



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ  
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ ΚΑΙ  
ΠΕΡΙΒΑΛΛΟΝΤΟΣ

# Η πιθανοτική κατανομή του κλιμακογράμματος με χρήση τεχνικής Monte Carlo

---

Εφαρμογή σε ανελίξεις  
μακροπρόθεσμης εμμονής

Διπλωματική Εργασία  
Παναγιώτα-Μαρίνα Γουρνάρη  
Επιβλέπων: Δημήτρης Κουτσογιάννης, Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2017



## ΕΥΧΑΡΙΣΤΙΕΣ

---

Φτάνοντας στο τέλος της φοιτητικής μου πορείας θα ήθελα να ευχαριστήσω τους ανθρώπους που διαδραμάτισαν καθοριστικό ρόλο και συνέβαλαν στην ομαλή περάτωση των σπουδών μου.

Θα ήθελα να ευχαριστήσω θερμά τον κ.Δημήτρη Κουτσογιάννη, Κοσμήτορα και Καθηγητή της Σχολής Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου και επιβλέποντα της διπλωματικής μου εργασίας. Αρχικά, για τον χρόνο και την καθοδήγησή του, καθώς και για την δυνατότητα που μου έδωσε, ώστε να εκπονήσω ένα θέμα που μου είχε κεντρίσει το ενδιαφέρον. Αποτέλεσε το κίνητρο για να ασχοληθώ με την Στοχαστική Υδρολογία, όπως επίσης μου έδωσε την ευκαιρία να συμμετάσχω στο συνέδριο E.G.U. το 2016 και 2017 με δύο ερευνητικές εργασίες.

Ευχαριστώ εξίσου θερμά τον υποψήφιο διδάκτορα κ.Παναγιώτη Δημητριάδη, ο οποίος με ενθάρρυνε να ασχοληθώ με αυτό το θέμα και η βοήθειά του ήταν ιδιαίτερα σημαντική για την εκπόνηση της διπλωματικής. Οι συμβουλές του καθώς και η συνεχής συνεργασία και επικοινωνία που είχαμε αποτέλεσαν το εφελτήριο για να επιτευχθεί το συγκεκριμένο αποτέλεσμα.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου, Αριστόδημο και Ειρήνη, και τον αδελφό μου Γιώργο για την συνεχή συμπαράσταση και δύναμη, που εισέπραξα, καθώς και τους φίλους μου για όλες τις όμορφες φοιτητικές στιγμές.

Παναγιώτα-Μαρίνα Γουρνάρη

Ιούλιος 2017



# ΠΕΡΙΛΗΨΗ

---

Πολλά γεωφυσικά φαινόμενα χαρακτηρίζονται από τυχαιότητα και δεν μπορούν να περιγραφούν με ντετερμινιστικό τρόπο. Η διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής στις γεωφυσικές διεργασίες έχει συγκεντρώσει το ενδιαφέρον πολλών επιστημόνων τις τελευταίες δεκαετίες. Αναφερόμαστε σε διεργασίες που ακολουθούν κανονική κατανομή για τον λόγο ότι ως υδρολόγοι, μελετώντας διεργασίες σε ετήσια και υπερετήσια κλίμακα, που μας ενδιαφέρει κυρίως για την διαχείριση των υδατικών πόρων, φαίνονται να ακολουθούν κανονική κατανομή κάτι που προκύπτει και από το κεντρικό οριακό θεώρημα. Χρησιμοποιούμε ως μοντέλο που περιγράφει την δομή αυτοσυσχέτισης των δεδομένων το μοντέλο απλής ομοιοθεσίας διότι έχει πλέον αποδειχθεί ότι η συμπεριφορά των φυσικών διεργασιών έχει απόκλιση από αυτήν που περιγράφουν τα απλά στοχαστικά μοντέλα. Για την εκτίμηση της μακροπρόθεσμης εμμονής εφαρμόστηκε η μέθοδος του κλιμακογράμματος δηλαδή της διασποράς συναρτήσεως της κλίμακας. Με αυτό τον τρόπο ποσοτικοποιείται η παράμετρος Hurst καθώς επίσης παρουσιάζει μικρότερη αβεβαιότητα συγκριτικά με άλλα στοχαστικά εργαλεία. Στην παρούσα διπλωματική εργασία μελετάται η εύρεση της πιθανοτικής κατανομής του κλιμακογράμματος σε κάθε κλίμακα. Επιπλέον μελετάται η επίδραση της μακροπρόθεσμης εμμονής στην εκτίμηση των στατιστικών χαρακτηριστικών των φαινομένων λόγω της μεροληψίας και της εξάρτησης μεταξύ των δεδομένων. Οι αναλύσεις έγιναν με χρήση τεχνικής Monte Carlo για την εξαγωγή των κλιμακογραμμάτων και των κατανομών τους. Αποδεικνύεται ότι για την περίπτωση του λευκού θορύβου που αντιστοιχεί σε παράμετρο  $H = 0.5$  η κατανομή του κλιμακογράμματος είναι η  $\chi^2$  κατανομή ενώ καθώς αυξάνεται η παράμετρος Hurst η κατανομή γίνεται ασύμμετρη και προσεγγίζει την κατανομή γάμμα. Επιπλέον παρουσιάζονται αναλυτικά τα αποτελέσματα των στατιστικών χαρακτηριστικών καταλήγοντας στο συμπέρασμα ότι η πιθανότερη τιμή συγκλίνει προς το κάτω τεταρτημόριο, ένα συμπέρασμα που μπορεί να φανεί χρήσιμο στις περιπτώσεις όπου έχουμε λίγα δεδομένα. Τέλος παρουσιάζονται κάποιες εφαρμογές όπου η συγκεκριμένη ανάλυση μπορεί να συμβάλλει στην ορθότερη προσομοίωση και πρόβλεψη των στοχαστικών διεργασιών.



# ABSTRACT

---

Many hydrological and geo-physical phenomena can be described by randomness instead of deterministic terms. The hypothesis testing of long range dependence structure in geophysical data has raised high scientific interest in the last decades. We refer to normally distributed processes due to the fact that as hydrologists we examine processes in annual and υπερετήσια scale, being interested in water resources management. The distribution of these processes can be easily derived from the central limit theorem. A simple scaling model known as the Fractional Gaussian noise is chosen to best describe the behavior of the geophysical processes, a stochastic model that was devised to represent the Hurst phenomenon. In order to estimate the long term persistence a stochastic tool is used known as the climacogram i.e. variance of the time averaged process over averaging time scale. In this way not only we can quantify the Hurst coefficient but also this tool compared to others appears to have the lowest uncertainty, an important advantage in stochastic model building. The scope of this thesis is the identification of the statistical distribution of the climacogram in every scale. Moreover, the effect of the long term persistence in the statistical properties of the processes is examined, due to the bias and the dependence between the data. The analysis was carried out using the Monte Carlo method for the generation of the climacograms and their distribution. It can be proved that in the case of  $H = 0.5$  (white noise), the distribution of the climacogram is the chi-square distribution whereas when the Hurst coefficient increases the distribution is getting more skewed and approaches the gamma distribution. Furthermore, the results of various statistical properties are presented analytically, leading to the conclusion that the most probable value is converging to the Q-25, a useful conclusion in case of limited data. Finally, some applications are presented where this particular analysis could contribute in the accurate simulation and prediction of stochastic processes.





# ΠΕΡΙΕΧΟΜΕΝΑ

---

1	Εισαγωγή.....	10
1.1	Γενικά .....	10
1.2	Αντικείμενο Διπλωματικής.....	10
2	Οριοί και Μεθοδολογίες.....	12
2.1	Στατιστική.....	12
2.1.1	Τυχαία μεταβλητή και συνάρτηση κατανομής .....	12
2.1.2	Αναμενόμενες τιμές και παράμετροι κατανομών .....	13
2.1.3	Εκτιμήτριες.....	14
2.1.4	Κανονική Κατανομή.....	15
2.1.5	Η κατανομή Γάμμα.....	16
2.1.6	Η κατανομή $\chi^2$ .....	16
2.1.7	Μέση τιμή .....	17
2.1.8	Διάμεσος.....	18
2.1.9	Πιθανότερη Τιμή.....	18
2.1.10	Ασυμμετρία .....	19
2.1.11	Κύρτωση.....	19
2.1.12	Διασπορά (και τυπική απόκλιση) .....	19
2.1.13	Ροπογεννήτρια συνάρτηση .....	21
2.2	Στοχαστική Ανέλιξη.....	22
2.2.2	Ανέλιξη στη συναθροισμένη κλίμακα .....	25
2.2.3	Απλά μοντέλα αναφοράς.....	26
2.2.4	Παραδείγματα αποκλίσεων από τα απλά μοντέλα αναφοράς.....	29
2.2.5	Μακροπρόθεσμη εμμονή .....	32
2.2.6	Εκτίμηση στατιστικών παραμέτρων με την υπόθεση ΗΚ ανέλιξης... 36	
2.3	Μεθοδολογίες .....	40
2.3.1	SMA .....	40

2.3.2	Κλιμακόγραμμα.....	41
2.3.3	Monte Carlo .....	44
3	Ανάλυση .....	48
3.1	Μεθοδολογική Προσέγγιση .....	48
3.2	Κατανομή κλιμακογράμματος.....	49
3.2.1	Η περίπτωση του λευκού θορύβου .....	49
3.2.2	Η περίπτωση ΗΚ.....	54
4	Εφαρμογές.....	76
4.1	Κλιμακόγραμμα ως στοχαστικό εργαλείο .....	76
4.1.1	Μεθοδολογία .....	76
4.1.2	Αποτελέσματα.....	79
4.2	Παράγωγα Καιρού .....	81
4.2.1	Μεθοδολογία .....	83
5	Συμπεράσματα .....	86
5.1	Σύνοψη Εργασίας.....	86
5.2	Συμπεράσματα.....	87
6	Βιβλιογραφία.....	90

# 1 ΕΙΣΑΓΩΓΗ

---

## 1.1 ΓΕΝΙΚΑ

Είναι ευρύτερα γνωστό ότι πολλά γεωφυσικά φαινόμενα χαρακτηρίζονται από τυχαιότητα και δεν μπορούν να περιγραφτούν με ντετερμινιστικό τρόπο. Η ύπαρξη χωρικής και χρονικής εξάρτησης μεταξύ των φαινομένων αναιρεί την ισχύ των κλασικών στατιστικών νόμων οι οποίοι βασίζονται στην ανεξαρτησία των δεδομένων. Αρκετοί στατιστικολόγοι και φυσικοί επιστήμονες έχουν παρατηρήσει αποκλίσεις στην δομή αυτοσυσχέτισης από τα απλά στοχαστικά μοντέλα λευκού θορύβου και των ανελιξων Markov. Η ύπαρξη μακροπρόθεσμης εμμονής στα δεδομένα επηρεάζει την εκτίμηση των στατιστικών χαρακτηριστικών τους και την προσομοίωση των αντίστοιχων φαινομένων. Η μελέτη της επίδρασης της μακροπρόθεσμης εμμονής στα φαινόμενα αυτά συγκεντρώνει μεγάλο επιστημονικό ενδιαφέρον τα τελευταία χρόνια.

Η μέθοδος της ανάλυσης χρονοσειρών αποτελεί σημαντικό εργαλείο για την επίλυση προβλημάτων που αφορούν ιδιαίτερα την διαχείριση υδατικών πόρων. Ένα βασικό εργαλείο για τον προσδιορισμό της δομής αυτοσυσχέτισης των δεδομένων είναι το κλιμακόγραμμα, δηλαδή η διασπορά συναρτήσεως της κλίμακας. Ο εντοπισμός της μακροπρόθεσμης εμμονής στα δεδομένα είναι πολλές φορές δύσκολος καθώς απαιτεί μεγάλο δείγμα (καταγραφές πολλών ετών) για να μπορέσει να μελετηθεί ικανοποιητικά.

## 1.2 ΑΝΤΙΚΕΙΜΕΝΟ ΔΙΠΛΩΜΑΤΙΚΗΣ

Σκοπός της διπλωματικής είναι η διερεύνηση της πιθανοτικής κατανομής του κλιμακογράμματος με χρήση τεχνικής Monte Carlo για εφαρμογή στις ανελιξεις μακροπρόθεσμης εμμονής. Όταν αναφερόμαστε σε λευκό θόρυβο ( $H = 0.5$ ) η κατανομή της διασποράς (του κλιμακογράμματος στην βασική κλίμακα) αποδεικνύεται ότι είναι η  $\chi^2$  κατανομή. Αυτό όμως δεν ισχύει στις ανελιξεις μακροπρόθεσμης εμμονής λόγω της επίδρασης της στις στατιστικές παραμέτρους της ανέλιξης.

Εφαρμόζεται Monte Carlo ανάλυση για διάφορα σενάρια μήκους δείγματος και παραμέτρων Hurst και αναλύονται στατιστικά τα αποτελέσματά της. Προκύπτουν

έτσι τα διαγράμματα για κάθε στατιστική παράμετρο (μέση τιμή, διασπορά, διάμεσος κ.λ.π.). Καταλήγουμε σε κάποια γενικά συμπεράσματα και αναφέρουμε ενδεικτικά δύο εφαρμογές όπου η λάθος εκτίμηση των στατιστικών παραμέτρων έχει σημαντικές επιπτώσεις.

Απώτερος σκοπός της διπλωματικής εργασίας είναι η κατανόηση του φαινομένου της μακροπρόθεσμης εμμονής και η επίδραση που μπορεί να έχει στα στατιστικά χαρακτηριστικά των γεοφυσικών διεργασιών. Έτσι λαμβάνοντας υπόψη την αβεβαιότητα της εκτίμησης και παρουσιάζοντας τα αποτελέσματα μπορεί να γίνει καλύτερος σχεδιασμός και προσομοίωσης του εκάστοτε μεγέθους.

## 2 ΟΡΙΣΜΟΙ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΕΣ

---

### 2.1 ΣΤΑΤΙΣΤΙΚΗ

#### 2.1.1 Τυχαία μεταβλητή και συνάρτηση κατανομής

Τυχαία μεταβλητή είναι μια συνάρτηση που αντιστοιχεί ένα αριθμό  $X(\omega)$  σε κάθε δυνατή έκβαση  $\omega$ . Γράφουμε  $\{X \leq x\}$  εννοώντας το γεγονός εκείνο που αποτελείται από όλα τα στοιχεία του  $\omega$  τέτοια ώστε οι τιμές του  $X(\omega)$  να είναι μικρότερες του αριθμού  $x$ . Την πιθανότητα του γεγονότος αυτού τη συμβολίζουμε με  $P(\{X \leq x\})$  ή απλούστερα  $P(X \leq x)$ . Ενώ το  $x$  παίρνει τιμές από  $-\infty$  έως  $+\infty$ , ισχύει ότι  $P(X = \infty) = 0$  και  $P(X = -\infty) = 0$ .

Η πιθανότητα  $P(X \leq x)$  ενός γεγονότος  $(X \leq x)$  είναι ένας αριθμός που εξαρτάται από το  $x$ . Αυτός ο αριθμός συμβολίζεται με  $F_x(x)$  και ονομάζεται (αθροιστική) συνάρτηση κατανομής της τυχαίας μεταβλητής  $X$ .

Συνάρτηση κατανομής είναι η συνάρτηση της πραγματικής μεταβλητής  $x$ , που δίνεται από την εξίσωση:

$$F_x(x) := P(X \leq x) \quad (2.1)$$

ορίζεται για κάθε πραγματικό αριθμό  $x$ , από  $-\infty$  έως  $+\infty$ .

$$F(x) = P[X \leq x] \equiv \int_{-\infty}^x f(y) dy \quad \forall x \in \mathbb{R} \quad (2.2)$$

Ισχύει ότι:

$$0 = F_x(-\infty) \leq F_x(x) \leq F_x(+\infty) = 1 \quad (2.3)$$

Η παράγωγος της συνάρτησης κατανομής είναι η:

$$f_x(x) := \frac{dF(x)}{dx} \quad (2.4)$$

και λέγεται συνάρτηση πυκνότητας πιθανότητας. Οι βασικές της ιδιότητες που προκύπτουν άμεσα από τον ορισμό της είναι:

$$f_x(x) \geq 0 \quad \int_{-\infty}^{\infty} f_x(x) dx = 1 \quad (2.5)$$

$$P[a < X \leq b] = \int_a^b f(x)dx \quad a, b \in \mathbb{R} \quad (a < b) \quad (2.6)$$

Είναι προφανές ότι η συνάρτηση πυκνότητας πιθανότητας δεν παριστάνει πιθανότητα και για αυτό μπορεί να πάρει τιμές μεγαλύτερες της μονάδας. Η σχέση της με την πιθανότητα προσδιορίζεται από την εξίσωση:

$$f_x(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} \quad (2.7)$$

### 2.1.2 Αναμενόμενες τιμές και παράμετροι κατανομών

Αν  $X$  είναι συνεχής τυχαία μεταβλητή και  $g(X)$  είναι συνάρτηση της  $X$ , τότε ορίζεται ως αναμενόμενη τιμή ή προσδοκία της  $g(X)$  το μέγεθος

$$E[g(X)] := \int_{-\infty}^{\infty} g(x)f_x(x) dx \quad (2.8)$$

Η αντίστοιχη σχέση για διακριτή τυχαία μεταβλητή  $X$ , που παίρνει τις τιμές  $x_1, x_2, \dots$ , είναι

$$E[g(X)] := \sum_{i=1}^{\infty} g(x_i)P(X = x_i) \quad (2.9)$$

Ειδικότερα:

1. Για  $g(X) = X^r$ , όπου  $r = 0, 1, 2, \dots$ , το μέγεθος

$$m_X^{(r)} := E[X^r] \quad (2.10)$$

ονομάζεται ροπή περί την αρχή (ή απλώς ροπή) τάξης  $r$  της  $X$ .

2. Για  $g(X) = X$ , το μέγεθος

$$m_X := E[X] \quad (2.11)$$

(δηλαδή η ροπή τάξης 1) ονομάζεται αναμενόμενη τιμή ή μέση τιμή της  $X$ .

Εναλλακτικά για τη μέση τιμή χρησιμοποιείται και το σύμβολο  $\mu_X$ .

3. Για  $g(X) = (X - m_X)^r$ , το μέγεθος

$$\mu_X^{(r)} := E[(X - \mu_X)^r] \quad (2.12)$$

ονομάζεται κεντρική ροπή τάξης  $r$  της  $X$ .

4. Για  $g(X) = (X - \mu_X)^2$ , το μέγεθος

$$\sigma_X^2 := \mu_X^{(2)} = E[(X - m_X)^2] \quad (2.13)$$

(δηλαδή η κεντρική ροπή τάξης 2) ονομάζεται διασπορά της  $X$ . Η διασπορά συμβολίζεται ακόμα και με  $Var[X]$ .

Εύκολα μπορεί να αποδειχθεί ότι:

$$\begin{aligned} V[X] &= E\{(X - \mu)^2\} = E(X^2 + \mu^2 - 2\mu X) \\ &= E(X^2) + E(\mu^2) - 2\mu E(X) = E(X^2) - \mu^2 \end{aligned}$$

Άρα

$$V(X) = E(X^2) - \{E(X)\}^2 \quad (2.14)$$

Η μηδενική ροπή, περί την αρχή ή κεντρική, είναι πάντα ίση με ένα ενώ η πρώτη κεντρική ροπή είναι μηδέν.

### 2.1.3 Εκτιμήτριες

Εκτιμήτρια είναι μια τυχαία μεταβλητή που χρησιμοποιείται για να εκτιμήσει ένα χαρακτηριστικό ενός πληθυσμού.

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

$$E[\hat{\theta}_n] = \theta \quad (2.15)$$

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

$$bias = E[\hat{\theta}_n] - \theta \quad (2.16)$$

Μας ενδιαφέρει να χρησιμοποιούμε αμερόληπτες εκτιμήτριες γιατί αλλιώς θα έχουμε υπερεκτίμηση ή υποεκτίμηση της παραμέτρου.

Η μεροληψία οφείλεται στην διαφορά μεταξύ της μέσης τιμής της εκτιμήτριας και της πραγματικής παραμέτρου, ενώ η διασπορά προκύπτει από τις διακυμάνσεις της εκτιμήτριας λόγω της στατιστικής δειγματοληψίας.

### 2.1.4 Κανονική Κατανομή

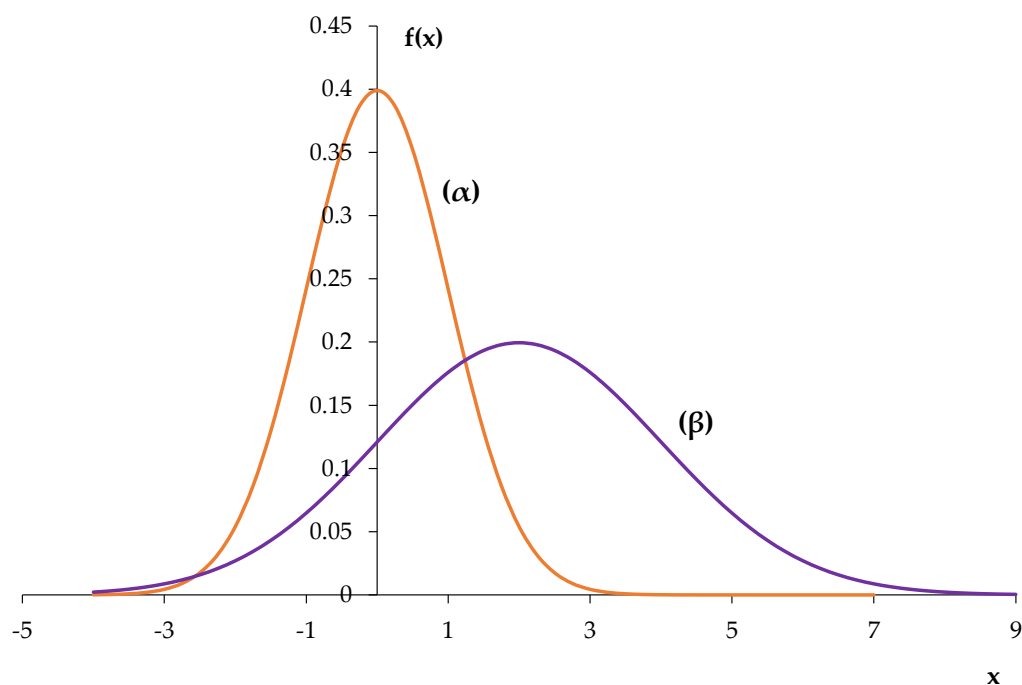
Λέμε ότι η μεταβλητή  $X$  είναι κανονική (ή ότι ακολουθεί κανονική κατανομή ή κατανομή Gauss) με παραμέτρους  $\mu$  και  $\sigma$  (συμβολικά  $N(\mu, \sigma)$ ) αν η πυκνότητα πιθανότητάς της είναι η συνάρτηση:

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.17)$$

Η αντίστοιχη συνάρτηση κατανομής είναι η:

$$F_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-\xi)^2}{2\sigma^2}} d\xi \quad (2.18)$$

Η μέση τιμή και η τυπική απόκλιση της κατανομής είναι  $\mu$  και  $\sigma$ , αντίστοιχα. Η κατανομή είναι συμμετρική και επομένως η τρίτη κεντρική ροπή είναι μηδέν (άρα  $C_s = 0$ ). Η τέταρτη κεντρική ροπή της είναι  $3\sigma^4$  (άρα  $C_k = 3$ ).



Σχήμα 1 Δύο παραδείγματα κανονικής πυκνότητας πιθανότητας (α)  $N(0,1)$  και (β)  $N(2,2)$ .

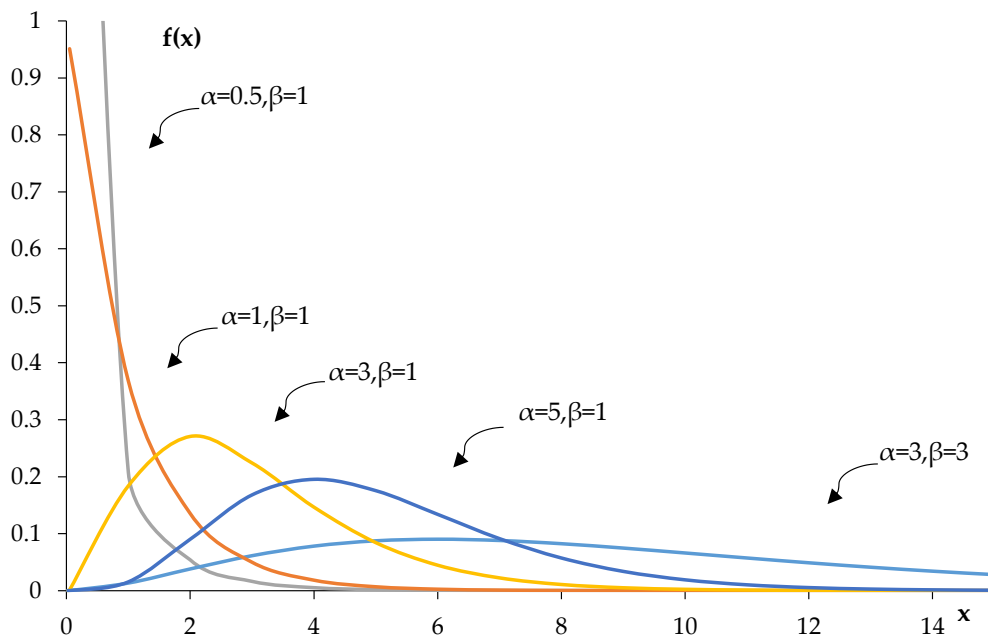


### 2.1.5 Η κατανομή Γάμμα

Λέμε ότι η μεταβλητή  $X$  ακολουθεί την κατανομή Γάμμα και θα συμβολίζεται με  $G(\alpha, \beta)$  όπου  $\alpha, \beta$  θετικές παράμετροι αν η πυκνότητα πιθανότητας της είναι η συνάρτηση:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0 \quad (2.19)$$

Η πρώτη παράμετρος  $\alpha$  είναι γνωστή ως παράμετρος μορφής (shape) και η δεύτερη  $\beta$  ως παράμετρος κλίμακας (scale).



Σχήμα 2 Κατανομές Γάμμα για διάφορες τιμές των παραμέτρων  $\alpha, \beta$ .

### 2.1.6 Η κατανομή $\chi^2$

Λέμε ότι η μεταβλητή  $X$  ακολουθεί την κατανομή  $\chi^2$  με  $n$  βαθμούς ελευθερίας (συμβολικά  $\chi^2(n)$ ) αν η πυκνότητα πιθανότητας της είναι η συνάρτηση

$$f_x(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2} \quad x \geq 0, n = 1, 2, \dots \quad (2.20)$$

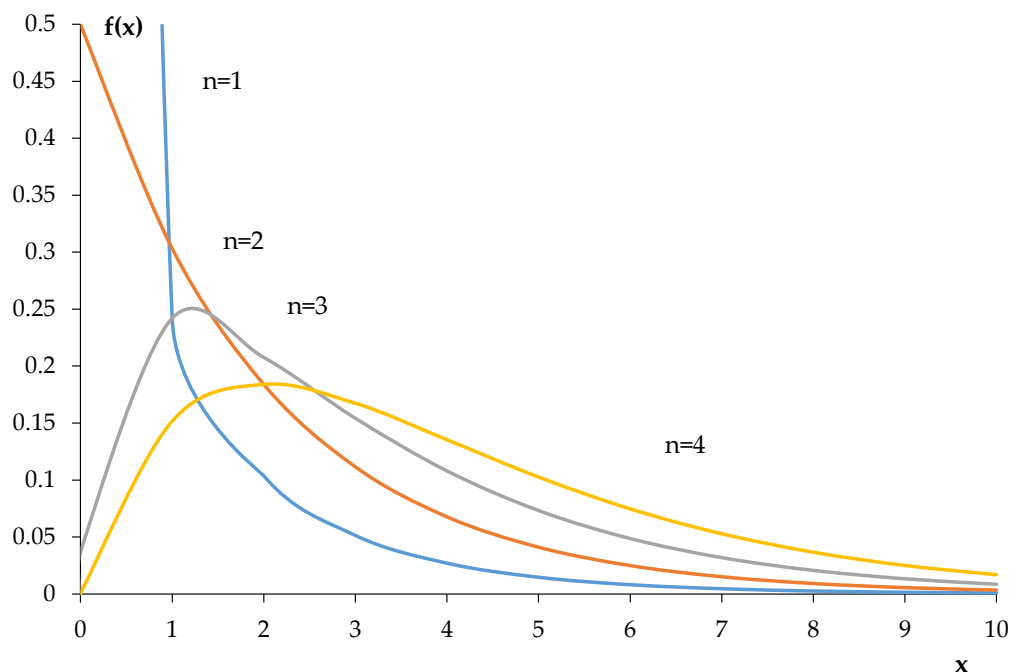
όπου  $\Gamma()$  η συνάρτηση γάμμα.

Πρόκειται για θετικά ασύμμετρη κατανομή μιας παραμέτρου ( $n$ ), η οποία αποτελεί ειδικότερη περίπτωση της κατανομής γάμα. Η μέση τιμή και η διασπορά της κατανομής είναι  $n$  και  $2n$ , αντίστοιχα. Οι συντελεστές ασυμμετρίας και κύρτωσης είναι  $C_s = 2\sqrt{2/n}$  και  $C_k = 3 + 12/n$ .

Αν οι μεταβλητές  $X_i(i, 1, \dots, n)$  ακολουθούν κανονική κατανομή  $N(0,1)$  τότε το άθροισμα των τετραγώνων τους

$$Q = \sum_{i=1}^n X_i^2 \quad (2.21)$$

ακολουθεί κατανομή  $\chi^2(n)$ . Με συνδυασμό αυτού του θεωρήματος με το κεντρικό οριακό θεώρημα προκύπτει ότι για μεγάλο  $n$  η  $\chi^2(n)$  τείνει στην κανονική κατανομή.



Σχήμα 3 Κατανομές  $\chi^2(n)$  για διάφορες τιμές του  $n$ .

### 2.1.7 Μέση τιμή

Η μέση τιμή αποτελεί το κέντρο ισορροπίας των δεδομένων. Περιγράφει τη θέση του κέντρου βάρους του σχήματος που ορίζει η συνάρτηση πυκνότητας πιθανότητας με

τον οριζόντιο άξονα. Έστω  $X_1, X_2, \dots, X_n$  οι τιμές των παρατηρήσεων του δείγματος για μια τ.μ.  $X$ . Η δειγματική μέση τιμή συμβολίζεται με  $\bar{X}$  και ορίζεται ως εξής:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.22)$$

Η δειγματική μέση τιμή αποτελεί εκτιμήτρια της μέσης τιμής και η αριθμητική της τιμή είναι:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.23)$$

και ονομάζεται παρατηρημένη ή αριθμητική μέση τιμή ή μέσος όρος του δείγματος. Τα μεγέθη  $\bar{X}$ ,  $\bar{x}$ ,  $m_X$  είναι τρία διαφορετικά μεγέθη και δεν πρέπει να συγχέονται εννοιολογικά. Όμως μπορούμε να βρούμε ότι:

$$E[\bar{X}] = E[X] \quad \text{Var}[\bar{X}] = \frac{\text{Var}[X]}{n} \quad (2.24)$$

και ισχύει ανεξάρτητα από την κατανομή της  $X$ . Έτσι η εκτιμήτρια της μέσης τιμής είναι αμερόληπτη.

### 2.1.8 Διάμεσος

Η διάμεσος ορίζεται ως η κεντρική τιμή όταν διατάξουμε τα δεδομένα σε αύξουσα σειρά και συμβολίζεται με  $x_{0,5}$ . Πρόκειται δηλαδή για την παρατήρηση εκείνη η οποία είναι ταυτόχρονα μεγαλύτερη ή ίση και μικρότερη ή ίση από το 50% των παρατηρήσεων. Ισχύει ότι  $P(X \leq x_{0,5}) = P(X \geq x_{0,5}) = \frac{1}{2}$  (για συνεχείς μεταβλητές).

Η διάμεσος δηλαδή αντιστοιχεί στο σημείο που χωρίζει την καμπύλη πυκνότητας πιθανότητας σε δύο ισοδύναμα τμήματα με εμβαδό  $\frac{1}{2}$ .

### 2.1.9 Πιθανότερη Τιμή

Η δειγματική πιθανότερη (επικρατούσα) τιμή χρησιμοποιείται επίσης για να δηλώσει την κεντρική τάση των δεδομένων κι ορίζεται ως η τιμή που εμφανίζεται με τη μεγαλύτερη συχνότητα μέσα στο δείγμα. Στη διακριτή περίπτωση είναι η τιμή με την μεγαλύτερη συχνότητα ενώ στη συνεχή είναι η κεντρική τιμή της κλάσης που

παρουσιάζει τη μεγαλύτερη συχνότητα. Αν υπάρχουν πάνω από μία τέτοιες τιμές, τότε όλες αυτές θεωρούνται πιθανότερες (επικρατούσες) τιμές. Είναι φανερό πως η πιθανότερη τιμή δεν έχει νόημα όταν το δείγμα δεν αποτελείται από διακεκριμένες επαναλαμβανόμενες τιμές.

Η μέση τιμή, η πιθανότερη τιμή και η διάμεσος ταυτίζονται μόνο αν η κατανομή είναι συμμετρική και μονοκόρυφη.

### 2.1.10 Ασυμμετρία

Η τρίτη κεντρική ροπή χρησιμοποιείται για την περιγραφ'της ασυμμετρίας της κατανομής. Στην περίπτωση που η κατανομή είναι μονοκόρυφη αυτό που παρατηρούμε είναι αν έχει συμμετρική μορφή ή έχει κλίση προς μια μεριά. Αν ο μεγαλύτερος όγκος των δεδομένων είναι στα αριστερά και η «ουρά» της κατανομής στα δεξιά τότε λέμε ότι η κατανομή είναι θετικά ασύμμετρη και στην αντίθετη περίπτωση αρνητικά ασύμμετρη. Όταν η κατανομή είναι θετικά ασύμμετρη ισχύει ότι  $x_p \leq x_{0.5} \leq m_x$  ενώ η ανάστροφη ανισότητα ισχύει για αρνητικά ασύμμετρη κατανομή. Μηδενική τιμή της τρίτης ροπής δείχνει συμμετρική κατανομή (όπως στην περίπτωση κανονικής κατανομής).

### 2.1.11 Κύρτωση

Ο όρος κύρτωση περιγράφει το πόσο «αιχμηρή» ή όχι είναι η συνάρτηση πυκνότητας πιθανότητας γύρω από την κορυφή της. Η άυξηση της κύρτωσης συνδέεται με την κίνηση της μάζας της συνάρτησης πυκνότητας πιθανότητας στην κορυφή της κατανομής. Για μεγάλες τιμές της κύρτωσης η κατανομή χαρακτηρίζεται ως λεπτόκυρτη (αιχμηρή) ενώ για μικρές τιμές πλατύκυρτη (επίπεδη).

### 2.1.12 Διασπορά (και τυπική απόκλιση)

Η διασπορά μιας τυχαίας μεταβλητής δείχνει το μέγεθος της συγκέντρωσης της πυκνότητας πιθανότητας γύρω από την μέση τιμή. Όσο μικρότερες οι τιμές της διασποράς, τόσο πιο συγκεντρωμένες γύρω από τη  $\mu_x$  βρίσκονται οι τιμές της μεταβλητής. Αν ορίσουμε την απόκλιση μιας παρατήρησης  $X_i$  από την μέση τιμή ως

$X_i - \bar{X}$  τότε το άθροισμα όλων των αποκλίσεων ισούται με 0 από τον ορισμό της δειγματικής μέσης τιμής:

$$\sum_1^n (X_i - \bar{X}) = \sum_1^n X_i - \sum_1^n \bar{X} = n\bar{X} - n\bar{X} = 0 \quad (2.25)$$

Για να μετρήσουμε οπότε την διακύμανση των τιμών γύρω από την μέση τιμή αθροίζουμε όχι τις ίδιες τις αποκλίσεις αλλά τα τετράγωνα των αποκλίσεων και διαιρούμε με το πλήθος  $n$  των παρατηρήσεων (μέση απόκλιση).

Συμβολίζεται με  $\sigma_X^2$  όταν αναφερόμαστε σε πληθυσμό και με  $S_X^2$  όταν αναφερόμαστε σε δείγμα.

$$\sigma_X^2 = \frac{\sum_{i=1}^n (X_i - \mu_X)^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - \mu_X^2 \quad (2.26)$$

Η μεροληπτική εκτιμήτρια της διασποράς  $\sigma_X^2$  του πληθυσμού είναι η ακόλουθη:

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad (2.27)$$

για την οποία αποδεικνύεται ότι (Kendall and Stewart, 1968) ότι:

$$E[S_X^2] = \frac{n-1}{n} \sigma_X^2 \quad (2.28)$$

$$Var[S_X^2] = \frac{\mu_X^{(4)} - \sigma_X^{(4)}}{n} - \frac{2(\mu_X^{(4)} - 2\sigma_X^{(4)})}{n^2} + \frac{\mu_X^{(4)} - 3\sigma_X^{(4)}}{n^3} \quad (2.29)$$

Όπου  $\mu_X^{(4)}$  η τέταρτη κεντρική ροπή του πληθυσμού.

Η αμερόληπτη (και συνεπής) εκτιμήτρια της  $\sigma_X^2$  είναι η ακόλουθη, γνωστή ως δειγματική διασπορά:

$$S_X^{*2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (2.30)$$

ισοδύναμα:

$$S_X^{*2} = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \quad (2.31)$$

Στην περίπτωση κανονικής κατανομής του πληθυσμού αποδεικνύεται ότι:

$$\text{Var}[S_X^2] = \frac{2(n-1)\sigma_x^4}{n^2} \qquad \text{Var}[S_X^{*2}] = \frac{2\sigma_x^4}{(n-1)} \qquad (2.32)$$

Η τυπική απόκλιση είναι η θετική ρίζα της διασποράς και εκφράζει την τυπική απόκλιση των δεδομένων από την μέση τιμή, δηλαδή μέχρι πόσο περιμένουμε μια τιμή της  $X$  να απέχει από την μέση τιμή και έχει τις ίδιες μονάδες μέτρησης με αυτές της μεταβλητής.

Οι  $n$  'ελεύθερες' παρατηρήσεις αποτελούν τους  $n$  βαθμούς ελευθερίας. Για τον υπολογισμό της  $\bar{x}$  σχηματίζουμε το μέσο όρο διαιρώντας το άθροισμα των παρατηρήσεων με τους βαθμούς ελευθερίας  $n$  αφού δεν έχουμε καμιά συνθήκη για τις  $n$  παρατηρήσεις που χρησιμοποιούμε. Για τον υπολογισμό της  $S^2$  όμως έχουμε της συνθήκη  $\sum_1^n (X_i - \bar{X}) = 0$ , δηλαδή αν ξέρουμε  $n - 1$  από τις αποκλίσεις μπορούμε να βρούμε αυτήν που απομένει. Άρα για τον υπολογισμό της  $S^2$  οι βαθμοί ελευθερίας είναι  $n - 1$  και γι αυτό διαιρούμε με  $n - 1$ .

### 2.1.13 Ροπογεννήτρια συνάρτηση

Ροπογεννήτρια συνάρτηση μιας τ.μ.  $X$  ονομάζεται η συνάρτηση μεταβλητής  $t$ ,  $M: \mathbb{R} \rightarrow \mathbb{R}$  με τύπο

$$M_x(t) = E[e^{tX}] = \begin{cases} \sum_{x \in X(\Omega)} e^{tX} P\{X = x\}, & X = \text{διακριτή} \\ \int_{-\infty}^{\infty} e^{tX} f_x(x) dx, & X = \text{συνεχής} \end{cases} \qquad (2.33)$$

με την προϋπόθεση ότι  $E(e^{tX}) < +\infty$ .

## 2.2 ΣΤΟΧΑΣΤΙΚΗ ΑΝΕΛΙΞΗ

Η θεωρία των στοχαστικών ανελιξεων είναι μια εφαρμογή της θεωρίας των πιθανοτήτων. Στοχαστική ανέλιξη ονομάζεται μια οικογένεια τυχαίων μεταβλητών  $X_t$ , ή εναλλακτικά  $X(t)$  αν δεν υπάρχει κίνδυνος παρανόησης, όπου  $t$  είναι παράμετρος που παίρνει τιμές από ένα κατάλληλο σύνολο  $T$ . Αν και γενικά το δεικτοσύνηλο  $T$  μπορεί να είναι οποιοδήποτε σύνολο, συχνότατα παριστάνει χρόνο. Σε περίπτωση που το δεικτοσύνηλο αντιστοιχεί σε διακριτές μονάδες χρόνου,  $T = \{0,1,2, \dots\}$ , έχουμε μια ανέλιξη σε διακριτό χρόνο. Αντίστοιχα, αν το δεικτοσύνηλο αντιστοιχεί σε συνεχή χρόνο, δηλαδή  $T = [0, \infty)$ , μιλούμε για ανέλιξη σε συνεχή χρόνο (Samuel Karlin, 1966).

Οι στοχαστικές ανελιξεις χρησιμοποιούνται για να περιγράψουν την χρονική εξέλιξη φυσικών φαινομένων. Οι φυσικές διεργασίες όπως και η μαθηματική τους έκφραση ως στοχαστικές ανελιξεις εξελίσσονται σε συνεχή χρόνο. Ωστόσο, οι παρατηρούμενες χρονοσειρές χαρακτηρίζονται από ένα δειγματικό χρονικό διάστημα  $D$  συνήθως ορισμένο από τον παρατηρητή και τον χρόνο απόκρισης  $\Delta$  της συσκευής παρατήρησης. Οι σταθερές  $D$ ,  $\Delta$  επηρεάζουν την εκτίμηση των στατιστικών παραμέτρων της ανέλιξης σε συνεχή χρόνο. Δύο περιπτώσεις έχουν μελετηθεί όπου  $\Delta \rightarrow 0$ , και  $D = \Delta$  (Koutsoyiannis, 2013a) με τελικό συμπέρασμα ότι οι διαφορές τους είναι μικρές οπότε θεωρούμε  $D = \Delta > 0$  που έχει και πρακτική σημασία για δείγματα με μικρό  $D$ . Η μαθηματική σχέση για την στοχαστική ανέλιξη σε διακριτό χρόνο  $\underline{x}_i^{(D)}$  για  $D = \Delta > 0$  είναι η εξής:

$$\underline{x}_i^{(D)} = \frac{\int_{(i-1)\Delta}^i \underline{x}(\xi) d\xi}{\Delta} \quad (2.34)$$

όπου  $i \in (1, n)$  είναι μια παράμετρος που αντιπροσωπεύει τον διακριτό χρόνο,  $n = [T/\Delta]$  είναι ο συνολικός αριθμός των παρατηρήσεων και  $T \in [0, \infty)$  είναι ο χρονικός (ορίζοντας) των παρατηρήσεων.

Η στοχαστική ανέλιξη χαρακτηρίζεται από ένα σύνολο στατιστικών παραμέτρων με κυριότερες τις εξής:

- Μέση τιμή:

$$\mu := E[X_i] \quad (2.35)$$

- Διασπορά:

$$Var[X_i] := E[(X_i - \mu)^2] \quad (2.36)$$

- Αυτοσυνδιασπορά της ανέλιξης:

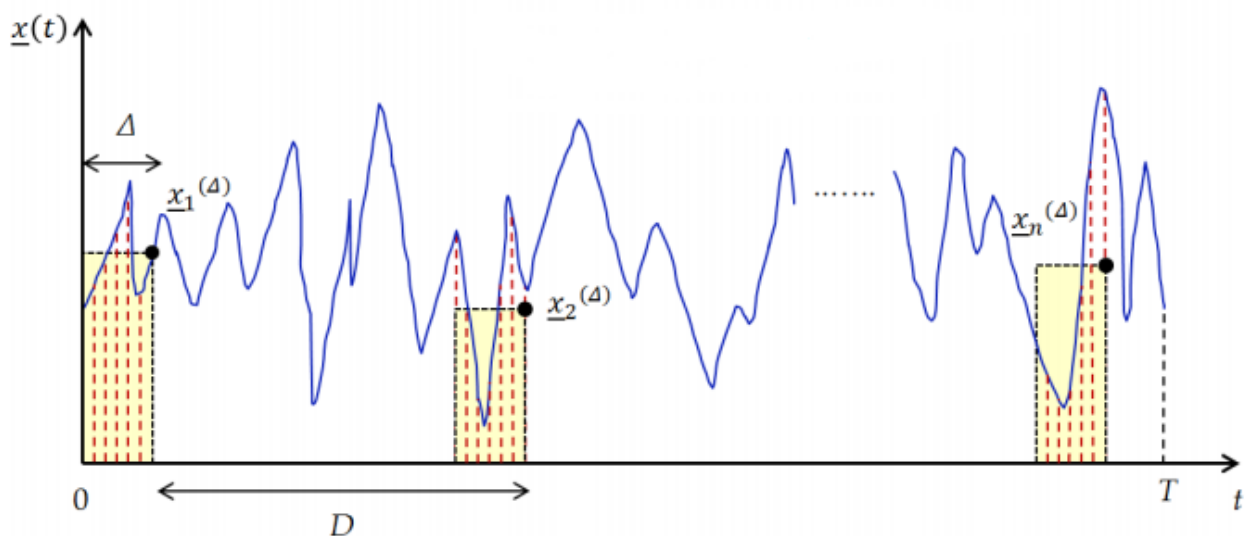
$$\gamma_j = Cov[X_i, X_{i+j}] = E[(X_i - \mu)(X_{i+j} - \mu)] \quad (2.37)$$

- Συντελεστής αυτοσυσχέτισης:

$$\rho(i, j) = \frac{Cov[X_i, X_{i+j}]}{\sqrt{Var[X_i]Var[X_{i+j}]}} \quad (2.38)$$

Ο συντελεστής αυτοσυσχέτισης παίρνει τιμές στο διάστημα  $[-1,1]$  (αρνητική έως θετική συσχέτιση) με την τιμή μηδέν να δηλώνει ασυσχέτιστα δεδομένα. Ο συντελεστής αυτοσυσχέτισης δεν εξαρτάται από τον χρόνο  $t$  αλλά από την υστέρηση (lag). Στην ανάλυση χρονοσειρών είναι σημαντικός ο συντελεστής αυτός γιατί δείχνει τόσο τον βαθμό όσο και το μήκος ή την χρονική διάρκεια της μνήμης της στοχαστικής διαδικασίας.

Μια υλοποίηση της στοχαστικής ανέλιξης, δηλαδή ένα σύνολο παρατηρήσεων  $x(t)$  της  $X(t)$ , για μεταβαλλόμενο χρόνο  $t$ , ονομάζεται δειγματοσυνάρτηση ή χρονοσειρά της ανέλιξης.



Σχήμα 4 Ένα παράδειγμα μια ανέλιξης σε συνεχή χρόνο (sampled) σε χρονικά διαστήματα  $D$  για περίοδο  $T$  και με (instrument response time)  $\Delta$  (Dimitriadis and Koutsoyiannis, 2015).



Το μεγαλύτερο πλεονέκτημα της χρήσης στοχαστικών ανελίξεων για την ανάλυση τυχαίων φαινομένων και όχι της κλασικής στατιστικής είναι ότι μπορούμε να λάβουμε υπόψη μας την χρονική και χωρική εξάρτηση μεταξύ των τυχαίων μεταβλητών.

### 2.2.1.1 Στασιμότητα

Μια στοχαστική ανέλιξη λέγεται στάσιμη με την αυστηρή έννοια όταν τα στατιστικά της χαρακτηριστικά είναι αμετάβλητα σε μια αλλαγή των αρχικών συνθηκών. Λέγεται δε στάσιμη με την ευρεία (ή ελαστική) έννοια αν η μέση τιμή της είναι σταθερή και η αυτοσυνδιασπορά της εξαρτάται μόνο από τη διαφορά του χρόνου (Paroulis, 1965), δηλαδή αν

$$E[X(t)] = \mu = \text{σταθερά} \quad E[(X(t) - \mu)(X(t + \tau) - \mu)] = C(\tau) \quad (2.39)$$

### 2.2.1.2 Εργοδικότητα

Μια στάσιμη στοχαστική ανέλιξη είναι εργοδική αν κάθε παράμετρος της κατανομής μπορεί να προσδιοριστεί από μια απλή δειγματοσυνάρτηση της ανελίξης. Μια ανέλιξη λοιπόν είναι εργοδική αν οι χρονικοί μέσοι είναι ίσοι με τους συνολικούς μέσους (δηλαδή τις αναμενόμενες τιμές). Αν ισχύει:

$$E[X(t)] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^N X(t) \text{ για ανέλιξη διακριτού χρόνου} \quad (2.40)$$

$$E[X(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) dt \text{ για ανέλιξη συνεχούς χρόνου} \quad (2.41)$$

τότε λέμε ότι η ανέλιξη είναι εργοδική ως προς την μέση τιμή.

Η εργοδικότητα είναι πολύ σημαντική ιδιότητα. Μόνο όταν μια στοχαστική ανέλιξη είναι εργοδική μπορούμε να κάνουμε μετρήσεις για να συμπεράνουμε τα στατιστικά χαρακτηριστικά της. Αν δεν είναι εργοδική τότε ανεξάρτητα από το πόσα δεδομένα έχουμε δεν μπορούμε να υπολογίσουμε τα στατιστικά χαρακτηριστικά της.

### 2.2.2 Ανέλιξη στη συναθροισμένη κλίμακα

Η συναθροισμένη ανέλιξη στην χρονική κλίμακα  $k$ , όπου  $k$  θετικός ακέραιος ορίζεται ως:

$$\underline{z}_i^{(k)} := \sum_{l=(i-1)k+1}^{ik} \underline{x}_l \quad (2.42)$$

προφανώς για  $k = 1$ ,  $\underline{z}_i^{(1)} \equiv \underline{x}_i$ , για  $k = 2$ ,  $\underline{z}_1^{(2)} := \underline{x}_1 + \underline{x}_2$ ,  $\underline{z}_2^{(2)} := \underline{x}_3 + \underline{x}_4$ ,  $\underline{z}_3^{(2)} := \underline{x}_5 + \underline{x}_6$  κ.ο.κ..

Οι κυριότερες στατιστικές παράμετροι της συναθροισμένης ανέλιξης  $\underline{z}_i^{(k)}$  αντίστοιχες με αυτές τις  $X_i$  είναι οι εξής:

- Μέση τιμή:

$$E(\underline{z}_i^{(k)}) = k\mu \quad (2.43)$$

- Αυτοσυνδιασπορά:

$$\gamma_j^{(k)} := Cov[\underline{z}_i^{(k)}, \underline{z}_{i+j}^{(k)}] = \sum_{l=1}^k \sum_{m=jk+1}^{(j+1)k} \gamma_{m-l} \quad j = 0, 1, 2 \quad (2.44)$$

- Αυτοσυσχέτιση:

$$\rho_j^{(k)} := Corr[\underline{z}_i^{(k)}, \underline{z}_{i+j}^{(k)}] = \frac{\gamma_j^{(k)}}{\gamma_0^{(k)}} \quad j = 0, 1, 2 \quad (2.45)$$

## 2.2.3 Απλά μοντέλα αναφοράς

### 2.2.3.1 Λευκός θόρυβος

Το μοντέλο του λευκού θορύβου είναι μια ακολουθία ανεξάρτητων τυχαίων και ομοιόμορφα κατανομημένων μεταβλητών έτσι ώστε  $\gamma_j = 0$  (και  $\rho_j = 0$ ) για  $j \neq 0$ . Η συναθροισμένη ανάλυση έχει τα εξής χαρακτηριστικά:

- Μέση τιμή:

$$E [Z_i^{(k)}] = kE[x_i] \quad (2.46)$$

- Διασπορά:

$$\gamma_0^{(k)} := Var[Z_i^{(k)}] = k\gamma_0 \quad (2.47)$$

- Αυτοσυνδιασπορά-Αυτοσυσχέτιση

$$\gamma_j^{(k)} := Cov [Z_i^{(k)}, Z_{i+j}^{(k)}] = 0, \quad \rho_j^{(k)} := Corr [Z_i^{(k)}, Z_{i+j}^{(k)}] = 0 \quad (2.48)$$

### 2.2.3.2 Γραμμικό Μαρκοβιανό μοντέλο AR(1)

Μια ανάλυση  $x(t)$  στην οποία, αν είναι γνωστό το παρόν, το μέλλον δεν εξαρτάται από το παρελθόν αλλά μόνο από το παρόν, λέγεται ανάλυση Markov.

Το μοντέλο αυτοπαλινδρόμησης πρώτης τάξης (autoregressive model of order 1) AR1 ή αλλιώς αλυσίδες Markov ονομάστηκε έτσι από τον Ρώσο μαθηματικό Andrey Markov που το ανακάλυψε τις αρχές του 20<sup>ου</sup> αιώνα. Το μοντέλο αυτό έχει το χαρακτηριστικό της «απώλειας μνήμης» και διατηρεί κάποια εξάρτηση μεταξύ δυο διαδοχικών τιμών. Η σχέση ορισμού του μοντέλου στην βασική κλίμακα θεωρώντας στάσιμη στοχαστική ανάλυση  $X_i$  είναι η εξής:

$$X_i = \alpha X_{i-1} + V_i \quad (2.49)$$

Αποδεικνύεται ότι:

$$\begin{aligned}
 X'_i &= \alpha X'_{i-1} + V'_i \\
 X'_{i-1} X'_i &= \alpha X'^2_{i-1} + X'_{i-1} V'_i \\
 E[X'_{i-1} X'_i] &= \alpha E[X'^2_{i-1}] + E[X'_{i-1} V'_i] \\
 \alpha &= \frac{Cov[X_{i-1}, X_i]}{Var[X_i]} = \rho
 \end{aligned} \tag{2.50}$$

όπου  $\rho$  ο συντελεστής αυτοσυσχέτισης πρώτης τάξης και  $V_i$  η ακολουθία λευκού θορύβου. Ο λευκός θόρυβος έχει:

- Μέση τιμή:

$$\begin{aligned}
 E[X_i] &= E[\alpha X_{i-1}] + E[V_i] \\
 E[X_i] &= \alpha E[X_{i-1}] + E[V_i] \\
 \mu_x &= \alpha \mu_x + \mu_v \\
 \mu_v &= (1 - \alpha) \mu_x
 \end{aligned} \tag{2.51}$$

- Διασπορά:

$$\begin{aligned}
 X'_i &= \alpha X'_{i-1} + V'_i \\
 X_i'^2 &= (\alpha X'_{i-1} + V'_i)^2 \\
 X_i'^2 &= \alpha^2 X'^2_{i-1} + 2\alpha X'_{i-1} V'_i + V_i'^2 \\
 E[X_i'^2] &= \alpha^2 E[X'^2_{i-1}] + 2\alpha E[X'_{i-1} V'_i] + E[V_i'^2] \\
 Var[X_i] &= \alpha^2 Var[X_i] + Var[V_i] \\
 \sigma_v^2 &= (1 - \alpha)^2 Var[X_i] \\
 \sigma_v^2 &= (1 - \alpha)^2 \gamma_0
 \end{aligned} \tag{2.52}$$

Η αυτοσυσχέτιση της  $X_i$  είναι:

$$\rho_j := Corr[X_i, X_{i+j}] = \rho^{|j|} \tag{2.53}$$

Η ανέλιξη είναι Markovian επειδή η εξάρτηση της μεταβλητής  $X_i$  από την μεταβλητή  $X_{i-1}$  αρκεί για να εκφραστεί πλήρως η εξάρτηση του παρόντος από το παρελθόν. Όσο αυξάνει η κλίμακα συναθροίσεως  $k$  τόσο η ανέλιξη τείνει σε λευκό θόρυβο.

Οι κυριότερες στατιστικές παράμετροι της συναθροισμένης ανέλιξης  $z_i^{(k)}$  είναι (Koutsoyiannis, 2002):

- Μέση τιμή:

$$E[z_i^{(k)}] = kE[x_i] \quad (2.54)$$

- Διασπορά:

$$\gamma_0^{(k)} := \gamma_0 \frac{k(1 - \rho^2) - 2\rho(1 - \rho^k)}{(1 - \rho)^2} \quad (2.55)$$

- Αυτοσυνδιασπορά:

$$\gamma_j^{(k)} := \gamma_0 \frac{\rho^{kj-k-1}(1 - \rho^k)^2}{(1 - \rho)^2} \quad (2.56)$$

- Αυτοσυσχέτιση:

$$\rho_j^{(k)} = \rho_1^{(k)} \rho^{k(j-1)} \text{ όπου } \rho_1^{(k)} = \frac{\rho(1 - \rho^k)^2}{k(1 - \rho^2) - 2\rho(1 - \rho^k)} \quad (2.57)$$

Συγκρίνοντας τις σχέσεις για την αυτοσυσχέτιση παρατηρούμε ότι η απλή ανέλιξη AR(1) είναι ανέλιξη Markov μόνο στην βασική της κλίμακα ενώ η συναθροισμένη ανέλιξη αποτελεί μια πιο πολύπλοκη δομή αυτοσυσχέτισης.

Για μεγάλες τιμές της κλίμακας  $k$  ο αριθμητής του κλάσματος της σχέσης της διασποράς επηρεάζεται από τον πρώτο όρο περισσότερο με αποτέλεσμα η διασπορά της συναθροισμένης ανέλιξης να είναι:

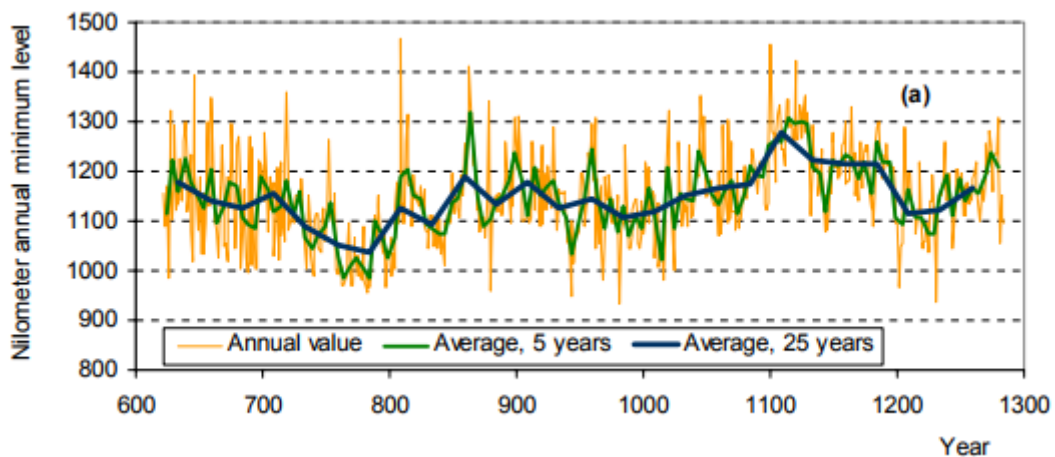
$$\gamma_0^{(k)} \approx k \frac{1 + \rho}{1 - \rho} \gamma_0 \quad (2.58)$$

και είναι πλέον ανάλογη με την κλίμακα  $k$  όπως και στον λευκό θόρυβο. Επίσης μικραίνει η τιμή  $\rho_j^{(k)}$  όπως και η  $\rho_1^{(k)}$ .

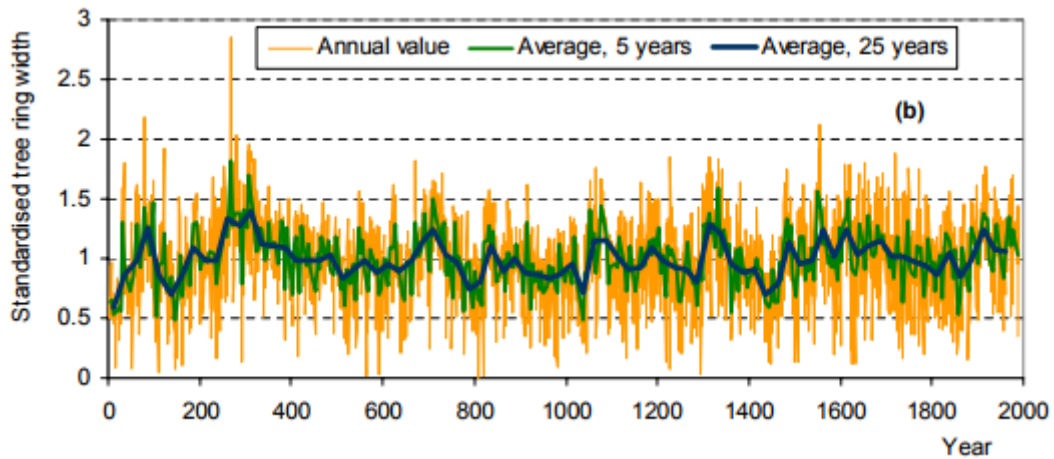
Το μοντέλο AR(1) αδυνατεί να διατηρήσει την δομή αυτοσυσχέτισης των γεωφυσικών φαινομένων επειδή η αυτοσυσχέτιση φθίνει εκθετικά και συγκλίνει στο μηδέν όσο αυξάνεται η κλίμακα συνάθροισης.

#### 2.2.4 Παραδείγματα αποκλίσεων από τα απλά μοντέλα αναφοράς

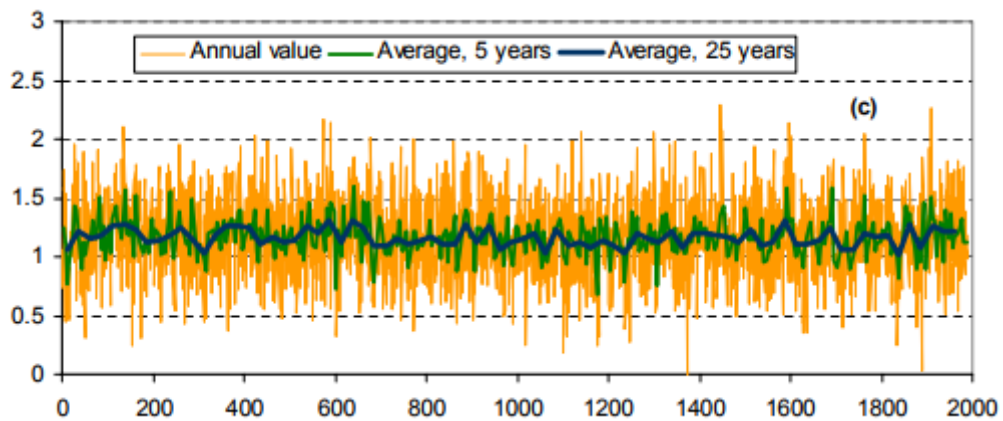
Υπάρχουν πλέον αποδείξεις ότι οι χρονοσειρές υδρολογικών φαινομένων παρουσιάζουν μια συμπεριφορά διαφορετική από αυτή που περιγράφουν τα απλά μοντέλα αναφοράς. Για να το αποδείξουμε αυτό αναφέρονται δυο πραγματικά παραδείγματα. Το πρώτο είναι η ετήσια χρονοσειρά της ελάχιστης στάθμης νερού του ποταμού Νείλου για τα χρόνια 622-1284 μ.Χ. (663 παρατηρήσεις), μια χρονοσειρά που έχει μελετηθεί από πολλούς επιστήμονες και οδήγησε και στην ανακάλυψη του φαινομένου Hurst. Το δεύτερο παράδειγμα έχει ακόμα μεγαλύτερο μήκος δείγματος: η χρονοσειρά των τυποποιημένων δεντροδακτυλίων της περιοχής Mammoth στην Γιούτα που αφορά τα χρόνια 1-1989 μ.Χ. (1990 τιμές).



Σχήμα 5 Διάγραμμα ετήσιας χρονοσειράς της στάθμης του νερού στον Νείλο ποταμό.

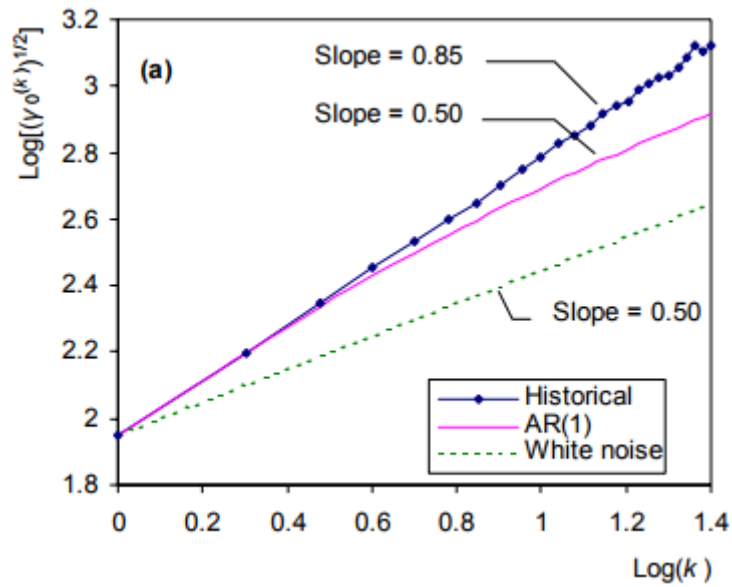


Σχήμα 6 Διάγραμμα χρονοσειράς τυποποιημένων δεντροδακτυλίων της περιοχής Mammoth.

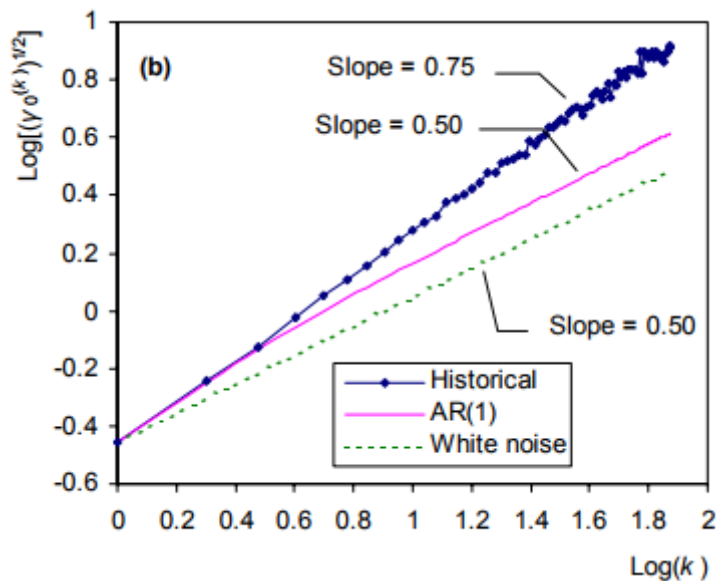


Σχήμα 7 Διάγραμμα λευκού θορύβου με τα ίδια στατιστικά χαρακτηριστικά της χρονοσειράς των δεντροδακτυλίων για λόγους σύγκρισης.

Στα σχήματα 5 και 6 παρουσιάζονται τα δεδομένα για τα δύο παραδείγματα. Επίσης φαίνονται και οι μέσοι 5-ετίας και 25-ετίας δηλαδή η συναθροισμένη ανέλιξη για την κλίμακα  $k = 5$  και  $k = 25$  αντιστοίχα. Στο σχήμα 7 παρουσιάζεται μια χρονοσειρά λευκού θορύβου με τα ίδια στατιστικά χαρακτηριστικά με αυτήν των δεντροδακτυλίων. Παρατηρείται ότι οι διακυμάνσεις της συναθροισμένης ανέλιξης ειδικά για την κλίμακα  $k = 25$  είναι πιο εμφανείς στα ιστορικά δεδομένα παρά στον λευκό θόρυβο. Αυτές οι διακυμάνσεις μπορεί να είναι μη-στασιμότητες, δηλαδή ντετερμινιστική τάση ανόδου ή καθόδου που διαρκεί 100-200 χρόνια. Μπορούμε να συμπεραίνουμε ότι η συνολική εικόνα και για τις δύο περιπτώσεις φαίνεται να αποτελεί μέρος μιας μεγάλης κλίμακας τυχαίων διακυμάνσεων παρά μια ντετερμινιστική τάση.



Σχήμα 8 Διάγραμμα τυπικής απόκλισης προς κλίμακα για την ετήσια χρονοσειρά της στάθμης του νερού στον Νείλο ποταμό.



Σχήμα 9 Διάγραμμα τυπικής απόκλισης προς κλίμακα για την χρονοσειρά τυποποιημένων δεντροδακτυλίων της περιοχής Mammoth (Koutsoyiannis, 2002).

Τα σχήματα 8 και 9 δείχνουν την τυπική απόκλιση της συναθροισμένης κλίμακας συναρτήσεως της χρονικής κλίμακας για τα δύο παραδείγματα (σε λογαριθμικούς άξονες). Για λόγους σύγκρισης παρουσιάζονται και οι θεωρητικές καμπύλες του λευκού θορύβου και του μοντέλου AR(1). Παρατηρείται ότι οι καμπύλες των δύο



παραδειγμάτων είναι σχεδόν ευθείες με κλίση 0.75-0.85. Το μοντέλο λευκού θορύβου και το AR(1) έχουν κλίση ίση με 0.5, και παρεκκλίνουν έντονα από την ιστορική χρονοσειρά.

### 2.2.5 Μακροπρόθεσμη εμμονή

Ο Harold Edwin Hurst (1 Ιανουαρίου 1880 – 7 Δεκεμβρίου 1978) ήταν Βρετανός υδρολόγος από το Leicester. Μελετώντας την διακύμανση της στάθμης του νερού του ποταμού Νείλου, εισήγαγε την έννοια της μακροπρόθεσμης εμμονής, (1951). Η ανάλυση της ετήσιας στάθμης του Νείλου για το διάστημα 622-1284 μ.Χ. (663 δεδομένα), μετρημένης στο «Νειλόμετρο» στο νησί Ρόντα στο Κάιρο (Toussoun, 1925) οδήγησε στην ανακάλυψη του ομώνυμου φαινομένου.

Η έννοια της μακροπρόθεσμης εμμονής έχει κινήσει το ενδιαφέρον πολλών από τότε που ο B.Mandelbrot την ανέφερε σε πολλές επιστημονικές δημοσιεύσεις του Mandelbrot (1965), Mandelbrot and Van Ness (1968), and Mandelbrot and Wallis (1968, 1969)) που, μεταξύ άλλων, εξήγησε και το φαινόμενο Hurst. Όπως χαρακτηριστικά αναφέρεται και στην Βίβλο «ίδου έπτα έτη έρχεται εϋθηνία πολλή εν πάση γῆ Αιγύπτου· ήξει δέ έπτα έτη λιμοϋ μετα ταϋτα» (Γένεσις 41, 29-30). Μάλιστα ο Mandelbrot ορμώμενος από την βιβλική φράση ονόμασε το φαινόμενο «φαινόμενο του Ιωσήφ» (J.Beran, 1994).

Ο συντελεστής του Hurst χρησιμοποιείται ως μέτρο της μακροπρόθεσμης μνήμης της χρονοσειράς. Αφορά τις αυτοσυσχετίσεις της χρονοσειράς, και τον ρυθμό με τον οποίο αυτές μειώνονται καθώς η υστέρηση μεταξύ ζεύγη τιμών αυξάνεται. Ποσοτικοποιεί δηλαδή τη σχετική τάση των χρονοσειρών είτε να συγκλίνουν έντονα στη μέση τιμή ή να συγκεντρώνονται σε μια κατεύθυνση.

Ο συντελεστής Hurst μπορεί να οριστεί ως (J.Beran, 2013):

$$H := 1 + \frac{1}{2} \lim_{m \rightarrow \infty} \gamma_m^\# \quad (2.59)$$

Οι τιμές Hurst  $0.5 < H \leq 1$  υποδεικνύουν μια χρονοσειρά με μακροπρόθεσμη θετική αυτοσυσχέτιση, που σημαίνει ότι μια υψηλή τιμή κατά πάσα πιθανότητα θα ακολουθήσει άλλη μεγάλη τιμή και ότι οι τιμές για μεγάλο χρονικό διάστημα στο

μέλλον θα τείνουν επίσης να είναι υψηλές. Παραδείγματα στην φύση μπορούν να βρεθούν πολλά όπως πχ. τα ξηρά και τα υγρά χρόνια μιας χρονοσειράς 30 ετών.

Μια τιμή  $0 < H < 0.5$  υποδεικνύει μια χρονοσειρά με μακροχρόνια εναλλαγή μεταξύ υψηλών και χαμηλών τιμών σε γειτονικά ζεύγη, που σημαίνει ότι μια υψηλή τιμή μάλλον θα ακολουθείται από μία χαμηλή τιμή και ότι η επόμενη τιμή μετά από αυτό θα τείνει να είναι υψηλή, με αυτή την τάση για εναλλαγή μεταξύ υψηλών και χαμηλών τιμών να διαρκεί για μεγάλο χρονικό διάστημα στο μέλλον. Η λεγόμενη αντεμμονή δεν έχει φυσική σημασία. Έχει παρατηρηθεί στο φυσικό φαινόμενο El Niño και στις ηλιακές κυλίδες.

Μια τιμή  $H = 0.5$  δείχνει ότι οι τιμές που εξετάζονται είναι χρονικά ανεξάρτητες και ότι υπάρχει τυχαία αύξηση ή μείωση των τιμών της χρονοσειράς (γνωστό ως λευκός θόρυβος).

### 2.2.5.1 FGN

Ο Mandelbrot και οι συνεργάτες του (1968) ανέπτυξαν το μοντέλο του κλασματικού Γκαουσιανού θορύβου, το πρώτο ουσιαστικά στατιστικό μοντέλο που αναπαρήγαγε μακροπρόθεσμη εμμονή. Το φαινόμενο της εμμονής ήταν ήδη γνωστό πρωτού περιγραφεί μαθηματικά. Είχε παρατηρηθεί ότι οι αυτοσυσχετίσεις γεγονότων απομακρυσμένων στον χρόνο φθίνουν προς το μηδέν με πολύ αργότερο ρυθμό από αυτόν που θα περιμέναμε σύμφωνα με τη δομή Markov. Σήμερα η έννοια αυτή είναι αρκετά σημαντική αφού χρησιμοποιείται σε ένα εύρος εφαρμογών σε πεδία όπως τα οικονομικά, τα δίκτυα επικοινωνιών, τις κλιματικές επιστήμες κ.α. Η αποτυχία των απλών μοντέλων αναφοράς να περιγράψουν την φυσική πραγματικότητα (AR1, λευκός θόρυβος) λόγω της τάσης ομαδοποίησης χαμηλών (ή ψηλών) τιμών σε μεγαλύτερες περιόδους ξηρασιών (ή υδρολογικά πλούσιων) ετών, αφού είχε παρατηρηθεί σημαντική αυτοσυσχέτιση ακόμα και για μεγάλες τιμές της υστέρησης (εξάρτηση μεγάλης εμβέλειας), και για μεγάλες τιμές της κλίμακας συνάθροισης οδήγησε στις ανελιξεις απλής ομοιοθεσίας (ή στάσιμα αυξήματα αυτό-όμοιων ανελιξεων).

Η ανέλιξη  $X_n = B_H(n + 1) - B_H(n), n \geq 0$  του κλασματικού θορύβου Brown (FBM) ονομάζεται κλασματικός Γκαουσιανός θόρυβος (FGN) και είναι στάσιμη ανέλιξη.

Αποδεικνύεται ότι:

$$\rho_n = \text{Corr}(X_0, X_n) \sim 2H(2H - 1)n^{-2(1-H)} \quad (2.60)$$

(Samorodnitsky, 2007).

Για κάθε  $0 < H \leq 1$  υπάρχει μια ακριβώς γκαουσιανή ακολουθία  $X_i$  (με μέση τιμή 0) η οποία είναι στάσιμη ανέλιξη της ομοιόθετης διαδικασίας FBM.

Ο κλασματικός Γκαουσιανός θόρυβος που αναπαράγει την ομοιοθεσία περιγράφεται μαθηματικά ως εξής:

$$\underline{z}_i^{(k)} - k\mu =_d \left(\frac{k}{l}\right)^H (\underline{z}_j^{(l)} - l\mu) \quad (2.61)$$

Όπου το σύμβολο  $=_d$  σημαίνει ισότητα στην (πεπερασμένης διάστασης από κοινού) κατανομή,  $i, j, k$  και  $l$  είναι οποιοιδήποτε ακέραιοι,  $\mu$  είναι η μέση τιμή του  $\underline{z}_i^{(1)}$  και  $H$  είναι μια σταθερά ( $0 < H < 1$ ) γνωστή ως συντελεστής (ή εκθέτης) Hurst.

Αν η  $X_i$  ακολουθεί κανονική (Gaussian) κατανομή τότε η ανέλιξη αυτή ονομάζεται FGN σε διακριτό χρόνο (Mandelbrot, 1965). Μια εναλλακτική ονομασία της ανέλιξης είναι στάσιμα αυξήματα αυτό-όμοιων ανελιξεων (stationary increments of self-similar process) ή αλλιώς ανέλιξη απλής ομοιοθεσίας (simple scaling process).

Η ανέλιξη HK (Hurst-Kolmogorov) ονομάστηκε έτσι από τον Hurst, 1951, ο οποίος ανέλυσε την έννοια της μακροπρόθεσμης εμμονής όταν την παρατήρησε σε χρονοσειρές γεωφυσικών διεργασιών και τον Kolmogorov, 1940, ο οποίος ενώ ασχολιόταν με την τύρβη έκφρασε την μαθηματική μορφή της ανέλιξης.

Βασικές ιδιότητες:

- Διασπορά (συνάρτηση δύναμης της κλίμακας συνάθροισης):

$$\gamma_0^{(k)} := \text{Var}[\underline{z}_i^{(k)}] = k^{2H}\gamma_0 \quad (2.62)$$

- Αυτοσυσχέτιση:

$$\rho_j^{(k)} = \rho_j = (1/2)[(j+1)^{2H} + (j-1)^{2H}] - j^{2H}, j > 0 \quad (2.63)$$

ή προσεγγιστικά (συνάρτηση δύναμης της υστέρησης)

$$\rho_j^{(k)} = \rho_j = H(2H - 1)j^{2H-2} \quad (2.64)$$

Η αυτοσυσχέτιση είναι ανεξάρτητη της κλίμακας  $k$  και σύμφωνα με την σχέση (2.63) είναι συνάρτηση δύναμης της υστέρησης.

Για  $H = 0.5$  η ανέλιξη απλής ομοιοθεσίας είναι πρακτικά μια ακολουθία ανεξάρτητων τυχαίων μεταβλητών.

### 2.2.5.2 Φυσική σημασία

Η θεωρία των στοχαστικών ανελιξεων είναι μια εφαρμογή της θεωρίας των πιθανοτήτων. Μια μεταβλητή μπορεί να χαρακτηριστεί τυχαία με την έννοια ότι δεν υπάρχει προσδιοριστική μέθοδος καθορισμού της τιμής της με πλήρη βεβαιότητα. Όμως, το γεγονός ότι μια φυσική διεργασία περιγράφεται από μια στοχαστική ανέλιξη δεν σημαίνει ότι η πρώτη δεν υπακούει σε κανενός είδους αιτιοκρατία. Η συνύπαρξη της τυχειότητας και της αιτιότητας στα φαινόμενα και η χρονική μεταξύ τους συσχέτιση εξηγεί γιατί η στοχαστική ανέλιξη έχει και στοχαστικές και προσδιοριστικές συνιστώσες. Ειδικότερα, το στοχαστικό μέρος της ανέλιξης δεν είναι πλήρως τυχαίο, αλλά έχει στοχαστική δομή (μνήμη). Αυτό σημαίνει ότι υπάρχει στοχαστική εξάρτηση των τιμών της ανέλιξης σε γειτονικές χρονικές στιγμές (ισχυρός συντελεστής αυτοσυσχέτισης της ανέλιξης). Η έννοια της μακροπρόθεσμης εμμονής χρησιμοποιείται για να υποδειξει την ήπια μείωση της αυτοσυσχέτισης.

Η συσχέτιση σημαίνει αλλαγή. Η αλλαγή (κυρίως ακανόνιστη και απρόβλεπτη σε ντετερμινιστικούς όρους) παράγει την μακροπρόθεσμη εμμονή ενώ η αυτοσυσχέτιση προκύπτει από αυτή (Climatedialogue.org, April 2014). Όπως για όλα τα φυσικά φαινόμενα, σε μικρές κλίμακες είναι προβλέψιμα ενώ όσο μεγαλώνει η κλίμακα δεν είναι (παράθυρο προβλεψιμότητας). Ο συντελεστής Hurst για μικρή κλίμακα που ακόμα δεν θα έχει διαμορφωθεί η ομαδοποίηση μπορεί να είναι πιο προβλέψιμος με μικρότερη αβεβαιότητα (μοιάζει με λευκό θόρυβο), ενώ καθώς μεγαλώνει η κλίμακα, μεγαλώνουν και τα όρια, μεγαλώνει η αυτοσυσχέτιση (εξάρτηση) άρα θα έχουμε και μεγαλύτερη αβεβαιότητα. Σε μικρές κλίμακες δηλαδή δεν διατηρείται η εμμονή.

(Αντιθέτως στο μοντέλο AR1 στις μικρές κλίμακες έχουμε μεγάλη αβεβαιότητα καθώς η αυτοσυσχέτιση πέφτει πολύ απότομα.)

Συγκρίνοντας τα δυο μοντέλα AR(1) και FGN μπορούμε να πούμε ότι χρησιμοποιούν και τα δυο από μια παράμετρο για να εκφράσουν την δομή αυτοσυσχέτισης. Το AR(1) χρησιμοποιεί την αυτοσυσχέτιση  $\rho$  ενώ το FGN την παράμετρο Hurst. Τελικά, συμπεραίνουμε ότι το FGN είναι ένα απλό μοντέλο που προσομοιάζει την πραγματικότητα, πιο αποτελεσματικά από το AR(1) ειδικά όσο αφορά τις υδρολογικές χρονοσειρές.

### 2.2.6 Εκτίμηση στατιστικών παραμέτρων με την υπόθεση ΗΚ ανέλιξης

Η πιο βασική στατιστική παράμετρος είναι η μέση τιμή της χρονοσειράς,  $\bar{X}$ , της οποίας η τυπική εκτιμήτρια είναι:

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i \quad (2.65)$$

η οποία είναι αμερόληπτη ανεξάρτητα από την κατανομή της ανέλιξης  $X_i$ , δηλαδή  $E[\bar{X}] = \mu$ . Στην κλασική στατιστική η διασπορά της είναι:

$$Var[\bar{X}] = \frac{\sigma^2}{n} \quad (2.66)$$

η οποία δεν ισχύει στην περίπτωση της ανέλιξης απλής ομοιοθεσίας.

Παρατηρώντας ότι  $\bar{X} = Z_1^{(n)}/n$  και χρησιμοποιώντας την εξίσωση [γ0], προκύπτει (J.Beran, 1994):

$$Var[\bar{X}] = \frac{\sigma^2}{n^{2-2H}} \quad (2.67)$$

Υπενθυμίζεται ότι η τετραγωνική ρίζα της  $Var[\bar{X}]$  που είναι η τυπική απόκλιση αποτελεί το τυπικό σφάλμα στην εκτίμηση της πραγματική μέσης τιμής των παρατηρήσεων μιας ιστορικής χρονοσειράς. Σε αυτό το σημείο θα πρέπει να σημειωθεί ότι για  $H = 0.5$  και οι δύο εξισώσεις καταλήγουν στο ίδιο τυπικό σφάλμα που είναι αντιστρόφως ανάλογο της ρίζας του μήκους  $n$  της χρονοσειράς. Ωστόσο για μεγάλες τιμές του  $H$  η διαφορά των δύο εξισώσεων γίνεται αρκετά μεγάλη. Για

παράδειγμα σε μια χρονοσειρά με  $n = 100$  χρόνια παρατηρήσεων, σύμφωνα με την κλασική στατιστική το τυπικό σφάλμα θα είναι  $\sigma/10$ . Αντιθέτως για  $H = 0.8$  το ορθό τυπικό σφάλμα σύμφωνα με την σχέση (2.67) θα είναι  $\sigma/2,5$ , δηλαδή 4 φορές μεγαλύτερο. Για να είχαμε τυπικό σφάλμα ίσο με  $\sigma/10$  τότε θα έπρεπε  $n = 100000$  χρόνια. Προφανώς αυτή η σημαντική διαφορά επηρεάζει και άλλα στατιστικά μεγέθη.

Η κλασική εκτιμήτρια της διασποράς:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (2.68)$$

δεν είναι πλέον αμερόληπτη εκτιμήτρια για ανέλιξη απλής ομοιοθεσίας. Έχει αποδειχθεί ότι μια συνεπής εκτιμήτρια της διασποράς για γνωστό  $H$  είναι:

$$\tilde{S}^2 := \frac{n-1}{n-n^{2H-1}} S^2 = \frac{1}{n-n^{2H-1}} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (2.69)$$

για  $H = 0.5$  η εξίσωση (2.68) γίνεται ίση με την (2.69).

Η στατιστική έννοια  $\tilde{S}$  (δηλαδή η τετραγωνική ρίζα της  $\tilde{S}^2$ ) μπορεί να θεωρηθεί ως εκτιμήτρια της τυπικής απόκλισης  $\sigma$ . Από την στιγμή όμως που η τετραγωνική ρίζα (καθώς και οποιοσδήποτε άλλος μη γραμμικός μετασχηματισμός) δεν διατηρεί την έλλειψη μεροληψίας, η  $\tilde{S}$  είναι μεροληπτική. Μια πιο συνεπής εκτιμήτρια είναι η εξής (σχεδόν αμερόληπτη για γνωστό  $H$  και κανονική κατανομή της ανέλιξης  $X_i$ ):

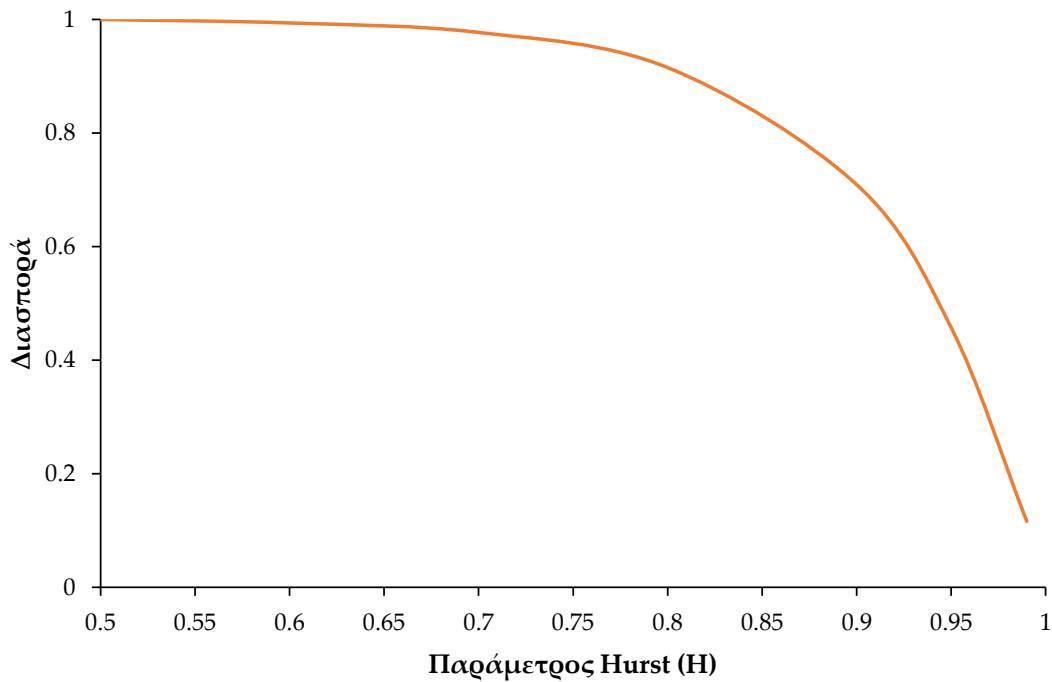
$$\ddot{S} := \sqrt{\frac{n-1/2}{n-n^{2H-1}}} S = \sqrt{\frac{n-1/2}{(n-1)(n-n^{2H-1})}} \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (2.70)$$

(Koutsoyiannis, 2003) αποτέλεσμα από Monte Carlo ανάλυση. Μπορεί εύκολα να επαληθευτεί ότι για  $H = 0.5$ ,  $\ddot{S} = [(n-1/2)/(n-1)]^{0.5} S$ . Αυτή είναι μια πολύ καλή προσέγγιση της πραγματικής αμερόληπτης εκτιμήτριας που για ανεξάρτητες Gaussian ανελίξεις  $X_i$  ισχύει:

$$\ddot{S} = \{[(n-1)/2]^{0.5} \Gamma[(n-1)/2] / \Gamma(n/2)\} S \quad (2.71)$$

Στο σχήμα 10 φαίνεται η προσομοίωση Monte Carlo κατά την οποία παρήχθησαν 10000 συνθετικές χρονοσειρές μήκους δείγματος  $n = 200$  με χρήση του μοντέλου

SMA θεωρώντας  $Var = 1$ . Παρατηρούμε ότι όσο αυξάνεται η τιμή του συντελεστή Hurst τόσο μεγαλώνει η αρνητική μεροληψία στην εκτίμηση της διασποράς.



Σχήμα 10 Διάγραμμα απόκλισης της διασποράς από την θεωρητική ( $\sigma^2 = 1$ ).

Η διασπορά της  $S$  στη κλασική στατιστική είναι:

$$Var[S] \approx \frac{\sigma^2}{2(n-c)} \quad (2.72)$$

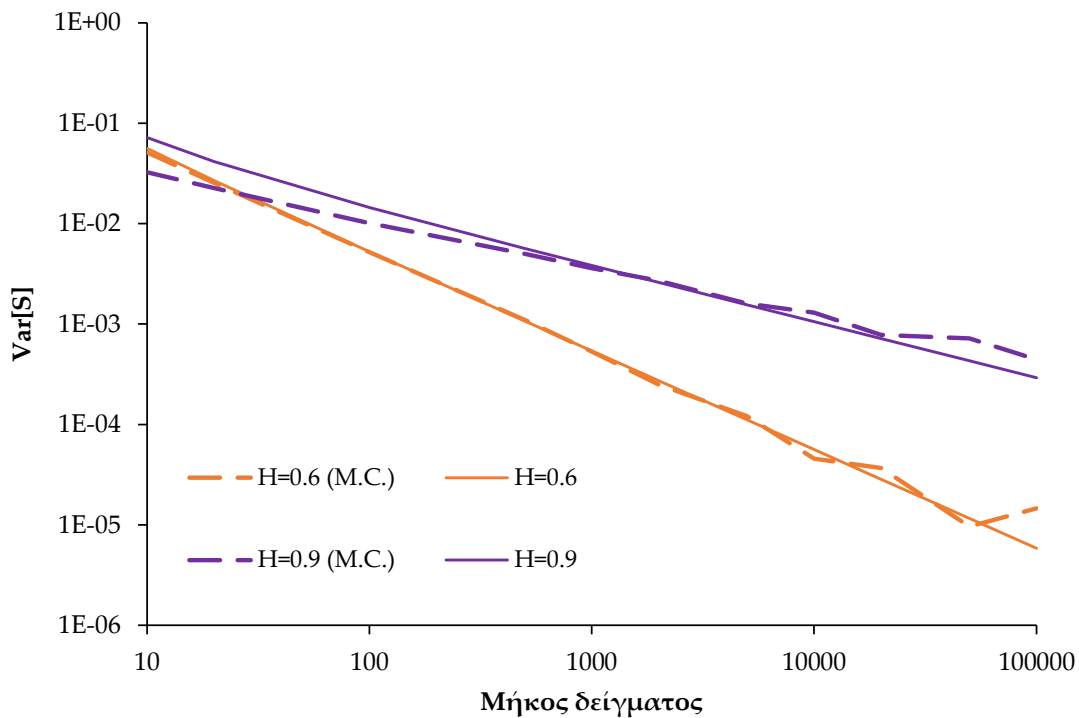
Όπου  $c$  ισούται με 0 ή 1, αλλά ακριβέστερα 0.75.

Η ακριβής διασπορά είναι  $\{1 - [2/(n-1)]\Gamma^2(n/2)/\Gamma^2[(n-1)/2]\}\sigma^2$ . Στην περίπτωση ανελίξεων απλής ομοιοθεσίας με την ίδια Monte Carlo ανάλυση βρέθηκε ότι:

$$Var[\hat{S}] \approx \frac{(0.1n + 0.8)^{\lambda(H)} \sigma^2}{2(n-1)} \quad \text{με } \lambda(H) := 0.088(4H^2 - 1)^2 \quad (2.73)$$

Μπορεί να επαληθευτεί ότι η σχέση (2.73) οδηγεί σε μεγαλύτερη διασπορά από αυτή της σχέσης (2.72) (εκτός από την περίπτωση  $H = 0.5$  όπου οι δυο αυτές σχέσεις ισούνται). Τονίζεται ότι οι εξισώσεις αυτές αφορούν κανονική κατανομή της ανελίξης.

Οι συνέπειες της λάθος εκτίμησης των στατιστικών παραμέτρων είναι σημαντικές. Η σχέση (2.72) οδηγεί σε υποτίμηση της διασποράς, κάτι που δεν γίνεται σοβαρά αντιληπτό στις μικρές κλίμακες αλλά αυξάνεται όσο μεγαλώνει η κλίμακα. Η υποτίμηση της τυπικής απόκλισης που αυξάνεται όσο αυξάνει η κλίμακα, όταν χρησιμοποιείται η κλασική εκτιμήτρια, έχει σημαντικές επιπτώσεις όπως: υποβαθμίζει την ύπαρξη του φαινομένου Hurst για μικρά δείγματα. Συγκεκριμένα σε ένα λογαριθμικό διάγραμμα της τυπικής απόκλισης συναρτήσει της κλίμακας η κλίση της καμπύλης δεν είναι σταθερή, αλλά μειώνεται με την αύξηση της κλίμακας. Έτσι, όχι μόνο υποεκτιμάται ο συντελεστής Hurst αλλά μπορεί να οδηγήσει και στο λανθασμένο συμπέρασμα ότι σε μεγάλες κλίμακες ο συντελεστής πλησιάζει την τιμή 0.5 κάτι που υποδηλώνει απουσία του φαινομένου (Koutsoyiannis, 2003).



Σχήμα 11 Επαλήθευση της σχέσης (2.73) με τα αποτελέσματα από την Monte Carlo ανάλυση για  $H=0.6$  και  $H=0.9$ .



## 2.3 ΜΕΘΟΔΟΛΟΓΙΕΣ

### 2.3.1 SMA

Είναι γνωστό (Box and Jenkins, 1970), ότι για κάθε συνάρτηση αυτοσυσχέτισης  $\gamma_j$ , η  $X_i$  μπορεί να γραφτεί ως το σταθμισμένο άθροισμα απείρου μήκους ανεξάρτητων τυχαίων και ομοιόμορφα καταναμημένων τυχαίων μεταβλητών λευκού θορύβου  $V_i$  γνωστό ως μοντέλο κυλιόμενου μέσου (MA).

Η γέννηση των συνθετικών χρονοσειρών γίνεται με χρήση ενός μοντέλου συμμετρικών κινούμενων μέσων όρων (SMA) (Koutsoyiannis, 2000) και προσεγγίζεται σύμφωνα με την ακόλουθη σχέση η οποία μετασχηματίζει μια ανέλιξη λευκού θορύβου  $\underline{v}_i$  σε μια ανέλιξη με αυτοσυσχέτιση  $\underline{x}_i$  :

$$\underline{X}_i = \sum_{j=-s}^s a_{|j|} V_{i+j} = a_s V_{i-s} + \dots + a_1 V_{i-1} + a_0 V_i + a_1 V_{i+1} + \dots + a_s V_{i+s} \quad (2.74)$$

όπου  $X_i$  η στοχαστική ανέλιξη,  $V_i$  ο λευκός θόρυβος και  $a_j$  οι συντελεστές βάρους όπου ο αριθμός τους  $q$  θεωρητικά είναι άπειρος αλλά στην πράξη λαμβάνει μια πεπερασμένη τιμή για τον λόγο ότι ο αριθμός των παραμέτρων που παράγουμε σε οποιοδήποτε πρόβλημα προσομοίωσης είναι πάντα πεπερασμένος και επειδή ο όρος  $a_j$  μειώνεται καθώς το  $j \rightarrow -\infty$ , έτσι πέρα από κάποιον αριθμό  $j = -s$  οι όροι αυτοί μπορούν να παραληφθούν χωρίς κάποια σοβαρή επίπτωση στην ακρίβεια του μοντέλου. Έτσι έχουμε υποθέσει ότι  $a_j = 0$  για  $|j| > s$ . Οι συντελεστές βάρους πλέον προσδιορίζονται από την ακολουθία των αυτοσυνδιασπορών  $\gamma_j$  ως εξής:

$$\gamma_i = \sum_{j=i-s}^s a_j a_{i-j}, \quad j = s+1, \dots, 2s \quad (2.75)$$

Με αυτό τον τρόπο το μοντέλο αναπαράγει τους  $s+1$  όρους της ακολουθίας αυτοσυνδιασπορών  $\gamma_j$  υποδηλώνοντας μη μηδενική αυτοσυσχέτιση και για τους επόμενους όρους μέχρι  $j = 2s$ . Έτσι αναπαράγεται η εμμονή της ανέλιξης. Η μέθοδος αυτή είναι κατάλληλη για τυχούσα συνάρτηση αυτοσυσχέτισης. Στην περίπτωση του μοντέλου απλής ομοιοθεσίας αποδεικνύεται ότι οι συντελεστές βάρους είναι:

$$a_j \approx \frac{\sqrt{(2-2H)\gamma_0}}{3-2H} [(j+1)^{H+0.5} + (j-1)^{H+0.5} - 2j^{H+0.5}], \quad j > 0 \quad (2.76)$$

και ο αναγκαίος αριθμούς βαρών:

$$q \geq \max \left[ m, \left( \frac{2\beta}{H^2 - 0.25} \right)^{\frac{1}{(H-1.5)}} \right] \quad (2.77)$$

Οι μεταβλητές αυτές σχετίζονται με τα στατιστικά χαρακτηριστικά της ανέλιξης και θεωρούνται ότι είναι στοχαστικά ανεξάρτητες μεταξύ τους, με διασπορά ίση με 1.

Η μέση τιμή  $\mu_v := E[V_i]$  του λευκού θορύβου  $V_i$  (από τον ορισμό του  $Var[V_i] = 1$ ) σχετίζεται με την αντίστοιχη παράμετρο της  $X_i$  ως εξής:

$$\left( a_0 + 2 \sum_{j=1}^s a_j \right) \mu_v = \mu_x \quad (2.78)$$

Η μέθοδος μπορεί να διατηρήσει και την ασυμμετρία  $\xi_x$  της  $X_i$  αν ο λευκός θόρυβος έχει ασυμμετρία  $\xi_v := E[(V_i - \mu_v)^3]$  που δίνεται από την σχέση:

$$\left( a_0^3 + 2 \sum_{j=0}^s a_j^3 \right) \xi_v = \xi_x \gamma_0^{3/2} \quad (2.79)$$

Η εκτίμηση των τιμών του λευκού θορύβου γίνεται μέσω μίας γεννήτριας τυχαίων αριθμών.

### 2.3.2 Κλιμακόγραμμα

Η διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής μπορεί να γίνει με διάφορους τρόπους όπως με την χρήση της αυτοσυνδιασποράς, του φάσματος ισχύος και του κλιμακογράμματος. Το κλιμακόγραμμα (Koutsoyiannis, 2013) είναι ένα στοχαστικό εργαλείο που χρησιμοποιείται για να διακρίνει την μακροπρόθεσμη εμμονή και συσχετίζει την διασπορά (ή τυπική απόκλιση) του μέσου όρου της ανέλιξης (υποθέτοντας στασιμότητα) με τον μέσο όρο της κλίμακας του χρόνου. Με αυτό τον τρόπο μπορούμε να ποσοτικοποιήσουμε τον συντελεστή Hurst ο οποίος ισούται με το μισό της κλίσης του κλιμακογράμματος σε log-log διάγραμμα, με την κλίμακα να τείνει στο άπειρο, προσθέτοντας 1.

Το κλιμακόγραμμα χρησιμοποιείται για την έκφραση της διασποράς στις διάφορες κλίμακες. Η ύπαρξη μεγάλης μεροληψίας στην εκτίμηση της αυτοσυνδιασποράς και

του φάσματος ισχύος, όπως επίσης οι επιπτώσεις της διαδικασίας διακριτοποίησης και η επίδραση του πεπερασμένου μεγέθους δεδομένων, είναι λόγοι που δεν χρησιμοποιούνται αυτές οι τεχνικές για να χτίσεις ένα στοχαστικό μοντέλο. Συγκρίνοντας λοιπόν του τρεις τρόπους φαίνεται πως το κλιμακόγραμμα ως στοχαστικό εργαλείο έχει το μικρότερο μέσο τετραγωνικό σφάλμα (mean squared error), άρα και μικρότερη αβεβαιότητα, και είναι πιο κοντά στην πιο πιθανή τιμή του (Dimitriadis and Koutsoyiannis, 2014).

Κλιμακόγραμμα	
Συνεχής χρόνος	$\gamma(m) := \frac{\text{Var} \left[ \int_t^{t+m} \underline{x}(\xi) d\xi \right]}{m^2} = \frac{\text{Var} \left[ \int_0^m \underline{x}(\xi) d\xi \right]}{m^2}$ όπου $m \in \mathbb{R}^+$
Διακριτός χρόνος	$\gamma_d^{(\Delta)}(k) := \frac{\text{Var} \left[ \sum_{l=k(i-1)+1}^{ki} \underline{x}_l^{(\Delta)} \right]}{k^2} = \frac{\text{Var} \left[ \sum_{l=1}^k \underline{x}_l^{(\Delta)} \right]}{k^2} = \gamma(k\Delta)$ όπου $k \in \mathbb{N}$ είναι η αδιάστατη κλίμακα για μια ανέλιξη σε διακριτό χρόνο
Κλασική εκτιμήτρια	$\underline{\gamma}(k) = \frac{1}{n/k - 1} \sum_{i=1}^n \left( \frac{1}{k} \left( \sum_{l=k(i-1)+1}^{ki} \underline{x}_l^{(\Delta)} \right) - \frac{\sum_{l=1}^n \underline{x}_l^{(\Delta)}}{n/k} \right)^2$
Αναμένουσα τιμή της κλασικής εκτιμήτριας	$E \left[ \underline{\gamma}_d^{(\Delta)}(k) \right] = \frac{1 - \gamma_d^{(\Delta)}(n)/\gamma_d^{(\Delta)}(k)}{1 - k/n} \gamma_d^{(\Delta)}(k)$
ΗΚ ανέλιξη	$\gamma(k) = \frac{\lambda}{k^{2-2H}}$ όπου $\lambda$ είναι η θεωρητική διασπορά της ανέλιξης
Μαρκον ανέλιξη	$\gamma(k) = \frac{2\lambda}{(k/q)^2} (k/q + e^{-k/q} - 1)$ με $\gamma(0) = \lambda$

Πίνακας 1 Σημαντικές σχέσεις κλιμακογράμματος

Υποθέτοντας ότι έχουμε  $n = T/k$  παρατηρήσεις της (μέσης) συναθροισμένης ανέλιξης στην κλίμακα  $k = \kappa D$ ,  $\underline{x}_i^{(k)}$ , όπου  $T := [T_0/k]k$  η περίοδος παρατηρήσεων στρογγυλοποιημένη σε ένα ακέραιο πολλαπλάσιο της κλίμακας  $k$ , με  $[ \ ]$  να δηλώνει την στρογγυλοποίηση και  $T_0$  την περίοδο παρατηρήσεων. Η (αμερόληπτη) εκτιμήτρια της μέσης τιμής της ανέλιξης σε διακριτό χρόνο  $\underline{x}_i^{(k)}$  είναι:

$$\bar{\underline{x}}^{(k)} := \frac{1}{n} \sum_{i=1}^n \underline{x}_i^{(D)} = \frac{\underline{X}(T)}{nD} = \frac{\underline{X}(T)}{T} = \underline{x}_1^{(T)} \quad (2.80)$$

Η κλασική εκτιμήτρια (αλλά μεροληπτική εκτός από τον λευκό θόρυβο)  $\hat{\underline{\gamma}}(k)$  της διασποράς  $\gamma(k)$  της μέσης ανέλιξης (averaged process)  $\underline{x}_i^{(k)}$  είναι:

$$\hat{\gamma}(k) := \frac{1}{n-1} \sum_{i=1}^n (x_i^{(k)} - \bar{x}^{(k)})^2 = \frac{1}{T/D-1} \sum_{i=1}^{T/D} (x_i^{(k)} - \bar{x}^{(k)})^2 \quad (2.81)$$

Η παρακάτω γενική εξίσωση μπορεί να χρησιμοποιηθεί για να εκτιμήσουμε την μεροληψία της  $\hat{\gamma}(D)$

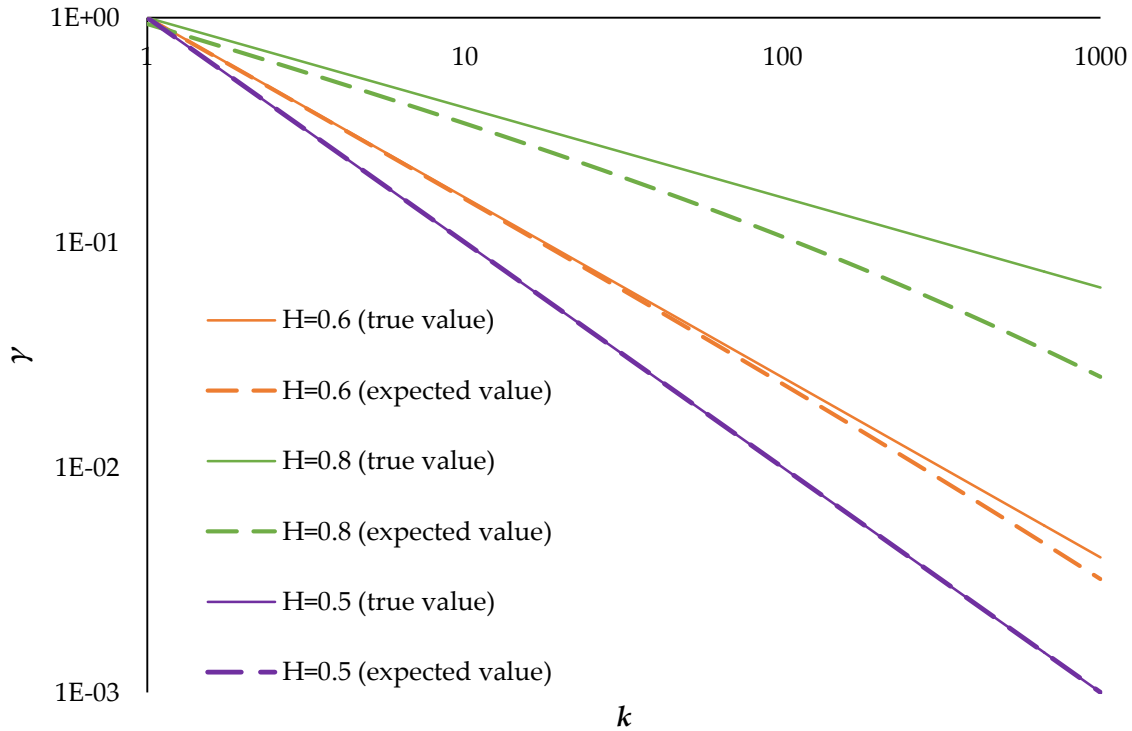
$$E[\hat{\gamma}(k)] = \frac{1}{1-1/n} (Var[x_i^{(k)}] - Var[\bar{x}^{(k)}]) \quad (2.82)$$

$$E[\hat{\gamma}(k)] = \frac{1}{1-k/T} (\gamma(k) - \gamma(T)) \quad (2.83)$$

Ενώ στην κλασική στατιστική ισχύει ότι  $\sigma^{(k)} = \sigma/\sqrt{k}$  στις χρονοσειρές που παρουσιάζουν μακροπρόθεσμη εμμονή παρατηρείται ότι (Koutsoyiannis, 2002):

$$\sigma^{(k)} = \frac{\sigma}{k^{1-H}} \quad (2.84)$$

Γνωρίζουμε από την κλασική στατιστική ότι για μια ανέλιξη ανεξάρτητη στο χρόνο (λευκός θόρυβος) ισχύει  $\sigma^{(k)} = \sigma/\sqrt{k}$  που σημαίνει κλίση  $1/2$  στο κλιμακόγραμμα. Θετικά αυτοσυσχετισμένες ανελίξεις αποφέρουν μεγαλύτερη  $\sigma^{(k)}$  και πιθανώς μικρότερες κλίσεις στο κλιμακόγραμμα.



Σχήμα 12 Η απόκλιση της αναμενόμενης τιμής από την πραγματική τιμή του κλιμακογράμματος για διάφορες παραμέτρους  $H$ .

Στο σχήμα 12 παρατηρούμε ότι η αναμενόμενη τιμή του κλιμακογράμματος είναι διαφορετική σε περιπτώσεις με συντελεστή *Hurst* μεγαλύτερο από 0.5. Μόνο στην περίπτωση λευκού θορύβου ( $H = 0.5$ ) το κλιμακόγραμμα είναι αμερόληπτο (μηδενική αυτοσυσχέτιση). Πολλές φορές στην στοχαστική προσομοίωση χρησιμοποιούμε την αναμενόμενη τιμή μιας ανέλιξης χωρίς να παίρνουμε υπόψη μας ότι αυτή είναι διαφορετική από την πραγματική. Αυτή η διαφορά οφείλεται στο γεγονός ότι δεν μπορούμε να έχουμε χρονοσειρές άπειρου μήκους ή, ισοδύναμα, άπειρες χρονοσειρές πεπερασμένου μήκους.

### 2.3.3 Monte Carlo

Προσομοίωση ονομάζεται η τεχνική μίμησης ενός πραγματικού συστήματος όπως αυτό εξελίσσεται στο χρόνο (Winston, 1994). Μοντέλο προσομοίωσης ονομάζεται ένα σύνολο υποθέσεων για την λειτουργία του συστήματος, εκφρασμένων από μορφή μαθηματικών ή λογικών σχέσεων μεταξύ των αντικειμένων του συστήματος (Koutsoyiannis, 2000).

Η μέθοδος Monte Carlo (ή πείραμα/προσομοίωση) είναι μια στοχαστική διαδικασία με την οποία μπορούμε να κατανοήσουμε ένα σύστημα με την προσομοίωση από τυχαίο δείγμα. Με την χρήση λοιπόν τυχαίων αριθμών (και τη στατιστική) μπορούμε να λύσουμε ένα πρόβλημα σε τομείς όπως τα μαθηματικά, τη φυσική, τη βιολογία, τα οικονομικά, τη μηχανική όπως επίσης έχει αξιοποιηθεί και στην θεωρία της μουσικής. Χρησιμοποιείται για (integrate) ενός αντιπροσωπευτικού δείγματος για μια ανέλιξη, για παραγωγή τελείως διαφορετικών γεγονότων/σεναρίων, για τον εντοπισμό προσομοιώσεων για την διόρθωση δεδομένων, ανάλυση και προσομοίωση δεδομένων και άλλα πολλά. Είναι απαραίτητη μέθοδος και για την προετοιμασία πειραμάτων όταν δεν υπάρχουν δεδομένα αλλά και για την ορθή επαλήθευση όταν διεξάγονται τα πειράματα ώστε να κατανοούμε περισσότερο τα αποτελέσματα.

Η μέθοδος Monte Carlo δίνει μια διαφορετική προσέγγιση στην επίλυση ενός ολοκληρώματος. Στην πιο απλή του μορφή η ολοκλήρωση με Monte Carlo γίνεται με το να πάρουμε την μέση τιμή της συνάρτησης υπολογιζόμενη σε τυχαία σημεία μέσα στο διάστημα ολοκλήρωσης. Ας υποθέσουμε ότι θέλουμε να υπολογίσουμε το ορισμένο ολοκλήρωμα μιας συνάρτησης  $f(x)$ :

$$I = \int_{x_1}^{x_2} f(x) dx = (x_2 - x_1)E(f(x)) \approx (x_2 - x_1) \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (2.85)$$

που σημαίνει ότι αν πάρουμε  $N$  τιμές του  $x$ , ομοιόμορφα κατανεμημένες στο διάστημα ολοκλήρωσης  $(x_1, x_2)$ , η μέση τιμή της  $f(x)$  θα αποτελεί μια καλή εκτίμηση του  $I$ .

Για τον παραπάνω υπολογισμό θα χρειαστούμε  $N$  τιμές του  $x$  ομοιόμορφα κατανεμημένες. Μπορούμε να χρησιμοποιήσουμε μια ακολουθία τυχαίων αριθμών ομοιόμορφα κατανεμημένων. Η χρήση τυχαίων αριθμών οδηγεί στην πιο σημαντική ιδιότητα της μεθόδου Monte Carlo, ότι δηλαδή για μεγάλο  $N$  η ακρίβεια της μεθόδου ακολουθεί το Κεντρικό Οριακό Θεώρημα. Αυτό σημαίνει ότι η κατανομή της  $f(i)$  τείνει σε Γκαουσιανή με τυπική απόκλιση  $\sigma_{MC} = \sigma/\sqrt{N}$ , με  $\sigma$  την τυπική απόκλιση της  $f(i)$  και από την στιγμή που η  $\sigma$  προσεγγίζει την τυπική απόκλιση της  $f(x)$ , τότε το σφάλμα ελλατώνεται ως  $1/\sqrt{N}$ .

Μια βασική αρχή της μεθόδου είναι η γέννηση τυχαίων αριθμών. Μία γεννήτρια είναι ένας αλγόριθμος, συνήθως αναδρομικός, ο οποίος μπορεί να παράγει διαδοχικά οσοδήποτε όρους της τυχαίας ακολουθίας. Σε μια ομοιόμορφη κατανομή τυχαίων αριθμών στο διάστημα  $[0,1)$  όλοι οι αριθμοί έχουν την ίδια πιθανότητα να επιλεχθούν. Προφανώς οι τυχαίοι αριθμοί δεν γεννώνται στην τύχη, αλλά βάσει ενός προσδιοριστικού αλγορίθμου. Οι γεννήτριες τυχαίων αριθμών μπορούν να δημιουργήσουν μια τέτοια ακολουθία αν και μπορεί να θεωρηθεί ως ψευδο-τυχαία (με αρκετά μεγάλη περίοδο) αφού για κάθε εκτέλεση του προγράμματος θα πάρουμε την ίδια ακολουθία για την ίδια αρχική τιμή του σπόρου (seed) της ακολουθίας.

Τα πλεονεκτήματα της μεθόδου είναι η ευκολία, η αμεσότητα και η ευελιξία σε συνδυασμό με την ακρίβεια στην περιγραφή του συστήματος. Στην φύση υπάρχουν συστήματα που δεν περιγράφονται από κάποιον μαθηματικό κανόνα (δεν είναι ντετερμινιστικά). Έτσι η μέθοδος Monte Carlo χρησιμοποιείται για την προσομοίωση φυσικών διεργασιών όταν η εφαρμογή αναλυτικών μεθόδων είναι ανέφικτη ή δυσχερής.

Η προσομοίωση του Monte Carlo θεμελιώθηκε επιστημονικά από την δεκαετία του 1940. Ο Stan Ulam παίζοντας πασιέντζα προσπάθησε να υπολογίσει την πιθανότητα να κερδίσει βασιζόμενος στην αρχική μοίρασιά των καρτών. Έπειτα από εξαντλητικούς υπολογισμούς για τους πιθανούς συνδυασμούς αποφάσισε να

ακολουθήσει μια άλλη τεχνική, δηλαδή να δοκιμάζει πολλές διαφορετικές μοιρασίες και να σημειώνει πόσα είναι τα πετυχημένα παιχνίδια. Συνειδητοποίησε ότι οι υπολογιστές μπορούν να χρησιμοποιηθούν για να λύσουν τέτοια παιχνίδια. Ο Stan Ulam συνεργάστηκε με τον John Von Neumann για να αναπτύξει αλγορίθμους σχετικά με τον έλεγχο υποθέσεων (importance sampling/rejection sampling). Ο Ulam και ο Von Neumann πρότειναν την ενίσχυση της έρευνας για την πυρηνική σχάση στο Los Alamos με την χρήση υπολογιστικών πειραμάτων βασισμένα στην τύχη. Το σχέδιο αυτό ήταν άκρως μυστικό για αυτό και ονομάστηκε Monte Carlo όπως το καζίνο στο Μονακό. Ο Nick Metropolis ανέπτυξε υπολογιστικές συσκευές που μπορούσαν να κάνουν τέτοιους υπολογισμούς και το 1953 με την δημοσίευσή του 'Η μέθοδος Monte Carlo' στο περιοδικό Journal of the American Statistics Association έκανε γνωστή την μέθοδο στην επιστημονική κοινότητα. Από το 1980 και μετά πολλές δημοσιεύσεις γράφτηκαν από τομείς όπως την φυσική, την στατιστική και την τεχνητή νοημοσύνη.





## 3 ΑΝΑΛΥΣΗ

---

### 3.1 ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

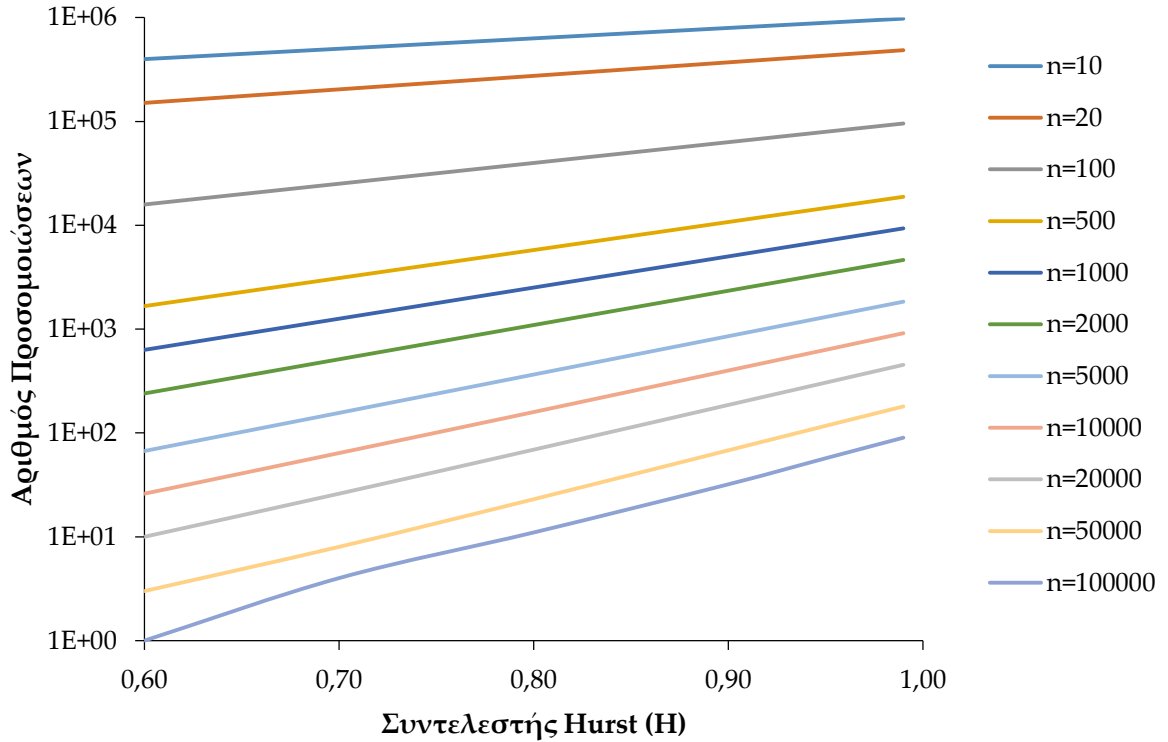
Υποθέτοντας ότι γνωρίζουμε την κατανομή του κλιμακογράμματος στην κλίμακα 1, έγινε γραφικός έλεγχος καταλληλότητας αυτής της υπόθεσης.

Με την χρήση του λογισμικού MATLAB (Matrix Laboratory) δημιουργήθηκε ένας κώδικας ο οποίος βοήθησε στην εξαγωγή των κλιμακογραμμάτων και της κατανομής τους. Το λογισμικό αυτό δίνει την δυνατότητα πολλών επαναλήψεων τα αποτελέσματα των οποίων μπορούν να αποθηκευτούν συγκεντρωτικά σε πίνακες ή γραφήματα και έπειτα μπορούν να περαστούν σε φύλλα του Excel για περαιτέρω ανάλυση.

Το πρώτο βήμα της προσομοίωσης είναι η παραγωγή της συνθετικής χρονοσειράς. Τα δεδομένα που χρησιμοποιούνται είναι ο συντελεστής Hurst και το μήκος της χρονοσειράς  $n$ . Οι δυο αυτές παράμετροι παίρνουν πολλές τιμές ( $H=0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99, m=10, 20, 100, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000$ ) για να εξεταστούν διαφορετικά σενάρια (77 στο σύνολο). Για το μοντέλο SMA ορίζεται μέση τιμή 0 και τυπική απόκλιση 1. Έπειτα υπολογίζεται το κλιμακόγραμμα του κάθε μοντέλου καθώς και κάποια στατιστικά χαρακτηριστικά του όπως η μέση τιμή, η διασπορά, η κύρτωση, η ασυμμετρία, η πιο πιθανή τιμή, η διάμεσος, το πρώτο και το τρίτο τεταρτημόριο σε κάθε κλίμακα  $k$ . Η ανάλυση αυτή έγινε πολλές φορές χρησιμοποιώντας την εμπειρική σχέση:

$$\text{floor}\left(\frac{10^7}{n^{2-H}}\right) + 1 \quad (3.1)$$

με την λογική ότι για ένα μικρό μήκος δείγματος (π.χ.  $n = 10$ ) και για  $H = 1$  ο μέγιστος αριθμός επαναλήψεων να είναι  $10^6$ , λόγω περιορισμού από τον υπολογιστικό φόρτο.



Σχήμα 13 Αριθμός προσομοιώσεων για κάθε συνδυασμό  $H, n$ .

### 3.2 ΚΑΤΑΝΟΜΗ ΚΛΙΜΑΚΟΓΡΑΜΜΑΤΟΣ

#### 3.2.1 Η περίπτωση του λευκού θορύβου

Στην περίπτωση του λευκού θορύβου ( $H = 0.5$ ) όπου τα δεδομένα είναι ανεξάρτητα μπορούμε να αποδείξουμε ότι το κλιμακόγραμμα στην βασική κλίμακα ( $k = 1$ ) σχετίζεται με την κατανομή  $\chi^2$ . Αποδεικνύεται ότι η δειγματική διασπορά πολλαπλασιασμένη με μια σταθερά ακολουθεί κατανομή  $\chi^2$ .

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \quad (3.2)$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (3.3)$$

Χρησιμοποιώντας την  $z = \frac{x_i - \mu}{\sigma}$

$$\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{\sigma} + \frac{\bar{x} - \mu}{\sigma} \right)^2 \quad (3.4)$$

$$= \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{\sigma} \right)^2 + \frac{2}{\sigma^2} (\bar{x} - \mu) \sum_{i=1}^n (x_i - \bar{x}) + \sum_{i=1}^n \left( \frac{\bar{x} - \mu}{\sigma} \right)^2 \quad (3.5)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 + n \left( \frac{\bar{x} - \mu}{\sigma} \right)^2 \quad (3.6)$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 + \left( \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right)^2 \quad (3.7)$$

όπου:

$$\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n-1}{\sigma^2} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} = \frac{(n-1)s^2}{\sigma^2} \quad (3.8)$$

για τον όρο:

$$\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 \quad (3.9)$$

είναι ένα άθροισμα από  $n$  ανεξάρτητες μεταβλητές της  $\chi^2$  κατανομής. Αυτό ισχύει γιατί έχουμε υποθέσει ότι οι  $X_1, X_2, \dots, X_n$  είναι παρατηρήσεις από τυχαίο δείγμα μεγέθους  $n$  από κανονική κατανομή  $N(\mu, \sigma^2)$ .

Άρα η:

$$\frac{X_i - \mu}{\sigma} \quad (3.10)$$

είναι τυποποιημένη κανονική μεταβλητή. Αν υψώσουμε μια τυποποιημένη κανονική μεταβλητή στο τετράγωνο τότε θα έχουμε μια τυχαία μεταβλητή της  $\chi^2$  κατανομής με 1 βαθμό ελευθερίας. Οπότε:

$$\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 \quad (3.11)$$

είναι ένα άθροισμα  $n$  ανεξάρτητων μεταβλητών  $\chi^2$  κατανομής με  $n$  βαθμούς ελευθερίας. Έτσι η ροπογεννήτρια του παραπάνω αθροίσματος είναι:

$$M_x(t) = (1 - 2t)^{-n/2} \quad (3.12)$$

Ο δειγματικός μέσος είναι ομοιόμορφα κατανομημένος με μέση τιμή  $\mu$  και διασπορά  $\sigma^2/n$ . Άρα η:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad (3.13)$$

είναι τυποποιημένη κανονική μεταβλητή. Αν υψώσουμε την  $Z$  στο τετράγωνο τότε θα έχουμε μια τυχαία μεταβλητή της  $\chi^2$  κατανομής με 1 βαθμό ελευθερίας.

$$Z^2 = \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2(1) \quad (3.14)$$

Έτσι η ροπογεννήτρια της  $Z^2$  είναι:

$$M_{Z^2}(t) = (1 - 2t)^{-1/2} \text{ για } t < \frac{1}{2} \quad (3.15)$$

τελικά θέτωντας

$$W = \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} \quad (3.16)$$

Παίρνοντας τον ορισμό της ροπογεννήτριας για την  $W$  έχουμε:

$$M_W(t) = E(e^{tW}) = E[e^{t((n-1)S^2/\sigma^2 + Z^2)}] \quad (3.17)$$

$$E(e^{tW}) = E[e^{t((n-1)S^2/\sigma^2)} e^{tZ^2}] = M_{(n-1)S^2/\sigma^2}(t) \cdot M_{Z^2}(t)$$

$$(1 - 2t)^{-n/2} = M_{(n-1)S^2/\sigma^2}(t) \cdot (1 - 2t)^{-1/2}$$

$$M_{(n-1)S^2/\sigma^2}(t) = (1 - 2t)^{-n/2} \cdot (1 - 2t)^{1/2}$$

οπότε η δειγματική διασπορά πολλαπλασιασμένη με την σταθερά  $(n-1)/\sigma^2$  ακολουθεί κατανομή  $\chi^2$  με  $n-1$  βαθμούς ελευθερίας.

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{(n-1)}^2 \quad (3.18)$$

Επίσης για την  $\chi^2$  κατανομή η μέση τιμή και η διασπορά είναι:

$$E(\chi_n^2) = n \quad (3.19)$$

$$Var(\chi_n^2) = 2n \quad (3.20)$$

Χρησιμοποιώντας αυτές τις ιδιότητες μπορούμε να υπολογίσουμε την μέση τιμή και διασπορά της  $S^2$ :

$$E(S^2) = E\left(\frac{\sigma^2 \chi_{n-1}^2}{n-1}\right) = \frac{\sigma^2(n-1)}{n-1} = \sigma^2 \quad (3.21)$$

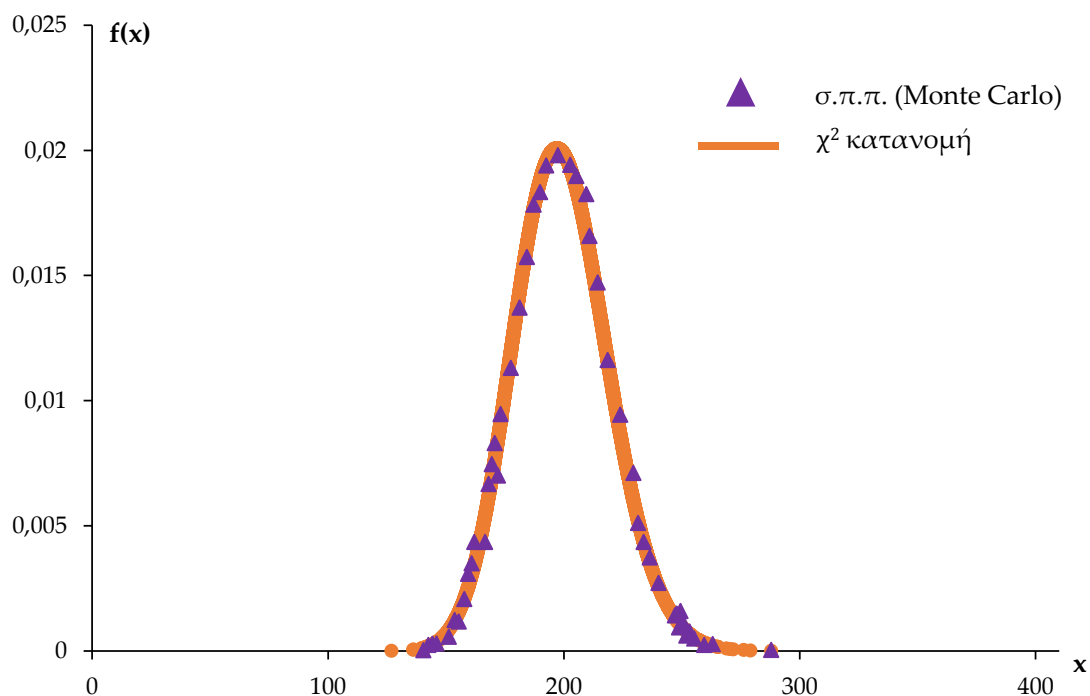
$$Var(S^2) = Var\left(\frac{\sigma^2 \chi_{n-1}^2}{n-1}\right) = \frac{\sigma^4}{(n-1)^2} Var(\chi_{n-1}^2) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1} \quad (3.22)$$

Επίσης:

$$\mu^{(3)} = \sqrt{\frac{8}{(n-1)}} \quad (3.23)$$

$$\mu^{(4)} = \frac{12}{n-1} + 3 \quad (3.24)$$

Για τον έλεγχο αυτής της υπόθεσης χρησιμοποιήθηκε το πρόγραμμα Excel. Πράγματι για  $H = 0.5$  το κλιμακόγραμμα πολλαπλασιασμένο με  $(n-1)$  ακολουθεί  $\chi^2$  κατανομή με  $n$  βαθμούς ελευθερίας για την κλίμακα  $k = 1$ .



Σχήμα 14 Συνάρτηση πυκνότητας πιθανότητας για τις δυο περιπτώσεις με  $n=200$ ,  $H=0.5$ ,  $k=1$ .

Ξεκινάμε με την προσομοίωση Monte Carlo και με την χρήση του μοντέλου SMA παράγουμε 10000 χρονοσειρές με μήκος δείγματος  $n = 200$  και συντελεστή  $H = 0.5$ . Ύστερα παράγουμε τα κλιμακογράμματα της κάθε χρονοσειράς. Χρησιμοποιώντας τα κλιμακογράμματα στην κλίμακα  $k = 1$  καταλήγουμε να έχουμε 10000 τιμές της διασποράς της συναθροισμένης ανέλιξης. Δημιουργώντας την συνάρτηση πυκνότητας πιθανότητας των τιμών αυτών και συγκρίνοντας αυτές με την σ.π.π. για την κατανομή  $\chi^2$  με  $n = 199$  βαθμούς ελευθερίας παίρνοντας ως τιμές τις 10000 τιμές της διασποράς συμπεραίνουμε ότι προσεγγίζει ακριβώς την  $\chi^2$ . Καταφέρνει και προσεγγίζει και τις 4 ροπές (μέση τιμή, διασπορά, ασυμμετρία, κύρτωση).

Η συνάρτηση πυκνότητας πιθανότητας της  $\chi^2$  κατανομής υπενθυμίζεται ότι ισούται με:

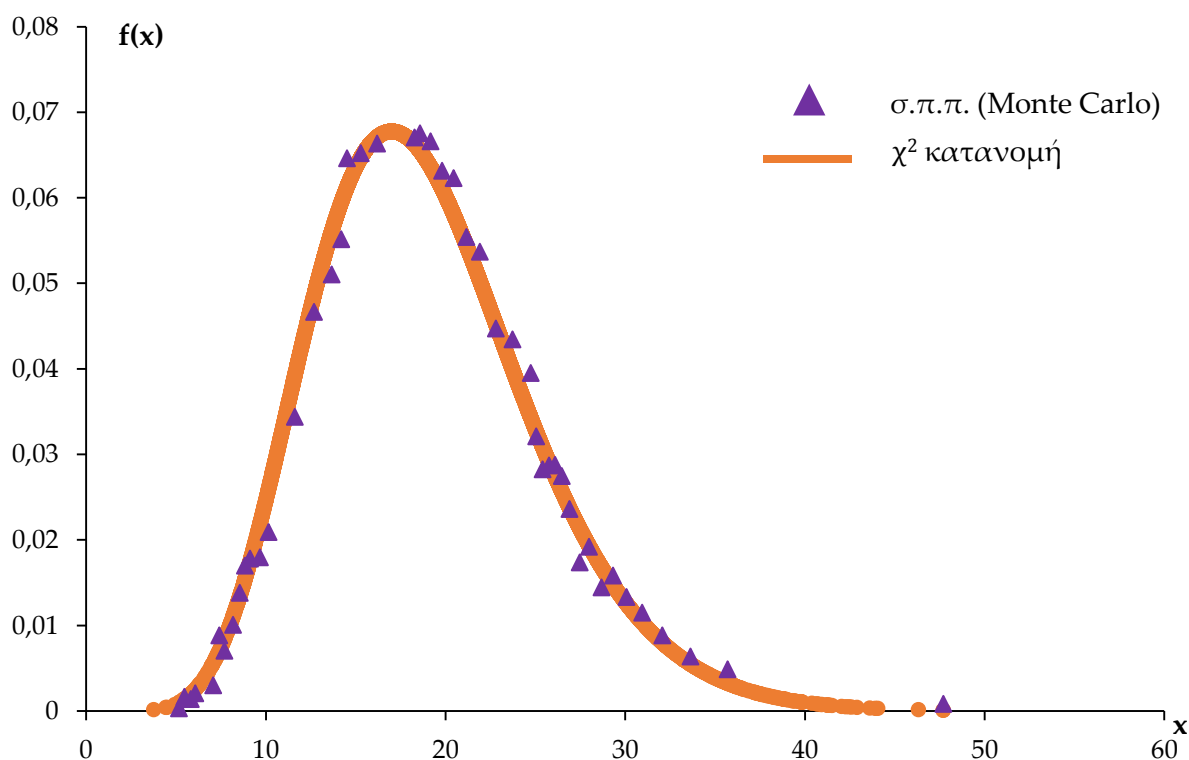
$$f_x(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2} \quad (3.25)$$

Για την κλίμακα 1 θεωρούμε  $n(1) = n - 1$ . Λόγω της σχέσης  $\sigma^{2(k)} = \sigma^2/k$  τότε οι βαθμοί ελευθερίας για  $k > 1$  υπολογίζονται από την σχέση:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^{2(k)} \cdot k} = \frac{n-1}{k} \cdot \frac{S^2}{\sigma^{2(k)}} \sim \chi^2_{\left(\frac{n-1}{k}\right)} \quad (3.26)$$

Στις υπόλοιπες κλίμακες οι βαθμοί ελευθερίας της  $\chi^2$  κατανομής ακολουθούν την σχέση:

$$n' = (n-1)/k \quad (3.27)$$



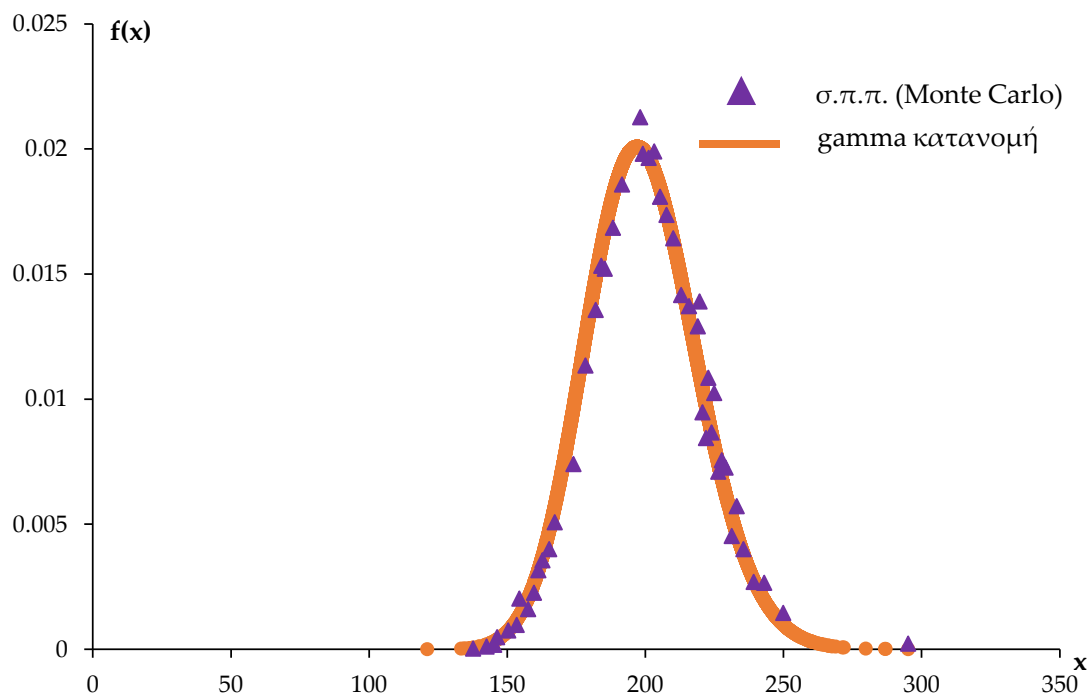
Σχήμα 15 Συνάρτηση πυκνότητας πιθανότητας για τις δυο περιπτώσεις με  $n=200$ ,  $H=0.5$ ,  $k=10$ .

Στο σχήμα φαίνονται οι σ.π.π. για την κλίμακα  $k = 10$  του κλιμακογράμματος με  $\sigma^2 = 0.1$  και  $n = 20$  βαθμούς ελευθερίας.

Παρατηρούμε ότι η κατανομή για  $H=0.5$  είναι ελάχιστα ασύμμετρη (σχεδόν συμμετρική)  $\mu_{\chi^2}^{(4)} \approx 0,2$ . Ελέγχθηκε ότι η κατανομή διατηρεί και τις τέσσερις ροπές.

### 3.2.2 Η περίπτωση HK

Η δοκιμή της  $\chi^2$  κατανομής σε δεδομένα με μακροπρόθεσμη εμμονή έδειξε ότι η κατανομή διατηρεί την μέση τιμή και τυπική απόκλιση αλλά όχι την ασυμμετρία και την κύρτωση. Το κλιμακόγραμμα για  $H > 0.5$  είναι ασύμμετρο και για αυτό τον λόγο η κατανομή αυτή απορρίφθηκε. Παρατηρείται ότι η διασπορά του κλιμακογράμματος στην κλίμακα 1 είναι διπλάσια από αυτή για  $H = 0.5$  (έχει να κάνει με τέταρτη ροπή).



Σχήμα 16 Συνάρτηση πυκνότητας πιθανότητας για gamma κατανομή  $H=0.6$  και  $n=200$  για  $k=1$ .

Το μοντέλο SMA βρέθηκε να μην επηρεάζει αυτό το αποτέλεσμα (για τον ορθό έλεγχο δοκιμάσαμε το μοντέλο SMA για μεγαλύτερο αριθμό συντελεστών βάρους). Θα μπορούσαμε επίσης να τροποποιήσουμε τους συντελεστές αυτούς για την ελαχιστοποίηση του σφάλματος αποκοπής (Koutsoyiannis, 2016). Το επόμενο βήμα θα ήταν η δοκιμή καταλληλότητας άλλης κατανομής. Η  $\chi^2$  κατανομή είναι ειδική περίπτωση της γάμμα κατανομής. Συγκεκριμένα για  $\alpha = n/2$  και  $\beta = 2$  τότε η γάμμα κατανομή γίνεται  $\chi_n^2$ . Επίσης δοκιμάστηκε και η γενικευμένη γάμμα κατανομή.

Ενδεικτικά για την περίπτωση  $n = 200$  και  $H = 0.6$  πολλαπλασιάζουμε το κλιμακόγραμμα στην κλίμακα 1 με  $\frac{(n-1)}{\sigma^2}$  και προσαρμόζουμε την κατανομή gamma με  $\alpha = 99.5$  και  $\beta = 2$ . Η κατανομή gamma προσεγγίζει την μέση τιμή, διασπορά, και ασυμμετρία ενώ στην περίπτωση της κύρτωσης έχει διαφορά η οποία αυξάνεται με την αύξηση της κλίμακας και της παραμέτρου  $H$ .



Monte Carlo ανάλυση	Gamma κατανομή
$\mu = 199$	$\mu = 199$
$var = 403.9$	$var = 398$
$\mu^{(3)} = 0.2148$	$\mu^{(3)} = 0.2005$
$\mu^{(4)} = 3.1576$	$\mu^{(4)} = 3.0603$

*Πίνακας 2 Σύγκριση των ροπών της γάμμα κατανομής και των αντίστοιχων αποτελεσμάτων από την Monte Carlo ανάλυση.*

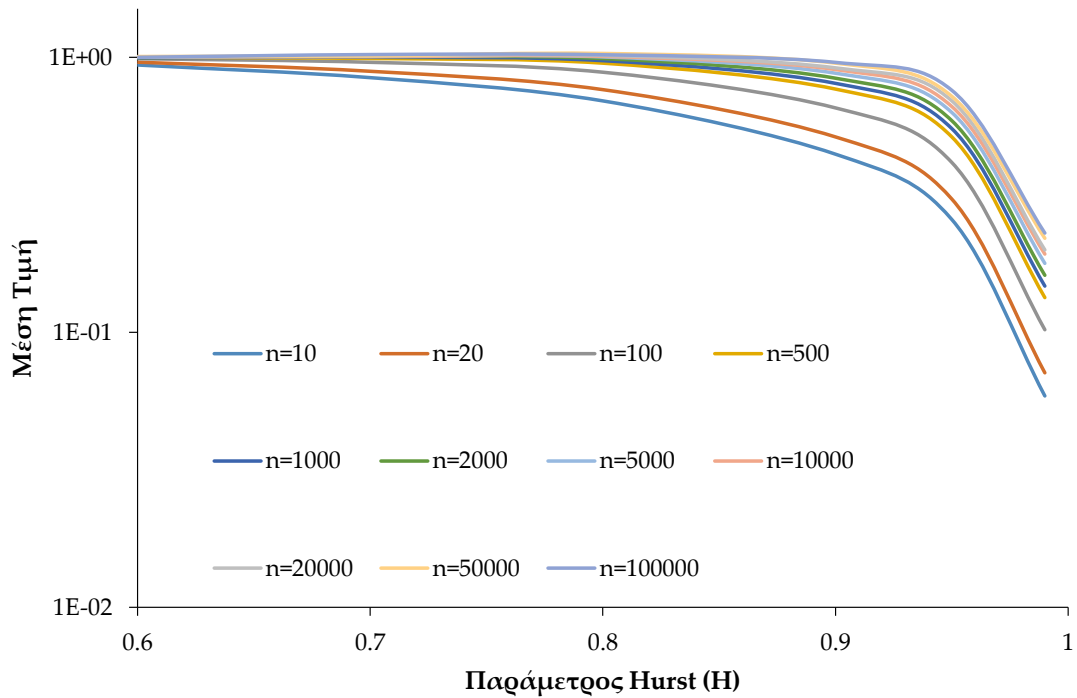
Αυτές οι κατανομές αδυνατούν να διατηρήσουν την τρίτη και τέταρτη ροπή οπότε έγινε ανάλυση Monte Carlo για να βγουν κάποια εμπειρικά συμπεράσματα.

Ελέγχθηκαν τα στατιστικά χαρακτηριστικά των κλιμακογραμμάτων στην κλίμακα  $k = 1$  και στην κλίμακα  $k = 0,1 * n$  ενδεικτικά.

### 3.2.2.1 Μέση τιμή

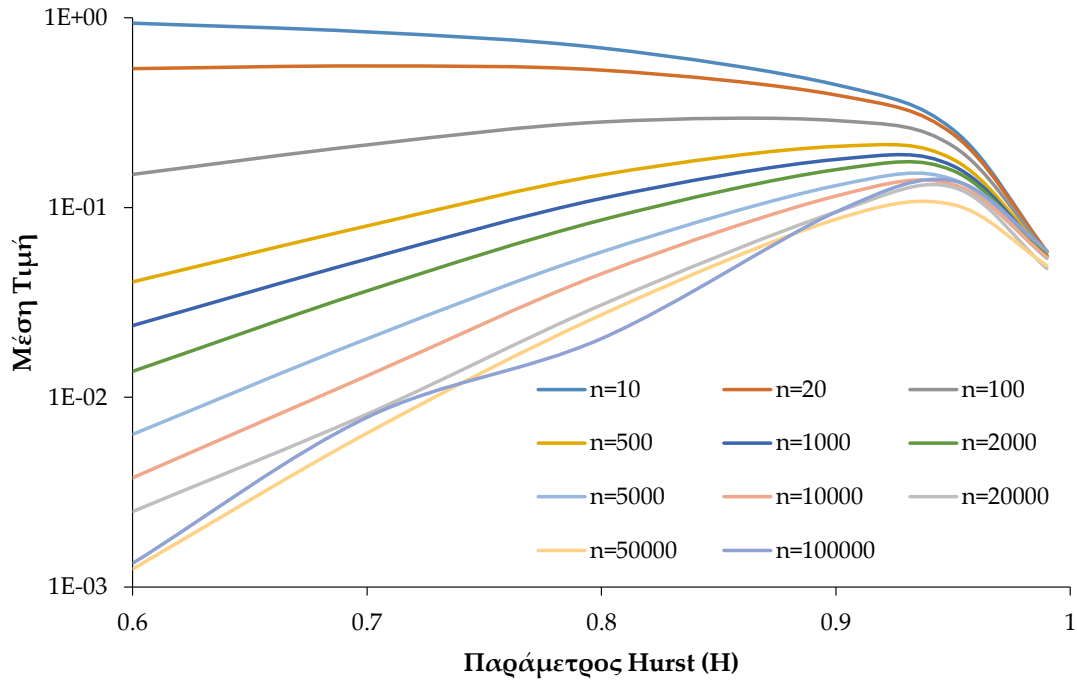
Αρχικά, ελέγχθηκε ότι η μέση τιμή του κλιμακογράμματος σε κάθε κλίμακα  $k$  που προκύπτει από την ανάλυση Monte Carlo συμπίπτει με την αναμενόμενη τιμή του σύμφωνα με τις θεωρητικές σχέσεις από το κεφάλαιο 2.

Έπειτα παρατείθεται το διάγραμμα για τα διαφορετικά μήκη δείγματος και τις διαφορετικές παραμέτρους *Hurst*.



Σχήμα 17 Μέση τιμή κλιμακογράμματος συναρτήσει της παραμέτρου  $H$  στην κλίμακα 1.

Παρατηρούμε ότι για  $H = 0.6$  η μέση τιμή του κλιμακογράμματος ισούται με 1 για όλες τις περιπτώσεις του μήκους δείγματος  $n$  στην κλίμακα  $k = 1$ . Καθώς αυξάνει η παράμετρος Hurst (άρα και η επίδραση της μακροπρόθεσμης εμμονής) τότε η μέση τιμή πέφτει λόγω της αρνητικής μεροληψίας.



Σχήμα 18 Μέση τιμή κλιμακογράμματος συναρτήσεως της παραμέτρου  $H$  στην κλίμακα  $0,1 * n$ .

Η μέση τιμή του κλιμακογράμματος στην κλίμακα  $k = 0.1 * n$  (π.χ. για  $n = 500$  στην κλίμακα  $k = 50$ , για  $n = 10000$  στην κλίμακα  $k = 1000$  κ.ο.κ.), παρατηρείται ότι καθώς αυξάνεται η παράμετρος Hurst τείνει στο μηδέν. Όπως και για την κλίμακα 1 καθώς αυξάνει η παράμετρος Hurst συγκλίνουν στο μηδέν.

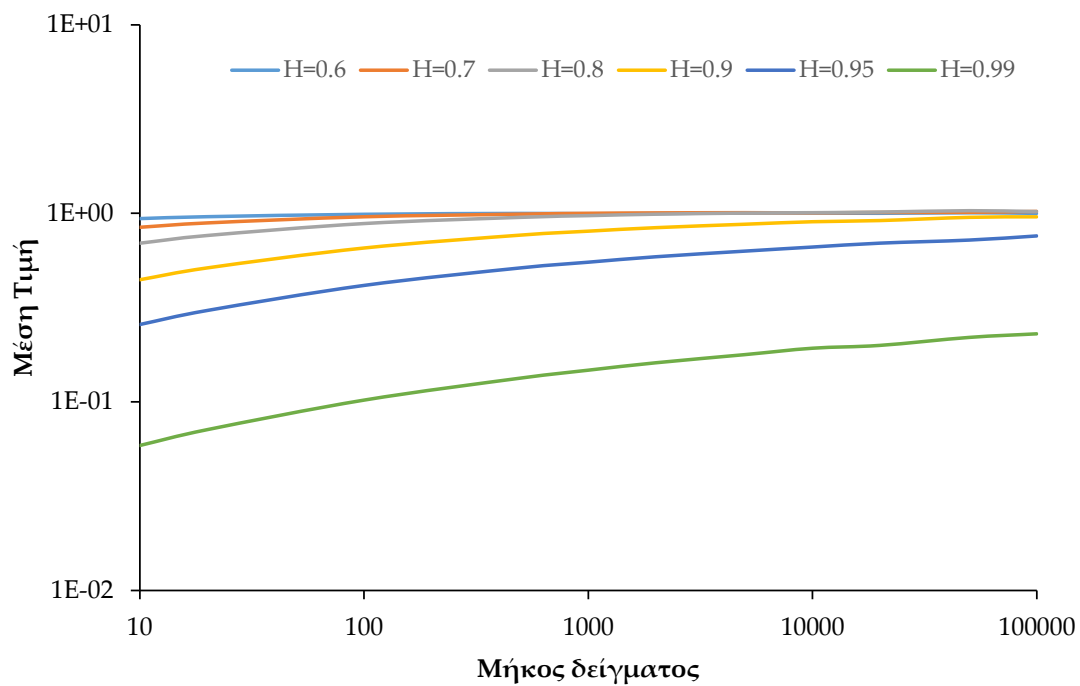
Από την θεωρητική σχέση της αναμενόμενης τιμής του κλιμακογράμματος προκύπτει ότι:

$$E[\gamma_d^{(\Delta)}(k)] = \frac{1 - \gamma_d^{(\Delta)}(n)/\gamma_d^{(\Delta)}(k)}{1 - k/n} \gamma_d^{(\Delta)}(k) \quad (3.28)$$

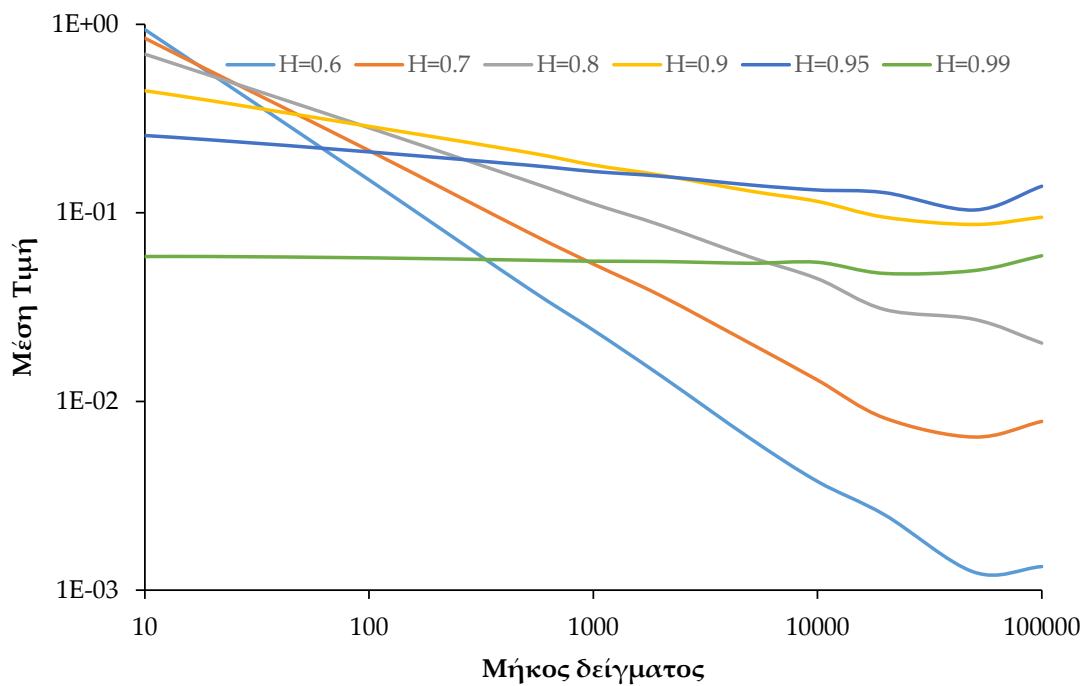
για  $H = 1$

$$= \frac{1}{k^{2-2H}} - \frac{1}{n^{2-2H}}}{1 - k/n} = 0$$

Ανεξάρτητα λοιπόν από την κλίμακα και το μήκος δείγματος, καθώς αυξάνει η παράμετρος Hurst τότε η μέση τιμή συγκλίνει στο μηδέν.



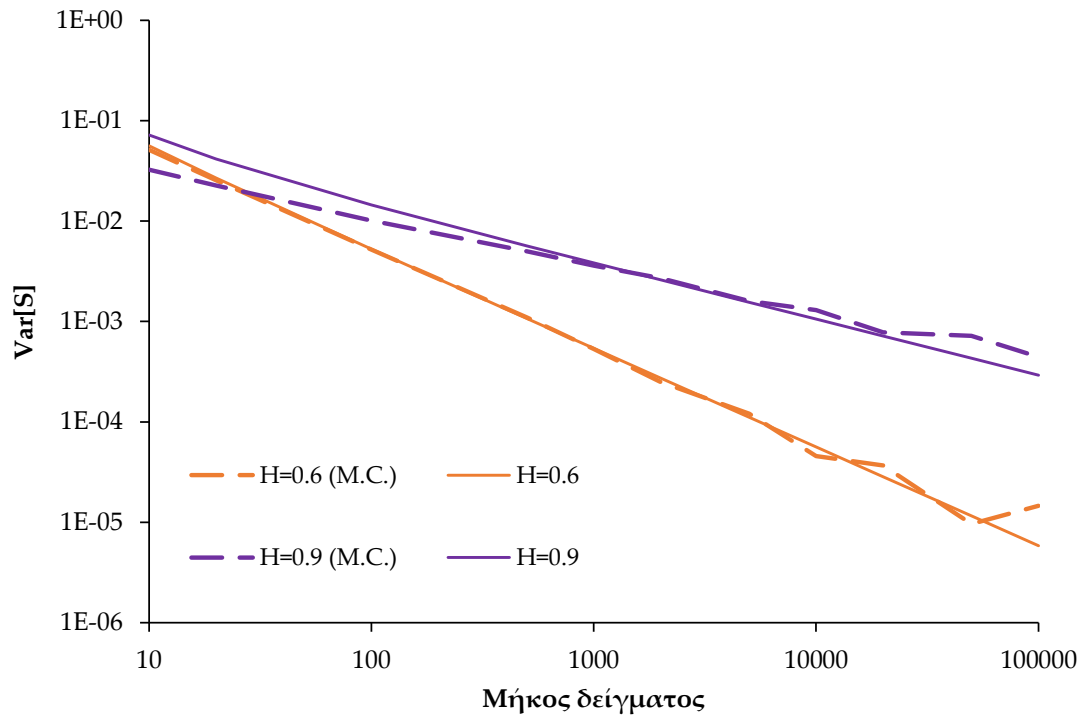
Σχήμα 19 Μέση τιμή κλιμακογράμματος συναρτήσει του μήκους δείγματος στην κλίμακα 1.



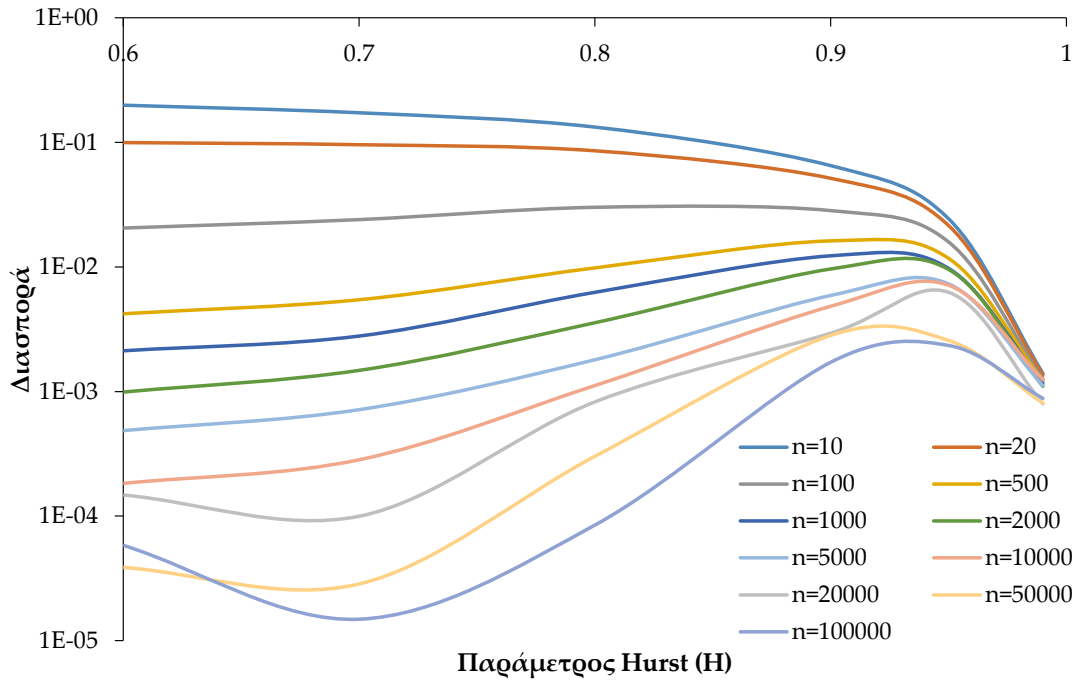
Σχήμα 20 Μέση τιμή κλιμακογράμματος συναρτήσει του μήκους δείγματος στην κλίμακα 0,1n.

### 3.2.2.2 Διασπορά

Για τον έλεγχο της διασποράς δεν υπάρχει θεωρητική σχέση οπότε χρησιμοποιήθηκε η εμπειρική σχέση από το κεφάλαιο 2 για τον έλεγχο της τυπικής απόκλισης του κλιμακογράμματος για δύο χαρακτηριστικές περιπτώσεις.

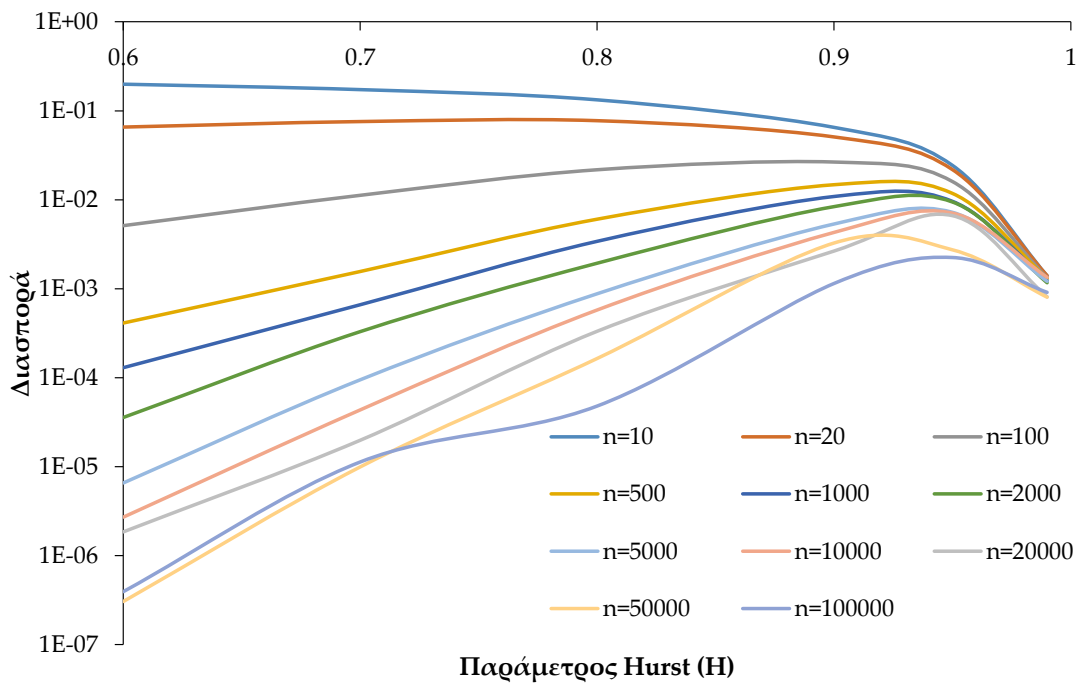


Σχήμα 21 Απόκλιση της τυπικής απόκλισης από την εμπειρική σχέση και τα αποτελέσματα Monte Carlo.

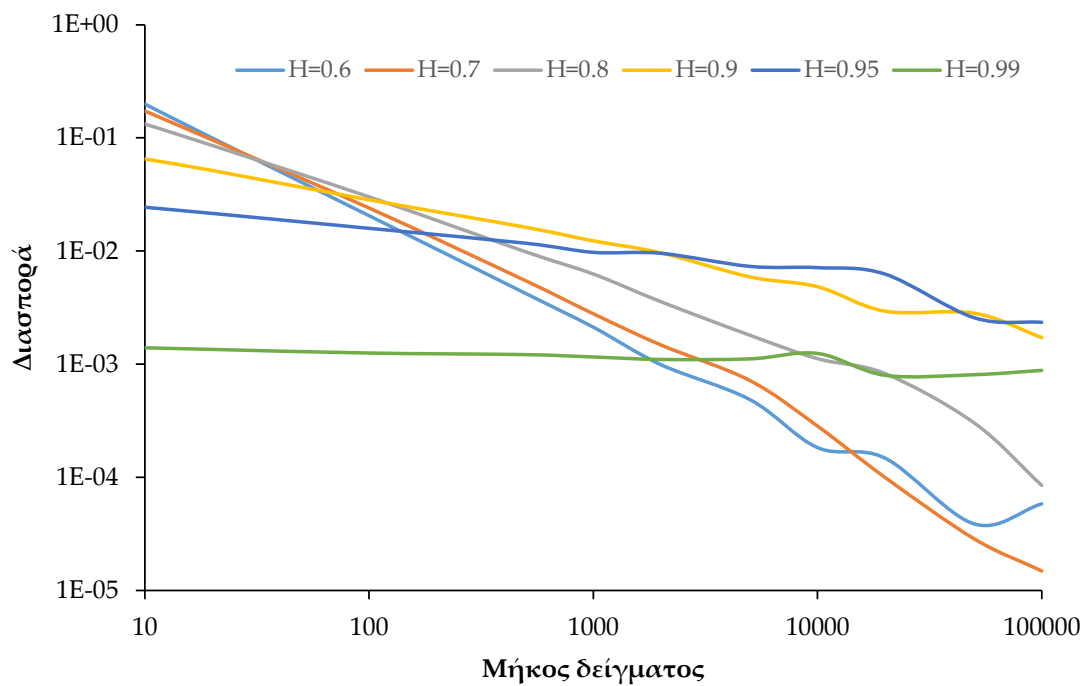


Σχήμα 22 Διασπορά κλιμακογράμματος συναρτήσεως της παραμέτρου  $H$  στην κλίμακα 1.

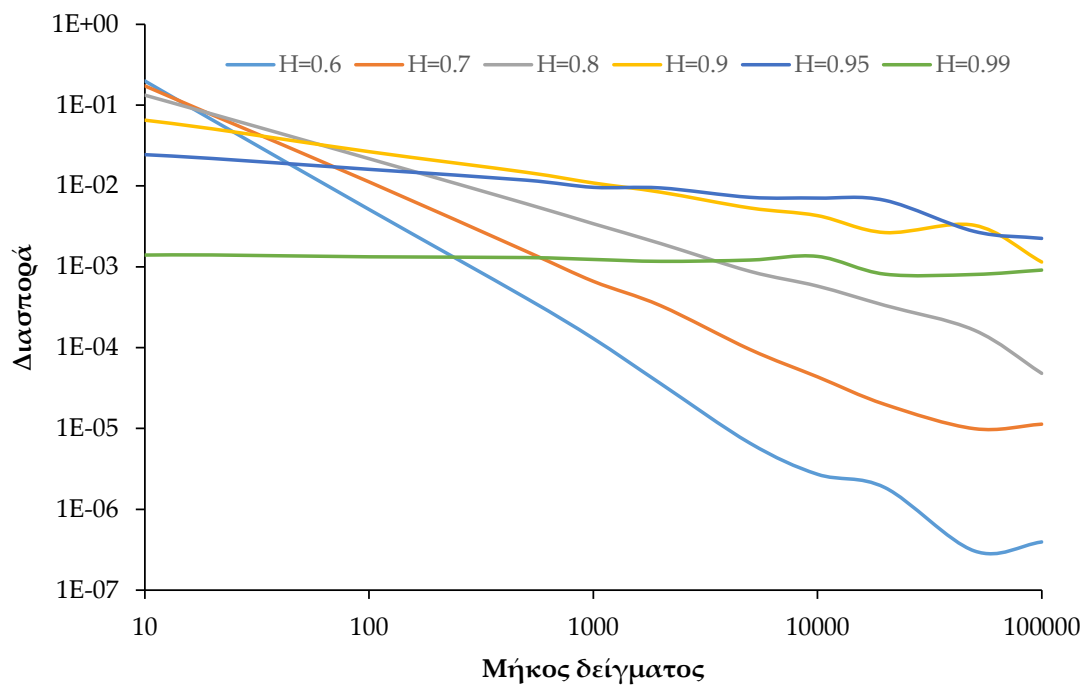
Παρατηρούμε ότι η διασπορά του κλιμακογράμματος για μεγάλο μήκος δείγματος ( $n = 50000, 100000$ ) έχει αρκετές διακυμάνσεις και αυτό οφείλεται στις λίγες προσομοιώσεις.



Σχήμα 23 Διασπορά κλιμακογράμματος συναρτήσεως της παραμέτρου  $H$  στην κλίμακα  $0, 1/n$ .

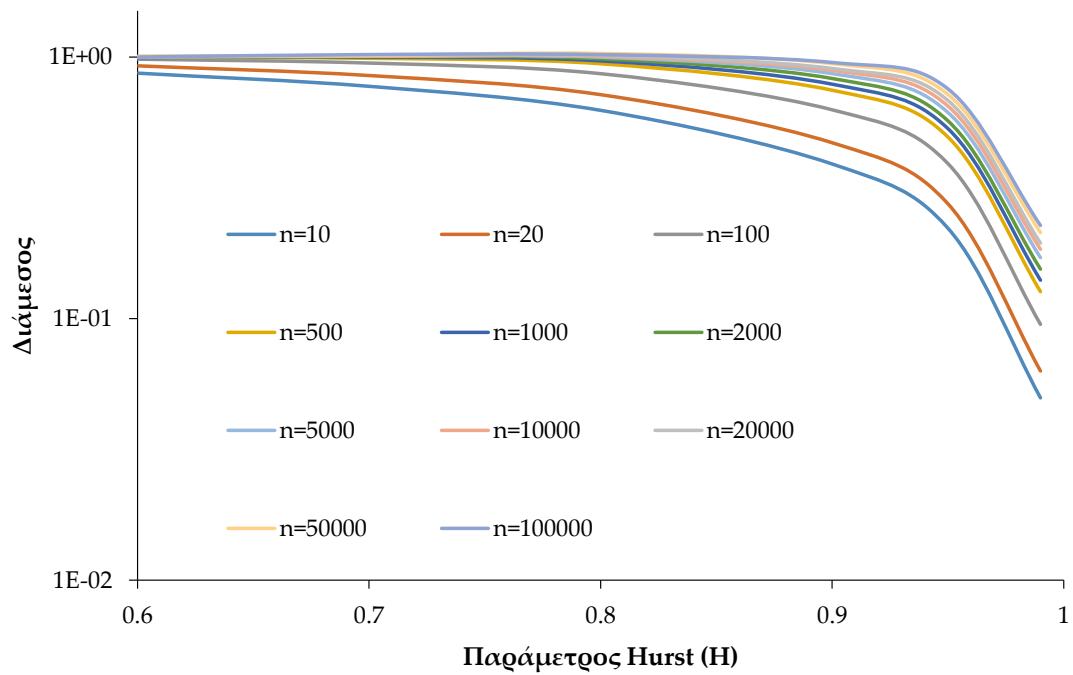


Σχήμα 24 Διασπορά κλιμακογράμματος συναρτήσει του μήκους δείγματος στην κλίμακα 1.

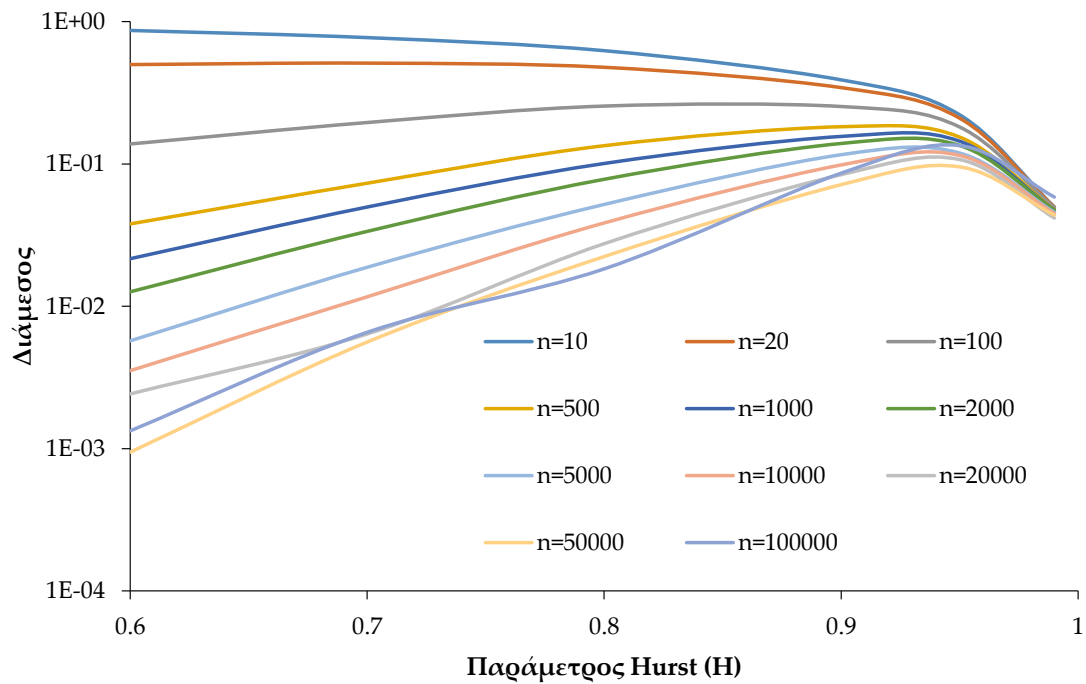


Σχήμα 25 Διασπορά κλιμακογράμματος συναρτήσει του μήκους δείγματος στην κλίμακα 0, 1n.

### 3.2.2.3 Διάμεσος

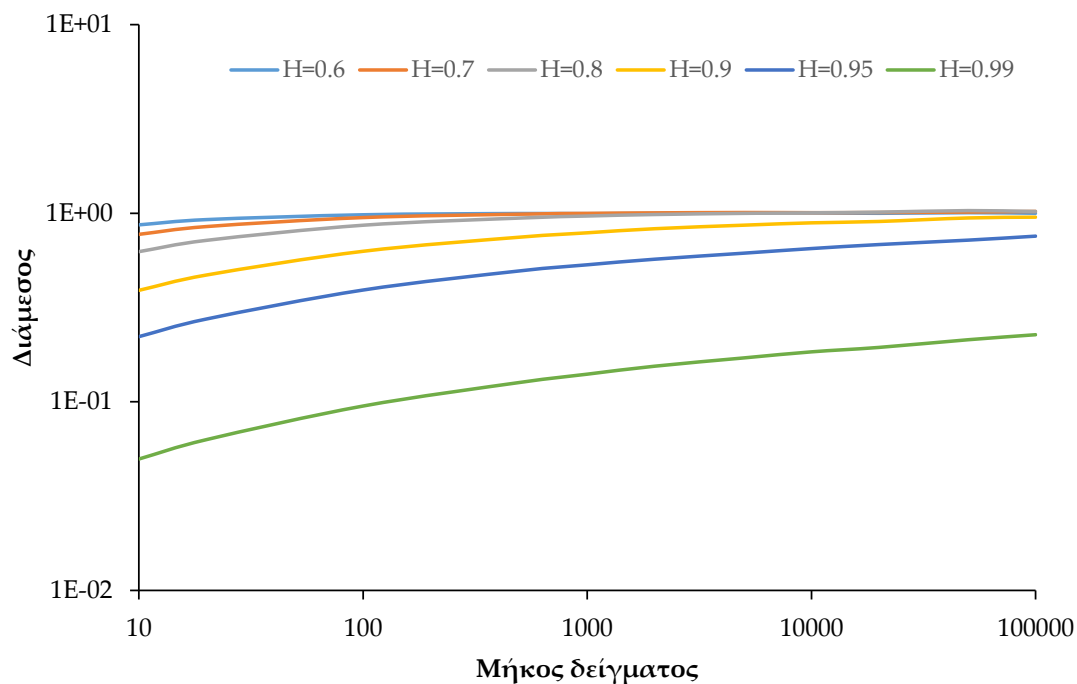


Σχήμα 26 Διάμεσος κλιμακογράμματος συναρτήσει της παραμέτρου H στην κλίμακα 1.



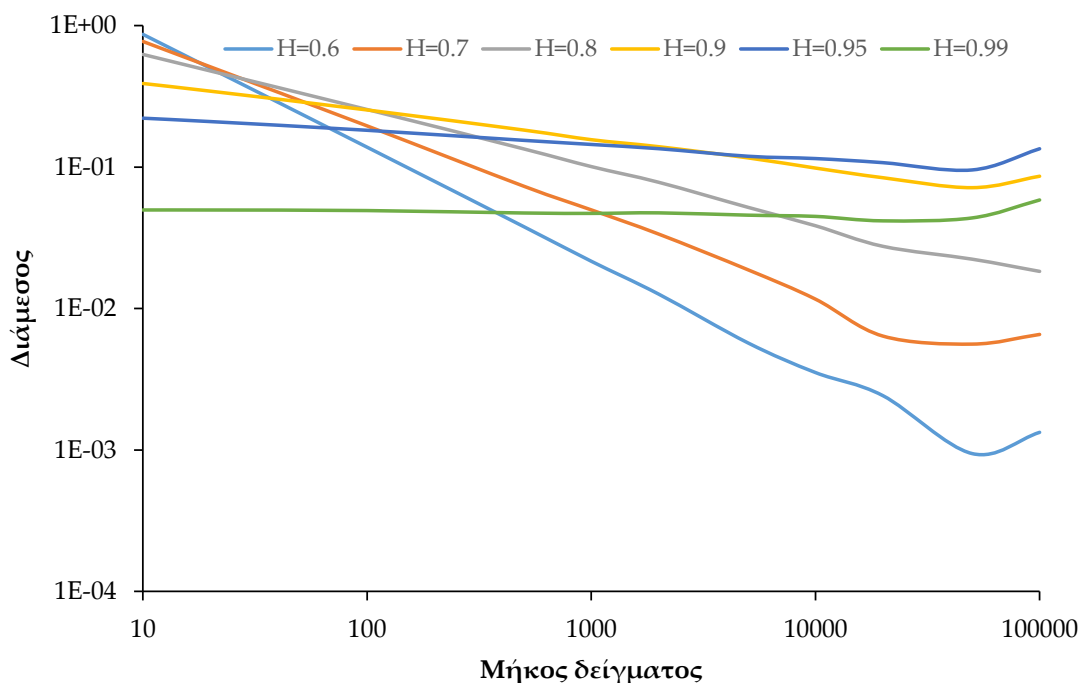
Σχήμα 27 Διάμεσος κλιμακογράμματος συναρτήσει της παραμέτρου H στην κλίμακα 0,1n.





Σχήμα 28 Διάμεσος κλιμακογράμματος συναρτήσει του μήκους δείγματος για διάφορες παραμέτρους Hurst στην κλίμακα 1.

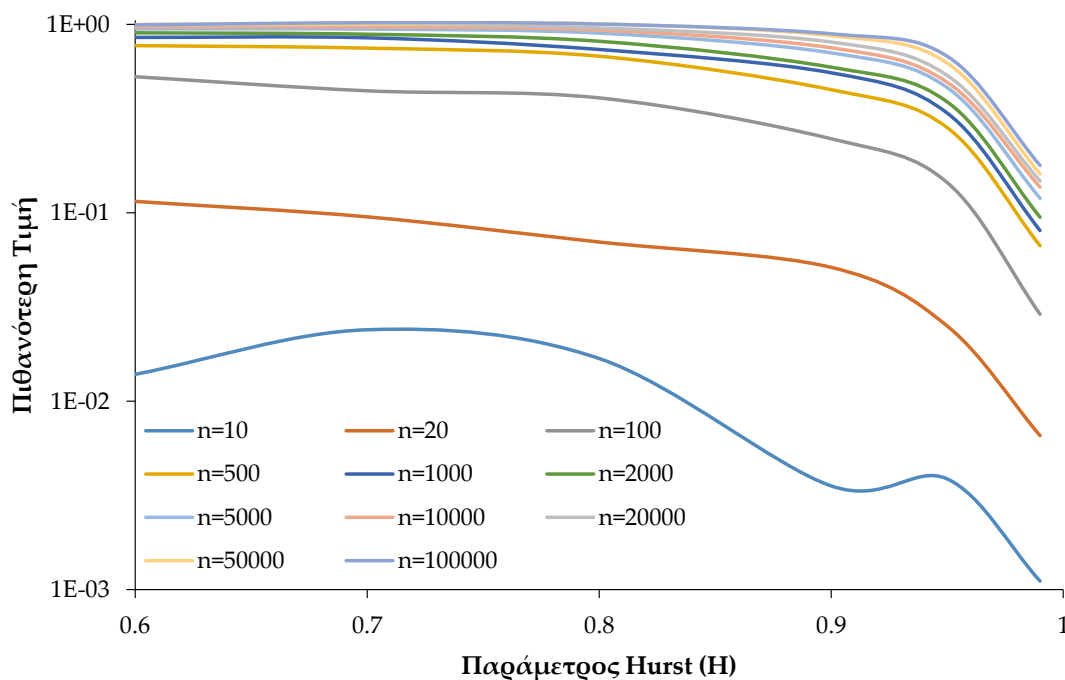
Η διάμεσος φαίνεται να διατηρεί μια σταθερή συμπεριφορά ειδικότερα για τις μικρές τιμές της παραμέτρου Hurst.



Σχήμα 29 Διάμεσος κλιμακογράμματος συναρτήσει του μήκους δείγματος στην κλίμακα 0, 1n.

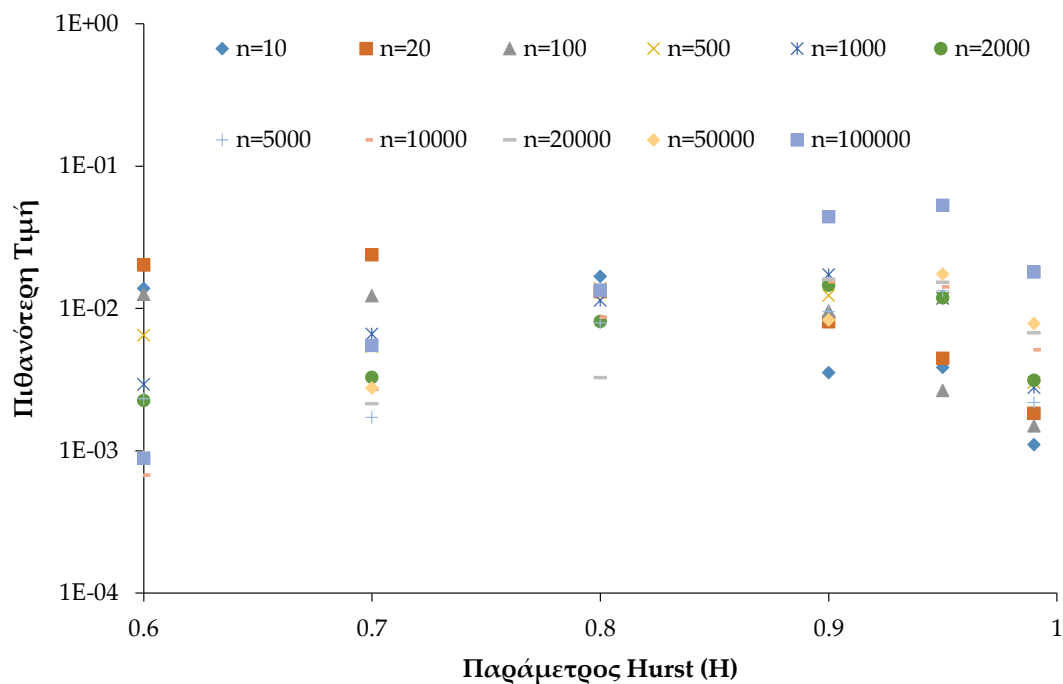
### 3.2.2.4 Πιθανότερη τιμή

Επειδή η κατανομή του κλιμακογράμματος είναι ασύμμετρη (θετική ασυμμετρία, μέση τιμή > πιο πιθανή τιμή) που σημαίνει ότι η μέση τιμή θα έχει μικρότερη πιθανότητα, αν η κατανομή διατηρεί την μέση τιμή τότε σίγουρα θα διατηρεί και την πιο πιθανή τιμή.



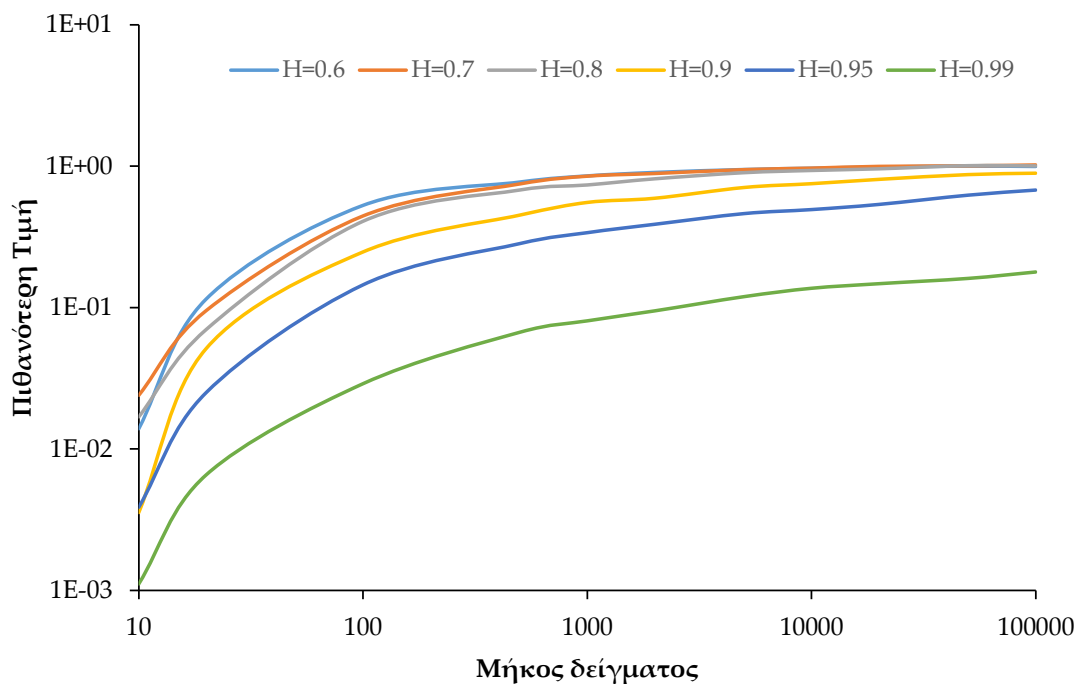
Σχήμα 30 Πιθανότερη τιμή κλιμακογράμματος για διάφορους συνδυασμούς  $n, H$  στην κλίμακα 1.

Ο λόγος που οι καμπύλες για  $n = 10$  και  $n = 20$  έχουν τέτοιες διακυμάνσεις σημαίνει ότι θα χρειαζόμασταν περισσότερες επαναλήψεις. Οι υπόλοιπες καμπύλες φαίνεται να παρουσιάζουν την ίδια συμπεριφορά.

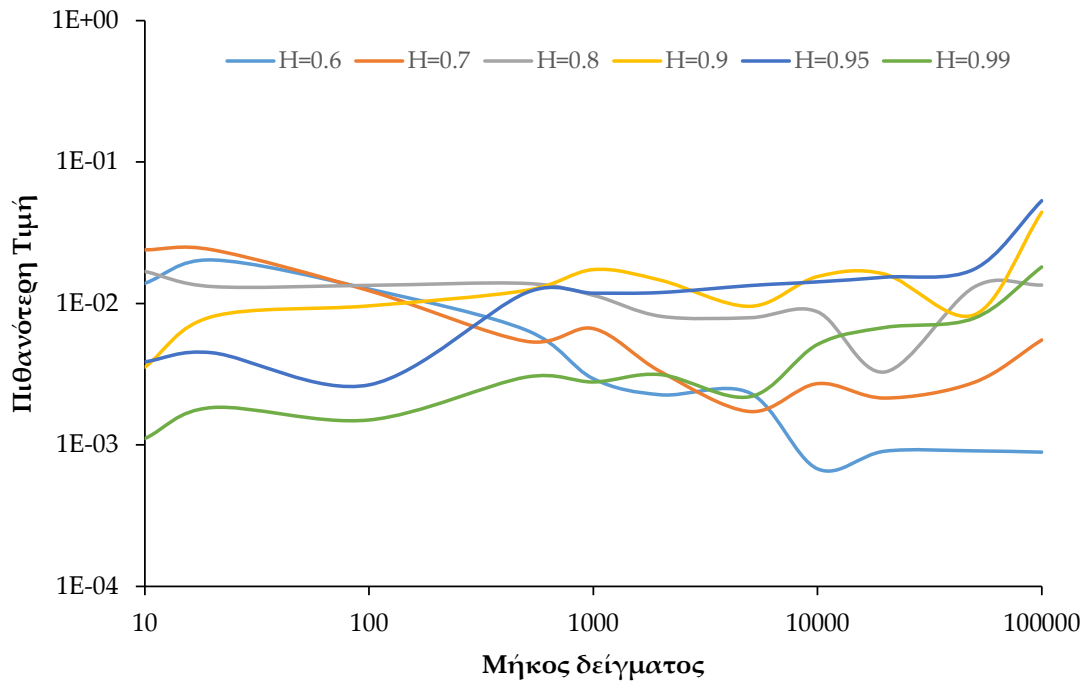


Σχήμα 31 Πιθανότερη τιμή κλιμακογράμματος συναρτήσει της παραμέτρου  $H$  στην κλίμακα  $0, 1n$ .

Παρατηρείται ένα νέφος σημείων τα οποία όμως φαίνονται να κινούνται σε κάποια όρια. Επίσης για  $n = 100000$  η πιθανότερη τιμή τείνει στην μέση τιμή καθώς η κατανομή στενεύει.



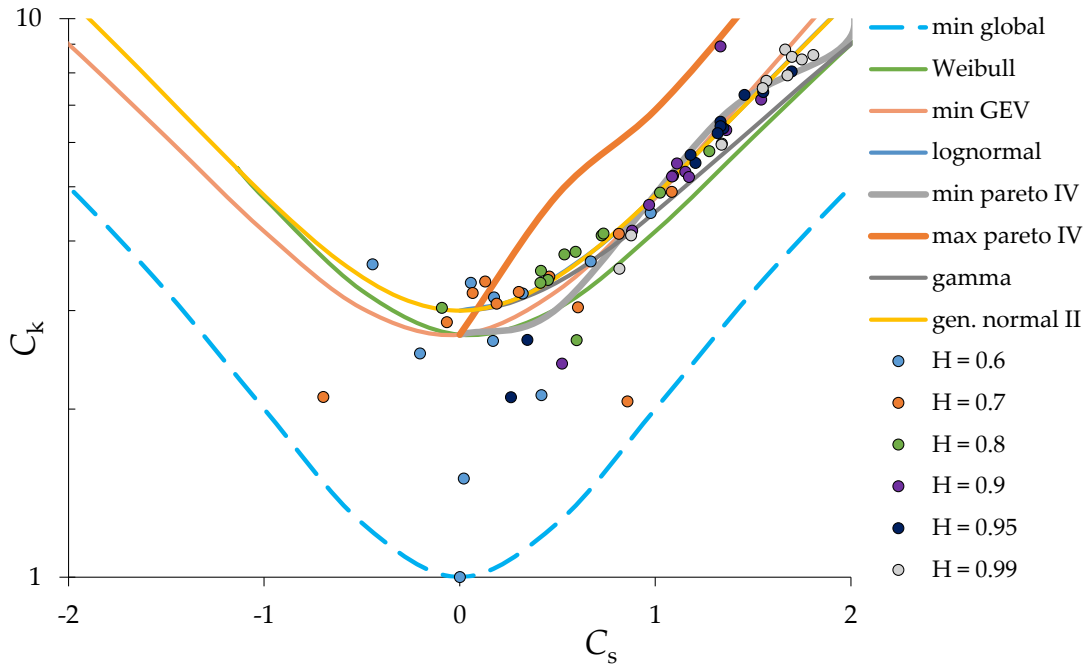
Σχήμα 32 Πιθανότερη τιμή κλιμακογράμματος συναρτήσει του μήκους δείγματος στην κλίμακα 1.



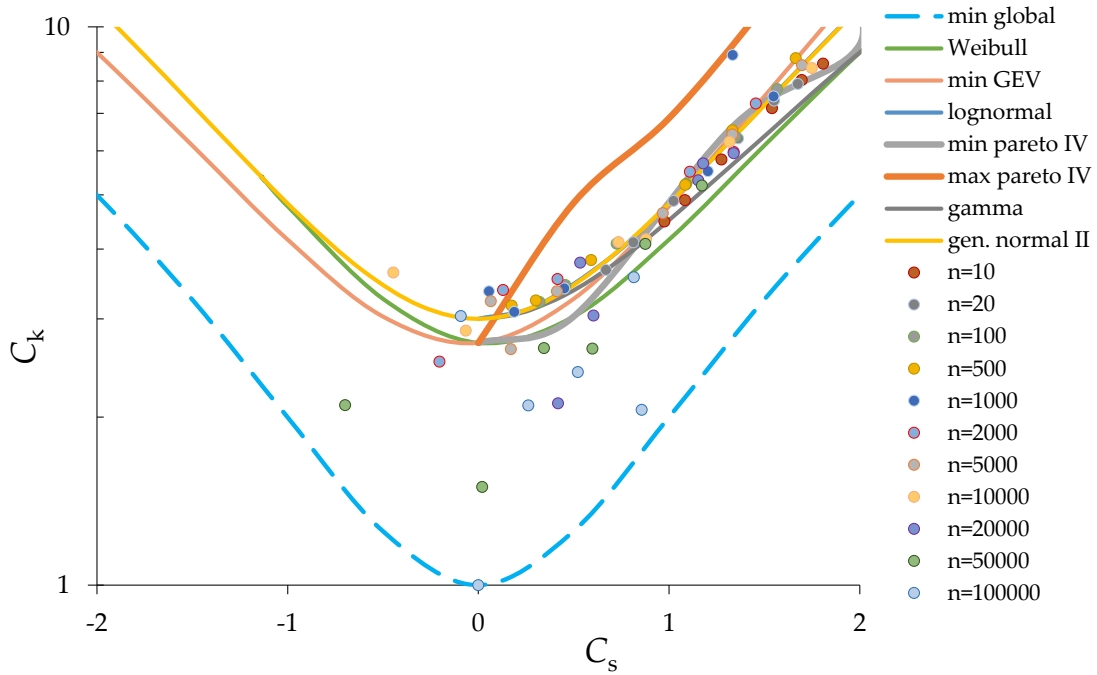
Σχήμα 33 Πιθανότερη τιμή κλιμακογράμματος συναρτήσει του μήκους δείγματος στην κλίμακα 0,1n.

### 3.2.2.5 Κύρτωση και Ασυμμετρία

Επειδή η κύρτωση παρουσιάζει αρκετές διακυμάνσεις, που σημαίνει ότι θα χρειαζόμασταν περισσότερες επαναλήψεις για να έχουμε την επιθυμητή ακρίβεια, παρουσιάζουμε το παρακάτω διάγραμμα στο οποίο εκφράζεται η κύρτωση συναρτήσει της ασυμμετρίας για διάφορες γνωστές κατανομές και πάνω στο διάγραμμα αυτό τοποθετούμε τα σημεία από την ανάλυση.



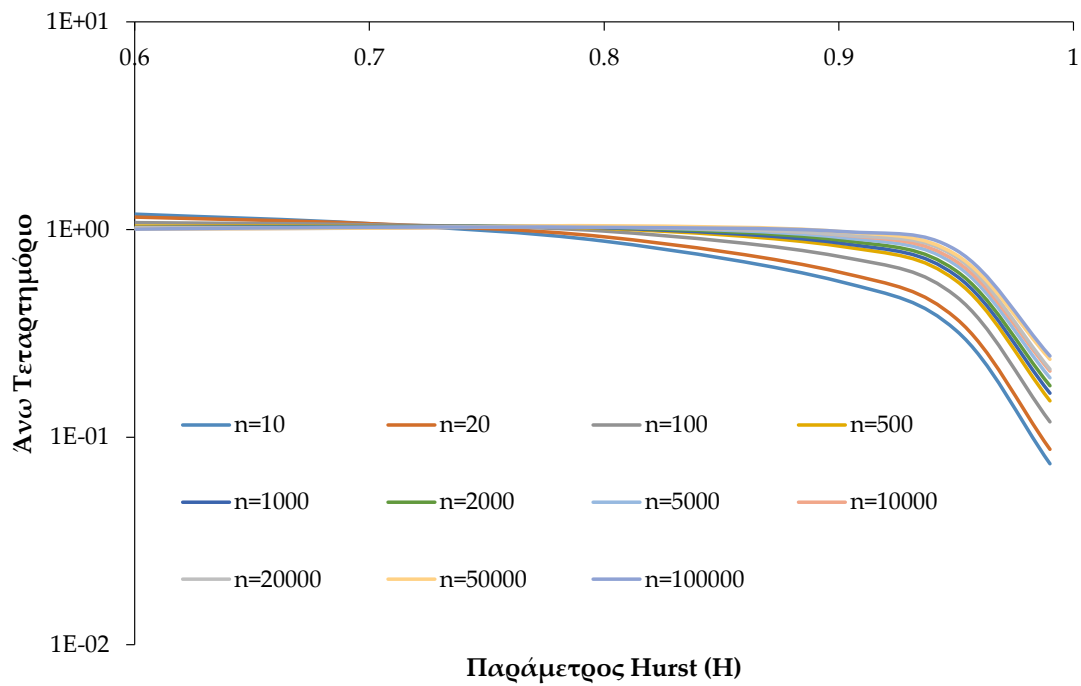
Σχήμα 34 Κύρτωση συναρτήσεσι ασυμμετρίας (H).



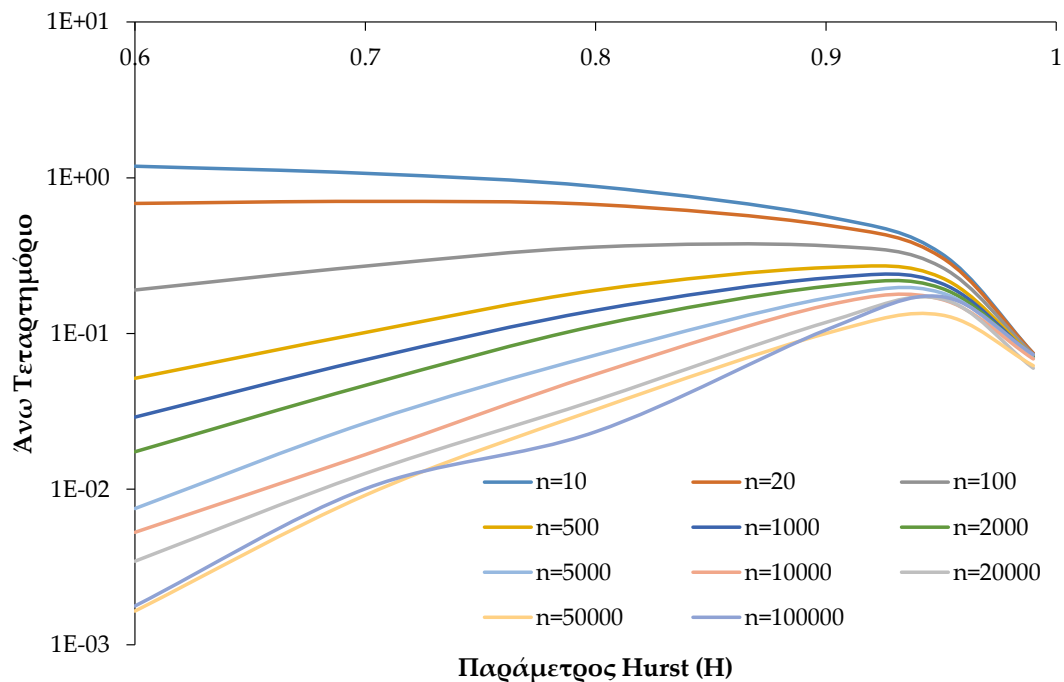
Σχήμα 35 Κύρτωση συναρτήσεσι ασυμμετρίας (n).

Τα σημεία θα περιμέναμε να ταυτίζονται με την καμπύλη της γάμμα κατανομής όπως στην περίπτωση του λευκού θορύβου κάτι που δεν ισχύει αλλά παρατηρείται ότι την προσεγγίζουν.

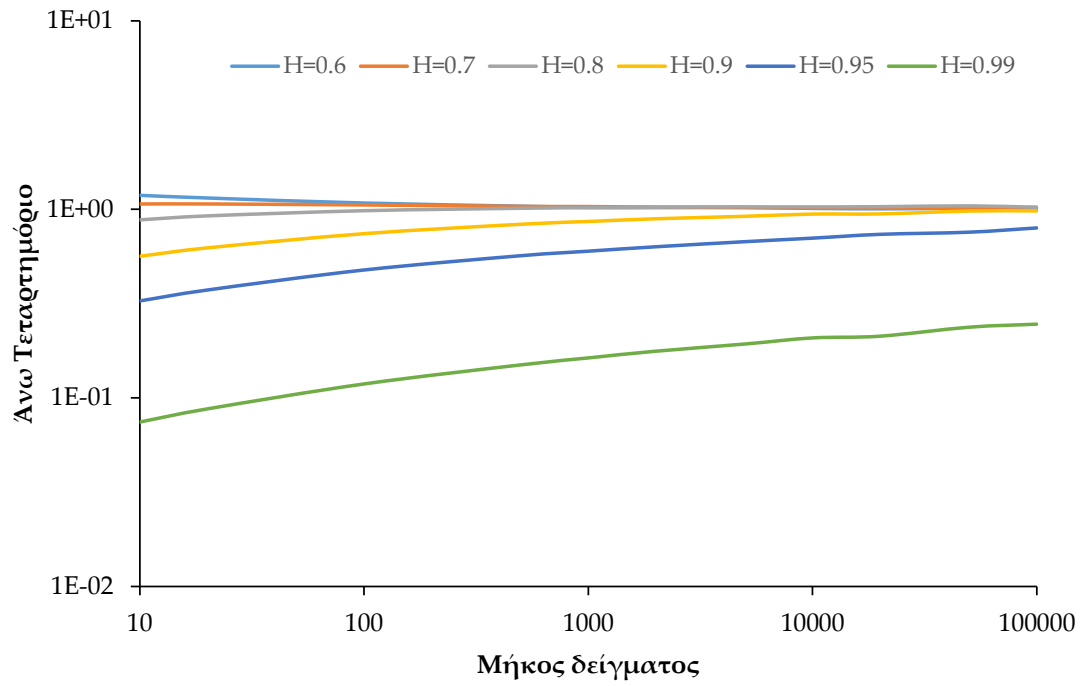
### 3.2.2.6 Άνω τεταρτημόριο



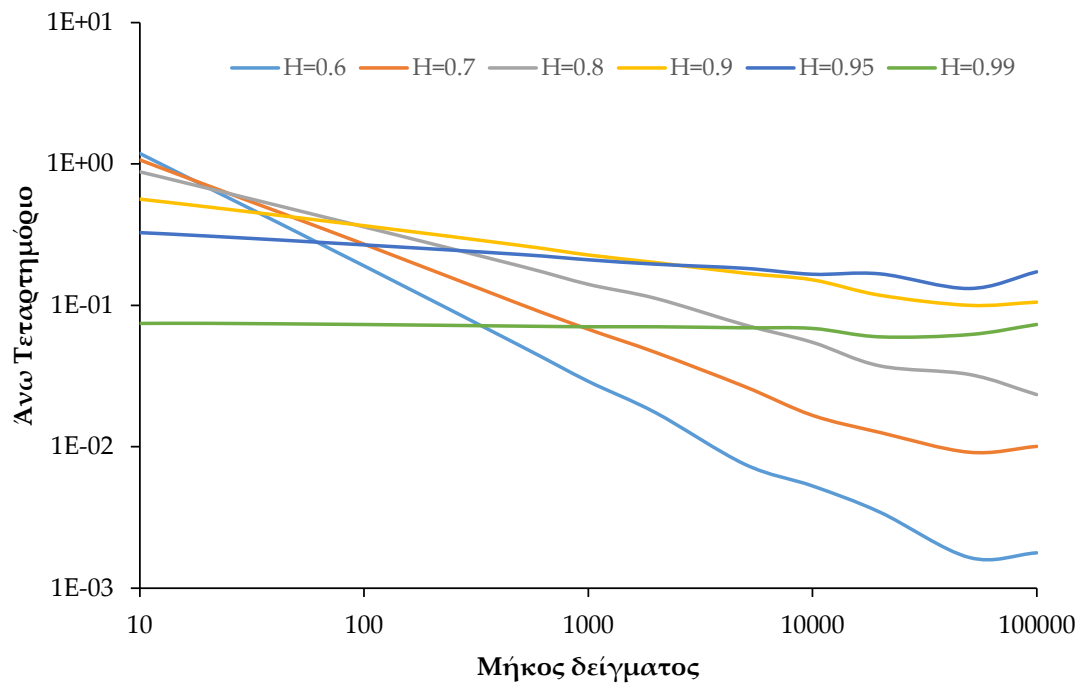
Σχήμα 36 Άνω τεταρτημόριο κλιμακογράμματος συναρτήσει της παραμέτρου  $H$  στην κλίμακα 1.



Σχήμα 37 Άνω τεταρτημόριο κλιμακογράμματος συναρτήσει της παραμέτρου  $H$  στην κλίμακα  $0,1n$ .

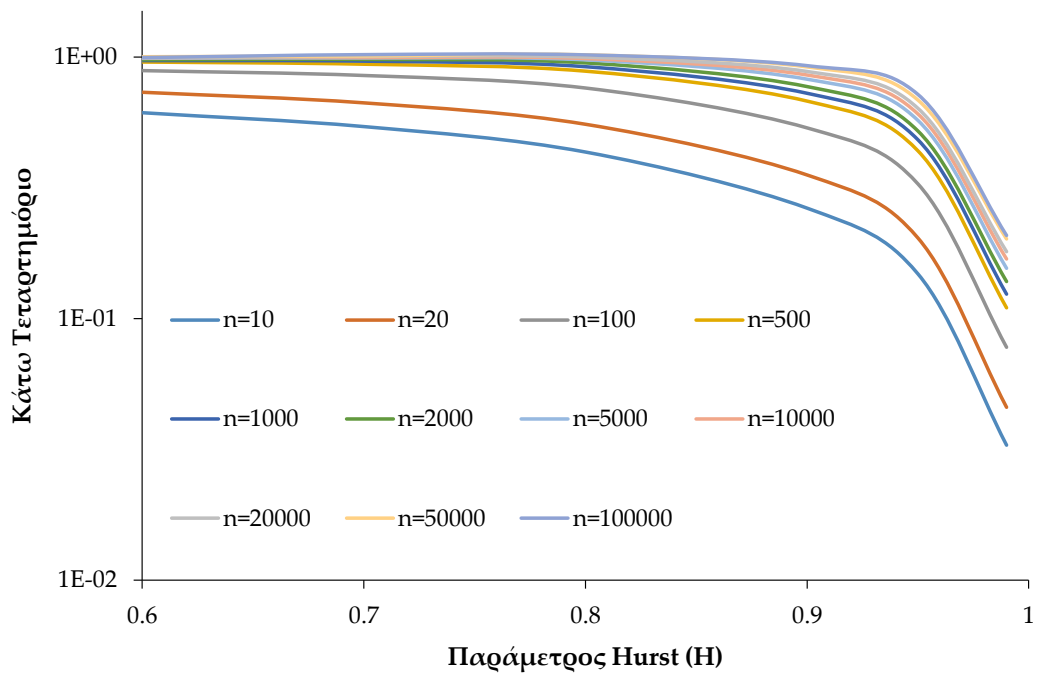


Σχήμα 38 Άνω τεταρτημόριο κλιμακογράμματος συναρτήσει του μήκους δείγματος  $n$  στην κλίμακα 1.

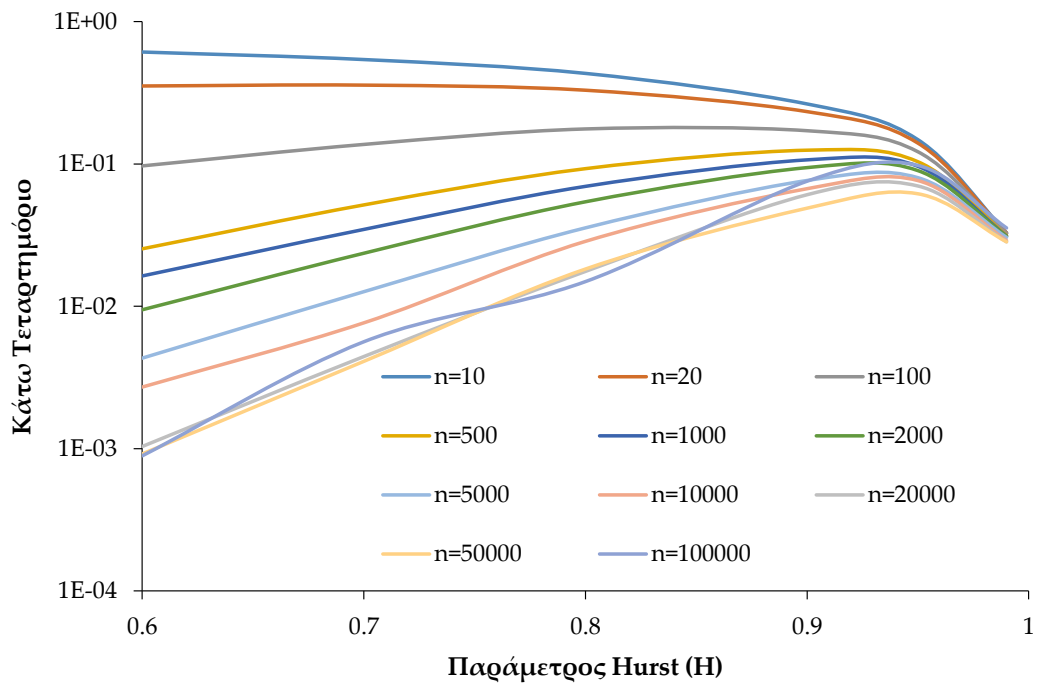


Σχήμα 39 Άνω τεταρτημόριο κλιμακογράμματος συναρτήσει του μήκους δείγματος  $n$  στην κλίμακα 0,1n.

### 3.2.2.7 Κάτω τεταρτημόριο

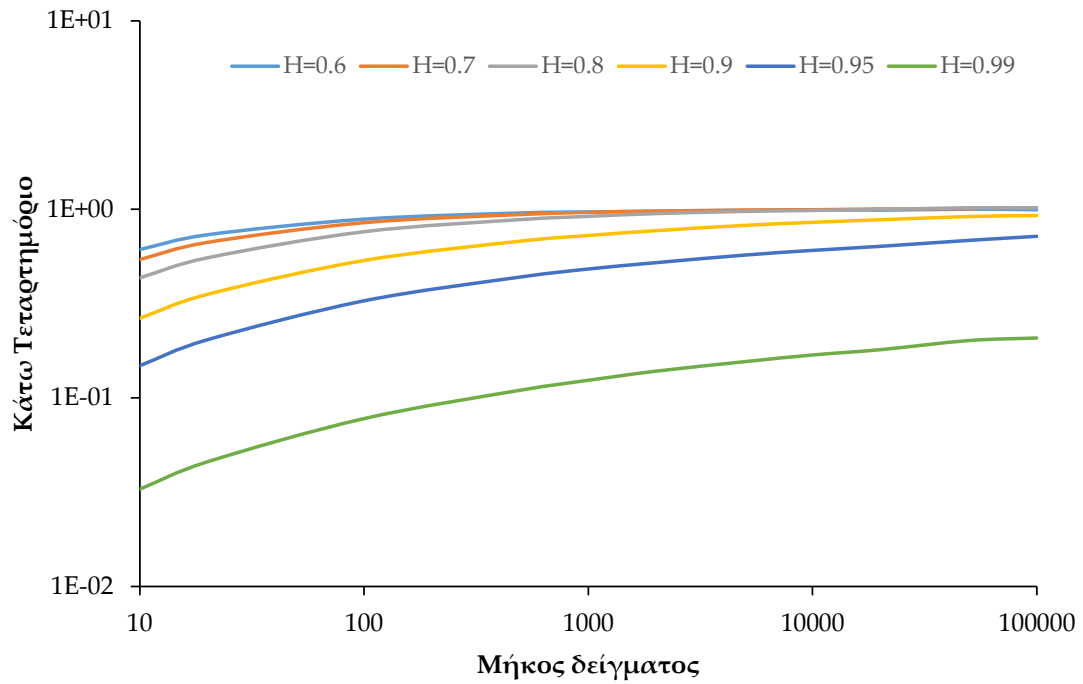


Σχήμα 40 Κάτω τεταρτημόριο κλιμακογράμματος συναρτήσει της παραμέτρου  $H$  στην κλίμακα 1.

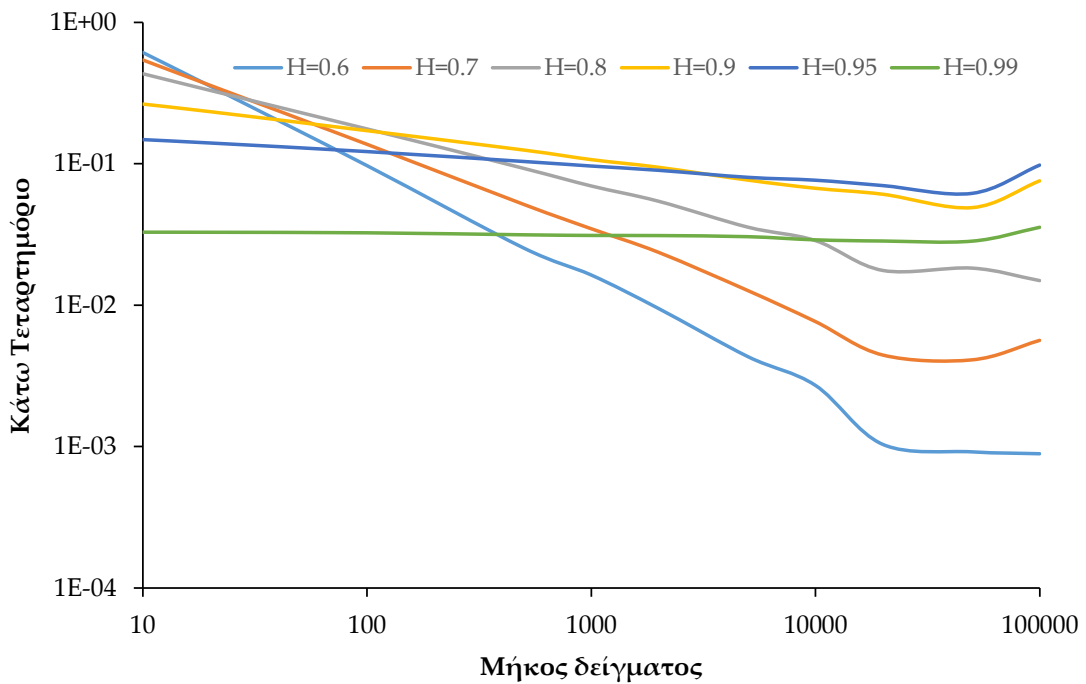


Σχήμα 41 Κάτω τεταρτημόριο κλιμακογράμματος συναρτήσει της παραμέτρου  $H$  στην κλίμακα  $0,1n$ .



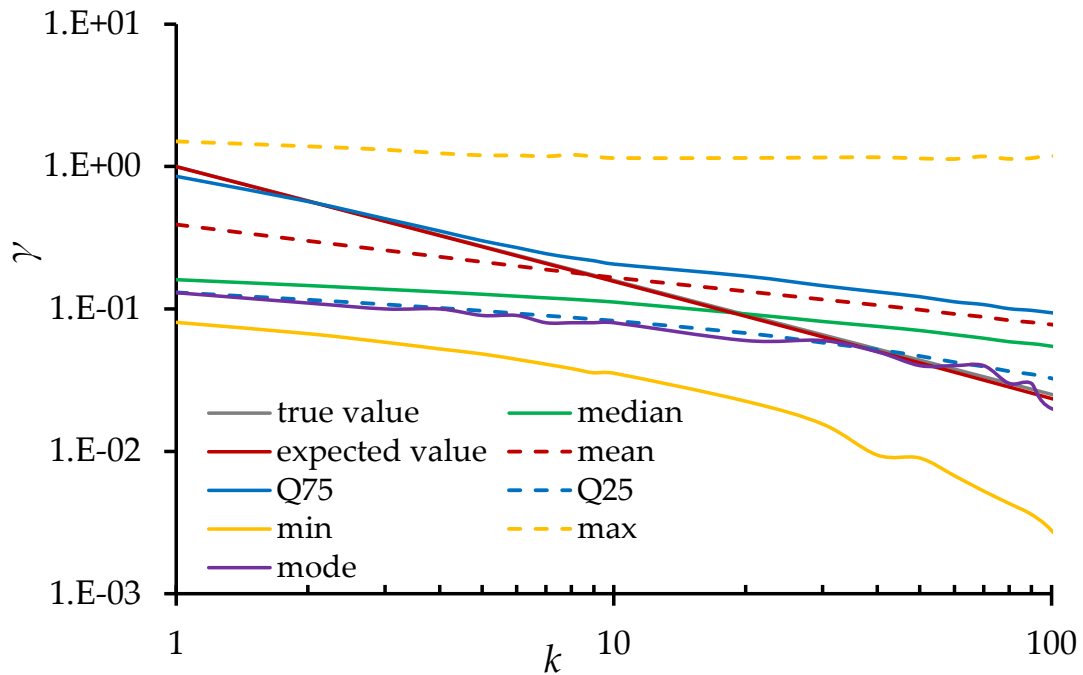


Σχήμα 42 Κάτω τεταρτημόριο κλιμακογράμματος συναρτήσει του μήκους δείγματος  $n$  στην κλίμακα 1.

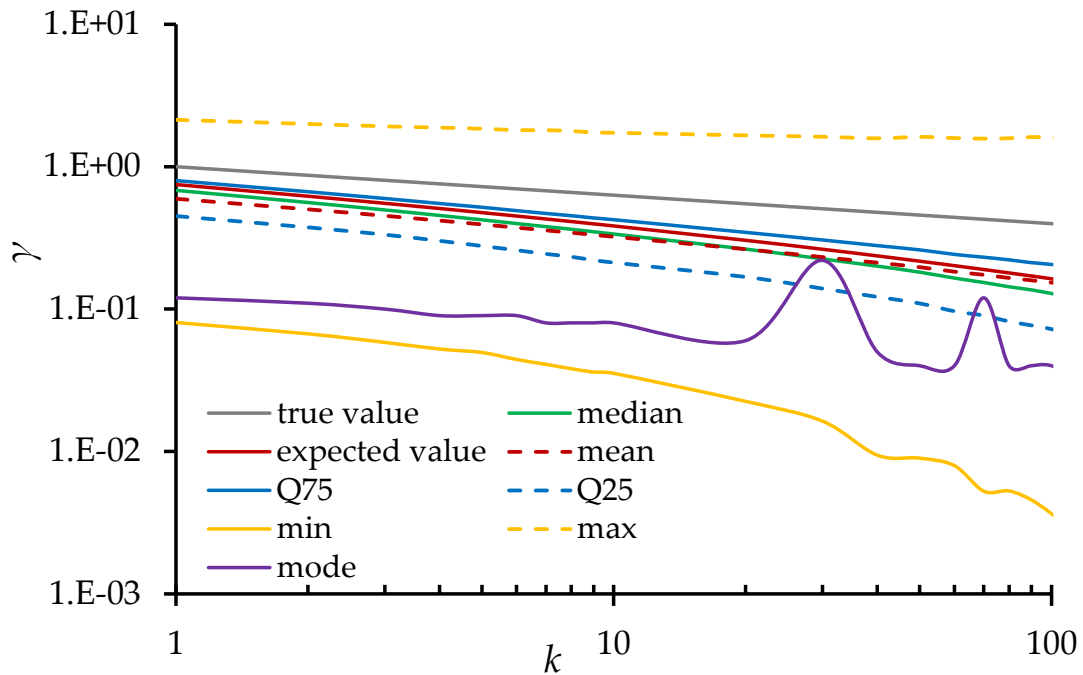


Σχήμα 43 Κάτω τεταρτημόριο κλιμακογράμματος συναρτήσει του μήκους δείγματος  $n$  στην κλίμακα 0,1n.

Επιπλέον, επειδή παρατηρείται από τα προηγούμενα διαγράμματα ότι κάποια μεγέθη έχουν ίδιες συμπεριφορές με κάποια άλλα, ακολουθούν τα διαγράμματα ενδεικτικά για  $H = 0.6$  και  $H = 0.9$  για μήκος δείγματος  $n = 1000$  παρουσιάζοντας τα στατιστικά χαρακτηριστικά των κλιμακογραμμάτων σε κάθε κλίμακα.

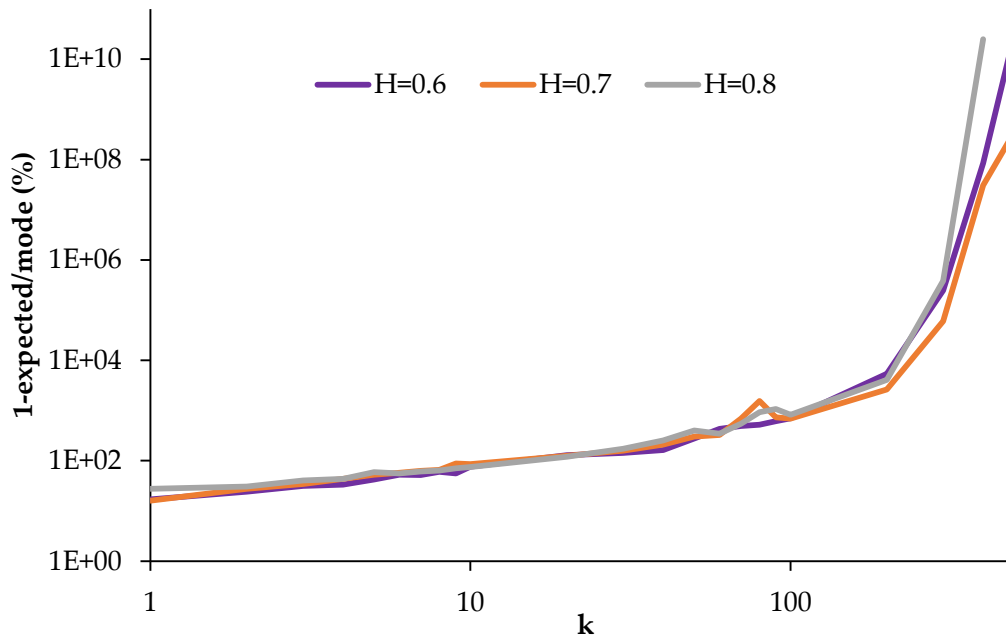


Σχήμα 44 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος ( $H=0.6$ ).

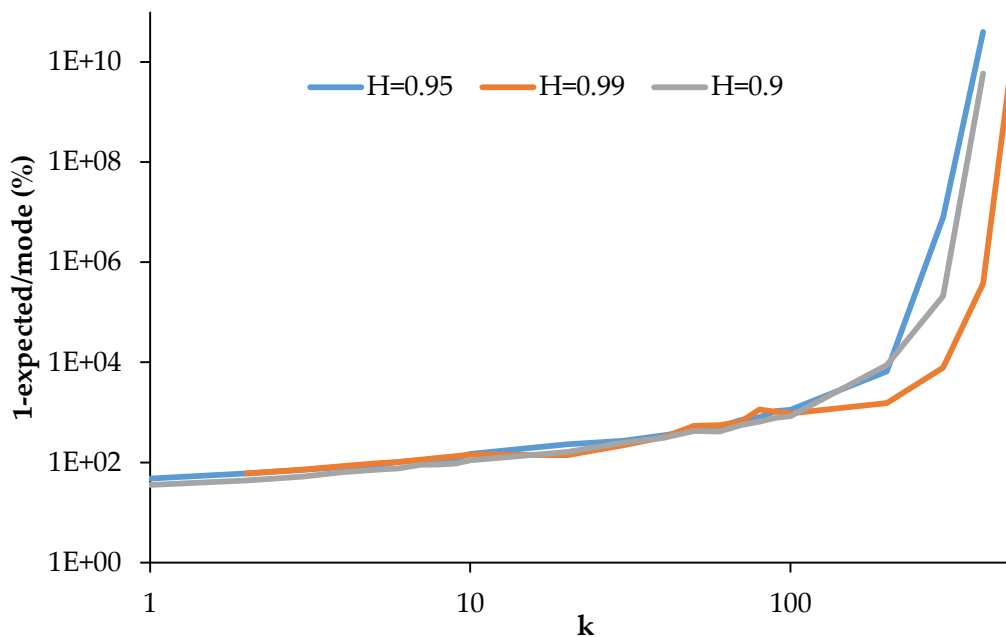


Σχήμα 45 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος ( $H=0.9$ ).

Παρατηρούμε ότι η πιο πιθανή τιμή πέφτει πολύ κοντά στο κάτω τεταρτημόριο για όλες σχεδόν τις κλίμακες. Επίσης η διάμεσος είναι πολύ κοντά στην μέση τιμή. Όσο αυξάνεται η ασυμμετρία τόσο αυξάνεται και η απόκλιση της μέσης τιμής από την πιθανότερη τιμή. Αυτό αποδुकνείται και από τα ακόλουθα διαγράμματα που είναι φανερό ότι ανεξάρτητα από την παράμετρο Hurst η διαφορά μεταξύ των δύο τιμών αυξάνεται με την αύξηση της κλίμακας.



Σχήμα 46 Διάγραμμα απόκλισης μέσης και πιθανότερης τιμής ( $H=0.6, 0.7, 0.8$ ).



Σχήμα 47 Διάγραμμα απόκλισης μέσης και πιθανότερης τιμής ( $H=0.9, 0.95, 0.99$ ).



## 4 ΕΦΑΡΜΟΓΕΣ

---

Στα πλαίσια του μαθήματος «Στοχαστικές Μέθοδοι στους Υδατικούς Πόρους» έχουν γίνει αναλύσεις που αφορούν την μακροπρόθεσμη εμμονή και τα αποτελέσματά τους παρουσιάστηκαν στο συνέδριο E.G.U. Συγκεκριμένα, τον Απρίλιο του 2016, μελετήθηκε η εξακρίβωση της συμπεριφοράς των μοντέλων Markov και Hurst-Kolmogorov στις υδρολογικές διεργασίες και τον Απρίλιο του 2017 μελετήθηκαν τα παράγωγα καιρού όπου και στις δύο περιπτώσεις χρησιμοποιήθηκε το κλιμακόγραμμα ως στοχαστικό εργαλείο.

### 4.1 ΚΛΙΜΑΚΟΓΡΑΜΜΑ ΩΣ ΣΤΟΧΑΣΤΙΚΟ ΕΡΓΑΛΕΙΟ

Οι υδρολογικές διεργασίες συνήθως προσομοιώνονται με εκθετικού τύπου φθίνουσα συνάρτηση αυτοσυσχέτισης, συμπεριφορά Markov, ή δυναμικού τύπου (μακροπρόθεσμη εμμονή), Hurst-Kolmogorov συμπεριφορά. Όπως έχει ήδη αναφερθεί για την ανέλιξη Markov το μέλλον εξαρτάται μόνο από το παρόν ενώ στην περίπτωση ανέλιξης HK εξαρτάται και από το παρόν και από το παρελθόν.

Αυτές οι δύο ανελιξεις εξαρτώνται από μια παράμετρο, την αυτοσυσχέτιση με υστέρηση 1 και τον συντελεστή Hurst αντίστοιχα. Ενώ μπορεί να φαίνεται απλό, είναι πολλές φορές δύσκολη η επιλογή της ανέλιξης που περιγράφει καλύτερα την ζητούμενη στοχαστική δομή αυτοσυσχέτισης. Ειδικά στις υδρολογικές διεργασίες όπου έχουμε περιορισμένο αριθμό δεδομένων η παραπάνω εξακρίβωση γίνεται ακόμα πιο δύσκολη και μερικές φορές στατιστικά αδύνατη.

Για την ταυτοποίηση και ποσοτικοποίηση τέτοιων συμπεριφορών χρησιμοποιήθηκε το κλιμακόγραμμα για τους λόγους που έχουν ήδη αναφερθεί.

#### 4.1.1 Μεθοδολογία

Εξετάστηκαν τρεις διαφορετικές περιπτώσεις. Για κάθε περίπτωση παρήχθησαν 10000 συνθετικές χρονοσειρές με  $n = 200$  μήκος δείγματος για κάθε ανέλιξη Markov ( $q = 0.5, 1, 2, 5, 10$ ), για τον λευκό θόρυβο ( $H = 0.5$ ) και για κάθε HK ανέλιξη ( $H = 0.6, 0.7, 0.8, 0.9$ ). Για την σύνθεση των χρονοσειρών απλής ομοιοθετικής ανέλιξης χρησιμοποιήθηκε η τεχνική  $3AR(1)$  (Koutsoyiannis, 2002).

Αυτή η τεχνική χρησιμοποιεί την λογική των τυχαίων διακυμάνσεων πολλαπλής κλίμακας. Το άθροισμα δηλαδή τριών διαφορετικών ανελιξων Markov (AR(1)) ορίζεται ως εξής:

$$X_i = A_i + B_i + C_i$$

Οι συντελεστές αυτοσυσχέτισης των τριών ανελιξων για υστέρηση 1 δίνονται παρακάτω:

$$\begin{aligned}\rho_A &= 1,52(H - 0,5)^{1,32} \\ \rho_B &= 0,953 - 7,69(1 - H)^{3,85} \\ \rho_C &= \begin{cases} 0,932 + 0,087H, & H \leq 0,76 \\ 0,993 + 0,007H, & H > 0,76 \end{cases}\end{aligned}$$

Οι διασπορές των τριών ανελιξων δίνονται παρακάτω:

$$\lambda_A = (1 - c_1 - c_2)\gamma_0$$

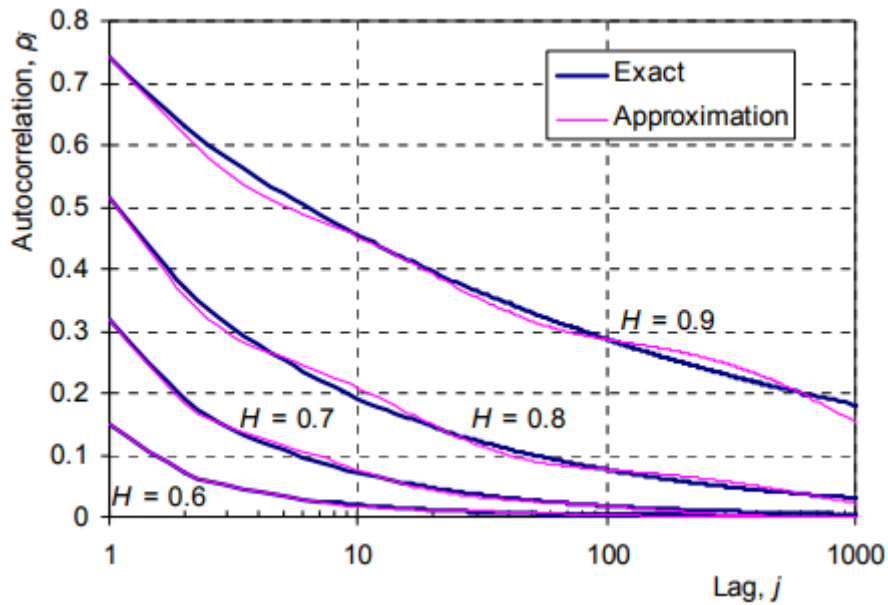
$$\lambda_B = c_1\gamma_0$$

$$\lambda_C = c_2\gamma_0$$

Οι συντελεστές  $c_1$ ,  $c_2$  υπολογίζονται με τέτοιο τρόπο ώστε η συνάρτηση αυτοσυσχέτισης του αθροίσματος των τριών ανελιξων να ταυτίζεται με την θεωρητική αυτοσυσχέτιση της ανέλιξης απλής ομοιοθεσίας για υστέρηση 1 και 100. Τελικά η συνάρτηση αυτοσυσχέτισης του αθροίσματος είναι:

$$\rho_j = (1 - c_1 - c_2)\rho_A + c_1\rho_B^j + c_2\rho_C^j$$

Στην συγκεκριμένη εργασία χρησιμοποιούμε το μοντέλο 3AR(1) γιατί θέλαμε να κάνουμε πολλές επαναλήψεις σε σύντομο χρονικό διάστημα και για εξικονόμηση χρόνου χρησιμοποιήσαμε αυτό από την θεωρητική λύση/μοντέλο που είναι το FGN.



Σχήμα 48 Η συνάρτηση αυτοσυσχέτισης του μοντέλου  $3AR(1)$  σε σύγκριση με την ακριβή αυτοσυσχέτιση του μοντέλου  $FGN$  για διάφορους συντελεστές Hurst (Koutsoyiannis, 2002)

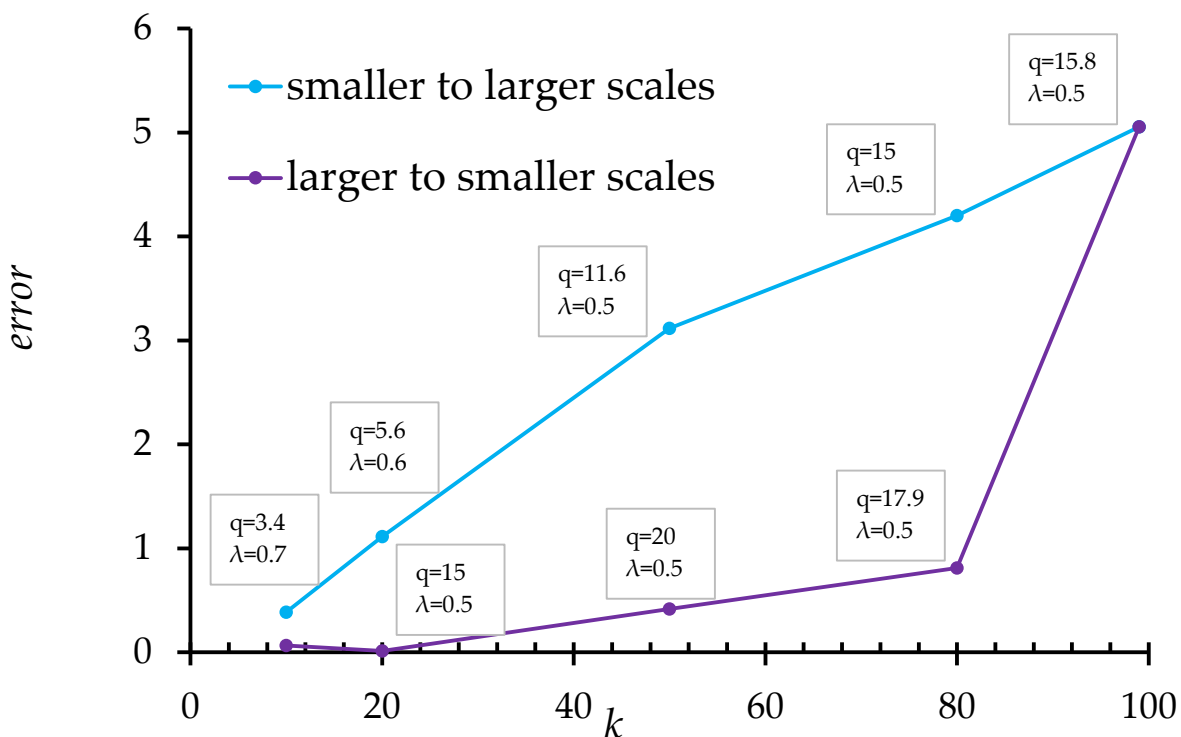
Αρχικά, μελετάμε την περίπτωση που έχουμε πολλές χρονοσειρές ίδιου μήκους, για παράδειγμα όταν κάνουμε επαναλαμβανόμενα πειράματα με σταθερές αρχικές συνθήκες. Για αυτή την περίπτωση η ανάλυση βασίζεται στην αναμενόμενη τιμή της ανέλιξης κάνοντας δύο δοκιμές. Μια για να δώσουμε έμφαση στην διαφορά μεταξύ των αναμενόμενων τιμών των δύο ανελίξεων και μια για τα όρια εμπιστοσύνης τους.

Η δεύτερη περίπτωση αφορά χρονοσειρές με διαφορετικό μήκος δείγματος, για παράδειγμα όταν έχουμε δεδομένα από διάφορους υδρομετεωρολογικούς σταθμούς παγκοσμίως με διαφορετική περίοδο παρατηρήσεων. Εδώ κάνουμε τις ίδιες δοκιμές αλλά προσαρμόζοντας ένα διορθωτικό παράγοντα για την αμεροληψία σε κάθε εμπειρικό κλιμακόγραμμα. Συγκεκριμένα το πολλαπλασιάζουμε με την αναλογία του εμπειρικού κλιμακογράμματος για το μήκος της αρχικής χρονοσειράς και της επιθυμητής.

Τέλος, στην τρίτη περίπτωση έχουμε μόνο μια χρονοσειρά και εξετάζουμε μια δοκιμή βασισμένη στην πιο πιθανή τιμή του κλιμακογράμματος.

#### 4.1.2 Αποτελέσματα

Για την πρώτη δοκιμή, εξετάζουμε το σφάλμα που προκύπτει από την προσαρμογή μιας ανέλιξης Markov σε μια ανέλιξη HK (ενδεικτικά  $H = 0.9$ ) σε όρους αναμενόμενης τιμής κλιμακογράμματος. Δηλαδή, χρησιμοποιώντας την μέση τιμή του κλιμακογράμματος αναζητούμε τις παραμέτρους  $\lambda, q$  της αναμενόμενης τιμής της ανέλιξης Markov για τις οποίες έχουμε το ελάχιστο σφάλμα. Ξεκινάμε εξετάζοντας το σφάλμα αυτό για τις πρώτες 10, 20, 50 κ.ο.κ. κλίμακες και έπειτα για τις τελευταίες 10, 20, 50 κ.ο.κ.



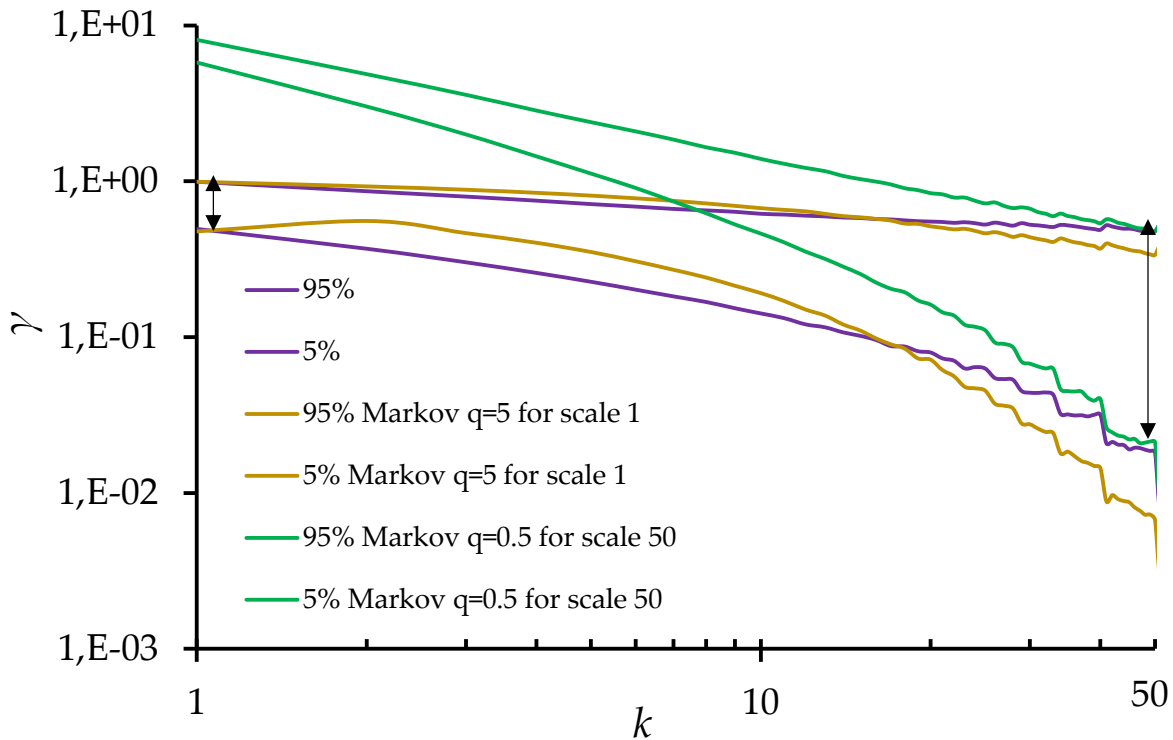
Σχήμα 49 Σφάλμα προσέγγισης του μέσου κλιμακογράμματος για HK ανέλιξη ( $H = 0.9$ ).

Παρατηρούμε ότι στις μικρές κλίμακες το σφάλμα προσαρμογής της ανέλιξης Markov στην ανέλιξη FGN είναι μικρό αλλά για μεγαλύτερες κλίμακες το σφάλμα γίνεται αρκετά σημαντικό.

Στην περίπτωση που έχουμε περιορισμένες κλίμακες μπορούμε να κάνουμε και μια δεύτερη δοκιμή που αφορά τα όρια εμπιστοσύνης των ανελιξεων. Γίνεται μια εκτίμηση για την προσέγγιση των ορίων εμπιστοσύνης (95% και 5%) του εμπειρικού

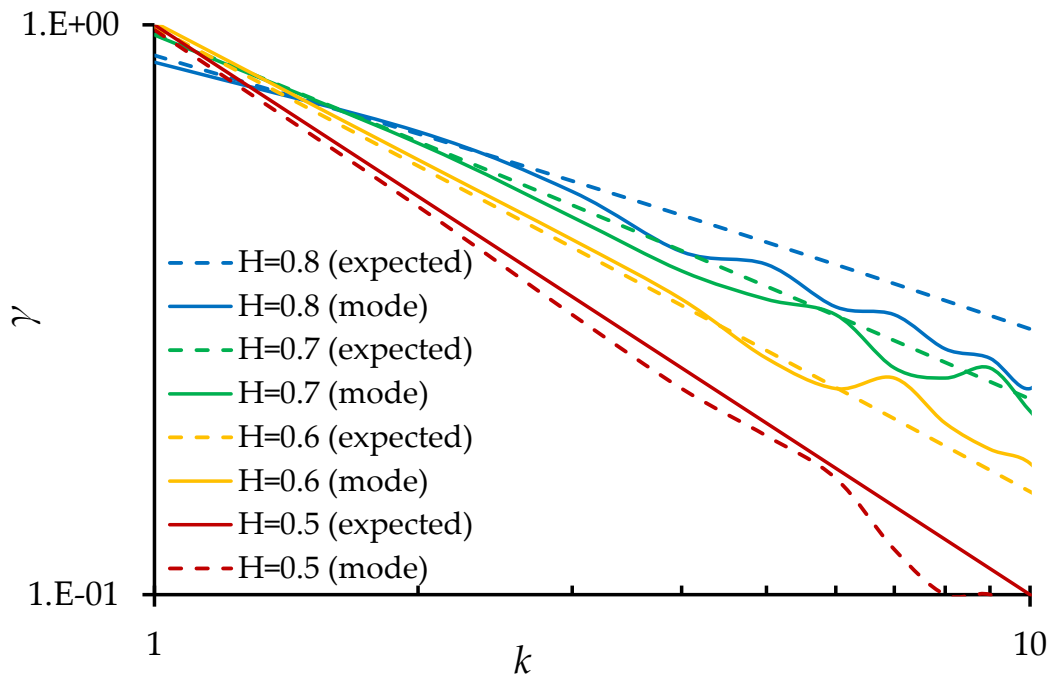


κλιμακογράμματος από την κάθε ανέλιξη. Στην περίπτωση των ορίων εμπιστοσύνης του εμπειρικού κλιμακογράμματος μιας FGN ανέλιξης (με  $H = 0,9$  για παράδειγμα), τα όρια εμπιστοσύνης της ανέλιξης Markov με παράμετρο  $q = 5$  προσεγγίζουν την FGN ανέλιξη στην κλίμακα 1 ενώ με παράμετρο  $q = 0.5$  την κλίμακα 50.



Σχήμα 50 Εύρος των ορίων εμπιστοσύνης για HK ανέλιξη ( $H=0,9$ ).

Στην τρίτη περίπτωση όπου έχουμε μια μόνο χρονοσειρά μπορούν να γίνουν οι ίδιες δοκιμές όπως αναφέρθηκαν παραπάνω μόνο που αντί να χρησιμοποιήσουμε την αναμενόμενη τιμή του κλιμακογράμματος, χρησιμοποιούμε την πιο πιθανή τιμή. Επειδή η κατανομή του κλιμακογράμματος είναι ασύμμετρη, η πιο πιθανή τιμή και η αναμενόμενη τιμή έχουν διαφορά μεταξύ τους (δεν ταυτίζονται) και μπορούν να οδηγήσουν σε διαφορετικά αποτελέσματα.



Σχήμα 51 Η διαφορά αναμενόμενης και πιο πιθανής τιμής για διάφορες παραμέτρους Hurst.

## 4.2 ΠΑΡΑΓΩΓΑ ΚΑΙΡΟΥ

Τα παράγωγα καιρού είναι ένα χρηματοοικονομικό εργαλείο που μπορεί να χρησιμοποιηθεί από οργανισμούς ή και ιδιώτες σαν μέρος μιας στρατηγικής διαχείρισης κινδύνου που σχετίζεται με τις αναπάντεχες και δυσμενείς επιπτώσεις που έχουν τα φαινόμενα του καιρού. Τα χρηματοοικονομικά παράγωγα είναι χρηματιστηριακές συμφωνίες (συμβόλαια) μιας μελλοντικής συναλλαγής, της οποίας η τιμή εξαρτάται από την τρέχουσα τιμή (spot) ενός υποκείμενου εμπορεύματος. Αφορούν, δηλαδή, στην προσδοκόμενη μελλοντική τιμή. Το πρώτο ιστορικά καταγεγραμμένο παράγωγο συμβόλαιο θεωρείται ότι συνήφθη από τον Θαλή τον Μιλήσιο (624-546 π.Χ.). Ο κίνδυνος από τις επιπτώσεις του καιρού μπορεί να ορισθεί ως η αβεβαιότητα που προκαλείται στα έσοδα, τα κόστη και τις χρηματικές ροές μια επιχείρησης από την μεταβλητότητα του καιρού. Ο καιρός αποτελεί ένα σημαντικό παράγοντα για την απόδοση μιας επιχείρησης. Οι καιρικές μεταβολές επηρεάζουν την τιμή ενός εμπορεύματος επηρεάζοντας έτσι την ζήτησή του π.χ. πετρέλαιο θέρμανσης, προϊόντα, φυσικό αέριο, ηλεκτρική ενέργεια.

Η διαφορά τους με την ασφάλιση, είναι ότι τα ασφαλιστικά συμβόλαια καλύπτουν υψηλού κινδύνου, χαμηλής συχνότητας γεγονότα (όπως τυφώνες, πλημμύρες) ενώ τα παράγωγα καιρού προσφέρουν προστασία από γεγονότα χαμηλότερου κινδύνου και υψηλής συχνότητας (ήπιος χειμώνες, βροχόπτωση, χιόνι). Η ασφαλιστική βιομηχανία αρνείται να καλύψει μικρούς κινδύνους που απλά δημιουργούν μειωμένα έσοδα. Η διαφορά που έχουν από τα κλασικά χρηματοοικονομικά παράγωγα είναι ότι τα δεύτερα έχουν σαν υποκείμενο δείκτη ένα διαπραγματεύσιμο στοιχείο (ή την τιμή του) όπως είναι μετοχές, δείκτες μετοχών, ομόλογα, συναλλαγματικές ισοτιμίες, τιμές εμπορευμάτων. Τα παράγωγα καιρού χρησιμοποιούν ως υποκείμενα στοιχεία καιρικές παραμέτρους όπως θερμοκρασία, βροχόπτωση, άνεμος. Αποτελούν λοιπόν συμφωνίες συναλλαγών επί μετεωρολογικών δεικτών.

Οι αγορές όπου οι κίνδυνοι καιρού έχουν μεγάλη ανάπτυξη είναι οι μονάδες παραγωγής ενέργειας, αγροτικές επιχειρήσεις, τουριστικές επιχειρήσεις, κατασκευών, λιανικών πωλήσεων και τράπεζες. Συμπεραίνουμε ότι ο καιρός επηρεάζει τις τιμές και το ύψος της ζήτησης και τα παράγωγα καιρού συμβάλλουν στην διαχείριση (αντιστάθμιση) του κινδύνου και στον οικονομικό προγραμματισμό.

Τα παράγωγα καιρού είναι ένα εργαλείο που μπορεί να περιορίσει την μεταβλητότητα των αποτελεσμάτων των επιχειρήσεων καθώς αποδεδειγμένα ο καιρός επηρεάζει την συμπεριφορά τους. Παράγωγα καιρού θα μπορούσαν να χρησιμοποιηθούν για την κάλυψη απώλειας εσόδων γεωργικών επιχειρήσεων από την μείωση της γεωργικής παραγωγής λόγω δυσμενών καιρικών συνθηκών. Όσο αφορά τα εποχικά προϊόντα, των οποίων οι πωλήσεις είναι στενά συνδεδεμένες με τις καιρικές συνθήκες, η αξιοποίηση των παραγώγων καιρού μπορούν να εξομαλύνουν τις πωλήσεις τους μέσα στο έτος. Έτσι επιτυγχάνεται πιο συνεπής παραγωγή κατά την διάρκεια του έτους και βελτίωση των επιπέδων τήρησης αποθεμάτων.

Τα παράγωγα καιρού έχουν συνήθως την μορφή δικαιωμάτων προαίρεσης (option) το οποίο είναι ένα συμβόλαιο που παρέχει στον κάτοχό του το δικαίωμα, αλλά όχι και την υποχρέωση να αγοράσει ή να πωλήσει ένα προϊόν (μετοχή, ομόλογο ή εμπόρευμα) σε μία προκαθορισμένη χρονική στιγμή και σε μία προσυμφωνημένη τιμή.

#### 4.2.1 Μεθοδολογία

Για να υπολογίσουμε τις απολαβές των παραγώγων καιρού χρειαζόμαστε δύο μεγέθη: τα συναθροισμένα HDDs, CDDs και την τιμή εξάσκησης του συμβολαίου. Για κάθε μήνα υπολογίζουμε τους καθαρούς αθροιστικούς βαθμούς θερμοκρασίας (Net Accumulated Degrees, NAD) σύμφωνα με την εξής σχέση:

$$NAD_i = \max(H_{MONTH} - C_{MONTH}; C_{MONTH} - H_{MONTH})$$

Ο πιο συχνός υποκείμενος δείκτης είναι η θερμοκρασία:

Βαθμοί υπολειπόμενης θερμοκρασίας (Heating Degree Days, HDDs): Αριθμός βαθμών που η μέση ημερήσια θερμοκρασία υπολείπεται των  $65^\circ\text{F} = (18,3^\circ\text{C})$ .

$$HDD_i \equiv \text{Max}(65^\circ\text{F} - \hat{t}_i; 0)$$

$$HDD_{MONTH} \equiv \sum_{i=1}^{n(M)} HDD_i$$

Βαθμοί υπερβάλλουσας θερμοκρασίας (Cooling Degree Days, HDDs): Αριθμός βαθμών που η μέση ημερήσια θερμοκρασία υπερβαίνει των  $65^\circ\text{F} = (18,3^\circ\text{C})$ .

$$CDD_i \equiv \text{Max}(\hat{t}_i - 65^\circ\text{F}; 0)$$

$$CDD_{MONTH} \equiv \sum_{i=1}^{n(M)} CDD_i$$

όπου  $18,3^\circ\text{C}$  το βιοκλιματικό βέλτιστο.

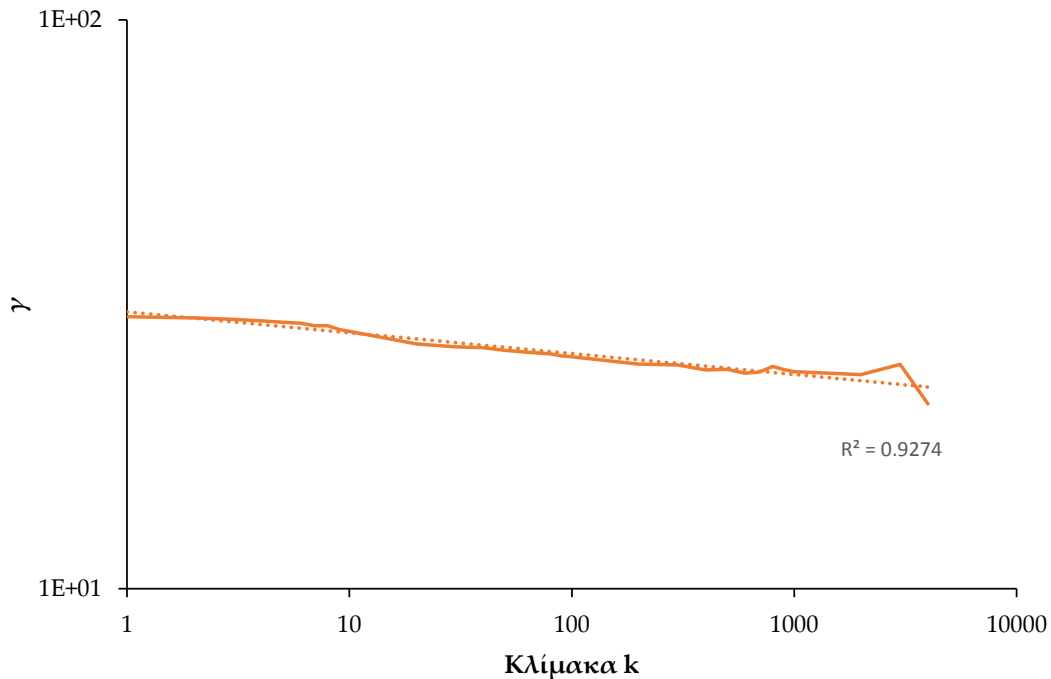
Η τιμή εξάσκησης (Strike Price) είναι η τιμή στην οποία ενεργοποιείται το συμβόλαιο (ανεξάρτητα από την τρέχουσα τιμή) και υπολογίζεται σε όρους βαθμών θερμοκρασίας. Η τελική απολαβή υπολογίζεται ως εξής:

$$\text{Payoff}_i = \text{Tick} \cdot \max(NAD_i - SP; 0)$$

θεωρώ  $\text{Tick} = 10 = \text{σταθερό}$ .

Για την τιμολόγηση των δικαιωμάτων προαίρεσης δεν μπορεί να εφαρμοσθεί το γνωστό μοντέλο Black-Scholes διότι προϋποθέτει την ύπαρξη ενός διαπραγματεύσιμου υποκειμένου, από την τιμή του οποίου παράγεται η τιμή του παραγώγου. Επειδή ο καιρός δεν μπορεί να τιμολογηθεί η προϋπόθεση αυτή δεν

ισχύει. Έτσι για την τιμολόγηση των παραγώγων καιρού συνήθως χρησιμοποιείται ένα μείγμα ιστορικών μετεωρολογικών στοιχείων και των γνωστών μοντέλων τιμολόγησης παραγώγων σε συνδυασμό με στοιχεία αναλογιστικών προβλέψεων (μαθηματικών και στατιστικών μοντέλων για την εκτίμηση των κινδύνων στον τομέα της ασφάλισης) του καιρού.



Σχήμα 52 Σχήμα 53 Κλιμακόγραμμα θερμοκρασίας Αστυπάλαιας.

Για αυτό τον λόγο κάνουμε Monte Carlo ανάλυση. Έχοντας ιστορικές χρονοσειρές θερμοκρασίας για δύο χρόνια, παράγουμε συνθετικές χρονοσειρές χρησιμοποιώντας το μοντέλο AR(1). Για κάθε μέρα υπολογίζουμε τους καθαρούς συναθροισμένους βαθμούς θερμοκρασίας για περίοδο 30 ημερών. Έπειτα υπολογίζουμε την απόδοση των συμβολαίων. Την διαδικασία αυτή την επαναλαμβάνουμε για αρκετές φορές για διαφορετικές τιμές εξάσκησης. Τέλος εκτιμούμε την αναμενόμενη απόδοση (απολαβή) ως τον μέσο όρο των πιθανών απολαβών για κάθε μέρα και την συγκρίνουμε με την πραγματική από την ιστορική χρονοσειρά.

Η στατιστική δομή μιας χρονοσειράς (θερμοκρασιών, βροχόπτωσης, ανέμου) έχει τεράστια σημασία για την επιτυχημένη σύναψη ενός συμβολαίου παραγώγου καιρού. Δεν αρκεί μονάχα η μελέτη των βασικών ροπών αλλά και η μελέτη ειδικών χαρακτηριστικών όπως η περιοδικότητα, η κυκλοστασιμότητα και η

μακροπρόθεσμη εμμονή. Συμπερασματικά, η χρήση μοντέλου HK θα ήταν πιο σωστή για να εκτιμήσουμε την ημερήσια θερμοκρασία με μεγαλύτερη βεβαιότητα. Όπως έχουμε ήδη αναφέρει μια υπέρθεση τριών ανεξάρτητων AR(1) με τα επιθυμητά χαρακτηριστικά πλησιάζει πολύ ικανοποιητικά το FGN μοντέλο, μπορούμε να πούμε ότι για μικρές κλίμακες μπορεί να χρησιμοποιηθεί το μοντέλο AR(1) για μια πρώτη προσέγγιση.

Έχουν γίνει αρκετές μελέτες για την επίδραση του φαινομένου Hurst στον κλάδο της οικονομίας και μακρο-οικονομίας όπως Orcutt and Irwin 1948; Ding, Granger, and Engle 1993; Lux 1996; Henry and Zaffaroni 2002; Cajueiro and Tabak 2004; Carbone, Castelli and Stanley 2004; Mikosch and Starica 2004; Grech and Pamula 2008.

## 5 ΣΥΜΠΕΡΑΣΜΑΤΑ

---

Η έννοια του λευκού θορύβου, όπως για παράδειγμα μια ακολουθία αποτελεσμάτων συνεχόμενων ρίψεων ενός ζαριού, μας είναι αρκετά οικεία. Υπό την υπόθεση ενός κλίματος που δεν αλλάζει, ακραία υδρολογικά φαινόμενα μπορούν να θεωρηθούν ως λευκός θόρυβος αφού δεν υπάρχει εξάρτηση μεταξύ των φαινομένων αυτών στα υδρολογικά έτη. Οι ανελίξεις AR(1) που έχουν κάποια εξάρτηση/ μνήμη είναι και αυτές σχετικά γνωστές καθώς δεν είναι δύσκολο να περιγραφούν.

Το φαινόμενο *Hurst* και η FGN ανέλιξη είναι πιο δυσνόητο. Οι Mesa και Poveda (1993) κατατάσσουν το φαινόμενο *Hurst* ως το πιο σημαντικό άλυτο πρόβλημα της υδρολογίας. Το FGN μοντέλο είναι πολύ διαφορετικό από το AR(1) λόγω της ισχυρής αυτοσυσχέτισης ακόμα και σε μεγάλες κλίμακες. Για παράδειγμα, στον ποταμό Νείλο αν ο συντελεστής *Hurst* είναι 0.85, τότε η αυτοσυσχέτιση για υστέρηση 100 (χρόνων) θα ισούται με 0.15 ενώ στην περίπτωση μοντέλου Markov η αυτοσυσχέτιση θα ήταν μηδέν ακόμα και για υστερήσεις 10 φορές μικρότερες.

Η μακροπρόθεσμη εμμονή παρουσιάζεται σε μια χρονοσειρά όταν οι αυτοσυσχετίσεις τείνουν στο μηδέν με πολύ αργό ρυθμό. Οι επιπτώσεις της ύπαρξης μακροπρόθεσμης εμμονής αφορούν την στατιστική εκτίμηση. Το φαινόμενο *Hurst* επιφέρει αύξηση των διαστημάτων εμπιστοσύνης στην εκτίμηση των στατιστικών χαρακτηριστικών όπως οι παράμετροι θέσης (μέση τιμή) και οι παράμετροι διασποράς (τυπική απόκλιση). Αξίζει να τονίσουμε ότι στην περίπτωση της τυπικής απόκλισης και διασποράς παρατηρείται μεγάλη αρνητική μεροληψία.

### 5.1 ΣΥΝΟΨΗ ΕΡΓΑΣΙΑΣ

Σκοπός της παρούσας διπλωματικής ήταν η διερεύνηση της κατανομής του κλιμακογράμματος ειδικότερα για εφαρμογή σε ανελίξεις μακροπρόθεσμης εμμονής. Αρχικά, μελετήθηκε η περίπτωση λευκού θορύβου που αντιστοιχεί σε παράμετρο *Hurst*  $H = 0.5$ . Στην περίπτωση αυτή τα δεδομένα έχουν μηδενική αυτοσυσχέτιση και έτσι είναι ανεξάρτητα. Όμως καθώς αυξάνεται η τιμή της παραμέτρου *Hurst* παρατηρείται ότι τα δεδομένα έχουν κάποια εξάρτηση μεταξύ τους και πλέον οι νόμοι της κλασικής στατιστικής δεν ισχύουν.

Το στοχαστικό μοντέλο που χρησιμοποιήθηκε για την παραγωγή συνθετικών χρονοσειρών είναι το SMA που διατηρεί την μακροπρόθεσμη εμμονή. Οι συνθετικές χρονοσειρές ακολουθούν κανονική κατανομή (Gaussian). Επιλέχθηκε αυτή η κατανομή επειδή στον τομέα της υδρολογίας και όλες οι μελέτες που αφορούν ετήσιες διεργασίες έχει παρατηρηθεί ότι ακολουθούν κανονική κατανομή. Για να μπορέσουμε να διεξάγουμε κάποια συμπεράσματα, επιλέχθηκε η μέθοδος του κλιμακογράμματος (συναθροισμένης διασποράς) αφού συγκρινόμενη με άλλα στοχαστικά εργαλεία για την εύρεση της δομής αυτοσυσχέτισης εμφανίζουν περισσότερη αβεβαιότητα. Επίσης χρησιμοποιήθηκε το λογισμικό Matlab που προσφέρει ένα εύκολο και γρήγορο τρόπο για να γίνουν αναλύσεις Monte Carlo. Η ανάλυση αυτή μας δίνει την δυνατότητα να επαναλάβουμε πολλές φορές το ίδιο πείραμα ώστε τα συμπεράσματα που θα βγάλουμε να είναι ακριβή. Οπότε για διάφορες τιμές του μήκους δείγματος και της παραμέτρου Hurst έγιναν για τους συνδυασμούς που προκύπτουν τα κλιμακογράμματα. Επομένως, τα αποτελέσματα της Monte Carlo ανάλυσης αφορούν την κατανομή του κλιμακογράμματος όπως και τα στατιστικά χαρακτηριστικά του σε δυο χαρακτηριστικές κλίμακες, την βασική  $k = 1$  και  $k = 0.1n$ .

## 5.2 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην περίπτωση λευκού θορύβου ( $H = 0,5$ ) προκύπτει ότι το κλιμακογράμμα στην βασική του κλίμακα ( $k = 1$ ) πολλαπλασιασμένο με  $\frac{(n-1)}{\sigma^2}$  ακολουθεί κατανομή  $\chi^2$  με  $n - 1$  βαθμούς ελευθερίας (λόγω της δειγματικής διασποράς). Η κατανομή διατηρεί και τις τέσσερις ροπές. Λόγω της σχέσης  $\sigma^{2(k)} = \sigma^2/k$  τότε οι βαθμοί ελευθερίας για  $k > 1$  υπολογίζονται από την σχέση:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^{2(k)} \cdot k} = \frac{n-1}{k} \cdot \frac{S^2}{\sigma^{2(k)}} \sim \chi^2_{\left(\frac{n-1}{k}\right)}$$

Στην περίπτωση FGN ανέλιξης φαίνεται να υπάρχει εξάρτηση μεταξύ των τιμών γιατί πλέον η κατανομή  $\chi^2$  δεν περιγράφει την συμπεριφορά των δεδομένων και παρατηρείται ότι η τρίτη και τέταρτη ροπή δεν προσεγγίζεται για  $H > 0.5$ . Η προσπάθεια προσέγγισης χρησιμοποιώντας την gamma κατανομή έδειξε ότι η



κατανομή δεν προσεγγίζει την κύρτωση του κλιμακογράμματος και όσο αυξάνει η παράμετρος Hurst οι διαφορές είναι μεγαλύτερες.

Τα αποτελέσματα της Monte Carlo ανάλυσης έδειξαν ότι η μέση τιμή του κλιμακογράμματος για όλα τα μήκη δείγματος  $n$  παρουσιάζει την ίδια συμπεριφορά δηλαδή ξεκινάει από την τιμή της θεωρητικής διασποράς (που για μικρές τιμές του  $H$  έχει μικρή μεροληψία) και τείνει στο μηδέν καθώς αυξάνεται η παράμετρος Hurst. Η διάμεσος φαίνεται να έχει μια σταθερή συμπεριφορά ανεξαρτήτως μήκους δείγματος στην κλίμακα 1 και είναι πολύ κοντά στην θεωρητική διασπορά ειδικά για  $H < 0.9$ .

Παρατηρούμε επίσης ότι για μεγάλο δείγμα  $n$  η πιο πιθανή τιμή προσεγγίζει την μέση τιμή επειδή η κατανομή στενεύει. Η πιθανότερη τιμή φαίνεται να είναι πολύ κοντά στο κάτω τεταρτημόριο σε όλες τις κλίμακες για όλες τις τιμές της παραμέτρου *Hurst*. Από τα διαγράμματα της κύρτωσης συναρτήσει της ασυμμετρίας φαίνεται ότι η κατανομή του κλιμακογράμματος προσεγγίζει την γάμμα κατανομή.

Στα περισσότερα αποτελέσματα φαίνεται ότι η ασυμμετρία είναι θετική που σημαίνει ότι η κατανομή του κλιμακογράμματος είναι θετικά ασύμμετρη. Επιπλέον η ασυμμετρία και η κύρτωση αυξάνονται καθώς αυξάνεται η παράμετρος Hurst και η κλίμακα  $k$ .

Το μήκος του δείγματος όπως και ο αριθμός των επαναλήψεων παίζει σημαντικό ρόλο στην ανάλυση καθώς αυξάνει την πληρότητα και εγκυρότητα των αποτελεσμάτων. Ειδικά όταν τα δεδομένα αφορούν γεωφυσικές διεργασίες ένα μικρό μήκος δεδομένων καθιστά δύσκολη την διαδικασία μελέτης καθώς εισάγει κάποια αβεβαιότητα. Είναι δύσκολο λοιπόν να μελετηθεί και για την ύπαρξη μακροχρόνιας εμμονής η οποία όπως αποδείξαμε επηρεάζει τα στατιστικά χαρακτηριστικά των δεδομένων.



## 6 ΒΙΒΛΙΟΓΡΑΦΙΑ

---

Κουτσογιάννης, Δ., Στατιστική Υδρολογία, έκδοση 3, Εθνικό Μετσόβιο Πολυτεχνείο, Τομέας Υδατικών Πόρων, Αθήνα, 1996.

Benjamin, J. R., and C. A. Cornell, Probability, Statistics and Decision for Civil Engineers, McGraw-Hill, 1970.

Beran, J., Statistics for Long-Memory Processes, Chapman and Hall, New York, 1994.

Bhat, U. N. Elements of Applied Stochastic Processes. New York: John Wiley & Sons, 1972.

Bobee, B., and F. Ashkar, The Gamma Family and Derived Distributions Applied in Hydrology, Water Resources Publications, Littleton, Colorado, 1991.

Breiman, L. Probability. Reading, Mass.: Addison-Wesley, 1968.

Cox, D. R., and H. D. Miller. The Theory of Stochastic Processes. New York: John Wiley & Sons, 1965.

Dimitriadis P., and Koutsoyiannis D., Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst-Kolmogorov processes, *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669.

Dimitriadis P., Gournari N., and D. Koutsoyiannis, Markov vs. Hurst-Kolmogorov behaviour identification in hydroclimatic processes, *European Geosciences Union General Assembly 2016, Geophysical Research Abstracts, Vol. 18*, Vienna, EGU2016-14577-4, doi:10.13140/RG.2.2.21019.05927, European Geosciences Union, 2016

Dynkin, E. B., and A. A. Yushkevich. Markov Processes: Theorems and Problems. New York: Plenum, 1969.

Eltahir, E.A.B. 1996. 'El Nino and the Natural Variability in the Flow of the Nile River'. *Water Resources Research* 32 (1): 131–137.

Feller, W. An Introduction to Probability Theory and Its Applications. 2 vols. New York: John Wiley & Sons, 1966 (vol. 2), 1968 (vol. 1, 3rd ed.).

Gournaris F., A Study of the Top Quark Production Threshold at a Future Electron-Positron Linear Collider, Department of Physics and Astronomy University College London, Thesis, September 2009.

Hoel, R. G., S. C. Port, and C. J. Stone. Introduction to Stochastic Processes. Boston: Houghton Mifflin, 1972.

Kendall, M.G., and A. Stuart, The Advanced Theory of Statistics, Vol. 1, Distribution theory, Charles Griffin & Co., London, 1963.

Kendall, M.G., and A. Stuart, The Advanced Theory of Statistics, Vol. 2, Inference and relationship, Third edition, Charles Griffin & Co., London, 1963.

Kolmogorov, A. 1941. 'The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds' Numbers'. In Akademiia Nauk SSSR Doklady, 30:301–305.

Koutsoyiannis, D. 2002. 'The Hurst Phenomenon and Fractional Gaussian Noise Made Easy'. Hydrological Sciences Journal 47 (4): 573–595.

Koutsoyiannis, D. 2003. 'Climate Change, the Hurst Phenomenon, and Hydrological Statistics'. Hydrological Sciences Journal 48 (1): 3–24.

Koutsoyiannis, D. 2010. 'A Random Walk on Water'. Hydrology and Earth System Sciences 14: 585–601.

Koutsoyiannis, D. 2011a. 'Hurst-Kolmogorov Dynamics and Uncertainty1'. JAWRA Journal of the American Water Resources Association 47 (3): 481–495.

Koutsoyiannis, D. 2011b. 'Hurst-Kolmogorov Dynamics as a Result of Extremal Entropy Production'. Physica A: Statistical Mechanics and Its Applications 390 (8): 1424–1432.

Koutsoyiannis D., Hydrology and Change, *Hydrological Sciences Journal*, 58 (6), 1177–1197, 2013.

Koutsoyiannis D., Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, 61 (2), 225–244, 2016

Koutsoyiannis, D., and A. Langousis. 2011. 'Precipitation'. In *Treatise on Water Science*, edited by P. Wilderer and S. Uhlenbrook, 2:27–78. Academic.

Koutsoyiannis, D., and A. Montanari. 2007. 'Statistical Analysis of Hydroclimatic Time Series: Uncertainty and Insights'. *Water Resources Research* 43 (5): W05429.

Lennartz, Sabine, and Armin Bunde. 2011. 'Distribution of Natural Trends in Long-term Correlated Records: A Scaling Approach'. *Physical Review E* 84 (2): 021129.

Lewis, P. A. *Stochastic Point Processes: Statistical Analysis, Theory, and Applications*. New York: John Wiley & Sons, Interscience, 1972.

Mandelbrot, Benoit B., and John W. Van Ness. 1968. 'Fractional Brownian Motions, Fractional Noises and Applications'. *SIAM Review* 10 (4): 422–437.

Papalexiou, S. M., D. Koutsoyiannis, and A. Montanari. 2013. 'Mind the Bias!'

Papoulis, A. 1990. 'Probability and Statistics', Prentice-Hall.

Papoulis, A. 1965. 'Probability, Random Variables and Stochastic Processes'. McGraw-Hill.

Rea, W., L. Oxley, M. Reale, and J. Brown. 2012. 'Not All Estimators Are Born Equal: The Empirical Properties of Some Estimators of Long Memory'. *Mathematics and Computers in Simulation*.

Ross, S. M. *Stochastic Processes*. New York: John Wiley & Sons, 1983.

Ross, S. M. *Introduction to Probability Models, 5/e* New York: Academic Press, 1993.

Roussis D., Parara I., Gournari P., Moustakis Y., Dimitriadis P., Iliopoulou T., Koutsoyiannis D., and Karakatsanis G., Energy, variability and weather finance engineering, *European Geosciences Union General Assembly 2017, Geophysical Research Abstracts, Vol. 19*, Vienna, 19, EGU2017-16919, European Geosciences Union, 2017

Samorodnitsky, G. *Long Memory and self-similar processes*, School of Operations Research and Industrial Engineering, and Department of Statistical Science, Cornell University, Ithaca, NY, 2007.

Spiegel, M. R., Theory and Problems of Probability and Statistics, Schaum's Outline Series, McGraw-Hill, New York, 1975.

Taylor, H.M., and S. Karlin, An Introduction to Stochastic Modelling, Academic Press, 1984.

Tyralis, H., and D. Koutsoyiannis. 2011. 'Simultaneous Estimation of the Parameters of the Hurst–Kolmogorov Stochastic Process'. Stochastic Environmental Research and Risk Assessment 25 (1): 21–33.

Yevjevich, V., Probability and Statistics in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1972.



# ΠΑΡΑΡΤΗΜΑ Α

## Κώδικας Matlab και μεθοδολογία

Ο βρόχος αυτός δημιουργεί όλους τους πιθανούς συνδυασμούς μεταξύ του μήκους χρονοσειράς  $n$  (10, 20, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000) και των συντελεστών Hurst (0.5, 0.6, 0.7, 0.8, 0.9, 0.99). Οι απαραίτητες επαναλήψεις (Monte Carlo) υπολογίζονται από την εμπειρική σχέση  $\text{floor}\left(\frac{10^7}{n^{2-H}}\right) + 1$  όπου  $\text{floor}$  σημαίνει στρογγυλοποίηση προς τα κάτω (στον πρώτο ακέραιο).

```
ScN=[10 20 100 200 500 1000 2000 5000 10000 20000 50000 10^5 ];
ScH=[0.6 0.7 0.8 0.9 0.95 0.99];
I=0;
for I2=1:numel(ScN);
    for I3=1:numel(ScH);
        I=I+1;
        ScenN(I)=ScN(I2);
        ScenH(I)=ScH(I3);
        ScenZ(I)=floor((10^7)/(ScenN(I)^(2-ScenH(I))))+1;
    end;
end;
```

Το υπόλοιπο μέρος του κώδικα χωρίζεται σε δυο τμήματα. Στο πρώτο τμήμα υπολογίζεται η συνθετική χρονοσειρά σύμφωνα με το μοντέλο SMA και τις αντίστοιχες σχέσεις του. Το μοντέλο αυτό θα έχει μέση τιμή 0 και τυπική απόκλιση 1. Οι συντελεστές βάρους υπολογίζονται από την σχέση (2.77). Η παραγωγή τυχαίων αριθμών από κανονική κατανομή με μέση τιμή και τυπική απόκλιση από τις σχέσεις (2.78), (2.79), χρησιμοποιείται για τον λευκό θόρυβο. Έτσι ο λευκός θόρυβος έχει  $n + 2S$  στοιχεία και οι συντελεστές βάρους  $2S + 1$ . Τελικά το SMA έχει  $n$  στοιχεία.

```
Z=ScenZ(Scen); In=[ScenN(Scen) ScenH(Scen) 0 1]
for L=1:Z
    Aj=@(x) (sqrt(2-2*x(2)).*x(4)./(3-2*x(2)).*(abs((-x(1):x(1))+1).^ (x(2)+0.5)+abs((-x(1):x(1))-1).^ (x(2)+0.5)-2*abs((-x(1):x(1))).^ (x(2)+0.5)));
    n=In(1); H=In(2); m=In(3); stdev=In(4);
    S=n;
    A=Aj([S H m stdev]); Mt=m./(A(S+1)+2*sum(A(S+2:end))); Var=1;
    R=norminv(rand(2*S+n,1),Mt,Var);
    SMA=[];
    for I=1:n;
        SMA(I)=A*R(I:2*S+I,1);
    end
```



Το δεύτερο τμήμα ασχολείται με την δημιουργία του κλιμακογράμματος. Ο πίνακας Μ είναι η κλίμακα και γράφεται με αυτό τον τρόπο για εξοικονόμηση χώρου. Το κλιμακόγραμμα (CC) υπολογίζεται με τις σχέσεις της συναθροισμένης κλίμακας.

```
M=[(1:9) 10^1*(1:1:9) 10^2*(1:1:9) 10^3*(1:1:5)]';
V=[]; KK=[]; CC=[];
for K=1:numel(M)
Y=[];
for J=1:floor(n/M(K))
l=M(K)*(J-1)+1;
Sum=0;
for I=l:M(K)*J
Sum=Sum+SMA(I);
end
Y(J)=Sum/M(K);
end
V(K,1)=var(Y);
end
CC(:,L)=V;
```

Τέλος, υπολογίζεται η συνάρτηση κατανομής και η συνάρτηση πυκνότητας πιθανότητας του κλιμακογράμματος για κάθε κλίμακα, όπως επίσης και κάποια στατιστικά χαρακτηριστικά του όπως η μέση τιμή, η διασπορά, η κύρτωση, η ασυμμετρία, η επικρατούσα τιμή, η διάμεσος, το πρώτο και το τρίτο τεταρτημόριο σε κάθε κλίμακα.

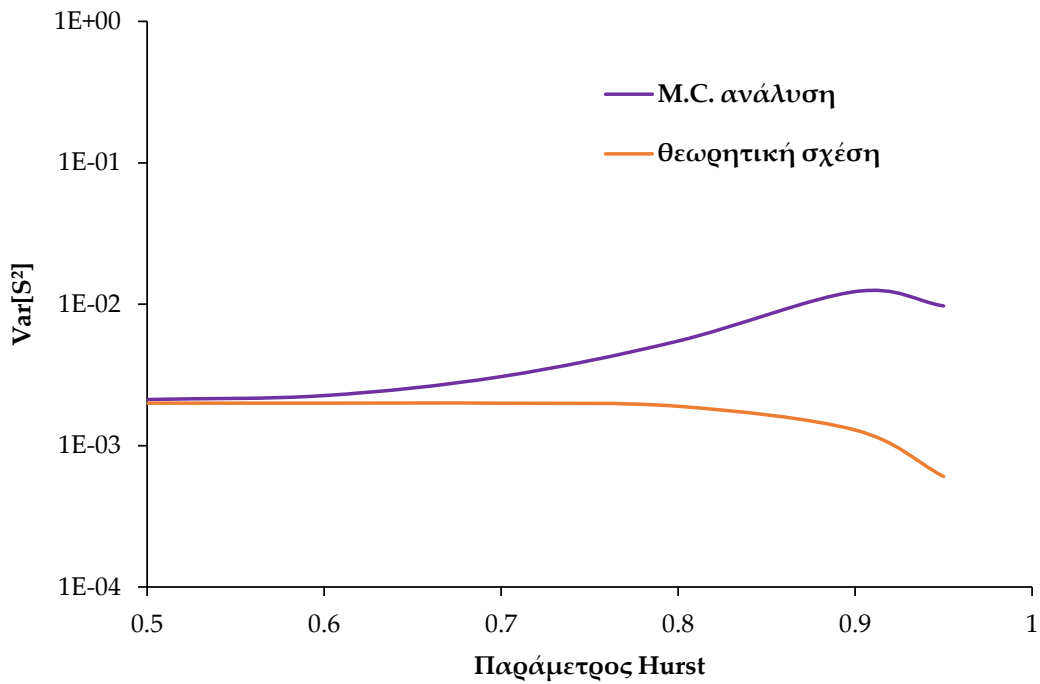
```
CDF=[]; PDF=[];
for IM=1:K;
B(IM,:)=floor(min(CC(IM,:))):floor(floor(max(CC(IM,:)))+1-
floor(min(CC(IM,:)))/100:floor(max(CC(IM,:)))+1;
for I=1:numel(B(IM,:)); S=numel(find(CC(IM,:)<=B(IM,I)));
CDF(IM,I)=S/(Z+1); end
PDF(IM,:)=(CDF(IM,2:end)-CDF(IM,1:end-1))./(B(IM,2:end)-
B(IM,1:end-1));
End
MCC=mean(CC)';
VCC=var(CC)';
KYCC=kurtosis(CC)';
SCC=skewness(CC)';
MOCC=mode(CC)';
MECC=median(CC)';
Q25=quantile(CC',0.25)';
Q75=quantile(CC',0.75)';
```

Ο κώδικας αυτός επαναλαμβάνεται όσες φορές απαιτείται για όλα τα εναλλακτικά σενάρια.

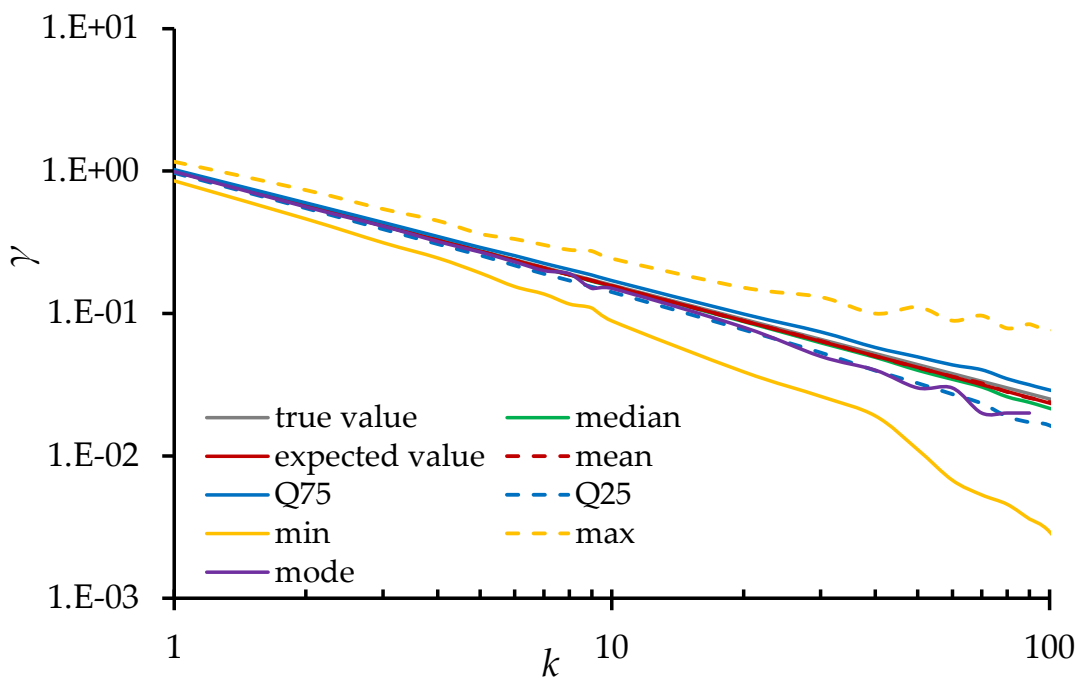
```
Scenarios=numel(ScenZ);
for Scen=1:Scenarios end
```

# ΠΑΡΑΡΤΗΜΑ Β

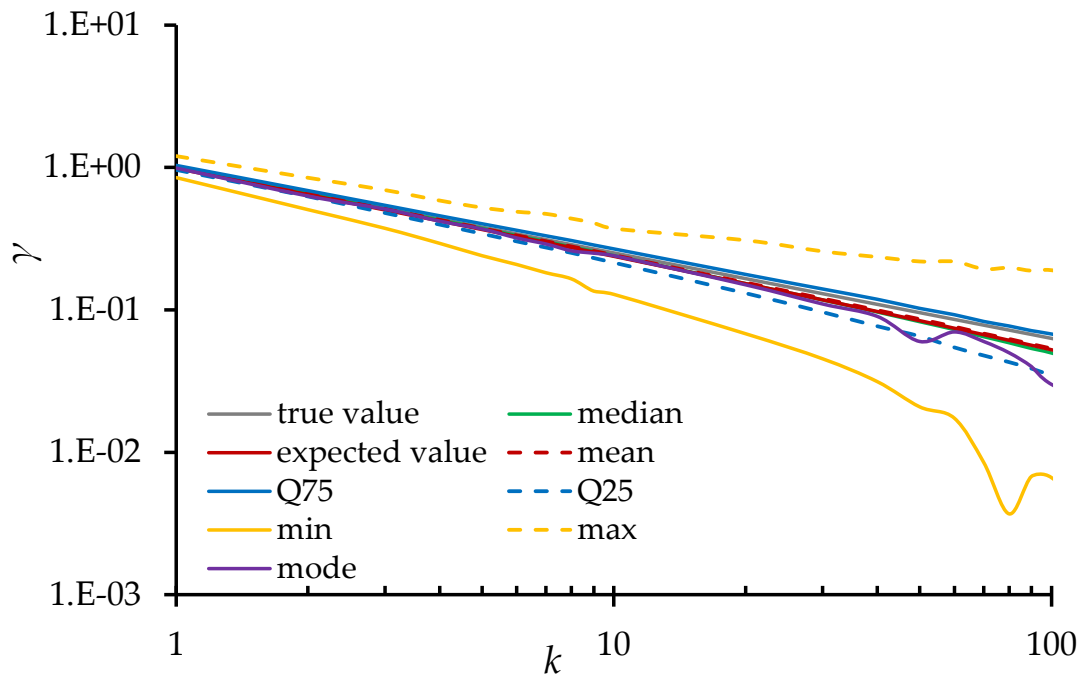
## Διαγράμματα αναλύσεων



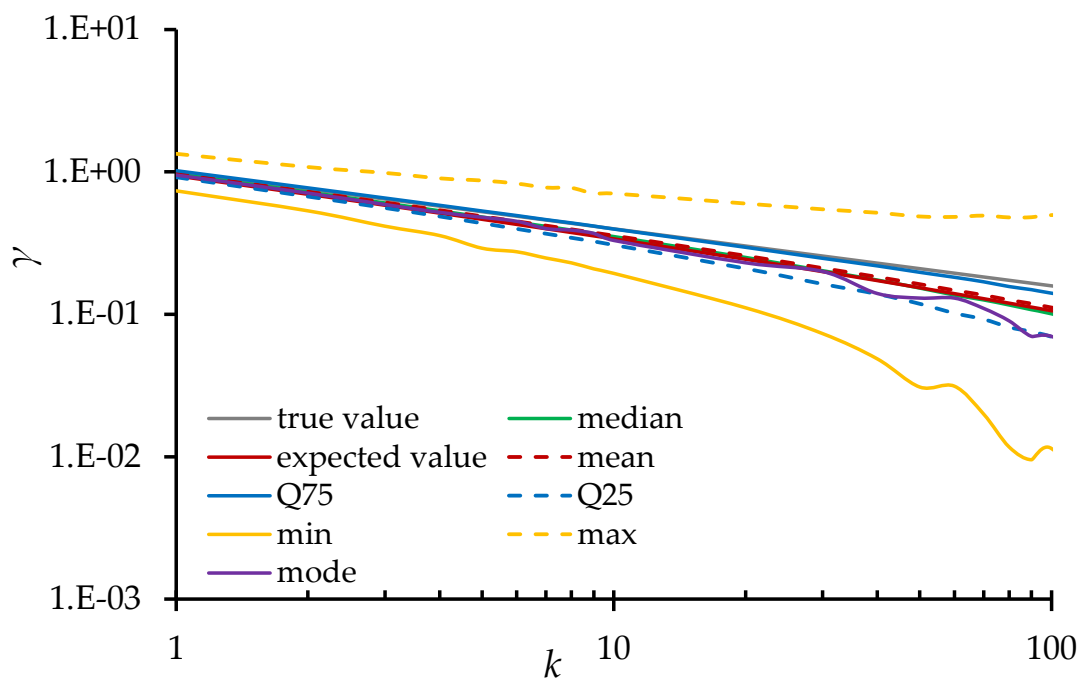
Σχήμα 54 Απόκλιση της διασποράς της δειγματικής διασποράς της θεωρητικής σχέσης (2.32) από τα αποτελέσματα από την Monte Carlo ανάλυση.



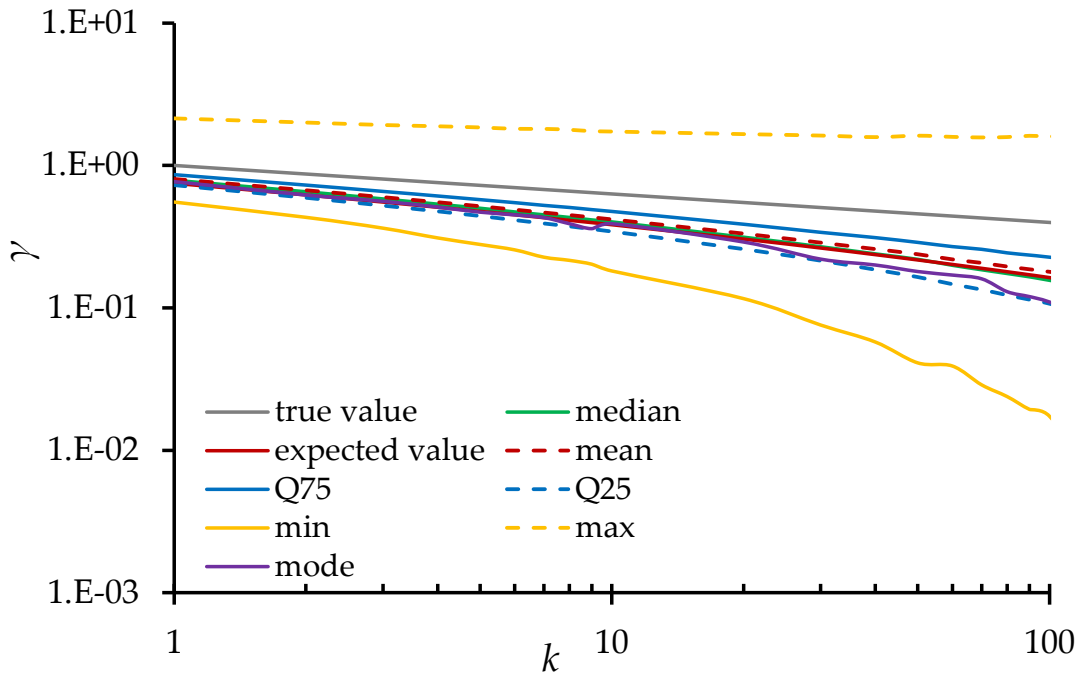
Σχήμα 55 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος για  $H=0.6$ ,  $n=1000$



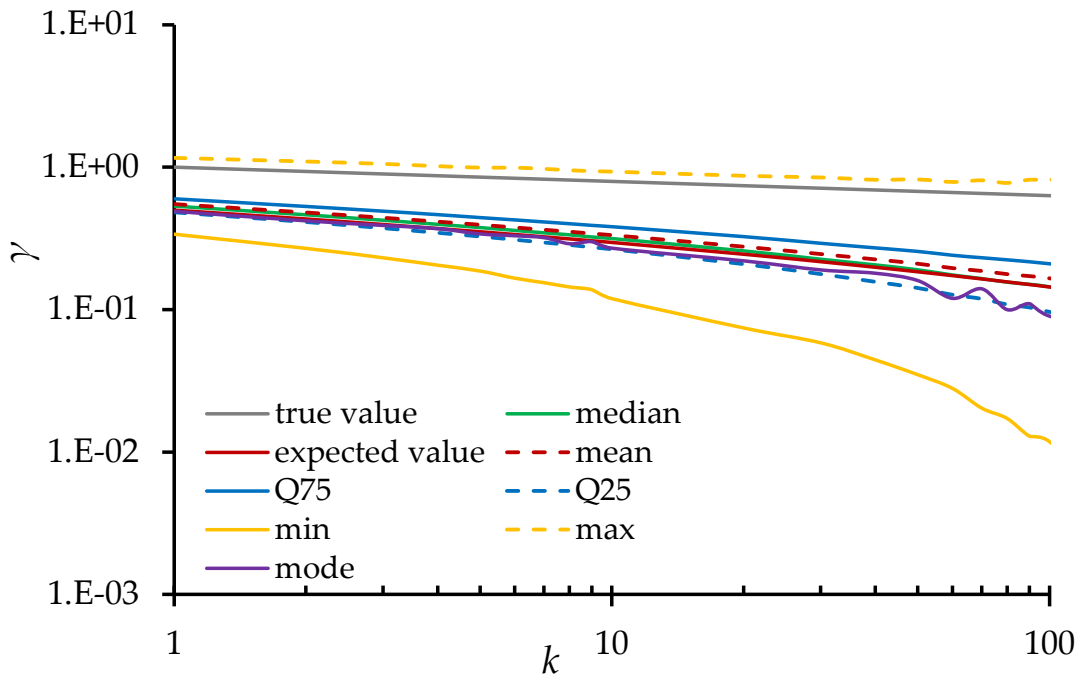
Σχήμα 56 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος για  $H=0.7$ ,  $n=1000$



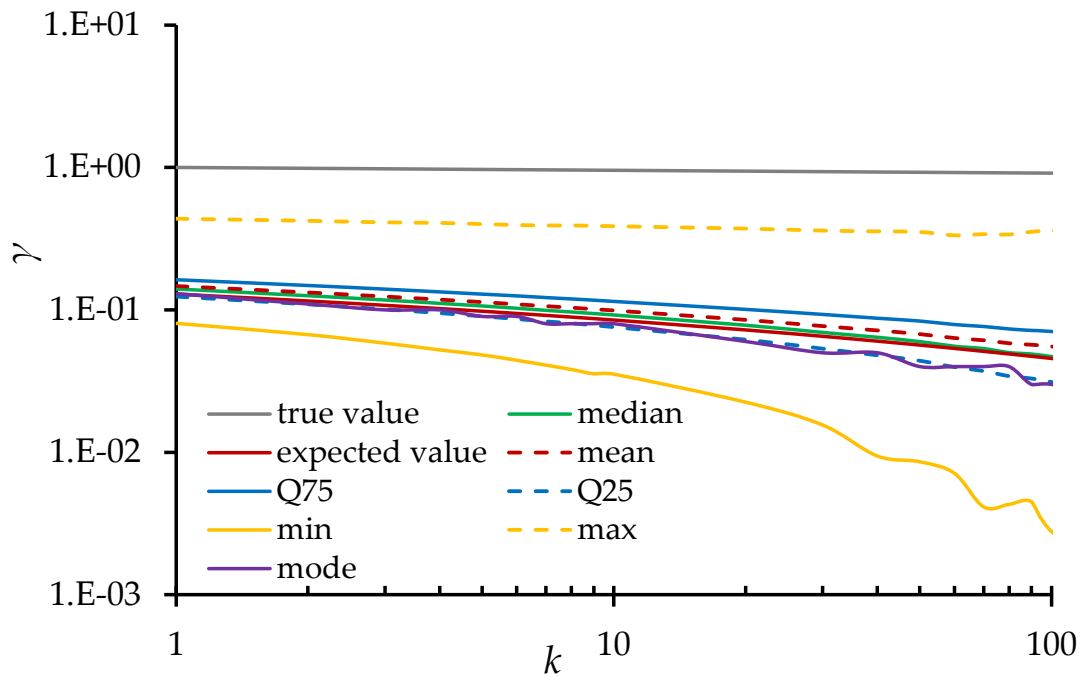
Σχήμα 57 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος για  $H=0.8$ ,  $n=1000$



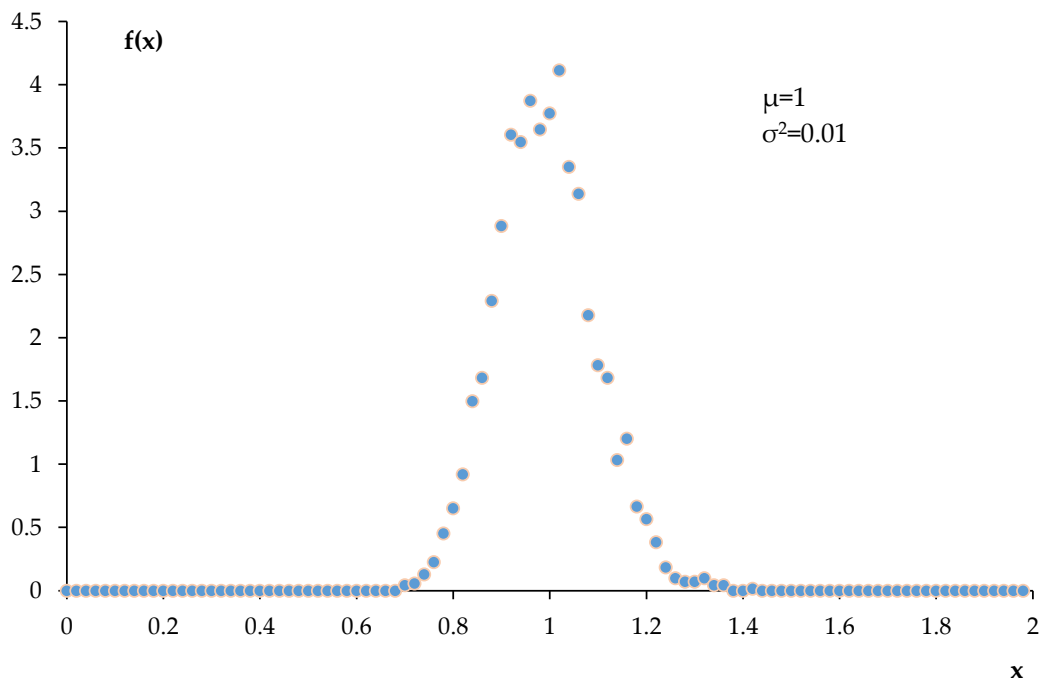
Σχήμα 58 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος για  $H=0.9$ ,  $n=1000$ .



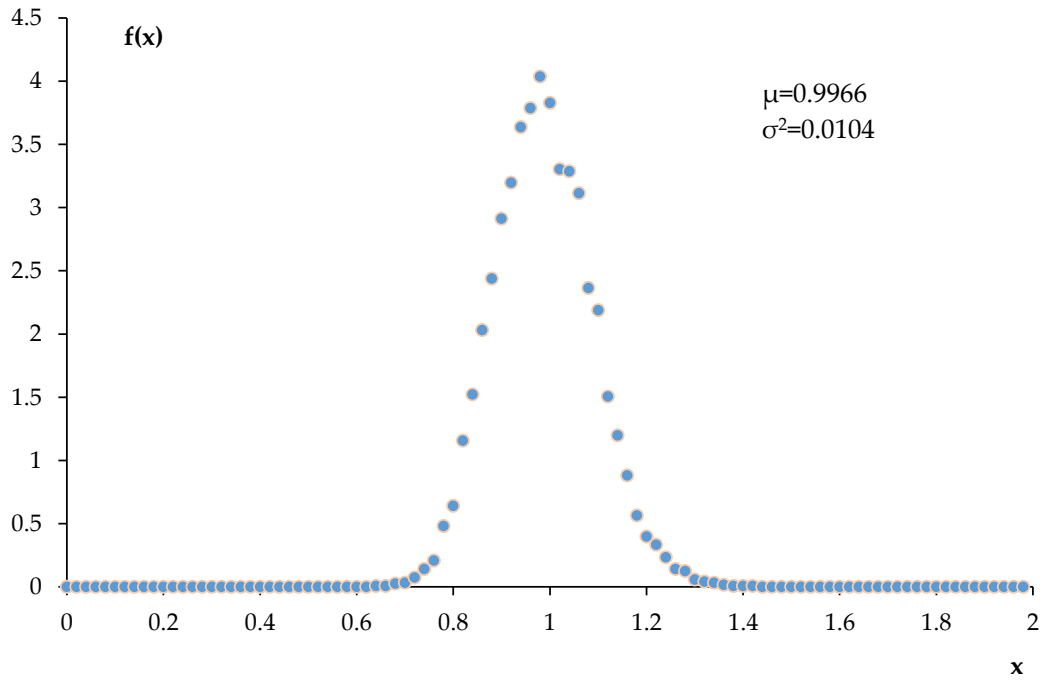
Σχήμα 59 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος για  $H=0.95$ ,  $n=1000$ .



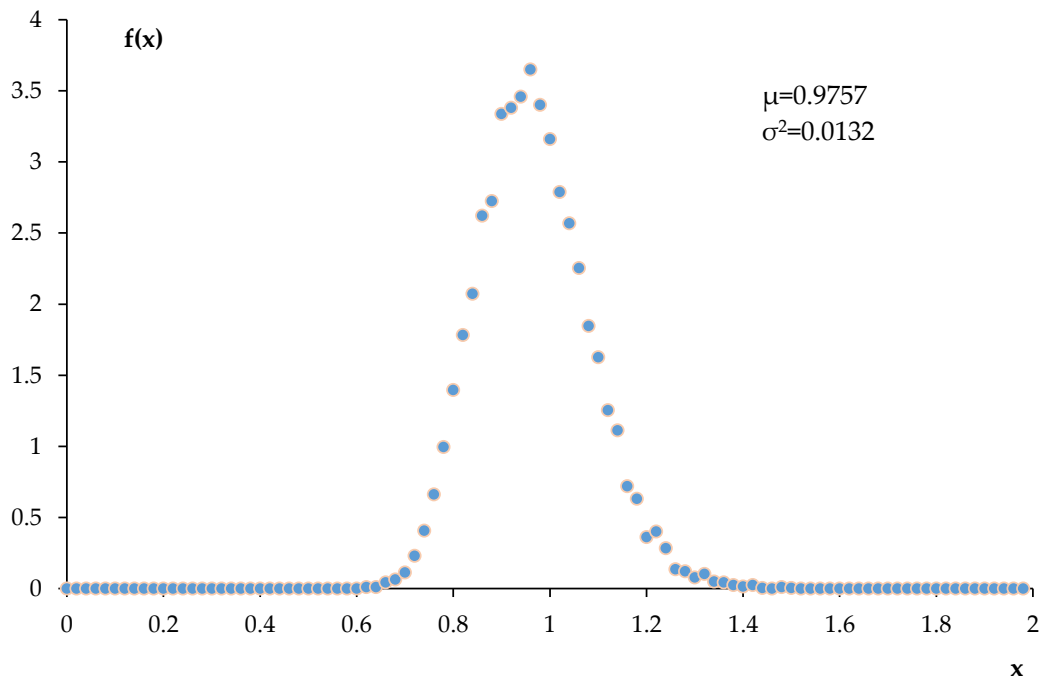
Σχήμα 60 Διάγραμμα στατιστικών χαρακτηριστικών κλιμακογράμματος για  $H=0.99$ ,  $n=1000$ .



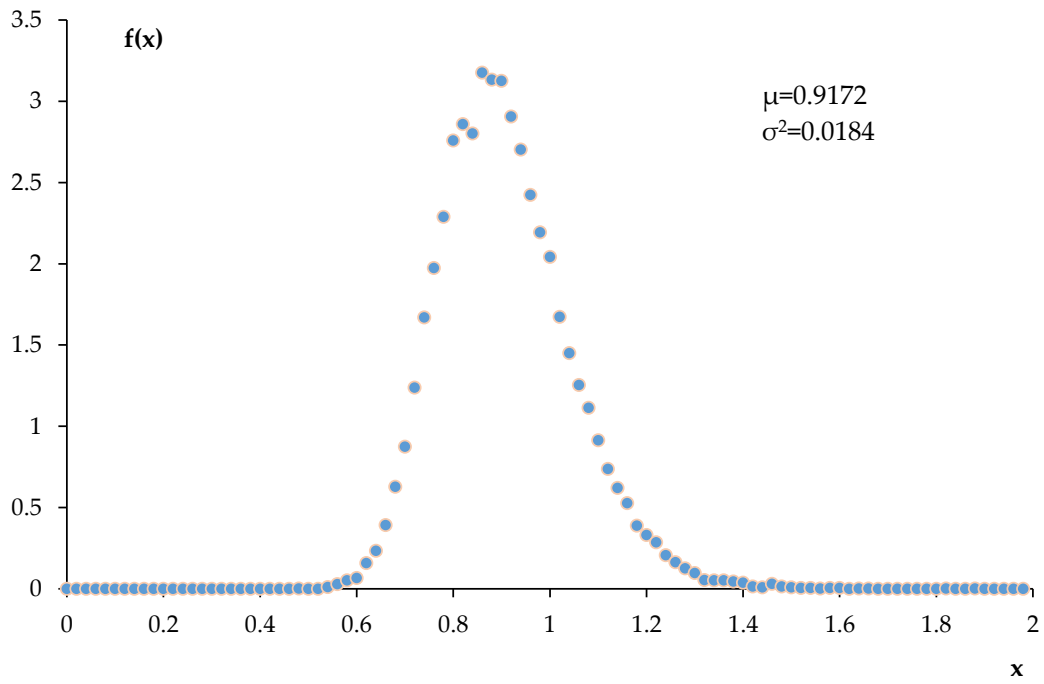
Σχήμα 61 Κατανομή κλιμακογράμματος για  $H=0.5$   $n=200$  στην κλίμακα  $k=1$ .



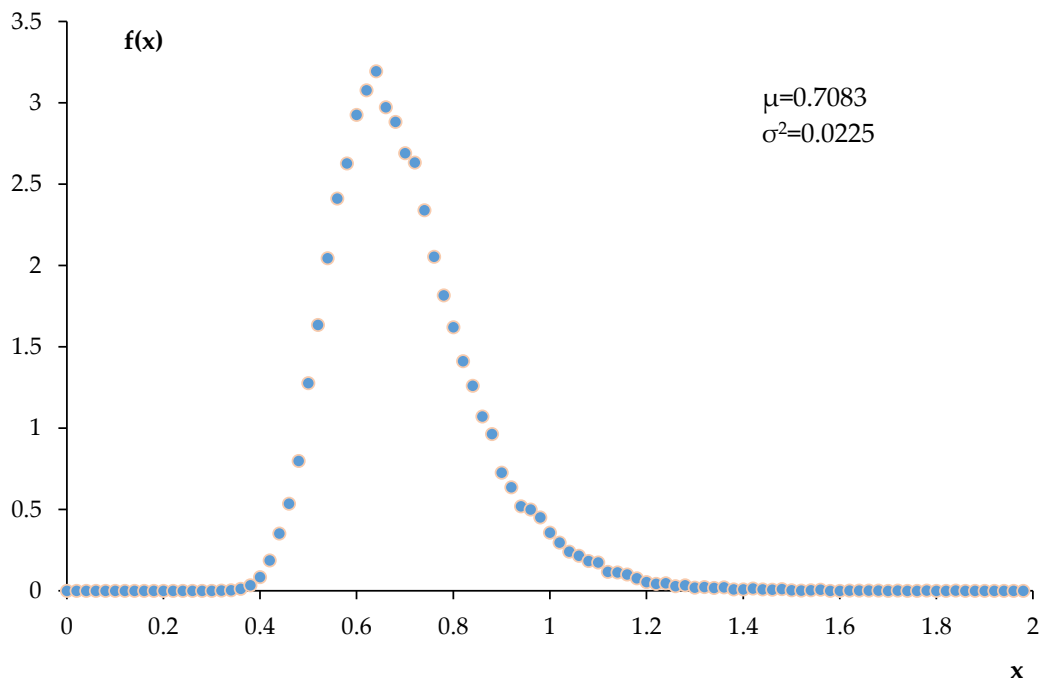
Σχήμα 62 Κατανομή κλιμακογράμματος για  $H=0.6$   $n=200$  στην κλίμακα  $k=1$ .



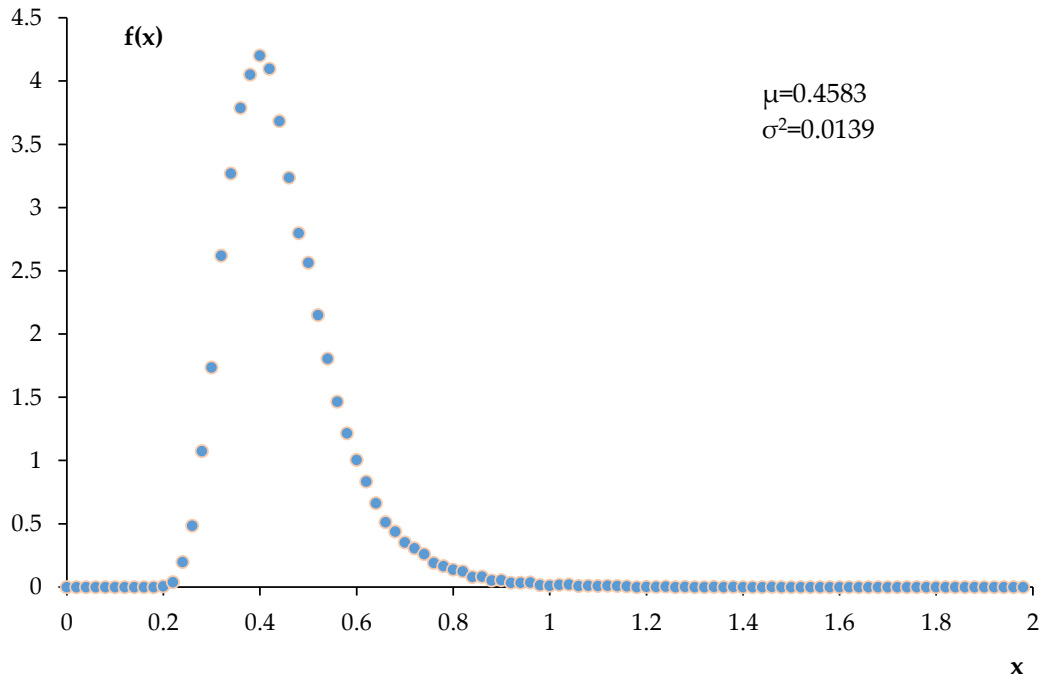
Σχήμα 63 Κατανομή κλιμακογράμματος για  $H=0.7$   $n=200$  στην κλίμακα  $k=1$ .



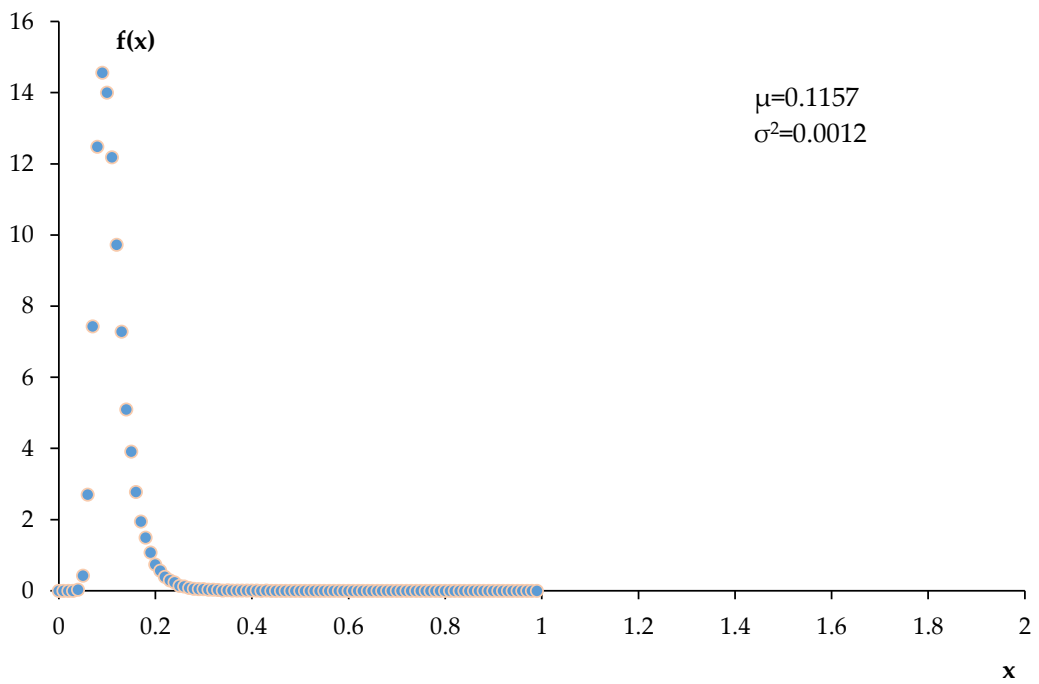
Σχήμα 64 Κατανομή κλιμακογράμματος για  $H=0.8$   $n=200$  στην κλίμακα  $k=1$ .



Σχήμα 65 Κατανομή κλιμακογράμματος για  $H=0.9$   $n=200$  στην κλίμακα  $k=1$ .

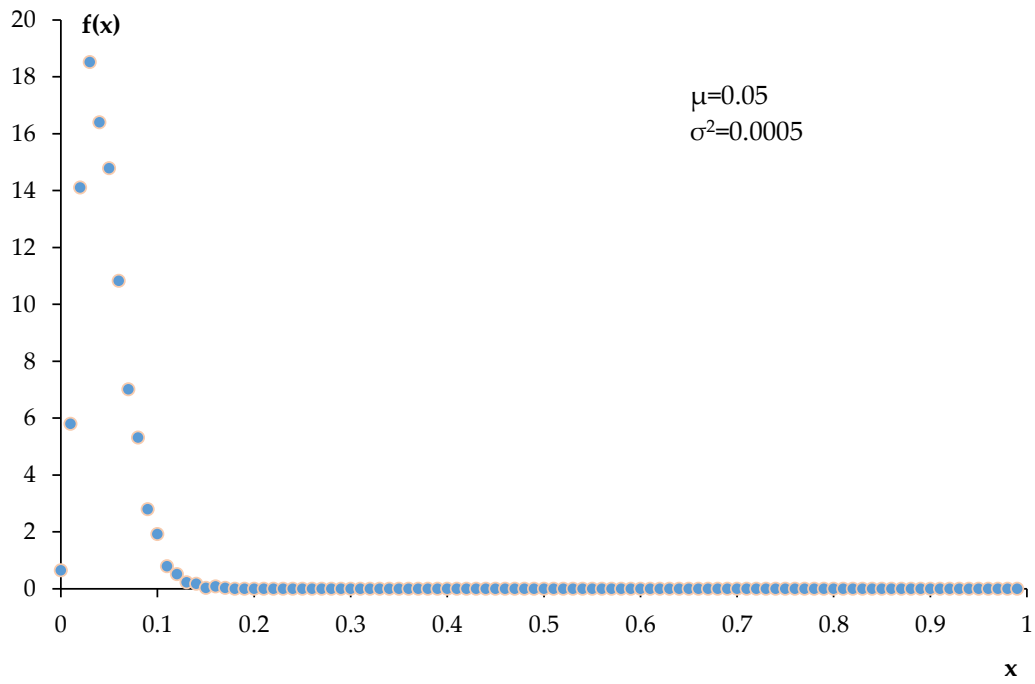


Σχήμα 66 Κατανομή κλιμακογράμματος για  $H=0.95$   $n=200$  στην κλίμακα  $k=1$ .

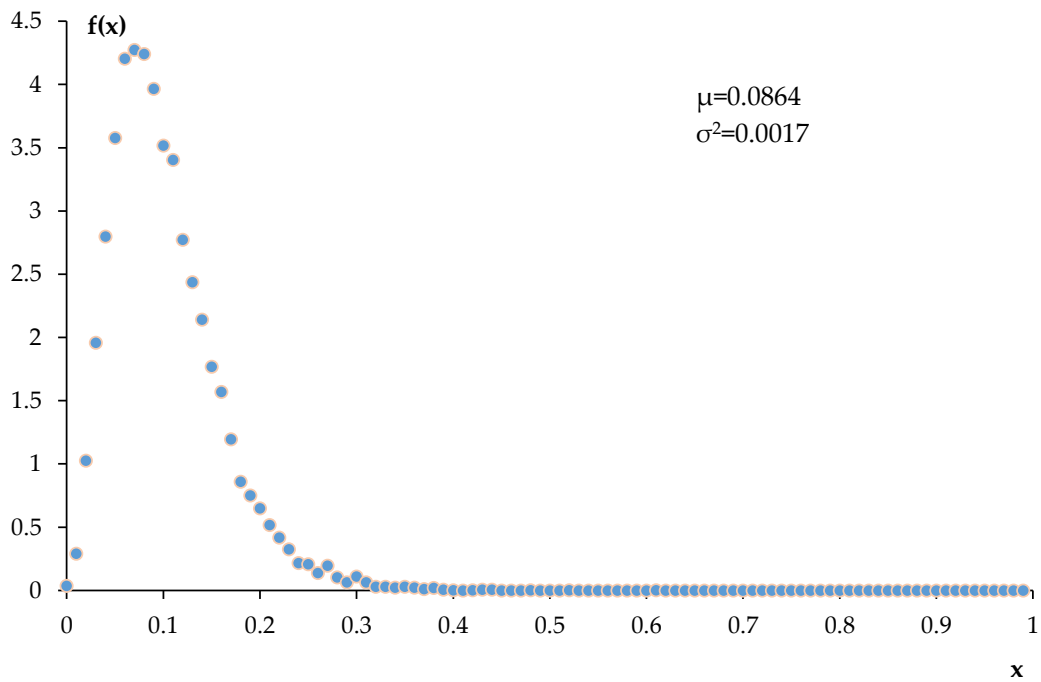


Σχήμα 67 Κατανομή κλιμακογράμματος για  $H=0.99$   $n=200$  στην κλίμακα  $k=1$ .

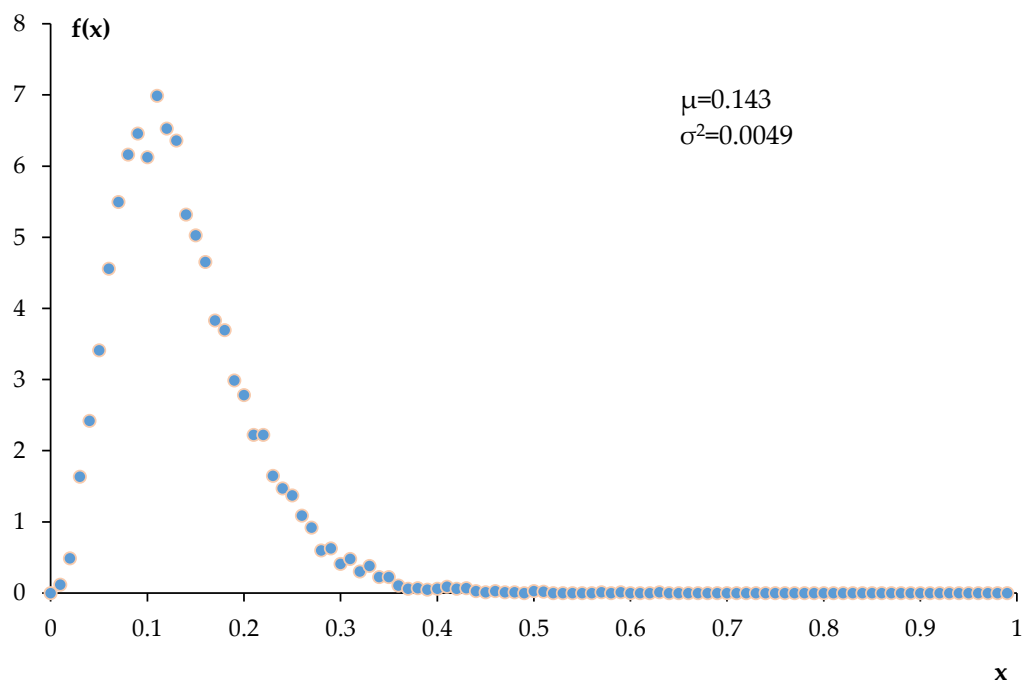




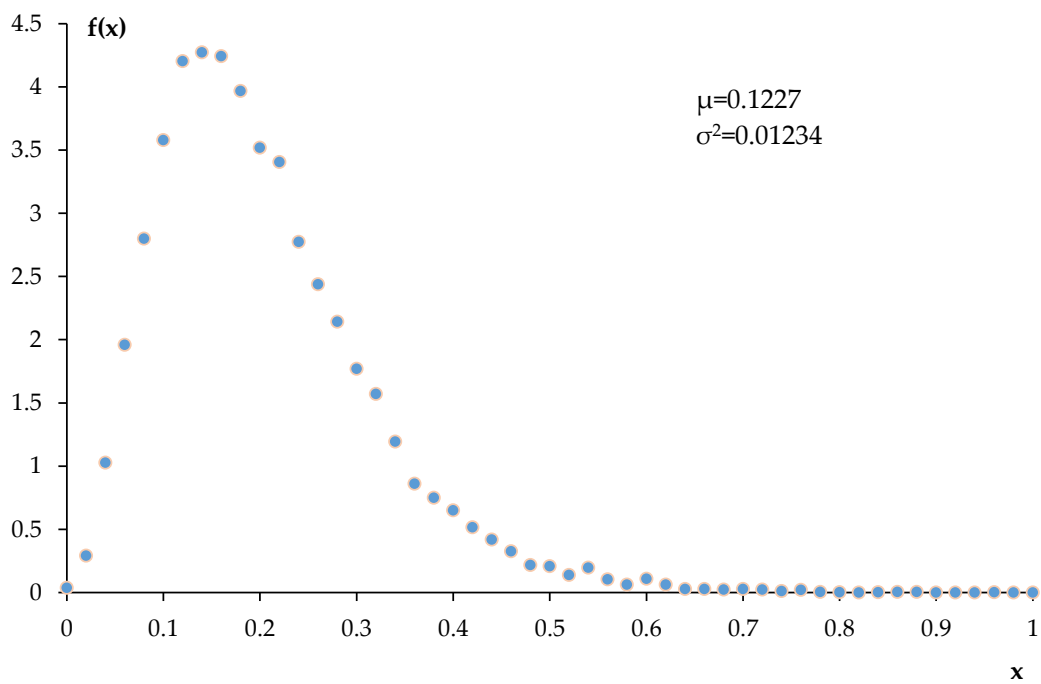
Σχήμα 68 Κατανομή κλιμακογράμματος για  $H=0.5$   $n=200$  στην κλίμακα  $k=20$ .



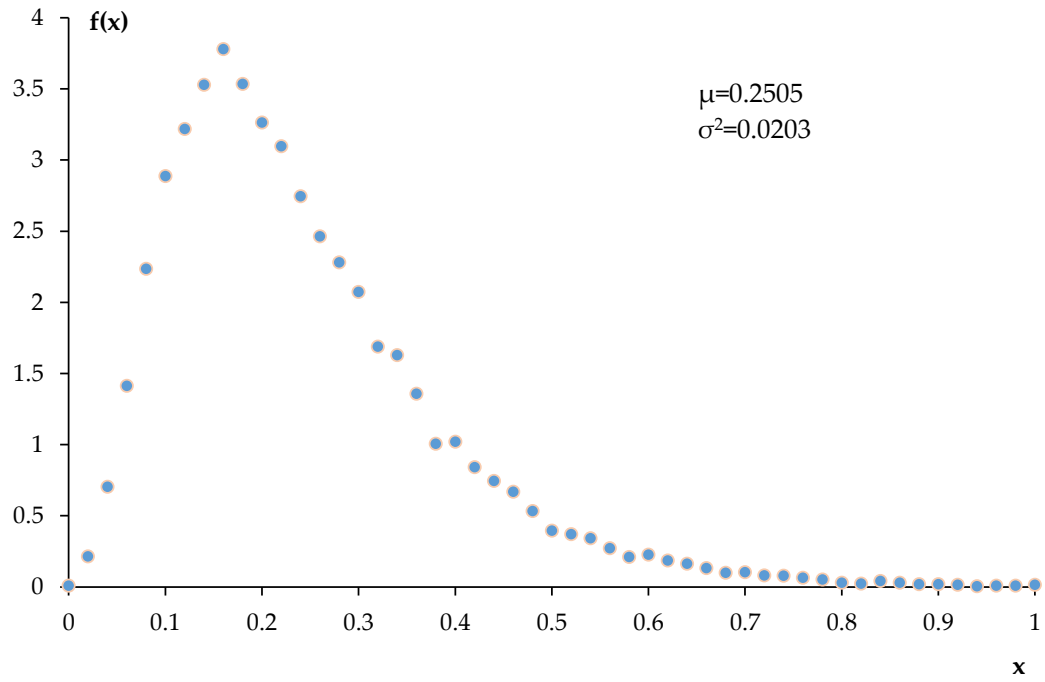
Σχήμα 69 Κατανομή κλιμακογράμματος για  $H=0.6$   $n=200$  στην κλίμακα  $k=20$ .



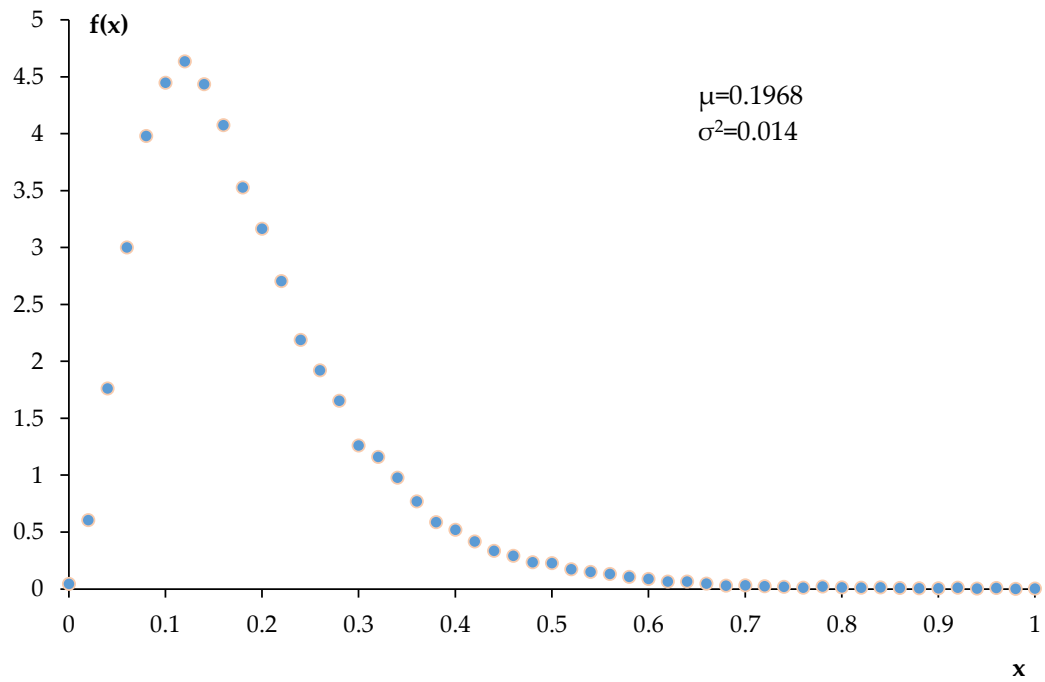
Σχήμα 70 Κατανομή κλιμακογράμματος για  $H=0.7$   $n=200$  στην κλίμακα  $k=20$ .



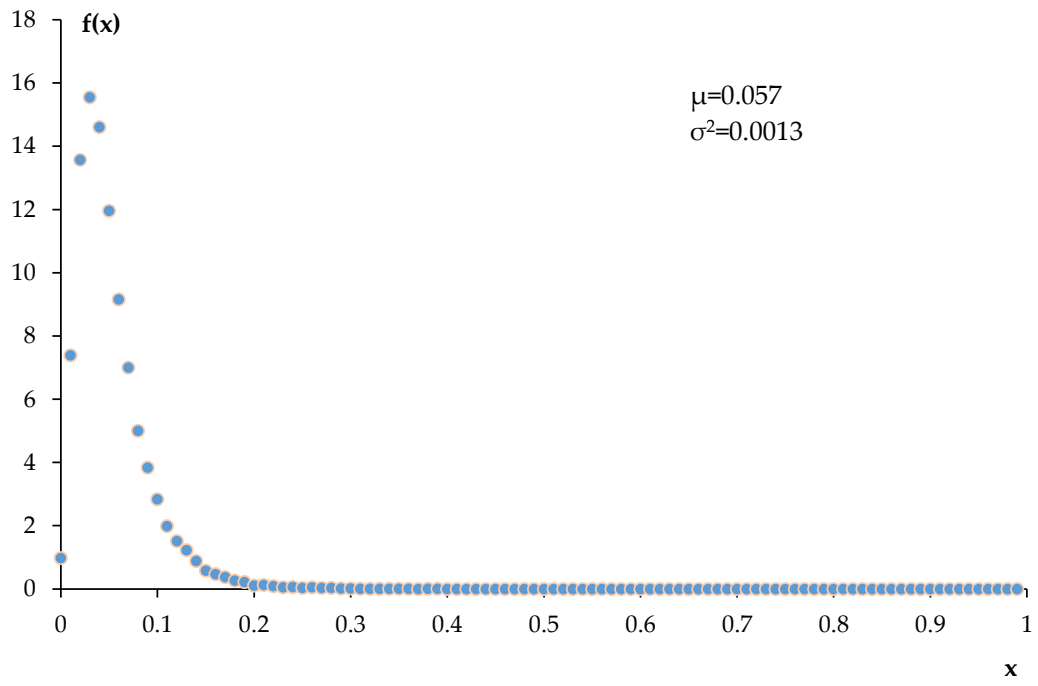
Σχήμα 71 Κατανομή κλιμακογράμματος για  $H=0.8$   $n=200$  στην κλίμακα  $k=20$ .



Σχήμα 72 Κατανομή κλιμακογράμματος για  $H=0.9$   $n=200$  στην κλίμακα  $k=20$ .



Σχήμα 73 Κατανομή κλιμακογράμματος για  $H=0.95$   $n=200$  στην κλίμακα  $k=20$ .



Σχήμα 74 Κατανομή κλιμακογράμματος για  $H=0.99$   $n=200$  στην κλίμακα  $k=20$ .