιμές όλων των παραμέτρων.

Θεωρήστε την οικογένεια των συναρτήσεων, $a_j(x) = x^j$, όπου $j=1,2,3,\dots$.

Οι αναμενόμενες τιμές αυτών των συναρτήσεων, $E(x^j) = \int x^j \cdot f(x;\underline{\theta}) dx = M^j(\underline{\theta})$, είναι συναρτήσεις των παραμέτρων $\theta_1, \theta_2, \dots, \theta_k$ και θα θεωρήσουμε ότι είναι γνωστές αναλυτικά.

Παράλληλα, εφαρμόζοντας τον νόμο των μεγάλων αριθμών, $\frac{1}{N} \cdot \sum_{i=1}^N x^j \xrightarrow{N \rightarrow \infty} \int x^j \cdot f(x;\underline{\theta}) dx$,

οι τιμές των ορμών που αντιστοιχούν στις πραγματικές τιμές των παραμέτρων, μπορούν να εκτιμηθούν με συνέπεια από τα πειραματικά δεδομένα, ως: $\hat{M}^j = \frac{1}{N} \cdot \sum_{i=1}^N x^j$. Η

εκτίμηση των πραγματικών τιμών των παραμέτρων επιτυγχάνεται με την λύση του συστήματος των k εξισώσεων:

$$\hat{M}^j = M^j(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k), \quad j = r_1, r_2, \dots, r_k \quad 6.3.4$$

Η επιλογή τιμών για τον δείκτη j (π.χ. $j=1,2,3,\dots,k$) γίνεται αυθαίρετα, έχουμε πλήρη ελευθερία να επιλέξουμε τις k ροπές. Η ελευθερία αυτή, αντανακλά μία απώλεια της ακρίβειας με την οποία εκτιμούνται οι παράμετροι.

Παράδειγμα 6.3.1

Ας αναζητήσουμε εκτιμητή για την παράμετρο μ της Poissonian συνάρτησης πιθανότητας, $P(n) = \frac{\mu^n \cdot e^{-\mu}}{n!}$, με την μέθοδο των ροπών.

Έστω ένα σύνολο τιμών, $n_1, n_2, \dots, n_N$, της τυχαίας μεταβλητής n , η οποία ακολουθεί την Poissonian κατανομή.

Οι (συνεπείς) εκτιμήσεις των ροπών πρώτης και δεύτερης ορμής είναι:

$$\hat{M}^1 = \frac{1}{N} \cdot \sum_{i=1}^N n_i \quad 6.3.5$$

$$\hat{M}^2 = \frac{1}{N} \cdot \sum_{i=1}^N n_i^2$$

Επειδή η Poissonian συνάρτηση, έχει μέση τιμή και τετραγωνική διασπορά ίσες με την παράμετρο μ , ισχύει ότι:

$$\hat{\mu}_1 = \hat{M}^1 \xrightarrow{N \rightarrow \infty} \mu = \frac{1}{N} \cdot \sum_{n=0}^{\infty} n \cdot P(n) \quad 6.3.6$$

$$\hat{\mu}_2 = \hat{M}^2 - \left(\hat{M}^1 \right)^2 \xrightarrow{N \rightarrow \infty} V(n) = \mu = \frac{1}{N} \cdot \sum_{n=0}^{\infty} n^2 \cdot P(n) - \left(\frac{1}{N} \cdot \sum_{n=0}^{\infty} n \cdot P(n) \right)^2$$

Συνεπώς, παράμετρος μ μπορεί να εκτιμηθεί από δύο διαφορετικούς εκτιμητές που κατασκευάζονται με την μέθοδο των ορμών. Επισημαίνεται ότι ο εκτιμητής $\hat{\mu}_2$, πάσχει από προκατάληψη η οποία αίρεται όταν πολλαπλασιαστεί με διορθωτικό πολλαπλασιαστικό παράγοντα: $\hat{\mu}_3 = \frac{N}{N-1} \cdot \hat{\mu}_2$ 6.3.7

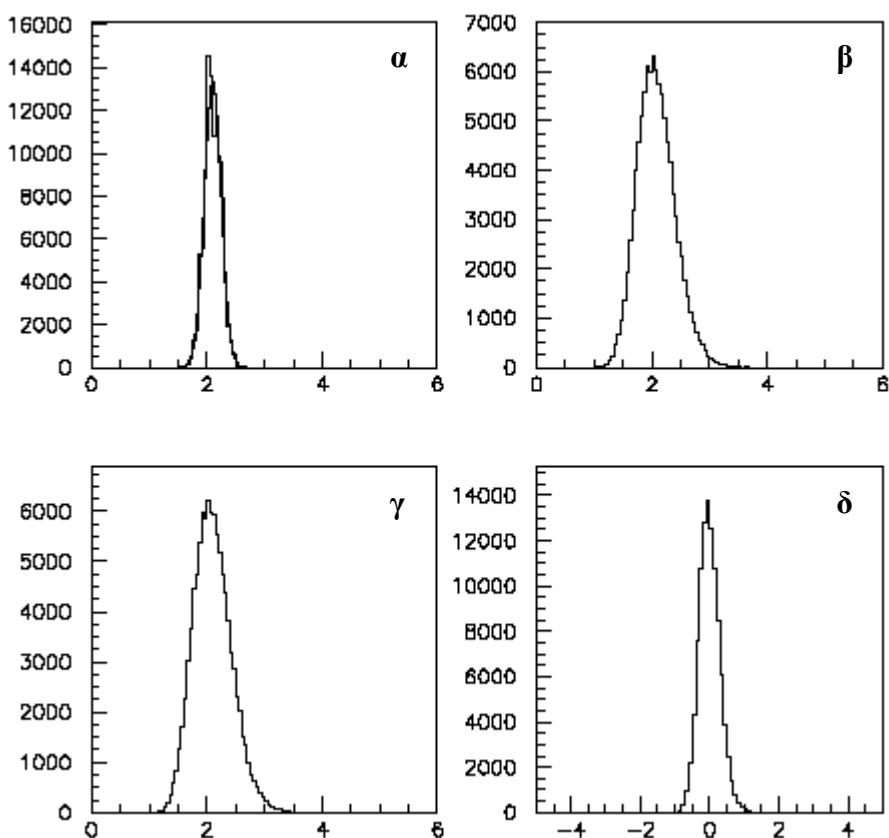
Παράδειγμα 6.3.2

Η εκτίμηση της παραμέτρου μ της Poissonian συνάρτησης πιθανότητας, από ένα δείγμα τιμών της τυχαίας μεταβλητής n , καταλήγει σε διαφορετικό αποτέλεσμα ανάλογα με τον εκτιμητή που χρησιμοποιείται.

Στο Σχήμα 6.3.1, παρίστανται υπό την μορφή ιστογράμματος τα αποτελέσματα 100000 επαναλήψεων του ίδιου πειράματος. Σε κάθε ένα από τα πειράματα, λαμβάνονται $N=100$ μετρήσεις, $n_1, n_2, n_3, \dots, n_{100}$, του ρυθμού των ραδιενεργών διασπάσεων (αριθμός διασπάσεων σε ένα δευτερόλεπτο) ενός δείγματος ραδιενεργού ισοτόπου. Ο καταμετρούμενος αριθμός ραδιενεργών διασπάσεων σε ένα δευτερόλεπτο, n , αυτού του ισοτόπου χαρακτηρίζεται από την Poissonian συνάρτηση πιθανότητας, με παράμετρο μ ίση με 2.1.

Μετά την εκτέλεση κάθε πειράματος εκτιμάται η τιμή της παραμέτρου μ , χρησιμοποιώντας τις πειραματικές μετρήσεις και τους εκτιμητές $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$.

Το θεώρημα του κεντρικού ορίου υπαγορεύει ότι η τυχαία μεταβλητή $\hat{\mu}_1$, ως άθροισμα τυχαίων μεταβλητών – ακόμα και εάν λαμβάνουν μόνο ακέραιες τιμές –, θα χαρακτηρίζεται από κανονική συνάρτηση πυκνότητας πιθανότητας καθώς το N τείνει προς το άπειρο.



Σχήμα 6.3.1: Γραφική παράσταση σε ιστογράμματα, των εκτιμήσεων: α) $\hat{\mu}_1$, β) $\hat{\mu}_2$ και γ) $\hat{\mu}_3$, της παραμέτρου μ καθώς και της διαφοράς: δ) $\hat{\mu}_3 - \hat{\mu}_1$. Οι εκτιμήσεις αντιστοιχούν σε 100000 δείγματα πειραματικών δεδομένων, κάθε ένα αποτελούμενο από $N=100$ γεγονότα, που ακολουθούν Poissonian συνάρτηση πιθανότητας:

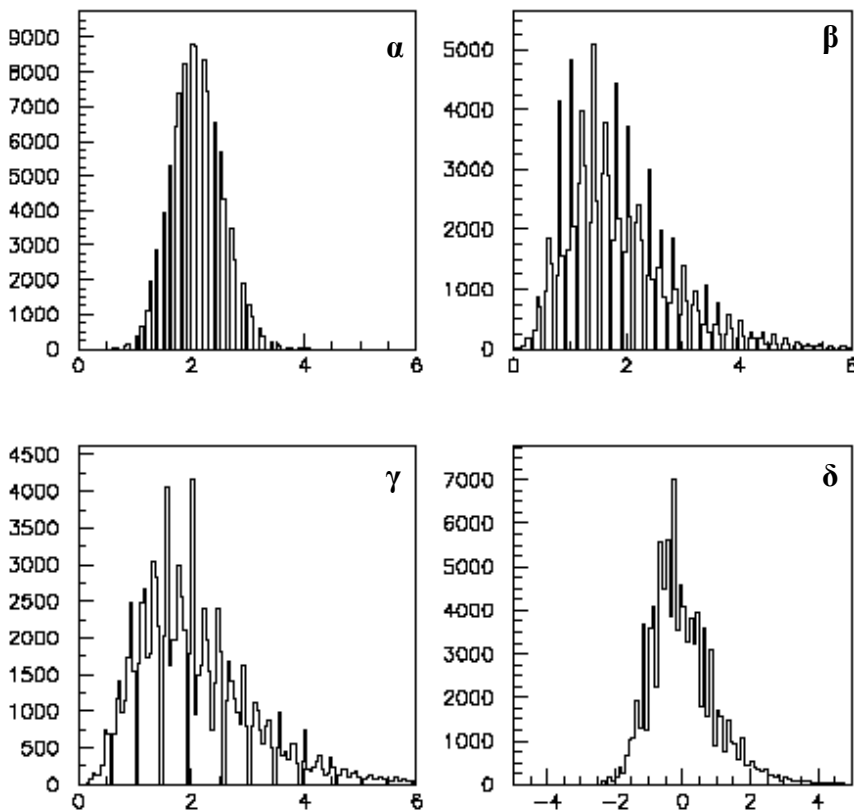
$$P(n) = \frac{\mu^n \cdot e^{-\mu}}{n!}$$
 Η συνεχής καμπύλη στο Σχήμα 6.3.1.α παριστά γραφικά, κανονική συνάρτηση.

Πράγματι, στο Σχήμα 6.3.1.α φαίνεται ότι, για πεπερασμένο πλήθος όρων, $N=100$, η συνάρτηση πυκνότητας πιθανότητας της εκτίμησης $\hat{\mu}_1$ προσεγγίζει την κανονική συνάρτηση και επιπλέον: $\langle \hat{\mu}_1 \rangle = 2.1$ ισούται με την πραγματική τιμή της παραμέτρου. Η ακρίβεια του εκτιμητή εκφράζεται από την διασπορά: $V(\hat{\mu}_1) = \sigma^2 = (0.15)^2$.

Στο Σχήμα 6.3.1.β παρίστανται υπό μορφή ιστογράμματος οι εκτιμήσεις της τιμής της παραμέτρου μ , σύμφωνα με τον εκτιμητή $\hat{\mu}_2$. Καθώς ο εκτιμητής είναι άθροισμα γραμμικών και τετραγωνικών ακέραιων όρων που συσχετίζονται (όροι όπως n και n^2), δεν ακολουθεί κανονική κατανομή. Επιπλέον, ο μέσος όρος των εκτιμήσεων $\langle \hat{\mu}_2 \rangle = 2.079$ αποκλίνει της πραγματικής τιμής, επιδεικνύοντας ότι η εκτίμηση πάσχει από προκατάληψη.

Ερώτηση: Παρατηρήσετε ότι $\langle \hat{\mu}_2 \rangle = 0.99 \cdot \mu$. Εξηγήστε την σχέση μεταξύ εκτίμησης και πραγματικής τιμής.

Υπόδειξη: Βλέπε Παράδειγμα 6.1.3



Σχήμα 6.3.2: Γραφική παράσταση, ως ιστογράμματα, των εκτιμήσεων: α) $\hat{\mu}_1$, β) $\hat{\mu}_2$ και γ) $\hat{\mu}_3$, της παραμέτρου μ καθώς και της διαφοράς: δ) $\hat{\mu}_3 - \hat{\mu}_1$. Οι εκτιμήσεις αντιστοιχούν σε 100000 δείγματα, κάθε ένα αποτελούμενο από $N=10$ γεγονότα που ακολουθούν Poissonian συνάρτηση πιθανότητας: $P(n) = \frac{\mu^n \cdot e^{-\mu}}{n!}$.

Στο Σχήμα 6.3.1.γ παρίστανται υπό μορφή ιστογράμματος οι εκτιμήσεις της τιμής της παραμέτρου μ , σύμφωνα με τον εκτιμητή $\hat{\mu}_3$. Ο μέσος όρος των εκτιμήσεων $\langle \hat{\mu}_3 \rangle = 2.1$

ισούται με την πραγματική τιμή της παραμέτρου και η διασπορά των εκτιμήσεων $\hat{\mu}_3$ είναι:

$V(\hat{\mu}_3) = (0.33)^2$, μεγαλύτερη της διασποράς $V(\hat{\mu}_1)$, υποδεικνύοντας ότι ο εκτιμητής $\hat{\mu}_3$ είναι ολιγότερο αποδοτικός

Τέλος στο Σχήμα 6.3.1.δ, παρίσταται η διαφορά στις εκτιμήσεις της τιμής της παραμέτρου, $\hat{\mu}_3 - \hat{\mu}_1$, που επιτυγχάνονται εφαρμόζοντας τους εκτιμητές $\hat{\mu}_3$ και $\hat{\mu}_1$ στο ίδιο σύνολο, $n_1, n_2, n_3, K, n_{100}$, πειραματικών δεδομένων: $\langle \hat{\mu}_3 - \hat{\mu}_1 \rangle = 0$

Στο Σχήμα 6.3.2 παρουσιάζονται τα αποτελέσματα 100000 εκτιμήσεων της παραμέτρου μ από τα δεδομένα ισάριθμων πειραμάτων, όταν σε κάθε πείραμα συλλέγονται μόνο $N=10$ μετρήσεις. Παρατηρήστε, ότι η δομή των ιστογραμμάτων παρουσιάζει περιοδικά κενά ως συνέπεια του γεγονότος ότι οι εκτιμητές εκφράζονται από αθροίσματα μικρού πλήθους ακεραίων. Ωστόσο, η γενική μορφή της κατανομής των εκτιμήσεων $\hat{\mu}_1$ είναι «κωδωνοειδής», ομοιάζουσα με κανονική, με $\langle \hat{\mu}_1 \rangle = 2.1$ ίση με την πραγματική τιμή της παραμέτρου. Η ακρίβεια του εκτιμητή εκφράζεται από την διασπορά: $V(\hat{\mu}_1) = \sigma^2 = (0.46)^2$. Ο εκτιμητής $\hat{\mu}_2$ πάσχει από προκατάληψη $b = \mu - E(\hat{\mu}_2) = \mu - \langle \hat{\mu}_2 \rangle = 2.1 - 1.88 = 0.22$, ενώ ο εκτιμητής $\hat{\mu}_3$ είναι απροκατάληπτος ($\langle \hat{\mu}_3 \rangle = 2.1$) αλλά λιγότερο αποδοτικός από τον εκτιμητή $\hat{\mu}_1$ ($V(\hat{\mu}_3) = (1.02)^2$).

Ερώτηση: Ερμηνεύσετε την αριθμητική σχέση μεταξύ των τιμών των στατιστικών μεγεθών $V(\hat{\mu}_1)$, $b = \mu - E(\hat{\mu}_2) = \mu - \langle \hat{\mu}_2 \rangle$ και $V(\hat{\mu}_3)$ για δείγματα των 100 και 1^ο γεγονότων.

Ερώτηση: Χρησιμοποιήσετε την μέθοδο των ορμών για να κατασκευάσετε τον εκτιμητή της διασποράς της τυχαίας μεταβλητής x , δεδομένου ότι διαθέτετε N τιμές, $\underline{x} = \{x_1, x_2, K, x_N\}$. Δείξτε ότι η εκτίμηση πάσχει από προκατάληψη.

Υπόδειξη:

$$V(x) = M^2 - (M^1)^2$$

$$\hat{V}(x) = \hat{M}^2 - (\hat{M}^1)^2 = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - \langle x \rangle)^2 \quad \text{βλέπε Παράδειγμα 6.1.3}$$

Όσον αφορά στο προσδιορισμό του σφάλματος εκτίμησης, μπορεί να αποδειχθεί εύκολα ότι:

$$V(\hat{M}^j) = V\left(\frac{1}{N} \cdot \sum_{i=1}^N (x_i)^j\right) = \frac{1}{N} \cdot (M^{2j} - (M^j)^2) \quad 6.3.5$$

$$\text{cov}(\hat{M}^j, \hat{M}^k) = \frac{1}{N} \cdot (M^{j+k} - M^j \cdot M^k)$$

Πράγματι:

$$\begin{aligned}
V(\hat{M}^j) &= V\left(\frac{1}{N} \cdot \sum_{i=1}^N (x_i)^j\right) = \frac{1}{N} \cdot V(x^j) \\
&= \frac{1}{N} \cdot E\left[\left(x^j - E(x^j)\right)^2\right] = \frac{1}{N} \cdot E\left[x^{2j} + E(x^j)^2 - 2 \cdot x^j \cdot E(x^j)\right] \\
&= \frac{1}{N} \cdot \left[E(x^{2j}) - E(x^j)^2\right] = \frac{1}{N} \cdot [M^j - M^{2j}]
\end{aligned}$$

Ερώτηση: Δείξτε ότι: $\text{cov}(M^j, M^k) = \frac{1}{N} \cdot (M^{j+k} - M^j \cdot M^k)$

Βεβαίως οι σχέσεις 6.3.5 δεν έχουν χρηστική αξία διότι θα πρέπει να υπολογισθούν οι ορμές $M^j = \int x^j \cdot f(x; \theta) dx$, όπου χρειάζεται να γνωρίζουμε τις τιμές των παραμέτρων.

Αντ' αυτών χρησιμοποιούμε τις εκτιμήσεις $\hat{M}^j = \frac{1}{N} \cdot \sum_{i=1}^N x_i^j$, για να ορίσουμε:

$$\hat{V}(\hat{M}^j) = \hat{V}\left(\frac{1}{N} \cdot \sum_{i=1}^N (x_i)^j\right) = \frac{1}{N-1} \cdot \left(\hat{M}^{2j} - (\hat{M}^j)^2\right) \quad 6.3.6$$

$$\hat{\text{cov}}(\hat{M}^j, \hat{M}^k) = \frac{1}{N-1} \cdot (\hat{M}^{j+k} - \hat{M}^j \cdot \hat{M}^k)$$

έχοντας αντικαταστήσει τον όρο $\frac{1}{N}$ με τον όρο $\frac{1}{N-1}$ προκειμένου να αποφευχθεί η προκατάληψη.

Η ακρίβεια της εκτίμησης της τιμής των παραμέτρων, $\theta_1, \theta_2, K, \theta_k$, εκφράζεται από την διασπορά και συνδιασπορά των τυχαίων μεταβλητών $\hat{\theta}_1, \hat{\theta}_2, K, \hat{\theta}_k$ οι οποίες ορίζονται ως λύσεις του συστήματος 6.3.4, συναρτήσεων των ορμών: $\hat{\theta}_i = f_i(\hat{M}^{r_1}, \hat{M}^{r_2}, K, \hat{M}^{r_k})$. Συνεπώς τα στοιχεία του πίνακα συνδιασποράς υπολογίζονται με την μέθοδο της μετάδοσης σφαλμάτων, χρησιμοποιώντας τις σχέσεις 6.3.6.

6.4 Μέθοδοι ορισμού συνεπών εκτιμητών – Μέθοδος της μεγίστης πιθανοφάνειας .

Ένας άλλος, γενικότερος τρόπος κατασκευής συνεπών εκτιμητών αναφέρεται στην επιλογή μίας συνάρτησης $a(x, \theta)$ η οποία είναι συνάρτηση, επιπλέον της τυχαίας μεταβλητής x , και της παραμέτρου θ . Η συνάρτηση $a(x, \theta)$ επιλέγεται ώστε το ολοκλήρωμα $h(\theta) = \int a(x, \theta) \cdot f(x; \theta) dx$ να έχει ως ρίζα, $h(\theta)_{\theta=\theta_0} = 0$, την πραγματική τιμή, θ_0 , της παραμέτρου.

Στη συνέχεια χρησιμοποιούνται οι N διαθέσιμες μετρήσεις για να ορισθεί η «πειραματική» συνάρτηση $\xi(\theta)$:

$$\xi(\theta) = \frac{1}{N} \sum_{i=1}^N a(x_i, \theta) \quad 6.4.1$$

Προφανώς, λόγω του θεωρήματος των μεγάλων αριθμών, ισχύει:

$$\xi(\theta_0) \xrightarrow{N \rightarrow \infty} E(a(x, \theta_0)) = h(\theta_0) = 0 \quad 6.4.2$$

και ο εκτιμητής της πραγματικής τιμής της παραμέτρου θ , ορίζεται ως η λύση της ακόλουθης εξίσωσης:

$$\xi(\hat{\theta}) = 0 \quad 6.4.3$$

Αποδεικνύεται ότι μία από τις ρίζες της 6.4.3 αποτελεί συνεπή εκτίμηση της πραγματικής τιμής της παραμέτρου, με την προϋπόθεση ότι ισχύουν μερικές ακόμα συνθήκες: ότι η συνάρτηση ξ είναι διαφορίσιμη και ότι υπάρχουν οι μέσες τιμές της ξ και της μερικής της παραγώγου ως προς θ , καθώς και ότι η 6.4.3 μπορεί να αντιστραφεί.

Στην περίπτωση όπου επιλέγουμε: $a(x, \theta) = \frac{\partial \ln f(x; \theta)}{\partial \theta}$, όπου $f(x; \theta)$ είναι η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής x , και με την επιπλέον συνθήκη να υπάρχουν οι αναμενόμενες τιμές των μεταβλητών: $\left. \frac{\partial \ln f(x; \theta)}{\partial \theta} \right|_{\theta=\theta_0}$ και $\left. \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} \right|_{\theta=\theta_0}$,

καθώς με την απαίτηση: οι τελεστές $\int dx$ και $\frac{\partial}{\partial \theta}$ να μετατίθενται, καταλήγουμε σε ένα εκτιμητή της παραμέτρου με σημαντικές ιδιότητες. Τον εκτιμητή μεγίστης πιθανοφάνειας - maximum likelihood estimator.

Χρησιμοποιώντας την συνθήκη κανονικοποίησης, εύκολα αποδεικνύουμε ότι:

$$\int_{\theta=\theta_0} f(x; \theta) dx = 1$$

$$\int_{\theta=\theta_0} \frac{\partial}{\partial \theta} (f(x; \theta)) dx = 0$$

$$\int_{\theta=\theta_0} \frac{\partial}{\partial \theta} (f(x; \theta)) \left(\frac{f(x; \theta)}{f(x; \theta)} \right)_{\theta=\theta_0} dx = 0$$

$$\int_{\theta=\theta_0} \frac{\partial}{\partial \theta} \ln(f(x; \theta)) \cdot f(x; \theta)_{\theta=\theta_0} dx = 0$$

Ή άλλως:

$$E \left[\frac{\partial}{\partial \theta} \ln(f(x; \theta)) \right]_{\theta=\theta_0} = 0 \quad 6.4.4$$

Πράγματι, η αναμενόμενη τιμή της επιλεγείσας συνάρτησης μηδενίζεται όταν η τιμή της παραμέτρου γίνει ίση της πραγματικής τιμής.

Επιπλέον, επειδή η «πειραματική συνάρτηση» $\xi(\theta) = N^{-1} \sum_{i=1}^N \frac{\partial}{\partial \theta} \ln f(x_i; \theta)_{\theta=\theta_0}$ συγκλίνει,

όταν το πλήθος των πειραματικών γεγονότων τείνει στο άπειρο, ως:

$$\xi(\theta_0) \xrightarrow{N \rightarrow \infty} E \left[\frac{\partial}{\partial \theta} \ln(f(x; \theta)) \right]_{\theta=\theta_0} = 0 \quad 6.4.5$$

η εκτίμηση της πραγματικής τιμής της παραμέτρου δίνεται ως λύση της εξίσωσης:

$$\xi(\hat{\theta}) = \sum_{i=1}^N \left(\frac{\partial \ln f(x_i; \theta)}{\partial \theta} \right)_{\theta=\hat{\theta}} = \left(\frac{\partial}{\partial \theta} \ln L(x_1, x_2, \dots, x_N; \theta) \right)_{\theta=\hat{\theta}} = 0 \quad \text{όπου} \quad 6.4.6$$

$L(x_1, x_2, \dots, x_N; \theta) = \prod_{i=1}^N f(x_i; \theta)$ είναι η συνάρτηση πιθανοφάνειας

Είναι εύκολο να αποδειχθεί ότι η λύση της 6.4.2 αντιστοιχεί σε μέγιστο της συνάρτησης πιθανοφάνειας. Πράγματι, παραγωγίζοντας τα δύο μέλη της 6.4.4 ως προς την παράμετρο θ , καταλήγουμε ότι:

$$\begin{aligned} \frac{\partial}{\partial \theta} \left(\int \frac{\partial \ln(f(x; \theta))}{\partial \theta} \cdot f(x; \theta) dx \right)_{\theta=\theta_0} &= 0 \\ \left(\int \left[\frac{\partial \ln(f(x; \theta))}{\partial \theta} \right]^2 \cdot f(x; \theta) dx \right)_{\theta=\theta_0} + \left(\int \frac{\partial^2 \ln(f(x; \theta))}{\partial \theta^2} \cdot f(x; \theta) dx \right)_{\theta=\theta_0} &= 0 \quad 6.4.7 \\ E \left[\left(\frac{\partial \ln(f(x; \theta))}{\partial \theta} \right)^2 \right]_{\theta=\theta_0} + E \left[\frac{\partial^2 \ln(f(x; \theta))}{\partial \theta^2} \right]_{\theta=\theta_0} &= 0 \end{aligned}$$

Επιπλέον, η παράγωγος της πειραματικής συνάρτησης $\xi(\theta)$, όταν το N τείνει στο άπειρο, συγκλίνει ως:

$$\frac{\partial \xi(\theta)}{\partial \theta} = N^{-1} \sum_{i=1}^N \frac{\partial^2}{\partial \theta^2} \ln f(x_i; \theta) \xrightarrow{N \rightarrow \infty} E \left[\frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) \right] \quad 6.4.8$$

Ωστε, από τις σχέσεις 6.4.7 και 6.4.8, καταλήγουμε ότι:

$$\frac{\partial \xi(\theta)}{\partial \theta} \xrightarrow{\theta=\theta_0, N \rightarrow \infty} -E \left[\left(\frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 \right]_{\theta=\theta_0} < 0 \quad 6.4.9$$

(η αναμενόμενη τυχαίας μεταβλητής, που δέχεται μόνο θετικές τιμές, είναι μεγαλύτερη από το μηδέν). Επειδή η εκτίμηση είναι συνεπής, βλέπε σχέση 6.4.4, χρησιμοποιώντας στην σχέση 6.4.9 ότι $\hat{\theta} \xrightarrow{N \rightarrow \infty} \theta_0$, καταλήγουμε:

$$\frac{\partial \xi(\theta)}{\partial \theta} \xrightarrow{\theta=\hat{\theta}, N \rightarrow \infty} -E \left[\left(\frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 \right]_{\theta=\theta_0} < 0 \quad 6.4.10$$

Δηλαδή ότι η εκτίμηση 6.4.6 αντιστοιχεί σε μέγιστο της συνάρτησης πιθανοφάνειας.

Αποδεικνύεται επίσης¹ ότι στο ασυμτωτικό όριο, $N \rightarrow \infty$, η εκτίμηση 6.4.6 αντιστοιχεί στο απόλυτο μέγιστο της συνάρτησης πιθανοφάνειας. Βεβαίως, για κάθε πρακτική χρήση της μεθόδου με πεπερασμένο πλήθος παρατηρήσεων, δεν υπάρχει τρόπος να εγγυάται ότι το μέγιστο που βρέθηκε αντιστοιχεί σε ένα απόλυτο ή ένα από τα τοπικά μέγιστα της συνάρτησης πιθανοφάνειας.

Ερώτηση: Δείξτε ότι η τυχαία μεταβλητή $\frac{\partial}{\partial \theta} \ln L(\underline{x}; \theta) = \frac{\partial}{\partial \theta} \left(\sum_{i=1}^N \ln f(x_i; \theta) \right)$ έχει μέση

τιμή ίση με το μηδέν και διασπορά ίση με $NI(\theta)$, όπου $I(\theta)$ είναι η πληροφορία που μεταφέρει κάθε ένα από τα πειραματικά δεδομένα αναφορικά με την παράμετρο θ .

Υπόδειξη: Χρησιμοποιώντας την σχέση 6.4.7, βρίσκουμε για την αναμενόμενη τιμή ότι:

¹ Eadie et al σελίδα 123 και υποενότητα 6.5

$$E \left[\frac{\partial}{\partial \theta} \ln L(\underline{x}; \theta) \right] = E \left[\sum_{i=1}^N \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right] = N \cdot E \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right] \quad 6.4.11$$

$$= N \cdot \int \left(\frac{\partial}{\partial \theta} \ln f(x; \theta) \right) \cdot f(x; \theta) dx = 0$$

Από την σχέση 6.4.11 και τις ιδιότητες της διασποράς καταλήγουμε ότι:

$$V \left[\frac{\partial}{\partial \theta} \ln L(\underline{x}; \theta) \right] = V \left[\sum_{i=1}^N \frac{\partial}{\partial \theta} \ln f(x_i; \theta) \right] = N \cdot V \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right]$$

$$= N \cdot \left(\int \left(\frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 \cdot f(x; \theta) dx - \left(\int \frac{\partial \ln f(x; \theta)}{\partial \theta} \cdot f(x; \theta) dx \right)^2 \right) \quad 6.4.12$$

$$= N \cdot I(\theta)$$

Παράδειγμα 6.4.1

Ας θεωρήσουμε την περίπτωση που μία φυσική σταθερά μετρείται σε N διαφορετικά πειράματα. Θα χρησιμοποιούμε το σύμβολο μ για να παραστήσουμε την αληθή τιμή της φυσικής σταθεράς. Θα υποθέσουμε επίσης ότι η συνάρτηση πυκνότητας πιθανότητας του αποτελέσματος κάθε πειράματος, x_i ($i = 1, 2, 3, \dots, N$), είναι κανονική συνάρτηση, με μέση τιμή την αληθή τιμή της φυσικής σταθεράς και διασπορά, σ_i^2 , που διαφέρει από πείραμα σε πείραμα.

Η πιθανοφάνεια που χαρακτηρίζει τα αποτελέσματα των N πειραμάτων, εκφράζεται συναρτήσει των μετρήσεων, ως:

$$L = \prod_{i=1}^N \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_i}} \cdot e^{-\frac{(x_i - \mu)^2}{2 \cdot \sigma_i^2}} \quad 6.4.13$$

$$\ln L = \sum_{i=1}^N \left[-\frac{1}{2} \cdot \ln(2\pi) - \ln \sigma_i - \frac{(x_i - \mu)^2}{2 \cdot \sigma_i^2} \right]$$

Η συνθήκη μεγιστοποίησης της πιθανοφάνειας:

$$\left(\frac{\partial \ln L}{\partial \mu} \right)_{\mu=\hat{\mu}} = \sum_{i=1}^N \frac{x_i - \hat{\mu}}{\sigma_i^2} = 0 \quad 6.4.14$$

καταλήγει στην εκτίμηση:

$$\hat{\mu} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \quad 6.4.15$$

Παρατηρήστε ότι η εκτίμηση 6.4.15 είναι το βεβαρημένο άθροισμα των μετρήσεων, όπου κάθε μέτρηση βαρύνεται με το αντίστροφο της διασποράς της.

Εύκολα μπορεί να δείξει κανείς ότι η εκτίμηση 6.4.15 είναι ελεύθερη προκατάληψης. Πράγματι η αναμενόμενη τιμή της εκτίμησης υπολογίζεται ως:

$$E[\hat{\mu}] = E\left[\frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}}\right] = \frac{\sum_{i=1}^N \frac{E[x_i]}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \frac{\sum_{i=1}^N \frac{\mu}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \mu \quad 6.4.16$$

Δηλαδή ίση με την αληθή τιμή.

Η ακρίβεια της εκτίμησης εκφράζεται από την διασπορά της (τυχαίας μεταβλητής) $\hat{\mu}$, ως:

$$V[\hat{\mu}] = E[\hat{\mu}^2] - (E[\hat{\mu}])^2 = \frac{E\left[\left(\sum_{i=1}^N \frac{x_i}{\sigma_i^2}\right)^2\right]}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^2} - \mu^2 \quad 6.4.17$$

$$= \frac{E\left[\sum_{i=1}^N \sum_{j=1}^N \frac{x_i \cdot x_j}{\sigma_i^2 \cdot \sigma_j^2}\right]}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^2} - \mu^2 = \frac{\sum_{i=1}^N \sum_{j=1}^N \frac{E[x_i \cdot x_j]}{\sigma_i^2 \cdot \sigma_j^2}}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^2} - \mu^2$$

Επειδή τα πειράματα είναι ανεξάρτητα μεταξύ τους, ισχύει ότι:

$$\text{cov}[x_i, x_j] = E[x_i \cdot x_j] - E[x_i] \cdot E[x_j] \begin{cases} = 0 \text{ για } i \neq j \\ = V(x) \text{ για } i = j \end{cases}$$

Συνεπώς:

$$E[x_i \cdot x_j] = \begin{cases} E[x_i] \cdot E[x_j] = \mu^2 \text{ για } i \neq j \\ E[x_i^2] = \sigma_i^2 + \mu^2 \text{ για } i = j \end{cases} \quad 6.4.18$$

Αντικαθιστώντας την σχέση 6.4.18 στην 6.4.17, καταλήγουμε ότι:

$$\begin{aligned} V[\hat{\mu}] &= \frac{\sum_{i=1}^N \frac{\sigma_i^2 + \mu^2}{\sigma_i^4} + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \frac{\mu^2}{\sigma_i^2 \cdot \sigma_j^2}}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^2} - \mu^2 \\ &= \frac{1}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^2} \cdot \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} + \mu^4 \cdot \sum_{i=1}^N \frac{1}{\sigma_i^4} + \mu^2 \cdot \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{\sigma_i^2 \cdot \sigma_j^2} \right) - \mu^2 \\ &= \frac{1}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)} + \mu^2 \cdot \left(\frac{\sum_{i=1}^N \frac{1}{\sigma_i^4} + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{\sigma_i^2 \cdot \sigma_j^2}}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^2} \right) - \mu^2 = \frac{1}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)} \end{aligned} \quad 6.4.19$$

Η πληροφορία, αναφορικά με την παράμετρο μ , που εμπεριέχεται στις N μετρήσεις υπολογίζεται ως:

$$I(\mu) = -E \left[\frac{\partial^2 \ln L}{\partial \mu^2} \right] = -E \left[\frac{\partial}{\partial \mu} \left(\sum_{i=1}^N \frac{x_i}{\sigma_i^2} - \sum_{i=1}^N \frac{\mu}{\sigma_i^2} \right) \right] \quad 6.4.20$$

$$= -E \left[-\sum_{i=1}^N \frac{1}{\sigma_i^2} \right] = \sum_{i=1}^N \frac{1}{\sigma_i^2}$$

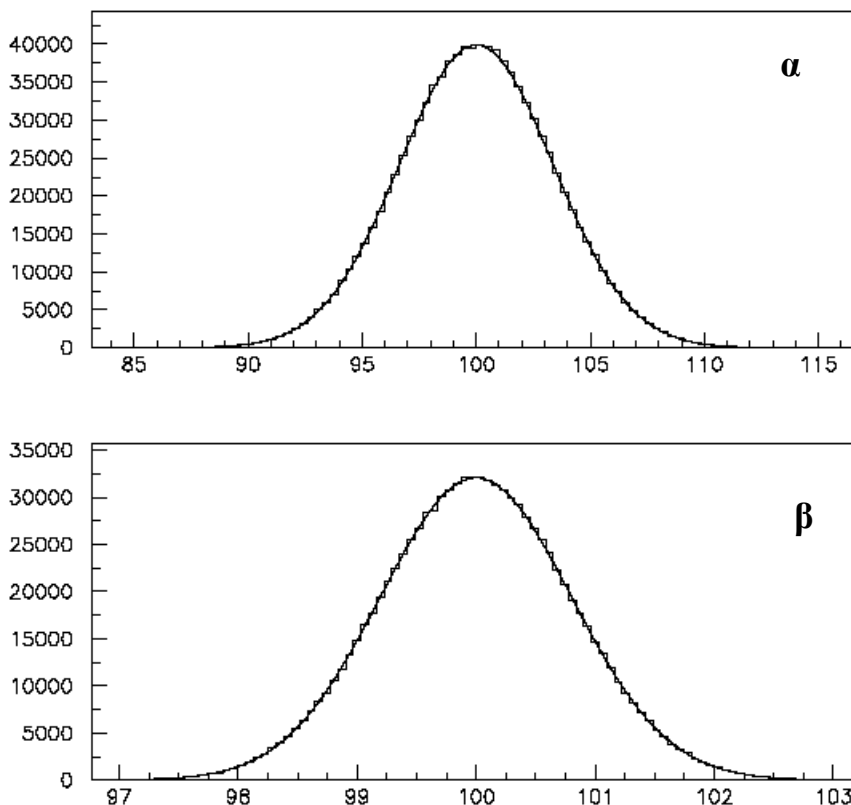
Συγκρίνοντας με την σχέση 6.4.19 καταλήγουμε ότι:

$$V[\hat{\mu}] = \frac{1}{I(\mu)} \quad 6.4.21$$

δηλαδή ότι ο εκτιμητής 6.4.15 έχει διασπορά ίση με το όριο που θέτει η ανισότητα των Cramer-Rao, για έναν εκτιμητή χωρίς προκατάληψη. Η σχέση 6.4.21 αποδεικνύει ότι ο εκτιμητής που ορίστηκε δια την συνθήκης μεγιστοποίησης της πιθανοφάνειας, σχέση 6.4.15, είναι αποδοτικός εκτιμητής.

Η ιδιότητα αυτή δεν ισχύει για όλους τους εκτιμητές μεγίστης πιθανοφάνειας. Θα αποδείξουμε όμως στην επόμενη υποενότητα ότι, εάν υπάρχει ένας αποδοτικός εκτιμητής χωρίς προκατάληψη τότε, ο εκτιμητής μεγίστης πιθανοφάνειας είναι αποδοτικός και δεν πάσχει από προκατάληψη.

Παράδειγμα 6.4.2



Σχήμα 6.4.1: Γραφική παράσταση, υπό μορφή ιστογράμματος, των αποτελεσμάτων εκτίμησης παραμέτρου με την σχέση 6.4.15. Οι συνεχείς γραμμές αναπαριστούν γραφικά κανονικές συναρτήσεις με παραμέτρους που περιγράφονται στο κείμενο.

Στο Σχήμα 6.4.1 παρίστανται, υπό μορφή ιστογράμματος τα αποτελέσματα δύο πειραματικών διαδικασιών, αντίστοιχων των διαδικασιών που περιγράφηκαν στο Παράδειγμα 6.4.1. Συγκεκριμένα, το Σχήμα 6.4.1.α αντιστοιχεί στη συλλογή μετρήσεων της τιμής κάποιας φυσικής ποσότητας, με αληθή τιμή 100, από τρία διαφορετικά πειράματα. Υποθέσαμε ότι η μέτρηση σε κάθε πείραμα χαρακτηρίζεται από κανονική συνάρτηση πυκνότητας πιθανότητας (δηλ. η διακριτική ικανότητα των μετρητικών οργάνων περιγράφεται από συνάρτηση Gauss) με διασπορά, σ^2 , που ποικίλει ανά πείραμα. Επιλέξαμε τις τιμές: $\sigma_1^2 = 5^2, \sigma_2^2 = 6^2, \sigma_3^2 = 7^2$, για τα τρία αντίστοιχα πειράματα.

Μετά από κάθε πειραματική διαδικασία η τιμή της φυσικής ποσότητας εκτιμήθηκε, σύμφωνα με την συνθήκη της μέγιστης πιθανοφάνειας, με την σχέση 6.4.15. Προκειμένου να δείξουμε τις στατιστικές ιδιότητες του εκτιμητή, επαναλάβουμε την πειραματική διαδικασία πολλές φορές (1000000) και τα αποτελέσματα κάθε εκτίμησης παρίστανται ως ιστόγραμμα στο Σχήμα 6.4.1.α σε σύγκριση με την γραφική παράσταση κανονικής

συνάρτησης με παραμέτρους: $\mu=100$ και $\sigma = \left[\left(\frac{1}{5^2} \right) + \left(\frac{1}{6^2} \right) + \left(\frac{1}{7^2} \right) \right]^{-\frac{1}{2}}$, σύμφωνα με τις σχέσεις 6.4.16 και 6.4.19.

Στο Σχήμα 6.4.2.β, παρίστανται τα αποτελέσματα άλλου αριθμητικού πειράματος όπου χρησιμοποιήθηκαν $N=10$ μετρήσεις που ακολουθούν κανονικές κατανομές με μέση (αληθή) τιμή $\mu=100$ και $\sigma_1^2 = 1^2, \sigma_2^2 = 2^2, \sigma_3^2 = 3^2, \dots, \sigma_{10}^2 = 10^2$. Τα αποτελέσματα από μεγάλο πλήθος εκτιμήσεων, σύμφωνα με την εκτίμηση μέγιστης πιθανοφάνειας 6.4.15, συγκρίνονται με την γραφική παράσταση κανονικής συνάρτησης με $\mu=100$ και

$$\sigma = \left[\sum_{i=1}^{10} \left(\frac{1}{i^2} \right) \right]^{-\frac{1}{2}}.$$

6.5 Ιδιότητες της μεθόδου μέγιστης πιθανοφάνειας .

Στην επόμενη υποενότητα θα μελετήσουμε τις ασυμπτωτικές ιδιότητες ($N \rightarrow \infty$) των εκτιμητών που ορίζονται με την μέθοδο της μέγιστης πιθανοφάνειας. Για πεπερασμένο πλήθος παρατηρήσεων, υπάρχει μόνο μία περίπτωση όπου ο εκτιμητής της μέγιστης πιθανοφάνειας έχει ιδανικές ιδιότητες. Αυτή είναι η περίπτωση όπου η συνάρτηση πυκνότητας πιθανότητας μπορεί να γραφεί σε εκθετική μορφή:

$$f(x; \theta) = \exp[\alpha(x) a(\theta) + \beta(x) + c(\theta)] \quad 6.5.1$$

Υπενθυμίζεται ότι αυτός ο συναρτησιακός τύπος πυκνότητας πιθανότητας (βλέπε υποενότητα 5.5 και σχέση 5.5.1), σύμφωνα με το Θεώρημα του Darms, επιδέχεται

επαρκείς στατιστικές συναρτήσεις. Συγκεκριμένα, η στατιστική συνάρτηση $R = \sum_{i=1}^N \alpha(x_i)$

και κάθε μονοτονική συνάρτηση της R (επί παραδείγματι η στατιστική συνάρτηση

$$t(\underline{x}) = t(x_1, x_1, K, x_N) = \frac{1}{N} \cdot \sum_{i=1}^N \alpha(x_i) \text{) είναι επαρκείς στατιστικές συναρτήσεις.}$$

Όταν η συνάρτηση πυκνότητας πιθανότητας ανήκει στην εκθετική κατηγορία 6.5.1, η πιθανοφάνεια γράφεται ως:

$$L(\underline{x};\theta) = \prod_{i=1}^N \exp [\alpha(x_i) \cdot a(\theta) + \beta(x_i) + c(\theta)] \quad 6.5.2$$

$$\ln L(\underline{x};\theta) = \sum_{i=1}^N \alpha(x_i) \cdot a(\theta) + \sum_{i=1}^N \beta(x_i) + c(\theta) \cdot N$$

όπου $\underline{x} = \{x_1, x_2, K, x_N\}$

Υπό τις συνθήκες ομαλότητας που έχουμε πολλαπλώς αναφέρει, ισχύει ότι:

$$\begin{aligned} E \left[\frac{\partial \ln L(\underline{x};\theta)}{\partial \theta} \right] &= \int_{\Omega} \frac{\partial \ln L(\underline{x};\theta)}{\partial \theta} \cdot L(\underline{x};\theta) dx_1 dx_2 K dx_N \\ &= \int_{\Omega} \frac{\partial L(\underline{x};\theta)}{\partial \theta} dx_1 dx_2 K dx_N \quad 6.5.3 \\ &= \frac{\partial}{\partial \theta} \int_{\Omega} L(\underline{x};\theta) dx_1 dx_2 K dx_N = 0 \end{aligned}$$

Ωστε για κάθε **αληθή** τιμή της παραμέτρου θ , οι σχέσεις 6.5.2 και 6.5.3 υποστηρίζουν ότι:

$$\begin{aligned} E \left[\frac{\partial \ln L(\underline{x};\theta)}{\partial \theta} \right] &= \left(\frac{da(\theta)}{d\theta} \right) \cdot E \left[\sum_{i=1}^N a(x_i) \right] + N \left(\frac{dc(\theta)}{d\theta} \right) = 0 \quad 6.5.4 \\ E[t(\underline{x})] &= E \left[\frac{1}{N} \sum_{i=1}^N a(x_i) \right] = - \frac{\left(\frac{da(\theta)}{d\theta} \right)_{\theta=\theta_0}}{\left(\frac{dc(\theta)}{d\theta} \right)_{\theta=\theta_0}} = r(\theta_0) \end{aligned}$$

Συνεπώς, εάν η συνάρτηση πυκνότητας πιθανότητας είναι της μορφής 6.5.1, η στατιστική συνάρτηση $t(\underline{x})$ είναι μη προκαταλημένος εκτιμητής της ποσότητας $r(\theta_0)$, **όχι κατ' ανάγκη της αληθούς τιμής της παραμέτρου θ , θ_0 . Μόνο στην ειδική περίπτωση όπου $r(\theta)=\theta$ ο εκτιμητής $\hat{\theta} = t(\underline{x}) = t(x_1, x_2, K, x_N) = \frac{1}{N} \cdot \sum_{i=1}^N \alpha(x_i)$ αναφέρεται στην αποδοτική και χωρίς προκατάληψη εκτίμηση της αληθούς τιμής θ_0 .**

Εύκολα μπορούμε να δείξουμε ότι η μέθοδος της μέγιστης πιθανοφάνειας καταλήγει ακριβώς στον ίδιο εκτιμητή για την ποσότητα-παραμέτρο $r(\theta)$. Πράγματι, σύμφωνα με την συνθήκη μέγιστης πιθανοφάνειας, ο εκτιμητής της πραγματικής τιμής της παραμέτρου θ

ορίζεται ως η λύση της εξίσωσης: $\left(\frac{d \ln L(\underline{x};\theta)}{d\theta} \right)_{\theta=\hat{\theta}} = 0$

Παραγωγίζοντας την σχέση 6.5.2 καταλήγουμε ότι:

$$\left(\frac{da(\theta)}{d\theta} \right)_{\theta=\hat{\theta}} \cdot \sum_{i=1}^N a(x_i) + N \left(\frac{dc(\theta)}{d\theta} \right)_{\theta=\hat{\theta}} = 0$$

$$\frac{1}{N} \cdot \sum_{i=1}^N a(x_i) = - \frac{\left(\frac{dc(\theta)}{d\theta} \right)_{\theta=\hat{\theta}}}{\left(\frac{da(\theta)}{d\theta} \right)_{\theta=\hat{\theta}}} = r(\hat{\theta}) \quad 6.5.6$$

Η εκτίμηση, $r(\hat{\theta}) = \frac{1}{N} \cdot \sum_{i=1}^N a(x_i)$, με την μέθοδο της μεγίστης πιθανοφάνειας, θα συμπεριφέρεται ως τυχαία μεταβλητή. Όπως αποδείχθηκε με την σχέση 6.5.5, η αναμενόμενη τιμή της εκτίμησης ισούται με την αληθή τιμή της ποσότητας: $E[r(\hat{\theta})] = E\left[\frac{1}{N} \cdot \sum_{i=1}^N a(x_i)\right] = r(\theta_0)$. Κατά συνέπεια, όταν η συνάρτηση πυκνότητας πιθανότητας έχει την εκθετική μορφή 6.5.1, η εκτίμηση, με την μέθοδο της μεγίστης πιθανοφάνειας, της ποσότητας $r(\theta)$ δεν πάσχει από προκατάληψη.

Επιπλέον, η δειγματική κατανομή του εκτιμητή t , $q(t; \theta)$, εκφράζεται από την υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας:

$$q(t; \theta) = \int_1^L \int_2^{\mathcal{B}} L(\underline{x}; \theta) dx_1 dx_2 \dots K dx_N = \int_L \int \delta\left(t - N^{-1} \sum_{i=1}^N a(x_i)\right) \cdot L(\underline{x}; \theta) dx_1 dx_2 \dots K dx_N$$

$$N^{-1} \sum_{i=1}^N a(x_i) = t$$

Εύκολα διαπιστώνεται ότι η δειγματική κατανομή είναι εκθετικής μορφής. Πράγματι:

$$q(t; \theta) = \int_1^K \int_2^{\mathcal{B}} \exp\left[\sum_{i=1}^N (a(x_i) \cdot a(\theta)) + \sum_{i=1}^N \beta(x_i) + N \cdot c(\theta)\right] dx_1 dx_2 \dots K dx_N$$

$$N^{-1} \sum_{i=1}^N a(x_i) = t$$

$$= \exp(N \cdot a(\theta)t + N \cdot c(\theta)) \cdot \int_1^K \int_2^{\mathcal{B}} \exp\left(\sum_{i=1}^N \beta(x_i)\right) dx^1 dx^2 \dots K dx^N \quad 6.5.7$$

$$\begin{matrix} 1 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ \exp(N\beta_1(t)) \end{matrix}$$

$$= \exp[N \cdot a(\theta)t + N \cdot c(\theta) + N \cdot \beta_1(t)]$$

Σύμφωνα με όσα δείξαμε στην παράγραφο 6.2, η εκτίμηση της ποσότητας $r(\theta)$ με την στατιστική συνάρτηση t είναι επί πλέον και εκτίμηση ελάχιστης διασποράς (Cramer-Rao). Η διασπορά του εκτιμητή, t , εκφράζεται συναρτήσει της πληροφορίας ως:

$$V(t) = \frac{1}{-E\left[\left(\frac{d^2 \ln L}{dr^2}\right)\right]_{r(\theta)=r(\theta_0)}} \quad 6.5.8$$

Εν κατακλείδι, εάν υπάρχει αποδοτικός και χωρίς προκατάληψη εκτιμητής της ποσότητας

$$r(\theta) = - \frac{\left(\frac{dc(\theta)}{d\theta} \right)}{\left(\frac{da(\theta)}{d\theta} \right)} \quad (\text{δηλαδή η συνάρτηση πυκνότητας πιθανότητας να γράφεται υπό την}$$

εκθετική μορφή 6.5.1) τότε ο εκτιμητής αυτός συμπίπτει με τον εκτιμητή μέγιστης πιθανοφάνειας. **Επισημαίνεται ότι στην περίπτωση που δεν υπάρχει τέτοιος εκτιμητής η μέθοδος της μέγιστης πιθανοφάνειας δεν οδηγεί απαραίτητα σε αποδοτική και χωρίς προκατάληψη εκτίμηση.**

Παράδειγμα 6.5.1

Στο Παράδειγμα 5.5.3 δείξαμε ότι η κανονική συνάρτηση πυκνότητας πιθανότητας γράφεται ως εκθετική συνάρτηση της μορφής 6.5.1. Εύκολα μπορεί να δει κανείς ότι η στατιστική συνάρτηση $t_1(\underline{x}) = \frac{R_1}{N} = \frac{1}{N} \cdot \sum_{i=1}^N x_i$ αποτελεί αποδοτικό και χωρίς προκατάληψη εκτιμητή της παραμέτρου μ . Πράγματι, η ποσότητα στην οποία αναφέρεται ο εκτιμητής t_1 , σύμφωνα με την σχέση 6.5.4 είναι:

$$-\frac{\frac{\partial c}{\partial \mu}}{\frac{\partial a_1}{\partial \mu}} = -\frac{\frac{\partial}{\partial \mu} \left(-\frac{\mu^2}{2\sigma^2} - \ln [\sqrt{2\pi}\sigma] \right)}{\frac{\partial}{\partial \mu} \left(\frac{\mu}{\sigma^2} \right)} = \mu \quad 6.5.9$$

Ωστόσο η στατιστική συνάρτηση R_2 δεν αναφέρεται αποκλειστικά στην παράμετρο σ όπως εύκολα μπορεί να δει κανείς από την σχέση 6.5.10.

$$-\frac{\frac{\partial c}{\partial \sigma}}{\frac{\partial a_2}{\partial \sigma}} = -\frac{\frac{\partial}{\partial \sigma} \left(-\frac{\mu^2}{2\sigma^2} - \ln [\sqrt{2\pi}\sigma] \right)}{\frac{\partial}{\partial \sigma} \left(\frac{1}{\sigma^2} \right)} \neq \sigma \quad 6.5.10$$

Συνεπώς, η εκθετική μορφή της κανονικής συνάρτησης πυκνότητας πιθανότητας δεν διασφαλίζει την ύπαρξη εκτιμητή μέγιστης πιθανοφάνειας, χωρίς προκατάληψη, για την παράμετρο σ .

Πράγματι, εάν κανείς διαθέτει N μετρήσεις της ίδιας φυσικής ποσότητας, οι οποίες λήφθηκαν με το ίδιο μετρητικό όργανο και χαρακτηρίζονται από κανονική συνάρτηση πυκνότητας πιθανότητας, θα μπορούσε να επιχειρήσει την σύγχρονη εκτίμηση της αληθούς τιμής της φυσικής ποσότητας, μ , και της διακριτικής ικανότητας του οργάνου, σ , με την μέθοδο της μέγιστης πιθανοφάνειας εκφρράοντας την πιθανοφάνεια ως:

$$L(\underline{x}; \mu, \sigma) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \cdot$$

Και απαιτώντας την σύγχρονη ικανοποίηση των ακόλουθων συνθηκών:

$$\left\{ \begin{array}{l} \left(\frac{\partial \ln L(\underline{x}; \mu, \sigma)}{\partial \mu} \right)_{\substack{\mu=\hat{\mu} \\ \sigma=\hat{\sigma}}} = 0 \\ \left(\frac{\partial \ln L(\underline{x}; \mu, \sigma)}{\partial \sigma} \right)_{\substack{\mu=\hat{\mu} \\ \sigma=\hat{\sigma}}} = 0 \end{array} \right\}$$

ορίζονται οι ακόλουθοι εκτιμητές μέγιστης πιθανοφάνειας:

$$\hat{\mu} = \frac{1}{N} \cdot \sum_{i=1}^N x_i$$

$$\hat{\sigma} = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - \hat{\mu})^2$$

Είναι προφανές, ότι η εκτίμηση της παραμέτρου σ πάσχει από προκατάληψη.

6.6 Ασυμπτωτικές ιδιότητες της μεθόδου μέγιστης πιθανοφάνειας.

Είναι σημαντικό να διακρίνουμε μεταξύ των ιδιοτήτων του εκτιμητή μέγιστης πιθανοφάνειας που αντιστοιχούν σε ένα πεπερασμένο δείγμα, πλήθους N , τιμών της ποσότητας x , και των ασυμπτωτικών ιδιοτήτων ($N \rightarrow \infty$) του εκτιμητή.

Ασυμπτωτικά, όπως δείξαμε στην υποενότητα 6.4, η εκτίμηση μέγιστης πιθανοφάνειας είναι συνεπής. Δηλαδή ένα από τα μέγιστα του λογαρίθμου της πιθανοφάνειας βρίσκεται αυθαίρετα κοντά στην πραγματική τιμή της παραμέτρου. Παράλληλα, όπως θα δείξουμε σ' αυτή την υποενότητα, αυτοί οι εκτιμητές είναι ασυμπτωτικά μη προκατειλημμένοι. Επιπλέον, θα αποδείξουμε ότι ο εκτιμητής μέγιστης πιθανοφάνειας πλησιάζει στο όριο μέγιστης ευαισθησίας, δηλαδή τό όριο που ορίζει η ανισότητα των Cramer-Rao.

Έστω η επανάληψις της ίδιας πειραματικής διαδικασίας. Σε κάθε πείραμα συλλέγονται N μετρήσεις, $\underline{x} = \{x_1, x_1, K, x_N\}$, της φυσικής ποσότητας X . Θα συμβολίζουμε ως $f(x, \theta)$ την συνάρτηση πυκνότητας πιθανότητας κάθε μέτρησης και ως ορίσουμε τις τυχαίες μεταβλητές $L(\underline{x}; \theta) = \prod_{i=1}^N f(x_i, \theta)$, $\partial L(\underline{x}; \theta) / \partial \theta$, $\hat{\theta}$. Οι τυχαίες μεταβλητές θα λάβουν, κατά την επανάληψη της πειραματικής διαδικασίας, τις ακόλουθες τιμές:

$$\begin{aligned} x^1 &= \{x_1^1, x_2^1, K, x_N^1\} \rightarrow L^1(\underline{x}^1 | \theta) = \prod_{i=1}^N f(x_i^1 | \theta), \quad \frac{\partial \ln L^1(\underline{x}^1 | \theta)}{\partial \theta}, \quad \left(\frac{\partial \ln L^1(\underline{x}^1 | \theta)}{\partial \theta} \right)_{\theta=\hat{\theta}^1} = 0 \\ x^2 &= \{x_1^2, x_2^2, K, x_N^2\} \rightarrow L^2(\underline{x}^2 | \theta) = \prod_{i=1}^N f(x_i^2 | \theta), \quad \frac{\partial \ln L^2(\underline{x}^2 | \theta)}{\partial \theta}, \quad \left(\frac{\partial \ln L^2(\underline{x}^2 | \theta)}{\partial \theta} \right)_{\theta=\hat{\theta}^2} = 0 \end{aligned} \quad 6.6.1$$

K K

$$x^n = \{x_1^n, x_2^n, K, x_N^n\} \rightarrow L^n(\underline{x}^n | \theta) = \prod_{i=1}^N f(x_i^n | \theta), \quad \frac{\partial \ln L^n(\underline{x}^n | \theta)}{\partial \theta}, \quad \left(\frac{\partial \ln L^n(\underline{x}^n | \theta)}{\partial \theta} \right)_{\theta=\hat{\theta}^n} = 0$$

όπου ο υπερυψωμένος δείκτης στις εκφράσεις 6.6.1 δηλώνει τον αύξοντα αριθμό της πειραματικής διαδικασίας προέλευσης των μετρήσεων.

Παράδειγμα 6.6.1

Έστω ότι σε ένα πείραμα μετρείται ο χρόνος ζωής, t , βραχύβιων φυσικών καταστάσεων. Για τις μετρήσεις χρησιμοποιούνται ιδανικές μετρητικές συσκευές, χωρίς πειραματικά σφάλματα και περιορισμούς στην γεωμετρική τους αποδοχή ή στην αποδοτικότητά τους. Σε κάθε επανάληψη του πειράματος παράγονται N τέτοιες καταστάσεις και μετρείται ο χρόνος που μεσολαβεί μέχρι την διάσπαση κάθε μίας από αυτές: $\underline{t} = \{t_1, t_1, K, t_N\}$. Το πείραμα επαναλαμβάνεται n φορές.

Η συνάρτηση πυκνότητας πιθανότητας του χρόνου ζωής, είτε πρόκειται για ραδιενεργούς πυρήνες ή βραχύβια σωματίδια ή έμβια συστήματα, εκφράζεται από την εκθετική συνάρτηση:

$$f(t; \lambda) = \frac{1}{\lambda} \cdot e^{-\frac{t}{\lambda}}, \quad t \geq 0 \quad 6.6.2$$

Εύκολα βρίσκει κανείς ότι η παράμετρος λ εκφράζει τον μέσο χρόνο ζωής (την αναμενόμενη τιμή της τυχαίας μεταβλητής t) των φυσικών καταστάσεων: $E[t] = \lambda$. Στο συγκεκριμένο παράδειγμα, η αληθινή τιμή του μέσου χρόνου ζωής, λ_0 , ισούται με 2 σε κάποιες αυθαίρετες μονάδες χρόνου.

Η τιμή της παραμέτρου λ εκτιμάται σε κάθε επανάληψη του πειράματος με την μέθοδο της μεγίστης πιθανοφάνειας. Οι τυχαίες μεταβλητές $L(\underline{t}; \lambda) = \prod_{i=1}^N f(t_i, \lambda)$, $\partial L(\underline{t}; \lambda) / \partial \lambda$, $\hat{\lambda}$ ορίζονται ως:

$$L^j(\underline{t}; \lambda) = \prod_{i=1}^N \frac{1}{\lambda} \cdot e^{-\frac{t_i^j}{\lambda}}$$

$$\frac{\partial \ln L^j(\underline{t}; \lambda)}{\partial \lambda} = -\frac{N}{\lambda} + \frac{\sum_{i=1}^N t_i^j}{\lambda^2} \quad 6.6.3$$

$$\left(\frac{\partial \ln L(\underline{t}; \lambda)}{\partial \lambda} \right)_{\lambda=\hat{\lambda}^j} = 0 \Rightarrow \hat{\lambda}^j = \frac{1}{N} \cdot \sum_{i=1}^N t_i^j$$

όπου $j=1, 2, 3, \dots, n$

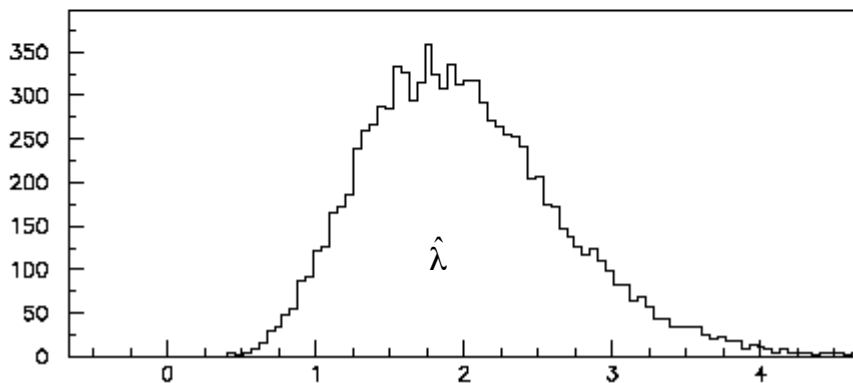
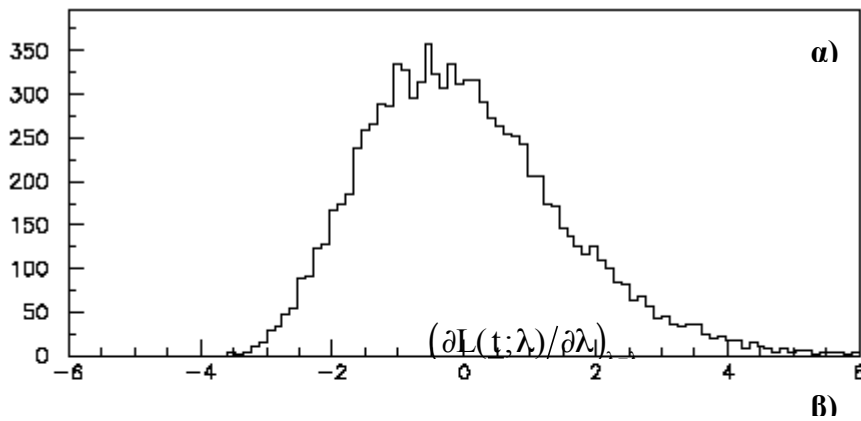
Στο Σχήμα 6.6.1 παρίστανται, υπό μορφή ιστογράμματος, οι τιμές των μεταβλητών $(\partial L(\underline{t}; \lambda) / \partial \lambda)_{\lambda=\lambda_0}$ και $\hat{\lambda}$, όπως προκύπτουν από $n=10000$ πειράματα, όπου σε κάθε πείραμα $N=9$. Η αναμενόμενη τιμή και η διασπορά των μεταβλητών, $(\partial L(\underline{t}; \lambda) / \partial \lambda)_{\lambda=\lambda_0}$ και $\hat{\lambda}$, υπολογίσθηκαν από τις διαθέσιμες τιμές και βρέθηκαν ίσες με:

$$E\left(\left(\partial L(\underline{t};\lambda)/\partial\lambda\right)_{\lambda=\lambda_0}\right); < \left(\partial L(\underline{t};\lambda)/\partial\lambda\right)_{\lambda=\lambda_0} > = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial L^i(\underline{t};\lambda)}{\partial\lambda} \right)_{\lambda=\lambda_0} = 0$$

$$E(\hat{\lambda}); < \hat{\lambda} > = \frac{1}{n} \sum_{i=1}^n (\hat{\lambda}_i) = 2$$

$$V\left(\left(\partial L(\underline{t};\lambda)/\partial\lambda\right)_{\lambda=\lambda_0}\right); \frac{1}{n-1} \sum_{i=1}^N \left(\frac{\partial L^i(\underline{t};\lambda)}{\partial\lambda} - < \left(\partial L(\underline{t};\lambda)/\partial\lambda\right)_{\lambda=\lambda_0} > \right)_{\lambda=\lambda_0}^2 = (1.5)^2 \quad 6.6.4$$

$$V(\hat{\lambda}); \frac{1}{n-1} \sum_{i=1}^n (\hat{\lambda}_i - < \hat{\lambda} >)^2 = (0.667)^2$$



Σχήμα 6.6.1: Γραφική αναπαράσταση σε ιστόγραμμα των τιμών των τυχαίων μεταβλητών α) $\left(\partial L(\underline{t};\lambda)/\partial\lambda\right)_{\lambda=\lambda_0}$ και β) $\hat{\lambda}$, για $N=9$, όπως υπολογίζονται από τα πειραματικά δεδομένα $(\underline{t}^j = \{t_1^j, t_2^j, K, t_9^j\}, j=1, 2, K, 10000)$ από τις σχέσεις 6.6.3.

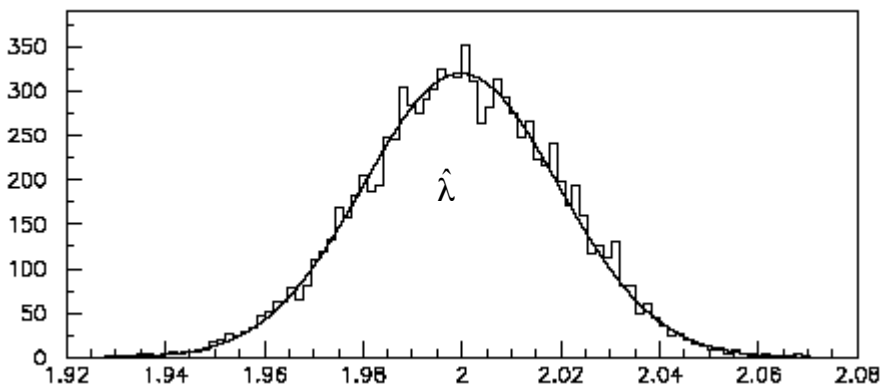
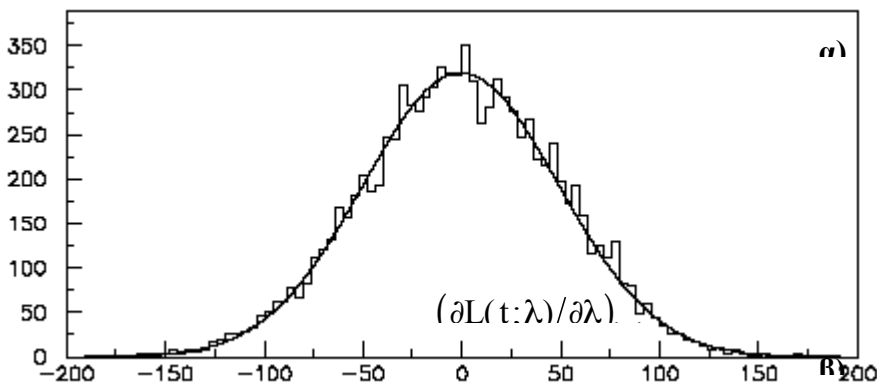
Αντίστοιχα, στο Σχήμα 6.6.2 παρίστανται οι τιμές των μεταβλητών $\left(\partial L(\underline{t};\lambda)/\partial\lambda\right)_{\lambda=\lambda_0}$ και $\hat{\lambda}$, όπως προκύπτουν από $n=10000$ πειράματα, όπου σε κάθε πείραμα $N=10000$. Η αναμενόμενη τιμή και η διασπορά των μεταβλητών, $\left(\partial L(\underline{t};\lambda)/\partial\lambda\right)_{\lambda=\lambda_0}$ και $\hat{\lambda}$, υπολογίσθηκαν από τις διαθέσιμες τιμές και βρέθηκαν ίσες με:

$$E\left(\left(\frac{\partial L(\underline{t};\lambda)}{\partial \lambda}\right)_{\lambda=\lambda_0}\right); \left\langle \left(\frac{\partial L(\underline{t};\lambda)}{\partial \lambda}\right)_{\lambda=\lambda_0} \right\rangle = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial L^i(\underline{t};\lambda)}{\partial \lambda} \right)_{\lambda=\lambda_0} = 0$$

$$E(\hat{\lambda}); \langle \hat{\lambda} \rangle = \frac{1}{n} \sum_{i=1}^n (\hat{\lambda}_i) = 2$$

$$V\left(\left(\frac{\partial L(\underline{t};\lambda)}{\partial \lambda}\right)_{\lambda=\lambda_0}\right); \frac{1}{n-1} \sum_{i=1}^n \left(\frac{\partial L^i(\underline{t};\lambda)}{\partial \lambda} - \left\langle \left(\frac{\partial L(\underline{t};\lambda)}{\partial \lambda}\right)_{\lambda=\lambda_0} \right\rangle \right)^2_{\lambda=\lambda_0} = (50)^2 \quad 6.6.5$$

$$V(\hat{\lambda}); \frac{1}{n-1} \sum_{i=1}^n (\hat{\lambda}_i - \langle \hat{\lambda} \rangle)^2 = (0.02)^2$$



Σχήμα 6.6.2: Γραφική αναπαράσταση σε ιστόγραμμα των τιμών των τυχαίων μεταβλητών α) $\left(\frac{\partial L(\underline{t};\lambda)}{\partial \lambda}\right)_{\lambda=\lambda_0}$ και β) $\hat{\lambda}$, για $N=10000$, όπως υπολογίζονται από τα πειραματικά δεδομένα $(\underline{t}^j = \{t_1^j, t_2^j, K, t_9^j\}, j=1,2,K,10000)$ από τις σχέσεις 6.6.3. Η συνεχής γραμμές παριστούν γραφικά κανονικές συναρτήσεις, με παραμέτρους ίσες με: α) $\mu=0$ και $\sigma=50$, β) $\mu=2$ και $\sigma=0.02$

Απλή σύγκριση μεταξύ των γραφικών αναπαραστάσεων στα Σχήματα 6.6.1 και 6.6.2 αρκεί να πείσει τον αναγνώστη ότι η συνάρτηση πυκνότητας πιθανότητας που περιγράφει τις τυχαίες μεταβλητές $(\partial L(\underline{x}; \lambda) / \partial \lambda)_{\lambda=\lambda_0}$ και $\hat{\lambda}$ εξαρτάται από το πλήθος, N , των μετρήσεων.

Επισημαίνεται επίσης ότι στην περίπτωση όπου $N=10000$, οι τυχαίες μεταβλητές περιγράφονται ικανοποιητικά από κανονικές συναρτήσεις με μέση τιμή και σ σε πλήρη συμφωνία με την αναμενόμενη τιμή και την διασπορά που υπολογίσθηκαν από τα πειραματικά δεδομένα με τις σχέσεις 6.6.5.

Σύμφωνα με τις σχέσεις 6.4.11 και 6.4.12, η αναμενόμενη τιμή της τυχαίας μεταβλητής $\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta}$, ισούται με μηδέν :

$$E\left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta}\right) = \int \frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \cdot L(\theta) dx_1 dx_2 \dots dx_N = 0 \quad 6.6.6$$

και έχει διασπορά ίση με:

$$\begin{aligned} V\left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta}\right) &= E\left(\left(0 - \frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta}\right)^2\right) = \int \left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta}\right)^2 \cdot L(\underline{x}; \theta) dx_1 dx_2 \dots dx_N \\ &= - \int \frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \cdot L(\underline{x}; \theta) dx_1 dx_2 \dots dx_N = -E\left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2}\right) \end{aligned} \quad 6.6.7$$

Παράδειγμα 6.6.2

Οι σχέσεις 6.6.6 και 6.6.7 ισχύουν για κάθε τιμή του N , του πλήθους των παρατηρήσεων.

Στην περίπτωση του παραδείγματος 6.6.1, η μέση τιμή της μεταβλητής $\frac{\partial \ln L(\underline{x}; \lambda)}{\partial \lambda}$

βρέθηκε να είναι ίση με μηδέν για $N=9$ και $N=10000$. Επιπλέον:

$$\begin{aligned} V\left(\frac{\partial \ln L(\underline{x}; \lambda)}{\partial \lambda}\right) &= - \int \frac{\partial^2 \ln L(\underline{x}; \lambda)}{\partial \lambda^2} \cdot L(\underline{x}; \lambda) dx_1 dx_2 \dots dx_N \\ &= - \int \sum_{i=1}^N \frac{\partial^2 \ln f(x_i; \lambda)}{\partial \lambda^2} \cdot \prod_{i=1}^N f(x_i; \lambda) dx_1 dx_2 \dots dx_N = -N \cdot \int \frac{\partial^2 \ln f(x; \lambda)}{\partial \lambda^2} \cdot f(x; \lambda) dx \\ &= \frac{N}{\lambda^2} \end{aligned}$$

σε άριστη συμφωνία με τις τιμές $(1.5)^2$ και $(50)^2$ που βρέθηκαν από τα δεδομένα του αριθμητικού παραδείγματος.

Παρατηρήστε επίσης ότι η έκφραση: $\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \ln \prod_{i=1}^N f(x_i; \theta) = \sum_{i=1}^N \frac{\partial \ln f(x_i; \theta)}{\partial \theta}$ αποτελεί το άθροισμα των N τυχαίων μεταβλητών $\frac{\partial \ln f(x_i; \theta)}{\partial \theta}$. Συνεπώς, το θεώρημα του κεντρικού

ορίου εγγυάται ότι η τυχαία μεταβλητή $\frac{\partial \ln L}{\partial \theta}$ θα χαρακτηρίζεται από κανονική συνάρτηση πυκνότητας πιθανότητας, καθώς το πλήθος των παρατηρήσεων, N , τείνει προς το άπειρο (στο ασυμπτωτικό όριο).

Πράγματι, στο παράδειγμα 6.6.1 η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $\frac{\partial \ln L}{\partial \theta}$, στην περίπτωση όπου το N είναι μεγάλος αριθμός ($N=10000$), περιγράφεται ικανοποιητικά από κανονική συνάρτηση με παραμέτρους που δίνονται από τις σχέσεις 6.6.6 και 6.6.7. Αντίθετα, όταν το πλήθος των παρατηρήσεων είναι μικρό οι τιμές της τυχαίας μεταβλητής $\frac{\partial \ln L}{\partial \theta}$ κατανομούνται ασυμμετρικά, όπως χαρακτηριστικά φαίνεται στο Σχήμα 6.6.1 για την περίπτωση με $N=9$

Αναπτύσσοντας την συνθήκη μέγιστης πιθανοφάνειας: $\left[\frac{\partial \ln L(\theta)}{\partial \theta} \right]_{\theta=\hat{\theta}} = 0$, σε ανάπτυγμα

σειράς Taylor γύρω από την πραγματική τιμή, θ_0 , της παραμέτρου θ , καταλήγουμε ότι:

$$\left[\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \right]_{\theta=\hat{\theta}} = \left[\sum_{k=1}^N \frac{(\hat{\theta} - \theta_0)^{k-1}}{(k-1)!} \cdot \left(\frac{\partial^k \ln L(\underline{x}; \theta)}{\partial \theta^k} \right)_{\theta=\theta_0} \right] \quad 6.6.8$$

$$\left[\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \right]_{\theta=\hat{\theta}} = \left[\left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \right)_{\theta=\theta_0} + (\hat{\theta} - \theta_0) \left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right)_{\theta=\theta_0} + \underbrace{\frac{1}{6} \frac{\partial^3 \ln L(\underline{x}; \theta)}{\partial \theta^3}}_{\text{όροι μεγαλύτερης τάξης}} \right] = 0 \quad 6.6.9$$

Επειδή η εκτίμηση με την μέθοδο της μέγιστης πιθανοφάνειας είναι συνεπής, η ποσότητα $\hat{\theta}$ βρίσκεται (στο ασυμπτωτικό όριο) απειροστά κοντά στην πραγματική τιμή. Συνεπώς, με πολλή καλή προσέγγιση, μπορούμε να παραλείψουμε όρους μεγαλύτερης της πρώτης τάξης στην σχέση 6.6.9 και να καταλήξουμε στην ακόλουθη σχέση:

$$\left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \right)_{\theta=\theta_0} + (\hat{\theta} - \theta_0) \cdot \left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right)_{\theta=\theta_0} \cong 0 \quad 6.6.10$$

Υπό την προϋπόθεση ότι: η δεύτερη παράγωγος στην σχέση 6.6.10 είναι σταθερά ή ότι μεταβάλλεται πολύ λίγο για τιμές της παραμέτρου θ γύρω από την τιμή θ_0 , η απόκλιση της εκτίμησης από την πραγματική τιμή, $(\hat{\theta} - \theta_0)$, θα είναι ανάλογη της πρώτης παραγώγου του λογαρίθμου της πιθανοφάνειας, δηλαδή:

$$\hat{\theta} - \theta_0 = - \frac{\left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \right)_{\theta=\theta_0}}{\left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right)_{\theta=\theta_0}} \quad 6.6.11$$

Ας συγκεντρώσουμε τα συμπεράσματα στα οποία καταλήξαμε για την εκτίμηση με την μέθοδο της μεγίστης πιθανοφάνειας, όταν το πλήθος των παρατηρήσεων τείνει στο άπειρο:

- Η ποσότητα $r = \left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \right)_{\theta=\theta_0}$ χαρακτηρίζεται από κανονική συνάρτηση

πυκνότητας πιθανότητας με μέση τιμή ίση με μηδέν (6.6.6) και διασπορά που δίνεται από την σχέση 6.6.7.

Συγκεκριμένα η συνάρτηση πυκνότητας πιθανότητας για αυτή την ποσότητα εκφράζεται ως:

$$P(r)dr = \frac{e^{\left[\frac{-(r)^2}{2 \times \left(-E \left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right) \right)} \right]}}{\sqrt{2\pi} \left[-E \left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right) \right]_{\theta=\theta_0}^{1/2}} dr$$

- Η απόκλιση της εκτίμησης από την πραγματική τιμή σχετίζεται με την παράγωγο $\left(\frac{\partial \ln L(\underline{x}; \theta)}{\partial \theta} \right)_{\theta=\theta_0}$ με την έκφραση 6.6.11

Κατά συνέπεια η εκτίμηση της παραμέτρου, $\hat{\theta}$ ακολουθεί κανονική συνάρτηση πυκνότητας πιθανότητας, με μέση τιμή την αληθή τιμή θ_0 και διασπορά που εκφράζεται ως:

$$V(\hat{\theta}) = - \frac{-E \left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right)_{\theta=\theta_0}}{\left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right)_{\theta=\theta_0}^2} \quad 6.6.12$$

Χρησιμοποιώντας τις ασυμπτωτικές ιδιότητες ($N \rightarrow \infty$) του νόμου των μεγάλων αριθμών, καταλήγουμε ότι:

$$\begin{aligned} E \left(\frac{\partial^2 \ln L(\underline{x}; \theta)}{\partial \theta^2} \right) &= E \left(\frac{\partial^2}{\partial \theta^2} \sum_{i=1}^N \ln f(x_i; \theta) \right) = N \cdot E \left(\frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) \right) \\ &= \lim_{N \rightarrow \infty} N \cdot \frac{1}{N} \sum_{j=1}^N \frac{\partial^2}{\partial \theta^2} \ln f(x_j; \theta) = \frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \end{aligned} \quad 6.6.13$$

Αντικαθιστώντας την σχέση 6.6.13 στη σχέση 6.6.12, η έκφραση της διασποράς της εκτίμησης, $V(\hat{\theta})$, απλοποιείται ως:

$$V(\hat{\theta}) = -\frac{1}{E\left(\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \Big|_{\theta=\theta^0}\right)} = -\frac{1}{I_{\theta}} = -\frac{1}{\frac{d^2 \ln L(\theta)}{d\theta^2} \Big|_{\theta=\theta^0}} \quad 6.6.14$$

Επιπλέον, συγκρίνοντας με την ανισότητα Cramer-Rao, παρατηρούμε ότι η εκτίμηση μέγιστης πιθανοφάνειας φθάνει, στο ασυμπτωτικό όριο, την μέγιστη δυνατή ακρίβεια.

Παρατηρήστε ότι πράγματι τα αποτελέσματα του αριθμητικού Παραδείγματος 6.6.1, ευρίσκονται σε συμφωνία με την σχέση 6.6.14. Επιπλέον η εκτίμηση της παραμέτρου λ χαρακτηρίζεται από κανονική συνάρτηση πιθανότητας, όταν ο αριθμός των παρατηρήσεων είναι μεγάλος. Είναι προφανές, από το Σχήμα 6.6.1, ότι ο εκτιμητής μέγιστης πιθανοφάνειας δεν χαρακτηρίζεται από «κανονική συμπεριφορά», όταν το πλήθος των παρατηρήσεων είναι μικρό (π.χ. $N=9$).

6.11 Θεώρημα Gauss-Markov

Σ' αυτή την υποενότητα παρουσιάζεται, χωρίς απόδειξη, το θεώρημα Gauss-Markov το οποίο προσφέρει την γενική θεμελίωση των ιδιοτήτων της μεθόδου ελαχίστων τετραγώνων.

Έστω τα πειραματικά δεδομένα $\underline{y} = \{y_1, y_2, y_3, \dots, y_N\}$ και το πρότυπο προσαρμογής $\phi(x; \underline{\theta})$, ώστε $y_i = \phi(x_i; \underline{\theta}) + \varepsilon_i$. Εάν ικανοποιούνται οι συνθήκες: $E[\varepsilon_i] = 0$ ($i = 1, 2, 3, \dots, N$) και τα στοιχεία του πίνακα συνδιασποράς $V_{0 \times 0}[\underline{\varepsilon}]$ δεν απειρίζονται και δεν εξαρτώνται από τις τιμές των παραμέτρων και τα πειραματικά δεδομένα, τότε η εκτίμηση με την μέθοδο των ελαχίστων τετραγώνων δεν πάσχει από προκατάληψη και έχει την μικρότερη διασπορά από όλους τους μη προκατηλημένους εκτιμητές που είναι γραμμικές συναρτήσεις των δεδομένων ($\underline{y} = \{y_1, y_2, y_3, \dots, y_N\}$), ανεξάρτητα από την συνάρτηση πυκνότητα πιθανότητας των αποκλίσεων ε_i .

Επισημαίνεται, ότι το θεώρημα αναφέρεται σε γραμμικούς μη προκατηλημένους εκτιμητές. Είναι δυνατόν να βρεθεί μη γραμμικός ή προκατηλημένος εκτιμητής με μικρότερη διασπορά από τον εκτιμητή ελαχίστων τετραγώνων.

Άσκησης Αυτοαξιολόγησης

Άσκηση 6.1

Προκειμένου σε ένα πείραμα να εκτιμηθεί ο χρόνος ημι-ζωής ενός τύπου ασταθών σωματίων, εξετάζεται η κατανομή του ιδίου χρόνου¹ που μεσολαβεί από την δημιουργία μέχρι την διάσπαση κάθε ενός από N τέτοια σωματίδια.

Τα σωματίδια κινούνται παράλληλα με τον άξονα x . Η πειραματική συσκευή έχει περιορισμένες διαστάσεις, όπως παρίσταται στο Σχήμα 1, ώστε μπορεί να μελετήσει την διάσπαση σωματίων που συμβαίνουν σε $L_1 \leq x \leq L_2$.

Η πειραματική πληροφορία, που συλλέγεται, για κάθε ένα από τα N σωματίδια, συνίσταται στο μήκος της τροχιάς, l_i , του σωματίου που διέγραψε μέχρι να διασπασθεί και την κινηματική παράμετρο $\beta_i = v_i/c$ (δηλαδή τον λόγο της ταχύτητας του σωματίου, v_i , ως προς την ταχύτητα του φωτός στο κενό, c).

Να εκτιμηθεί ο χρόνος ημιζωής από τα πειραματικά δεδομένα

Λύση της άσκησης 6.1

Ο ιδιόχρονος ζωής ενός σωματίου t_i (δηλαδή η «ηλικία θανάτου» του σωματίου όπως μετράται από κάποιον παρατηρητή ο οποίος έχει την ίδια ακριβώς ταχύτητα με αυτό το σωματίο και κατά συνέπεια αυτός ο παρατηρητής βλέπει το σωματίο να ηρεμεί) δίνεται από την σχέση $t_i = l_i / (\beta_i \gamma_i c)$ ².

¹Χρόνος ανηγμένος στο σύστημα ηρεμίας του σωματίου.

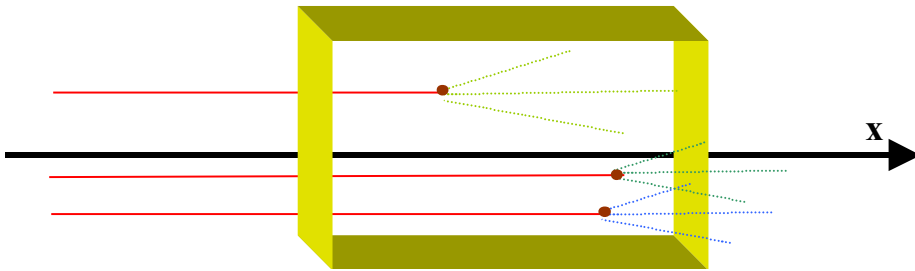
² Όπου $\gamma = [1 - \beta^2]^{-1/2}$

Για απλούστευση θεωρούμε ότι όλα τα σωματίδια έχουν τροχιές παράλληλες προς τον άξονα x και γεννώνται σε σημεία που αντιστοιχούν στην ίδια συντεταγμένη $x=x_0$, όπως φαίνεται στο Σχήμα 1.

Επειδή: α) η πειραματική διάταξη έχει περιορισμένες γεωμετρικές διαστάσεις και β) συνήθως τα σωματίδια γεννούνται έξω αλλά διασπώνται μέσα στην πειραματική διάταξη, μπορούμε να μετρήσουμε μήκη τροχιών μεταξύ L_1 και L_2 .

Προφανώς το όριο L_1 αντιστοιχεί στο μήκος της τροχιάς του σωματίου που διασπάσθηκε αμέσως μετά την είσοδο του στην πειραματική διάταξη, ενώ το όριο L_2 αντιστοιχεί στο μήκος της τροχιάς εκείνου του σωματίου που διασπάσθηκε αμέσως πριν την έξοδο αυτού του σωματίου από τον ανιχνευτή.

Ο ιδιοχρόνος, t_i , που αντιστοιχεί στο υπ' αριθμό i σωματίο με μήκος τροχιάς l_i , θα ικανοποιεί την σχέση $t_1^i \leq t_i \leq t_2^i$, όπου $t_{1,2}^i = L_{1,2} / (\beta_i \gamma_i c)$.



Σχήμα 1: Αναπαράσταση τροχιών σωματίων από το σημείο παραγωγής τους μέχρι το σημείο διάσπασης τους. Το πλαίσιο παριστάνει τον ενεργό χώρο της πειραματικής διάταξης.

Η συνάρτηση πυκνότητας πιθανότητας των ιδιοχρόνων των σωματίων που διασπώνται μέσα στα γεωμετρικά όρια της πειραματικής διάταξης ορίζεται από τον γνωστό εκθετικό νόμο διάσπασης:

$$P(t_i; \tau) = R \cdot \frac{1}{\tau} \exp\left(-\frac{t_i}{\tau}\right) \quad 1.$$

όπου ο παράγων R είναι παράγων κανονικοποίησης και ο παράγων τ αντιστοιχεί στον μέσο χρόνο ζωής.

Επειδή μελετούμε σωματίδια που διασπώνται μέσα στα γεωμετρικά όρια της πειραματικής διάταξης (υπό συνθήκη πιθανότητα) το ολοκλήρωμα της συνάρτησης πυκνότητας πιθανότητας του χρόνου ζωής του σωματίου i στα χρονικά όρια $[t_1^i, t_2^i]$ θα πρέπει να ισούται με την μονάδα.

Συνεπώς η συνάρτηση πυκνότητας θα έχει την αναλυτική μορφή:

$$P(t_i; \tau) = \frac{\frac{1}{\tau} \exp\left(-\frac{t_i}{\tau}\right)}{\exp\left(-\frac{t_1^i}{\tau}\right) - \exp\left(-\frac{t_2^i}{\tau}\right)} \quad 2.$$

Ο λογάριθμος της πιθανοφάνειας που αντιστοιχεί στα N σωματίδια θα είναι:

$$\ln L = \sum_{i=1}^N \left\{ -\frac{t_i}{\tau} - \ln \left[\exp\left(\frac{-t_1^i}{\tau}\right) - \exp\left(\frac{-t_2^i}{\tau}\right) \right] \right\} - N \ln \tau \quad 3.$$

και ο εκτιμητής του χρόνου ημιζωής θα ευρίσκεται από την συνθήκη: $\left(\frac{\partial \ln L}{\partial \tau} \right)_{\tau=\hat{\tau}} = 0$, ως

λύση της εξίσωσης:

$$\sum_{i=1}^N \frac{t_i}{\hat{\tau}^2} - \frac{1}{\hat{\tau}^2} \cdot \sum_{i=1}^N \frac{t_1^i \cdot \exp\left(\frac{-t_1^i}{\hat{\tau}}\right) - t_2^i \cdot \exp\left(\frac{-t_2^i}{\hat{\tau}}\right)}{\exp\left(\frac{-t_1^i}{\hat{\tau}}\right) - \exp\left(\frac{-t_2^i}{\hat{\tau}}\right)} - \frac{N}{\hat{\tau}} = 0 \quad 4.$$

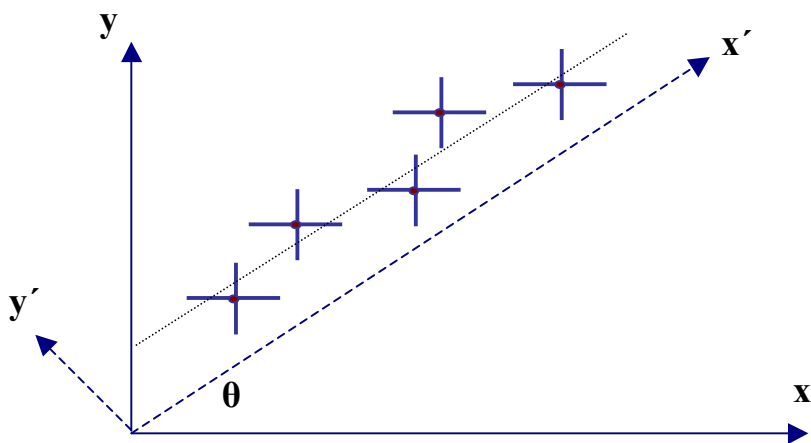
Στην περίπτωση όπου: $L_1 = 0$ και $L_2 = \infty$, όταν δηλαδή η πειραματική διάταξη μπορεί να ανιχνεύσει όλα τα σωμάτια, τότε για κάθε σωματίο ισχυει: $t_1^i = 0$, $t_2^i = \infty$ και η σχέση 6.1.4 ανάγεται στην απλή εκτίμηση:

$$\hat{\tau} = \frac{1}{N} \sum t_i \quad 5.$$

Άσκηση 6.2

Έστω ένα σύνολο μετρήσεων των τιμών των μεγεθών X και Y : $\{y_i, x_i\}, i = 1, 2, 3, \dots, N$.

Θα υποθέσουμε ότι ο προσδιορισμός αμφοτέρων των μετρήσεων x_i και y_i πάσχουν από μετρητικά σφάλματα. Για απλότητα θα υποθέσουμε ότι όλες οι μετρήσεις των ποσοτήτων x έχουν το ίδιο σφάλμα, σ_x , ενώ οι μετρήσεις y έχουν το ίδιο σφάλμα σ_y όπως παρίσταται γραφικά στο Σχήμα 1. Να προσαρμοσθεί στα πειραματικά δεδομένα το γραμμικό πρότυπο: $y(x) = m \cdot x + c$ (να εκτιμηθούν οι παράμετροι m και c) με την μέθοδο των ελαχίστων τετραγώνων.



Σχήμα 1.

Λύση της άσκησης 6.2

Ας υποθέσουμε ότι με κάποιο τρόπο γνωρίζουμε τις παραμέτρους της «αληθούς» ευθείας που εκφράζει την συναρτησιακή σχέση μεταξύ y και x .

Ορίζουμε ένα σύστημα ορθογωνίων αξόνων ώστε ο άξονας x' να είναι παράλληλος της «αληθούς» ευθείας όπως φαίνεται στο Σχήμα 1. Οι σχέσεις μετασχηματισμού από το ένα καρτεσιανό σύστημα στο άλλο θα είναι:

$$\begin{aligned} y' &= y \cos\theta - x \sin\theta \\ x' &= y \sin\theta + x \cos\theta \end{aligned} \quad 1.$$

Στο σύστημα $x' y'$, η γραμμική σχέση $y=mx+c$ εκφυλίζεται στην σχέση $y'=c'$. Κατά συνέπεια κάθε σφάλμα κατά την διεύθυνση του x' δεν επηρεάζει πλέον το πρόβλημα. Στην συνέχεια ορίζουμε το άθροισμα τετραγώνων, Q^2 στο σύστημα συντεταγμένων $x' y'$, ως:

$$Q^2 = \frac{(y'_i - c')^2}{\sigma_{y'_i}^2} \text{ όπου } \sigma_{y'_i}^2 = \sigma_y^2 \cdot \cos^2\theta + \sigma_x^2 \cdot \sin^2\theta \quad 2.$$

Και μετά από αντικατάσταση βρίσκουμε ότι:

$$\begin{aligned} Q^2 &= \sum \frac{[y_i \cdot \cos\theta - x_i \cdot \sin\theta - c']^2}{\sigma_y^2 \cdot \cos^2\theta + \sigma_x^2 \cdot \sin^2\theta} \\ &= \sum \frac{\left[y_i - x_i \cdot \frac{\sin\theta}{\cos\theta} - \frac{c'}{\cos\theta} \right]^2}{\sigma_y^2 + \sigma_x^2 \cdot \frac{\sin^2\theta}{\cos^2\theta}} \quad 3. \\ &= \sum \frac{[y_i - x_i m - c]^2}{\sigma_y^2 + m^2 \sigma_x^2} \end{aligned}$$

Η ελαχιστοποίηση της ποσότητας Q^2 καταλήγει στις ακόλουθες εκτιμήσεις:

$$\hat{m} = \frac{\sigma_y}{\sigma_x} \left[A + \frac{\text{cov}[x,y]}{|\text{cov}[x,y]|} \sqrt{A^2 + 1} \right] \text{ και } \hat{c} = \frac{\sum y_i}{N} - \hat{m} \frac{\sum x_i}{N} \quad 4.$$

όπου

$$A = \frac{\sigma_x^2 V[y] - \sigma_y^2 V[x]}{2\sigma_x \sigma_y \text{cov}[x,y]} \quad 5.$$

$$V[x] = \langle x^2 \rangle - \langle x \rangle^2, \quad V[y] = \langle y^2 \rangle - \langle y \rangle^2, \quad \text{cov}[x,y] = \langle x \cdot y \rangle - \langle x \rangle \cdot \langle y \rangle$$

Αξίζει να σημειωθεί ότι αναλυτική λύση υπάρχει μόνο όταν το μετρητικό σφάλμα είναι το ίδιο για όλα τα ζεύγη x και y . Σ' αντίθετη περίπτωση πρέπει να αναζητηθούν αριθμητικές μέθοδοι ελαχιστοποίησης του Q^2 .