

ΜΕΡΟΣ Β

Στατιστική Συμπερασματολογία

ΚΕΦΑΛΑΙΟ 12

ΕΚΤΙΜΗΤΙΚΗ: ΣΗΜΕΙΑΚΗ ΕΚΤΙΜΗΣΗ

ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΕΚΤΙΜΗΤΙΚΗΣ

Σε πολλές περιπτώσεις στην στατιστική, συναντώνται προβλήματα για τα οποία απαιτείται να εκτιμηθεί μία παράμετρος του υπό μελέτη πληθυσμού.

Η μέθοδος που ακολουθεί στις περιπτώσεις αυτές κανείς από διαίσθηση είναι να χρησιμοποιεί την τιμή μιας αντίστοιχης ποσότητας από ένα τυχαίο δείγμα ως εκτίμηση της άγνωστης παραμέτρου του πληθυσμού.

Παραδείγματα όπου η εκτίμηση των παραμέτρων του πληθυσμού είναι απαραίτητη είναι, μεταξύ άλλων τα εξής:

1. Μια εταιρεία έρευνας αγοράς ενδιαφέρεται να εκτιμήσει την μέση μετακίνηση προτίμησης των καταναλωτών που προέρχεται από μια διαφημιστική καμπάνια στην τηλεόραση και υπολογίζεται μέσω μιας δειγματοληπτικής έρευνας.
2. Μια αποθήκη χονδρικής πώλησης, προκειμένου να καθορίσει μια πολιτική για το απόθεμα που θα διατηρεί, χρειάζεται να εκτιμήσει το επίπεδο ζήτησης για προϊόντα τα οποία διακινεί, με βάση ένα δείγμα, για κάποια χρονική περίοδο.

Η περιοχή της στατιστικής που ασχολείται με το πρόβλημα αυτό λέγεται *εκτιμητική*. Ένα από τα προβλήματα που απασχολούν την περιοχή αυτή είναι ο καθορισμός κριτηρίων με βάση τα οποία ο ερευνητής θα αποφασίζει πόσο “καλή” είναι η εκτιμήτρια μιας παραμέτρου του υπό μελέτη πληθυσμού.

Άλλο πρόβλημα είναι ο καθορισμός των κριτηρίων εκείνων βάσει των οποίων ο ερευνητής θα αποφασίζει αν μια εκτιμήτρια είναι η καλύτερη δυνατή από αυτές που έχει στην διάθεσή του.

Πρωταρχικό, βέβαια, πρόβλημα είναι η ανεύρεση μεθόδων καθορισμού εκτιμητριών παραμέτρων.

Η μεθοδολογία που θα αναπτύξουμε στηρίζεται σε δείγματα που έχουν ληφθεί με τυχαίο τρόπο (τυχαία δείγματα). Αν ένα δείγμα δεν είναι τυχαίο, κανένα ακριβές συμπέρασμα δεν είναι δυνατόν

να προέλθει από το δείγμα αυτό όσον αφορά τον πληθυσμό. Είναι ενδεχόμενο βέβαια να εξαχθούν κάποια συμπεράσματα σε μερικές περιπτώσεις τα οποία όμως είναι τετριμμένα. Το πρόβλημα, στην απλούστερη μορφή του, είναι να εκτιμηθεί η τιμή της παραμέτρου του πληθυσμού από πληροφορίες που παρέχονται από ένα τυχαίο δείγμα. Η απλούστερη περίπτωση είναι εκείνη που δίνει μια μόνο τιμή ως εκτίμηση της παραμέτρου που θέλουμε να εκτιμήσουμε. Στην περίπτωση αυτή μιλάμε για *σημειακή εκτίμηση*.

ΣΗΜΕΙΑΚΗ ΕΚΤΙΜΗΣΗ

Ας διευκρινίσουμε κάπως περισσότερο τι εννοούμε με τον όρο *εκτιμητική*. Υποθέτουμε κατ' αρχήν ότι ο πληθυσμός τον οποίο θέλουμε να μελετήσουμε έχει μορφή που είναι πλήρως καθορισμένη εκτός από την τιμή κάποιας παραμέτρου θ (ή κάποιων παραμέτρων θ_i , $i=1,2,\dots,k$). Στην συνέχεια, θεωρούμε τις παρατηρήσεις x_1, x_2, \dots, x_n ενός τυχαίου δείγματος. Εκείνο που θέλουμε, χρησιμοποιώντας τις παρατηρήσεις αυτές, είναι να καθορίσουμε ένα αριθμό ο οποίος μπορεί να θεωρηθεί ως η τιμή της συγκεκριμένης παραμέτρου θ . (Όταν μιλάμε για ένα διάστημα τιμών της παραμέτρου, η αντίστοιχη στατιστική μεθοδολογία οδηγεί στα ονομαζόμενα *διαστήματα εμπιστοσύνης*, τα οποία θα εξετάσουμε αργότερα).

Είναι προφανές ότι οι παρατηρήσεις είναι αυτές καθαυτές τυχαίες μεταβλητές (δοθέντος ότι ένα άλλο τυχαίο δείγμα θα οδηγήσει, πιθανότατα, σε διαφορετικές παρατηρήσεις). Επομένως, και οποιαδήποτε συνάρτηση των παρατηρήσεων του δείγματος θα είναι κι αυτή μια τυχαία μεταβλητή.

Ορισμός: Μια συνάρτηση των παρατηρήσεων ενός δείγματος που εξαρτάται μόνο από τις παρατηρήσεις λέγεται *στατιστική συνάρτηση (statistic)*.

Εάν χρησιμοποιήσουμε μια στατιστική συνάρτηση για να εκτιμήσουμε μια παράμετρο θ , είναι ενδεχόμενο, σε κάποιες περιπτώσεις, η εκτίμηση αυτή να διαφέρει σημαντικά από την πραγματική τιμή της παραμέτρου θ . Ενδέχεται, επομένως, να μην είναι δυνατόν να καταλήξουμε σε κάποια μέθοδο εκτίμησης η οποία

να εγγυάται την πλησιέστερη εκτίμηση του θ σε κάθε δυνατή περίπτωση και για κάθε τυχαίο δείγμα.

Θα πρέπει επομένως, να περιορισθούμε στον καθορισμό ενός κανόνα που θα δίνει “καλά” αποτελέσματα “μακροπρόθεσμα” ή κατά “μέσο όρο” ή που “θα έχει υψηλή πιθανότητα επιτυχίας”. Όλες οι φράσεις που χρησιμοποιήσαμε προηγουμένως εκφράζουν το βασικό γεγονός ότι οποιαδήποτε μέθοδο εκτίμησης χρησιμοποιήσουμε θα πρέπει να οδηγεί στην δημιουργία κατανομών εκτιμητριών. Οι μέθοδοι αυτές, έτσι, θα αξιολογούνται ανάλογα με τις ιδιότητες των κατανομών αυτών.

Πριν προχωρήσουμε στην ανάπτυξη της θεωρίας αυτής, είναι απαραίτητο να διευκρινίσουμε την διαφορά ανάμεσα στην μεθοδολογία ή τους κανόνες εκτίμησης που θα χρησιμοποιήσουμε την οποία θα ονομάζουμε *εκτιμήτρια* (ή εκτιμητή) (*estimator*) και την τιμή που η εκτιμήτρια παίρνει σε κάποια συγκεκριμένη περίπτωση την οποία θα ονομάζουμε *εκτίμηση* (*estimate*).

Η διαφορά μεταξύ των δύο εννοιών είναι η ίδια που υπάρχει μεταξύ μιας συνάρτησης $f(x)$, η οποία θεωρείται ορισμένη για μια σειρά τιμών της μεταβλητής x και της συγκεκριμένης τιμής που η συνάρτηση αυτή παίρνει, έστω $f(a)$ για κάποια καθορισμένη τιμή του x που είναι ίση με το a .

Το πρόβλημά μας δεν είναι να βρούμε εκτιμήσεις, αλλά να βρούμε εκτιμήτριες. Για τον λόγο αυτό, δεν απορρίπτουμε μια εκτιμήτρια διότι σε κάποια συγκεκριμένη περίπτωση έδωσε ένα άσχημο αποτέλεσμα (με την έννοια ότι η εκτίμηση διαφέρει σημαντικά από την πραγματική τιμή). Απορρίπτουμε μία εκτιμήτρια μόνο εάν δίνει άσχημα αποτελέσματα μακροπρόθεσμα, αν, δηλαδή, η κατανομή όλων των πιθανών τιμών της εκτιμήτριας αποκλίνει σημαντικά από την πραγματική τιμή της παραμέτρου θ . Η αξία μιας εκτιμήτριας αξιολογείται από την κατανομή των εκτιμήσεων τις οποίες παράγει, δηλαδή, με άλλα λόγια, από τις ιδιότητες της *δειγματικής της κατανομής*.

Ορισμός: *Εκτιμήτρια (estimator)* είναι μια τυχαία μεταβλητή που χρησιμοποιείται για να εκτιμήσει ένα χαρακτηριστικό του πληθυσμού

(π.χ. μια παράμετρο). Η αριθμητική τιμή που η εκτιμήτρια παίρνει για κάποιο συγκεκριμένο δείγμα ονομάζεται *εκτίμηση* (*estimate*).

Σε πολλές περιπτώσεις, διαισθητικά, χρησιμοποιούμε ως εκτιμήτρια της παραμέτρου ενός πληθυσμού την αντίστοιχη συνάρτηση του δείγματος. Για παράδειγμα, ως εκτιμήτρια της μέσης τιμής μ ενός πληθυσμού χρησιμοποιούμε συνήθως τον δειγματικό μέσο \bar{X} . (Δηλαδή ο δειγματικός μέσος θεωρείται ως μια εκτιμήτρια της μέσης τιμής του πληθυσμού). (Η μέθοδος αυτή προσδιορισμού μιας εκτιμήτριας λέγεται *μέθοδος των ροπών* και θα εξετασθεί αργότερα.)

Παράδειγμα: Έστω ότι μια εταιρεία θέλει να εκτιμήσει τον μέσο χρόνο μ υπηρεσίας στην εταιρεία του συνόλου των 3580 εργαζομένων στην εταιρεία προκειμένου να καθορίσει την στάση της στις επόμενες διαπραγματεύσεις για υπογραφή συλλογικής σύμβασης εργασίας.

Ένας τρόπος για να εκτιμηθεί η άγνωστη παράμετρος μ είναι να χρησιμοποιηθεί ένα τυχαίο δείγμα από εργαζόμενους στην εταιρεία. Ένας προφανής τρόπος για εκτίμηση του μ είναι να χρησιμοποιηθεί ο δειγματικός μέσος \bar{X} . Έστω ότι το γραφείο προσωπικού της εταιρείας, χρησιμοποιώντας ένα απλό τυχαίο δείγμα 50 εργαζομένων και καθορίζοντας τον χρόνο προϋπηρεσίας στην εταιρεία καθενός από αυτούς, κατέληξε στο συμπέρασμα ότι ο μέσος χρόνος προϋπηρεσίας των εργαζομένων στο δείγμα είναι $\bar{X}=6.3$ χρόνια. Το \bar{X} είναι η εκτιμήτρια του μ ενώ το $\bar{X}=6.3$ είναι η εκτίμηση του μ που προκύπτει από το συγκεκριμένο δείγμα.

Σε πολλές περιπτώσεις η ακρίβεια των εκτιμητριών αυξάνεται όσο μεγαλύτερο είναι το μέγεθος του δείγματος από το οποίο έχουν προέλθει. Στην περίπτωση αυτή λέμε ότι η ακρίβεια της εκτιμήτριας αυξάνει με το n . Η έννοια της διατύπωσης αυτής είναι ότι, όσο αυξάνει το n , η διασπορά της κατανομής της εκτιμήτριας ελαττώνεται, ενώ η μέση τιμή της ή ταυτίζεται ή πλησιάζει πάρα πολύ στην πραγματική τιμή της υπό εκτίμηση παραμέτρου. (Παράδειγμα είναι ο δειγματικός μέσος \bar{X} ως εκτιμήτρια του μέσου ενός πληθυσμού μ).

Η κατάσταση αυτή ισχύει για τις παραμέτρους των περισσοτέρων από τους πληθυσμούς με τους οποίους θα ασχοληθούμε. Υπάρχουν όμως και εξαιρέσεις όπως π.χ. πληθυσμοί που ακολουθούν την κατανομή Cauchy.

Τα χαρακτηριστικά των σημειακών εκτιμητριών μπορούν να συνοψισθούν στα εξής:

1. Υπάρχει κάποιο άγνωστο χαρακτηριστικό του πληθυσμού, ή μια παράμετρος, την οποία συμβολίζουμε με θ , η οποία θα πρέπει να εκτιμηθεί.
2. Για να βρούμε μια σημειακή εκτίμηση του θ , παίρνουμε ένα τυχαίο δείγμα από n παρατηρήσεις X_1, X_2, \dots, X_n από τον πληθυσμό. Στην συνέχεια χρησιμοποιούμε μια στατιστική συνάρτηση $\hat{\theta}$, η οποία είναι συνάρτηση των παρατηρήσεων του δείγματος, έστω $\hat{\theta} = h(X_1, X_2, \dots, X_n)$, για να εκτιμήσουμε το θ .
3. Πριν να γίνει η επιλογή του δείγματος, τα X_1, X_2, \dots, X_n είναι τυχαίες μεταβλητές. Επομένως, το $\hat{\theta}$ είναι επίσης μια τυχαία μεταβλητή. Η κατανομή πιθανότητας του $\hat{\theta}$ είναι αυτή που ονομάζουμε *δειγματική κατανομή* του $\hat{\theta}$.

ΙΔΙΟΤΗΤΕΣ ΤΩΝ ΣΗΜΕΙΑΚΩΝ ΕΚΤΙΜΗΤΡΙΩΝ

Μια ιδιότητα η οποία είναι ιδιαίτερα επιθυμητή για μια εκτιμήτρια είναι αυτή της αυξανόμενης ακρίβειας. Αυτό γιατί εάν η διασπορά της δειγματικής κατανομής μιας εκτιμήτριας ελαττώνεται όσο αυξάνει το n , είναι απαραίτητο η κεντρική τιμή της να προσεγγίζει την εκτιμώμενη παράμετρο. Διαφορετικά, η εκτιμήτρια θα έχει τιμές που θα διαφέρουν συστηματικά από την τιμή της παραμέτρου.

Συνέπεια (*consistency*)

Ορισμός: Μια εκτιμήτρια (μια στατιστική συνάρτηση) $\hat{\theta}_n$ είναι μια *συνεπής εκτιμήτρια* μιας παραμέτρου θ του πληθυσμού (*consistent estimator*) αν για μια οποιαδήποτε πολύ μικρή θετική τιμή ε :

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$$

Ο ορισμός αυτός δηλώνει ότι με πιθανότητα 1 (δηλαδή τελικά ή σχεδόν με βεβαιότητα (*almost surely*)), μια συνεπής εκτιμήτρια της παραμέτρου θα δώσει μια εκτίμηση πάρα πολύ κοντά στην πραγματική τιμή της παραμέτρου (με απόκλιση $\pm \varepsilon$) όταν το n είναι μεγάλο. Αυτό εναλλακτικά σημαίνει ότι, για μια καθορισμένη ποσότητα ε οσοδήποτε μικρή, μπορούμε να βρούμε ένα δείγμα τόσο μεγάλο, ώστε για όλα τα δείγματα με μεγαλύτερο μέγεθος από αυτό, η πιθανότητα ότι η εκτιμήτρια $\hat{\theta}_n$ διαφέρει από την πραγματική τιμή της παραμέτρου περισσότερο από την ποσότητα ε είναι πολύ κοντά στο 0. Στην θεωρία πιθανοτήτων λέμε ότι η εκτιμήτρια $\hat{\theta}_n$ συγκλίνει κατά πιθανότητα (*converges in probability*) στο θ .

Αυτό σημαίνει ότι μια εκτιμήτρια $\hat{\theta}_n$ είναι συνεπής εκτιμήτρια μιας παραμέτρου θ αν συγκλίνει κατά πιθανότητα στο θ .

Σημείωση: Ο αρχικός ορισμός της συνέπειας δόθηκε από τον Fisher το 1921, ο οποίος όρισε ως συνεπή την εκτιμήτρια εκείνη που, αν η τιμή της υπολογισθεί από ολόκληρο τον πληθυσμό θεωρούμενο ως δείγμα, η τιμή αυτή θα ισούται με την τιμή της υπό εκτίμηση παραμέτρου.

Παράδειγμα: Ο δειγματικός μέσος είναι μια συνεπής εκτιμήτρια της μέσης τιμής ενός πληθυσμού.

Αμεροληψία (*unbiasedness*)

Αν θεωρήσουμε την δειγματική κατανομή μιας εκτιμήτριας $\hat{\theta}_n$ μιας παραμέτρου θ , από τα προηγούμενα γίνεται αντιληπτό ότι, αν η εκτιμήτρια είναι συνεπής, η κατανομή της, για μεγάλα δείγματα, θα έχει μια κεντρική τιμή στην περιοχή της υπό εκτίμηση παραμέτρου που θα πλησιάζει πολύ την πραγματική τιμή της. Είναι λογικό στην περίπτωση αυτή να διαλέξουμε από την κλάση των συνεπών εκτιμητριών εκείνη της οποίας η κεντρική τιμή θα ισούται ακριβώς με την υπό εκτίμηση παράμετρο θ όχι μόνο για μεγάλα δείγματα ($n \rightarrow \infty$), αλλά για όλα τα δείγματα.

Ορισμός: Μια εκτιμήτρια $\hat{\theta}_n$ θα λέγεται *αμερόληπτη (unbiased)* αν ο μέσος της δειγματικής της κατανομής ισούται με την υπό εκτίμηση παράμετρο θ του πληθυσμού. Αν δηλαδή:

$$E(\hat{\theta}_n) = \theta$$

Σημείωση: Αν μια εκτιμήτρια $\hat{\theta}$ δεν είναι αμερόληπτη θα λέμε ότι είναι μεροληπτική (*biased*) και η ποσότητα μεροληπτικότητάς της (*bias*) θα είναι:

$$\text{bias} = E(\hat{\theta}) - \theta$$

Σημείωση: Ο λόγος που επιλέγουμε την μέση τιμή ως κριτήριο μεροληπτικότητας ή αμεροληψίας είναι καθαρά λόγος ευκολίας. Θα ήταν εξίσου δυνατό να είχε επιλεγεί για τον ορισμό της αμερόληπτης εκτιμήτριας η διάμεσος ή η επικρατούσα τιμή. Η μέση τιμή όμως χρησιμοποιείται κυρίως λόγω της ευκολίας στις μαθηματικές πράξεις.

Παράδειγμα: Ο δειγματικός μέσος

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

είναι μια αμερόληπτη εκτιμήτρια του μέσου ενός πληθυσμού, διότι

$$E(\bar{X}) = \frac{\sum E(X_i)}{n} = \frac{n\mu}{n} = \mu$$

Αντίθετα, η δειγματική διασπορά

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$$

δεν είναι αμερόληπτη εκτιμήτρια της διασποράς σ^2 ενός πληθυσμού, διότι

$$E(S^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

Επομένως, αμερόληπτη εκτιμήτρια του σ^2 είναι η στατιστική συνάρτηση

$$\frac{nS^2}{n-1} = \frac{\sum (X_i - \bar{X})^2}{(n-1)}$$

Αυτός είναι και ο λόγος που οι περισσότεροι συγγραφείς στην Στατιστική ορίζουν την διασπορά του δείγματος την στατιστική συνάρτηση

$$S^{*2} = \frac{\sum(X_i - \bar{X})^2}{n - 1}$$

(την αμερόληπτη δηλαδή εκτιμήτρια S^{*2} του σ^2 και όχι το S^2 όπως ορίστηκε παραπάνω).

Σημείωση: Από την προηγούμενη συζήτηση προκύπτει ότι συνεπείς εκτιμήτριες δεν είναι υποχρεωτικά και αμερόληπτες εκτιμήτριες. Υπάρχουν επίσης και αμερόληπτες εκτιμήτριες οι οποίες δεν είναι συνεπείς. Καμιά από τις δύο αυτές ιδιότητες δεν συνεπάγεται υποχρεωτικά την άλλη. Παρ' όλα αυτά, μια συνεπής εκτιμήτρια της οποίας η ασυμπτωτική κατανομή έχει πεπερασμένη μέση τιμή είναι και ασυμπτωτικά αμερόληπτη.

Υπάρχουν επίσης και περιπτώσεις στις οποίες δεν υφίσταται αμερόληπτη εκτιμήτρια μιας παραμέτρου.

Αποτελεσματικότητα (*efficiency*)

Αν δύο εκτιμήτριες έχουν μικρή ή καθόλου μεροληπτικότητα είναι φυσικό να προτιμήσει κανείς μεταξύ τους την εκτιμήτρια εκείνη που έχει την μικρότερη διασπορά για το συγκεκριμένο μέγεθος δείγματος (διότι τα ενδεχόμενα αποτελέσματα που θα δίνει θα τείνουν να βρίσκονται πλησιέστερα προς την παράμετρο του πληθυσμού που θέλουμε να εκτιμήσουμε).

Η παρατήρηση αυτή οδηγεί στον ορισμό της *σχετικής αποτελεσματικότητας* (*relative efficiency*).

Ορισμός: Εάν δύο εκτιμήτριες $\hat{\theta}_1$ και $\hat{\theta}_2$ που προέρχονται από δείγματα του αυτού μεγέθους είναι αμερόληπτες εκτιμήτριες της παραμέτρου θ , θα λέμε ότι η μία από αυτές έχει μεγαλύτερη *σχετική αποτελεσματικότητα* (*relative efficiency*) αν έχει μικρότερη διακύμανση.

Αν δηλαδή

$$V(\hat{\theta}_1) < V(\hat{\theta}_2) \text{ και } E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$$

θα λέμε ότι η εκτιμήτρια $\hat{\theta}_1$ είναι σχετικά πιο αποτελεσματική από την εκτιμήτρια $\hat{\theta}_2$ στην εκτίμηση του θ .

Παράδειγμα: Έστω ότι μας ενδιαφέρει να εκτιμήσουμε τον μέσο χρόνο που ένα νέο προϊόν είναι δυνατόν να διατηρηθεί πριν χρησιμοποιηθεί από τον καταναλωτή. Από προηγούμενες εμπειρίες, είναι δυνατόν να υποθέσουμε ότι η κατανομή του χρόνου που προϊόντα, εν γένει, διατηρούνται πριν την κατανάλωσή τους είναι κανονική.

Το ερώτημα είναι αν θα πρέπει να χρησιμοποιήσουμε τον μέσο χρόνο που διατηρούνται τα προϊόντα της συγκεκριμένης κατηγορίας ενός τυχαίου δείγματος από τέτοια προϊόντα ως εκτιμήτρια του μ ή, εναλλακτικά, αν είναι προτιμότερο να χρησιμοποιήσουμε την διάμεσο του δείγματος. Είναι γνωστό ότι, για τυχαία δειγματοληψία από ένα κανονικό πληθυσμό, τόσο το \bar{X} όσο και η διάμεσος είναι αμερόληπτες εκτιμήτριες του μ .

Επίσης, από την στατιστική θεωρία προκύπτει ότι για δοθέν δείγμα μεγέθους n

$$V(\bar{X}) = \frac{\sigma^2}{n} \text{ και } V(\text{Διαμέσου}) \cong 1.57 \left(\frac{\sigma^2}{n} \right)$$

Επομένως το \bar{X} είναι μία σχετικά πιο αποτελεσματική εκτιμήτρια από ότι η διάμεσος για την εκτίμηση του μέσου.

Κάτω φράγμα των Cramér-Rao: Έχει αποδειχθεί ότι, κάτω από αρκετά γενικές συνθήκες, υπάρχει ένα κατώτερο φράγμα για την διακύμανση μιας αμερόληπτης εκτιμήτριας (ένα φράγμα δηλαδή από το οποίο η διασπορά μιας αμερόληπτης εκτιμήτριας δεν μπορεί να πάρει μικρότερη τιμή). Η ανισότητα από την οποία προκύπτει ο προσδιορισμός του ορίου αυτού, όσον αφορά την διακύμανση μιας αμερόληπτης εκτιμήτριας, είναι γνωστή ως ανισότητα των *Cramér-Rao* (από τα ονόματα των δύο επιστημόνων οι οποίοι και το ανακάλυψαν το 1946 και το 1948, αντίστοιχα). Συγκεκριμένα, μπορεί να αποδειχθεί ότι αν X_1, X_2, \dots, X_n είναι ανεξάρτητες παρατηρήσεις πάνω σε μια μεταβλητή X με συνάρτηση πιθανότητας (ή συνάρτηση πυκνότητας πιθανότητας στην συνεχή περίπτωση) $P_X(x;\theta)$ και

$\hat{\theta} \equiv \hat{\theta}(X_1, X_2, \dots, X_n)$ είναι μια αμερόληπτη εκτιμήτρια της παραμέτρου θ , τότε

$$V(\hat{\theta}) \geq \frac{1}{E \left[\left(\frac{\partial}{\partial \theta} \ln P_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta) \right)^2 \right]}$$

$$\equiv \frac{1}{E \left[\left(\frac{\partial}{\partial \theta} \ln \prod_{i=1}^n P_{X_i}(x_i; \theta) \right)^2 \right]} = \frac{1}{n E \left[\left(\frac{\partial}{\partial \theta} \ln P_X(x; \theta) \right)^2 \right]}$$

Το όριο αυτό ονομάζεται πολλές φορές *όριο ελάχιστης διακύμανσης (minimum variance bound)* της εκτιμήτριας του θ ή *κάτω φράγμα των Cramér-Rao*.

Η ποσότητα

$$E \left[\left(\frac{\partial}{\partial \theta} \ln P_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta) \right)^2 \right]$$

ονομάστηκε από τον Fisher *ποσότητα πληροφορίας (amount of information)* σχετικά με την παράμετρο θ , η οποία περιέχεται στις παρατηρήσεις X_1, X_2, \dots, X_n και συχνά συμβολίζεται με I_θ . Για την ποσότητα αυτή, στην βιβλιογραφία, έχει επικρατήσει ο όρος *πληροφορία του Fisher (Fisher's information)*.

Είναι προφανές ότι η ποσότητα πληροφορίας που περιέχεται σε μια μοναδική παρατήρηση X , είναι

$$i_\theta = E \left[\left(\frac{\partial}{\partial \theta} \ln P_X(x; \theta) \right)^2 \right]$$

Επομένως, η ανισότητα των Cramér-Rao γράφεται

$$V(\hat{\theta}) \geq I_\theta^{-1} \equiv (ni_\theta)^{-1}$$

Ορισμός: Μια εκτιμήτρια της οποίας η διασπορά επιτυγχάνει αυτό το κατώτερο όριο για όλα τα θ ονομάζεται *αμερόληπτη εκτιμήτρια ελάχιστης διασποράς (minimum variance unbiased estimator)*.

Οι περιπτώσεις ύπαρξης τέτοιων εκτιμητριών είναι ελάχιστες και αφορούν κατανομές για τις οποίες η $\frac{\partial}{\partial \theta} \ln P_X(x; \theta)$ έχει την μορφή

$\alpha(\theta)[\hat{\theta}(\underline{x}) - \theta]$, όπου τότε $\alpha(\theta) = I_\theta$, όπως στην περίπτωση n ανεξάρτητων δοκιμών Bernoulli, όπου

$$P_{\underline{X}}(\underline{x}; \theta) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{n-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}$$

και, επομένως,

$$\frac{\partial}{\partial \theta} \ln P_{\underline{X}}(\underline{x}; \theta) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1-\theta} = \frac{n}{\theta(1-\theta)} (\hat{\theta}(\underline{x}) - \theta)$$

με $\hat{\theta}(\underline{x}) = \sum_{i=1}^n x_i / n$ (=το ποσοστό επιτυχιών στις n δοκιμές).

Η παραπάνω συνθήκη έχει παρ' όλα αυτά πολλές πρακτικές εφαρμογές, γιατί, στην περίπτωση μεγάλων δειγμάτων, ικανοποιείται κατά προσέγγιση και, επομένως, μπορούμε να έχουμε εκτιμήτριες που είναι περίπου ή σχεδόν αμερόληπτες με διασπορές σχεδόν ίσες με το φράγμα των Cramér-Rao.

Η σύγκριση της διασποράς μιας αμερόληπτης εκτιμήτριας $\hat{\theta}$ με την ποσότητα I_θ^{-1} , δηλαδή με μια ποσότητα που δεν μπορεί να βελτιωθεί περισσότερο, οδηγεί στην έννοια της *αποτελεσματικότητας* της εκτιμήτριας $\hat{\theta}$.

Ορισμός: Ορίζουμε ως *αποτελεσματικότητα (efficiency)* μιας αμερόληπτης εκτιμήτριας $\hat{\theta}$ μιας πραγματικής παραμέτρου θ και συμβολίζουμε με $\text{eff}(\hat{\theta})$ τον λόγο

$$\text{eff}(\hat{\theta}) = \frac{I_\theta^{-1}}{V(\hat{\theta})}$$

Εάν η τιμή της $\text{eff}(\hat{\theta})$ είναι πολύ κοντά στην μονάδα, αυτό συνηγορεί υπέρ της επιλογής της $\hat{\theta}$ ως εκτιμήτριας.

Σημείωση: Ο υπολογισμός της ποσότητας I_θ γίνεται ευκολότερος αν χρησιμοποιηθεί η σχέση

$$E \left[\frac{\partial}{\partial \theta} \ln P_{\underline{X}}(\underline{x}; \theta) \right]^2 = -E \left(\frac{\partial^2}{\partial \theta^2} \ln P_{\underline{X}}(\underline{x}; \theta) \right), \text{ όπου } \underline{X} = (X_1, X_2, \dots, X_n)$$

Επάρκεια (*sufficiency*)

Το γενικό πρόβλημα της στατιστικής ανάλυσης είναι, ως γνωστόν, η περιγραφή ενός πληθυσμού που εκφράζεται με μια τυχαία μεταβλητή X και ο προσδιορισμός της οικογένειας όλων των πιθανών κατανομών της τυχαίας μεταβλητής X .

Εκείνο που ενδιαφέρει είναι η επιλογή της πιο κατάλληλης κατανομής από την οικογένεια αυτή των κατανομών για να περιγράψει καλύτερα τον υπό μελέτη πληθυσμό.

Η επιλογή αυτή είναι ισοδύναμη με την επιλογή (εκτίμηση) μιας, ή περισσοτέρων, παραμέτρων του υπό μελέτη πληθυσμού. Πολλές φορές συμβαίνει να παρατηρείται το φαινόμενο ότι ένα μέρος από τα δεδομένα που περιλαμβάνονται σε ένα τυχαίο δείγμα δεν προσφέρει (δεν περιέχει) καμιά πληροφορία για την άγνωστη κατανομή (για την άγνωστη παράμετρο που θέλουμε να εκτιμήσουμε). Είναι δε δυνατόν όλες οι πληροφορίες για την υπό μελέτη κατανομή (παράμετρο) να περιέχονται σε μια συνάρτηση T παρατηρήσεων του δείγματος. Στην περίπτωση αυτή το δείγμα μπορεί να αντικατασταθεί από την συνάρτηση T χωρίς απώλεια οποιασδήποτε πληροφορίας.

Ορισμός: Μια στατιστική συνάρτηση T ονομάζεται *επαρκής* (*sufficient*) για την παράμετρο θ ενός πληθυσμού (ή, ισοδύναμα, για τον πληθυσμό που εκφράζει η τυχαία μεταβλητή X , ή ισοδύναμα, για την οικογένεια όλων των δυνατών κατανομών της X όσον αφορά το θ) αν η στατιστική αυτή συνάρτηση T περιέχει όλες τις πληροφορίες στο δείγμα γύρω από την παράμετρο θ (ή, ισοδύναμα, αν η δεσμευμένη κατανομή του X δοθέντος ($T=t$) είναι ανεξάρτητη από το θ για όλα τα t).

Για να αντιληφθούμε καλύτερα την έννοια της επάρκειας, ας υποθέσουμε ότι ένας ερευνητής αναφέρει την τιμή της στατιστικής συνάρτησης T , αλλά όταν του ζητηθούν τα πλήρη δεδομένα του δείγματος παραδέχεται ότι δεν τα έχει κρατήσει.

Σε μια προσπάθεια ανάκτησης των απωλεσθέντων δεδομένων, ο ερευνητής είναι δυνατόν να χρησιμοποιήσει ένα τυχαίο μηχανισμό (όπως, για παράδειγμα, ένα πίνακα τυχαίων αριθμών) για να

κατασκευάσει μία τυχαία ποσότητα X' σύμφωνα με την δεσμευμένη κατανομή του X δοθέντος ($T=t$). (Αυτό βέβαια δεν είναι δυνατόν εάν η δεσμευμένη αυτή κατανομή εξαρτάται από την άγνωστη παράμετρο θ). Λόγω της ιδιότητας της T , η κατανομή της X' είναι η ίδια με την κατανομή της X ανεξάρτητα από την τιμή της παραμέτρου θ . Επομένως, από την γνώση της στατιστικής συνάρτησης T και μόνο είναι δυνατόν να κατασκευασθεί μια ποσότητα X' , η οποία ισοδυναμεί με την αρχική X .

Οι ποσότητες X και X' , δοθέντος ότι έχουν την ίδια κατανομή για όλα τα θ , δίνουν ακριβώς τις ίδιες πληροφορίες για την παράμετρο θ .

Με τον τρόπο αυτό είναι προφανές ότι μια επαρκής στατιστική συνάρτηση (μια επαρκής εκτιμήτρια) επιτυγχάνει μια απλοποίηση (ένα περιορισμό ή μία συρρίκνωση) των δεδομένων του δείγματος χωρίς την απώλεια οποιασδήποτε πληροφορίας που το δείγμα παρέχει. Η ιδιότητα αυτή βέβαια ισχύει εφόσον περιοριζόμαστε στην ίδια οικογένεια κατανομών, τα μέλη της οποίας διαφοροποιούνται μόνο από την τιμή της υπό εκτίμηση παραμέτρου.

Παράδειγμα: Έστω ότι θέλουμε να μελετήσουμε τον αριθμό των ατυχημάτων σε ένα συγκεκριμένο σημείο της εθνικής οδού. Είναι γνωστό ότι, κάτω από ορισμένες προϋποθέσεις, είναι δυνατόν να υποθέσουμε ότι ο αριθμός των ατυχημάτων στην μονάδα του χρόνου ακολουθεί την κατανομή Poisson με παράμετρο λ . (Το λ εκφράζει τον μέσο αριθμό των ατυχημάτων στην μονάδα του χρόνου).

Αν θεωρήσουμε ως μονάδα χρόνου την μία ημέρα ας υποθέσουμε ότι σε δύο ημέρες που έχουν επιλεγεί τυχαία παρατηρήθηκαν X_1 και X_2 ατυχήματα αντίστοιχα στο συγκεκριμένο αυτό σημείο της εθνικής οδού.

Σε κάποιο ερευνητή που ενδιαφέρεται να μελετήσει τα ατυχήματα στο σημείο αυτό της εθνικής οδού δίνεται από την αστυνομία η πληροφορία ότι ο συνολικός αριθμός των ατυχημάτων τις δύο αυτές ημέρες ήταν 10, ($X_1 + X_2 = 10$).

Είναι δυνατόν, με βάση την πληροφορία αυτή, να εκτιμήσουμε την τιμή της παραμέτρου λ ; (με άλλα λόγια να καθορίσουμε την

κατανομή του αριθμού των ατυχημάτων στο συγκεκριμένο σημείο της εθνικής οδού;)

Λύση: Είναι προφανές ότι η στατιστική συνάρτηση του δείγματος

$$T = X_1 + X_2$$

αποτελεί ένα περιορισμό του αρχικού δείγματος που είναι το ζευγάρι (X_1, X_2) .

Η T αποτελεί πράγματι μια στατιστική συνάρτηση που προέρχεται από το τυχαίο δείγμα X_1, X_2 και αποτελεί επίσης ένα περιορισμό του τυχαίου αυτού δείγματος (αφού γνώση των X_1, X_2 συνεπάγεται την T , ενώ αντίστροφα γνώση της T δεν συνεπάγεται γνώση των X_1, X_2).

Θα αποδείξουμε ότι η $T = X_1 + X_2$ αποτελεί μια επαρκή στατιστική συνάρτηση για την παράμετρο λ (ή ισοδύναμα για την κατανομή του αριθμού των ατυχημάτων στο συγκεκριμένο σημείο).

Η δεσμευμένη κατανομή των X_1 και X_2 δοθέντος ότι $T=10$ δίδεται από την σχέση:

$$P(X_1 = x_1, X_2 = x_2 | T = X_1 + X_2 = 10) = \frac{P(X_1 = x_1, X_2 = x_2, T = 10)}{P(T = 10)}$$

Δοθέντος ότι, όπως έχουμε υποθέσει, ο αριθμός των ατυχημάτων στο σημείο που μας ενδιαφέρει ακολουθεί την κατανομή Poisson με παράμετρο λ και δεδομένου ότι X_1 και X_2 είναι ανεξάρτητες μεταβλητές (αφού το δείγμα είναι τυχαίο), έχουμε ότι και οι τυχαίες μεταβλητές X_1 και X_2 ακολουθούν κατανομές Poisson με την ίδια παράμετρο του λ (αφού από τον ορισμό της κατανομής Poisson το λ εκφράζει τον μέσο αριθμό των ατυχημάτων ανά ημέρα).

Είναι προφανές ότι

$$P(X_1 = x_1, X_2 = x_2, T = 10) = P(X_1 = x_1, X_2 = x_2)$$

και λόγω της ανεξαρτησίας των X_1 και X_2 η παράσταση αυτή ισούται με:

$$P(X_1 = x_1)P(X_2 = x_2) = e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!}$$

$$= e^{-(\lambda+\lambda)} \frac{\lambda^{x_1} \lambda^{x_2}}{x_1! x_2!} = e^{-2\lambda} \frac{\lambda^{(x_1+x_2)}}{x_1! x_2!}$$

Εξάλλου:

$$P(T = 10) = P(X_1 + X_2 = 10) = e^{-2\lambda} \frac{(2\lambda)^{10}}{10!}$$

Έτσι, τελικά, η δεσμευμένη κατανομή X_1 και X_2 δοθέντος T είναι:

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2 | T = 10) &= \frac{e^{-2\lambda} \frac{\lambda^{10}}{x_1! x_2!}}{e^{-2\lambda} \frac{(2\lambda)^{10}}{10!}} = \frac{10!}{x_1! x_2!} \left(\frac{1}{2}\right)^{10} \\ &= \frac{10!}{x_1! (10 - x_1)!} \left(\frac{1}{2}\right)^{x_1} \left(\frac{1}{2}\right)^{10-x_1} \end{aligned}$$

Είναι προφανές ότι η δεσμευμένη αυτή κατανομή είναι μια διωνυμική κατανομή με παραμέτρους $n = 10$ και $p = 1/2$, ανεξάρτητη από το λ .

Επομένως, πράγματι η στατιστική συνάρτηση $T = X_1 + X_2$ είναι μία επαρκής συνάρτηση για την παράμετρο λ του πληθυσμού.

Η διωνυμική μορφή κατανομής στην οποία καταλήξαμε για την δεσμευμένη κατανομή των X_1 και X_2 δοθέντος του T επιβεβαιώνει την προηγούμενη παρατήρησή μας ότι, αν και η επαρκής στατιστική συνάρτηση είναι μια μείωση των παρατηρήσεων του δείγματος, παρ' όλα αυτά και επειδή ακριβώς είναι επαρκής, μας δίνει την δυνατότητα να ανακτήσουμε τις χαμένες "πληροφορίες" από το δείγμα.

Πράγματι, στο συγκεκριμένο πρόβλημα δοθέντος ότι $X_1 + X_2 = 10$ (ότι δηλαδή το άθροισμα του αριθμού των ατυχημάτων για τις δύο μέρες είναι 10), μπορούμε να παράγουμε το ζεύγος (X_1, X_2) , δηλαδή να ανακτήσουμε τις πληροφορίες και το δείγμα, προσομοιώνοντας μια διωνυμική κατανομή. (Η προσομοίωση αυτή μπορεί να γίνει π.χ. με το στρίψιμο ενός αμερόληπτου νομίσματος δέκα φορές, οπότε X_1 θα είναι ο αριθμός των φορών που το αποτέλεσμα ήταν γράμματα και

$X_2 = 10 - X_1$ θα είναι ο αριθμός των φορών που το αποτέλεσμα ήταν κορώνα).

Σημείωση: Για την κατανομή Poisson μπορεί να αποδειχθεί γενικότερα ότι, δοθέντος ενός τυχαίου δείγματος (X_1, X_2, \dots, X_n) από μια κατανομή Poisson, η στατιστική συνάρτηση $T(X_1, X_2, \dots, X_n) = X_1 + X_2 + \dots + X_n$ είναι επαρκής για την παράμετρο λ της κατανομής Poisson ή, ισοδύναμα, για την οικογένεια των κατανομών Poisson με παράμετρο λ .

Εύκολα αποδεικνύεται ότι στην γενική αυτή περίπτωση η δεσμευμένη κατανομή των X_1, X_2, \dots, X_n δοθέντος $T = X_1 + X_2 + \dots + X_n$ είναι η πολυωνυμική κατανομή με παραμέτρους $p_1 = p_2 = \dots = p_n = 1/n$. Προφανώς, η κοινή αυτή τιμή των παραμέτρων είναι ανεξάρτητη του λ .

Σημείωση: Το γεγονός ότι το άθροισμα των παρατηρήσεων του δείγματος είναι μια επαρκής συνάρτηση για την παράμετρο λ της Poisson εξηγείται εύκολα από το γεγονός ότι ως εκτιμητήρια του λ μπορεί να θεωρηθεί ο δειγματικός μέσος

$$\hat{\lambda} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Από αυτό φαίνεται ότι η γνώση του αθροίσματος των παρατηρήσεων του δείγματος είναι επαρκής για την εκτίμηση της παραμέτρου λ χωρίς να απαιτείται γνώση των μεμονωμένων παρατηρήσεων του δείγματος.

Το θεώρημα των Rao-Blackwell: Η έννοια της επάρκειας συνδέεται με το πρόβλημα της ύπαρξης αμερόληπτων εκτιμητριών ελάχιστης διασποράς. Συγκεκριμένα, αν έχουμε μια αμερόληπτη εκτιμήτρια $\tilde{\theta}$ μιας παραμέτρου θ και μια επαρκή στατιστική συνάρτηση T για την παράμετρο θ , τότε η αναζήτηση μιας “καλύτερης” (με την έννοια της μικρότερης διασποράς) αμερόληπτης εκτιμήτριας $\hat{\theta}$ μπορεί να περιορισθεί μεταξύ συναρτήσεων της T που έχουν την μορφή $\hat{\theta} \equiv \hat{\theta}(T) = E(\tilde{\theta} | T)$. Το αποτέλεσμα αυτό αποδείχθηκε από τον Rao το 1945 και τον Blackwell το 1947 και είναι γνωστό ως θεώρημα Rao-Blackwell.

Θεώρημα: (Rao-Blackwell): Έστω ότι $\tilde{\theta}$ είναι μια αμερόληπτη εκτιμήτρια μιας πραγματικής παραμέτρου θ . Τότε, αν T είναι μια επαρκής στατιστική συνάρτηση για την παράμετρο θ , η εκτιμήτρια

$$\hat{\theta} \equiv \hat{\theta}(T) = E(\tilde{\theta} | T)$$

είναι επίσης μια αμερόληπτη εκτιμήτρια της παραμέτρου θ με διασπορά που δεν υπερβαίνει την διασπορά της $\tilde{\theta}$, δηλαδή $V(\hat{\theta}) \leq V(\tilde{\theta})$.

Αν η $\hat{\theta} \equiv \hat{\theta}(T)$ είναι η μοναδική συνάρτηση της T που είναι αμερόληπτη εκτιμήτρια της θ , τότε η $\hat{\theta}(T)$ είναι μια *αμερόληπτη εκτιμήτρια ελάχιστης διασποράς (minimum variance unbiased estimator)* της παραμέτρου θ .

Αυτό μπορεί εύκολα να διαπιστωθεί ως εξής: Αν υπάρχει μια άλλη αμερόληπτη εκτιμήτρια του θ , έστω $\hat{\theta}_1$, τότε, σύμφωνα με το θεώρημα Rao-Blackwell, η $\hat{\theta}_2 = E(\hat{\theta}_1 | T)$ είναι επίσης αμερόληπτη εκτιμήτρια της θ με διασπορά $V(\hat{\theta}_2)$ μικρότερη ή ίση της διασποράς $V(\hat{\theta}_1)$ της $\hat{\theta}_1$. Είναι προφανές ότι η $\hat{\theta}_2 \equiv E(\hat{\theta}_1 | T)$ είναι συνάρτηση της T . Τότε όμως, αφού η $\hat{\theta} \equiv \hat{\theta}(T)$ είναι η μοναδική συνάρτηση της T που είναι αμερόληπτη εκτιμήτρια της θ , ισχύει αναγκαστικά ότι $\hat{\theta} = E(\hat{\theta}_1 | T) \equiv \hat{\theta}_2$. Επομένως, $V(\hat{\theta}) = V(\hat{\theta}_2) \leq V(\hat{\theta}_1)$. Άρα, η $\hat{\theta}$ έχει διασπορά μικρότερη από την διασπορά οποιασδήποτε άλλης αμερόληπτης εκτιμήτριας. Είναι, επομένως, αμερόληπτη εκτιμήτρια ελάχιστης διασποράς.

Το ερώτημα που προκύπτει είναι πώς μπορούμε να διαπιστώσουμε αν για μια επαρκή στατιστική συνάρτηση T υπάρχει μία και μόνο μία συνάρτησή της, $\hat{\theta}(T)$ που είναι αμερόληπτη εκτιμήτρια της θ . Μερικές φορές αυτό μπορούμε να το επιτύχουμε χρησιμοποιώντας την έννοια της *πληρότητας* της κατανομής της T (ή της οικογένειας κατανομών στην οποία ανήκει η κατανομή της T).

Η πλευρά αυτή του θέματος δεν αναπτύσσεται εδώ γιατί είναι πέρα από τους σκοπούς αυτού του βιβλίου. Ο ενδιαφερόμενος αναγνώστης μπορεί να βρει περισσότερες πληροφορίες σε εγχειρίδια

Θεωρητικής Στατιστικής (βλέπε, π.χ. Ε. Ξεκαλάκη (1993): *Εισαγωγή στην Θεωρητική Στατιστική*).

ΜΕΘΟΔΟΙ ΣΗΜΕΙΑΚΗΣ ΕΚΤΙΜΗΣΗΣ

Στην προηγούμενη ενότητα εξετάσαμε επιθυμητές ιδιότητες σημειακών εκτιμητριών. Δεν μιλήσαμε όμως για τρόπους εύρεσης σημειακών εκτιμητριών.

Υπάρχει μια σειρά μεθόδων που χρησιμοποιούνται για τον καθορισμό σημειακών εκτιμητριών.

Οι κυριότερες από τις μεθόδους αυτές είναι:

- α) Μέθοδος των ροπών (*method of moments*)
- β) Μέθοδος των ελαχίστων τετραγώνων (*method of least squares*)
- γ) Μέθοδος μέγιστης πιθανοφάνειας (*method of maximum likelihood*)
- δ) Minimax μέθοδος.

Καθεμιά από τις παραπάνω μεθόδους οδηγεί σε εκτιμήτριες που έχουν κάποιες επιθυμητές ιδιότητες. Σε πολλές περιπτώσεις, οι εκτιμήτριες που προκύπτουν από την εφαρμογή της μιάς μεθόδου συμπίπτουν με τις εκτιμήτριες που προκύπτουν από την εφαρμογή μιας άλλης μεθόδου.

Στη συνέχεια, θα αναφερθούμε περιληπτικά στην μέθοδο των ροπών και την μέθοδο των ελαχίστων τετραγώνων και θα μελετήσουμε αναλυτικά την μέθοδο μέγιστης πιθανοφάνειας. Δεν θα αναφερθούμε στην μέθοδο minimax γιατί αυτή αναφέρεται σε μια άλλη προσέγγιση της Στατιστικής (Στατιστική Θεωρία κατά Bayes), η οποία θα μελετηθεί χωριστά.

Μέθοδος των Ροπών

Όπως είδαμε στο πρώτο μέρος, και γνωρίζουμε από τις πιθανότητες (βλέπε, για παράδειγμα, Ε. Ξεκαλάκη και Ι. Πανάρετος *Πιθανότητες και Στοιχεία Στοχαστικών Ανελιζέων*, Κεφάλαιο 5), ως ροπή r τάξης μιας τυχαίας μεταβλητής ορίζεται η συνάρτηση

$$\mu_r = E(X^r)$$

με την προϋπόθεση βέβαια ότι η μέση αυτή τιμή υπάρχει.

Είναι προφανές ότι, για ένα τυχαίο δείγμα X_1, X_2, \dots, X_n από ένα πληθυσμό X , μπορούμε να ορίσουμε αντίστοιχα την ροπή r τάξης του δείγματος ως

$$m_r = \frac{\sum_{i=1}^n X_i^r}{n}$$

Η μέθοδος των ροπών, η οποία είναι και η παλαιότερη από τις γνωστές μεθόδους, συνίσταται στην εξίσωση μερικών από τις πρώτες ροπές του πληθυσμού με τις τιμές των αντιστοιχών ροπών ενός τυχαίου δείγματος. Σύμφωνα με τη μέθοδο αυτή, θεωρούμε τόσες ροπές (τόσες εξισώσεις) όσος είναι και ο αριθμός των παραμέτρων που θέλουμε να εκτιμήσουμε. Με την εξίσωση των ροπών του πληθυσμού με τις αντίστοιχες ροπές του δείγματος θα έχουμε ένα σύστημα k εξισώσεων με k αγνώστους (όπου k είναι ο αριθμός των παραμέτρων που θέλουμε να εκτιμήσουμε). Θα έχουμε δηλαδή,

$$\mu_r = m_r, r = 1, 2, \dots, k$$

Η λύση του συστήματος αυτού των εξισώσεων ως προς τις άγνωστες παραμέτρους θα δώσει τις εκτιμήτριες της μεθόδου των ροπών.

Παράδειγμα: Έστω ότι θέλουμε να εκτιμήσουμε τις παραμέτρους μ και σ^2 της κανονικής κατανομής με τη μέθοδο των ροπών.

Έχουμε ότι

$$\mu_1 = E(X) = \mu \quad \text{και} \quad \mu_2 = E(X^2)$$

Θεωρούμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_n . Οι αντίστοιχες δειγματικές ροπές θα είναι

$$m_1 = \frac{\sum_{i=1}^n X_i}{n} = \bar{X} \quad \text{και} \quad m_2 = \frac{\sum_{i=1}^n X_i^2}{n}$$

Επομένως, θεωρούμε την εξίσωση

$$\mu = \bar{X} \quad \text{και} \quad \text{επειδή} \quad \sigma^2 = E(X^2) - [E(X)]^2 = \mu_2 - \mu^2, \text{ την εξίσωση}$$

$$\mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

Λύνοντας το σύστημα των δύο αυτών εξισώσεων ως προς μ και σ^2 έχουμε

$$\hat{\mu} = \bar{X}$$

και

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \\ &= \frac{1}{n} (\sum_{i=1}^n X_i^2 - n\bar{X}^2) \\ &= \frac{1}{n} (\sum_{i=1}^n X_i^2 + n\bar{X}^2 - 2n\bar{X}^2) \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

Δηλαδή τελικά,

$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$
--

Μέθοδος των Ελαχίστων Τετραγώνων (Least Squares Method)

Η μέθοδος των ελαχίστων τετραγώνων βασίζεται, όπως προκύπτει και από το όνομά της, στον καθορισμό εκτιμητριών με ελαχιστοποίηση κάποιου αθροίσματος τετραγώνων.

Για παράδειγμα, όταν μας ενδιαφέρει η εκτίμηση της μέσης τιμής μ ενός πληθυσμού με βάση ένα τυχαίο δείγμα X_1, X_2, \dots, X_n , επιλέγουμε ως σημειακή εκτιμήτρια την τιμή $\hat{\mu} = h(X_1, X_2, \dots, X_n)$ του μ που για το δοθέν δείγμα ελαχιστοποιεί την παράσταση

$$Q = \sum_{i=1}^n (X_i - \mu)^2$$

Στην συγκεκριμένη περίπτωση, είναι προφανές ότι η εκτιμήτρια ελαχίστων τετραγώνων του μ είναι

$$\hat{\mu} = \bar{X}$$

Γενικότερα, η μέθοδος των ελαχίστων τετραγώνων χρησιμοποιείται, κυρίως, όταν θέλουμε να εκτιμήσουμε μία παράμετρο θ (ή ένα διάνυσμα παραμέτρων) με βάση ένα σύνολο n παρατηρήσεων από ένα τυχαίο δείγμα X_1, X_2, \dots, X_n , οι οποίες συνδέονται με την υπό εκτίμηση παράμετρο θ μέσω των μέσων τιμών

$$E(X_i) = f(\theta)$$

Στις περιπτώσεις αυτές επιλέγουμε ως εκτιμήτρια $\hat{\theta}$ της παραμέτρου θ εκείνη που ελαχιστοποιεί το άθροισμα των τετραγωνικών αποκλίσεων μεταξύ των τιμών X και των αντιστοίχων εκτιμωμένων τιμών. Δηλαδή, αυτή που ελαχιστοποιεί την παράσταση:

$$\begin{aligned} S &= \sum_{i=1}^n [X_i - E(X_i)]^2 \\ &= \sum_{i=1}^n [X_i - f(\theta)]^2 \end{aligned}$$

Η μέθοδος των ελαχίστων τετραγώνων χρησιμοποιείται κυρίως, για την εκτίμηση των παραμέτρων της ανάλυσης παλινδρόμησης.

Μέθοδος Μέγιστης Πιθανοφάνειας

Έστω ότι το μοντέλο πιθανότητας (*probability model*) για κάποιο τυχαίο πείραμα αναφέρεται σε κάποια (περιέχει κάποια) παράμετρο θ .

Στην συνέχεια, το πείραμα εκτελείται και παρατηρείται ότι έχει συμβεί κάποιο ενδεχόμενο (*event*) E . (Αυτό σημαίνει ότι έχουν συγκεντρωθεί κάποια δεδομένα). Αυτό που ενδιαφέρει είναι, με την χρήση των δεδομένων αυτών, να εκτιμηθεί η τιμή της παραμέτρου θ .

Με χρησιμοποίηση του συγκεκριμένου πιθανοθεωρητικού μοντέλου (που έχουμε υποθέσει ότι ισχύει) και με τους νόμους των

πιθανοτήτων, είναι δυνατόν να καθορίσουμε την πιθανότητα (πιθανοφάνεια) του ενδεχομένου E που έχουμε παρατηρήσει.

Η πιθανότητα αυτή θα είναι, συνήθως, συνάρτηση της άγνωστης παραμέτρου θ , έστω $P(E; \theta)$.

Θα πρέπει τότε να υπάρχουν κάποιες τιμές του θ για τις οποίες το παρατηρηθέν ενδεχόμενο E είναι αρκετά πιθανό (έχει μεγάλη πιθανότητα να συμβεί) και άλλες για τις οποίες είναι ελάχιστα πιθανό να συμβεί.

Είναι, επομένως, λογικό να επιλέξουμε ως εκτιμητήρια του θ μια τιμή του θ για την οποία το ενδεχόμενο E είναι περισσότερο πιθανό να συμβεί παρά μια άλλη για την οποία το E είναι λίγο πιθανό να συμβεί.

Με άλλα λόγια, τιμές του θ για τις οποίες το ενδεχόμενο E έχει μία σχετικά υψηλή πιθανότητα να συμβεί είναι προτιμότερες από άλλες τιμές του θ για τις οποίες το E είναι ελάχιστα πιθανό.

Συνήθως, υπάρχει μια μοναδική τιμή του θ η οποία μεγιστοποιεί την συνάρτηση $P(E; \theta)$. Η τιμή αυτή συμβολίζεται με $\hat{\theta}$ και ονομάζεται *εκτιμητήρια μέγιστης πιθανοφάνειας* του θ (*EMΠ*) (*maximum likelihood estimator of θ (MLE)*).

Η EMΠ (MLE) του θ είναι επομένως η τιμή εκείνη του θ για την οποία το παρατηρηθέν ενδεχόμενο E έχει τη μεγαλύτερη δυνατή πιθανότητα κάτω από το συγκεκριμένο μοντέλο.

Σημείωση: Η συνάρτηση πιθανοφάνειας (*likelihood function*) ορίζεται πολλές φορές ως ένα πολλαπλάσιο της $P(E; \theta)$, δηλαδή

$$L(\theta) = k P(E; \theta),$$

όπου k είναι μια οποιαδήποτε σταθερά σε σχέση με το θ . Το k δηλαδή, δεν είναι συνάρτηση του θ . (Μπορεί όμως να είναι συνάρτηση των δεδομένων).

Ο φυσικός λογάριθμος της συνάρτησης πιθανοφάνειας (*log likelihood function*)

$$\ell(\theta) = \ln L(\theta)$$

έχει την ιδιότητα ότι η τιμή του θ που μεγιστοποιεί την $\ell(\theta)$ μεγιστοποιεί και την $L(\theta)$ (και επομένως και την $P(E; \theta)$).

Επομένως, η ΕΜΠ (MLE) $\hat{\theta}$ της θ είναι αυτή που μεγιστοποιεί την $L(\theta)$ ή, ισοδύναμα, την $\ell(\theta)$.

Ο παραμετρικός χώρος Θ των τιμών μιας παραμέτρου θ , όταν μιλάμε για την εκτίμηση μιας μόνο παραμέτρου, είναι ένα διάστημα πραγματικών τιμών. Επίσης η πρώτη και η δεύτερη παράγωγος

$$\ell'(\theta) = \frac{\partial}{\partial \theta} \ell(\theta) \text{ και } \ell''(\theta) = \frac{\partial^2}{\partial \theta^2} \ell(\theta)$$

θα υπάρχουν σε εσωτερικά σημεία του Θ .

Στην περίπτωση αυτή η ΕΜΠ (MLE) θα προσδιορίζεται, συνήθως, ως η ρίζα της εξίσωσης μεγίστης πιθανοφάνειας:

$$\ell'(\theta) = 0$$

Σε μερικές απλές περιπτώσεις (όπως αυτή της διωνυμικής και κανονικής κατανομής πληθυσμού), η εξίσωση αυτή μπορεί να λυθεί αλγεβρικά και να δώσει κάποιο μαθηματικό τύπο για την $\hat{\theta}$. Σε περισσότερο πολύπλοκες περιπτώσεις όμως είναι συνήθως απαραίτητο να λυθεί η εξίσωση αυτή αριθμητικά, ή με την βοήθεια του υπολογιστή.

Μια ρίζα της παραπάνω εξίσωσης για την οποία $\ell''(\theta) < 0$ είναι σημείο τοπικού μέγιστου.

Πολλές φορές ως ρίζες της εξίσωσης $\ell'(\theta) = 0$ προκύπτουν και σημεία που είναι τοπικά ελάχιστα, ή και σημεία καμπής (*points of inflection*).

Είναι επομένως απαραίτητο να καθορίσουμε το πρόσημο της δεύτερης παραγώγου της συνάρτησης αυτής, ή εναλλακτικά, να επιβεβαιώσουμε ότι η ρίζα που βρήκαμε είναι ένα τοπικό μέγιστο.

Υπάρχουν εξάλλου και περιπτώσεις στις οποίες δεν είναι δυνατόν να υπολογισθεί το $\hat{\theta}$ με λύση της εξίσωσης μεγίστης πιθανοφάνειας. Για παράδειγμα, το απόλυτο μέγιστο (*overall maximum*) της συνάρτησης πιθανοφάνειας είναι δυνατόν να είναι

σημείο (να συμβαίνει) στο σύνορο (*boundary*) του παραμετρικού χώρου Θ και επομένως η $\ell'(\theta)=0$ να μην ισχύει στο μέγιστο. Επίσης, αν οι τιμές του θ περιορίζονται σε ένα σύνολο διακριτών τιμών (όπως π.χ. οι ακέραιοι αριθμοί), η εξίσωση $\ell'(\theta)=0$ δεν έχει εφαρμογή.

Παράδειγμα: Έστω ότι μας ενδιαφέρει να εκτιμήσουμε το ποσοστό των ανθρώπων ενός μεγάλου ομογενούς πληθυσμού που πάσχουν από AIDS. Για τον λόγο αυτό επιλέγεται ένα τυχαίο δείγμα n ατόμων του πληθυσμού αυτού. Τα άτομα αυτά υποβάλλονται στο σχετικό τεστ. Έστω ότι από τον έλεγχο αυτό προκύπτει ότι x από τα άτομα αυτά πάσχουν από την ασθένεια.

Λύση: Λόγω της υπόθεσης ότι ο πληθυσμός είναι μεγάλος και ομοιογενής, μπορούμε να υποθέσουμε ότι τα άτομα του δείγματος που υποβλήθηκαν στον έλεγχο είναι ανεξάρτητα και κάθε ένα από αυτά έχει πιθανότητα p να πάσχει από την ασθένεια.

Στην περίπτωση αυτή η πιθανότητα του παρατηρηθέντος ενδεχομένου (του παρατηρηθέντος δείγματος) είναι

$$P(E; \theta) = P(x \text{ από } n \text{ άτομα έχουν AIDS})$$

$$= \binom{n}{x} p^x (1-p)^{n-x}$$

Ο παραμετρικός χώρος στην περίπτωση αυτή είναι

$$\Theta = [0, 1] \text{ (δηλ. } 0 \leq p \leq 1)$$

Ως συνάρτηση πιθανοφάνειας μπορεί να ορισθεί οποιοδήποτε θετικό πολλαπλάσιο της $P(E; p)$ που μας διευκολύνει. Για ευκολία, θεωρούμε την

$$L(p) = p^x (1-p)^{n-x}, \quad 0 \leq p \leq 1$$

Τότε

$$\ell(p) = \ln L(p) = x \ln p + (n-x) \ln(1-p)$$

και

$$\ell'(p) = \frac{x}{p} - \frac{n-x}{1-p}, \quad \ell''(p) = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2}, \quad p \neq 0, p \neq 1$$

Για $1 \leq x \leq n-1$, η εξίσωση $\ell'(p) = 0$ έχει ως μοναδική λύση την $p = \frac{x}{n}$.

Δεδομένου ότι $\ell''(p) < 0$ για $p = \frac{x}{n}$ το σημείο αυτό είναι ένα (τοπικό) μέγιστο.

Εξάλλου, δοθέντος ότι $L(p) = 0$ για $p = 0$ και για $p = 1$, το $p = \frac{x}{n}$ είναι το (ολικό) απόλυτο μέγιστο και επομένως

$$\hat{p} = \frac{x}{n}$$

Επομένως, για να μεγιστοποιηθεί η πιθανότητα των δεδομένων (η πιθανότητα να παρατηρηθούν δεδομένα όπως αυτά που πήραμε από το πείραμα) θα πρέπει να εκτιμάμε το ποσοστό p του πληθυσμού με το αντίστοιχο δειγματικό ποσοστό $\hat{p} = \frac{x}{n}$.

Σημείωση: Αν $x=0$, η εξίσωση $\ell'(p) = 0$ δεν έχει λύση και το μέγιστο εμφανίζεται (συμβαίνει) στο σύνορο του παραμετρικού χώρου $\Theta = [0, 1]$.

Στην περίπτωση αυτή έχουμε

$$P(E; \theta) = (1 - p)^n \quad \text{για } 0 \leq p \leq 1$$

Προφανώς, η συνάρτηση αυτή παίρνει την μέγιστη τιμή της όταν $p=0$. Ομοίως, για $x=n$ θα είναι $\hat{p} = 1$ και έτσι, τελικά,

$$\hat{p} = \frac{x}{n} \quad \text{για } x = 0, 1, 2, \dots, n$$

Παράδειγμα: Προκειμένου να ελεγχθεί η καταλληλότητα του νερού της λίμνης του Μαραθώνα ως ποσίμου, γίνονται κάποιοι εργαστηριακοί έλεγχοι σε δείγματα νερού. Συγκεκριμένα, ενδιαφέρει η συγκέντρωση μιας ορισμένης κατηγορίας μυκήτων στο νερό.

Για να ελεγχθεί η συγκέντρωση των μυκήτων αυτών, προσδιορίζεται στο εργαστήριο ο αριθμός των μυκήτων σε n δείγματα νερού από τη λίμνη. Κάθε ένα από τα δείγματα αυτά έχει όγκο μία μονάδα όγκου. Συγκεκριμένα, προσδιορίζεται ο αριθμός των μυκήτων σε κάθε ένα από τα δείγματα αυτά. (Με τον τρόπο αυτό υπάρχουν οι παρατηρηθείσες μετρήσεις x_1, x_2, \dots, x_n). Το πρόβλημα είναι να εκτιμηθεί ο μέσος αριθμός λ μυκήτων ανά μονάδα όγκου της λίμνης.

Λύση: Υποθέτουμε ότι οι μύκητες είναι διασκορπισμένοι τυχαία σε όλη την έκταση της λίμνης. Επομένως, η θέση των μυκήτων είναι τυχαία σημεία στο χώρο. Έτσι, η πιθανότητα να βρεθούν x_i μύκητες στο δείγμα νερού i μοναδιαίου όγκου δίνεται από την κατανομή Poisson με παράμετρο λ .

$$P(X_i = x_i) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \quad x_i=0, 1, 2, \dots; 0 \leq \lambda < \infty$$

Δεδομένου ότι ξένοι μεταξύ τους όγκοι νερού είναι ανεξάρτητοι, η πιθανότητα των n παρατηρηθεισών μετρήσεων x_1, x_2, \dots, x_n είναι

$$\begin{aligned} P(E; \lambda) &= P(X_1 = x_1, \dots, X_n = x_n) = \prod_{j=1}^n e^{-\lambda} \frac{\lambda^{x_j}}{x_j!} \\ &= \frac{e^{-n\lambda} \lambda^{\sum x_j}}{x_1! x_2! \dots x_n!} \end{aligned}$$

Η συνάρτηση πιθανοφάνειας για το λ είναι $kP(E; \lambda)$. Διαλέγοντας κατάλληλα το k μπορούμε να έχουμε

$$L(\lambda) = e^{-n\lambda} \lambda^{\sum x_j}, \quad 0 \leq \lambda < \infty$$

Η λογαριθμοσυνάρτηση πιθανοφάνειας (*loglikelihood function*) και οι παράγωγοί της είναι:

$$\ell(\lambda) = \sum x_j \ln \lambda - n\lambda, \quad \ell'(\lambda) = \frac{1}{\lambda} \sum x_j - n, \quad \ell''(\lambda) = -\frac{\sum x_j}{\lambda^2}$$

Αν $\sum x_j > 0$, η εξίσωση μεγίστης πιθανοφάνειας $\ell'(\lambda) = 0$ έχει ως μοναδική λύση την

$$\lambda = \frac{\sum x_j}{n} = \bar{x}$$

Η δεύτερη παράγωγος στο σημείο αυτό είναι αρνητική, στοιχείο που αποτελεί ένδειξη ότι έχουμε ένα (τοπικό) μέγιστο.

Δοθέντος ότι $L(0)=0$ και $L(\lambda) \xrightarrow{\lambda \rightarrow \infty} 0$ το μέγιστο αυτό είναι (ολικό) γενικό μέγιστο (*overall maximum*).

Εξάλλου, αν $\sum x_j = 0$, η εξίσωση $\ell'(\lambda) = 0$ δεν έχει λύση και το μέγιστο επιτυγχάνεται στο σύνορο (όριο) του παραμετρικού χώρου: $\hat{\lambda} = 0$.

Επομένως, τελικά, σε κάθε περίπτωση έχουμε

$$\hat{\lambda} = \bar{x}$$

(Δηλαδή, η πιθανότητα (πιθανοφάνεια) του επιλεγέντος δείγματος είναι μέγιστη αν ο μέσος λ του πληθυσμού εκτιμηθεί από τον αντίστοιχο μέσο \bar{x} του δείγματος).

Παράδειγμα: (Συνέχεια του προηγούμενου παραδείγματος).

Στην πράξη, συνήθως, δεν είναι δυνατόν να μετρηθούν οι μύκητες σε ένα δείγμα νερού. Εκείνο που είναι δυνατόν να εξακριβωθεί είναι αν υπάρχουν ή όχι μύκητες στο συγκεκριμένο δείγμα νερού. Για τον λόγο αυτό, γεμίζονται οι δοκιμαστικοί σωλήνες χωρητικότητας όγκου v ο καθένας και ελέγχονται στο εργαστήριο. Το αποτέλεσμα του τεστ θεωρείται θετικό αν στο νερό του υπό έλεγχο σωλήνα περιέχονται μύκητες και αρνητικό αν δεν περιέχονται. Έστω ότι σε m από τους n σωλήνες που ελέγχθηκαν δεν βρέθηκαν μύκητες (αρνητικό τεστ). Να βρεθεί η εκτιμήτρια μεγίστης πιθανοφάνειας της μέσης συγκέντρωσης μυκήτων ανά μονάδα νερού στην λίμνη.

Λύση: Η πιθανότητα να υπάρχουν x μύκητες σε μια ποσότητα νερού της λίμνης όγκου v ακολουθεί την κατανομή Poisson με παράμετρο λv .

$$P(X=x) = e^{-\lambda v} \frac{(\lambda v)^x}{x!} \quad x=0, 1, \dots$$

Η πιθανότητα να μην υπάρχουν καθόλου μύκητες σε όγκο νερού v είναι

$$p = P(X=0) = e^{-\lambda v}$$

Η πιθανότητα να είναι το τεστ θετικό (τουλάχιστον ένας μύκητας στον σωλήνα) είναι

$$P(X \geq 1) = 1 - P(X=0) = 1 - e^{-\lambda v}$$

Δεδομένου ότι ξένοι μεταξύ τους όγκοι νερού είναι ανεξάρτητοι, οι n δοκιμαστικοί σωλήνες αποτελούν μια ακολουθία από ανεξάρτητες “δοκιμές”. Για τον λόγο αυτό η πιθανότητα m αρνητικών δοκιμών (σε σύνολο n δοκιμών) ακολουθεί την διωνυμική κατανομή με παράμετρο p . Επομένως,

$$P(E; \lambda) = \binom{n}{m} p^m (1-p)^{n-m}$$

με $p = e^{-\lambda v}$ και $0 \leq \lambda < \infty$.

Αγνοώντας τον σταθερό συντελεστή $\binom{n}{m}$ ορίζουμε ως συνάρτηση πιθανοφάνειας την

$$L(\lambda) = p^m (1-p)^{n-m}$$

Έχουμε όμως ήδη δει ότι η συνάρτηση αυτή επιτυγχάνει την μέγιστη τιμή της όταν $p = \frac{m}{n}$

Η ζητούμενη τιμή του λ βρίσκεται ως η λύση της εξίσωσης $e^{-v\lambda} = \frac{m}{n}$ και είναι

$$\lambda = -\frac{1}{v} \ln p$$

Τελικά, επομένως

$$\hat{\lambda} = -\frac{1}{v} \ln \frac{m}{n} \Rightarrow \hat{\lambda} = -\frac{\ln m - \ln n}{v}$$

Αν, για παράδειγμα πάρουμε 40 δείγματα (σωλήνες) που το καθένα έχει περιεκτικότητα 10 ml νερού και δούμε ότι το τεστ για 28 από αυτά είναι αρνητικό, θα έχουμε ότι

$$\hat{\lambda} = \frac{\ln 40 - \ln 28}{10} = 0.0357$$

Επομένως, η μέση συγκέντρωση μυκήτων στο νερό της λίμνης εκτιμάται σε 0.0357 μύκητες/ml νερού.

Πιθανοφάνειες Βασισμένες σε Πίνακες Συχνότητας

Έστω ότι έχουμε n ανεξάρτητες επαναλήψεις ενός πειράματος. Με το πείραμα αυτό ο δειγματικός χώρος S , για μια απλή επανάληψη του πειράματος, διαμερίζεται σε k ενδεχόμενα (ή κλάσεις) που είναι αμοιβαία ξένα μεταξύ τους.

Δηλαδή,

$$S = A_1 \cup A_2 \cup \dots \cup A_k$$

Έστω f_j ο αριθμός των φορών που το ενδεχόμενο A_j εμφανίζεται στις n επαναλήψεις του πειράματος ($\sum f_j = n$). Έστω P_j η πιθανότητα του ενδεχομένου A_j [$P_j = P(A_j)$] σε μια οποιαδήποτε από τις επαναλήψεις του πειράματος. (Προφανώς, $\sum P_j = 1, j=1, 2, \dots, k$). Οι τιμές των P_j μπορούν να καθορισθούν από το μοντέλο πιθανότητας που ισχύει στο συγκεκριμένο πρόβλημα. Εάν το μοντέλο περιλαμβάνει μια άγνωστη παράμετρο θ , οι πιθανότητες P_j θα είναι, εν γένει, συναρτήσεις του θ .

Το πείραμα αυτό συνοψίζεται στον πίνακα συχνότητας που ακολουθεί.

Ενδεχόμενα	$A_1 \quad A_2 \quad \dots \quad A_k$	Σύνολο
Παρατηρηθείσα Συχνότητα	$f_1 \quad f_2 \quad \dots \quad f_k$	n
Αναμενόμενη Συχνότητα	$np_1 \quad np_2 \quad \dots \quad np_k$	n

Η πιθανότητα να παρατηρήσουμε ένα συγκεκριμένο πίνακα συχνότητας δίνεται ως γνωστόν, από την πολυωνυμική κατανομή.

$$P(E; \theta) = \binom{n}{f_1 f_2 \dots f_k} P_1^{f_1} P_2^{f_2} \dots P_k^{f_k}$$

Η συνάρτηση πιθανοφάνειας του θ που βασίζεται στον συγκεκριμένο πίνακα συχνότητας είναι ανάλογη του $P(E; \theta)$.

Επομένως μπορούμε να ορίσουμε ως συνάρτηση πιθανοφάνειας την

$$L(\theta) = k P_1^{f_1} P_2^{f_2} \dots P_k^{f_k},$$

όπου k είναι μια κατάλληλη σταθερά.

Η εκτιμήτρια μέγιστης πιθανοφάνειας προσδιορίζεται με την μεγιστοποίηση της συνάρτησης $L(\theta)$.

Χρησιμοποιώντας την εκτιμήτρια αυτή $\hat{\theta}$ είναι δυνατόν να υπολογίσουμε τις αναμενόμενες συχνότητες και να τις συγκρίνουμε με τις παρατηρηθείσες συχνότητες.

Παράδειγμα: Στον στατιστικό έλεγχο ποιότητας ενός εργοστασίου, προκειμένου να ελεγχθεί η ποιότητα των προϊόντων, επιλέγονται καθημερινά και επί 200 συνεχείς ημέρες 10 αντικείμενα με τυχαίο τρόπο και ελέγχονται για πιθανές ατέλειες. Τα αποτελέσματα δίνονται στον πίνακα που ακολουθεί:

Αριθμός ελαττωματικών	0	1	2	3	≥ 4	Σύνολο
Παρατηρηθείσα συχνότητα	133	52	12	3	0	200

Έστω p η πιθανότητα ένα αντικείμενο να είναι ελαττωματικό. Να βρεθεί η εκτιμήτρια μέγιστης πιθανοφάνειας του p .

Λύση: Μπορούμε να υποθέσουμε ότι ο αριθμός των ελαττωματικών αντικειμένων σε κάθε παρτίδα 10 αντικειμένων ακολουθεί την διωνυμική κατανομή με παράμετρο p . Στην περίπτωση αυτή, η πιθανότητα P_j στα 10 αντικείμενα να παρατηρηθούν j ελαττωματικά δίνεται από την διωνυμική κατανομή.

$$P_j = P(X = j) = \binom{10}{j} p^j (1-p)^{10-j} \quad j=0, 1, \dots, 10$$

Η πιθανότητα να παρατηρηθούν 4, ή περισσότερα ελαττωματικά αντικείμενα στο δείγμα των 10 αντικειμένων δίνεται από τον τύπο

$$P(X \geq 4) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$$

Για τα συγκεκριμένα δεδομένα έχουμε ότι η συνάρτηση πιθανοφάνειας της παραμέτρου p που βρήκαμε προηγουμένως παίρνει την μορφή

$$\begin{aligned} L(p) &= k [P(X=0)]^{133} [P(X=1)]^{52} [P(X=2)]^{12} [P(X=3)]^3 [P(X \geq 4)]^0 \\ &= k P_0^{133} P_1^{52} P_2^{12} P_3^3 P_4^0, \quad 0 \leq p \leq 1 \end{aligned}$$

Αντικαθιστώντας τις τιμές των P_0, P_1, P_3, P_4 και διαλέγοντας ένα κατάλληλο k , έχουμε

$$\begin{aligned} L(p) &= [(1-p)^{10}]^{133} [p(1-p)^9]^{52} [p^2(1-p)^8]^{12} [p^3(1-p)^7]^3 \\ &= p^{85} (1-p)^{1915} \end{aligned}$$

Η συνάρτηση πιθανοφάνειας αυτής της μορφής είναι η ίδια με τη συνάρτηση πιθανοφάνειας παρόμοιου προβλήματος που έχουμε ήδη συναντήσει και από την οποία έχουμε βρει ότι εκτιμήτρια μέγιστης πιθανοφάνειας είναι ο λόγος των παρατηρηθεισών ευνοϊκών περιπτώσεων διά του συνολικού αριθμού των δοκιμών

$$\hat{p} = \frac{x}{n}$$

Στην περίπτωσή μας $x = 85, n = 2000$.

Επομένως, $\hat{p} = \frac{85}{2000} = 0.0425$.

Χρησιμοποιώντας την τιμή αυτή του \hat{p} μπορούμε να βρούμε τις αναμενόμενες συχνότητες

$$n P_j = 100 P_j, \quad j=0, 1, 2, 3$$

Η αναμενόμενη συχνότητα για την τελευταία κλάση ($X \geq 4$) μπορεί τότε να βρεθεί με αφαίρεση από το 200.

Οι αναμενόμενες συχνότητες δίνονται στον πίνακα που ακολουθεί:

Αριθμός Ελαττωματικών	0	1	2	3	≥ 4	Σύνολο
Παρατηρούμενη Συχνότητα	133	52	12	3	0	200
Αναμενόμενη Συχνότητα	129.54	57.50	11.48	1.36	0.12	200

Όπως παρατηρούμε από τον πίνακα αυτό, οι αναμενόμενες τιμές, με την χρησιμοποίηση του συγκεκριμένου μοντέλου πλησιάζουν πολύ τις παρατηρηθείσες τιμές.

Δεδομένου ότι οι παρατηρηθείσες συχνότητες f_j είναι τυχαίες μεταβλητές, είναι αναμενόμενο ότι θα διαφέρουν από τις αντίστοιχες αναμενόμενες συχνότητες. Υπάρχει στατιστική μεθοδολογία (ο έλεγχος καλής προσαρμογής) η οποία μας δίνει τρόπους και μεθόδους για να ελέγξουμε την καλή προσαρμογή του μοντέλου που έχουμε επιλέξει.

Υπολογιστικές Μέθοδοι Καθορισμού της Εκτιμήτριας Μέγιστης Πιθανοφάνειας

Όπως έχουμε ήδη αναφέρει, η εκτιμήτρια $\hat{\theta}$ μιας παραμέτρου είναι εκείνη που μεγιστοποιεί τον λογάριθμο της συνάρτησης πιθανοφάνειας $\ell(\theta)$.

Σε συγκεκριμένες περιπτώσεις, όπως αυτές που έχουμε ήδη συναντήσει, η εξίσωση $\ell'(\theta) = 0$ είναι δυνατόν να λυθεί αλγεβρικά και να δώσει έναν τύπο ο οποίος να προσδιορίζει την εκτιμήτρια $\hat{\theta}$.

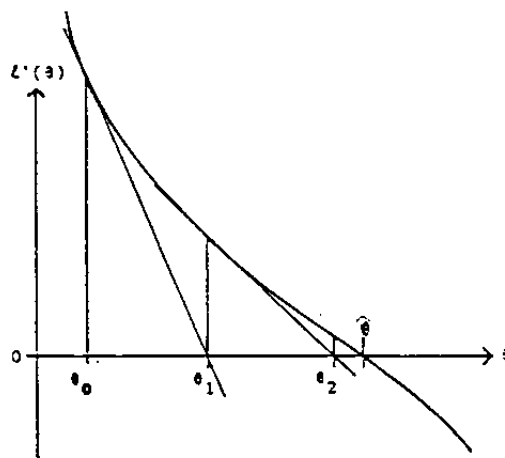
Στις περισσότερες όμως περιπτώσεις αυτό δεν είναι δυνατό. Είμαστε τότε υποχρεωμένοι να καταφύγουμε σε αριθμητικές μεθόδους επίλυσης της εξίσωσης αυτής. Τα τελευταία χρόνια, οι μέθοδοι αριθμητικής επίλυσης της εξίσωσης μέγιστης πιθανοφάνειας έχουν απλοποιηθεί πολύ με την χρήση των υπολογιστών και διάφορες

μέθοδοι έχουν αναπτυχθεί για την επίλυση της εξίσωσης αυτής. Για παράδειγμα, αναφέρουμε μία μόνο από τις μεθόδους αυτές που είναι από τις περισσότερο διαδεδομένες.

Η μέθοδος αυτή είναι η λεγόμενη μέθοδος του Newton.

Μέθοδος του Newton: Το σχήμα που ακολουθεί περιγράφει πώς λειτουργεί η μέθοδος του Newton για τον καθορισμό της ρίζας της εξίσωσης μέγιστης πιθανοφάνειας

$$l'(\theta) = 0$$



Λύση της $l'(\theta) = 0$ με την μέθοδο Newton

Αρχίζουμε με μια τιμή του θ , έστω $\theta = \theta_0$ που πιστεύουμε ότι ενδεχομένως είναι η εκτιμήτρια μέγιστης πιθανοφάνειας και στη συνέχεια προσπαθούμε να την βελτιώσουμε. Αν θ_i είναι η προσέγγιση του $\hat{\theta}$ την οποία πήραμε στην i επανάληψη του πειράματος, θα έχουμε

$$\theta_{i+1} = \theta_i - l'(\theta_i)/l''(\theta_i)$$

Αυτό γιατί η κλίση της καμπύλης στο $l'(\theta_i)$ είναι η παράγωγος της l' στο θ_i .

Είναι δε (όπως προκύπτει από το σχήμα)

$$l''(\theta_1) = \frac{\text{εφ } \omega}{\theta_1 - \theta_0} = - \frac{\text{εφ } (180 - \omega)}{\theta_1 - \theta_0} = - \frac{l'(\theta_1)}{\theta_1 - \theta_0}$$

Επομένως,

$$\theta_1 - \theta_0 = \ell'(\theta_1) / \ell''(\theta_1)$$

Όπως δείχνει το διάγραμμα, το θ_{i+1} είναι το σημείο στο οποίο η εφαπτομένη της $\ell'(\theta)$ στο σημείο $\theta = \theta_i$ τέμνει τον οριζόντιο άξονα. Αν η αρχική επιλογή θ_0 είναι αρκετά καλή, η μέθοδος θα οδηγήσει σε μια ακριβή προσέγγιση του $\hat{\theta}$ μετά από μερικές μόνο επαναλήψεις.

Αν η εξίσωση $\ell'(\theta) = 0$ έχει περισσότερες από μια λύσεις, η μέθοδος του Newton δεν είναι βέβαιο ότι θα συγκλίνει στην επιθυμητή λύση.

Δυσκολίες επίσης παρουσιάζονται εάν το μέγιστο της συνάρτησης εμφανίζεται στο σύνορο ή κοντά στο σύνορο του παραμετρικού χώρου.

Για να προφυλαχθεί κανείς από τέτοια ενδεχόμενα, είναι καλό να εξετάζει πρώτα τη γραφική παράσταση της $\ell(\theta)$ ή της $\ell'(\theta)$ πριν εφαρμόσει την μέθοδο του Newton

Μια γενίκευση της μεθόδου του Newton είναι η μέθοδος Newton-Raphson η οποία χρησιμοποιείται όταν υπάρχουν δύο ή περισσότερες άγνωστες παράμετροι που θα πρέπει να εκτιμηθούν. Η μέθοδος αυτή βρίσκεται στο μενού MATHEMATICAL AND USER PROCEDURES ως επιλογή 3 (Root finding (Εύρεση ριζών)). Μια εφαρμογή της διαδικασίας αυτής θα συναντήσουμε σε παράδειγμα που ακολουθεί.

Χρησιμοποίηση Ανεξαρτήτων Πιθανοφαιών

Σε πολλές περιπτώσεις μελέτης ενός προβλήματος ο ερευνητής έχει στην διάθεσή του τα αποτελέσματα δύο ή περισσότερων ανεξαρτήτων επαναλήψεων του πειράματος και όχι μόνο μιας (όπως έχουμε μέχρι τώρα δει).

Στις περιπτώσεις αυτές υπάρχουν πληροφορίες για την ίδια παράμετρο θ από δύο, ή περισσότερες, διαφορετικές πηγές που όπως είναι φυσικό ο ερευνητής θα ήθελε να αξιοποιήσει.

Ας υποθέσουμε για παράδειγμα ότι στην πρώτη επανάληψη ενός πειράματος παρατηρείται ένα ενδεχόμενο E_1 . Το ενδεχόμενο αυτό έχει συνάρτηση πιθανοφάνειας για το θ

$$L_1(\theta) = k_1 P(E_1; \theta)$$

όπου k_1 είναι μια θετική σταθερά.

Σε μια δεύτερη επανάληψη του πειράματος, έστω ότι παρατηρείται το ενδεχόμενο E_2 το οποίο θα οδηγήσει σε συνάρτηση πιθανοφάνειας

$$L_2(\theta) = k_2 P(E_2; \theta)$$

Είναι δυνατό να θεωρήσουμε τις δύο επαναλήψεις του πειράματος ως συνιστώσες ενός απλού συνθέτου πειράματος. Στο σύνθετο αυτό πείραμα ως ενδεχόμενο που παρατηρείται μπορεί να θεωρηθεί η τομή των ενδεχομένων E_1 και E_2 . Η αντίστοιχη συνάρτηση πιθανοφάνειας του πειράματος αυτού θα είναι

$$L(\theta) = k P(E_1 E_2; \theta)$$

Δοθέντος ότι τα ενδεχόμενα E_1 και E_2 είναι ανεξάρτητα η συνάρτηση πιθανοφάνειας γράφεται

$$P(E_1 E_2; \theta) = P(E_1; \theta) P(E_2; \theta)$$

Επομένως, προκύπτει ότι η συνάρτηση πιθανοφάνειας $L(\theta)$ είναι

$$L(\theta) = k' L_1(\theta) L_2(\theta),$$

όπου k' είναι μια θετική σταθερά.

Επειδή το k' μπορεί να επιλεγεί με αυθαίρετο τρόπο, μπορούμε να γράψουμε τη συνάρτηση πιθανοφάνειας ως

$$L(\theta) = L_1(\theta) L_2(\theta)$$

Θεωρώντας τους λογαρίθμους και των δύο πλευρών της εξίσωσης αυτής, έχουμε

$$\ell(\theta) = \ell_1(\theta) + \ell_2(\theta)$$

Επομένως, προκειμένου να συνδυάσουμε τις πληροφορίες για το θ από δύο, ή περισσότερα, ανεξάρτητα πειράματα πολλαπλασιάζουμε, απλώς, τις συναρτήσεις πιθανοφάνειας ή αντίστοιχα, αθροίζουμε τους λογαρίθμους των συναρτήσεων πιθανοφανειών.

Αν $\hat{\theta}_1$ είναι η εκτιμήτρια μέγιστης πιθανοφάνειας του θ που προκύπτει από την πρώτη επανάληψη του πειράματος και $\hat{\theta}_2$ είναι η αντίστοιχη εκτιμήτρια μέγιστης πιθανοφάνειας που προκύπτει από την δεύτερη επανάληψη του πειράματος, ενώ $\hat{\theta}$ είναι η γενική εκτιμήτρια μέγιστης πιθανοφάνειας του θ , θα έχουμε ότι το $\hat{\theta}_1$ μεγιστοποιεί την $\ell_1(\theta)$, το $\hat{\theta}_2$ μεγιστοποιεί την $\ell_2(\theta)$ και το $\hat{\theta}$ μεγιστοποιεί την $\ell(\theta)$.

Αν $\hat{\theta}_1 = \hat{\theta}_2$, τότε και οι δύο όροι στο δεξιό μέλος της τελευταίας εξίσωσης παίρνουν την μέγιστη τιμή τους στο ίδιο σημείο και επομένως

$$\hat{\theta} = \hat{\theta}_1 = \hat{\theta}_2$$

Διαφορετικά, το ολικό μέγιστο $\hat{\theta}$ είναι συνήθως μια τιμή που θα βρίσκεται μεταξύ του $\hat{\theta}_1$ και του $\hat{\theta}_2$.

Παράδειγμα: (Συνέχεια του παραδείγματος για το AIDS)

Έστω ότι στο παράδειγμα με το AIDS m άτομα επιλέγονται τυχαία από τον υπό εξέταση πληθυσμό. Έστω ότι y από αυτούς βρίσκεται ότι πάσχουν από AIDS.

Να βρεθεί η ΕΜΠ (MLE) του p βάσει των στοιχείων που υπάρχουν από τα δύο δείγματα.

Λύση: Από το πρώτο πείραμα, είχαμε δει ότι ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι

$$\ell_1(p) = x \ln p + (n - x) \ln (1 - p)$$

με εκτιμήτρια μέγιστης πιθανοφάνειας για το p την $\hat{p}_1 = \frac{x}{n}$.

Για το δεύτερο πείραμα (δεύτερο δείγμα) θα έχουμε, αντίστοιχα

$$\ell_2(p) = y \ln p + (m - y) \ln(1 - p)$$

με

$$\hat{p}_2 = \frac{y}{m}$$

Δεδομένου ότι ο πληθυσμός είναι μεγάλος τα δύο δείγματα είναι δυνατόν να θεωρηθούν σχεδόν ανεξάρτητα και επομένως ο λογάριθμος της συνάρτησης πιθανοφάνειας για τα δύο δείγματα μπορεί να γραφεί

$$\begin{aligned} \ell(p) &= \ell_1(p) + \ell_2(p) \\ &= (x+y) \ln p + (n + m - x - y) \ln(1 - p) \end{aligned}$$

Η συνάρτηση αυτή είναι της ίδιας μορφής με τον λογάριθμο της συνάρτησης πιθανοφάνειας για κάθε ένα από τα δείγματα, και επομένως η ολική εκτιμήτρια μέσης πιθανοφάνειας είναι:

$$\hat{p} = \frac{x + y}{n + m}$$

Δοθέντος ότι,

$$x = n \hat{p}_1, \quad y = m \hat{p}_2$$

θα έχουμε

$$\hat{p} = \frac{n}{n + m} \hat{p}_1 + \frac{m}{n + m} \hat{p}_2$$

(που είναι ο σταθμισμένος μέσος των \hat{p}_1 και \hat{p}_2)

Για παράδειγμα, αν στο πρώτο δείγμα είχαν εξετασθεί 90 άτομα ($n=90$) και στο δεύτερο δείγμα είχαν εξετασθεί 10 άτομα ($m=10$) θα είχαμε

$$\hat{p} = 0.9 \hat{p}_1 + 0.1 \hat{p}_2$$

Είναι προφανές ότι η εκτιμήτρια μεγίστης πιθανοφάνειας θα βρίσκεται πλησιέστερα προς την εκτιμήτρια μεγίστης πιθανοφάνειας που έχει προέλθει από το μεγαλύτερο από τα δείγματα από ότι προς την εκτιμήτρια μεγίστης πιθανοφάνειας που θα έχει προέλθει από το μικρότερο από τα δύο δείγματα.

Μπορούμε επίσης να παρατηρήσουμε ότι η λογαριθμοσυνάρτηση πιθανοφάνειας είναι η ίδια με εκείνη στην οποία θα είχαμε καταλήξει εάν είχαμε θεωρήσει ένα και μόνο δείγμα που θα περιλάμβανε $n + m$ άτομα από τα οποία $x + y$ θα είχαν βρεθεί να έχουν την ασθένεια.

Ο διαχωρισμός των αποτελεσμάτων σε δύο διαφορετικά πειράματα δεν προσφέρει καμιά πρόσθετη πληροφορία όσον αφορά την εκτίμηση της παραμέτρου p .

Παράδειγμα: (συνέχεια του παραδείγματος για τον έλεγχο της ποιότητας του νερού της λίμνης του Μαραθώνα).

Στο στάδιο του πειράματος του προβλήματος εκείνου, που αφορούσε την δειγματοληπτική λήψη νερού από την λίμνη για έλεγχο, ήταν απαραίτητο να καθορισθεί ο όγκος v του νερού από τη λίμνη το οποίο θα ετοποθετείτο σε κάθε ένα δοκιμαστικό σωλήνα για έλεγχο. Είναι προφανές ότι αν ο όγκος αυτός ήταν πολύ μεγάλος όλα τα δείγματα προς έλεγχο θα περιείχαν μύκητες και επομένως θα κατέληγαν σε θετικό αποτέλεσμα του ελέγχου.

Από το άλλο μέρος, αν το v ήταν πολύ μικρό είναι πολύ πιθανό να είχαμε μόνο αρνητικά αποτελέσματα του ελέγχου. Σε οποιαδήποτε από τις ακραίες αυτές περιπτώσεις το πείραμα δεν θα έδινε ικανοποιητικές πληροφορίες για την μέση συγκέντρωση μυκήτων στη λίμνη.

Ένας τρόπος για να αποφύγει κανείς αυτή τη δυσκολία είναι να ετοιμάσει δύο (ή περισσότερα) είδη δοκιμαστικών σωλήνων που να μπορούν να χωρέσουν διαφορετικούς όγκους νερού από τη λίμνη.

Έστω, ότι στο συγκεκριμένο πρόβλημα που έχουμε αναφερθεί ο ερευνητής είχε στη διάθεσή του δοκιμαστικούς σωλήνες που είχαν τη δυνατότητα να περιέχουν 10ml νερού και 1 ml αντίστοιχα. Έστω ότι χρησιμοποιήθηκαν 40 δοκιμαστικοί σωλήνες χωρητικότητας 10ml και από αυτούς βρέθηκε στο σχετικό έλεγχο ότι 28 έδωσαν αρνητικό αποτέλεσμα (δεν περιείχαν μύκητες). Επίσης, έστω ότι χρησιμοποιήθηκαν 40 δοκιμαστικοί σωλήνες όγκου 1ml από τους οποίους, στον σχετικό έλεγχο, 37 έδωσαν αρνητικό αποτέλεσμα. Με ποιο τρόπο στην περίπτωση αυτή θα υπολογίσουμε την εκτιμήτρια

μέγιστης πιθανοφάνειας της μέσης συγκέντρωσης λ μυκήτων στα νερά της λίμνης;

Λύση: Όπως είχαμε αναφέρει στο παράδειγμα, η συνάρτηση πιθανοφάνειας για το λ , η οποία θα βασίζεται στις πληροφορίες από τους 40 δοκιμαστικούς σωλήνες που περιέχουν 10 ml νερού, είναι

$$L_1(\lambda) = p_1^{28}(1-p_1)^{12}$$

όπου, όπως έχει αναφερθεί, p_1 είναι η πιθανότητα αρνητικού αποτελέσματος στον έλεγχο για ένα δοκιμαστικό σωλήνα που προέρχεται από το πρώτο πείραμα με

$$p_1 = e^{-10\lambda}$$

Όπως είχαμε δει στο παράδειγμα εκείνο η εκτίμηση μέγιστης πιθανοφάνειας ήταν

$$\hat{\lambda}_1 = 0.0357$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι

$$\ell_1(\lambda) = 28\ln p_1 + 12\ln(1-p_1)$$

Για τους ίδιους λόγους ο λογάριθμος της συνάρτησης πιθανοφάνειας που στηρίζεται στους 40 δοκιμαστικούς σωλήνες του ενός ml είναι

$$\ell_2(\lambda) = 37\ln p_2 + 3\ln(1-p_2)$$

όπου, για τους ίδιους λόγους

$$p_2 = e^{-\lambda}$$

Η εκτιμήτρια μέγιστης πιθανοφάνειας του λ είναι η

$$\hat{\lambda} = \frac{\ln 40 - \ln 37}{1} = 0.078$$

Εξάλλου, σύμφωνα με τα όσα είδαμε, ο λογάριθμος της συνάρτησης πιθανοφάνειας που αναφέρεται και στους 80 δοκιμαστικούς σωλήνες είναι:

$$\ell(\lambda) = \ell_1(\lambda) + \ell_2(\lambda)$$

$$\begin{aligned}
&= 28 \ln p_1 + 12 \ln (1 - p_1) + 37 \ln p_2 + 3 \ln (1 - p_2) \\
&= -317\lambda + 12 \ln (1 - e^{-10\lambda}) + 3 \ln (1 - e^{-\lambda})
\end{aligned}$$

Η ολική εκτίμηση μέγιστης πιθανοφάνειας $\hat{\lambda}$ της παραμέτρου λ είναι εκείνη για την οποία $\ell(\lambda)$ έχει μέγιστο.

Η πρώτη παράγωγος της συνάρτησης $\ell(\lambda)$ η οποία θα μας οδηγήσει σε ακρότατο, ως προς λ είναι:

$$\begin{aligned}
\ell'(\lambda) &= -317 + \frac{120e^{-10\lambda}}{1 - e^{-10\lambda}} + \frac{3e^{-\lambda}}{1 - e^{-\lambda}} \\
&= -317 + \frac{120}{e^{10\lambda} - 1} + \frac{3}{e^{\lambda} - 1}
\end{aligned}$$

Εδώ έχουμε μία περίπτωση που η εξίσωση μέγιστης πιθανοφάνειας

$$\ell'(\lambda) = 0$$

δεν είναι δυνατόν να λυθεί αλγεβρικά. Για τον λόγο αυτό και προκειμένου να καθορίσουμε το λ , θα πρέπει να χρησιμοποιήσουμε υπολογιστικές μεθόδους. Αν χρησιμοποιήσουμε τη μέθοδο Newton που περιγράψαμε προηγουμένως, μπορούμε να ξεκινήσουμε με αρχική τιμή την

$$\frac{1}{2}(\hat{\lambda}_1 + \hat{\lambda}_2) = 0.057$$

Για να συνεχίσουμε χρειαζόμαστε τη δεύτερη παράγωγο της συνάρτησης $\ell(\lambda)$. Είναι:

$$\ell''(\lambda) = -\frac{1200e^{10\lambda}}{(e^{10\lambda} - 1)^2} - \frac{3e^{\lambda}}{(e^{\lambda} - 1)^2}$$

Οι υπολογισμοί συνοψίζονται στον πίνακα που ακολουθεί.

Πίνακας
(Λύση της εξίσωσης $\ell'(\lambda)=0$ με τη μέθοδο Newton)

i	λ_i	$\ell'(\lambda_i)$	$\ell''(\lambda_i)$	$\ell'(\lambda_i)/\ell''(\lambda_i)$
0	0.057	-109.66	-4518.16	0.02427
1	0.03273	83.07	-13902.58	-0.00598
2	0.03871	12.87	-9910.74	-0.00130
3	0.04001	0.41	-9270.86	-0.00004
4	0.04005	0.04	-9252.15	-0.00000

Όπως παρατηρούμε στον πίνακα, μετά από τέσσερις επαναλήψεις της διαδικασίας παίρνουμε

$$\hat{\lambda} = 0.04005$$

Σημείωση: Από τον πίνακα, μπορούμε να παρατηρήσουμε ότι η δεύτερη παράγωγος είναι αρνητική, γεγονός που αποτελεί ένδειξη ότι έχουμε καταλήξει σε ένα τοπικό μέγιστο.

Ο προσδιορισμός της λύσης της εξίσωσης μέγιστης πιθανοφάνειας με το Statgraphics γίνεται με τον εξής τρόπο:

Μετά την επιλογή από το μενού της διαδικασίας ROOT FINDING εμφανίζεται στην οθόνη του υπολογιστή ένας πίνακας όπως αυτός που ακολουθεί:

ROOT FINDING

FUNCTION:

INITIAL GUESS: 0

INITIAL STEP SIZE: 0.1

ABSOLUTE CONVERGENCE CRITERION: 1E-3

SIGNIFICANT DIGITS DESIRED: 5

MAXIMUM FUNCTION CALLS: 50

Προκειμένου να βρούμε τη λύση της εξίσωσης, χρειάζεται να συμπληρώσουμε τα πεδία αυτά. Στο πεδίο FUNCTION γράφουμε

τη συνάρτηση της οποίας θέλουμε να βρούμε τη ρίζα με τον τρόπο που γράφονται οι συναρτήσεις στο Statgraphics. Στη συνέχεια, επιλέγουμε κατάλληλες τιμές για την πρώτη προσέγγιση (INITIAL GUESS), για το μέγεθος του βήματος του αλγορίθμου (INITIAL STEP SIZE), για το κριτήριο απόλυτης σύγκλισης (ABSOLUTE CONVERGENCE CRITERION), για τον αριθμό των δεκαδικών ψηφίων που επιθυμούμε (SIGNIFICANT DIGITS DESIRED), και τέλος για το μέγιστο αριθμό των επαναλήψεων που επιθυμούμε (MAXIMUM FUNCTION CALLS). Στον προηγούμενο πίνακα, φαίνονται οι τιμές που δίνει αρχικά ο υπολογιστής. Αν, για το συγκεκριμένο πρόβλημα που αντιμετωπίσαμε, γράψουμε την συνάρτηση και θεωρήσουμε ως αρχική τιμή του αλγορίθμου το 0.4 αφήνοντας τα υπόλοιπα πεδία με τις τιμές που δίνει ο υπολογιστής, βλέπουμε στον πίνακα που ακολουθεί ότι η ρίζα (το μέγιστο) είναι στο σημείο 0.040054 (ROOT FOUND AT 0.040054).

ROOT FINDING

FUNCTION: $-317+(120/\text{EXP } 10^*X)-1)+3/(\text{EXP } X)-1$

INITIAL GUESS: 0.4

INITIAL STEP SIZE: 0.1

ABSOLUTE CONVERGENCE CRITERION: 1E-3

SIGNIFICANT DIGITS DESIRED: 5

MAXIMUM FUNCTION CALLS: 50

ROOT FOUND AT: 0.040054

Σημείωση: Προφανώς, η δυνατότητα αυτή του Statgraphics μπορεί να χρησιμοποιηθεί για τον καθορισμό της ρίζας οποιασδήποτε εξίσωσης και όχι μόνο για την εύρεση της ρίζας της συνάρτησης μέγιστης πιθανοφάνειας.

Σημείωση: Είναι προφανές ότι όσο μεγαλύτερη είναι η συγκέντρωση μυκήτων στην λίμνη, τόσο μεγαλύτερη γίνεται η πιθανότητα ότι και οι δοκιμαστικοί σωλήνες θα δώσουν θετικό αποτέλεσμα στο τεστ.

Αυτό συνεπάγεται ότι αυξανόμενου του λ μειώνεται το m και επομένως αυξάνει η πιθανότητα του γεγονότος $m=0$. Εάν παρατηρήσουμε $m=0$, τότε η ΕΜΠ του λ θα είναι $+\infty$. Στην

περίπτωση αυτή, το να δώσει ο ερευνητής μια απλή (σημειακή) εκτίμηση για το λ δεν έχει καμμία πρακτική σημασία. Στην περίπτωση αυτή, ζητούμενο είναι μια ένδειξη για ένα πεδίο τιμών του λ που είναι πιθανές (για τις δεδομένες παρατηρήσεις) παρά μια “πιθανή” τιμή του λ .

Αυτό μπορεί να γίνει είτε με την μέθοδο του διαστήματος εμπιστοσύνης που θα εξετασθεί αργότερα ή με την χρήση της συνάρτησης σχετικής πιθανοφάνειας (*relative likelihood function*) που θα δούμε στη συνέχεια.

Σχετική Πιθανοφάνεια (*Relative Likelihood*)

Ας υποθέσουμε και πάλι ότι για κάποιο συγκεκριμένο πείραμα υπάρχει ένα μοντέλο πιθανότητας που το περιγράφει, το οποίο περιέχει και μια παράμετρο θ .

Στη συνέχεια εκτελείται το πείραμα και παρατηρείται κάποιο ενδεχόμενο E . Η πιθανότητα του ενδεχομένου E μπορεί να καθοριστεί από το συγκεκριμένο μοντέλο ως συνάρτηση της παραμέτρου θ .

Έστω

$$P(E; \theta)$$

Η συνάρτηση πιθανοφάνειας του θ είναι, όπως έχουμε δει, ένα σταθερό πολλαπλάσιο της $P(E; \theta)$

$$L(\theta) = k P(E; \theta)$$

όπου k είναι μια σταθερά που δεν εξαρτάται από το θ .

Στις μέχρι τώρα συζητήσεις μας έχουμε χρησιμοποιήσει τη συνάρτηση πιθανοφάνειας με μοναδικό σκοπό να καθορίσουμε την εκτιμήτρια $\hat{\theta}$ της άγνωστης παραμέτρου θ . Αυτή, όπως έχουμε δει, είναι η τιμή της παραμέτρου για την οποία η πιθανότητα να παρατηρηθούν τα δεδομένα E μεγιστοποιείται.

Στην πιο γενική περίπτωση η συνάρτηση πιθανοφάνειας μπορεί να χρησιμοποιηθεί για να εξετασθεί ολόκληρο το πεδίο των πιθανών τιμών της παραμέτρου θ και να καθορισθεί ποια τιμή είναι

περισσότερο πιθανή και ποια λιγότερο πιθανή κάτω από το φως των δεδομένων του διεξαχθέντος πειράματος.

Έστω ότι θ_1 και θ_2 είναι δύο πιθανές τιμές της παραμέτρου θ .

Ορισμός: Ορίζουμε ως *λόγο πιθανοφάνειας (likelihood ratio)* του θ_1 έναντι του θ_2 το λόγο

$$\frac{L(\theta_1)}{L(\theta_2)} = \frac{kP(E; \theta_1)}{kP(E; \theta_2)}$$

$$\frac{\text{Πιθανότητα των δεδομένων για } \theta=\theta_1}{\text{Πιθανότητα των δεδομένων για } \theta=\theta_2}$$

Είναι προφανές ότι, αν ο λόγος αυτός είναι μεγαλύτερος από 1, τα δεδομένα του πειράματός μας είναι περισσότερο πιθανό να έχουν προέλθει από το μοντέλο με τιμή της παραμέτρου $\theta=\theta_1$ παρά από αυτό με τιμή παραμέτρου $\theta=\theta_2$.

Στην περίπτωση αυτή, λέμε ότι η τιμή θ_1 είναι περισσότερο *ευλογοφανής (plausible)* ή περισσότερο πιθανή (*more likely*) τιμή της παραμέτρου από ότι η θ_2 . Το μέγεθος του λόγου αυτού δίνει ένα μέτρο του πόσο περισσότερο πιθανή είναι η τιμή θ_1 από την τιμή θ_2 . Για παράδειγμα, εάν

$$L(\theta_1) / L(\theta_2) = 100$$

τότε τα δεδομένα είναι 100 φορές περισσότερο πιθανό να έχουν προέλθει από ένα πληθυσμό με τιμή παραμέτρου $\theta=\theta_1$ από ότι από πληθυσμό με την ίδια κατανομή και τιμή παραμέτρου $\theta=\theta_2$. Στην περίπτωση αυτή, λέμε ότι με βάση τα δεδομένα, η παραμετρική τιμή θ_1 είναι 100 φορές περισσότερο πιθανή από ότι η θ_2 .

Για λόγους ομοιομορφίας, διευκολύνει να διαλέξουμε μια τιμή με την οποία να συγκρίνουμε όλες τις άλλες ενδεχόμενες τιμές του θ . Η φυσική και προφανής επιλογή είναι το $\hat{\theta}$, η περισσότερο πιθανή τιμή του θ .

Ορισμός: Ορίζουμε ως *συνάρτηση σχετικής πιθανοφάνειας* (*relative likelihood function*) (RLF) της παραμέτρου θ τον λόγο:

$$R(\theta) = L(\theta) / L(\hat{\theta})$$

όπου $\hat{\theta}$ είναι η εκτιμήτρια μέγιστης πιθανοφάνειας του θ .
 Δεδομένης της ιδιότητας του $\hat{\theta}$ είναι προφανές ότι:

$$0 \leq R(\theta) \leq 1$$

για όλες τις δυνατές τιμές του θ .

Εάν $\hat{\theta}$ είναι κάποια συγκεκριμένη τιμή του θ , τότε $R(\theta_1)$ είναι ο λόγος πιθανοφάνειας της θ_1 σε σχέση με την $\hat{\theta}$

$$R(\theta_1) = \frac{L(\theta_1)}{L(\hat{\theta})} = \frac{kP(E; \theta_1)}{kP(E; \hat{\theta})} =$$

Πιθανότητα των δεδομένων για $\theta = \theta_1$
 Μέγιστη πιθανότητα των δεδομένων
 για οποιαδήποτε τιμή του θ

Αν $R(\theta_1)$ είναι σχετικά μικρό, π.χ. $R(\theta_1) \leq 0.1$, το θ_1 είναι μια μάλλον απίθανη τιμή της παραμέτρου, δεδομένου ότι υπάρχουν άλλες τιμές του θ για τις οποίες τα δεδομένα είναι 10 φορές περισσότερο πιθανά.

Εξάλλου, αν $R(\theta_1)$ είναι σχετικά μεγάλο, π.χ. $R(\theta_1) \geq 0.5$, το θ_1 είναι μια αρκετά πιθανή τιμή της παραμέτρου, δοθέντος ότι δίνει στα δεδομένα τουλάχιστον το 50% της μέγιστης πιθανότητας που μπορεί να επιτευχθεί κάτω από το συγκεκριμένο μοντέλο. Επομένως, η συνάρτηση σχετικής πιθανοφάνειας ιεραρχεί όλες τις δυνατές τιμές της παραμέτρου ανάλογα με την πιθανότητα που έχουν κάτω από το φως των δεδομένων που παρατηρήθηκαν.

Στα περισσότερα παραδείγματα που συναντάμε στην πράξη το $\hat{\theta}$ υπάρχει, είναι μονοσήμαντα ορισμένο και επομένως ο ορισμός της συνάρτησης σχετικής πιθανοφάνειας μπορεί να εφαρμοσθεί.

Γενικότερα, η συνάρτηση σχετικής πιθανοφάνειας μπορεί να ορισθεί ως ο λόγος του $L(\theta)$ προς το supremum της $L(\theta)$ ως προς όλες τις τιμές της παραμέτρου. Δηλαδή,

$$R(\theta) = L(\theta) / \sup_{\theta} L(\theta)$$

Δοθέντος ότι

$$L(\theta) = k P(E; \theta), \quad \text{όπου} \quad P(E; \theta) \leq 1$$

το supremum αυτό είναι πεπερασμένο.

Επομένως, η συνάρτηση σχετικής πιθανοφάνειας υπάρχει πάντοτε και μπορεί να χρησιμοποιηθεί για τον καθορισμό του πόσο εύλογη είναι μια τιμή της παραμέτρου ακόμα και όταν η εκτιμήτρια μέγιστης πιθανοφάνειας $\hat{\theta}$ δεν υπάρχει.

ΠΙΘΑΝΟΦΑΝΕΙΕΣ ΔΥΟ ΠΑΡΑΜΕΤΡΩΝ

Στην προηγούμενη ενότητα, εξετάσαμε την χρησιμοποίηση της μεθόδου μέγιστης πιθανοφάνειας για την εκτίμηση μιας παραμέτρου ενός πληθυσμού. Θα εξετάσουμε τώρα την περίπτωση που το μοντέλο πιθανότητας που περιγράφει τον υπό μελέτη πληθυσμό έχει δύο παραμέτρους, έστω α και β . Η πιθανότητα του παρατηρηθέντος ενδεχομένου E (δηλαδή τα δεδομένα) θα είναι, εν γένει, συνάρτηση των α και β , έστω $P(E; \alpha, \beta)$. Η (από κοινού) συνάρτηση πιθανοφάνειας των α και β θα είναι ανάλογη αυτής της πιθανότητας,

$$L(\alpha, \beta) = k P(E; \alpha, \beta),$$

όπου k είναι μια θετική ποσότητα ανεξάρτητη από τα α και β .

Η από κοινού εκτιμήτρια μέγιστης πιθανοφάνειας των α και β είναι το ζεύγος των τιμών των παραμέτρων $(\hat{\alpha}, \hat{\beta})$ που μεγιστοποιεί τη συνάρτηση $L(\alpha, \beta)$. Οι τιμές τους μπορούν να βρεθούν με την λύση του συστήματος των εξισώσεων,

$$\frac{\partial l}{\partial \alpha} = \frac{\partial l}{\partial \beta} = 0$$

όπου, ως συνήθως, $\ell(\alpha, \beta) = \ln L(\alpha, \beta)$ είναι ο λογάριθμος της συνάρτησης πιθανοφάνειας. Η εξίσωση μεγίστης πιθανοφάνειας ισχύει μόνο σε κάποιο σημείο σχετικού μεγίστου (*relative maximum*). Προκειμένου να καθορισθεί ένα maximum σημείο που παρατηρείται στο σύνορο (*boundary*) του παραμετρικού χώρου ή όταν μία από τις παραμέτρους περιορίζεται σε τιμές διακριτές, χρειάζονται άλλες μέθοδοι για τον καθορισμό αυτού του μεγίστου σημείου.

Η από κοινού συνάρτηση σχετικής πιθανοφάνειας ορίζεται ως

$$R(\alpha, \beta) = L(\alpha, \beta) / L(\hat{\alpha}, \hat{\beta})$$

έτσι ώστε $0 \leq R(\alpha, \beta) \leq 1$ για όλα τα ζεύγη των τιμών των παραμέτρων (α, β) .

Όπως και στην περίπτωση της μιας παραμέτρου έχουμε,

$$R(\alpha_1, \beta_1) = \frac{\text{Πιθανότητα των δεδομένων όταν } \alpha = \alpha_1 \text{ και } \beta = \beta_1}{\text{Μέγιστη πιθανότητα των δεδομένων για όλα τα } \alpha, \beta}$$

Αν η ποσότητα αυτή είναι μικρή, τότε το ζεύγος των τιμών των παραμέτρων (α_1, β_1) δεν είναι πιθανό να ισχύει, αφού υπάρχει άλλο ζεύγος τιμών των παραμέτρων τέτοιο ώστε τα δεδομένα να είναι πολύ πιο πιθανά. Αυτό, βέβαια, δεν σημαίνει ότι το α_1 δεν αποτελεί μια πιθανή τιμή για το α ή ότι το β_1 δεν αποτελεί μια πιθανή τιμή για το β . Σημαίνει όμως ότι τα α_1 και β_1 ως ζευγάρι δεν εκτιμούν καλά ταυτόχρονα τα α και β . Η από κοινού σχετική συνάρτηση πιθανοφάνειας είναι κατάλληλη για να βγάξει κανείς συμπεράσματα για τις δύο παραμέτρους ταυτόχρονα. Υπάρχουν μέθοδοι με τις οποίες είναι δυνατόν να αποκλείσει κανείς τη μία παράμετρο, έτσι ώστε να έχει τη δυνατότητα να βγάλει συμπέρασμα μόνο για την άλλη παράμετρο.

Ως βοήθεια για την ερμηνεία του $R(\alpha, \beta)$, μπορεί κανείς να κατασκευάσει τη γραφική παράσταση των καμπύλων σταθερού επιπέδου σχετικής πιθανοφάνειας ως προς α και β . (Κατασκευή δηλαδή της γραφικής παράστασης της καμπύλης $R(\alpha, \beta) = c$ για επιλεγμένες τιμές του c π.χ. $c=0.01$, $c=0.1$, $c=0.5$, $c=1$, κ.λ.π.). Η

διαδικασία αυτή οδηγεί στη κατασκευή ενός χάρτη καμπύλων του “βουνού” της σχετικής πιθανοφάνειας. Οι καμπύλες αυτές είναι συνήθως ένα σύνολο κλειστών καμπύλων με σχήμα περίπου ελλειπτικό.

Επάρκεια

Ορισμός: Μια οικογένεια στατιστικών συναρτήσεων T_1, T_2, \dots, T_k λέγεται *επαρκής* (*sufficient*) για τις παραμέτρους α και β αν γνώση των παρατηρηθεισών τιμών των T_1, T_2, \dots, T_k για το δείγμα που έχει επιλεγεί είναι επαρκής για τον καθορισμό της συνάρτησης πιθανοφάνειας $L(\alpha, \beta)$ μέχρι μια σταθερά αναλογικότητας.

Αν θεωρήσουμε ως θ το διάνυσμα των παραμέτρων (α, β) , τότε ο ορισμός της επάρκειας για δύο παραμέτρους είναι η φυσική επέκταση του ορισμού της επάρκειας για μία παράμετρο που έχει δοθεί στην αρχή του κεφαλαίου. Συγκεκριμένα, η εκτιμήτρια μεγίστης πιθανοφάνειας $(\hat{\alpha}, \hat{\beta})$ μπορεί να καθορισθεί από οποιοδήποτε σύνολο επαρκών στατιστικών συναρτήσεων. Σε μερικές περιπτώσεις, όπως π.χ. στην περίπτωση της κανονικής κατανομής, το διάνυσμα $(\hat{\alpha}, \hat{\beta})$ είναι αυτό καθαυτό επαρκές για το διάνυσμα (α, β) .

ΕΚΤΙΜΗΤΡΙΕΣ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ μ ΚΑΙ σ^2 ΤΗΣ ΚΑΝΟΝΙΚΗΣ ΚΑΤΑΝΟΜΗΣ

Θεωρούμε ένα τυχαίο δείγμα X_1, X_2, \dots, X_n n ανεξαρτήτων παρατηρήσεων από μία κανονική κατανομή με άγνωστη μέση τιμή μ και άγνωστη διασπορά σ^2 .

Ο παραμετρικός χώρος στην περίπτωση αυτή είναι

$$\Theta = \{(\mu, \sigma^2), -\infty < \mu < +\infty, 0 < \sigma^2 < \infty\}$$

Η συνάρτηση πιθανοφάνειας της κανονικής κατανομής είναι

$$L(\mu, \sigma^2) = k \prod_{i=1}^n f(x_i)$$

$$= k \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\}$$

(Με την υπόθεση ότι τα διαστήματα των μετρήσεων είναι μικρά).
Με κατάλληλη επιλογή του k , έχουμε ότι η συνάρτηση πιθανοφάνειας μπορεί να γραφεί ως,

$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{\sigma^2})^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \right\}$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι,

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2}$$

Οι μερικές παράγωγοι πρώτης τάξεως της συνάρτησης αυτής ως προς μ και σ^2 δίνουν,

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^4}$$

Επομένως, οι εκτιμήτριες μέγιστης πιθανοφάνειας των μ και σ^2 αντίστοιχα είναι οι λύσεις των εξισώσεων,

$$\sum_{i=1}^n (X_i - \mu) = 0$$

$$-\frac{n}{2} + \frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2} = 0$$

Δηλαδή,

$$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{και} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Αν κανείς θεωρήσει τις μερικές παραγώγους δευτέρας τάξεως, μπορεί να αποδείξει ότι τα ακρότατα αυτά πράγματι αποτελούν ένα μέγιστο. Επομένως, οι εκτιμήτριες μεγίστης πιθανοφάνειας των μ και σ^2 είναι

$$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = S^2$$

Σημείωση: Μια πρώτη ματιά ίσως οδηγήσει στο συμπέρασμα ότι η συνάρτηση πιθανοφάνειας εξαρτάται από όλες τις παρατηρηθείσες τιμές x_1, x_2, \dots, x_n . Εύκολα όμως μπορούμε να διαπιστώσουμε ότι στην πραγματικότητα μόνο δύο συναρτήσεις των παρατηρήσεων αυτών απαιτούνται για τον καθορισμό της συνάρτησης πιθανοφάνειας.

Πράγματι, δοθέντος ότι,

$$\begin{aligned} (X_i - \mu)^2 &= (X_i - \bar{X} + \bar{X} - \mu)^2 \\ &= (X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + 2(X_i - \bar{X})(\bar{X} - \mu) \end{aligned}$$

έχουμε,

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \bar{X})$$

Αλλά,

$$\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n X_i - n\bar{X} = 0$$

Επομένως,

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 = n\hat{\sigma}^2 + n(\hat{\mu} - \mu)^2$$

Και επομένως, η συνάρτηση πιθανοφάνειας είναι,

$$L(\mu, \sigma^2) = \sigma^{-n} \exp \left\{ -\frac{n}{2\sigma^2} (\hat{\sigma}^2 + (\hat{\mu} - \mu)^2) \right\},$$

$$-\infty < \mu < +\infty, \quad 0 < \sigma^2 < \infty$$

Υποθέτοντας ότι το n είναι γνωστό εκ των προτέρων, είναι προφανές ότι γνώση των $\hat{\mu}, \hat{\sigma}^2$ είναι επαρκής για τον καθορισμό της συνάρτησης πιθανοφάνειας. Επομένως, $(\hat{\mu}, \hat{\sigma}^2)$ είναι ένα ζευγάρι επαρκών στατιστικών συναρτήσεων για τις παραμέτρους (μ, σ^2) . Το μέγιστο της συνάρτησης πιθανοφάνειας είναι,

$$L((\hat{\mu}, \hat{\sigma}^2)) = \frac{1}{(\hat{\sigma}^2)^{n/2}} e^{-n/2}$$

και, επομένως, η από κοινού συνάρτηση σχετικής πιθανοφάνειας των μ και σ^2 είναι

$$\begin{aligned} R(\mu, \sigma^2) &= \frac{L(\mu, \sigma^2)}{L(\hat{\mu}, \hat{\sigma}^2)} \\ &= \left(\frac{\hat{\sigma}^2}{\sigma^2} \right)^{n/2} \exp \left\{ \frac{n}{2} \left[1 - \frac{\hat{\sigma}^2}{\sigma^2} - \frac{(\hat{\mu} - \mu)^2}{\sigma^2} \right] \right\} \end{aligned}$$

ΥΠΟΛΟΓΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ

Σε μερικές περιπτώσεις, όπως αυτή που μόλις εξετάσαμε, είναι δυνατόν να καταλήξει κανείς σε αλγεβρικές λύσεις των εξισώσεων μεγίστης πιθανοφάνειας. Σε άλλες περιπτώσεις, είναι ενδεχόμενο να εξαλείψει την μία παράμετρο από τις εξισώσεις και στην συνέχεια να χρησιμοποιήσει την μέθοδο Raphson, που ήδη εξετάσαμε, για την απομένουσα εξίσωση. Άλλες φορές πάλι ένας αμφιμονοσήμαντος μετασχηματισμός των παραμέτρων μπορεί να βοηθήσει στο να εκφραστούν οι εξισώσεις σε μια πιο απλή έκφραση.

Εάν οι εξισώσεις δεν είναι εύκολο να απλοποιηθούν, χρησιμοποιούμε συνήθως την μέθοδο Newton-Raphson. Σύμφωνα με την μέθοδο αυτή, αρχίζουμε με μια πρώτη προσέγγιση την οποία διαρκώς βελτιώνουμε. Αν (α_i, β_i) είναι η προσέγγιση στο $(\hat{\alpha}, \hat{\beta})$ την οποία βρήκαμε στην i επανάληψη, τότε,

$$(\alpha_{i+1}, \beta_{i+1}) = (\alpha_i, \beta_i) - \left(\frac{\partial \ell}{\partial \alpha}, \frac{\partial \ell}{\partial \beta} \right) C^{-1}$$

όπου, $C = (c_{jk})$ είναι ο πίνακας των μερικών παραγώγων δευτέρας

τάξης:

$$c_{11} = \frac{\partial^2 \ell}{\partial \alpha^2}; \quad c_{12} = c_{21} = \frac{\partial^2 \ell}{\partial \alpha \partial \beta}; \quad c_{22} = \frac{\partial^2 \ell}{\partial \beta^2}$$

Όλες οι παράγωγοι υπολογίζονται στα σημεία $\alpha = \alpha_j$ και $\beta = \beta_j$. Όταν οι παράγωγοι είναι πολύ πολύπλοκες ή όταν η συνάρτηση πιθανοφάνειας δεν συμπεριφέρεται "καλά" κοντά στο μέγιστο, οι εκτιμήτριες $\hat{\alpha}$, $\hat{\beta}$ μπορούν να υπολογισθούν με διπλή εφαρμογή της μεθόδου που αναπτύξαμε στη προηγούμενη ενότητα. Αυτό γιατί,

$$\max_{\alpha, \beta} \ell(\alpha, \beta) = \max_{\alpha} \left\{ \max_{\beta} \ell(\alpha, \beta) \right\} = \max_{\alpha} g(\alpha)$$

όπου g είναι μια συνάρτηση μόνο του α

$$g(\alpha) = \max_{\beta} \ell(\alpha, \beta)$$

Έτσι γράφουμε, ή χρησιμοποιούμε, στην αρχή ένα πρόγραμμα στον υπολογιστή για να υπολογίσουμε το $g(\alpha)$ για όλες τις τιμές του α . Αυτό συνεπάγεται μεγιστοποίηση ως προς το β μόνο. Όταν είμαστε σε θέση να προσδιορίσουμε το $g(\alpha)$, μπορούμε να προσδιορίσουμε το μέγιστό του είτε με γραφικές μεθόδους είτε με τη μέθοδο που αναπτύξαμε στη προηγούμενη ενότητα.

ΓΕΝΙΚΑ ΠΕΡΙ ΣΗΜΕΙΑΚΩΝ ΕΚΤΙΜΗΤΡΙΩΝ

Η επιλογή της μεθόδου καθορισμού της σημειακής εκτιμήτριας μίας ή περισσότερων παραμέτρων ενός πληθυσμού εξαρτάται από τις ιδιότητες που είναι επιθυμητό να έχει η συγκεκριμένη εκτιμήτρια στο υπό εξέταση πρόβλημα. Μερικές από τις ιδιότητες αυτές έχουν αναφερθεί στην αρχή του βιβλίου.

Εν γένει, θεωρείται ότι η μέθοδος μέγιστης πιθανοφάνειας είναι η καλύτερη από τις μεθόδους που αναφέρθηκαν. Αυτό, κυρίως, γιατί οι εκτιμήτριες μέγιστης πιθανοφάνειας ακολουθούν, ασυμπτωτικά, την κανονική κατανομή και επομένως είναι εύκολο να μελετηθούν. Η κυριότερη "δικαίωση" της μεθόδου της μέγιστης πιθανοφάνειας, σε ότι αφορά τα κριτήρια της αμεροληψίας και της ελάχιστης διασποράς, είναι ότι μπορεί να αποδειχθεί ότι, για μεγάλα δείγματα και κάτω από ήπιες γενικές συνθήκες, οι εκτιμήτριες

μέγιστης πιθανοφάνειας είναι σχεδόν αμερόληπτες και έχουν διασπορές σχεδόν ίσες με το φράγμα των Cramér-Rao. Το μειονέκτημα των εκτιμητριών μέγιστης πιθανοφάνειας είναι ότι για να προσδιορισθούν θα πρέπει να είναι γνωστή η μορφή της κατανομής του πληθυσμού.

Αντίθετα, η μέθοδος των ελαχίστων τετραγώνων δεν απαιτεί κάποιες υποθέσεις, ή κάποια γνώση της κατανομής του πληθυσμού.

Η μέθοδος των ελαχίστων τετραγώνων έχει, όπως και η μέθοδος μέγιστης πιθανοφάνειας, μερικές ασυμπτωτικές βέλτιστες ιδιότητες. Εν γένει, όμως δεν δίνει αμερόληπτες εκτιμήτριες.

Για τις εκτιμήτριες μέγιστης πιθανοφάνειας ισχύει ότι κάτω από πολύ γενικές συνθήκες είναι συνεπείς. Δεν είναι όμως πάντα αμερόληπτες. Πολλές φορές επίσης η εκτιμήτρια μέγιστης πιθανοφάνειας δεν είναι μονοσήμαντα ορισμένη.

Η μέθοδος των ελαχίστων τετραγώνων χρησιμοποιείται, κυρίως, στη θεωρία της γραμμικής παλινδρόμησης όπου και θα εξετασθεί αναλυτικότερα.

Το πρόβλημα της επιλογής της “καλύτερης” εκτιμήτριας $\hat{\theta}$ μιας παραμέτρου θ δεν είναι δυνατόν να ορισθεί μονοσήμαντα εκτός και αν ξέρουμε την ακριβή χρήση της εκτιμήτριας που θέλουμε να υπολογίσουμε. Για παράδειγμα, δεν είναι πάντοτε απαραίτητο, και μερικές φορές δεν έχει και νόημα, να απαιτούμε μια εκτιμήτρια να είναι αμερόληπτη. Αυτό γιατί πολλές φορές η απαίτηση της αμεροληψίας είναι πολύ ισχυρή έτσι ώστε να αποκλείει πολλές άλλες επιθυμητές ιδιότητες. Παρότι η ιδιότητα της αμεροληψίας έχει μαθηματικό ενδιαφέρον, δεν έχει πάντοτε το ίδιο ενδιαφέρον στα πρακτικά προβλήματα. Όταν μιλάμε για εκτιμήτριες παραμέτρων οι οποίες δεν είναι αμερόληπτες θα ήταν ενδιαφέρον, από πρακτικής πλευράς, να ξέρουμε αν οι εκτιμήτριες αυτές υπερεκτιμούν ή υποεκτιμούν την υπό εκτίμηση παράμετρο θ και το μέγεθος της απόκλισης αυτής. Η χρησιμοποίηση της μεθόδου των ελαχίστων τετραγώνων όμως δεν παίρνει υπόψη της τις πλευρές αυτές, αφού μεταχειρίζεται με τον ίδιο τρόπο και δίνει την ίδια σοβαρότητα σε υπερεκτίμηση και υποεκτίμηση. Υπάρχουν όμως βέβαια πολλές περιπτώσεις όπου οι δύο αυτές καταστάσεις δεν είναι εξίσου

σημαντικές. Στις περιπτώσεις βέβαια αυτές, δεν έχει έννοια να χρησιμοποιήσει κανείς τη μέθοδο των ελαχίστων τετραγώνων.

Οι δυσκολίες αυτές που προαναφέραμε όπως και άλλες παρόμοιες σε συνδυασμό με τη μεγάλη πρόοδο των ηλεκτρονικών υπολογιστών έχουν οδηγήσει την επιστήμη της Στατιστικής να χρησιμοποιεί πολλές φορές ad hoc μεθόδους εκτίμησης που είναι κατάλληλες για το συγκεκριμένο πρόβλημα και όχι τις κλασικές τις κυριότερες από τις οποίες αναφέραμε.

Ως τελικό συμπέρασμα θα μπορούσε κανείς να πει ότι η επιλογή της “καλύτερης” εκτιμήτριας είναι ένα πρόβλημα της Στατιστικής Θεωρίας Αποφάσεων μάλλον παρά της Στατιστικής Συμπερασματολογίας.