ΕΚΤΙΜΗΣΗ ΜΕΤΡΟΥΜΕΝΟΥ ΜΕΓΕΘΟΥΣ ΑΠΟ ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ: BAYESIAN ΕΡΜΗΝΕΙΑ ΤΗΣ ΤΥΠΟΥ Α ΑΒΕΒΑΙΟΤΗΤΑΣ

Ε. Μαθιουλάκης και Β. Μπελεσιώτης

Εργαστήριο Ηλιακών & άλλων Ενεργειακών Συστημάτων – ΕΚΕΦΕ «ΔΗΜΟΚΡΙΤΟΣ»
15310 Αγία Παρασκευή Αττικής
E-mail: math@ipta.demokritos.gr

Περίληψη

Στην εργασία αυτή περιγράφεται μια μεθοδολογία χειρισμού επαναλαμβανόμενων παρατηρήσεων της μετρούμενης ποσότητας, ενταγμένη σε ένα Bayesian ερμηνευτικό πλαίσιο. Οι παρατηρούμενες τιμές εκλαμβάνονται ως μέτρο της πιθανοφάνειας (*likelihood*) και αξιοποιούνται για την αναβάθμιση του υφιστάμενου επιπέδου γνώσεων και την οικοδόμηση μιας νέας συνάρτησης πυκνότητας πιθανότητας, χρησιμοποιώντας τον νόμο του Bayes.

Τα αποτελέσματα συγκρίνονται με αυτά που προκύπτουν από την εφαρμογή της κλασικής Τύπου Α προσέγγισης, ειδικότερα στην περίπτωση ολιγάριθμων δειγμάτων όπου οι δύο μέθοδοι οδηγούν σε διαφορετικά αποτελέσματα.

Το βασικό πλεονέκτημα της προσέγγισης αυτής συνίσταται στην υιοθέτηση μιας ενιαίας Bayesian λογικής για όλες τις συνιστώσες αβεβαιότητας, είτε αυτές προέρχονται από επαναλαμβανόμενες μετρήσεις είτε από άλλες πηγές, επιτρέποντας την χρήση ενός συνεκτικού ερμηνευτικού πλαισίου βασισμένου στις κατανομές πιθανοτήτων, διευκολύνοντας παράλληλα την εξαγωγή συμπερασμάτων όσον αφορά τα διαστήματα εμπιστοσύνης και την πιθανότητα κάλυψης.

Λέξεις κλειδιά: *Επαναλαμβανόμενες παρατηρήσεις, Τύπου Α αβεβαιότητα, Bayesian στατιστική*

1. Εισαγωγή

Πολύ συχνά στην τρέχουσα εργαστηριακή πρακτική, οι διαθέσιμες πληροφορίες για μια ποσότητα Q προέρχονται από επαναλαμβανόμενες ανεξάρτητες παρατηρήσεις της ποσότητας αυτής, παίρνοντας βέβαια υπόψη και την ποιότητα του οργάνου παρατήρησης, δηλαδή της διάταξης μέτρησης. Η ανάγκη για επανάληψη της μέτρησης προκύπτει από το ότι η μετρούμενη τιμή μεταβάλλεται από τη μια παρατήρηση στην άλλη, χωρίς να είναι δυνατόν να διορθωθεί ή να περιοριστεί η μεταβλητότητα αυτή με τα δεδομένα μέσα που χρησιμοποιούνται. Στην πράξη, η τιμή του μετρούμενου μεγέθους εμφανίζεται να είναι "θολή" λόγω μεταβλητότητας, όχι αναγκαστικά γιατί είναι πράγματι θολή, αλλά πιθανά γιατί το όργανο παρατήρησης (μέτρησης) δεν οδηγεί σε σταθερότερες τιμές ή ακόμα γιατί οι συνθήκες (μέθοδος) παρατήρησης δεν επιτρέπουν καλύτερες μετρολογικές επιδόσεις. Σε τελική όμως ανάλυση, η μεταβλητότητα μπορεί να οφείλεται σε μια από τις παραπάνω αιτίες ή σε συνδυασμό τους, χωρίς να έχουν ιδιαίτερη πρακτική σημασία τα πραγματικά αίτια που την προκαλούν. Σημασία έχει η ποσοτική αποτύπωση, στις συγκεκριμένες συνθήκες, του πόσο "θολά" διακρίνεται η αληθής τιμή, δηλαδή της μεταβλητότητας του μεγέθους Q.

Η περίπτωση αυτή αντιμετωπίζεται από τον GUM [1] από τη σκοπιά της συχνοτικής στατιστικής, θεωρώντας δηλαδή ότι οι επιμέρους παρατηρήσεις διαφέρουν λόγω των τυχαίων μεταβολών των παραμέτρων επιρροής και ότι η στατιστική επεξεργασία των παρατηρήσεων μπορεί να οδηγήσει στη συνάρτηση κατανομής πιθανοτήτων (ΣΚΠ) της μεταβλητής, άρα και στην καλύτερή της εκτίμηση, καθώς και στην τυπική απόκλιση ή τυπική αβεβαιότητα που χαρακτηρίζει την εκτίμηση αυτή. Η προσέγγιση αυτή αποκαλείται, στα πλαίσια του GUM, Τύπου Α εκτίμηση της αβεβαιότητας.

Πιο συγκεκριμένα, για την Τύπου Α προσέγγιση, η τιμή του μετρούμενου μεγέθους 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. Η κατανομή αυτή χαρακτηρίζεται από μια εκτιμούμενη καλύτερη τιμή (συνήθως τη μέση τιμή) και μια τυπική απόκλιση. Στην πράξη ο μετρολόγος, αφού υποθέσει ένα είδος κατανομής, παράγει μια εκτίμηση της ΤΜΜ 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).

Η μεθοδολογική προσέγγιση του GUM δεν είναι μονομερώς προσανατολισμένη στην κλασική ή στην Bayesian ανάλυση, αλλά βασίζεται σε στοιχεία τόσο της μιας όσο και της άλλης. Η Τύπου Α αβεβαιότητα για παράδειγμα [1, section 4.2], ή ο υπολογισμός των βαθμών ελευθερίας [1, Annex G] βασίζεται στην κλασική στατιστική, ενώ ο ορισμός του Διαστήματος Εμπιστοσύνης καθώς και οι Τύπου Β αβεβαιότητες είναι προσανατολισμένες στην Bayesian συλλογιστική. Η συνύπαρξη δύο διαφορετικών οπτικών σε συνδυασμό με τις αδυναμίες της συχνοτικής θεώρησης και την αναθέρμανση της συζήτησης σχετικά με τις Bayesian πιθανότητες, οδήγησαν πολλούς συγγραφείς στην αναζήτηση τρόπων εναρμόνισης της Τύπου Α προσέγγισης με την λογική των πιθανοτήτων ως γνωσιακών μοντέλων [2]. Η εναρμόνιση αυτή είναι εξάλλου ένα από τα ζητούμενα της αναθεώρησης του GUM που βρίσκεται σε εξέλιξη [3]. Κεντρικό σημείο των προβληματισμών αυτών συνιστά η αξιοποίηση των επαναλαμβανόμενων παρατηρήσεων για την οικοδόμηση μοντέλων κατανομής πιθανοτήτων μέσω μιας Bayesian οπτικής, θέμα το οποίο συζητείται και από την παρούσα εργασία.

2. Η κλασική συχνοτική προσέγγιση - Οι Τύπου Α αβεβαιότητες

Όπως ήδη αναφέρθηκε, σκοπός είναι η εξαγωγή συμπερασμάτων σχετικά με την τιμή ενός μεγέθους, όταν οι διαθέσιμες πληροφορίες για το μέγεθος αυτό συνίσταται σε μια σειρά τιμές προερχόμενες από επαναλαμβανόμενες παρατηρήσεις του μεγέθους αυτού. Έστω λοιπόν ότι η διαθέσιμη πληροφορία για ένα μέγεθος Q συνίσταται σε μια σειρά από ανεξάρτητες παρατηρήσεις $\boldsymbol{q}=(q_1, q_2, \ldots, q_n)$ οι οποίες προέκυψαν κάτω από συνθήκες ελεγχόμενης επαναληψιμότητας.

Η συχνοτική προσέγγιση του GUM, βασισμένη στη θεωρία της δειγματοληψίας, αντιμετωπίζει τη ζητούμενη τιμή του μετρούμενου μεγέθους ως μια σταθερά, συχνά αποκαλούμενη αληθή τιμή. Τα αποτελέσματα των μετρήσεων θεωρούνται αντίθετα τυχαίες μεταβλητές χαρακτηριζόμενες από μια κατανομή πιθανοτήτων δειγματοληψίας (sampling probability distribution). Η κατανομή αυτή είναι στην πράξη μια συνάρτηση που αποτυπώνει τη σχετική συχνότητα εμφάνισης όλων των δυνατών αποτελεσμάτων ως αποτέλεσμα της επίδρασης τυχαίων παραμέτρων σε συνθήκες ελεγχόμενης επαναληψιμότητας, για αυτό και ορισμένοι συγγραφείς την αποκαλούν "κατανομή συχνοτήτων" (frequency distribution). Η καλύτερη εκτίμηση (expectation) μ_q και η μεταβλητότητα σ_q^2 της κατανομής είναι άγνωστες, μπορούν όμως να εκτιμηθούν από την αριθμητική μέση τιμή $\overline{q}$ και την πειραματική μεταβλητότητα s_q^2 των $q_1, q_2, ..., q_n$:

$$\overline{q} = \frac{1}{n}\sum_{i=1}^{n} q_i \tag{1}$$

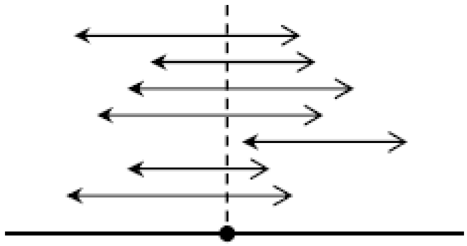
$$s_q^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(q_i - \overline{q}\right)^2 \tag{2}$$

Ως καλύτερη εκτίμηση της expectation μ_q του μετρούμενου μεγέθους Q λαμβάνεται λοιπόν το $\overline{q}$, με την μεταβλητότητα $\sigma_{\overline{q}}^2 = s_q^2 / n$ της μέσης τιμής, ως δείκτη της διασποράς των παρατηρούμενων τιμών, να θεωρείται μια καλή εκτίμηση της μεταβλητότητας της κατανομής συχνοτήτων και την τετραγωνική της ρίζα $u_{\overline{q}} = \sqrt{\sigma_{\overline{q}}^2}$ να ορίζεται ως η τυπική αβεβαιότητα που χαρακτηρίζει την καλύτερη εκτίμηση $\overline{q}$.

Απομένει να αξιολογηθεί πόσο καλά εκτιμάται το μ_q από το $\overline{q}$, δηλαδή πόση εμπιστοσύνη μπορεί να δείξει κάποιος στον εκτιμητή $\overline{q}$. Δεδομένου ότι το δείγμα $\mathbf{q}=(q_1, q_2, ..., q_n)$ που χρησιμοποιήθηκε είναι ένα από τα άπειρα δείγματα που θα μπορούσαν να έχουν ληφθεί, τα μεγέθη $\overline{q}$ και s_q θεωρούνται τυχαίες μεταβλητές. Εάν επιπλέον υποτεθεί ότι η κατανομή συχνοτήτων από την οποία έχουν ληφθεί τα $q_1, q_2, ..., q_n$ είναι κανονική, τότε η μεταβλητή $t = \frac{\sqrt{n}}{s_q}\left(\overline{q} - \mu_q\right)$ ακολουθεί μια κατανομή *t-student* με $v=n-1$ βαθμούς ελευθερίας.

Εάν λοιπόν οριστεί μια πιθανότητα κάλυψης p, η διευρυμένη αβεβαιότητα που χαρακτηρίζει το $\overline{q}$ υπολογίζεται από τη σχέση $U_{\overline{q}} = t_{pv} u_{\overline{q}}$, με το t_{pv} να είναι το αντίστοιχο ποσοστιαίο σημείο (quantile) της κατανομής *t-student*. Επιπρόσθετα, το διάστημα ($\overline{q} - U_{\overline{q}}$, $\overline{q} + U_{\overline{q}}$) εκλαμβάνεται ως ένα Διάστημα Εμπιστοσύνης για το μ_q *για* πιθανότητα κάλυψης p, υπό την έννοια του "πόσο συχνά το διάστημα αυτό θα περιέχει την άγνωστη expectation μ_q. Εδώ άλλωστε υπάρχει ένα πρόβλημα με τον GUM, από τη στιγμή που επισημαίνεται ότι η πιθανότητα p, αποκαλούμενη "confidence level", δεν ταυτίζεται με το "level of confidence" (το οποίο λαμβάνεται στα πλαίσια του GUM ως ισοδύναμο με το "coverage probability").

Στην πραγματικότητα μπορούμε, με βάση το Τύπου Α αποτέλεσμα, ότι υπάρχει $p\%$ πιθανότητα το τυχαία μεταβαλλόμενο διάστημα ($\overline{q} - U_{\overline{q}}, \overline{q} + U_{\overline{q}}$) να περιέχει την τιμή του μετρούμενου μεγέθους (η οποία θεωρείται σταθερή αλλά άγνωστη). Πρόκειται δηλαδή για μια πιθανότητα που χαρακτηρίζει τη διαδικασία εύρεσης του διαστήματος και όχι το διάστημα αυτό καθεαυτό. Από τη στιγμή που προσδιορίστηκε το διάστημα, δεν υπάρχει πλέον καμία τυχαιότητα στη όλη διαδικασία. Το διάστημα αυτό είναι δυνατόν να περιέχει ή να μην περιέχει την αληθή τιμή, αυτό όμως παραμένει άγνωστο [4, 5].



Σχήμα 1: Η συμβατική συχνοτική προσέγγιση Τύπου Α: στο p% των περιπτώσεων το τυχαία μεταβαλλόμενο Διάστημα Εμπιστοσύνης περιέχει την τιμή του μετρούμενου μεγέθους

3. Η Bayesian προσέγγιση

Η διαφορά ανάμεσα στην κλασική συχνοτική στατιστική και την Bayesian προσέγγιση θα μπορούσε να αποδοθεί σχηματικά από τη διαφορά ανάμεσα στις εκφράσεις "τυχαία" (random) και "αβέβαια" (uncertain) που χρησιμοποιούνται για να χαρακτηρίσουν 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. Στα πλαίσια της Bayesian ανάλυσης, τα αποτελέσματα E μιας εμπειρικής παρατήρησης μπορεί 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] αξιοποιώντας το γνωστό θεώρημα του Bayes το οποίο, στη γενική του μορφή, 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 [6, 7, 8]:

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

Στην Bayesian λογική, οι παρατηρήσεις $\boldsymbol{q}=(q_1, q_2, ..., q_n)$ μπορούν να οδηγήσουν σε διαφορετικές τιμές για την καλύτερη εκτίμηση και τη μεταβλητότητα του παρατηρούμενου μεγέθους, ανάλογα με το είδος της πληροφορίας που συνοδεύει το σετ των ανεξάρτητων παρατηρήσεων. Είναι σαφές ότι η πληροφορία αυτή από μόνη της δεν επιτρέπει την εφαρμογή του θεωρήματος του Bayes, από τη στιγμή που δεν είναι επαρκής για την διατύπωση της συνάρτησης πιθανοφάνειας, όταν δηλαδή δεν είναι διαθέσιμο ένα "μοντέλο" - ποιοτικής έστω - συσχέτισης παρατηρήσεων με πραγματικότητα. Το κενό αυτό δεν μπορεί να καλυφθεί παρά μόνο με τη αποδοχή λιγότερο ή περισσότερο εύλογων υποθέσεων, βασισμένων στην εμπειρία ή ακόμα τη διαίσθηση του μετρολόγου. Στις περισσότερες περιπτώσεις φαίνεται ως πλέον εύλογη η υπόθεση ότι οι ανεξάρτητες τιμές ανήκουν σε μια δεδομένη κατανομή πιθανοτήτων, συνήθως κανονική, ορθογώνια ή Poison. Από την κατανομή αυτή προκύπτει και η αντίστοιχη συνάρτηση πιθανοφάνειας.

Υπάρχει βέβαια η πιθανότητα οι επιμέρους τιμές να προέρχονται από παρατηρήσεις που έγιναν κάτω από ανόμοιες συνθήκες, χωρίς δηλαδή να διασφαλίζεται η συνθήκη της επαναληψιμότητας. Τέτοια είναι η περίπτωση διεργαστηριακών μετρήσεων, όπου από κάθε εργαστήριο προέκυψε μια τιμή q_i, συνοδευόμενη από μια τυπική αβεβαιότητα u_i, ή ακόμα ένα διάστημα $(q_i\text{-}e_i,\ q_i\text{+}e_i)$ στο οποίο περιέχεται η τιμή q_i.

Επανερχόμενοι στην περίπτωση των ανεξάρτητων παρατηρήσεων σε συνθήκες επαναληψιμότητας, με την τυχαιότητα των παρατηρούμενων τιμών να παραπέμπει σε μια κανονική κατανομή. Η υπόθεση της κανονικότητας είναι βάσιμη στις περισσότερες περιπτώσεις, τουλάχιστον όταν υπάρχει λόγος να πιστεύει κάποιος ότι η εμφανιζόμενη μεταβλητότητα οφείλεται σε πολλές τυχαίες, ανεξάρτητες μεταξύ τους, επιδράσεις.

Γίνεται επομένως η υπόθεση ότι οι παρατηρούμενες τιμές ανήκουν σε ένα απείρου μεγέθους πληθυσμό τιμών, η κατανομή πιθανοτήτων του οποίου είναι κανονική, με κέντρο την άγνωστη τιμή του Q και μεταβλητότητα U. Σημειώνεται ότι αν και επιδιώκεται η εξαγωγή συμπερασμάτων για τις δύο μεταβλητές Q και U, μόνο η πρώτη έχει φυσική υπόσταση.

Για ένα ζεύγος τιμών q και u, η συνάρτησης πιθανοφάνειας $l(q,u|q_i)$ είναι η κατανομής πιθανοτήτων για κάθε δεδομένη τιμή q_i και γράφεται ως ακολούθως:

$$f(q_i \mid q,u) \propto \frac{1}{u^{0.5}} exp\left[-\frac{1}{2}\frac{(q_i - q)^2}{u} \right] \tag{4}$$

Η πιθανοφάνεια $l(q,u|\boldsymbol{q})$ είναι επομένως ίση με το γινόμενο των επιμέρους κατανομών πιθανοτήτων:

$$l(q,u \mid \boldsymbol{q}) \propto \frac{1}{u^{n/2}} exp\left[-\frac{1}{2u}\sum_n (q_i - q)^2 \right] \tag{5}$$

Η κοινή prior πιθανότητα των Q και U, λόγω της μεταξύ τους ανεξαρτησίας, είναι ίση με το γινόμενο των αντίστοιχων prior τα οποία, στο βαθμό που δεν υφίσταται καμία προηγούμενη πληροφορία είναι non-informative prior. Για το Q το οποίο εδώ παρουσιάζεται σαν μια αναμενόμενη τιμή ή ακόμα ως παράμετρος θέσης (location parameter), η prior κατανομή του είναι μια σταθερά (Bayes' prior). Αντίστοιχα για τη μεταβλητότητα η οποία παρουσιάζεται ως παράμετρος κλίμακας (scale parameter), η prior κατανομή του U είναι του τύπου Jeffreys prior, δηλαδή ανάλογη του $1/u$ [9, 10]. Επομένως, η κοινή posterior συνάρτηση κατανομής πιθανοτήτων γίνεται:

$$f(q,u \mid \boldsymbol{q}) \propto \frac{1}{u^{1+n/2}} exp\left[-\frac{1}{2u}\sum_n (q_i - q)^2 \right] \tag{6}$$

και παίρνοντας υπόψη ότι $\sum_n (q_i - q)^2 = \sum_n (q_i - \overline{q})^2 + n(q - \overline{q})^2$:

$$f(q,u \mid \boldsymbol{q}) \propto \frac{1}{u^{1+n/2}} exp\left[-\frac{(n-1)s^2 + n(q-\overline{q})^2}{2u} \right] \tag{7}$$

όπου $s^2 = \frac{1}{n-1} \sum_n (q_i - \overline{q})^2$

Πρέπει να σημειωθεί ότι, στην περίπτωση αυτή, τα μέγεθος *s* είναι απλά ένα σύμβολο και δεν αντιπροσωπεύει μια στατιστική εκτίμηση της μεταβλητότητας.

Σύμφωνα με τη θεωρία των πιθανοτήτων, από την ολοκλήρωση της σχέσης (7) ως προς *u* προκύπτει η οριακή ΣΚΠ του *Q* (marginal probability density function), από την οποία είναι δυνατόν να προσδιοριστεί η καλύτερη εκτίμηση και η μεταβλητότητά του:

$$f(q \mid \boldsymbol{q}) \propto \left[-\frac{1}{2u} \sum_n (q_i - q)^2 \right]^{-n/2} \tag{8}$$

Εάν γίνει ο μετασχηματισμός $Z = n^{0.5} s^{-1} (Q - \overline{q})$ και *m=n-1*, η κατανομή του *Q* μετασχηματίζεται σε κατανομή *f(z)* του *z* και γράφεται:

$$f(z) \propto \left(1 + \frac{z^2}{m} \right)^{-\frac{m+1}{2}} \tag{9}$$

Η κατανομή *f(z)* της σχέσης (9) έχει, όσον αφορά το *z*, τη μορφή μιας κατανομή *t-Student* με παράμετρο (βαθμούς ελευθέριας) *m*. Όπως είναι γνωστό, η καλύτερη εκτίμηση μιας κατανομής *Student* είναι *E(Z)=0*, δηλαδή ίση με το μηδέν και η μεταβλητότητα που χαρακτηρίζει την εκτίμηση αυτή είναι *V(Z)=m/(m-2)*. Κατά συνέπεια, εάν οριστεί ότι $r = \sqrt{(n-1)/(n-3)}$:

$$E(Z \mid \boldsymbol{q}) = 0 \tag{10}$$

$$V(Z \mid \boldsymbol{q}) = \frac{m}{m-2} = \frac{n-1}{n-3} = r^2 \tag{11}$$

Επιχειρώντας αναστροφή του μετασχηματισμού από Z σε Q, συνάγεται εύκολα ότι:

$$E(Q \mid \boldsymbol{q}) = \bar{q} \tag{12}$$

$$V(Q \mid \boldsymbol{q}) = \frac{n-1}{n-3}\frac{s^2}{n} = r^2\,\frac{s^2}{n} \tag{13}$$

Η ολοκλήρωση της σχέσης (7) ως προς q οδηγεί στην οριακή ΣΚΠ του U, από την οποία επίσης είναι δυνατόν να προσδιοριστεί η καλύτερη εκτίμηση και η μεταβλητότητά του:

$$f(u \mid \boldsymbol{q}) \propto \frac{1}{u^{\frac{n+1}{2}}}\,exp\left[-\frac{(n-1)s^2}{2u}\right] \tag{14}$$

Από την παραπάνω ΣΚΠ, προκύπτουν, όσον αφορά την καλύτερη εκτίμηση και τη μεταβλητότητα του U, οι ακόλουθες σχέσεις:

$$E(U \mid \boldsymbol{q}) = (rs)^2, \quad V(U \mid \boldsymbol{q}) = \frac{2(rs)^4}{n-5}$$

(13)

$$E(U \mid \boldsymbol{q}) = (rs)^2, \quad V(U \mid \boldsymbol{q}) = \frac{2(rs)^4}{n-5} \tag{15}$$

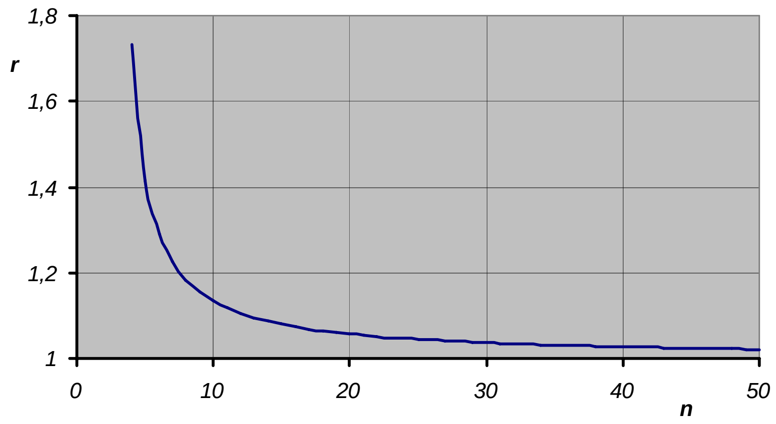
Η πληροφορία που σχετίζεται με τη μεταβλητότητα του U δεν παρουσιάζει ιδιαίτερο ενδιαφέρον στη συγκεκριμένη περίπτωση. Επισημαίνεται όμως ότι η πληροφορία αυτή αντιπροσωπεύει την αβεβαιότητας της αβεβαιότητας και δεν είναι δυνατόν να υπάρξει για πληθυσμό μετρήσεων μικρότερο του 6.

Επανερχόμενοι στο μέγεθος που κυρίως ενδιαφέρει, δηλαδή στην καλύτερη εκτίμηση του Q, η τυπική αβεβαιότητα που χαρακτηρίζει την εκτίμηση αυτή είναι:

$$u_Q = \sqrt{\frac{\sum_n (q_i - \bar{q})^2}{n(n-3)}} = r\,u_{\bar{q}} \tag{14}$$

όπου $u_{\bar{q}}$ είναι η τυπική αβεβαιότητα που προκύπτει από την Τύπου Α συχνοτική προσέγγιση που εισάγεται από τον GUM.

Είναι προφανές ότι η σχέση αυτή δεν μπορεί να χρησιμοποιηθεί παρά μόνο για $n>3$. Είναι επίσης φανερό ότι η Bayesian προσέγγιση οδηγεί σε μεγαλύτερες τυπικές αβεβαιότητες συγκρινόμενες με αυτές που προκύπτουν από την κλασική συχνοτική προσέγγιση, με την διαφορά τους να εξασθενεί όταν το μέγεθος n του δείγματος γίνεται μεγάλο (σχήμα 1).



Σχήμα 2: Λόγος Bayesian προς Τύπου Α αβεβαιότητα συναρτήσει του μεγέθους του δείγματος

Η διευρυμένη αβεβαιότητα, για μια πιθανότητα κάλυψης p, γίνεται με τη σειρά της:

$$U_{Q,p} = u_Q \frac{t_{p,n-1}}{r} \tag{8}$$

όπου $t_{p,n-1}$ είναι η ΣΚΠ Student με n-1 βαθμούς ελευθερίας.

Η ερμηνεία του διαστήματος ($q-U_{Q,p}, q-U_{Q,p}$) είναι στην περίπτωση αυτή διαφορετική από την ερμηνεία που δόθηκε στην Τύπου Α διευρυμένη αβεβαιότητας. Η Bayesian προσέγγιση οδήγησε σε μια ΣΚΠ για τις πιθανές τιμές του μετρούμενου μεγέθους και μπορούμε αν πούμε ότι "η πιθανότητα να είναι η αληθής τιμή στο διάστημα ($q-U_{Q,p}, q-U_{Q,p}$) είναι p. Το διάστημα αυτό αποκαλείται συνήθως "διάστημα αξιοπιστίας" (credibility interval) η πιθανότητα p "επίπεδο αξιοπιστίας" (credibility level).

Μια άλλη διαφορά ανάμεσα στην Τύπου Α και την Bayesian προσέγγιση είναι το ότι στην περίπτωση της δεύτερης δεν απαιτείται καμία ειδική αναφορά στο πλήθος του δείγματος, αντίθετα με την πρώτη όπου το πλήθος καθορίζει τους βαθμούς ελευθερίας. Με την Bayesian ανάλυση δεν γίνεται εξάλλου καμία αναφορά σε βαθμούς ελευθερίας και δεν έχει εφαρμογή η σχέση Welch–Satterthwaite, βάσει της οποίας υπολογίζονται οι ενεργοί βαθμού ελευθερίας που χαρακτηρίζουν τη συνδυασμένη αβεβαιότητα [8].

Αξίζει τέλος να σημειωθεί ότι, παρά τις διαφορές τους όσον αφορά στην εκτίμηση της αβεβαιότητας, οι δύο προσεγγίσεις καταλήγουν στην ίδια διευρυμένη αβεβαιότητας. Αυτό εξηγείται από το ότι η Τύπου Α προσέγγιση εισάγει, για τα μικρού πληθυσμού δείγματα, μια «αβεβαιότητα της αβεβαιότητας» (μικρός αριθμός βαθμών ελευθερίας) η οποία καταλήγει ουσιαστικά, στο επόμενο στάδιο υπολογισμού του Διαστήματος Εμπιστοσύνης, στην διεύρυνση της «αβέβαιης» τυπικής αβεβαιότητας.

4. Συμπεράσματα

Η κύρια διαφορά μεταξύ συχνοτικής Τύπου Α προσέγγισης και Bayesian ανάλυσης εστιάζεται στο ερμηνευτικό πλαίσιο που χρησιμοποιείται από την μια ή από την άλλη. Η Τύπου Α ανάλυση εμφανίζεται ως πιο πρακτική στην εφαρμογή αλλά υστερεί στην ερμηνεία του τελικού αποτελέσματος. Αντίθετα η Bayesian ανάλυση εισάγει ορισμένα στοιχεία πολυπλοκότητας, παρέχοντας όμως ένα συνεκτικό ερμηνευτικό πλαίσιο, κοινό με αυτό των Τύπου Β αβεβαιοτήτων.

Στα πλαίσια της συχνοτικής προσέγγισης, το Διάστημα Εμπιστοσύνης αναφέρεται περισσότερο στην αξιοπιστία της στατιστικής επεξεργασίας και λιγότερο στην πιθανότητα η αληθής τιμή να περιέχεται στο συγκεκριμένο διάστημα. Με την Bayesian ανάλυση αντίθετα, ένα *p%* Διάστημα Εμπιστοσύνης, αποκαλούμενο επίσης Διάστημα Αξιοπιστίας, σημαίνει ακριβώς αυτό που φαίνεται να σημαίνει, δηλαδή ότι η πιθανότητα να περιέχεται η αληθής τιμή στο συγκεκριμένο διάστημα είναι *p%*.

Απούσης οποιασδήποτε άλλης προηγούμενης πληροφορίας σχετικά με το μετρούμενο μέγεθος, οι δύο προσεγγίσεις καταλήγουν στο ίδιο Διάστημα Εμπιστοσύνης (ή αξιοπιστίας). Μόνο όμως η Bayesian ανάλυση εμφανίζεται ικανή να ενσωματώσει τυχόν προϋπάρχουσες πληροφορίες και να παράγει μια επικαιροποιημένη (post-data) εκτίμηση.

Αν και η Bayesian τυπική αβεβαιότητα είναι μεγαλύτερη από την αντίστοιχη Τύπου Α, ειδικότερα για δείγματα μικρού πληθυσμού, οι δύο τιμές συγκλίνουν με την αύξηση του εν λόγω πληθυσμού.

ΑΝΑΦΟΡΕΣ

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

[2] Kacker R and Jones A, On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent Metrologia 40 (2003), 235–48.

[3] Bich W., Cox M. and Harris P., Evolution of the ‘Guide to the Expression of Uncertainty in Measurement', Metrologia 43 (2006) S161–S166.

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

[5] ISO 3534-1, Statistics - Vocabulary and symbols - Part 1: Probability and general statistical terms, ISO ed., Geneva, 1993

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

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

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

[9] Box G. and Tiao G., Bayesian Inference in Statistical Analysis, wiley, New York, 1992.

[10] Lira I., Evaluating the measurement uncertainty, IoP ed., Bristol, 2002.