ΤΕΤΡΑΔΙΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ, ΤΕΥΧΟΣ 16 (*σσ. 14-22*)
DATA ANALYSIS BULLETIN, ISSUE 16 (*pp. 14-22)*

# Συνεχείς συναρτήσεις και συντελεστής μεταβλητότητας: Προσέγγιση μέσα από δειγματοληψία

**Νικόλαος Φαρμάκης**

*Τμήμα Μαθηματικών*
*Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης*

*Περίληψη*

Ο Συντελεστής Μεταβλητότητας είναι μία πολύ χρήσιμη έννοια για διάφορες εφαρμογές στη Στατιστική. Έχει ήδη χρησιμοποιηθεί στη διόρθωση μεροληψίας δεδομένων, στην πολυωνυμική απόδοση της συνάρτησης πυκνότητας πιθανότητας (σ.π.π.) *f(x)* μιας τυχαίας μεταβλητής *X*, Farmakis (2003,2010), και σε πολλές άλλες περιπτώσεις ερευνών. Εδώ επιχειρείται μία προσέγγιση πολυωνυμικής μορφής διαφόρων ειδών συνεχών συναρτήσεων *f(x)* (όχι κατ' ανάγκη σ.π.π.) μέσα από ένα συστηματικό δείγμα τιμών της εκάστοτε συνάρτησης. Ακολουθείται στην συνέχεια μία διαδικασία ανάλογη με την, ήδη γνωστή, διαδικασία που ακολουθήθηκε στην περίπτωση που η συνάρτηση *f(x)* ήταν μία σ.π.π., Farmakis (2003). Η νέα ιδέα στην παρούσα φάση είναι η επέκταση της χρήσης του Συντελεστή Μεταβλητότητας και σε περιπτώσεις όπου η υπό μελέτη συνάρτηση *f(x)* είναι απλά κάποια συνεχής συνάρτηση *f(x)* ορισμένη στο διάστημα [*α, β*] της ευθείας ***R***, των Πραγματικών Αριθμών. Γίνεται η μελέτη πάνω στις συμμετρικές συναρτήσεις *f(x)* και με παραδείγματα. Η εφαρμογή λειτουργεί ως αλγόριθμος με χρήση πληροφορικών προγραμμάτων (EXCEL) για την επιτυχή και γρήγορη μετάβαση από τα δειγματικά δεδομένα στις παραμέτρους της συνάρτησης *f(x)* και τελικά στην ίδια τη συνάρτηση *f(x)*.

## *1. Εισαγωγή – Θεωρητική προσέγγιση*

Θεωρούμε μία συνάρτηση *f(x)* συνεχή και ορισμένη στο διάστημα [*α, β*] της ευθείας ***R***, των Πραγματικών Αριθμών. Η συνάρτηση μπορεί να είναι συμμετρική ή όχι. Εδώ θα μελετήσουμε την περίπτωση των συμμετρικών συναρτήσεων διεξοδικά: θεωρητικά και με παραδείγματα. Όταν η συνάρτηση *f(x)* δεν είναι συμμετρική χρειάζονται ειδικοί χειρισμοί που δεν είναι μέσα στους στόχους της εργασίας αυτής.

Ειδικότερα θα παρουσιαστεί μία μέθοδος προσέγγισης της συνάρτησης *f(x)* μέσα από δεδομένα δείγματος και με τη βοήθεια του Συντελεστή Μεταβλητότητας (ΣΜ).

Θα χρησιμοποιήσουμε ένα δείγμα τιμών $y_0, y_1, y_2, \ldots, y_n$ της *f(x)* που αντιστοιχούν στις θέσεις $a = x_0, x_1, x_2, \ldots, x_n = \beta \in [\alpha,\ \beta]$ που ισαπέχουν μεταξύ τους (συστηματικό δείγμα). Άρα είναι $x_i = a + i \cdot w,\ i = 0,1,2,\ldots,n,\ w = (\beta - a)/n$ και παρατηρούμε ότι το όλο πεδίο ορισμού έχει διαμελιστεί σε *n* ισομήκη διαστήματα $\delta_i = [x_{i-1}, x_i],$ με μήκος *w* έκαστο και με μέσο το σημείο $a + \left(i - \frac{1}{2}\right) \cdot w = \kappa_i,\ i = 1,2,3,\ldots,n$ . Στο καθένα διάστημα $\delta_i$ αντιστοιχεί το ολοκλήρωμα $I_i = \int_{x_{i-1}}^{x_i} f(x)dx$ της συνάρτηση *f(x),* που δεν ξέρουμε την τιμή του ακριβώς αλλά μπορεί να προσεγγιστεί από την ποσότητα $f_i = w \cdot \frac{y_{i-1} + y_i}{2}$ . Θεωρούμε τώρα τρία *n*-διάστατα διανύσματα: $\boldsymbol{k}=(\kappa_1, \kappa_2, \ldots,\kappa_n)$, $\boldsymbol{\varphi}=(f_1, f_2, \ldots,f_n)$ και $\mathbf{1}=(1,1,\ldots,1)$. Αν στιγμιαία θεωρήσουμε ότι η συνάρτηση *f(x)* παριστάνει την απόλυτη συχνότητα των τιμών του ***X***, τότε η μέση τιμή του *X* στο [*α*, *β*] προσεγγίζεται από $\hat{\mu} = \bar{x} = \frac{1}{\boldsymbol{\varphi} \cdot \mathbf{1}'} \sum_{i=1}^{n} \kappa_i \cdot f_i = \frac{1}{\boldsymbol{\varphi} \cdot \mathbf{1}'} \cdot <\boldsymbol{k},\boldsymbol{\varphi}>$ , και $<\boldsymbol{k}\ ,\ \boldsymbol{\varphi}>$ είναι το εσωτερικό γινόμενο των δύο συμβαλλομένων διανυσμάτων. Η τιμή της μέσης τιμής συμπίπτει θεωρητικά με τη μέση τιμή των άκρων *α* και *β* του διαστήματος λόγω και της συμμετρίας της συνάρτησης *f(x).* Στην πράξη όμως είναι μία πολύ κοντινή τιμή της μέσης τιμής τους. Με ανάλογο τρόπο και κάνοντας χρήση γνωστών μεθόδων (Φαρμάκης, 2001) βρίσκουμε τη διασπορά $s^2$ και την τυπική απόκλιση *s* του *X* και με βάση το προαναφερόμενο δείγμα. Από τα παραπάνω υπολογίζουμε τον συντελεστή μεταβλητότητας (ΣΜ) από τη σχέση

$$Cv = \frac{\sigma}{\mu} \quad \text{με εκτιμήτρια συνάρτηση} \quad \hat{C}v = \frac{\hat{\sigma}}{\hat{\mu}} = \frac{s}{\bar{x}}. \tag{1.1}$$

Θεωρώντας συμμετρική τη συνάρτηση *f(x),* της οποίας έχουμε γνωστές τις *n* τιμές των ολοκληρωμάτων της (ανά διάστημα) που προαναφέρθηκαν, μπορούμε να παραστήσουμε με την παρακάτω πολυωνυμική μορφή δύο κλάδων μία άλλη, επίσης συμμετρική συνάρτηση που θα χρησιμοποιηθεί βοηθητικά:

$$g(x) = \begin{cases} h \cdot (x - a)^{\nu}, & x \in [a, \frac{\beta+\alpha}{2}) \\ h \cdot (\beta - x)^{\nu}, & x \in [\frac{\beta+\alpha}{2}, \beta) \\ 0\ , & x \notin [\alpha, \beta] \end{cases} \tag{1.2}$$

Αν θεωρήσουμε ότι α=0 τότε έχουμε την απλούστερη μορφή της:

$$g(x) = \begin{cases} h \cdot x^{\nu} & x \in [0, \frac{\beta}{2}) \\ h \cdot (\beta - x)^{\nu} & x \in [\frac{\beta}{2}, \beta] \\ 0 & x \notin [0, \beta] \end{cases}. \tag{1.3}$$

Αυτό που συνδέει τις δύο συναρτήσεις *f(x)*και *g(x)* είναι η συμμετρική τους ιδιότητα. Θα προσδιορίσουμε την *g(x)* και μετά την *f(x)* έτσι ώστε η δεύτερη να προκύπτει από την πρώτη με προσέγγιση ενός αριθμητικού συντελεστή. Το ολοκλήρωμα της *g(x)* στο [*α*, *β*] είναι ίσο με μονάδα, Farmakis (2003).

Η *g(x)* είναι εύκολο να προσδιοριστεί από τα δεδομένα που αναφέρθηκαν παραπάνω, στην αρχή της παραγράφου. Τα δεδομένα είναι τα *n*+1 ζεύγη $\left(x_i, y_i\right), i = 0,1,2,...,n$ και τα τρία διανύσματα *n*-διαστάσεων, τα ***k***, ***φ*** και **1**. Ο προσδιορισμός της *g(x)* γίνεται με βάση μέθοδο προσδιορισμού σ.π.π. τυχαίας μεταβλητής *X*, από δείγμα τιμών της *X*, με τη μορφή πολυωνυμικής εκτιμήτριας, Farmakis (2003). Η μέθοδος αξιοποιεί ιδιότητες του ΣΜ της εκάστοτε τυχαίας μεταβλητής (τμ) *X*. Δίνουμε περιγραφή της μεθόδου σε αδρές γραμμές:

$1^o$) Από δείγμα της τμ *X*, μεγέθους *M*>45 βρίσκουμε τη μέση τιμή, τη διασπορά, την τυπική απόκλιση και τελικά τον ΣΜ. Επειδή η τμ (εδώ) είναι συνεχής ομαδοποιούμε τα δεδομένα σε μικρό αριθμό τάξεων και έχουμε τις συντεταγμένες του διανύσματος ***k***, τα μέσα των τάξεων. Η συχνότητα της κάθε τιμής(=μέσο τάξης) είναι η αντίστοιχη συντεταγμένη τιμή του ***φ***. Το μέγεθος του δείγματος είναι το *M*= ***φ*.1′**, όπου το **1′** είναι το διάνυσμα-στήλη με στοιχεία μονάδες.

$2^o$) Έχει αποδειχτεί ότι (Farmakis, 2003) οι παράμετροι της *g(x)* των σχέσεων (1.2), (1.3) δίνονται από:

$$\nu = \frac{-5+\sqrt{1+8\cdot\lambda}}{2} \quad \text{και} \quad h = \frac{2^{\nu}\cdot(\nu+1)}{(\beta-\alpha)^{\nu+1}} \text{ ή } h = \frac{2^{\nu}\cdot(\nu+1)}{\beta^{\nu+1}} \qquad (1.4)$$

όπου είναι $\lambda = Cv^{-2}$.

$3^o$) Σύμφωνα με όσα περιλαμβάνονται στην εργασία Farmakis (2003) η συνάρτηση *g(x)* έχει ολοκλήρωμα ίσο με τη μονάδα επειδή είναι σ.π.π. της τμ *X*. Η συνάρτηση όμως *f(x)* έχει ολοκλήρωμα διάφορο της μονάδας και που προσεγγίζεται δειγματοληπτικά από:

$$\int_a^{\beta} f(x)dx \cong \hat{I} = \boldsymbol{\varphi}'\cdot\mathbf{1} = M\ , \quad a \ge 0 \qquad (1.5)$$

Η σχέση άρα *f(x)* = *M* · *g(x)* δίνει την εκτίμηση της *f(x)* από τα δεδομένα του δείγματος (συστηματικό) τιμών της.

<u>*Σημείωση 1η*</u>: Το αρχικό συστηματικό δείγμα έχει *n*+1 ζεύγη τιμών *(x, f(x))*. Από αυτά προκύπτουν τα *n* ζεύγη $(\kappa_i, f_i)$, i=1,2,…,*n*. Το λειτουργικό όμως (για την μέθοδο αυτή) δείγμα έχει $\boldsymbol{\varphi}'\cdot\mathbf{1} = M$ στοιχεία-τιμές. Αυτό είναι το μέγεθος του δείγματος από το οποίο εξαρτάται και η ακρίβεια της προσέγγισης. Το μέγεθος *M* προκύπτει από τη διαδικασία προσδιορισμού της μέσης τιμής και της διασποράς και εξαρτάται και από τις τιμές της προς προσδιορισμό συνάρτησης *f(x)* και από το *n* μέσω του εύρους *w*. Είναι γνωστό από τη βιβλιογραφία ότι αυτά τα δύο μεγέθη συνδέονται με μια προσεγγιστική σχέση της μορφής $n \cong q\ln M,\ q \in (1.3, 1.4)$, Φαρμάκης (2001). Άρα μία τιμή του *M*>45, απαιτεί το *n* να είναι τουλάχιστο 5.

*Σημείωση 2η:* Στις περιπτώσεις που είναι *α*>0 αντί της τμ *X* παίρνουμε την *X-α*. Άρα μας ενδιαφέρει πάντα η σχέση (1.3) και αυτό αρκεί για την *g(x)*. Γενικότερα μπορούμε για όποιες τιμές άκρων του πεδίου ορισμού [*ρ*, *ξ*] να έχουμε το μετασχηματισμό [0, *ξ-ρ*], οπότε η μέθοδος επεκτείνεται για όλες τις συμμετρικές συναρτήσεις ανεξάρτητα από το αν έχουμε και αρνητικές τιμές της *X*. Υιοθετείται δηλαδή το [0, *β*]= [0, *ξ-ρ*] γενικά.

*Σημείωση 3η:* Στην περίπτωση που η *f(x)* είναι σ.π.π. μιας τμ *X*, τότε η μέθοδος προσδιορισμού της είναι αυτή που περιγράφεται στην εργασία Farmakis (2003).

## 2. Παραδείγματα – Πρακτική προσέγγιση

Θα δώσουμε παραδείγματα που θα υλοποιούν την παραπάνω μέθοδο και θα φωτίζουν το θέμα με πρακτικές εφαρμογές.

*Παράδειγμα 1ο:* Δίνεται δείγμα τιμών της συνάρτησης *f(x)/*[*α*, *β*] που είναι συμμετρική και συνεχής. Τα δεδομένα του δείγματος που είναι συστηματικό ως προς x είναι στον επόμενο Πίνακα 1:

Πίνακας 1: Συστηματικό δείγμα ως προς x για τη συνάρτηση *y=f(x)*

| **i** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8=*n* |
|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 5 | 9 | 13 | 17 | 21 | 25 | 29 | 33 | 37 |
| $y_i$ | 18 | 23 | 28 | 32 | 36 | 32 | 28 | 22 | 17 |

Να βρεθεί μία προσέγγισή της συνάρτησης πολυωνυμικής μορφής.

*Λύση:* Από τη γραμμή των *x* διαπιστώνουμε ότι είναι *α*=5 και *β*=37. Άρα *w*=4=(*β-α*)/*n*.
Μετασχηματίζουμε σε *X*-5 το *X* και παίρνουμε τον παρακάτω Πίνακα 2, επεξεργασίας ομαδοποιημένων στοιχείων, με 8 κλάσεις στοιχείων, Φαρμάκης (2001):

Πίνακας 2: Επεξεργασία ταξινομημένων στοιχείων σε 8 τάξεις εύρους *w*=4

| Τάξεις | $\kappa_i$ | $f_i = w\cdot(y_{i-1}+y_i)/2$ | $\kappa'=(\kappa_i-m)/w$ | $f_i\kappa'$ | $f_i\kappa'^2$ | $f_i'$*(θεωρητικά)* |
|---|---|---|---|---|---|---|
| [0, 4) | 2 | 82 | -3 | -246 | 738 | *79.29* |
| [4, 8) | 6 | 102 | -2 | -204 | 408 | *106.86* |
| [8, 12) | 10 | 120 | -1 | -120 | 120 | *120.51* |
| [12, 16) | ***14*** | 136 | 0 | 0 | 0 | *130.34* |
| [16, 20) | 18 | 136 | 1 | 136 | 136 | *130.34* |
| [20, 24) | 22 | 120 | 2 | 240 | 480 | *120.51* |
| [24, 28) | 26 | 100 | 3 | 300 | 900 | *106.51* |
| [28, 32] | 30 | 78 | 4 | 312 | 1248 | *79.29* |
| Σύνολα | | *M*=874 | | $T_1$=418 | $T_2$=4030 | *M=874.00* |

Πήραμε προσωρινή μέση τιμή την $m$=14. Από τα αποτελέσματα του Πίνακα 2 έχουμε:

Μέση τιμή: $\overline{x} = 14 + \dfrac{4 \cdot 418}{874} = 15.91$

Διασπορά: $s^2 = \dfrac{4^2}{873} \cdot \left( 4030 - \dfrac{418^2}{874} \right) = 70.1963$ και άρα

Τυπική απόκλιση: $s = \sqrt{70.1963} = 8.38$

Συντελεστής μεταβλητότητας: $Cv = \dfrac{8.38}{15.91} = 0.5267$ και $\lambda = Cv^{-2} = 3.6047$

Άρα εκθέτης $v = \dfrac{-5 + \sqrt{1 + 8 \cdot 3.6047}}{2} = 0.2312$ και συντελεστής $h = 0.0202$.

Από τα ανωτέρω και τον τύπο (1.3) έχουμε

$$g(x) = \begin{cases} 0.020267 \cdot x^{0.2312} & x \in [0,16) \\ 0.020267 \cdot (32 - x)^{0.2312} & x \in [16,32]. \\ 0 & x \notin [0,32] \end{cases} \qquad (2.1)$$

και τελικά

$$f(x) = \begin{cases} 17.713358 \cdot x^{0.2312} & x \in [0,16) \\ 17.713358 \cdot (32 - x)^{0.2312} & x \in [16,32]. \\ 0 & x \notin [0,32] \end{cases} \qquad (2.2)$$

είναι η εκτιμήτρια της συνάρτησης *f(x)* από την οποία είχαμε τα αρχικά δεδομένα σε συστηματική παρατήρηση με βήμα $w$=4. Οίκοθεν νοείται ότι για κάθε x (π.χ. x=13) που μας δίνεται παίρνουμε το αντίστοιχο x-5 και το θέτουμε στην (2.2). Έτσι για x=13 θα μπει η τιμή 13-5=8 και η τιμή της συνάρτησης θα είναι: $f(8)$=28.65.

Στη συνέχεια επιχειρούμε να ελέγξουμε αν η συνάρτηση (2.2) δίνει δεδομένα σε καθεμία από τις 8 τάξεις (θεωρητικά δεδομένα) με τέτοιο τρόπο κατανομής ώστε να είναι συγκρίσιμα με τα δεδομένα του δείγματος στην 3η στήλη του Πίνακα 2. Τα θεωρητικά αυτά μεγέθη τα υπολογίζουμε με τα αντίστοιχα ολοκληρώματα της *f(x)*στα οικεία διαστήματα και τα καταγράφουμε στην τελευταία στήλη του Πίνακα 2. Μία δοκιμασία $X^2$ δίνει τιμή της παραμέτρου $X^2$=1.2709 <14.00671=$X^2_{7;0.05}$. Άρα οι δειγματικές συχνότητες προσαρμόζονται καλά στις θεωρητικές συχνότητες. Έχουμε ικανοποιητική προσέγγιση.

<u>*Παράδειγμα 2ο:*</u> Δίνεται δείγμα τιμών της συνάρτησης *f(x)/*[*α*, *β*] που είναι συμμετρική και συνεχής. Τα δεδομένα του δείγματος που είναι συστηματικό ως προς x είναι:

Πίνακας 3: Δεδομένα από συστηματικό δείγμα ως προς x για τη συνάρτηση *y=f(x)*

| **i** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8=*n* |
|---|---|---|---|---|---|---|---|---|---|

| **$x_i$** | 5 | 9 | 13 | 17 | 21 | 25 | 29 | 33 | 37 |
|---|---|---|---|---|---|---|---|---|---|
| **$y_i$** | 22 | 22 | 25 | 24 | 25 | 23 | 24 | 24 | 23 |

Να βρεθεί μία προσέγγισή της συνάρτησης πολυωνυμικής μορφής.

Λύση: Από τη γραμμή των $x$ διαπιστώνουμε ότι είναι $\alpha$=5 και $\beta$=37. Άρα $w$=4=($\beta$-$\alpha$)/$n$. Μετασχηματίζουμε σε $X$-5 το $X$ και παίρνουμε τον παρακάτω Πίνακα 4 επεξεργασίας ομαδοποιημένων στοιχείων, με 8 κλάσεις στοιχείων, Φαρμάκης (2001):

Πίνακας 4: Επεξεργασία ταξινομημένων στοιχείων σε 8 τάξεις εύρους $w$=4

| Τάξεις | $\kappa_i$ | $f_i = w\cdot(y_{i-1}+y_i)/2$ | $\kappa'=(\kappa_i-m)/w$ | $f_i\kappa'$ | $f_i\kappa'^2$ | $f_i'$*(θεωρητικά)* |
|---|---|---|---|---|---|---|
| [0, 4) | 2 | 88 | -3 | -264 | 792 | *87.41* |
| [4, 8) | 6 | 94 | -2 | -188 | 376 | *94.60* |
| [8, 12) | 10 | 98 | -1 | -98 | 98 | *97.52* |
| [12, 16) | ***14*** | 98 | 0 | 0 | 0 | *99.47* |
| [16, 20) | 18 | 96 | 1 | 96 | 96 | *99.47* |
| [20, 24) | 22 | 94 | 2 | 188 | 376 | *97.52* |
| [24, 28) | 26 | 96 | 3 | 288 | 864 | *94.60* |
| [28, 32] | 30 | 94 | 4 | 376 | 1504 | *87.41* |
| Σύνολα | | $M$=758 | | $T_1$=398 | $T_2$=4106 | *M=758.00* |

Πήραμε προσωρινή μέση τιμή την $m$=14. Από τα αποτελέσματα του Πίνακα 4 προκύπτει:

Μέση τιμή: $\bar{x} = 14 + \dfrac{4\cdot 398}{758} = 16.10$

Διασπορά: $s^2 = \dfrac{4^2}{757}\cdot\left(4106 - \dfrac{398^2}{758}\right) = 82.3677$ και άρα

Τυπική απόκλιση: $s = \sqrt{82.3677} = 9.08$

Συντελεστής μεταβλητότητας: $Cv = \dfrac{9.08}{16.10} = 0.5637$ και $\lambda = Cv^{-2} = 3.1471$

Άρα εκθέτης $\nu = \dfrac{-5+\sqrt{1+8\cdot 3.1471}}{2} = 0.0582$ και συντελεστής $h = 0.028143$.

Από τα ανωτέρω και τον τύπο (1.3) έχουμε

$$g(x) = \begin{cases} 0.028143 \cdot x^{0.0582} & x \in [0,16) \\ 0.028143 \cdot (32 - x)^{0.0582} & x \in [16,32]. \\ 0 & x \notin [0,32] \end{cases} \tag{2.3}$$

και τελικά η

$$f(x) = \begin{cases} 20.15995 \cdot x^{0.0582} & x \in [0,16) \\ 20.15995 \cdot (32 - x)^{0.0582} & x \in [16,32]. \\ 0 & x \notin [0,32] \end{cases} \tag{2.4}$$

είναι η εκτιμήτρια της συνάρτησης *f(x)* από την οποία είχαμε τα αρχικά δεδομένα σε συστηματική παρατήρηση με βήμα *w*=4. Οίκοθεν νοείται ότι για κάθε x (π.χ. x=13) που μας δίνεται παίρνουμε το αντίστοιχο x-5 και το θέτουμε στην (2.4). Έτσι για x=13 θα μπει η τιμή 13-5=8 και η τιμή της συνάρτησης θα είναι: *f(*8*)*=22.75

Στη συνέχεια επιχειρούμε να ελέγξουμε αν η συνάρτηση (2.4) δίνει δεδομένα σε καθεμία από τις 8 τάξεις (θεωρητικά δεδομένα) με τέτοιο τρόπο κατανομής ώστε να είναι συγκρίσιμα με τα δεδομένα του δείγματος στην 3η στήλη του Πίνακα 4. Τα θεωρητικά αυτά μεγέθη τα υπολογίζουμε με τα αντίστοιχα ολοκληρώματα της *f(x)* στα οικεία διαστήματα και τα καταγράφουμε στην τελευταία στήλη του Πίνακα 4. Μία δοκιμασία $X^2$ δίνει τιμή της παραμέτρου $X^2$=0.797084<14.00671=$X^2_{7;0.05}$. Άρα οι δειγματικές συχνότητες προσαρμόζονται καλά στις θεωρητικές συχνότητες. Έχουμε μια πολύ ικανοποιητική προσέγγιση.

Στο παράδειγμα αυτό οι απόλυτες συχνότητες φαίνονται περίπου ίσες κυμαινόμενες ελαφρά γύρω από το 95. Αυτό υποβάλει την ιδέα της ομοιόμορφης κατανομής, που είναι συμμετρική επίσης, ως προέλευση των δεδομένων. Θεωρώντας άλλωστε τη διασπορά $s^2 = 82.3677$ του δείγματος και εκείνη της ομοιόμορφης κατανομής 85.333=$(\beta\text{-}\alpha)^2/12$ διαπιστώνουμε ότι είναι περίπου ισοδύναμες. Αυτή η παρατήρηση και το αποτέλεσμα της δοκιμασίας $X^2$ μας βεβαιώνουν ότι η κατανομή που μελετάμε είναι η ομοιόμορφη.

## *3. Συμπεράσματα*

Η σημασία της χρήσης του ΣΜ σε διάφορες εφαρμογές είναι από παλιά γνωστή, κύρια στις Βιολογικές και Οικονομικές Επιστήμες. Στην παρούσα εργασία αναδεικνύεται ο ρόλος του ΣΜ στην επίτευξη ενός πιο κεντρικού στόχου στα Μαθηματικά και ιδιαίτερα στη Στατιστική: «Ο ΣΜ βοηθάει στον προσδιορισμό των συναρτήσεων μέσα από δειγματικά δεδομένα, αρκεί αυτές να πληρούν κάποιες συνθήκες, όπως να είναι συμμετρικές. Όταν δεν είναι συμμετρικές υπάρχουν δυσκολίες αλλά υπάρχουν κι εκεί κάποιες ευκαιρίες χρήσης του. Ο προσδιορισμός των συναρτήσεων γίνεται φυσικά κατ' εκτίμηση μέσα από την πολύ εύχρηστη μορφή του πολυωνύμου, πράγμα που βοηθάει να χρησιμοποιείται η μέθοδος και από ερευνητές που δεν είναι Μαθηματικοί κατ' ανάγκη. Μπορούν μάλιστα οι ερευνητές να κάνουν χρήση και των δυνατοτήτων της Πληροφορικής κατασκευάζοντας κάποιο πρόγραμμα που να βρίσκει αυτόματα τη συνάρτηση με το που θα δίνονται τα δειγματικά δεδομένα. Αυτό το τελευταίο μάλιστα είναι μία καλή ιδέα που προτείνεται η χρήση της από τους διδάσκοντες Στατιστική και ειδικά στο κεφάλαιο περί κατανομών».

Σημειώνεται η ανάγκη να συσχετιστεί η παρούσα μελέτη με άλλες μελέτες άλλων κατανομών, π.χ. αύξουσας μορφής, ώστε να αντιμετωπίζονται και πιο σύνθετες σ.π.π. ή έστω απλά συναρτήσεις. Αυτό όμως ξεφεύγει από το πλαίσιο της παρούσας εργασίας και θα αποτελέσει αντικείμενο άλλων άρθρων στο μέλλον, μόλις αντιμετωπιστούν και κάποια ευαίσθητα σημεία τέτοιων συναρτήσεων.

# Continuous functions and coefficient of variation: Reaching via sampling

**Nikolaos Farmakis**

*Department of Mathematics*
*Aristotle University of Thessaloniki, Greece*

*Abstract*

Coefficient of Variation (Cv) is a very useful tool for several statistical applications. It has been used for bias correction of some kind of data, Also it is used for a polynomial expression, estimating the probability density function (pdf) f(x), of a random variable X in Statistics, Farmakis (2003, 2010), and in too many other Economical or Biological studies. In this paper we try to reach with a polynomial form some kind of continuous functions f(x)(not only pdf) via sampling data of this functions. We try via Systematic sample. After the data collection we try to work with a process like the known from Farmakis (2003) for the pdf. So the new idea here is to extend the use of the process for the pdf to any other continuous function f(x) defined in a space [α, β] subset of the Real Numbers set **R**. At first we try to work on symmetrical functions and a more general goal of ours is to work afterwards on some other forms of functions f(x) like Increasing and decreasing ones or some combination of them. Some illustrative examples are presented and with the proposed method we reach the studied function f(x), via sampled data as a polynomial function. Note that the above proposed method includes also the idea of an algorithm and some computer programs can be used, like EXCEL, etc. Therefore, a successful and fast way is established in order to obtain the parameters of the studied function f(x) from sampling data.

## *Βιβλιογραφία*

Sachs, L., (1984). *Applied Statistics,* Springer-Verlang Inc, New York.

Φαρμάκης, Ν., (2001). *Στατιστική, Περιληπτική Θεωρία-Ασκήσεις*, Α & Π Χριστοδουλίδη, Θεσσαλονίκη

Farmakis, N., (2003). Estimation of Coefficient of Variation: Scaling of Symmetric Continuous Distributions, *Statistics in Transition,* Vol. 6, Nr 1, pp *83–96.*

Φαρμάκης, Ν., (2009). *Εισαγωγή στη Δειγματοληψία*, Α & Π Χριστοδουλίδη, Θεσσαλονίκη

Farmakis, N., (2010). Coefficient of Variation: Connecting Sampling with some Increasing Distribution Models, *Proceedings of Stochastic Modelling Techniques and Data Analysis International Conference*, SMTDA-2010, Chania, Crete, Greece, part A-H, pp 259-267.

Ευχαριστίες:

Εκφράζονται ευχαριστίες στους κριτές για τις εύστοχες και εποικοδομητικές παρατηρήσεις και διορθώσεις.