Η BAYESIAN ΑΝΑΛΥΣΗ ΣΤΗ ΜΕΤΡΟΛΟΓΙΑ ΚΑΙ ΣΤΗΝ ΕΚΤΙΜΗΣΗ ΤΩΝ ΑΒΕΒΑΙΟΤΗΤΩΝ

Ε. Μαθιουλάκης και Β. Μπελεσιώτης
Εργαστήριο Ηλιακών & άλλων Ενεργειακών Συστημάτων – ΕΚΕΦΕ «ΔΗΜΟΚΡΙΤΟΣ»
15310 Αγ. Παρασκευή Αττικής
sollab@ipta.demokritos.gr

Περίληψη
Παρά το ότι η ρητή αναφορά στην Bayesian ανάλυση είναι φαινομενικά αρκετά περιορισμένη στην τρέχουσα βιβλιογραφία, είναι δεδομένη η σημαντική της επιρροή στη διαμόρφωση των σύγχρονων μεθοδολογικών προσεγγίσεων για την εκτίμηση της ποιότητας των μετρητικών αποτελεσμάτων και των αβεβαιοτήτων. Στην ανάλυση αυτή βασίζονται για παράδειγμα καίριας σημασίας έννοιες, όπως οι αβεβαιότητες Τύπου Β και ο νόμος διάχυσης σφαλμάτων. Παράλληλα όμως, ειδικά τα τελευταία χρόνια, η στατιστική Bayesian βρίσκεται στο επίκεντρο μιας συνεχώς διευρυνόμενης συζήτησης σχετικά με τις αδυναμίες και τις ανεπάρκειες της συχνοτικής προσέγγισης της μέτρησης και, κατ' επέκταση, του GUM (*Guide to the expression of Uncertainty in Measurements,* [1]).
Σκοπός αυτής της εργασίας είναι να παρουσιάσει τις βασικές αρχές της Bayesian ανάλυσης και να συζητήσει τις επιπτώσεις από την εφαρμογή της στον τομέα της ερμηνείας των εμπειρικών παρατηρήσεων γενικά και της μετρολογίας ειδικότερα. Η συζήτηση αυτή τοποθετείται στο ευρύτερο πλαίσιο που διαμορφώθηκε από τη διαχρονική διαμάχη για τη σχέση πιθανοτήτων και στατιστικής.
Πιο συγκεκριμένα, η εργασία αυτή εξετάζει τις προϋποθέσεις και τα εργαλεία αποτελεσματικής αξιοποίησης των παρατηρούμενων πειραματικών αποτελεσμάτων σε μια διαδικασία συνεχούς βελτίωσης και αναβάθμισης της εικόνας που διαθέτει ο παρατηρητής για την αληθή τιμή του μετρούμενου μεγέθους, με βάση τις επικαιροποιημένες πληροφορίες που απορρέουν από την τρέχουσα μετρητική διαδικασία.. Στα εργαλεία αυτά εντάσσεται επιπλέον και η δυνατότητα που προσφέρει η Bayesian ανάλυση σχετικά με την αξιοποίηση των συχνά παρεξηγημένων υποκειμενικών εκτιμήσεων ή κρίσεων, μέσω μιας φυσικής διεργασίας κατά την οποία οι πιθανότητες εκλαμβάνονται ως βαθμός πεποίθησης ότι ένα γεγονός θα συμβεί και, επομένως, ως συνάρτησης των γνώσεων που διαθέτουμε για το γεγονός αυτό.

1. Εισαγωγή
Είναι σήμερα κοινά αποδεκτό ότι μια ποσοτική δήλωση σχετική με το αποτέλεσμα μιας μέτρησης δεν μπορεί να θεωρηθεί πλήρης εάν δεν συμπεριλαμβάνει μια αναφορά στην αβεβαιότητα που χαρακτηρίζει το αποτέλεσμα αυτό. Αν και η μεθοδολογία εκτίμησης των αβεβαιοτήτων στις μετρήσεις ήταν για ένα μεγάλο διάστημα αντικείμενο διαμάχης, τα τελευταία χρόνια έγινε μια συστηματική προσπάθεια σύγκλισης των απόψεων σε διεθνές επίπεδο γύρω από το ζήτημα αυτό, προσπάθεια που κατέληξε το 1995 στη δημοσίευση από τον ISO και άλλους διεθνείς οργανισμούς του Οδηγού GUM.
Παρά το ότι ο GUM συνιστά αναμφίβολα σημείο αναφοράς, η συζήτηση γύρω από τις μεθοδολογικές προσεγγίσεις στην εκτίμηση των αβεβαιοτήτων παραμένει επίκαιρη. Ορισμένα μάλιστα από τα ζητήματα που τίθενται σε δημόσια συζήτηση στα πλαίσια της μετρολογικής κοινότητας αφορούν διαχρονικές αντιπαραθέσεις με σημαντικές πρακτικές

και φιλοσοφικές προεκτάσεις. Ένα από τα ζητήματα αυτά, ίσως το σημαντικότερο, είναι το πεδίο και ο τρόπος εφαρμογής μεθοδολογιών που εντάσσονται στην συλλογιστική της κλασικής στατιστικής ή της bayesian προσέγγισης, καθώς και η μεταξύ τους σχέση.
Η όποια αντιπαράθεση μεταξύ κλασικής στατιστικής και bayesian προσέγγισης δεν είναι καινούργια, ούτε αφορά μόνο την εκτίμηση των αβεβαιοτήτων. Παραπέμπει ευθέως στην συζήτηση για τη σχέση μεταξύ, από τη μια, της θεωρίας των πιθανοτήτων όπως αυτή διατυπώθηκε στα τέλη του 17^{ου} αιώνα ως θεωρία του αβέβαιου στις εργασίες των Fermat, Pascal, Bernoulli και, από την άλλη, της στατιστικής. Η συζήτηση αυτή προσέλαβε νέες διαστάσεις στις αρχές του 20^{ου} αιώνα [2] και παραμένει ανοικτή ακόμα και στις μέρες μας [3, 4], με προνομιακό πεδίο αντιπαράθεσης την επεξεργασία των εμπειρικών παρατηρήσεων γενικά και τη μετρολογία ειδικότερα.
Η συζήτηση σχετικά με το πλαίσιο στο οποίο εντάσσονται οι διαφορετικές προσεγγίσεις που προαναφέρθηκαν δεν είναι ακαδημαϊκή, αλλά έχει σημαντικές πρακτικές επιπτώσεις τόσο στον υπολογισμό της αβεβαιότητας αυτής καθεαυτής, όσο και στην ερμηνεία της αβεβαιότητας ως δείκτη αξιοπιστίας του αποτελέσματος.
Με την εργασία αυτή επιχειρείται η ανάδειξη των ιδιαιτεροτήτων της μιας ή της άλλης προσέγγισης, να συζητήσει τα σχετικά τους μειονεκτήματα και πλεονεκτήματα και να προτείνει δυνατότητες πρακτικής αξιοποίησης της bayesian συλλογιστικής στα προβλήματα εκτίμησης της αβεβαιότητας και της αξιοπιστίας του αποτελέσματος, ειδικότερα όταν η εκτίμηση αυτή βασίζεται σε υποκειμενικές κρίσεις και πεποιθήσεις του μετρολόγου.

2. Πιθανότητες και αβεβαιότητα – Η Bayesian προσέγγιση και η σχέση της με την κλασική στατιστική ανάλυση

Σε ένα κόσμο αβεβαιότητας, οι εμπειρικές παρατηρήσεις δεν επιτρέπουν την διαμόρφωση απολύτως σίγουρων συμπερασμάτων. Όλες οι εκδοχές δεν είναι όμως εξίσου αβέβαιες, ορισμένες από αυτές είναι πιθανότερες από άλλες, γεγονός που επιβάλει ένα είδος διαβάθμισης της αβεβαιότητας που χαρακτηρίζει την κάθε μια από αυτές. Σε πολλές περιπτώσεις, όπως για παράδειγμα στο χώρο της μετρολογίας, η διαβάθμιση αυτή, για να είναι πρακτικά αξιοποιήσιμη, είναι αναγκαίο να είναι ποσοτική. Η ποσοτικοποιημένη διαβάθμιση των πιθανοτήτων που χαρακτηρίζουν τις πιθανές εκδοχές ενός συμβάντος έχει επικρατήσει να αποκαλείται συνάρτηση πυκνότητας πιθανότητας ή απλούστερα Συνάρτηση Κατανομής Πιθανοτήτων.
Δύο είναι τα κύρια ρεύματα κατανόησης και ερμηνείας των πιθανοτήτων, όπως έχουν διαμορφωθεί ιστορικά με την πάροδο του χρόνου: η κλασική στατιστική και η bayesian ανάλυση. Η πρώτη συνδέει την πιθανότητα με τη συχνότητα εμφάνισης ενός γεγονότος σε μια σειρά ανεξάρτητων επαναλήψεων σε παρόμοιες συνθήκες (π.χ. η εμφάνιση της μιας ή της άλλης όψης κατά τη ρίψη ενός νομίσματος). Η δεύτερη παραπέμπει στην πιθανότητα ως μέτρο, αναγκαστικά υποκειμενικό, της πεποίθησης ότι μπορεί ένα συμβάν να αποδειχθεί αληθινό ή όχι (π.χ. η πιθανότητα να βρέξει αύριο), όπου η πεποίθηση αυτή βασίζεται όχι στην επανάληψη της εμπειρικής παρατήρησης αλλά σε άλλες σχετικές πληροφορίες.
Είναι γεγονός ότι η λεγόμενη "*συχνοτική*" προσέγγιση (*frequentist*) που βασίζεται στην κλασική στατιστική είναι περισσότερο οικεία στο χώρο της μετρολογίας, εν μέρει λόγω της έμφασης που δίδεται σε αυτή από τον ίδιο το GUM. Είναι για παράδειγμα εμφανής η άμεση συγγένεια των Τύπου Α αβεβαιοτήτων με την προσέγγιση αυτή. Είναι αντίθετα λιγότερο εμφανής, αν και απολύτως υπαρκτή, η σχέση των Τύπου Β αβεβαιοτήτων με την bayesian προσέγγιση [5, 6, 7].

2.1. Η κλασική συχνοτική στατιστική

Η κλασική «συχνοτική» στατιστική επεξεργάζεται στατιστικές πιθανότητες, δηλαδή συχνότητες εμφάνισης συμβάντων. Αντίθετα με τον υποκειμενικό χαρακτήρα της *πεποίθησης* που χαρακτηρίζει τις bayesian πιθανότητες, η στατιστική προσέγγιση έχει ως πρόθεση τον αντικειμενικό χαρακτηριστικό επαναλαμβανόμενων παρατηρήσεων. Στα πλαίσια της συλλογιστικής αυτής αναπτύχθηκαν τεχνικές που αποσκοπούν είτε στο να δοκιμαστούν υποθέσεις χρησιμοποιώντας δοκιμές σημαντικότητας (significance tests), είτε στο να εκτιμηθούν οι τιμές αγνώστων παραμέτρων που χαρακτηρίζουν στατιστικά το δείγμα των παρατηρήσεων, όπως η πιθανότερη τιμή (ή καλύτερη εκτίμηση), η τυπική απόκλιση και τα διαστήματα εμπιστοσύνης.

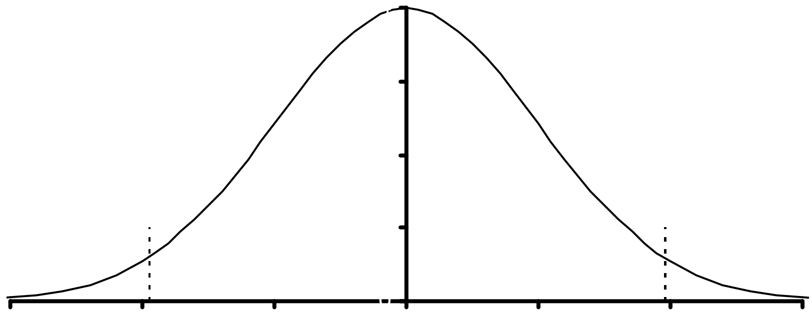
Σύμφωνα με την κλασική στατιστική, η τιμή του μετρούμενου μεγέθους, αποκαλούμενη συμβατικά *Αληθής Τιμή*, θεωρείται ως μια άγνωστη σταθερά και το αποτέλεσμα της μέτρησης ως τυχαία μεταβλητή που ακολουθεί μια ΣΚΠ. Αυτό που παράγεται από την επανάληψη της μέτρησης σε συνθήκες ελεγχόμενης επαναληψιμότητας είναι μια κατανομή δειγματοληψίας που περιγράφει τη συχνότητα εμφάνισης της μιας ή της άλλης τιμής. Στην πράξη ο μετρολόγος, αφού υποθέσει ένα είδος κατανομής, παράγει μια εκτίμηση της ΑΤ και μια εκτίμηση για την τυπική της απόκλιση. Βασικό χαρακτηριστικό της προσέγγισης αυτής είναι ότι οι εκτιμήσεις που παράγει ισχύουν μόνο για τις συνθήκες της μέτρησης και δεν λαμβάνουν υπόψη πιθανές άλλες πηγές πληροφορίας (προηγούμενη εμπειρία, δημοσιευμένα δεδομένα κλπ).

Ας υποθέσουμε ότι για τον προσδιορισμό του άγνωστου μεγέθους μ, οι τιμές του οποίου εμφανίζουν μια διασπορά λόγω τυχαίων επιρροών ακολουθώντας μια κανονική κατανομή, λαμβάνονται τυχαία δείγματα τιμών. Οι ανεξάρτητες μετρήσεις του κάθε δείγματος, εάν το πλήθος τους είναι επαρκές, ακολουθούν επίσης μια κανονική κατανομή δειγνατοληψίας γύρω από τη μέση τιμή $\overline{X}$, με τυπική απόκλιση ίση με την τυπική απόκλιση του συνολικού πληθυσμού των πιθανών τιμών του μεγέθους. Να σημειωθεί ότι η *κατανομή δειγματοληψίας* είναι μια ειδική κατανομή πιθανοτήτων που περιγράφει τη σχετική συχνότητα εμφάνισης όλων των πιθανών τιμών στις συνθήκες λήψης του δείγματος.

Ας υποθέσουμε ακόμα ότι ένα συγκεκριμένο δείγμα n ανεξάρτητων μετρήσεων $x_1,\ldots,x_n$ έχει μέση τιμή $\bar{x}$ και τυπική απόκλιση s. Είναι σαφές ότι κάθε δείγμα παράγει τη δική του μέση τιμή και επομένως η μεταβλητή "μέση τιμή του δείγματος" (με το σύμβολο $\overline{X}$) είναι και αυτή μια τυχαία μεταβλητή που μπορεί να λάβει πολλές διαφορετικές τιμές, μερικές πιο πιθανές από άλλες. Μια άλλη παρατήρηση, δηλαδή η λήψη ενός άλλου δείγματος τιμών, θα οδηγήσει σε μια διαφορετική μέση τιμή. Η διασπορά αυτή των τιμών του $\overline{X}$ μπορεί να περιγραφεί επίσης από μια κανονική κατανομή γύρω από το μ. Σύμφωνα με την κλασική στατιστική (Τύπου Α προσέγγιση), το $\bar{x}$ είναι η καλύτερη εκτίμηση του μ, ενώ η τυπική του απόκλιση $\sigma = s/\sqrt{n}$ εκλαμβάνεται ως τυπική απόκλιση ή τυπική αβεβαιότητα (Τύπου Α) που χαρακτηρίζει την εκτίμηση αυτή. Το αποτέλεσμα της μέτρησης διατυπώνεται συνήθως ως $\mu = \bar{x} \pm k_p\sigma$. Η ποσότητα $k_p\sigma$, αποκαλούμενη επίσης διευρυμένη αβεβαιότητα, είναι συνάρτηση της επιθυμητής πιθανότητας κάλυψης *p%* και ορίζει ένα διάστημα ($\bar{x} - k_p\sigma$, $\bar{x} + k_p\sigma$) το οποίο προτείνεται συχνά ως διάστημα εμπιστοσύνης για την αληθή τιμή [8].

Αυτό που δημιουργεί πρόβλημα στην περίπτωση της Τύπου Α προσέγγισης, είναι η ερμηνεία που αποδίδεται στο διάστημα ($\bar{x} - k_p\sigma$, $\bar{x} + k_p\sigma$) και η δήλωση που γίνεται συνήθως ότι κάποιος μπορεί να είναι «95% σίγουρος» (95% confident), ότι το μ περιέχεται στο διάστημα αυτό.

Στην πραγματικότητα, ο μόνος θεωρητικά έγκυρος ισχυρισμός, με βάση τις υπάρχουσες πληροφορίες αφορά την κατανομή της μέσης τιμής $\overline{X}$ ως τυχαίας μεταβλητής, και συγκεκριμένα την πιθανότητα *p%* η μέση τιμή $\overline{X}$ να περιέχεται σε ένα καθορισμένο διάστημα γύρω από την αληθή τιμή μ:



Σχήμα 1: Κανονική κατανομή μέσης τιμής δείγματος

$$p(\mu - k_p\sigma \leq \overline{X} \leq \mu + k_p\sigma) = p\% \qquad (1)$$

Πολλές από τις συχνά προτεινόμενες ερμηνείες του αποτελέσματος $\mu = \bar{x} \pm k_p\sigma$ είναι εσφαλμένες, με χαρακτηριστικά παραδείγματα τις ακόλουθες:

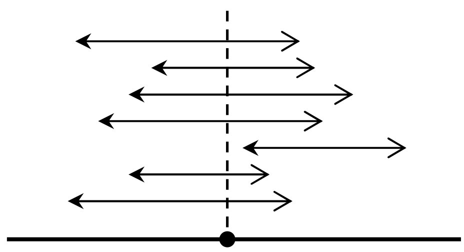
- ότι η πιθανότητα η αληθής τιμή να περιέχεται διάστημα $(\bar{x} - k_p\sigma, \bar{x} + k_p\sigma)$ είναι *p%*

$$p(\bar{x} - k_p\sigma \leq \mu \leq \bar{x} + k_p\sigma) = p\% \qquad (2)$$

- ότι εάν η μέτρηση επαναληφθεί πολλές φορές, στις περίπου *p%* των περιπτώσεων η μέση τιμή του κάθε δείγματος θα περιέχεται στο διάστημα $(\bar{x} - k_p\sigma, \bar{x} + k_p\sigma)$.

Στην πράξη προτείνεται συχνά ως ένα εύλογο διάστημα εμπιστοσύνης για την αληθή τιμή το $\bar{x} - k_p\sigma \leq \mu \leq \bar{x} + k_p\sigma$, στη βάση δύο υποθέσεων οι οποίες δεν τεκμηριώνονται πάντα επαρκώς:

- αντικαθιστώντας το $\overline{X}$ από τη συγκεκριμένη του εκτίμηση $\bar{x}$, όπως αυτή υπολογίστηκε από το συγκεκριμένο δείγμα τιμών
- θεωρώντας ότι μια εκτίμηση της απόστασης του μ από το $\bar{x}$ οδηγεί αυτόματα σε μια ανάλογη εκτίμηση της απόστασης του $\bar{x}$ από το μ, πράγμα όμως που προϋποθέτει την ύπαρξη συμμετρικής συνάρτησης κατανομής.



Σχήμα 2: Η τιμή του μετρούμενου μεγέθους περιέχεται στα p% των ανεξάρτητων διαστημάτων εμπιστοσύνης $\left(\mu - k_p s,\ \mu + k_p s\right)$

Στην πραγματικότητα, το μόνο ασφαλές συμπέρασμα είναι ότι, εάν η μέτρηση (δειγματοληψία) επαναλαμβανόταν 100 φορές, στις *p* από αυτές το διάστημα $\left(\bar{x} - k_p \sigma, \bar{x} + k_p \sigma\right)$ θα περιέχει το μ (σχήμα 2). Πολύ περισσότερο που, όπως είναι φυσικό, το κάθε συγκεκριμένο διάστημα $\left(\bar{x} - k_p \sigma, \bar{x} + k_p \sigma\right)$ εξαρτάται από την κάθε παρατήρηση (διαφορετικό πλήθος δείγματος, διαφορετικοί κανόνες δειγματοληψίας κλπ).

Ενδεικτικό της σύγχυσης που μπορεί να δημιουργηθεί είναι ότι ο GUM ξεκαθαρίζει ότι το διάστημα $\left(\bar{x} - k_p \sigma, \bar{x} + k_p \sigma\right)$ ή αλλιώς η *διευρυμένη αβεβαιότητα*, δεν πρέπει να εκλαμβάνεται ως δείκτης εμπιστοσύνης και ότι η *πιθανότητα κάλυψης* (*coverage probability* ή *level of confidence*) δεν ταυτίζεται με την κλασική έννοια του *επιπέδου εμπιστοσύνης* (*confidence level*), ακόμα και εάν η διάκριση μεταξύ *level of confidence* και *confidence level* είναι εξαιρετικά προβληματική σε ορισμένες γλώσσες όπως άλλωστε είναι τα ελληνικά.

Σε τελική ανάλυση, με την Τύπου Α προσέγγιση δεν τεκμηριώνεται καμιά πληροφορία σχετικά με το εάν η αληθής τιμή περιέχεται ή όχι στο παραγόμενο από τη μέτρηση διάστημα. Η έννοια του *επιπέδου εμπιστοσύνης* ορίζεται για την περίπτωση που λαμβάνονται ανεξάρτητα δείγματα τιμών της ίδιας μεταβλητής (ίδιος πληθυσμός) και υπολογίζεται μια διευρυμένη αβεβαιότητα για το καθένα από αυτά. Στην περίπτωση λοιπόν αυτή, ένα ορισμένο ποσοστό (confidence level) του συνολικού αριθμού διαστημάτων αναμένεται να περιέχει την άγνωστη αληθή τιμή. Με άλλα λόγια αποτελεί μια ένδειξη για το πόσο συχνά αναμένεται το αποτέλεσμα (δηλαδή το διάστημα εμπιστοσύνης) να είναι σωστό (δηλαδή να περιέχει την αληθή τιμή), εάν η μέτρηση επαναληφθεί πολλές φορές (με τη βοήθεια λήψης ενός δείγματος τιμών κάθε φορά).

Πρέπει να τονιστεί ότι ένα διάστημα εμπιστοσύνης 95% (ή διευρυμένη αβεβαιότητα σε επίπεδο κάλυψης 95%) δεν σημαίνει ότι υπάρχει 95% πιθανότητα το *συγκεκριμένο αυτό διάστημα* να περιέχει την αληθή τιμή, ούτε ότι το 95% των τιμών του δείγματος θα περιέχεται στο εν λόγω διάστημα, ούτε βέβαια ότι η αληθής τιμή είναι περισσότερο πιθανό να βρίσκεται κοντά στο κέντρο του διαστήματος. Όπως επίσης είναι εσφαλμένη η αντίληψη ότι ένα μεγάλο διάστημα σημαίνει χαμηλή εμπιστοσύνη στο αποτέλεσμα.

Από την άλλη βέβαια το Τύπου Α αποτέλεσμα αποτελεί αναμφίβολα ένα δείκτη ποιότητας της μέτρησης στο βαθμό που μεγάλη πιθανότητα κάλυψης σημαίνει ότι, σε περίπτωση επανάληψης της δειγματοληψίας, η πιθανότητα να περιέχεται η αληθής τιμή σε κάποιο από τα παραγόμενα διαστήματα είναι μεγάλη, και άρα είναι μεγάλη και η πιθανότητα κάθε ένα ξεχωριστά να περιέχει την αληθή τιμή.

Ένα από τα σημαντικότερα μειονεκτήματα της κλασικής συχνοτικής Τύπου Α προσέγγισης είναι ότι στην πραγματικότητα αποφαίνεται σχετικά με τη διασπορά των παρατηρήσεων και όχι για την ορθότητα τους. Εάν τα σφάλματα που προκαλούν αυτή τη διασπορά και απομακρύνουν τις παρατηρούμενες τιμές από την αληθή τιμή είναι τυχαία, η μέση τιμή συνιστά μια αποτελεσματική εκτίμηση της αληθούς τιμής. Στις περιπτώσεις όμως που υπάρχουν συστηματικές συνιστώσες σφάλματος, η εκτίμηση δεν είναι αποτελεσματική και παράγεται μια εσφαλμένη εικόνα για το μετρούμενο μέγεθος.

2.2. Η bayesian προσέγγιση

Οι πιθανότητες εμφανίστηκαν αρχικά ως θεωρητικό υπόβαθρο της αβεβαιότητας. Έτσι, μπορεί τα αποτελέσματα *E* μιας εμπειρικής παρατήρησης να μην επιτρέπουν την απόλυτη επιβεβαίωση μιας υπόθεσης *H*, μπορούν όμως να βοηθήσουν στην εκτίμηση της πιθανότητας να είναι αληθής η υπόθεση αυτή, μέσω μιας επαγωγικής λογικής διεργασίας. Η πιθανότητα «η *H* να είναι αληθής με δεδομένα τα αποτελέσματα *E*» συμβολίζεται *p(H/E)* και αντανακλά τη γνώμη που σχημάτισε ο παρατηρητής μετά την παρατήρηση. Σε τελική ανάλυση, σκοπός της πειραματικής – εμπειρικής δραστηριότητας είναι η εκτίμηση αυτής ακριβώς της πιθανότητας [5].

Πολλές από τις δυσκολίες στην κατανόηση των πιθανοτήτων ως θεωρίας της αβεβαιότητας σχετίζονται με την ερμηνεία των πιθανοτήτων ως έννοιας. Η στροφή προς τη στατιστική των τελευταίων δεκαετιών, βοήθησε ώστε να εξασθενίσει η πρωταρχική ερμηνεία των πιθανοτήτων ως ένα μέτρο του <u>*βαθμού της πεποίθησης*</u> ότι ένα γεγονός θα συμβεί, δηλαδή του πόσο σίγουρος είναι κάποιος σχετικά με την αλήθεια μιας εκδοχής, ή ακόμα του πόσο πιστεύει την εκδοχή αυτή. Έτσι, λαμβάνοντας υπόψη τις γνώσεις που διαθέτουμε για ότι επηρεάζει την πραγμάτωση μιας εκδοχής, η αριθμητική πιθανότητα *p* αυτής της εκδοχής είναι ένας πραγματικός αριθμός ο οποίος μπορεί να χρησιμοποιηθεί ως ποσοτικό μέτρο της έντασης της εικασίας ή της προσμονής μας, στη βάση των εν λόγω γνώσεων, ότι το ενδεχόμενο θα πραγματοποιηθεί. Και όπως σημειώνει ο Poincaré:

> *«Εάν δεν υπήρχε άγνοια δεν θα υπήρχαν πιθανότητες αλλά μόνο βεβαιότητες. Η άγνοια μας όμως δεν μπορεί να είναι απόλυτη, ειδάλλως δεν θα υπήρχε επίσης λόγος επίκλησης των πιθανοτήτων. Κατά συνέπεια, το πρόβλημα των πιθανοτήτων μπορεί να κατηγοριοποιηθεί ανάλογα με τη μικρότερη ή μεγαλύτερη έκταση της άγνοιας μας».*

Η απόσταση που χωρίζει την κλασική στατιστική από τις πιθανότητες σχετίζεται και με τη διαφορά ανάμεσα στις εκφράσεις "τυχαία" (random) και "αβέβαια" (uncertain) που χρησιμοποιούνται για μια μεταβλητή: η πρώτη χαρακτηρίζει τον τρόπο εμφάνισης (και παραπέμπει στη συχνότητα εμφάνισης), ενώ η δεύτερη σχετίζεται με τη γνώση ή την έλλειψη γνώσης σχετικά με τη μεταβλητή αυτή.

Αυτό όμως που ενδιαφέρει εδώ δεν είναι το γενικό πρόβλημα των πιθανοτήτων, αλλά η θέση της παρατήρησης γενικά και της μέτρησης ειδικότερα στα πλαίσια ενός ερμηνευτικού σχήματος που επικεντρώνεται στο χειρισμό των αβεβαιοτήτων με βάση τις πιθανότητες. Δεν πρέπει να ξεχνάμε ότι η πειραματική παρατήρηση αποσκοπεί στον προσδιορισμό της τιμής μιας συγκεκριμένης παραμέτρου ή στην επιβεβαίωση μιας θεωρίας που περιγράφει κατά το δυνατόν αποτελεσματικότερα το παρατηρούμενο φαινόμενο ή ακόμα και τα δύο ταυτόχρονα. Από την άποψη αυτή, το ενδιαφέρον εστιάζεται αυτονόητα στην αξιοποίηση του αποτελέσματος της παρατήρησης-μέτρησης σε μια διαδικασία αναβάθμισης και διεύρυνσης των γνώσεων για το μετρούμενο μέγεθος, διαδικασία που περιγράφεται με εντυπωσιακή απλότητα από το νόμο του Bayes.

Το 1763 δημοσιεύτηκε μια μεταθανάτια εργασία του αντικομφορμιστή κληρικού και μαθηματικού Thomas Bayes (1702-1761) στην οποία, βάσει των αξιωμάτων της θεωρίας των πιθανοτήτων, διατυπώθηκε το ομώνυμο θεώρημα που έμελλε να αποτελέσει σημείο εκκίνησης για μια συγκεκριμένη σχολή σκέψης στα προβλήματα επεξεργασίας της πληροφορίας [3, 6]:

$$p(H/E) = \frac{p(E/H)p(H)}{P(E)} \qquad (3)$$

Στην γενική του μορφή το θεώρημα του Bayes συσχετίζει τη προηγούμενη πιθανότητα *p(H)* να είναι αληθής μια υπόθεση *H*, με την αντίστοιχη μεταγενέστερη πιθανότητα

p(H/E), όπως αυτή διαμορφώθηκε υπό το φως της παρατήρησης *E*, συνυπολογίζοντας και την αληθοφάνεια *p(E)* της παρατήρησης αυτής (σχήμα 2).
Στην πραγματικότητα, το θεώρημα του Bayes κωδικοποιεί μια πολύ φυσική διαδικασία κατά την οποία οι υποθέσεις εκτιμώνται υπό το φως τόσο της υφιστάμενης πεποίθησης, όσο και της εμπειρικής παρατήρησης, χωρίς να είναι καμιά από τις δύο αυτές παραμέτρους από μόνη της αποφασιστική, αντανακλώντας τη διαδικασία αναβάθμισης της γνώσης για το παρατηρούμενο μέγεθος, όπως ακριβώς γίνεται και στην πραγματική ζωή. Η πεποίθηση ελέγχεται από την εμπειρική παρατήρηση ενώ η αξιοπιστία της εμπειρικής παρατήρησης ελέγχεται από την υφιστάμενη γνώση. Η bayesian προσέγγιση εμφανίζεται να είναι πιο φυσική και περισσότερο εναρμονισμένη με την συνήθη επιστημονική δραστηριότητα.

$$p(E/H)$$

*Γνώση για το Η **πριν** την παρατήρηση Ε*

$$p(E/H)$$

*Γνώση για το Η **από** την παρατήρηση Ε*

$$\xrightarrow{p(H/E)=\frac{p(E/H)p(H)}{P(E)}}$$

$$p(H/E)$$

*Γνώση για το Η **μετά** την παρατήρηση Ε*

Σχήμα 2: Αναβάθμιση της γνώσης μέσω της παρατήρησης

Η πιθανότητα *p(E/H)*, αποκαλούμενη και *πιθανοφάνεια (likelihood)* αποτελεί μέτρο του πόσο πιθανή είναι η παρατήρηση *E* εάν ισχύει η υπόθεση *H* ή του βαθμού κατά τον οποίο η *H* εξηγεί την εμφάνιση του *E*. Η ποσότητα αυτή προκύπτει επομένως από την παρατήρηση και αντιπροσωπεύει τη γνώση που απορρέει από την παρατήρηση αυτή.

3. Μαθαίνοντας από τα δεδομένα των μετρήσεων

Αυτό που ενδιαφέρει πρακτικά στη μετρολογία είναι το ερώτημα «*πώς μπορούμε να μάθουμε από τα δεδομένα των παρατηρήσεων σε ένα πιθανολογικό πλαίσιο;*». Η αλλιώς «*πώς μπορούμε να πάμε από τα παρατηρούμενα αποτελέσματα στα βαθύτερα αίτια, δηλαδή τα πραγματικά φυσικά μεγέθη;»*.
Η εφαρμογή του θεωρήματος του Bayes σε όλη την κλίμακα των τιμών που είναι δυνατόν να πάρει το μετρούμενο μέγεθος *Q* οδηγεί στο συνδυασμό κατανομών πιθανοτήτων αντί για συνδυασμό πιθανοτήτων μεμονωμένων τιμών, με την πιθανότητα να περιέχεται η αληθής τιμή *Q* μια μεταβλητής στο διάστημα *(q, q+dq)* να είναι *f(q)*. Το θεώρημα του Bayes γράφεται στην περίπτωση αυτή:

$$f(q/d)=\frac{f(d/q)f(q)}{f(d)} \qquad (4)$$

Σαν σημείο εκκίνησης λαμβάνεται η υπάρχουσα γνώση για το παρατηρούμενο (μετρούμενο) μέγεθος, όπως αυτή αντιπροσωπεύεται από την *prior* κατανομή πιθανοτήτων *f(q)*, πριν την μέτρηση. Εάν αυτή η γνώση είναι πολύ αβέβαιη ή ανύπαρκτη,

η κατανομή *f(q)* θα είναι πολύ «ανοικτή» με μεγάλη ή άπειρη τυπική απόκλιση. Το θεώρημα του Bayes επιτρέπει την ενσωμάτωση της γνώσης που προκύπτει από τα δεδομένα *d* της παρατήρησης και τον υπολογισμό μιας νέας *posterior* κατανομής *f(q/d)*, χαρακτηριζόμενης από μικρότερη τυπική απόκλιση. Εισάγεται με άλλα λόγια μια λογική επαναπροσδιορισμού της αβεβαιότητας υπό το φως της νέας γνώσης που παρέχει η παρατήρηση.
Πρέπει να τονιστεί ότι οι εκφράσεις «πριν» και «μετά» (prior και posterior) δεν πρέπει να εκλαμβάνονται με τη στενή χρονική τους έννοια, αλλά ως προσδιοριστικά της αλληλουχίας των ενεργειών στην πορεία επεξεργασίας της πληροφορίας (σχήμα 3).
Η κατανομή *f(d/q)* είναι κάπως δύσκολο να εκληφθεί ως μια συμβατική συνάρτηση κατανομής, δεδομένου ότι αφορά συγκεκριμένα δεδομένα *d* τα οποία σχετίζονται με μια μεταβλητή *q*. Για το λόγο αυτό, η κατανομή αυτή γράφεται συνήθως με διαφορετικό τρόπο, ως *πιθανοφάνεια (likelihood)* $\ell(q/d)$ των δεδομένων *d*:

$$f(q/d) = K\ f(d/q) f(q) \propto \ell(q/d) f(q) \tag{5}$$

όπου *K* είναι μια σταθερά μπορεί να υπολογιστεί από τη συνθήκη κανονικοποίησης της μεταγενέστερης συνάρτησης *f(q/d)* (ολοκλήρωμα ίσο με τη μονάδα).
Εάν η υφιστάμενη γνώση είναι σχεδόν ανύπαρκτη (μεγάλη αβεβαιότητα σχετικά με πιθανές τιμές του *q*), η ποσότητα *f(q)* είναι πρακτικά μια σταθερά που απορροφάται κατά την κανονικοποίηση και η αβεβαιότητα μετά τη μέτρηση είναι πρακτικά ίση με την αβεβαιότητα της παρατήρησης:

$$f(q|d) \propto f(d|q)\, f(q) \propto f(x|\mu) \tag{6}$$

Η *πιθανοφάνεια (likelihood)* $\ell(q/d) = f(d/q)$ αντιστοιχεί στην πιθανότητα να εμφανιστούν οι τιμές δεδομένων *d* εάν μετράται ένα μέγεθος με αληθή τιμή *q*. Η πιθανοφάνεια αντιστοιχεί επομένως στην απόκριση του μηχανισμού παρατήρησης, συμπεριλαμβανομένου του εξοπλισμού, των ικανοτήτων του χειριστή και της μεθόδου μέτρησης. Πρόκειται με άλλα λόγια για την εικόνα που προκύπτει από τη μέτρηση, συνυπολογίζοντας τη διακρίβωση του εξοπλισμού.
Η σημασία της bayesian ανάλυσης έγκειται κυρίως στο διαφορετικό ερμηνευτικό σχήμα που εισάγει σε σχέση με την κλασική στατιστική, ωστόσο μπορεί να βρει σημαντικές πρακτικές εφαρμογές, όπως φαίνεται και από τα δύο παράδειγμα που ακολουθούν.

α) Βέλτιστη αξιοποίηση αποτελέσματος μεμονωμένης μέτρησης.

Έστω η υφιστάμενη γνώση για ένα μέγεθος *X* αντιπροσωπεύεται από μια κανονική κατανομή επικεντρωμένη στο x_1 με τυπική απόκλιση u_1:

$$f(x) = \frac{1}{\sqrt{2\pi}u_1} exp\left(-\frac{1}{2}\frac{(x - x_1)^2}{u_1^2} \right) \tag{7}$$

Έστω επίσης ότι είναι διαθέσιμη μια μέθοδος μέτρησης που οδηγεί σε μια τιμή x_2, τιμή που μπορεί να θεωρηθεί (με βάση τα χαρακτηριστικά της μεθόδου) ότι ανήκει σε κατανομή συχνοτήτων με κέντρο την αληθή τιμή και τυπική απόκλιση u_2. Η τυπική απόκλιση u_2 αντιστοιχεί στην αβεβαιότητα της μεθόδου. Η πιθανοφάνεια είναι τότε [5]:

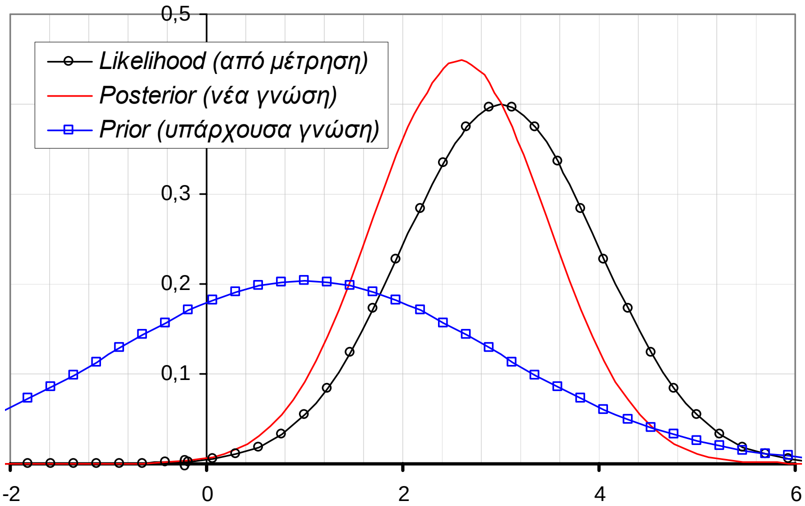
$$f(x_2, u_2 / x) = l(x / x_2, u_2) \propto exp\left(-\frac{1}{2}\frac{(x - x_2)^2}{u_2^2}\right) \tag{8}$$

Το θεώρημα του Bayes μας επιτρέπει να αναβαθμίσουμε τη γνώση μας για το μέγεθος *X*, σχηματίζοντας μια νέα εικόνα:

$$f(x / x_2, u_2) = \frac{1}{\sqrt{2\pi u_3}} exp\left(-\frac{1}{2}\frac{(x - x_3)^2}{u_3^2}\right) \tag{9}$$

Από το συνδυασμό των δύο κατανομών υπολογίζεται εύκολα ότι:

$$\frac{1}{u_3^2} = \frac{1}{u_1^2} + \frac{1}{u_2^2} \;, \quad \frac{x_3}{u_3^2} = \frac{x_1}{u_1^2} + \frac{x_3}{u_2^2} \tag{10}$$



Σχήμα 3: Αναβάθμιση γνώσης για το μετρούμενο μέγεθος

Από τις σχέσεις (9) και (10) προκύπτει ότι η τελική εικόνα για το μετρούμενο μέγεθος επηρεάζεται τόσο από τις υφιστάμενες γνώσεις, όσο και από την τρέχουσα παρατήρηση. Εάν όμως μια από τις δύο είναι πιο σίγουρη (μικρότερη αβεβαιότητα), τότε είναι αυτή που θα καθορίσει στο μεγαλύτερο βαθμό την νέα κατανομή, άρα και τη νέα εικόνα που σχηματίζεται για το μετρούμενο μέγεθος. Έτσι, μια μέτρηση με μια μέθοδο χαμηλών μετρολογικών αξιώσεων αναμένεται να μην επηρεάσει ουσιαστικά την ήδη διαμορφωμένη αντίληψη για το μετρούμενο μέγεθος και αντίστροφα.

β) Ερμηνεία αποτελεσμάτων από μεθόδους δεδομένης αξιοπιστίας

Το αποτέλεσμα μιας μέτρησης δεν είναι ποτέ εντελώς σίγουρο στο βαθμό που υπάρχει πάντα η πιθανότητα, έστω μικρή, η πραγματική τιμή να βρίσκεται έξω από το διάστημα εμπιστοσύνης. Ο τυπικός χρήστης του αποτελέσματος υποχρεώνεται συχνά να αξιολογήσει στην πράξη την πιθανότητα αστοχίας, συσχετίζοντάς την με τη γενικότερη εικόνα που έχει σχηματίσει για το μετρούμενο μέγεθος. Η bayesian ανάλυση επιτρέπει μια ρεαλιστική και συνάμα επιστημονικά έγκυρη συσχέτιση, όπως φαίνεται και από το παράδειγμα που ακολουθεί (παρμένο από την αναφορά [4]).

Έστω μια κλινική μέθοδος εξέτασης της μόλυνσης από τον ιό του AIDS η οποία έχει ποσοστό επιτυχίας, σύμφωνα με τα δεδομένα της επικύρωσης της, 100% όταν ο εξεταζόμενος είναι όντως φορέας. Η ίδια μέθοδος αστοχεί στα 0.2% των περιπτώσεων υγιών ασθενών, δείχνει δηλαδή μόλυνση όταν αυτή δεν υπάρχει. Ποια είναι η πιθανότητα ένα άτομο με θετικό αποτέλεσμα στην εξέταση αυτή να είναι όντως φορέας όταν ξέρουμε ότι μόνο το 0.15% του γενικού πληθυσμού είναι φορείς του ιού;
Τα δεδομένα του προβλήματος μπορούν να διατυπωθούν με όρους πιθανοτήτων με τις σχέσεις *p(θετικός/φορέας)=1*, *p(θετικός/όχι φορέας)=0.002* και *p(φορέας)=0.0015*. Το ζητούμενο είναι να εκτιμηθεί η πιθανότητα *p(φορέας/θετικός)*. Είναι εύκολο να διαπιστωθεί ότι η εφαρμογή του θεωρήματος του Bayes οδηγεί στη σχέση:

$$\frac{p(\varphi o\rho\acute{\varepsilon}\alpha\varsigma/\theta\varepsilon\tau\iota\kappa\acute{o}\varsigma)}{1\text{-}p(\varphi o\rho\acute{\varepsilon}\alpha\varsigma/\theta\varepsilon\tau\iota\kappa\acute{o}\varsigma)}=\frac{p(\varphi o\rho\acute{\varepsilon}\alpha\varsigma/\theta\varepsilon\tau\iota\kappa\acute{o}\varsigma)}{p(\acute{o}\chi\iota\ \varphi o\rho\acute{\varepsilon}\alpha\varsigma/\theta\varepsilon\tau\iota\kappa\acute{o}\varsigma)}=\frac{p(\theta\varepsilon\tau\iota\kappa\acute{o}\varsigma/\varphi o\rho\acute{\varepsilon}\alpha\varsigma)\ P(\varphi o\rho\acute{\varepsilon}\alpha\varsigma)}{p(\theta\varepsilon\tau\iota\kappa\acute{o}\varsigma/\acute{o}\chi\iota\ \varphi o\rho\acute{\varepsilon}\alpha\varsigma)p(\acute{o}\chi\iota\ \varphi o\rho\acute{\varepsilon}\alpha\varsigma)}=\frac{1*0{,}0015}{0{,}002*0{,}9985}=0.75$$

Από τη σχέση αυτή υπολογίζεται εύκολα ότι η ζητούμενη πιθανότητα είναι *p(φορέας/θετικός)=43%.* Είναι προφανές ότι η εικόνα αυτή θα ήταν διαφορετική εάν επρόκειτο για εξέταση μέλους μιας ευπαθούς ομάδας πληθυσμού, όπως για παράδειγμα χρήστες ναρκωτικών, για τους οποίους η πιθανότητα *p(φορέας)* θα ήταν ασφαλώς μεγαλύτερη, δηλαδή εάν η υπάρχουσα γνώση ήταν περισσότερο εστιασμένη. Γενικότερα αλώστε διαπιστώνεται μια διαρκώς εντεινόμενη τάση αξιοποίησης της bayesian στατιστικής για την αποτίμηση της αξιοπιστίας των κλινικών αποτελεσμάτων [9].

4. Συμπεράσματα

Η bayesian ανάλυση έχει επανέλθει τα τελευταία χρόνια στο επίκεντρο των συζητήσεων σχετικά με την αξιοπιστία και την ερμηνεία των αποτελεσμάτων των μετρήσεων. Όντας πιο κοντά στο φυσικό τρόπο επεξεργασίας της πληροφορίας, δίνει τη δυνατότητα για μια ρεαλιστική αναβάθμιση των πληροφοριών για το μετρούμενο μέγεθος στη βάση των εμπειρικών παρατηρήσεων. Η έμφαση στην αξιολόγηση των παρατηρούμενων αποτελεσμάτων υπό το φως των υφιστάμενων γενικότερων γνώσεων επαναφέρει στο επίκεντρο της προσοχής την πεποίθηση του μετρολόγου, αναβαθμίζοντας έτσι την σημασία της κατανόησης των, συχνά αόρατων, μηχανισμών που διαμορφώνουν το παρατηρούμενο αποτέλεσμα. Αναβαθμίζει με τον τρόπο αυτό την κριτική σκέψη, σε αντιπαράθεση με την αλόγιστη χρησιμοποίηση των κλασικών στατιστικών εργαλείων τα οποία οδηγούν συχνά σε παρανοήσεις και εσφαλμένες ερμηνείες.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] ISO, *Guide to the Expression of Uncertainty in Measurement*, Geneva, International Organization for Standardization, 1993.

[2] B. de Finetti, *Theory of probability*, J. Wiley & Sons, 1974

[3] C. Howson and P. Urbach, *Bayesian reasoning in science*, Nature, Vol. 350, 1991

[4] G. D'Agostini, *Bayesian reasoning in data analysis* - A Critical Introduction, World Scientific Publishing, 2003

[5] K. Weise, W. Woger, *A Bayesian theory of measurement uncertainty*, Meas. Science Tech., 1993, **4**, 1-11.

[6] P. Lee, *Bayesian Statistics: an Introduction*, London, Edward Arnold, 1989.

[7] R. Willink, I. Lira, *A united interpretation of different uncertainty intervals*, Measurement 38 (2005) 61–66

[8] Ε. Μαθιουλάκης, *Μέτρηση, ποιότητα μέτρησης και αβεβαιότητα*, HellasLab, 2004

[9] R. A. Matthews, *Methods for assessing the credibility of clinical trial outcomes*, Drug Information Journal, Vol. 35, pp. 1469–1478, 2001