5. ΜΕΘΟΔΟΙ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

5.1 Η συνάρτηση μέγιστης πιθανοφάνειας

Έστω μία τυχαία μεταβλητή d η οποία αντιπροσωπεύει την μέτρηση κάποιας συγκεκριμένης ποσότητας με πραγματική αλλά άγνωστη τιμή θ σε ένα πείραμα που εκτελείται Ν φορές. Οι μετρήσεις που λαμβάνονται d_i αποτελούν τα στοιχεία ενός διανύσματος $\vec{d}$. Εάν οι μετρήσεις γίνονται σε περιβάλλον θορύβου ή με όργανα που ενδέχεται να εμφανίζουν λάθος μέτρηση, τότε οι τιμές d_i δεν θα είναι κατ΄ ανάγκη ίσες με την πραγματική τιμή της ποσότητας. Εάν η πιθανότητα η μέτρηση d_i να πάρει την τιμή θ είναι $P_d(d_i;\theta)$, η πιθανότητα όλες οι μετρήσεις να πάρουν την τιμή θ, με δεδομένο ότι η πιθανότητα κάθε μέτρηση να πάρει την εν λόγω τιμή είναι ανεξάρτητη οποιασδήποτε άλλης είναι :

$$P_d(d_1, d_2, ...d_N;\theta) = P_d(d_1;\theta)\cdots P_d(d_N;\theta) \tag{5.1}$$

Ορίζουμε ως *συνάρτηση πιθανοφάνειας (likelihood function)* $L(\theta;\vec{d})$ την ανωτέρω έκφραση της από κοινού πιθανότητας των μετρήσεων $\vec{d}$.

$$L(\theta;\vec{d}) = P_d(d_1, d_2, ...d_N;\theta) = P_d(d_1;\theta)\cdots P_d(d_N;\theta) \tag{5.2}$$

Η τιμή της θ που μεγιστοποιεί την συνάρτηση πιθανοφάνειας, ονομάζεται *εκτίμηση μέγιστης πιθανοφάνειας*; (*maximum likelihood estimate*) μια μπορεί να βρεθεί παραγωγίζοντας την L ως προς θ και ελέγχοντας ότι η δεύτερη παράγωγος είναι μικρότερη του 0 :

$$\frac{dL}{d\theta} = 0, \frac{d^2L}{d\theta^2} < 0 \tag{5.3}$$

Εάν η $\vec{d}$ ακολουθεί κανονική κατανομή με μέση τιμή m και διασπορά σ^2, η έκφραση της συνάρτησης πιθανοφάνειας γίνεται :

$$\begin{gathered} L(m;\sigma^2/\vec{d}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(d_1-m)}{2\sigma^2})\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(d_2-m)}{2\sigma^2})\right)\cdots \\ \left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(d_N-m)}{2\sigma^2})\right) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}}\exp\left[-\frac{1}{2\sigma^2}\sum_{\iota=1}^{\mathrm{N}}(d_i-m)^2\right] \end{gathered} \tag{5.4}$$

Η εκτίμηση μέγιστης πιθανοφάνειας για την μέση τιμή *m* προκύπτει από την 5.3. Παράλληλα μπορούμε να ζητήσουμε και την εκτίμηση μέγιστης πιθανοφάνειας για την διασπορά, από τον ίδιο τύπο.

Η εκτίμηση μέγιστης πιθανοφάνειας εκφράζει την μέση τιμή της κατανομής πιθανότητας των μετρήσεων που δικαιολογούν με τον καλύτερο τρόπο τις μετρήσεις. Με άλλα λόγια, γνωρίζοντας την συνάρτηση πυκνότητας πιθανότητας των μετρήσεων ως προς τη μορφή της (εδώ έχομε κανονική κατανομή), αναζητούμε τις παραμέτρους της κατανομής που να δίνει μετρήσεις όσο το δυνατόν κοντύτερα σε αυτή που αναμένεται να συγκεντρώνει τη μέγιστη πιθανότητα.

Πολλές φορές χρησιμοποιούμε τον λογάριθμο της συνάρτησης πιθανοφάνειας ως την προς μεγιστοποίηση συνάρτηση, καθώς ο λογάριθμος της L είναι συνάρτηση με ίδια χαρακτηριστικά μονοτονίας σε σχέση με την L. Επομένως μπορούμε να εκφράσουμε τις παραγώγους της 5.3 ως προς $\ln(L)$. Έτσι παίρνομε :

$$L' = \ln(L) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{\iota=1}^{\mathrm{N}}(d_i - m)^2 \tag{5.5}$$

$$\frac{\partial L'}{\partial m} = \frac{1}{\sigma^2}\sum_{\iota=1}^{\mathrm{N}}(d_i - m) = \frac{1}{\sigma^2}\left(\sum_{\iota=1}^{\mathrm{N}} d_i - Nm\right) \tag{5.6}$$

$$\frac{\partial L'}{\partial \sigma^2} = -\frac{N}{\sigma^2} + \frac{1}{\left(\sigma^2\right)^2}\sum_{\iota=1}^{\mathrm{N}}(d_i - m)^2 = \frac{N}{\left(\sigma^2\right)^2}\left(\sigma^2 - \frac{1}{N}\sum_{\iota=1}^{\mathrm{N}}(d_i - m)^2\right) \tag{5.7}$$

Θέτοντας τις παραγώγους από τις 5.6 και 5.7 ίσες με 0, παίρνομε τις εκτιμήσεις μέγιστης πιθανοφάνειας για m και σ^2 που είναι αντίστοιχα :

$$m^{est} = \frac{1}{N}\sum_{i=1}^{N} d_i \tag{5.8}$$

και

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(d_i - m^{est})^2 \tag{5.9}$$

που αποτελούν τις γνωστές εκφράσεις για την μέση αριθμητική τιμή και την τυπική απόκλιση (standard deviation) της δειγματοληψίας. Επισημαίνεται ότι η εκτίμηση μέγιστης πιθανοφάνειας των μετρήσεων συμπίπτει με τον αριθμητικό μέσο, λόγω της υπόθεσης που κάναμε για κανονική κατανομή.

5.2 Εκτιμήσεις μέγιστης πιθανοφάνειας για το γραμμικό αντίστροφο πρόβλημα.

Θα δούμε τώρα πως υλοποιούνται οι ανωτέρω έννοιες στο γραμμικό αντίστροφο πρόβλημα της μορφής $\mathbf{Gm} = \mathbf{d}$. Υποθέτομε ότι οι μετρήσεις του προβλήματος υπακούουν σε μία κανονική κατανομή της μορφής :

$$P(\mathbf{d}) \propto \exp\left[-\tfrac{1}{2}(\mathbf{d} - \mathbf{Gm})^T [\operatorname{cov}\mathbf{d}]^{-1}(\mathbf{d} - \mathbf{Gm})\right] \tag{5.10}$$

Ο όρος $\mathbf{Gm}$ μπορεί να θεωρηθεί ότι αντιπροσωπεύει κάτι ανάλογο με τον μέσο των μετρήσεων ($\langle \mathbf{d} \rangle$) (δείτε εξίσωση 2.14). Ωστόσο εδώ έχει τη σημασία των εκτιμήσεων των μετρήσεων για δεδομένο **m.** Επομένως μπορεί να δει κανείς την διαφορά $\mathbf{d} - \mathbf{Gm} = \mathbf{d}^{obs} - \mathbf{d}^{pre} = e_i$.

Η μεγιστοποίηση της $P(\mathbf{d})$ αντιστοιχεί σε ελαχιστοποίηση της ποσότητας $(\mathbf{d} - \mathbf{Gm})^T [\operatorname{cov}\mathbf{d}]^{-1}(\mathbf{d} - \mathbf{Gm})$. Η εκτίμηση μέγιστης πιθανοφάνειας λοιπόν στην περίπτωσή μας θα δώσει ένα $\mathbf{m}^{est}$ το οποίο επιβεβαιώνει με τον καλύτερο τρόπο τις μετρήσεις. Με άλλα λόγια η βέλιστη λύση για τις παραμέτρους **m**, είναι η λύση ελαχίστων τετραγώνων για ένα καθαρά υπερορισμένο πρόβλημα, στην οποία η έκφραση του «λάθους» των μετρήσεων έχει ζυγιστεί με τον αντίστροφο του πίνακα συνδιακύμανσης. Δηλαδή $\mathbf{W}_e = [\operatorname{cov}\mathbf{d}]^{-1}$. Εάν όλες οι μετρήσεις είναι ασυσχέτιστες και έχουν την ίδια διασπορά, τότε $[\operatorname{cov}\mathbf{d}] = \sigma_d^2 \mathbf{I}$ και η λύση μέγιστης πιθανοφάνειας είναι η απλή λύση ελαχίστων τετραγώνων. Εάν έχομε μετρήσεις ασυσχέτιστες αλλά με διαφορετική διασπορά (σ_{di}^2) τότε το λάθος εκτίμησης είναι

$$E = \sum_{i=1}^{N} \sigma_{di}^{-2} e_i^2 \tag{5.11}$$

5.3 Εκ προοιμίου κατανομές

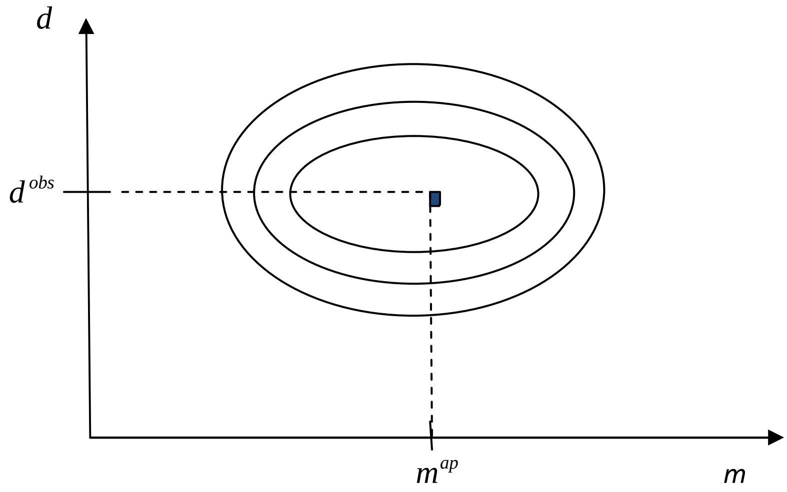
Εάν το πρόβλημα είναι υποορισμένο, δεν υφίσταται η λύση ελαχίστων τετραγώνων. Όπως έχουμε δει, θα πρέπει να εισαχθούν αρχικές συνθήκες ως προς τις παραμέτρους, που εδώ μπορούν να πάρουν τη μορφή πιθανοθεωρητικών κατανομών. Στην περίπτωση αυτή θα πρέπει να αξιοποιήσουμε τη συνάρτηση κατανομής πιθανότητας $P_A(\mathbf{m})$ εφ’ όσον είναι γνωστή, οπότε έχομε πληροφορία για την μέση τιμή των παραμέτρων αλλά και για τη διασπορά τους. Συνδυάζοντας αυτή την πληροφορία με την αντίστοιχη για τις μετρήσεις, μπορούμε να βρούμε την εκτίμηση μέγιστης πιθανοφάνειας που βελτιστοποιεί εκτιμήσεις παραμέτρων και μετρήσεων.

Συνήθως οι κατανομές πιθανότητας παραμέτρων και μετρήσεων είναι ασυσχέτιστες, συνεπώς, η από κοινού συνάρτηση πιθανότητας εκφράζεται μέσω της σχέσης :

$$P_A(\mathbf{m},\mathbf{d}) = P_A(\mathbf{m})P(\mathbf{d}) \tag{5.12}$$

Εάν εφαρμοστεί μέθοδος μέγιστης πιθανοφάνειας σε συνάρτηση που προκύπτει από την 5.12, σημειώνουμε ότι δεν έχει ληφθεί υπ' όψιν της το «μοντέλο» που χαρακτηρίζει το αντίστροφο πρόβλημα, δηλαδή η σχέση που συνδέει παραμέτρους και μετρήσεις, αλλά μόνο η πληροφορία για τις κατανομές πιθανοτήτων.

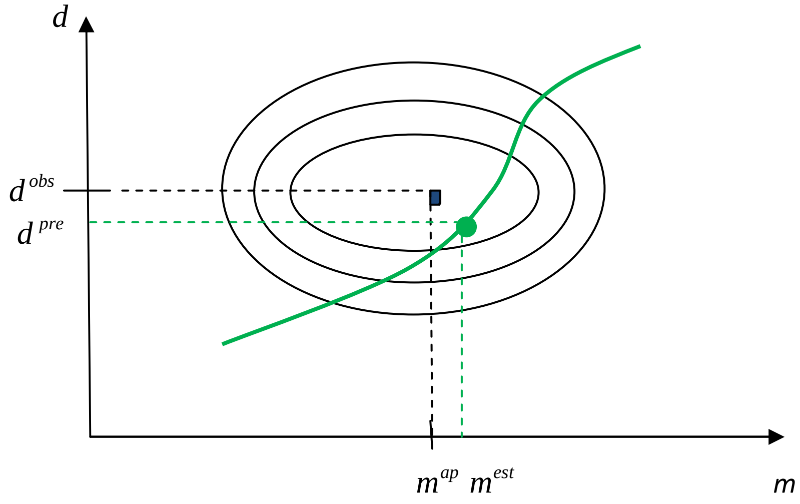
Στο σχήμα 5.1 παρουσιάζονται καμπύλες ίσης πιθανότητας για μία παράμετρο και μία μέτρηση σε ένα σχετικό πρόβλημα. Οι καμπύλες εκτείνονται γύρω από το σημείο (d^{obs}, m^{ap}) που δίνει τη μέγιστη πιθανότητα συνδυασμού μέτρησης και παραμέτρου.



Σχήμα 5.1 Από κοινού συνάρτηση κατανομής πιθανότητας για μετρήσεις και παραμέτρους

5.4 Εκτιμήσεις μέγιστης πιθανοφάνειας για ακριβή θεωρία.

Εάν στην πληροφορία για κατανομές πιθανοτήτων μετρήσεων και παραμέτρων προστεθεί και η σχέση που συνδέει τις δύο ποσότητες (είτε γραμμική είτε μη γραμμική), έχομε και μία επί πλέον πληροφορία που θα πρέπει να αξιοποιήσομε. Εάν η θεωρία αυτή είναι ακριβής, γνωρίζομε με σιγουριά την επιφάνεια (σε ένα πολυεπίπεδο χώρο) στην οποία πρέπει να αναζητηθεί η βέλτιστη λύση. Η θεωρία εκφράζεται στη γενική περίπτωση των διακριτών προβλημάτων από μία σχέση της μορφής : $\mathbf{g(m) = d}$ (που δεν είναι κατ' ανάγκην γραμμική). Σχηματικά, βλέπομε τη διαφοροποίηση σε σχέση με την προηγούμενη περίπτωση στο σχήμα 5.2 όπου στις καμπύλες ίσης πιθανότητας παραμέτρων και δεδομένων όπως αυτές έχουν δοθεί «εκ προοιμίου), έχει προστεθεί και η επιφάνεια (γραμμή σε δύο διαστάσεις) που εκφράζει τη σχέση ανάμεσα σε παραμέτρους και μετρήσεις. Η βέλτιστη λύση πρέπει να αναζητηθεί πάνω σε αυτή τη γραμμή και μπορεί να μας δώσει βέλτιστο σημείο (d^{pre}, m^{est}) διαφορετικό από το προηγούμενο.



Σχήμα 5.2 Από κοινού συνάρτηση κατανομής πιθανότητας για μετρήσεις και παραμέτρου ςμε ακριβή θεωρία (μοντέλο).

5.5 Εκτιμήσεις μέγιστης πιθανοφάνειας για μη ακριβή θεωρία.

Εάν το μοντέλο που συνδέει παραμέτρους και μετρήσεις δεν είναι ακριβές, τότε μπορεί να εκφραστεί μέσω οικογένειας επιφανειών διαφορετικής πιθανότητας, εφ΄ όσον βέβαια γνωρίζομε μία κατανομή πιθανότητας $P_g(\mathbf{m}|\mathbf{d})$ με κέντρο μία σχέση της μορφής $\mathbf{g(m) = d}$. Πρέπει να προσέξομε στην έκφραση της $P_g(\mathbf{m}|\mathbf{d})$ ότι, έχομε μία πιθανότητα υπό συνθήκη, δηλαδή έχομε την πιθανότητα η θεωρία να προβλέπει τις μετρήσεις **d** όταν δίδονται οι παράμετροι **m.** Στην περίπτωση αυτή ζητάμε να υπολογίσομε παραμέτρους και μετρήσεις με την αρχή της μέγιστης πιθανοφάνειας, ορίζοντας μία νέα συνάρτηση πιθανοφάνειας :

$$P_T(\mathbf{m},\mathbf{d}) = P_g(\mathbf{m}|\mathbf{d})P_A(\mathbf{m},\mathbf{d}) \tag{5.13}$$

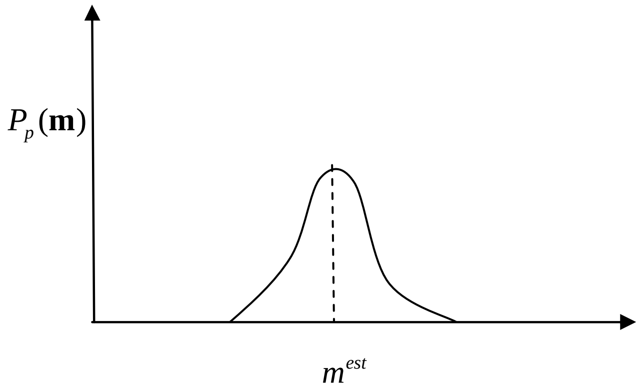
που δικαιολογείται από το γεγονός ότι η εκ προοιμίου κατανομή πιθανότητας παραμέτρων και μετρήσεων είναι ανεξάρτητη από την κατανομή πιθανότητας της θεωρίας.

Αξίζει να προσέξομε ότι με βάση την αρχή της μέγιστης πιθανοφάνειας, υπολογίζομε ταυτόχρονα βέλτιστες παραμέτρους και μετρήσεις, και όχι μόνο πατραμέτρους όπως κάναμε π.χ. στην περίπτωση των ελαχίστων τετραγώνων. Έτσι οι εκτιμήσεις των παραμέτρων με τις δύο προσεγγίσεις μπορεί να είναι διαφορετικές.

Εάν ολοκληρώσουμε την έκφραση της συνάρτησης πιθανοφάνειας από την 5.13 ως προς τις μετρήσεις **d,** προβάλουμε ουσιαστικά την συνάρτηση στο επίπεδο **d**=0 :

$$P_p(\mathbf{m}) = \int P_T(\mathbf{m}, \mathbf{d}) \partial \mathbf{d} \tag{5.14}$$

Σε περίπτωση ακριβούς θεωρίας, η εκτίμηση μέγιστης πιθανοφάνειας που δίδει η 5.14 δεν είναι διαφορετική από αυτή που δίδεται εάν μεγιστοποιήσομε την 5.12 στην επιφάνεια που ορίζει η ακριβής θεωρία.



Σχήμα 5.3 Η κατανομή $P_p(\vec{m})$. *Το μέγιστο της κατανομής ορίζει την εκτίμηση μέγιστης πιθανοφάνειας* m^{est}.