

Κεφάλαιο 5

Πληροφορία

5.1 Ορισμοί.

Η διεξαγωγή πειραμάτων και η συλλογή πειραματικών δεδομένων αποσκοπεί στην συλλογή πληροφορίας αναφορικά με την ποσοτική περιγραφή φυσικών φαινομένων. Σε ένα πείραμα κατά το οποίο μετρώνται συγχρόνως, η πίεση, η θερμοκρασία και ο όγκος ενός αερίου, τα πειραματικά δεδομένα θα χρησιμοποιηθούν για να ελεγχθεί ο ποσοτικός νόμος που περιγράφεται από την καταστατική εξίσωση των αερίων ή για να υπολογισθεί η τιμή των φυσικών σταθερών που υπεισέρχονται στην ποσοτική περιγραφή του φαινομένου. Τα πειραματικά δεδομένα εμπεριέχουν την απαραίτητη πληροφορία για τον έλεγχο υποθέσεων (θεωρητικών προτύπων) ή/και για την εκτίμηση της τιμής φυσικών σταθερών. Συνεπώς, η αντίστοιχη πληροφορία θα πρέπει να εξαχθεί από την ανάλυση των πειραματικών δεδομένων. Πριν ορίσουμε αυστηρά την εξάρτηση της πληροφορίας από τις διαθέσιμες παρατηρήσεις (μετρήσεις) ας επιχειρήσουμε μία περιγραφή της πληροφορίας, καταγράφοντας τις ιδιότητες που θα θέλαμε να την χαρακτηρίζουν.

- Η πληροφορία που εμπεριέχεται σε ένα σύνολο παρατηρήσεων πρέπει να αυξάνει με το πλήθος των παρατηρήσεων.
- Δεδομένα που είναι άσχετα προς την υπόθεση που εξετάζουμε δεν περιέχουν πληροφορία.
- Η πληροφορία πρέπει να σχετίζεται με την ακρίβεια. Όσο περισσότερη είναι η πληροφορία που συλλέγεται, τόσο μεγαλύτερη θα πρέπει να είναι και η ακρίβεια του πειράματος.

Έστω $\mathbf{x} = \{x_1, x_2, K, x_p\}$ ένα σύνολο p τυχαίων μεταβλητών και η κοινή συνάρτηση πυκνότητας πιθανότητας $f(\mathbf{x}; \boldsymbol{\theta})$, όπου οι συνιστώσες του διανύσματος $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \dots, \theta_k\}$ εκφράζει τις k παραμέτρους¹ που απαιτούνται για την συναρτησιακή έκφραση της πυκνότητας πιθανότητας.

Παράδειγμα 5.1.1:

Ας υποθέσουμε ότι $\mathbf{x} = \{\delta\}$ παριστά την απόκλιση του βέλους από το κέντρο του στόχου, σε έναν αγώνα τοξοβολίας (προφανώς σ' αυτή την περίπτωση $p=1$). Έστω δε ακόμη ότι η συνάρτηση πυκνότητας πιθανότητας για την ποσότητα (τυχαία μεταβλητή) δ είναι η κανονική συνάρτηση:

$$f(\mathbf{x}; \boldsymbol{\theta}) = f(\delta; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi} \cdot \theta_2} e^{-\frac{(\delta - \theta_1)^2}{2\theta_2^2}}$$

Σ' αυτό το παράδειγμα οι παράμετροι θ_1 και θ_2 είναι η μέση τιμή και το σίγμα της συνάρτησης Gauss, αντίστοιχα.

¹ Για παράδειγμα, εάν η συνάρτηση κατανομής $f(\mathbf{x}; \boldsymbol{\theta})$ ήταν κανονική συνάρτηση (Gauss), το διάνυσμα $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \dots, \theta_k\}$ περιέχει τις μέσες τιμές των τυχαίων μεταβλητών x_1, x_2, K, x_p και τα στοιχεία του πίνακα συνδιασποράς τους.

Η κοινή συνάρτηση κατανομής πιθανότητας, σύμφωνα με την 3.3.2, θα είναι:

$$\mathbf{r}_X = \{X_1, X_2, X_3, X_4\}, \quad \mathbf{r}_\mu = \{\mu_1, \mu_2, \mu_3, \mu_4\}$$

$$\mathbf{V}_0 = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{V}_{13} & \mathbf{V}_{14} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{V}_{23} & \mathbf{V}_{24} \\ \mathbf{V}_{31} & \mathbf{V}_{32} & \mathbf{V}_{33} & \mathbf{V}_{34} \\ \mathbf{V}_{41} & \mathbf{V}_{42} & \mathbf{V}_{43} & \mathbf{V}_{44} \end{bmatrix}, \quad \mathbf{V}_{ij} = \mathbf{V}_{ji} = \text{cov}(\mathbf{x}_i, \mathbf{x}_j), \quad \mathbf{V}_{ii} = \sigma_i^2 = \mathbf{V}(\mathbf{x}_i)$$

Ας υποθέσουμε ότι σε ένα πείραμα (π.χ. σκέδαση Compton φωτονίου), μετρούμε συγχρόνως το σύνολο των p φυσικών ποσοτήτων (την γωνία, ϕ_c , μεταξύ των διευθύνσεων του προσπίπτοντος και σκεδαζομένου φωτονίου και την ενέργεια, E_c , του σκεδαζομένου φωτονίου: $p=2$), x_1, x_2, K, x_p . Επαναλαμβάνοντας το πείραμα N φορές (π.χ. μετρώντας την γωνία σκέδασης και την ενέργεια του σκεδαζομένου φωτονίου σε N σκεδάσεις Compton με τις ίδιες αρχικές συνθήκες και την ίδια πειραματική συσκευή), συλλέγουμε τα ακόλουθα πειραματικά δεδομένα:

$$\mathbf{r}_{\mathbf{x}_2} = \{\mathbf{x}_{21}, \mathbf{x}_{22}, \mathbf{x}_{23}, \mathbf{K}, \mathbf{x}_{2p}\}$$

K K K K K K K K K K 5.1.1

$$\mathbf{r}_{\mathbf{x}_N} = \{\mathbf{x}_{N1}, \mathbf{x}_{N2}, \mathbf{x}_{N3}, \mathbf{K}, \mathbf{x}_{Np}\}$$

² Στην υποενότητα 4.1 συζητήσαμε τους λόγους για τους οποίους οι πειραματικές μετρήσεις «συμπεριφέρονται» ως τυχαίες μεταβλητές, ακόμα και στην περίπτωση «ιδανικών» μετρητικών συσκευών.

$$L(\vec{x}; \vec{\theta}) = L(\vec{x}_1, \vec{x}_2, \vec{x}_3, K, \vec{x}_N; \vec{\theta}) = \prod_{i=1}^N f(\vec{x}_i; \vec{\theta}) \quad 5.1.2$$

Συνηθίζεται, η συνάρτηση $L(\vec{x}; \vec{\theta})$ να καλείται συνάρτηση πιθανοφάνειας (Likelihood function) και θεωρείται μόνο ως συνάρτηση των παραμέτρων $\vec{\theta} = \{\theta_1, \theta_2, \theta_3, \dots, \theta_k\}$, διότι οι μεταβλητές $\vec{x}_i$ έχουν ήδη αποκτήσει συγκεκριμένες τιμές (αυτές που καταγράφηκαν κατά την πειραματική διαδικασία). Η πιθανοφάνεια, ως συνάρτηση μόνο των παραμέτρων, αντιστοιχεί στην πιθανότητα: οι πειραματικές παρατηρήσεις, $\vec{x}_1, \vec{x}_2, \vec{x}_3, K, \vec{x}_N$, σε N επαναλήψεις του ίδιου πειράματος, να περιγράφονται από την 5.1.2, με τιμές των παραμέτρων ίσες με $\theta_1, \theta_2, \theta_3, \dots, \theta_k$. Συνεπώς, η πιθανοφάνεια εμπεριέχει πληροφορία αναφορικά με τις τιμές των παραμέτρων.

Εν τούτοις, για τις ανάγκες αυτής της ενότητας θα θεωρήσουμε την πιθανοφάνεια ως συνάρτηση των παραμέτρων $\vec{\theta}$ και των τυχαίων μεταβλητών $\vec{x}_1, \vec{x}_2, \vec{x}_3, K, \vec{x}_N$. Με τον τρόπο αυτό η συνάρτηση 5.1.2 αποκτά ευρύτερο περιεχόμενο. Επί παραδείγματι, θεωρήσετε ένα μεγάλο πλήθος πειραματιστών που εκπονούν, εντελώς ανεξάρτητα, ακριβώς την ίδια πειραματική διαδικασία συλλέγοντας N μετρήσεις ο κάθε ένας. Σ' αυτή την περίπτωση, η πιθανοφάνεια 5.1.2 εκφράζει την πιθανότητα: ένας πειραματιστής να συλλέξει τις μετρήσεις $\vec{x}_1, \vec{x}_2, \vec{x}_3, K, \vec{x}_N$ όταν οι τιμές των παραμέτρων είναι ίσες με $\theta_1, \theta_2, \theta_3, \dots, \theta_k$.

5.2 Πληροφορία σύμφωνα με τον R.A. Fisher

Για απλούστευση, ας υποθέσουμε ότι η συνάρτηση πυκνότητας πιθανότητας των τυχαίων μεταβλητών $\vec{x} = \{x_1, x_2, x_3, x_4\}$ εξαρτάται από μόνο μία παράμετρο: $\vec{\theta} = \{\theta\}$ και $f(\vec{x}; \vec{\theta}) = f(\vec{x}; \theta)$.

Η ποσότητα πληροφορίας, σχετικά με την παράμετρο θ , που περιέχεται **σε μία** παρατήρηση ορίζεται, σύμφωνα με τον R.A. Fisher, με την ακόλουθη έκφραση (εάν βέβαια υπάρχει):

$$I(\theta) = E \left[\left(\frac{\partial \ln L(\vec{x}; \theta)}{\partial \theta} \right)^2 \right] \quad 5.2.1$$

$$= \int_{\Omega_\theta} \left(\frac{\partial \ln L(\vec{x}; \theta)}{\partial \theta} \right)^2 \cdot L(\vec{x}; \theta) dx_1 dx_2 \dots dx_p$$

όπου το $\vec{x} = \{x_1, x_2, x_3, x_4\}$ αντιπροσωπεύει το σύνολο των πειραματικών δεδομένων που αντιστοιχεί σε μία παρατήρηση (μία εκτέλεση του πειράματος) και με το σύμβολο Ω_θ δηλώνεται το σύνολο όλων των επιτρεπτών τιμών για τα αποτελέσματα των μετρήσεων $x_1, x_2, \dots, x_p$. Εν γένει το πεδίο τιμών των x_i μπορεί να εξαρτάται από την τιμή της παραμέτρου θ .

Παράδειγμα 5.2.1:

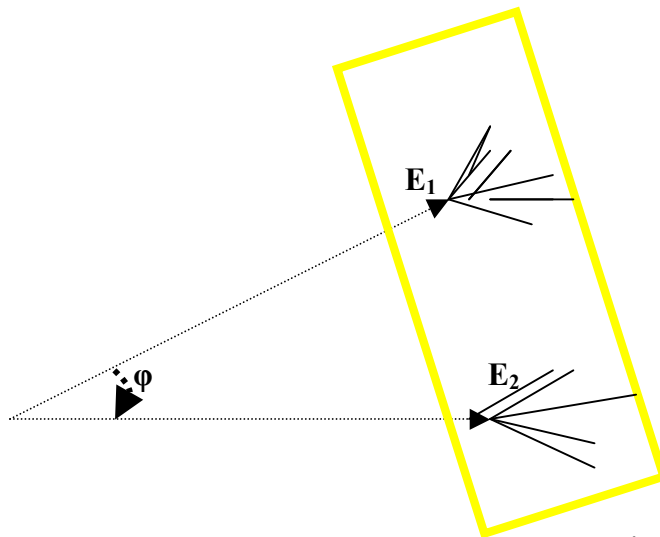
Σ' ένα πείραμα, για τον πειραματικό προσδιορισμό της μάζας του ουδέτερου πιονίου, π^0 , μελετάται η διάσπασή του σε δύο φωτόνια, $\pi^0 \rightarrow \gamma + \gamma$. Ουδέτερα πιόνια παράγονται κατά την πρόσπτωση φορτισμένων σωματιδίων σε (πρακτικά) σημειακό στόχο και διασπώνται, σχεδόν αμέσως, σε δύο φωτόνια.

Στη συνέχεια, τα θυγατρικά φωτόνια προσπίπτουν σε ειδική μετρητική διάταξη (ηλεκτρομαγνητικό θερμιδόμετρο) η οποία μετρά αφ' ενός μεν την ενέργεια τους, E_1 και E_2 αντίστοιχα, αφ' ετέρου την διεύθυνση κίνηση των φωτονίων και κατ' επέκταση την γωνία, φ , μεταξύ των τροχιών τους (βλέπε Σχήμα 5.2.1). Η μάζα, m , του ουδέτερου πιονίου υπολογίζεται (σύμφωνα με την ειδική σχετικότητα) ως:

$$m = 2E_1E_2(1-\cos\varphi) \quad 5.2.2$$

Η πειραματική εκτίμηση της μάζας, προφανώς, πάσχει από στατιστικό σφάλμα, λόγω της περιορισμένης διακριτικής ικανότητας της πειραματικής διάταξης στην μέτρηση των ενεργειών και των διευθύνσεων των φωτονίων.

Ας υποθέσουμε ότι η πυκνότητα πιθανότητας της εκτιμώμενης τιμής της μάζας, m , εκφράζεται από κανονική συνάρτηση (σε αυτό συνηγορεί και το θεώρημα του Κεντρικού Ορίου), με μέση τιμή ίση με τη πραγματική τιμή της μάζας του π^0 , m_{π^0} .



Σχήμα 4.2.1. Σχηματική παράσταση της πειραματικής μελέτης της διάσπασης $\pi^0 \rightarrow \gamma\gamma$

Σ' αυτή την περίπτωση, θα μπορούσαμε να θεωρήσουμε γνωστή (από τα αποτελέσματα άλλων πειραμάτων) την τιμή της πραγματικής μάζας και να μελετήσουμε την διακριτική ικανότητα της πειραματικής συσκευής αναφορικά με την μέτρηση μαζών³.

$$f(\vec{x};\vec{\theta}) = f(m;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(m - m_{\pi^0})^2}{2\sigma^2}\right] \quad 5.2.3$$

Παρατηρήσετε ότι το πεδίο τιμών της μέτρησης της μάζας, $0 \leq \Omega \leq \infty$, είναι ανεξάρτητο της παραμέτρου σ .

³ Πράγματι, επειδή η μάζα του ουδέτερου πιονίου είναι γνωστή με μεγάλη ακρίβεια, η μέθοδος αυτή χρησιμοποιείται στην αξιολόγηση της διακριτικής ικανότητας των ηλεκτρομαγνητικών θερμιδομέτρων, σε πειράματα σωματιδιακής φυσικής.

Επισημαίνεται, ότι το ουδέτερο πιόνιο, όπως όλα τα σωματίδια του μικρόκοσμου, δεν έχει αυστηρά καθορισμένη μάζα. Η κβαντομηχανική προβλέπει την συνάρτηση πυκνότητα πιθανότητας για την μάζα 5.2.2, έστω και στην περίπτωση ιδανικών ανιχνευτικών διατάξεων. Επειδή όμως το φυσικό εύρος μαζών είναι κατά πολύ μικρότερο από τα ανιχνευτικά σφάλματα, για κάθε πρακτική εφαρμογή, ταυτίζουμε την μέση τιμή μαζών του σωματίου με την «πραγματική» τιμή της μάζας του.

Για μία μόνο παρατήρηση, η πιθανοφάνεια και η πληροφορία αναφορικά με την διακριτική ικανότητα της διάταξης έχουν ως εξής:

$$L(x;\theta) = L(m_1;\sigma) = f(m_1;\sigma)$$

$$I_{\sigma}(\sigma) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln(f(m_1|\sigma))}{\partial \sigma} \right)^2 \cdot f(m_1;\sigma) dm_1 \quad 5.2.4$$

Γενικά, εάν η συνάρτηση πυκνότητας πιθανότητας εξαρτάται από k παραμέτρους, η πληροφορία ορίζεται ως εξής:

$$\begin{aligned} [I(\theta)]_{ij} &= E \left[\frac{\partial \ln L(\mathbf{x};\theta)}{\partial \theta_i} \cdot \frac{\partial \ln L(\mathbf{x};\theta)}{\partial \theta_j} \right] \\ &= \int_{\Omega_{\theta}} \left[\frac{\partial \ln L(\mathbf{x};\theta)}{\partial \theta_i} \cdot \frac{\partial \ln L(\mathbf{x};\theta)}{\partial \theta_j} \right] \cdot L(\mathbf{x};\theta) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3 \dots d\mathbf{x}_p \end{aligned} \quad 5.2.5$$

Στη γενική της μορφή, η έκφραση 5.2.5 για την πληροφορία αποτελεί ένα πίνακα $k \times k$ διαστάσεων, όπου τα διαγώνια στοιχεία ($[I(\theta)]_{ii}$ $i=1,2,\dots,k$) εμπεριέχουν την πληροφορία για κάθε μία από τις παραμέτρους, $\theta_1, \theta_2, \theta_3, \dots, \theta_k$, ενώ τα μη διαγώνια στοιχεία εκφράζουν την πληροφορία για τον συσχετισμό μεταξύ των παραμέτρων.

Επισημαίνεται ότι, ο φαινομενικά αυθαίρετος, ορισμός της πληροφορίας, 5.2.5, προσφέρει δύο βασικά πλεονεκτήματα:

- Χαρακτηρίζεται από τις ιδιότητες που απαιτούνται από την πληροφορία.
- Σχετίζεται στενά (όπως θα δούμε στα επόμενα) με την μέγιστη δυνατή ακρίβεια εκτίμησης της τιμής των παραμέτρων.

5.3 Ιδιότητες της πληροφορίας.

Η πληροφορία, όπως ορίστηκε με την σχέση 5.2.5, αποκτά μεγαλύτερη ευχρηστία όταν ικανοποιούνται οι ακόλουθες συνθήκες-απαιτήσεις:

- Το πεδίο τιμών των τυχαίων μεταβλητών (μετρήσεων), Ω_{θ} , να είναι ανεξάρτητο των παραμέτρων, $\theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_k\}$.
- Η πιθανοφάνεια εκφράζεται από αρκετά «ομαλή» συνάρτηση, , ώστε να επιτρέπει την αμοιβαία μετάθεση των τελεστών: $\partial^2 / \partial \theta_1 \partial \theta_2$ και $\int dx$

Υποθέτοντας ότι οι παραπάνω συνθήκες ικανοποιούνται, μετά από απλές πράξεις, καταλήγουμε:

$$\begin{aligned}
\int_{\Omega} L(\vec{x}; \vec{\theta}) d\vec{x}_1 d\vec{x}_2 d\vec{x}_3 \dots d\vec{x}_N &= \int_{\Omega} L(\vec{x}; \vec{\theta}) d\vec{x} = 1 \\
\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \int_{\Omega} L(\vec{x}; \vec{\theta}) d\vec{x} &= \int_{\Omega} \frac{\partial^2 L(\vec{x}; \vec{\theta})}{\partial \theta_1 \partial \theta_2} d\vec{x} = 0 \\
\int_{\Omega} \frac{\partial}{\partial \theta_1} \left[\frac{\partial L(\vec{x}; \vec{\theta})}{\partial \theta_2} \right] d\vec{x} &= \int_{\Omega} \frac{\partial}{\partial \theta_1} \left[L(\vec{x}; \vec{\theta}) \cdot \frac{\partial \ln L(\vec{x}; \vec{\theta})}{\partial \theta_2} \right] d\vec{x} = 0
\end{aligned} \tag{5.3.1}$$

Συνεπώς:

$$\begin{aligned}
\int_{\Omega} \frac{\partial L(\vec{x}; \vec{\theta})}{\partial \theta_1} \cdot \frac{\partial \ln L(\vec{x}; \vec{\theta})}{\partial \theta_2} d\vec{x} &= - \int_{\Omega} L(\vec{x}; \vec{\theta}) \cdot \frac{\partial^2 \ln L(\vec{x}; \vec{\theta})}{\partial \theta_1 \partial \theta_2} d\vec{x} \\
\int_{\Omega} \frac{\partial \ln L(\vec{x}; \vec{\theta})}{\partial \theta_1} \cdot \frac{\partial \ln L(\vec{x}; \vec{\theta})}{\partial \theta_2} \cdot L(\vec{x}; \vec{\theta}) d\vec{x} &= - \int_{\Omega} L(\vec{x}; \vec{\theta}) \cdot \frac{\partial^2 \ln L(\vec{x}; \vec{\theta})}{\partial \theta_1 \partial \theta_2} d\vec{x}
\end{aligned} \tag{5.3.2}$$

Χρησιμοποιώντας την σχέση 5.3.2, η πληροφορία εκφράζεται ως:

$$\begin{aligned}
\left[I(\vec{\theta}) \right]_{12} &= \int_{\Omega} \frac{\partial \ln L(\vec{x}; \vec{\theta})}{\partial \theta_1} \cdot \frac{\partial \ln L(\vec{x}; \vec{\theta})}{\partial \theta_2} \cdot L(\vec{x}; \vec{\theta}) d\vec{x} \\
&= - \int_{\Omega} \frac{\partial^2 \ln L(\vec{x}; \vec{\theta})}{\partial \theta_1 \partial \theta_2} \cdot L(\vec{x}; \vec{\theta}) d\vec{x} \\
&= -E \left[\frac{\partial^2 \ln L(\vec{x}; \vec{\theta})}{\partial \theta_1 \partial \theta_2} \right]
\end{aligned} \tag{5.3.3}$$

Στα επόμενα θα δείξουμε την εξάρτηση της πληροφορίας από το πλήθος των παρατηρήσεων. Χρησιμοποιώντας την 5.1.2, η πληροφορία, για μία μόνο παρατήρηση, εκφράζεται ως:

$$I_1(\vec{\theta})_{ij} = \int_{\Omega} \frac{\partial^2 \ln L(\vec{x}; \vec{\theta})}{\partial \theta_i \partial \theta_j} \cdot L(\vec{x}; \vec{\theta}) d\vec{x} = \int_{\Omega} \frac{\partial^2 \ln f(\vec{x}_1; \vec{\theta})}{\partial \theta_i \partial \theta_j} \cdot f(\vec{x}_1; \vec{\theta}) d\vec{x}_1 \tag{5.3.4}$$

Όπου $\vec{x}_1 = \{x_{11}, x_{12}, x_{13}, \dots, x_{1p}\}$ αναφέρεται στα αποτελέσματα του μοναδικού πειράματος και με το υπόθεμα “1” στον συμβολισμό της πληροφορίας ($I_1(\vec{\theta})_{ij}$) δηλώνεται ότι συλλέχθηκε από μία μία παρατήρηση .

Κατ’ αναλογία, πληροφορία που μεταφέρουν N ανεξάρτητες παρατηρήσεις ($L(\vec{x}; \vec{\theta}) = \prod_{i=1}^N f(\vec{x}_i; \vec{\theta})$) εκφράζεται ως:

$$\begin{aligned}
I_N(\theta)_{ij} &= \int_{\Omega} \int_{\Omega} \frac{\partial^2 \ln L(\bar{x}; \theta)}{\partial \theta_i \partial \theta_j} \cdot L(\bar{x}; \theta) d\bar{x}_1 d\bar{x}_2 \dots d\bar{x}_N \\
&= \int_{\Omega} \int_{\Omega} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \sum_r \ln f(\bar{x}_r; \theta) \right\} \cdot \left\{ \prod_{m=1}^N f(\bar{x}_m; \theta) \right\} d\bar{x}_1 d\bar{x}_2 \dots d\bar{x}_N \\
&= \sum_{r=1}^N \left\{ \int_{\Omega} \frac{\partial^2 \ln f(\bar{x}_r; \theta)}{\partial \theta_i \partial \theta_j} \cdot f(\bar{x}_r; \theta) d\bar{x}_r \right\} \cdot \int_{\Omega} \int_{\Omega} \prod_{m=1}^N f(\bar{x}_m; \theta) d\bar{x}_1 d\bar{x}_2 \dots d\bar{x}_N
\end{aligned} \quad 5.3.5$$

όπου με το υπόθεμα “N” στον συμβολισμό της πληροφορίας ($I_N(\theta)_{ij}$) δηλώνεται ότι αναφέρεται σε N παρατηρήσεις. Ο όρος $\int_{\Omega} \frac{\partial^2 \ln f(\bar{x}_r; \theta)}{\partial \theta_i \partial \theta_j} \cdot f(\bar{x}_r; \theta) d\bar{x}_r$ στην 5.3.5 εκφράζει την πληροφορία που μεταφέρει μία μόνο παρατήρηση (η υπ’ αριθμόν r) η οποία θα πρέπει να είναι ίση με την πληροφορία που μεταφέρει κάθε άλλη παρατήρηση ($= I_1(\theta)_{ij}$). Συνεπώς, η 5.3.5 καταλήγει στην σχέση:

$$I_N(\theta)_{ij} = \sum_{r=1}^N I_1(\theta)_{ij} = N \cdot I_1(\theta)_{ij} \quad 5.3.6$$

ότι δηλαδή: η πληροφορία που μεταφέρεται από N πανομοιότυπες παρατηρήσεις (πειράματα, μετρήσεις...), σχετικά με την τιμή των παραμέτρων $\theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_k\}$, είναι N φορές πολλαπλάσια της πληροφορίας που εμπεριέχεται σε μία μόνο παρατήρηση.

Παράδειγμα 5.3.1:

Έστω η τυχαία μεταβλητή x, οι τιμές της οποίας κατανέμονται σύμφωνα με κανονική συνάρτηση πυκνότητας πιθανότητας: $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$, $x \in \mathbb{R}$. Ας υποθέσουμε ότι η τυχαία μεταβλητή x αναφέρεται στο αποτέλεσμα της μέτρησης κάποιας φυσικής ποσότητας της οποίας η πραγματική τιμή ταυτίζεται με την μέση τιμή, μ , της συνάρτησης πυκνότητας πιθανότητας. Η άλλη παράμετρος, σ , εξαρτάται από την διακριτική ικανότητα του οργάνου.

Το αποτέλεσμα κάθε μέτρησης εμπεριέχει πληροφορία αναφορικά με την παράμετρο μ . Η πληροφορία από μία μοναδική παρατήρηση x είναι:

$$\begin{aligned}
I_1(\mu) &= - \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \mu^2} \left[-\frac{(x-\mu)^2}{2\sigma^2} - \ln \sigma \sqrt{2\pi} \right] \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma^2} \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \frac{1}{\sigma^2}
\end{aligned} \quad 5.3.7$$

Η πληροφορία, όσον αφορά στην τιμή του μ , που περιέχεται σε N παρατηρήσεις είναι:

$$I_N(\mu) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \mu^2} \left[\sum_{i=1}^N -\frac{(x_i - \mu)^2}{2\sigma^2} - N \ln \sigma \sqrt{2\pi} \right] \cdot \prod_r \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x_r - \mu)^2}{2\sigma^2} \right] dx_1 \dots dx_N$$

$$= \frac{N}{\sigma^2}$$

όπως αναμένεται από την σχέση 5.3.6

Στις επόμενες υποενότητες θα συζητήσουμε ένα σημαντικό θεώρημα (την ανισότητα των Cramer-Rao) που δηλώνει ότι η τετραγωνική διασπορά μίας εκτίμησης (που δεν πάσχει από στατιστική πόλωση - σφάλμα) της τιμής μίας παραμέτρου είναι μεγαλύτερη ή ίση προς το αντίστροφο της πληροφορίας. Συνεπώς, στην περίπτωση αυτού του παραδείγματος,

$$V(\hat{\mu}) \geq \frac{\sigma^2}{N}$$

Υπενθυμίζεται ότι έχουμε αποδείξει ότι η τετραγωνική απόκλιση του μέσου όρου N τιμών της τυχαίας μεταβλητής x ($\langle x \rangle = \sum x_i / N$), ισούται με $V(x)/N$.

Παράδειγμα 5.3.2:

Ας υποθεθεί ότι θέλουμε να ζυγίσουμε τέσσερα αντικείμενα με μία ατελή ζυγαριά. Ας υποθέσουμε ακόμα ότι η ακρίβεια ζύγισης είναι ανεξάρτητη από το βάρος.

- Η απλούστερη διαδικασία θα ήταν να ζυγίσουμε κάθε αντικείμενο χωριστά και να επαναλάβουμε τις μετρήσεις αρκετές φορές για να αυξήσουμε την ακρίβεια. Θυμηθείτε ότι $I_N(\theta) = N I_1(\theta)$
- Ένας άλλος τρόπος είναι να ζυγίζουμε συνδυασμούς αντικειμένων

$$x_i = \sum_{j=1}^4 a_{ij} y_j \quad 5.3.7$$

όπου y_j , είναι το βάρος του j αντικειμένου ($j=1,2,3,4$), και x_i είναι τα βάρη των γραμμικών συνδυασμών. Στην περίπτωση αυτή, το αποτέλεσμα της ζύγισης αναφέρεται στο βάρος που θα πρέπει να τοποθετηθεί στο δεξιό σκέλος της ζυγαριάς για να ισορροπήσει. Οι συντελεστές a_{ij} μπορούν να παίρνουν τις ακόλουθες τιμές:

- $a_{ij} =$
- 1 όταν το αντικείμενο j βρίσκεται στην αριστερή πλευρά της ζυγαριάς όταν ζυγίζουμε τον συνδυασμό i .
 - 0 όταν είναι το j αντικείμενο είναι εκτός του συνδυασμού i
 - 1 όταν το αντικείμενο j βρίσκεται στην δεξιά πλευρά της ζυγαριάς όταν ζυγίζουμε τον συνδυασμό i .

Προφανώς εάν επιλέξουμε $a_{ij} = \delta_{ij}$ τότε $x_i = y_i$ και οι συνδυασμοί εκφυλίζονται σε ζυγίσεις καθ' ενός από τα αντικείμενα.

Έστω x_i^m είναι το αποτέλεσμα μίας ζύγισης του συνδυασμού i .

Προφανώς το x_i^m είναι τυχαία μεταβλητή και θα υποθέσουμε ότι η συνάρτηση πυκνότητας πιθανότητας του αποτελέσματος της ζύγισης εκφράζεται από την ακόλουθη κανονική συνάρτηση:

$$f(x_i^m | y_1, y_2, y_3, y_4, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{\left(x_i^m - \sum_j a_{ij} y_j \right)^2}{2\sigma^2} \right] \quad 5.3.8$$

όπου η παράμετρος σ εξαρτάται από την μετρητική συσκευή και σύμφωνα με τις αρχικές υποθέσεις είναι ανεξάρτητη από το βάρος των αντικειμένων που ζυγίζονται.

Εάν εκτελέσουμε τέσσερις ανεξάρτητες ζυγίσεις, τεσσάρων ανεξάρτητων γραμμικών συνδυασμών, η πιθανοφάνεια (likelihood) των μετρήσεων θα είναι

$$L(x_1^m, x_2^m, x_3^m, x_4^m; y_1, y_2, y_3, y_4, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^4} \exp \left(\sum_{i=1}^4 \left[-\frac{\left(x_i^m - \sum_j a_{ij} y_j \right)^2}{2\sigma^2} \right] \right) \quad 5.3.9$$

Επειδή το πεδίο δυνατών τιμών των μετρήσεων, Ω , δεν εξαρτάται από τις παραμέτρους και επιπλέον $\partial^2 / \partial y_1 \partial y_2$ μετατίθεται με το $\int dx^m$ η πληροφορία σχετικά με τα βάρη των αντικειμένων εκφράζεται ως:

$$I(y_1, y_2, y_3, y_4, \sigma^2)_{lk} = \int \frac{\partial^2 \ln L(x_1^m, x_2^m, x_3^m, x_4^m; y_1, y_2, y_3, y_4, \sigma)}{\partial y_l \partial y_k} \cdot L(x_1^m, x_2^m, x_3^m, x_4^m; y_1, y_2, y_3, y_4, \sigma) dx_1^m dx_2^m dx_3^m dx_4^m$$

Μετά από απλές πράξεις η πληροφορία εκφράζεται ως:

$$I(y_1, y_2, y_3, y_4, \sigma^2) = \frac{1}{\sigma^2} A^T A \quad 5.3.10$$

όπου οι πίνακες $I(y_1, y_2, y_3, y_4, \sigma^2)$ και A έχουν στοιχεία τις ποσότητες $I(y_1, y_2, y_3, y_4, \sigma^2)_{ij}$ και a_{ij} αντίστοιχα.

Εάν επιλέξουμε την πρώτη μέθοδο ζύγισης, τότε:

$$I(y_1, y_2, y_3, y_4, \sigma^2)_{ij} = \delta_{ij} / \sigma^2 \quad 5.3.11$$

Αλλά εάν επιλέξουμε την δεύτερη μέθοδο μπορούμε να αναζητήσουμε συνδυασμούς που να μεγιστοποιούν το γινόμενο $a^T a$, της έκφραση 5.3.10

Όπως είναι εύκολο να αποδειχθεί ο πίνακας συνδυασμών:

$$A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad 5.3.12$$

αντιστοιχεί σε πληροφορία, αναφορικά με τα αληθή βάρη, ίση με $4\delta_{ij}/\sigma^2$, δηλαδή τέσσαρις φορές περισσότερη από αυτήν που εμπεριέχεται στα αποτελέσματα της πρώτης (5.3.11) ζύγισης.

5.4 Η Στατιστική συνάρτηση (Statistic) - Υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας

Έστω $x_1, x_2, x_3, \dots, x_N$ είναι ένα σύνολο N τυχαίων μεταβλητών και ότι y είναι μία άλλη τυχαία μεταβλητή ορισμένη σαν συνάρτηση των μεταβλητών x_i ($i=1,2,3,\dots,N$):

$$T = \psi(x_1, x_2, x_3, \dots, x_N) \quad 5.4.1.$$

Κάθε τέτοια συνάρτηση T λέγεται στατιστική συνάρτηση⁴.

Έστω $f(x_1, x_2, x_3, \dots, x_N)$ είναι η κοινή συνάρτηση πυκνότητας πιθανότητας των N τυχαίων μεταβλητών. Επί πλέον, έστω $T = \phi(x_1, x_2, x_3, \dots, x_N)$ είναι μία στατιστική συνάρτηση των N μεταβλητών. Η κανονικοποιημένη συνάρτηση

$$g(x_1, x_2, x_3, \dots, x_N | T) = \frac{f(x_1, x_2, x_3, \dots, x_N)}{R} \quad 5.4.2$$

$T = \text{σταθερά}$

με πεδίο ορισμό τέτοιο ώστε η στατιστική συνάρτηση $T = \phi(x_1, x_2, x_3, \dots, x_N)$ να ισούται με σταθερή τιμή, εκφράζει την **υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας των $x_1, x_2, x_3, \dots, x_N$ δοθείσης της τιμής του T .**

Προφανώς πρόκειται για μία γενίκευση της υπό συνθήκη πιθανότητας.

Παράδειγμα 5.4.1:

Έστω N τυχαίες μεταβλητές, $x_1, x_2, x_3, \dots, x_N$. Κάθε μία από αυτές τις μεταβλητές παίρνει τιμές σύμφωνα με την κανονική συνάρτηση πυκνότητας πιθανότητας :

$$f(x_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right] \quad i = 1, 2, \dots, N \quad 5.4.3$$

Ας θεωρήσουμε την στατιστική συνάρτηση:

$$\langle x \rangle = \frac{\sum_{i=1}^N x_i}{N} \quad 5.4.4$$

Προφανώς πρόκειται για τον μέσο όρο των τυχαίων μεταβλητών, $x_1, x_2, x_3, \dots, x_N$. Ο μέσος όρος, $\langle x \rangle$, είναι τυχαία μεταβλητή. Αναζητούμε την συνάρτηση πυκνότητας πιθανότητας του $\langle x \rangle$, δοθέντων των παραμέτρων μ και σ .

⁴ Ο κλάδος της επιστήμης που είναι γνωστός σαν "Στατιστική ανάλυση" ή "Statistics" ονομάστηκε έτσι επειδή ασχολείται με την μελέτη των στατιστικών συναρτήσεων (statistics).

Ας υποθέσουμε ότι διαλέγουμε τιμές για κάθε μία από τις τυχαίες μεταβλητές, σύμφωνα με την συνάρτηση πυκνότητας πιθανότητας 5.4.3, και υπολογίζουμε τον μέσο όρο τους. Επαναλαμβάνουμε την διαδικασία επιλογής τιμών k συνολικά φορές και κάθε φορά υπολογίζουμε τον μέσο όρο των επιλεγμένων τιμών των τυχαίων μεταβλητών.

$$\begin{aligned} \{x_1^1, x_2^1, x_3^1, \dots, x_N^1\} &\rightarrow \langle x \rangle^1 = \frac{\sum_{i=1}^N x_i^1}{N} \\ \{x_1^2, x_2^2, x_3^2, \dots, x_N^2\} &\rightarrow \langle x \rangle^2 = \frac{\sum_{i=1}^N x_i^2}{N} \end{aligned} \quad 5.4.5$$

K K K K K K K K K K K K K K K

$$\{x_1^k, x_2^k, x_3^k, \dots, x_N^k\} \rightarrow \langle x \rangle^k = \frac{\sum_{i=1}^N x_i^k}{N}$$

Η συνάρτηση πυκνότητας πιθανότητας επιλογής των τιμών $\bar{x}_1^j, x_2^j, x_3^j, \dots, x_N^j$ εκφράζεται από την πιθανοφάνεια:

$$\begin{aligned} L(x_1^j, x_2^j, x_3^j, \dots, x_N^j; \mu, \sigma) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x_i^j - \mu)^2}{2\sigma^2} \right] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp \left[\sum_{i=1}^N \left(-\frac{(x_i^j - \mu)^2}{2\sigma^2} \right) \right] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp \left[-\frac{N}{2\sigma^2} (\langle x \rangle^j - \mu)^2 \right] \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i^j - \langle x \rangle^j)^2 \right] \end{aligned} \quad 5.4$$

.6

Η συνάρτηση πυκνότητας πιθανότητας, επιλογής των τιμών $\bar{x}_1^j, x_2^j, x_3^j, \dots, x_N^j$, ώστε ο μέσος όρος, $\langle x \rangle^j$, να ισούται με c , εκφράζεται ως:

$$g(x_1, x_2, x_3, \dots, x_N | \langle x \rangle = c) = \frac{\int_1^K \int_2^K \int_3^K L(x_1, x_2, x_3, \dots, x_N; \mu, \sigma) dx_1 dx_2 \dots dx_N}{\int_{-\infty}^{\infty} \left[\int_1^K \int_2^K \int_3^K L(x_1, x_2, x_3, \dots, x_N; \mu, \sigma) dx_1 dx_2 \dots dx_N \right] dc} \quad 5.4.7$$

Ο αριθμητής του δεξιού μέρους της σχέσης 5.4.7, $\int_1^K \int_2^K \int_3^K L(x_1, x_2, x_3, \dots, x_N; \mu, \sigma) dx_1 dx_2 \dots dx_N$

εκφράζεται με την βοήθεια της 5.4.6 ως εξής:

$$\left\{ \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp \left[-\frac{N}{2\sigma^2} (\langle x \rangle - \mu)^2 \right] \cdot \int K \prod_{i=1}^N \exp \left(-\frac{(x_i - \langle x \rangle)^2}{2\sigma^2} \right) dx_1 \dots dx_N \right\}_{\langle x \rangle = c}$$

$$= \exp \left[-\frac{N}{2\sigma^2} (c - \mu)^2 \right]$$

Κατ' ακολουθία, ο παρανομαστής, ο παράγοντας κανονικοποίησης R, θα είναι:

$$R = \int_{-\infty}^{+\infty} \exp \left[-\frac{N}{2\sigma^2} (c - \mu)^2 \right] d\mu = \sqrt{\frac{2\pi}{N}} \sigma$$

Συνεπώς, η συνάρτηση πυκνότητας πιθανότητας, επιλογής τιμών για τις N τυχαίες μεταβλητές (που χαρακτηρίζονται από κανονική συνάρτηση πυκνότητας πιθανότητας, με την ίδια μέση τιμή και σίγμα) υπό την συνθήκη: ο μέσος όρος τους να ισούται με c, εκφράζεται, πολύ απλά, ως:

$$\frac{1}{\sqrt{2\pi/N} \sigma} \exp \left[-\frac{N}{2\sigma^2} (c - \mu)^2 \right]$$

Γενικεύοντας το παραπάνω συμπέρασμα καταλήγουμε ότι η συνάρτηση πυκνότητας πιθανότητας του μέσου όρου των τιμών N τυχαίων μεταβλητών με κανονική συνάρτηση πιθανότητας, δοθέντος του μ και σ, είναι η ακόλουθη κανονική συνάρτηση:

$$g(\langle x \rangle; \mu, \sigma) = \frac{1}{\sqrt{2\pi/N} \sigma} \exp \left[-\frac{N}{2\sigma^2} (\langle x \rangle - \mu)^2 \right] \quad 5.4.8$$

Ερώτηση: Γιατί δεν αρκούσε το Θεώρημα του Κεντρικού Ορίου για την απόδειξη της 5.4.8 ;

Υπόδειξη: Ποίες είναι οι προϋποθέσεις εφαρμογής του Θεωρήματος του Κεντρικού Ορίου;

Ερώτηση: Δείξτε ότι ο μέσος όρος N τιμών της τυχαίας μεταβλητής x, με κανονική συνάρτηση πυκνότητα πιθανότητας, είναι τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας που δίνεται από την 5.4.8.

Υπόδειξη: Οι περιπτώσεις: α) μίας μέτρησης ενός φυσικού μεγέθους που επαναλαμβάνεται υπό τις ίδιες συνθήκες N φορές στο ίδιο πείραμα και β) N διαφορετικών πειραματιστών που εκτελούν μία μέτρηση της ίδιας φυσικής ποσότητας με πανομοιότυπες μετρητικές συσκευές και υπό τις ίδιες συνθήκες, είναι απόλυτα ταυτόσημες.

5.5 Επάρκεια (Sufficiency)

Έστω $\underline{y} = \{y_1, y_2, y_3, \dots, y_k\}$ είναι τυχαίες μεταβλητές με κοινή συνάρτηση πυκνότητας πιθανότητας με κοινή συνάρτηση πυκνότητας πιθανότητας, $f(\underline{y}; \theta) = f(y_1, y_2, y_3, \dots, y_k; \theta)$ και έστω η στατιστική συνάρτηση $T(y_1, y_2, y_3, \dots, y_k)$. Εάν η υπό συνθήκη συνάρτηση πυκνότητας $q(y_1, y_2, y_3, \dots, y_k | T)$ δεν εξαρτάται από τις παραμέτρους θ , τότε η στατιστική συνάρτηση T είναι επαρκής (sufficient) όσον αφορά την παράμετρο θ . Εν άλλους λόγους, η στατιστική συνάρτηση $T(y_1, y_2, y_3, \dots, y_k)$ εμπεριέχει όλη την πληροφορία αναφορικά με την παράμετρο θ .

Ο ορισμός ισχύει και στην περίπτωση πολυπαραμετρικών προβλημάτων ($\theta \rightarrow \underline{\theta}$, $T \rightarrow \underline{T}$)

Ενδιαφερόμαστε ιδιαίτερα για περιπτώσεις όπου οι τυχαίες μεταβλητές αντιστοιχούν παρατηρήσεις (μετρήσεις). Έστω λοιπόν οι N ανεξάρτητες παρατηρήσεις, $\underline{x} = \{x_1, x_2, x_3, \dots, x_N\}$ με κοινή συνάρτηση πυκνότητας πιθανότητας που εκφράζεται από την συνάρτηση πιθανοφάνειας $L(\underline{x}; \theta)$. Ποια θα είναι η συνθήκη που θα εξασφαλίζει ότι η στατιστική συνάρτηση $t(x_1, x_2, x_3, \dots, x_N)$ είναι επαρκής όσον αφορά στην παράμετρο θ ;

Παράδειγμα 5.5.1

Έστω $\bar{x}_1, x_2, x_3, \dots, x_N$ είναι N τιμές της τυχαίας μεταβλητής, x , της οποίας η συνάρτηση πυκνότητας πιθανότητας είναι η κανονική συνάρτηση $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$.

Επί παραδείγματι, έστω ότι $x_1, x_2, x_3, \dots, x_N$ είναι τα αποτελέσματα N επαναλήψεων της ίδιας μέτρησης. Ας υποθεθεί ακόμα ότι η τιμή της παραμέτρου σ είναι γνωστή, ώστε να θεωρούμε το σ ως σταθερά και όχι ως παράμετρο. Σ' αυτή την περίπτωση η πιθανοφάνεια είναι συνάρτηση μόνο της τυχαίας μεταβλητής x και της παραμέτρου, μ , που εκφράζει τη μέση τιμή.

Προφανώς το σύνολο των N παρατηρήσεων, αυτό καθ' εαυτό, περιλαμβάνει την μέγιστη διαθέσιμη πληροφορία αναφορικά με το μ . Υπάρχει όμως, μία και μόνο στατιστική συνάρτηση, η οποία περιέχει όλη την διαθέσιμη πληροφορία σχετικά με την μέση τιμή. Αυτή η στατιστική συνάρτηση είναι ο μέσος όρος:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i \quad 5.5.1$$

Στην υποενότητα 5.4 δείξαμε ότι πιθανοφάνεια ενός συνόλου N τυχαίων μεταβλητών⁵, με την ίδια κανονική συνάρτηση πυκνότητας πιθανότητας 5.4.3, εκφράζεται ως:

$$L(x_1, x_2, x_3, \dots, x_N; \mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left[-\frac{N}{2\sigma^2} (\langle x \rangle - \mu)^2\right] \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \langle x \rangle)^2\right]$$

και ότι η έκφραση

⁵ Η αντίστοιχα, η πιθανοφάνεια N τιμών της ίδιας τυχαίας μεταβλητής, με κανονική συνάρτηση πυκνότητας πιθανότητας.

$$g(< x > ; \mu) = \left(\frac{N}{2\pi \sigma^2} \right)^{1/2} \exp \left[-\frac{N}{2\sigma^2} \cdot (< x > - \mu)^2 \right] \quad 5.5.3$$

είναι η πυκνότητα πιθανότητας του μέσου όρου $<x>$ δοθέντος⁶ του μ .

Χρησιμοποιώντας τις εκφράσεις 5.5.2 και 4.5.2, η πιθανοφάνεια μπορεί να γραφεί ως το γινόμενο της υπό συνθήκης πιθανότητας, $g(< x > | \mu)$, και κάποιας άλλης συνάρτησης που δεν εξαρτάται από την παράμετρο μ .

$$L(x_1, x_2, x_3, K, x_N; \mu) = g(< x > | \mu) \cdot h(x) \quad 5.5.4$$

Η σχέση 5.5.4 εκφράζει την αναγκαία και ικανή συνθήκη ώστε μία στατιστική συνάρτηση να περιέχει όλη την διαθέσιμη πληροφορία αναφορικά με την παράμετρο μ . Στην περίπτωση αυτή, η στατιστική συνάρτηση χαρακτηρίζεται ως επαρκής

Μία ικανή και αναγκαία συνθήκη⁷, για την στατιστική συνάρτηση $t(x_1, x_2, x_3, K, x_N)$ να είναι επαρκής όσον αφορά την παράμετρο θ , είναι ότι η πιθανοφάνεια μπορεί να παραγοντοποιηθεί ως:

⁶ Υπενθυμίζεται ότι το σ είναι γνωστό.

⁷ **Η (μη πλήρης) απόδειξη έχει ως εξής:** Από τις N διαθέσιμες παρατηρήσεις (x_1, x_2, x_3, K, x_N) μπορεί κανείς να κατασκευάσει (με άπειρους τρόπους) $m-1 \leq N-1$ αμοιβαία ανεξάρτητες στατιστικές συναρτήσεις :

$$t_1 = t_1(x_1, x_2, K, x_N), K, t_{m-1} = t_{m-1}(x_1, x_2, K, x_N), t_m = x_m, K, t_{N-1} = x_{N-1}$$

και ανεξάρτητες με την στατιστική συνάρτηση $t = t(x_1, x_2, K, x_N)$, για την οποία η συνάρτηση πιθανοφάνειας παραγοντοποιείται όπως στην σχέση 5.5.5

Στη συνέχεια, η πιθανοφάνεια εκφράζεται ως συνάρτηση των μεταβλητών t, t_1, t_2, K, t_{N-1} με τη βοήθεια του μετασχηματισμού:

$$L(\underline{x}; \theta) dx_1 dx_2 \dots dx_N = g(t|\theta) \cdot h(\underline{x}) \left| J \left(\frac{t_1, K, t_{N-1}}{x_1, x_2, K, x_N} \right) \right| dt_1 dt_2 \dots dt_{N-1} \quad (\text{όπου } J \text{ δηλώνει την Jacobian})$$

Ολοκληρώνοντας αμφότερα τα μέλη ως προς t_m, t_{m+1}, K, t_{N-1} καταλήγουμε ότι:

$$\begin{aligned} \int_{\Omega} L(\underline{x}; \theta) dx_m L \dots dx_{N-1} dx_1 L \dots dx_{m-1} &= \Phi(t, t_1, K, t_{m-1}; \theta) dt dt_1 L \dots dt_{m-1} \\ &= g(t; \theta) \cdot \left[\int_{\Omega} h(\underline{x}) \left| J \left(\frac{t_1, K, t_{N-1}}{x_1, x_2, K, x_N} \right) \right| dt_m L \dots dt_{N-1} \right] dt dt_1 L \dots dt_{m-1} \\ &= g(t; \theta) \cdot R(t_1, K, t_{m-1} | t) dt dt_1 L \dots dt_{m-1} \end{aligned}$$

όπου η συνάρτηση $\Phi(t, t_1, K, t_{m-1}; \theta)$ είναι η προβολή της πιθανοφάνειας πάνω στα t, t_1, K, t_{m-1} και υποθέσαμε ότι τα όρια ολοκλήρωσης είναι ανεξάρτητα της παραμέτρου θ . Αποδεικνύεται όμως (βλέπε M.G. Kendal and A. Stuart, "The Advanced Theory of Statistics") ότι ακόμα και εάν τα όρια ολοκλήρωσης εξαρτώνται από την παράμετρο θ , το ολοκλήρωμα

$$\int_{\Omega} h(\underline{x}) \left| J \left(\frac{t_1, K, t_{N-1}}{x_1, x_2, K, x_N} \right) \right| dt_m L \dots dt_{N-1}$$

στατιστικών συναρτήσεων, $t_1 = t_1(x_1, x_2, K, x_N), K, t_{m-1} = t_{m-1}(x_1, x_2, K, x_N), t_m = x_m, K, t_{N-1} = x_{N-1}$ και για

κάθε $m-1 \leq N-1$, το ολοκλήρωμα $\int_{\Omega} h(\underline{x}) \left| J \left(\frac{t_1, K, t_{N-1}}{x_1, x_2, K, x_N} \right) \right| dt_m L \dots dt_{N-1}$ θα είναι ανεξάρτητο της παραμέτρου θ ,

$$L(\underline{x};\theta) = g(t|\theta) \cdot h(\underline{x}) \quad 5.5.5$$

όπου α) η συνάρτηση $h(\underline{x})$ δεν εξαρτάται από την παράμετρο, θ , και β) η συνάρτηση $g(t|\theta) \propto A(t|\theta)$, όπου $A(t|\theta)$ είναι υπό συνθήκη πυκνότητα πιθανότητας για την στατιστική συνάρτηση t , δοθέντος του θ

Ερώτηση: Δείξτε ότι, εάν τα πεδία τιμών για τα $x_1, x_2, x_3, \dots, x_N$ είναι ανεξάρτητα της παραμέτρου, θ , τότε η συνθήκη β) είναι συνέπεια της συνθήκης α).

$$\text{Υπόδειξη: } A(t|\theta) = \int_{\Omega} L(\underline{x};\theta) d\underline{x} = g(t|\theta) \cdot \int_{\Omega} h(\underline{x}) d\underline{x}$$

Γιά ένα σύνολο N παρατηρήσεων, όταν η συνάρτηση πυκνότητας πιθανότητας εξαρτάται από k ($k \leq N$) παραμέτρους $\underline{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$, μπορούν να ευρεθούν S ($S \leq N$) επαρκείς στατιστικές συναρτήσεις, όπου το S μπορεί να είναι μεγαλύτερο, ίσο ή μικρότερο του πλήθους των παραμέτρων k .

Το σύνολο των επαρκών στατιστικών συναρτήσεων με το μικρότερο πλήθος, S , λέγεται ότι είναι ελάχιστο επαρκές (minimal sufficient). Προφανώς, με το ελάχιστο επαρκές σύνολο στατιστικών συναρτήσεων επιτυγχάνουμε την μέγιστη συμπίκνωση πληροφορίας, αφού όλη η διαθέσιμη πληροφορία για τις παραμέτρους εμπεριέχεται στο μικρότερο πλήθος στατιστικών συναρτήσεων.

Επισημαίνεται, ότι στις περισσότερες των περιπτώσεων το πλήθος των επαρκών στατιστικών συναρτήσεων εξαρτάται από το πλήθος, N , των παρατηρήσεων. Το απλούστερο παράδειγμα αναφέρεται στην περίπτωση όπου οι επαρκείς συναρτήσεις, t_i , μπορούν να ορισθούν μόνο ως: $t_i = x_i$, όπου x_i , $i = 1, 2, 3, \dots, N$, είναι οι πειραματικές παρατηρήσεις.

Μόνο μία πολύ περιορισμένη τάξη συναρτήσεων πυκνότητας πιθανότητας μπορεί να δεχθεί έναν αριθμό, S , από επαρκείς στατιστικές συναρτήσεις ανεξάρτητα από τον αριθμό των παρατηρήσεων (το S ανεξάρτητο του N). Συναρτήσεις πυκνότητας πιθανότητας που επιδέχονται ένα τέτοιο σύνολο επαρκών στατιστικών συναρτήσεων, ορίζονται μέσω του θεωρήματος του **Darmois**.

5.6 Το θεώρημα του Darmois (χωρίς απόδειξη)

Το θεώρημα, για μία μόνο παράμετρο ($k=1$) εκφράζεται ως εξής:

Για οποιοδήποτε Ω_θ , εάν υπάρχει ένα σύνολο τιμών της τυχαίας μεταβλητής x , $\{x_1, x_2, x_3, \dots, x_N\}$ με $N > 1$, ώστε να υπάρχει μία επαρκής στατιστική συνάρτηση για την παράμετρο θ , τότε η συνάρτηση πυκνότητας πιθανότητας του x , $f(x;\theta)$, έχει την ακόλουθη εκθετική συναρτησιακή μορφή:

$$f(x;\theta) = \exp[\alpha(x) \cdot a(\theta) + \beta(x) + c(\theta)] \quad 5.6.1$$

η συνάρτηση $R(t_{1:k}, t_{m-1}|t)$ δεν μεταφέρει καμία πληροφορία αναφορικά με την παράμετρο.

Συνεπώς, οι στατιστικές $t_{1:k}, t_{m-1}$ που επιλέγονται κάθε φορά, δεν συνεισφέρουν ουδόλως σε πληροφορία για το θ , αλλά όλη η πληροφορία εμπεριέχεται στην στατιστική συνάρτηση t , η οποία είναι επαρκής αναφορικά με την παράμετρο θ .

Αντίστροφα, οι N τιμές της τυχαίας μεταβλητής x , $\{x_1, x_2, x_3, \dots, x_N\}$, επιδέχονται μία επαρκή στατιστική συνάρτηση της παραμέτρου θ , για όλα τα $N > 1$ (αλλά μόνο εάν το Ω_θ δεν εξαρτάται από το θ) εάν η συνάρτηση πυκνότητας πιθανότητας, $f(x; \theta)$, έχει εκθετική την μορφή 5.6.1. Επιπλέον απαιτείται η απεικόνιση:

$$\{x_1, x_2, x_3, \dots, x_N\} \rightarrow \{x_1, x_2, x_3, \dots, x_N\} : R = \sum_{i=1}^N \alpha(x_i)$$

να είναι ένα προς ένα και συνεχώς διαφορίσιμη για όλα τα x_i .

Η στατιστική συνάρτηση R είναι επαρκής για το θ , και το ίδιο συμβαίνει για οποιαδήποτε άλλη μονοτονική συνάρτηση του R .

Παράδειγμα 5.5.2

Σαν παράδειγμα η κανονική συνάρτηση πυκνότητας πιθανότητας:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] \text{ είναι εκθετικής μορφής με } \Omega_\theta = [-\infty, \infty] \text{ ανεξάρτητο από}$$

τα μ και σ .

$$\text{Επιπλέον } f(x; \mu, \sigma) = \exp \left(x \cdot \frac{\mu}{\sigma^2} - \frac{x^2}{2 \cdot \sigma^2} - \frac{\mu^2}{2 \cdot \sigma^2} - \ln [\sigma \cdot \sqrt{2 \cdot \pi}] \right) \quad 5.6.2$$

Εάν υποθέσουμε ότι σ^2 είναι γνωστό και μ άγνωστο, συγκρίνοντας με την 5.6.1 θα έχουμε

$$a = \mu \quad \alpha = \frac{x}{\sigma^2} \quad c = -\frac{\mu^2}{2\sigma^2} \quad \beta = -\frac{x^2}{2\sigma^2} - \ln [\sigma \sqrt{2\pi}]$$

$$R = \sum_{i=1}^N \frac{x_i}{\sigma^2} = \frac{N}{\sigma^2} \cdot \left[\frac{1}{N} \sum_{i=1}^N x_i \right] = \frac{N}{\sigma^2} \langle x \rangle$$

Προφανώς, ο μέσος όρος, $\langle x \rangle$ είναι μονοτονική συνάρτηση του R

Ερώτηση: Τι συμβαίνει όταν το μ είναι γνωστό και το σ^2 άγνωστο;

Υπόδειξη: Από την σχέση 5.6.2, συνάγεται ότι:

$$a = \frac{1}{\sigma^2} \quad \alpha = x \cdot \left[\mu - \frac{x}{2} \right] \quad c = -\frac{\mu^2}{2\sigma^2} - \ln [\sigma \sqrt{2\pi}] \quad \beta = 0$$

$$R = \sum_{i=1}^N x_i \cdot \left[\mu - \frac{x_i}{2} \right]$$

Στην συνέχεια δείξτε ότι $N \cdot A = \sum_{i=1}^N [\mu - x_i]^2 = 2 \cdot R - N \cdot \mu^2$ και συνεπώς A είναι μονοτονική συνάρτηση του R .

Στην περίπτωση όπου $k > 1$ ($\underline{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$) υπάρχουν S επαρκείς στατιστικές συναρτήσεις για τις παραμέτρους $\underline{\theta}$, η γενίκευση της εκθετικής μορφής, 5.6.1, για την συνάρτηση πυκνότητας πιθανότητας γράφεται ως εξής:

$$f(x|\underline{\theta}) = \exp \left[\sum_{j=1}^S \alpha_j(x) \cdot a_j(\underline{\theta}) + \beta(x) + c(\underline{\theta}) \right]$$

και οι S επαρκείς συναρτήσεις είναι οι: $R_j = \sum_{i=1}^N \alpha_j(x_i) \quad (j=1,2,\dots,S)$

Παράδειγμα 5.5.3:

Στην περίπτωση, κανονικής συνάρτησης πιθανότητας όπου αμφότεροι οι παράμετροι, μ και σ^2 , είναι άγνωστοι:

$$f(x;\mu,\sigma) = \exp \left\{ x \frac{\mu}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \ln [\sqrt{2\pi}\sigma] \right\}$$

$$a_1 = \frac{\mu}{\sigma^2} \quad \alpha_1 = x \quad \beta = 0$$

$$a_2 = \frac{1}{\sigma^2} \quad \alpha_2 = -\frac{x^2}{2} \quad c = -\frac{\mu^2}{2\sigma^2} - \ln [\sqrt{2\pi}\sigma]$$

$$R_1 = \sum_{i=1}^N \alpha_1(x_i) = \sum_{i=1}^N x_i$$

$$R_2 = \sum_{i=1}^N \alpha_2(x_i) = \sum_{i=1}^N \frac{x_i^2}{2}$$

Πρόβλημα: Έστω N πειραματικές παρατηρήσεις $\{x_1, x_2, x_3, \dots, x_N\}$ με συνάρτηση πυκνότητας πιθανότητας $f(x;\theta)$ και έστω η στατιστική συνάρτηση, $T = T(x_1, x_2, x_3, \dots, x_N)$. Δείξτε ότι η πληροφορία ($I_T(\theta)$) που εμπεριέχεται στην στατιστική συνάρτηση T , αναφορικά με την παράμετρο θ , είναι μικρότερη ή ίση από την αντίστοιχη πληροφορία ($I_{\{x_1, x_2, x_3, \dots, x_N\}}(\theta)$) που εμπεριέχεται στο σύνολο των πειραματικών δεδομένων. Στην συνέχεια δείξτε ότι η ισότητα ισχύει όταν ισχύουν οι προϋποθέσεις που εκφράζονται με την σχέση 5.5.5. (Υποθέσετε ότι το πεδίο τιμών της μεταβλητής x δεν εξαρτάται από την παράμετρο θ .)

Απάντηση: Η πιθανοφάνεια $L(\underline{x};\theta) = \prod_{i=1}^N f(x_i;\theta)$ ($\underline{x} = \{x_1, x_2, x_3, \dots, x_N\}$) πάντοτε μπορεί να παραγοντοποιηθεί ως: $L(\underline{x};\theta) = g(T|\theta) \cdot H(\underline{x};\theta)$, όπου $g(T|\theta)$ είναι η υπό συνθήκη συνάρτηση πιθανότητας της στατιστικής συνάρτησης, T , και η συνάρτηση $H(\underline{x};\theta)$ εξαρτάται και από την παράμετρο θ .

Χρησιμοποιώντας την σχέση 5.3.3 γράφουμε για την συνολική πληροφορία $I_{\{x_1, x_2, x_3, \dots, x_N\}}(\theta)$:

$$I_{\{x_1, x_2, x_3, \dots, x_N\}}(\theta) = E \left[\frac{\partial^2}{\partial \theta^2} \ln L(\underline{x}; \theta) \right]$$

$$= E \left[\frac{\partial^2}{\partial \theta^2} \ln g(T | \theta) \right] +$$

Πρόβλημα: Αποδείξτε ότι, για τις τυχαίες μεταβλητές $x = \{x_1, x_2, x_3, \dots, x_N\}$ με συνάρτηση πυκνότητας πιθανότητας Gauss με γνωστή μέση τιμή, μ , και άγνωστη τετραγωνική διασπορά, σ^2 , η στατιστική συνάρτηση

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

είναι επαρκής όσον αφορά στο σ^2 . Η υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας για το σ^2 είναι

$$g(\hat{\sigma}^2 | \sigma^2) = \frac{1}{\sqrt{N/2}} \left(\frac{N}{2\sigma^2} \right)^{\frac{N}{2}} (\hat{\sigma}^2)^{\frac{N-2}{2}} \exp \left(-\frac{N}{2\sigma^2} < \sigma^2 > \right)$$