Πολύγωνο αθροιστικών σχετικών συχνοτήτων και διάμεσος μιας τυχαίας μεταβλητής

Δρ. Παναγιώτης Λ. Θεοδωρόπουλος
πρώην Σχολικός Σύμβουλος ΠΕ03
e-mail@p-theodoropoulos.gr

Πρόλογος

Στην εργασία αυτή μελετώνται η γραφική παράσταση *των αθροιστικών σχετικών συχνοτήτων* ενός δείγματος τιμών και η *διάμεσος* μιας τυχαίας μεταβλητής με αφορμή κάποιες αστοχίες που παρατηρούνται στα σχολικά βιβλία των Μαθηματικών (θεωρίας και λύσεων των ασκήσεων) της Γενικής Παιδείας της Γ΄ Λυκείου. Με την διαλεύκανση των σημείων αυτών γίνεται και η ανάλογη ανάλυση και εμβάθυνση της αντίστοιχης θεωρίας.

1. Γραφική παράσταση των αθροιστικών σχετικών συχνοτήτων

Στο βιβλίο λύσεων των ασκήσεων του βιβλίου των Μαθηματικών Γενικής Παιδείας της Γ΄ Λυκείου, που έχει δοθεί στους μαθητές, για την άσκηση 1 των Γενικών Ασκήσεων της σελίδας 126, της οποίας η εκφώνηση είναι:

Ο αριθμός των παιδιών σε ένα δείγμα 80 οικογενειών μιας πόλης δίνεται στον παρακάτω πίνακα:

Αριθμός παιδιών	0	1	2	3	4	5	6
Οικογένειες	10	25	20	12	6	5	2

α) Να βρείτε τη μέση τιμή, τη διάμεση τιμή, την επικρατούσα τιμή και την τυπική απόκλιση του αριθμού των παιδιών.
β) Να κατασκευάσετε το διάγραμμα σχετικών συχνοτήτων και το πολύγωνο αθροιστικών σχετικών συχνοτήτων.
γ) Από το πολύγωνο αθροιστικών σχετικών συχνοτήτων να εκτιμήσετε τα τρία τεταρτημόρια.

δίνεται η παρακάτω λύση[1]:

[1] Είναι ακριβώς η λύση που υπάρχει στο βιβλίο αυτό.

1. Κατασκευάζουμε τον πίνακα συχνοτήτων

x_i	v_i	N_i	$x_i v_i$	$x_i^2 v_i$	f_i%	F_i%
0	10	10	0	0	12,50	12,50
1	25	35	25	25	31,25	43,75
2	20	55	40	80	25,00	68,75
3	12	67	36	108	15,00	83,75
4	6	73	24	96	7,50	91,25
5	5	78	25	125	6,25	97,50
6	2	80	12	72	2,50	100,00
Σύν.	80	—	162	506	100,00	—

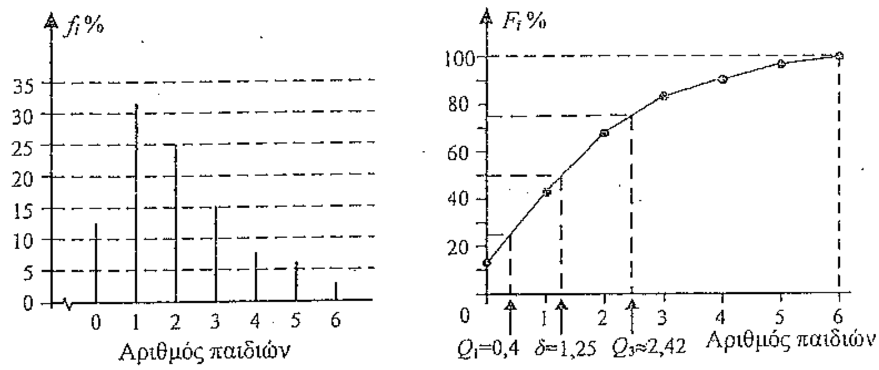
από τον οποίο βρίσκουμε

α) $\bar{x} = \dfrac{\sum x_i v_i}{\sum v_i} = \dfrac{162}{80} = 2{,}025$, $\delta = \dfrac{2+2}{2} = 2$, $M_0 = 1$

$$s^2 = \frac{1}{v}\left\{\sum x_i^2 v_i - \frac{\left(\sum x_i v_i\right)^2}{v}\right\} = \frac{1}{80}\left\{506 - \frac{162^2}{80}\right\} = 2{,}224 .$$

Άρα $s = \sqrt{2{,}224} = 1{,}49$.

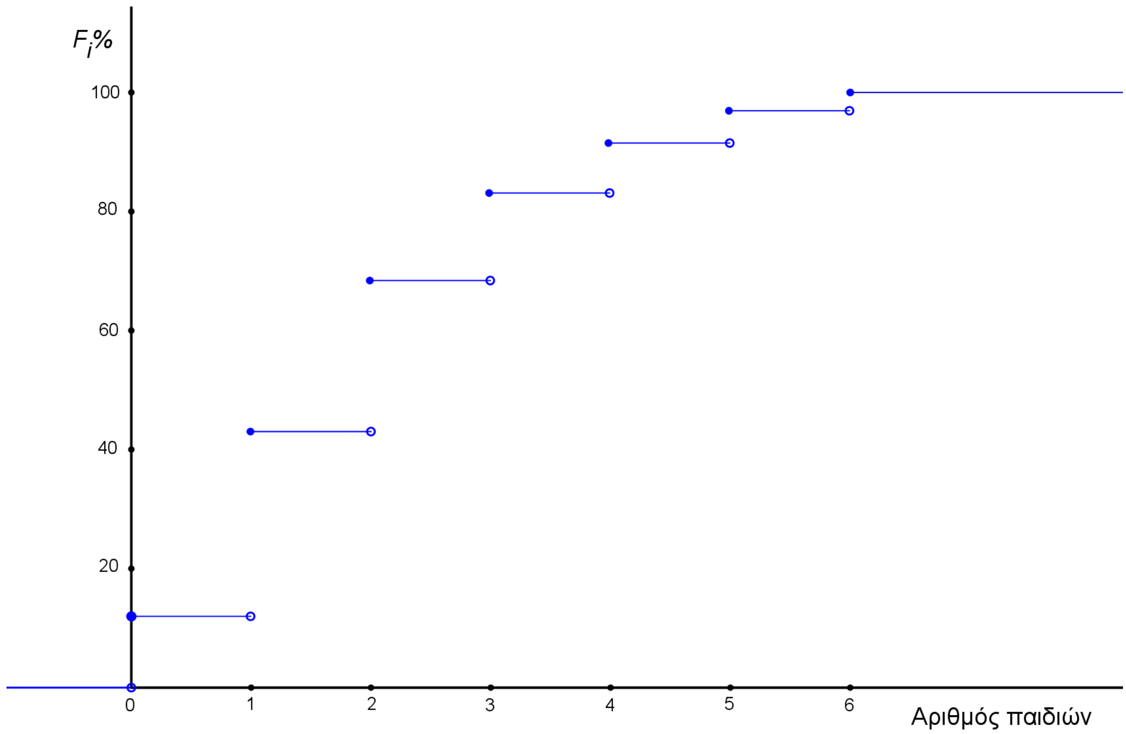
β)



Από το πολύγωνο αθροιστικών σχετικών συχνοτήτων εκτιμούμε
$Q_1 \approx 0{,}4$, $\delta \approx 1{,}25$ *και* $Q_3 \approx 2{,}4$.

Στην άσκηση ζητείται να κατασκευαστεί “το πολύγωνο αθροιστικών σχετικών συχνοτήτων” και στη λύση που υπάρχει στο βιβλίο λύσεων των ασκήσεων έχει κατασκευαστεί το πολύγωνο που βλέπουμε παραπάνω, το οποίο δημιουργεί προβληματισμό ως προς την ορθότητα.

Γνωρίζουμε ότι οι *αθροιστικές σχετικές συχνότητες* ενός δείγματος τιμών αποτελούν τη *δειγματική συνάρτηση κατανομής* της αντίστοιχης τυχαίας μεταβλητής. Επίσης, είναι γνωστό ότι η *συνάρτηση κατανομής* μιας *διακριτής* τυχαίας μεταβλητής είναι μια *κλιμακωτή συνάρτηση* (δείτε οποιοδήποτε βιβλίο Θεωρίας Πιθανοτήτων). Έτσι λοιπόν για τα δεδομένα της άσκησης και σύμφωνα με τον παραπάνω πίνακα η γραφική παράσταση των αθροιστικών σχετικών συχνοτήτων ή της δειγματικής συνάρτησης κατανομής της μεταβλητής είναι το κλιμακωτό διάγραμμα που παρουσιάζεται στο επόμενο σχήμα.



Γίνεται φανερό ότι με το πολύγωνο που έχει κατασκευαστεί στη λύση που υπάρχει στο βιβλίο λύσεων δεν αποδίδεται σωστά η κατανομή των σχετικών συχνοτήτων της μεταβλητής. Συνεπώς δεν μπορεί να θεωρηθεί ως σωστό. Αν είχε γίνει μια ομαδοποίηση των τιμών του δείγματος, τότε θα ήταν δικαιολογημένη η κατασκευή πολυγώνου. Σε ομαδοποιημένα δεδομένα δεχόμαστε ότι η κατανομή συχνοτήτων σε κάθε κλάση είναι η *ομοιόμορφη συνεχής*, οπότε η αντίστοιχη συνάρτηση αθροιστικών σχετικών συχνοτήτων είναι *γραμμική* με γραφική παράσταση ένα *ευθύγραμμο τμήμα*. Σχεδιάζοντας αυτά τα ευθύγραμμα τμήματα για όλες τις κλάσεις προκύπτει το πολύγωνο αθροιστικών σχετικών συχνοτήτων. Αν ο αριθμός των κλάσεων τείνει στο άπειρο και το πλάτος τους στο μηδέν, τότε το πολύγωνο των αθροιστικών σχετικών συχνοτήτων τείνει στη γραφική παράσταση της συνάρτησης κατανομής της μεταβλητής (παραβλέπουμε τα κατακόρυφα τμήματα στην περίπτωση που η μεταβλητή είναι διακριτή).

2. Η διάμεσος μιας τυχαίας μεταβλητής

Σχετικά τώρα με την διάμεσο ενός δείγματος τιμών, στο σχολικό βιβλίο στη σελίδα 88 αναφέρονται τα εξής:

«Παρατηρούμε ότι, η διάμεσος είναι η τιμή που χωρίζει ένα σύνολο παρατηρήσεων σε δύο ίσα μέρη όταν οι παρατηρήσεις αυτές τοποθετηθούν με σειρά τάξης μεγέθους. Ακριβέστερα, η διάμεσος είναι η τιμή για την οποία το πολύ 50% των παρατηρήσεων είναι μικρότερες από αυτήν και το πολύ 50% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτήν.»

Η *διάμεσος ενός δείγματος* ή η *δειγματική διάμεσος* ορίζεται *μονοσήμαντα*, γιατί ορίζεται με τη βοήθεια μιας *στατιστικής συνάρτησης*. Όμως, ο τρόπος που επιχειρείται στο σχολικό βιβλίο να περιγραφεί η ιδιότητα της δειγματικής διαμέσου δεν είναι σωστός, διότι με το οριστικό άρθρο "**η**", που χρησιμοποιείται στην περιγραφή (η τιμή …), δηλώνεται ότι η διάμεσος ενός δείγματος είναι ο μοναδικός αριθμός που έχει αυτή την ιδιότητα, ενώ μπορεί αυτό να μην ισχύει. Για παράδειγμα, αν οι παρατηρήσεις ενός δείγματος μεγέθους ν = 20 διαταγμένες κατά αύξουσα σειρά είναι:

2, 2, 3, 3, 3, 4, 4, 5, 5, 6, 7, 7, 7, 8, 8, 9, 9, 9, 9, 10

τότε η διάμεσος του δείγματος αυτού είναι **6,5**. Παρατηρούμε ότι το πολύ 50% (εδώ το 50%) των παρατηρήσεων είναι μικρότερες του 6,5 και το πολύ 50% (εδώ το 50%) των παρατηρήσεων είναι μεγαλύτερες του 6,5. Είναι όμως ο αριθμός 6,5 ο μοναδικός που έχει αυτή την ιδιότητα; Ασφαλώς όχι, αφού κάθε αριθμός του διαστήματος [6, 7], όπως μπορούμε εύκολα να επαληθεύσουμε, έχει την ιδιότητα αυτή.

Γι' αυτό λοιπόν μία σωστή διατύπωση της παραπάνω ιδιότητας της διαμέσου ενός δείγματος τιμών είναι η εξής:

Η διάμεσος ενός δείγματος τιμών *έχει την ιδιότητα: το πολύ 50% των παρατηρήσεων είναι μικρότερες από αυτήν και το πολύ 50% των παρατηρήσεων μεγαλύτερες.*

Ενώ η δειγματική διάμεσος ορίζεται μονοσήμαντα, επειδή όπως αναφέρθηκε ορίζεται με τη βοήθεια μιας στατιστικής συνάρτησης, η διάμεσος μιας τυχαίας μεταβλητής X στη Θεωρία Πιθανοτήτων δεν είναι απαραίτητα μοναδικός αριθμός, αφού δεν ορίζεται με συναρτησιακό τρόπο. Η διάμεσος μιας τυχαίας μεταβλητής X ορίζεται ως εξής:

«***Διάμεσος** ή **διχοτόμος** μιας τυχαίας μεταβλητής X με συνάρτηση κατανομής $F(x)$ λέγεται κάθε αριθμός δ για τον οποίο ισχύει:*

$$P(X < \delta) = F(\delta -) \ \le \ \frac{1}{2} \ \le \ P(X \le \delta) = F(\delta) \text{ ».}$$

(Δείτε Θ. Κάκουλλου: *Μαθήματα Θεωρίας Πιθανοτήτων* (1975), σελίδα 58).

Εάν η F είναι συνεχής, τότε από τον παραπάνω ορισμό προκύπτει ότι διάμεσος της X είναι κάθε λύση δ της εξίσωσης:

$$F(\delta) = \frac{1}{2} \ .$$

Κάποιες τυχαίες μεταβλητές έχουν μοναδική διάμεσο, ενώ κάποιες άλλες όχι. Ας δούμε μερικά χαρακτηριστικά παραδείγματα.

i. Έστω η διακριτή τυχαία μεταβλητή X με τιμές **1**, **2**, **3**, **4** και **5** και με πιθανότητες:

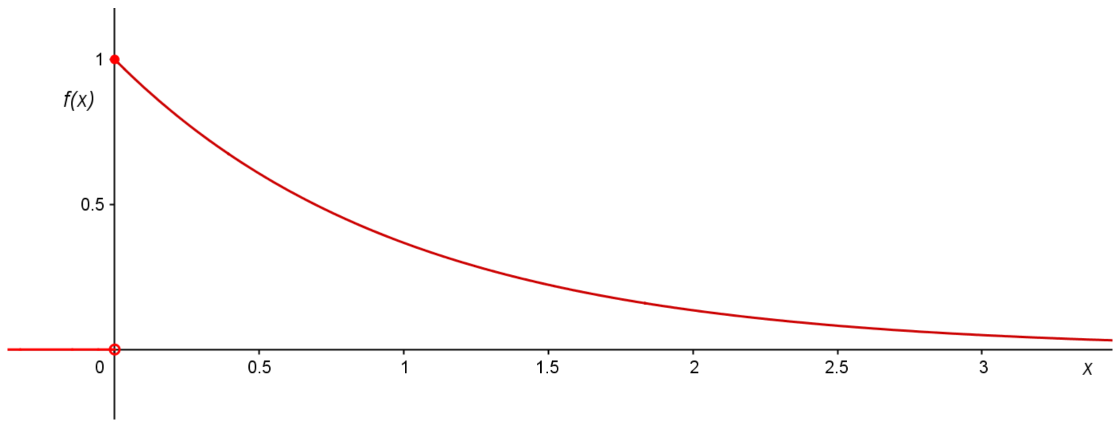
$$P(1) = \frac{3}{9},\ P(2) = \frac{2}{9},\ P(3) = \frac{1}{9},\ P(4) = \frac{2}{9} \ \text{ και } \ P(5) = \frac{1}{9}.$$

Παρατηρούμε ότι ισχύει:

$$P(X < 2) = F(2-) = \frac{3}{9} \le \frac{1}{2} \le P(X \le 2) = F(2) = \frac{5}{9}$$

και ότι ο αριθμός **2** είναι ο μοναδικός που έχει αυτή την ιδιότητα. Άρα η τιμή 2 είναι η μοναδική διάμεσος της μεταβλητής αυτής.

ii. Έστω μία συνεχής τυχαία μεταβλητή X που ακολουθεί την εκθετική κατανομή με συνάρτηση πυκνότητας πιθανότητας $f(x) = \theta e^{-\theta x}$, αν $x \ge 0$ και $f(x) = 0$, αν $x < 0$, όπου θ μία θετική παράμετρος.
Για $\theta = 1$ η f έχει την παρακάτω γραφική παράσταση[2]:



Η συνάρτηση κατανομής της εκθετικής είναι η εξής: $F(x) = 1 - e^{-\theta x}$ για $x \ge 0$ και $F(\mathrm{x}) = 0$ για $x < 0$.

Εύκολα βρίσκουμε ότι η εξίσωση $F(\delta) = \frac{1}{2}$, δηλαδή η εξίσωση $1 - e^{-\theta\delta} = \frac{1}{2}$

έχει μοναδική λύση, που είναι ο αριθμός $\delta = \frac{1}{\theta}\ln 2$.

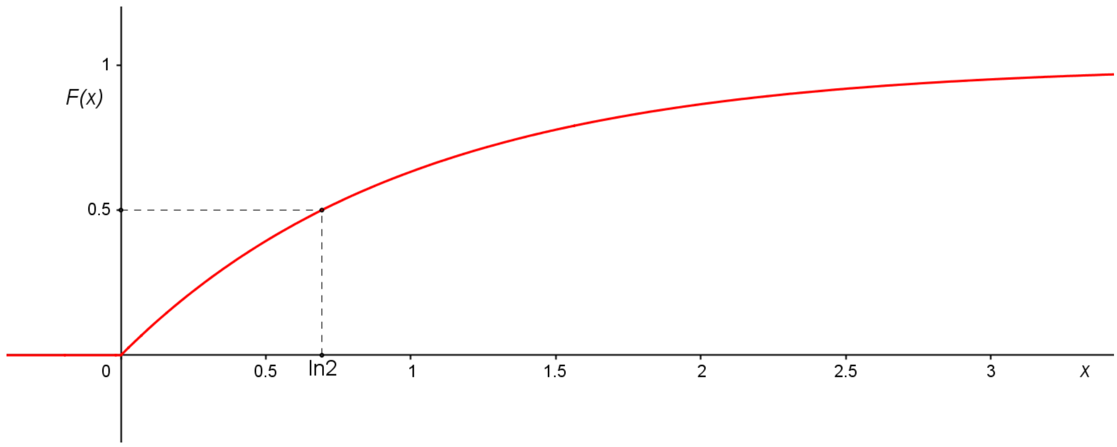
Άρα, αν μία τυχαία μεταβλητής X ακολουθεί την εκθετική κατανομή με συνάρτηση πυκνότητας πιθανότητας $f(x) = \theta e^{-\theta x}$, αν $x \ge 0$ και $f(x) = 0$, αν $x < 0$, τότε έχει μοναδική διάμεσο, που είναι ο αριθμός:

$$\delta = \frac{1}{\theta}\ln 2 .$$

Στο παρακάτω σχήμα στο οποίο απεικονίζεται η γραφική παράσταση της συνάρτησης κατανομής της εκθετικής για $\theta = 1$ φαίνεται καθαρά η μοναδικότητα της διαμέσου της εκθετικής κατανομής.

---

[2] Όλες οι γραφικές παραστάσεις της εργασίας έχουν γίνει με το λογισμικό GeoGebra.



iii. Έστω η διακριτή τυχαία μεταβλητή X με τιμές **2**, **3**, **4**, **5** και **6** και με πιθανότητες:

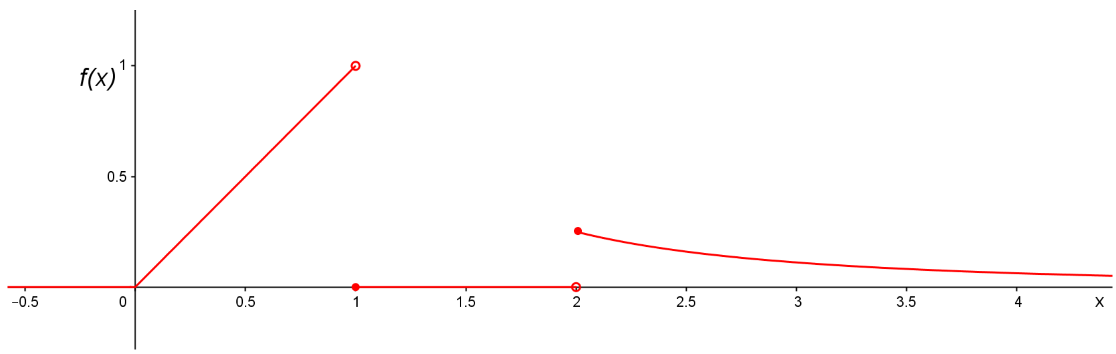
$$P(2)=\frac{3}{10},\ P(3)=\frac{2}{10},\ P(4)=\frac{2}{10},\ P(5)=\frac{2}{10}\ \text{ και }\ P(6)=\frac{1}{10}.$$

Ως διάμεσος της μεταβλητής αυτής μπορεί να θεωρηθεί κάθε αριθμός του διαστήματος [3, 4], αφού κάθε αριθμός του διαστήματος αυτού ικανοποιεί τον παραπάνω ορισμό της διαμέσου μιας τυχαίας μεταβλητής.

iv. Έστω η συνεχής τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας:

$$f(x)=\begin{cases}0, & x<0\\ x, & 0\le x<1\\ 0, & 1\le x<2\\ \dfrac{1}{x^2}, & x\ge 2\end{cases}$$

της οποίας η γραφική παράσταση είναι:



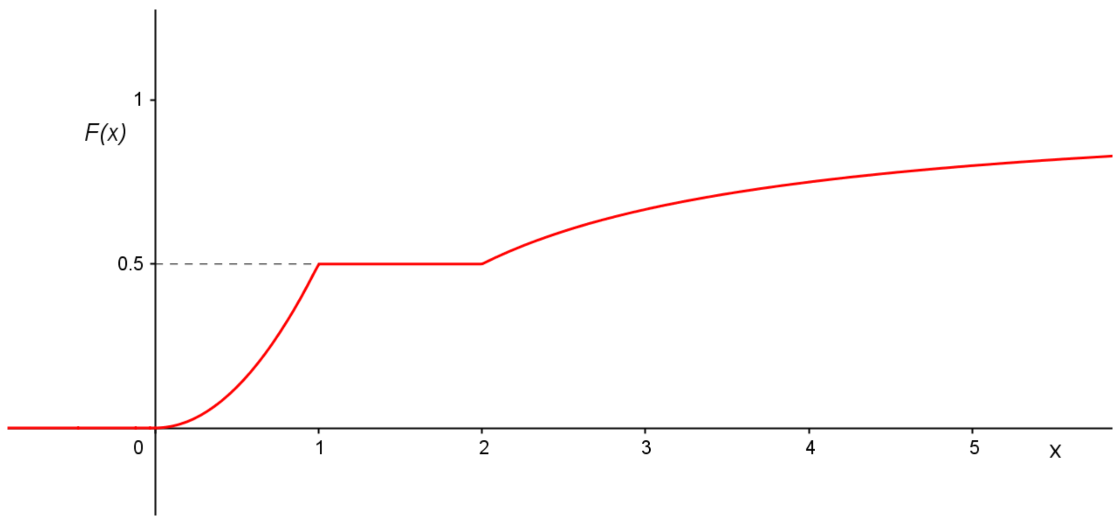
Εύκολα επαληθεύεται ότι η f είναι συνάρτηση πυκνότητας πιθανότητας, δηλαδή ότι έχει τις ιδιότητες:

$$1)\quad f(x) \geq 0,\ x \in IR \quad \text{και} \quad 2)\quad \int_{-\infty}^{+\infty} f(x)dx = 1 .$$

Η συνάρτηση κατανομής της παραπάνω τυχαίας μεταβλητής X είναι η:

$$F(x) = \begin{cases} 0,\ x < 0 \\ \dfrac{x^2}{2},\ 0 \leq x < 1 \\ \dfrac{1}{2},\ 1 \leq x < 2 \\ \dfrac{x-1}{x},\ x \geq 2 \end{cases}$$

με γραφική παράσταση:



Παρατηρούμε ότι κάθε αριθμός του διαστήματος [1, 2] είναι λύση της εξίσωσης $F(\delta) = 0{,}5$. Άρα κάθε αριθμός του διαστήματος αυτού μπορεί να θεωρηθεί ως διάμεσος της παραπάνω τυχαίας μεταβλητής X.

<u>Σημείωση:</u> Αν δεν ορίζεται μονοσήμαντα η διάμεσος μιας τυχαίας μεταβλητής, τότε συνήθως για πρακτικούς λόγους θεωρούμε ως διάμεσο της μεταβλητής αυτής την κεντρική τιμή των τιμών που ικανοποιούν τον ορισμό της διαμέσου μιας τυχαίας μεταβλητής. Έτσι λοιπόν, για την τυχαία μεταβλητή του παραδείγματος iii μπορούμε να θεωρήσουμε συμβατικά ως διάμεσο τον αριθμό 3,5 και για την τυχαία μεταβλητή του παραδείγματος iv τον αριθμό 1,5.