



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

ΠΙΘΑΝΟΤΙΚΕΣ ΚΑΤΑΝΟΜΕΣ ΑΚΡΑΙΩΝ ΒΡΟΧΟΠΤΩΣΕΩΝ
ΕΦΑΡΜΟΓΗ ΣΕ ΠΑΓΚΟΣΜΙΑ ΚΛΙΜΑΚΑ

Διπλωματική Εργασία

Σοφία Νεραντζάκη

Επιβλέπων Καθηγητής

Δημήτρης Κουτσογιάννης

Αθήνα, 30 Οκτωβρίου 2012

Εισαγωγή

- Βροχόπτωση θεωρείται τυχαία μεταβλητή
- Ορίζεται από μια συνάρτηση κατανομής – Περίοδος επαναφοράς
- Συνήθης μεθοδολογία για επιλογή κατανομής
 - ✓ Επιλογή οικογενειών κατανομών βάσει εμπειρίας
 - ✓ Εκτίμηση παραμέτρων με μεθόδους προσαρμογής
 - ✓ Επιλογή καταλληλότερης με τεστ εφαρμογής
- Σφάλμα: μέθοδοι προσαρμογής αφορούν στο σύνολο των δεδομένων και μεροληπτούν εις βάρος της ουράς- κακή εκπροσώπηση ακραίων
- Επιπτώσεις στον υδρολογικό σχεδιασμό
- Στόχος: σωστή προσαρμογή κατανομών στα ακραία

Θεωρία Ακραίων Τιμών

- Σύμφωνα με τη Θεωρία Ακραίων Τιμών:
- Το μέγιστο του ορθά κανονικοποιημένου δείγματος συγκλίνει σε μια μη εκφυλισμένη κατανομή που ανήκει σε μια από τις τρεις:
 - ✓ Κατανομή τύπου Gumbel
 - ✓ Κατανομή τύπου Fréchet
 - ✓ Αντίστροφη Weibull κατανομή
- Οι υπερβάσεις πάνω από ένα κατώφλι περιγράφονται από τη Γενικευμένη κατανομή Pareto
- Όμως οι παραπάνω κατανομές είναι οριακοί νόμοι – αναφέρονται σε άπειρο δείγμα

Κατάταξη κατανομών βάσει ουράς

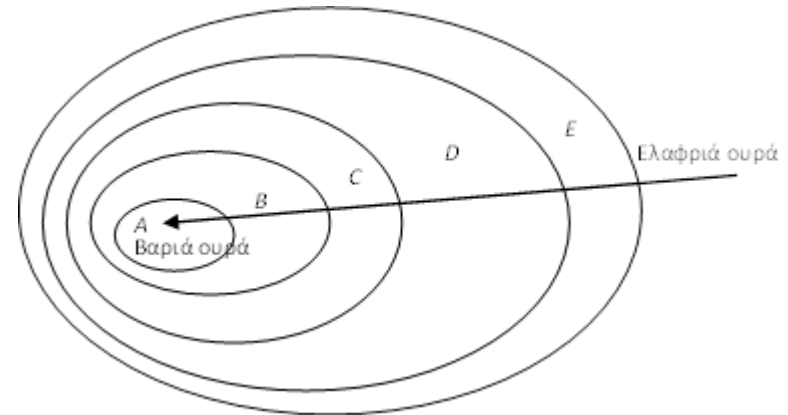
- Ως ουρά ορίζεται το ανώτερο μέρος της κατανομής - οι ακραίες τιμές της Ορίζει το μέγεθος και τη συχνότητα των ακραίων γεγονότων.
 - ✓ «Χονδρές» ή «Βαριές»
 - ✓ «Λεπτές» ή «Ελαφριές»
- Βάσει της συμπεριφοράς της ουράς έχουν προταθεί διάφοροι τρόποι κατάταξης κατανομών
- Ο κυριότερος διαχωρισμός είναι σε υποεκθετικές και υπερεκθετικές
- Οι πρώτες έχουν πιο χονδρή ουρά από οποιαδήποτε εκθετική (κατανομές τύπου δύναμης - Pareto, Λογαριθμοκανονική, Weibull με $\gamma < 1$)
- Οι δεύτερες πιο λεπτή ουρά από τις εκθετικές (Κανονική, Γάμμα, Weibull $\gamma > 1$)

Ταξινόμηση κατανομών (1)

- Κατά Werner και Upprer (2002)
- ✓ Πέντε τάξεις κατανομών που περιέχουν η μία την άλλη



- *E*: κατανομές με μη ύπαρξη εκθετικών ροπών
- *D*: υπο-εκθετικές κατανομές
- *C*: κατανομές ομαλής κύμανσης
- *B*: κατανομές τύπου Pareto
- *A*: σταθερές κατανομές



Ταξινόμηση κατανομών (2)

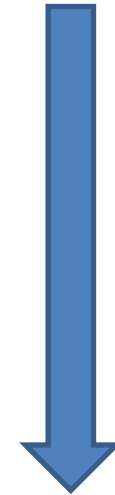
- Κατά Ouarda et al. (2002)

και σύμφωνα με τη σχέση $T \sim_{x_T \rightarrow \infty} \left(\frac{1}{f(x_T)} \right)' = -\frac{f'(x_T)}{f^2(x_T)}$

➔ τέσσερις τάξεις κατανομών

ισοδύναμες με Werner και Upper

Ταξινόμηση	Κατανομές
«Λεπτές» Ουρές	Κανονική
Τάξη $E \setminus D$	Εκθετική
Τάξη $D \setminus C$	Gumbel, Halphen τύπου A, Halphen τύπου B, Pearson τύπου B
Ομάδα II	Λογαριθμοκανονική
Τάξη $C \setminus B$	Fréchet, Halphen B, Αντίστροφη Γάμμα, Λογαριθμική Pearson τύπου 3
Τάξη $B \setminus A$	Pareto
Τάξη A	Σταθερές Κατανομές



Αύξηση
«βαρύτητας»
ουράς

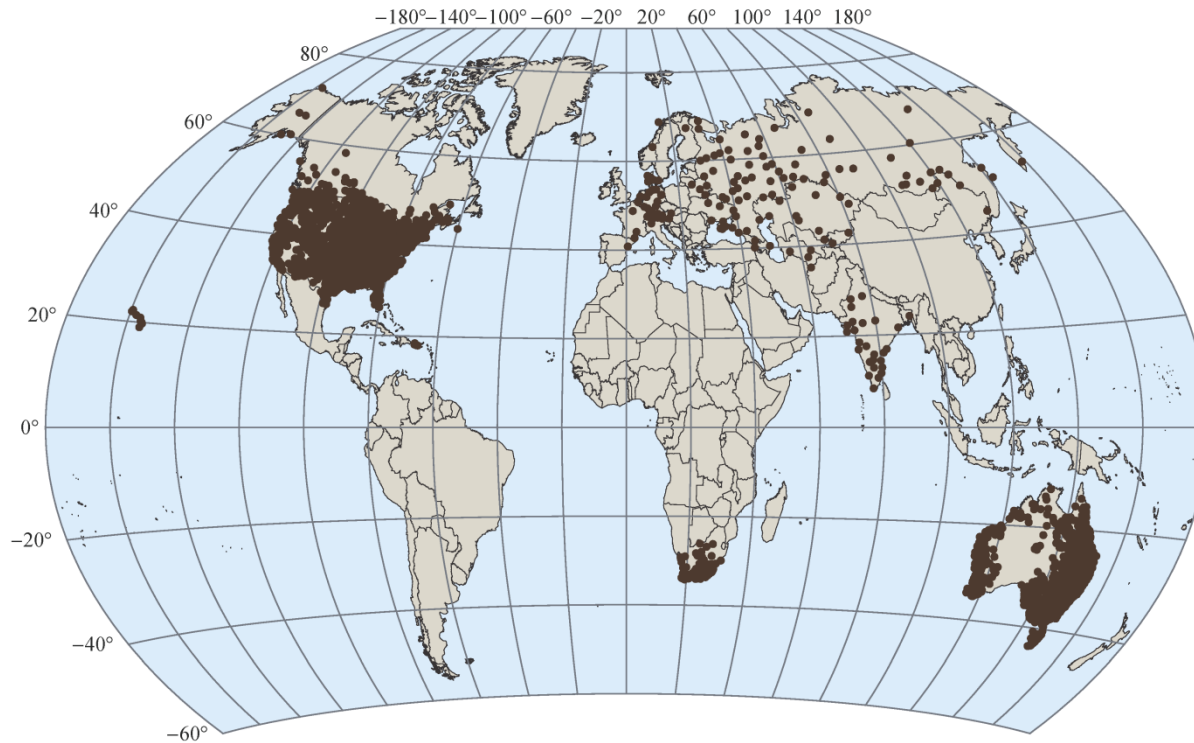
- Επίσης κατά Embrechts et al. (1997) και Cline και Samorodnitsky (1994)

Γραφικές Μέθοδοι

- Γενικά υπάρχουν διάφορες γραφικές μέθοδοι για την επιλογή κατάλληλου μοντέλου ουράς
- Εδώ αναπτύσσονται και εφαρμόζονται δύο:
 - Η μέθοδος απεικόνισης ουράς σε λογαριθμικούς άξονες (Log-Log plot) που ελέγχει αν η ουρά είναι τύπου δύναμης
 - Η μέθοδος βασιζόμενη στη Συνάρτηση Μέσης Υπέρβασης που ελέγχει αν η ουρά είναι εκθετική (Embrechts et al., 2003)
- Επίσης πραγματοποιείται απευθείας εφαρμογή γνωστών ουρών στα δεδομένα μέσω ελαχιστοποίησης τύπου σφάλματος.

Δεδομένα

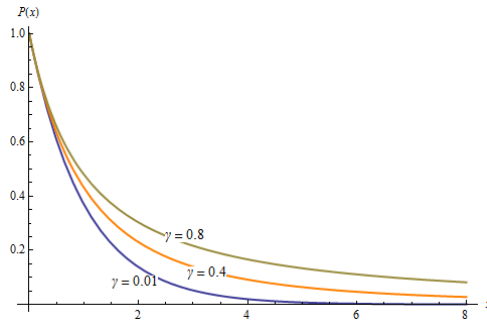
- Ημερήσιες τιμές βροχόπτωσης από το Παγκόσμιο Ιστορικό Δίκτυο Κλιματολογίας - Global Historical Climatology Network, version 2.60
 - Σταθμοί που είχαν καταγραφές 100 ετών και άνω
 - Δεν τους έλειπε πάνω από το 20% των δεδομένων
 - «Σημαίες» ποιότητας λιγότερες από 0.1%
- Τα δεδομένα προς ανάλυση προήλθαν από 3 477 σταθμούς



Υπό εξέταση ουρές

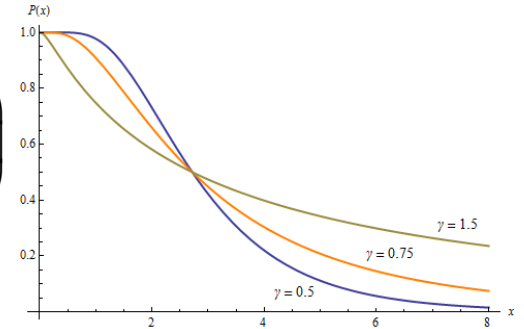
Pareto τύπου II

$$\bar{F}(x) = \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}}$$



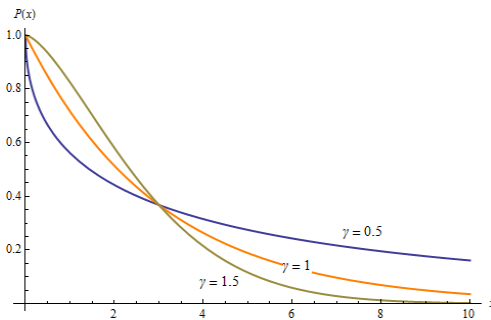
Λογαριθμοκανονική

$$\bar{F}(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{\ln x - \beta}{\sqrt{2}\gamma}\right)$$



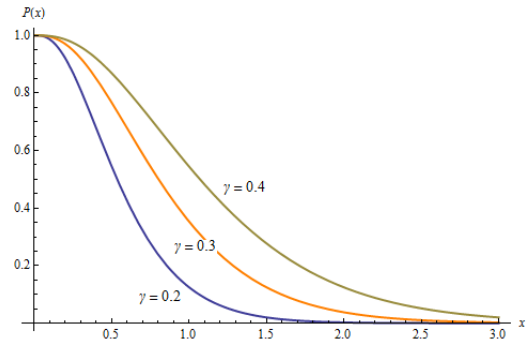
Weibull

$$\bar{F}(x) = \exp\left(-\frac{x^\gamma}{\beta}\right)$$



Γάμμα

$$\bar{F}(x) = \Gamma\left(\gamma, \frac{x}{\beta}\right) / \Gamma(\gamma)$$



Log-Log Plot

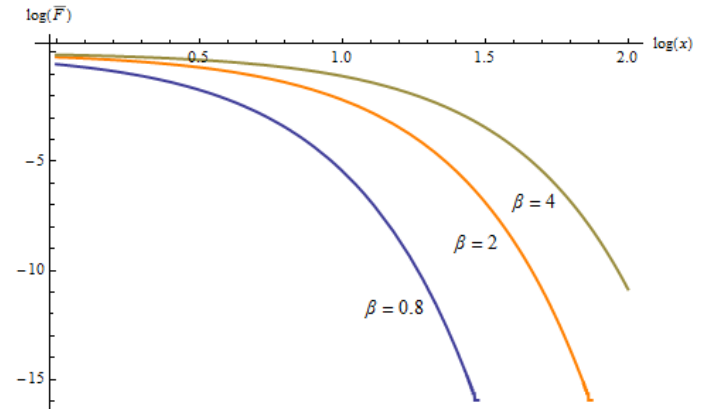
Θεωρητικό Υπόβαθρο

- Ουρά Εκθετικής κατανομής με παράμετρο β :

$$\bar{F}(u) = P(X > u) = \exp\left(-\frac{u}{\beta}\right)$$

- Λογάριθμος ουράς

$$\log[P(X > u)] = -u/\beta$$

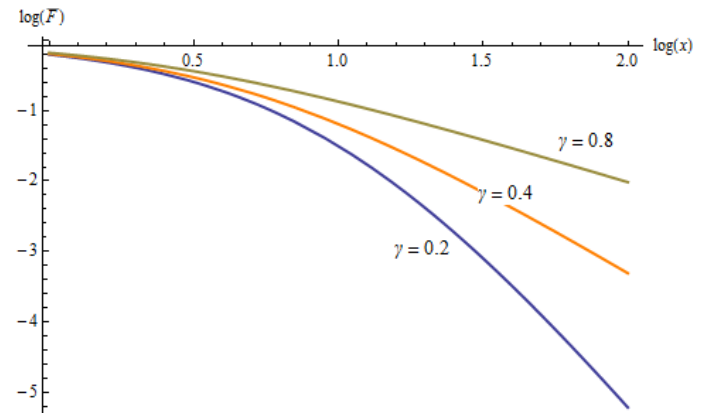


- Ουρά κατανομής δύναμης με παράμετρο α για μεγάλα x :

$$\bar{F}(u) = P(X > u) \approx cx^{-\alpha}$$

- Λογάριθμος ουράς

$$\log[P(X > u)] \approx \log(c) - \alpha \log(u)$$

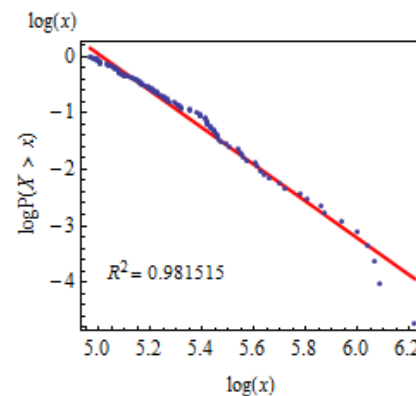
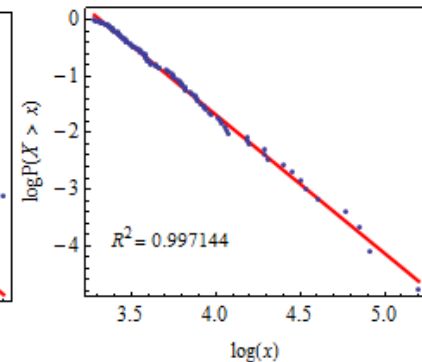
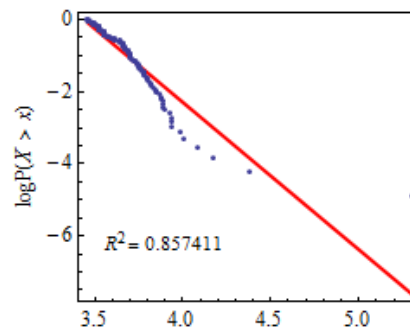
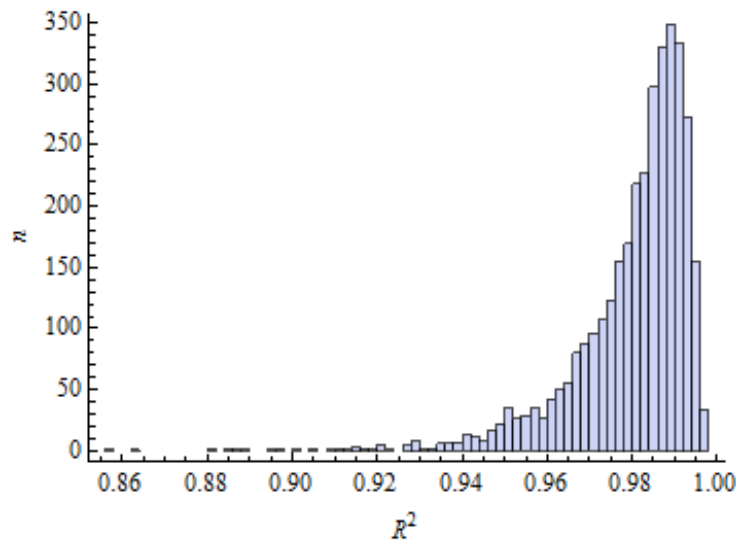


Log-Log Plot

Εφαρμογή στα δεδομένα

- Η εφαρμογή γίνεται για τις N μεγαλύτερες παρατηρήσεις κάθε σταθμού όπου N το πλήθος των ετών κάθε σταθμού (Partial duration series method)
- Προσαρμογή ευθείας στο γράφημα
- Υψηλές τιμές συντελεστή

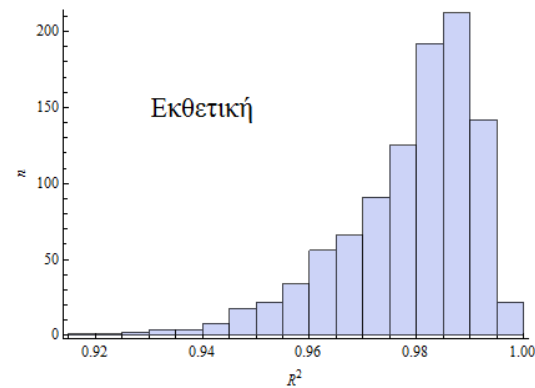
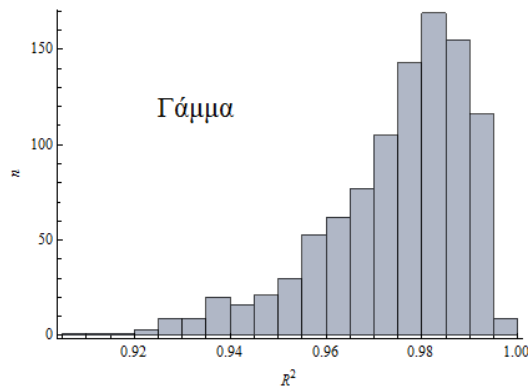
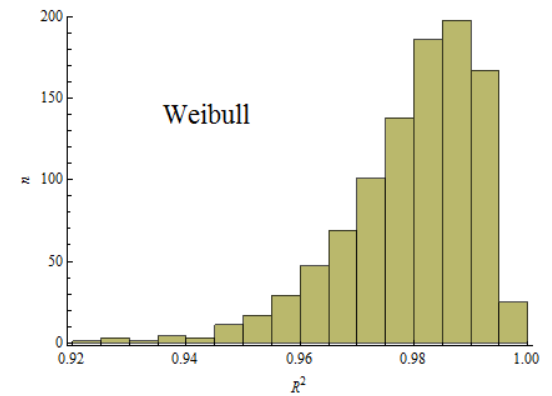
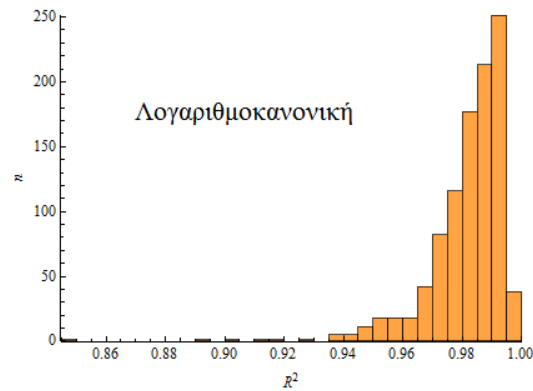
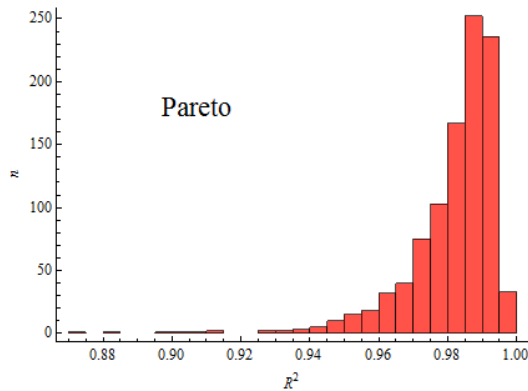
Max	Min	Mean	StDev
0.997	0.857	0.98	0.014



Log-Log Plot

Προσομοίωση Monte Carlo

- 1 000 τυχαία δείγματα για κάθε κατανομή με $n = 9500$ και $N = 116$
- Αποτελέσματα

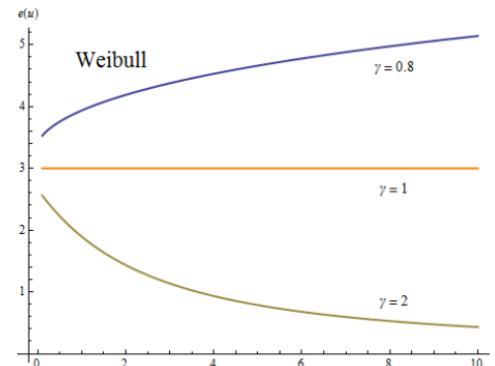
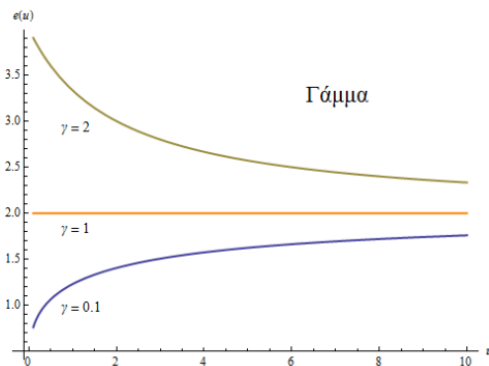
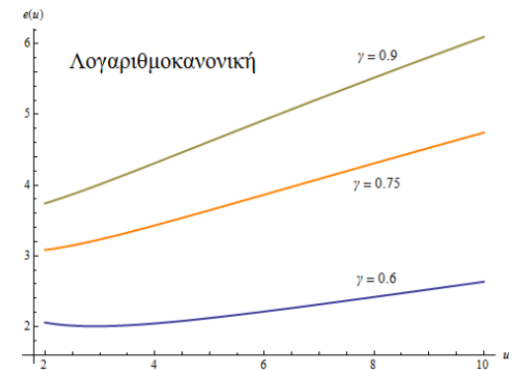
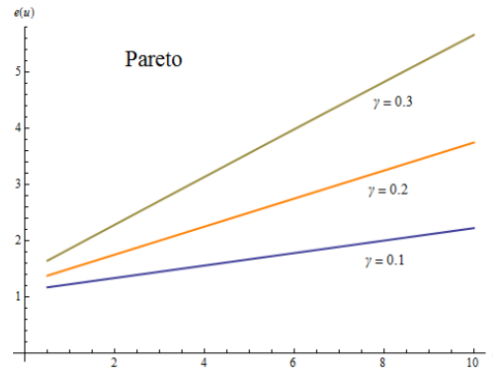
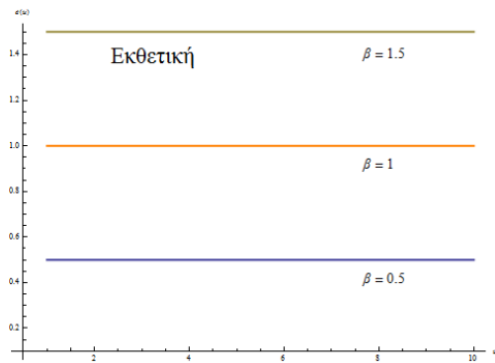


- Ο συντελεστής R^2 δε διαφοροποιείται στις θεωρητικές κατανομές – η μέθοδος εφαρμόστηκε και για $N = 10\% n$ με τα ίδια αποτελέσματα
- Log-Log Plot αναξιόπιστο εργαλείο για διάκριση

Mean Excess Function

Θεωρητικό Υπόβαθρο

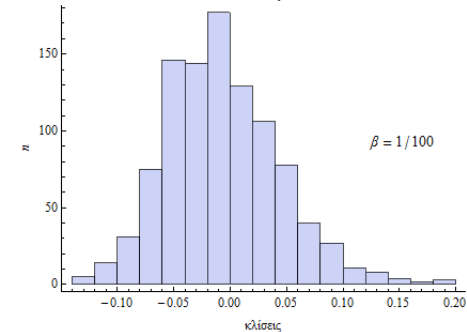
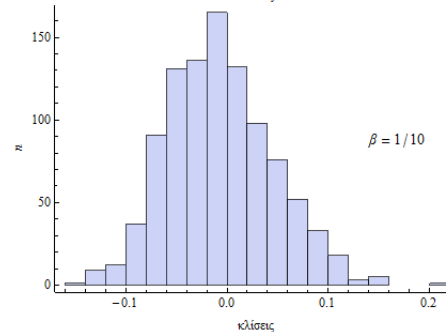
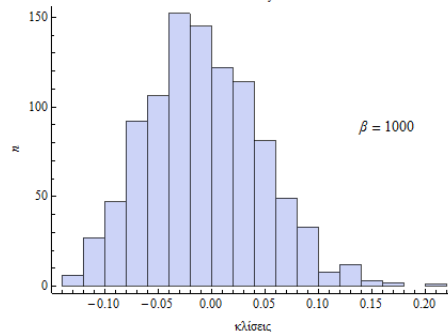
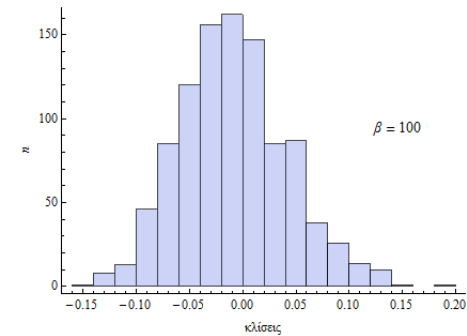
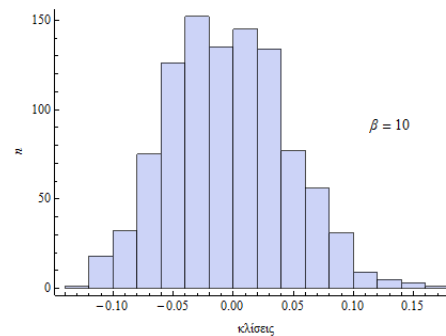
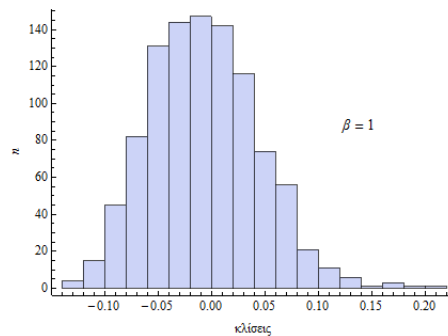
- $F_u(x) = P(X - u \leq x | X > u)$, $x \geq 0$ \Rightarrow Πιθανότητα Υπέρβασης
- $e(u) = E(X - u | X > u)$ \Rightarrow Συνάρτηση Μέσης Υπέρβασης
- Για την Εκθετική κατανομή με παράμετρο β είναι $e(u) = \beta$
- Για δυναμικού τύπου κατανομές με $\alpha > 2$ είναι $e(u) = \frac{u}{(\alpha - 2)}$



Mean Excess Function

Προσομοίωση Monte Carlo (1)

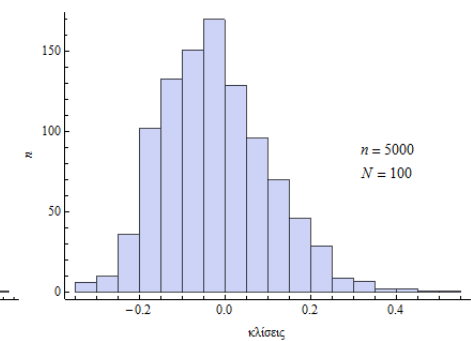
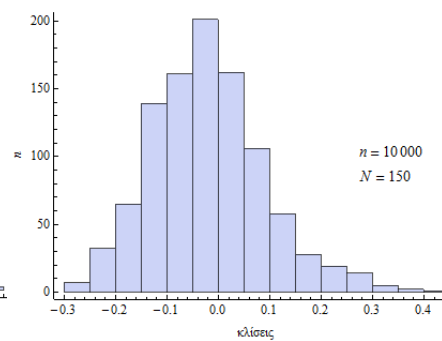
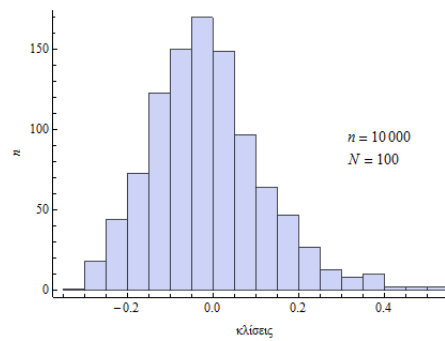
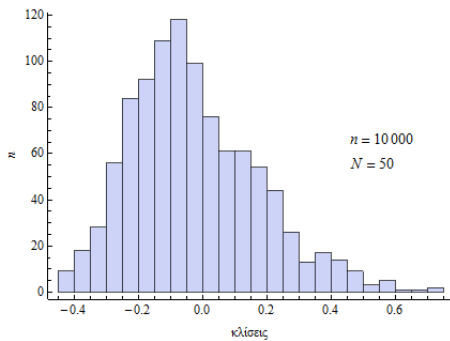
- Έλεγχος για το βαθμό που επηρεάζονται τα διαστήματα των κλίσεων από τη β .
- 1000 τυχαία δείγματα Εκθετικής κατανομής μήκους 10000 με παραμέτρους 1, 10, 100, 0.1, 0.01, 0.001.
- Σ.Μ.Υ. \Rightarrow Προσαρμογή ευθείας \Rightarrow Εύρεση κλίσης.
- Το διάστημα των κλίσεων δεν μεταβάλλεται με το β και για διάστημα εμπιστοσύνης 95% οι κλίσεις είναι στο $[-0.1, 0.1]$



Mean Excess Function

Προσομοίωση Monte Carlo (2)

- Έλεγχος για το βαθμό που επηρεάζονται οι κλίσεις της Σ.Μ.Υ. της Εκθετικής κατανομής από το μήκος ολικού δείγματος και ουράς



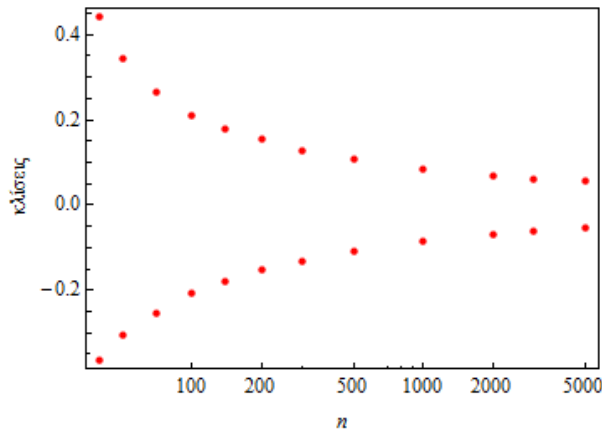
- Προκύπτει ότι μεταβάλλοντας το πλήθος των τιμών των δεδομένων ουράς, το διάστημα των κλίσεων παρουσιάζει σημαντικές αλλαγές ενώ είναι ανεξάρτητο από το μήκος του ολικού δείγματος.

$n = 10\,000$		$n = 5\,000$	
$N = 50$	$N = 100$	$N = 150$	$N = 100$
$[-0.35, 0.43]$	$[-0.24, 0.25]$	$[-0.20, 0.22]$	$[-0.24, 0.25]$

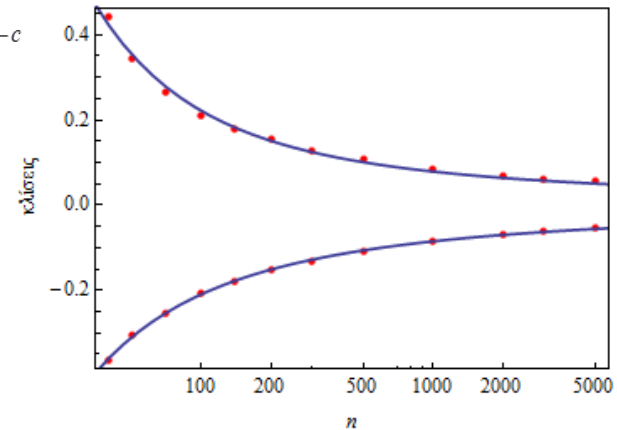
Mean Excess Function

Προσομοίωση Monte Carlo (3) – Διαστήματα εμπιστοσύνης

- Μετάφραση διαδικασίας αλγοριθμικά για να μπορεί να εφαρμοστεί αυτόματα
- 10 000 προσομοιώσεις για διάφορα μήκη ουράς. Για διάστημα εμπιστοσύνης 90%



$$y = b \log \left(1 + \frac{x}{d} \right)^{-c}$$



Δ.Ε.	όρια	b	c	d
90%	άνω όριο	0.41	1.25	23.86
	κάτω όριο	-0.92	1.49	7.14

- Οι κλίσεις ανάμεσα στις καμπύλες είναι κλίσεις εκθετικών ουρών
- Δημιουργία «τεστ» και εφαρμογή στα δεδομένα
- Αυστηρό τεστ

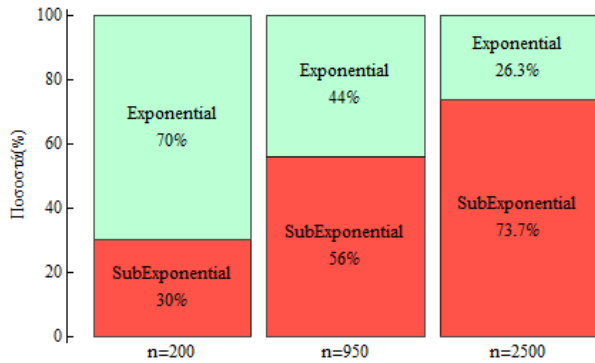
Mean Excess Function

Προσομοίωση Monte Carlo (4)

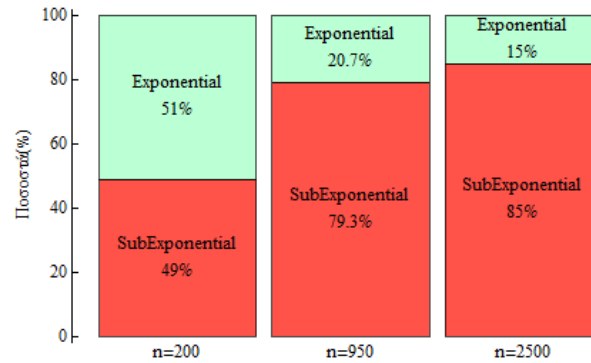
Έλεγχος σε θεωρητικές κατανομές

- Pareto

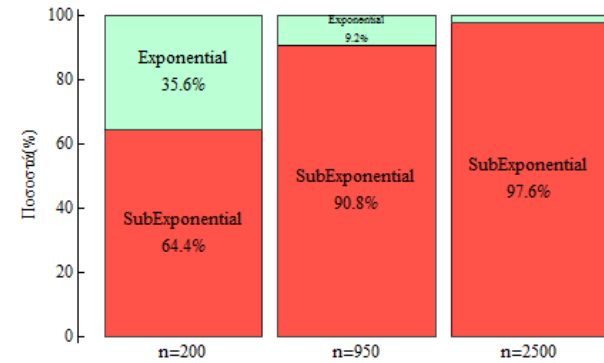
$\gamma = 0.1$



$\gamma = 0.15$

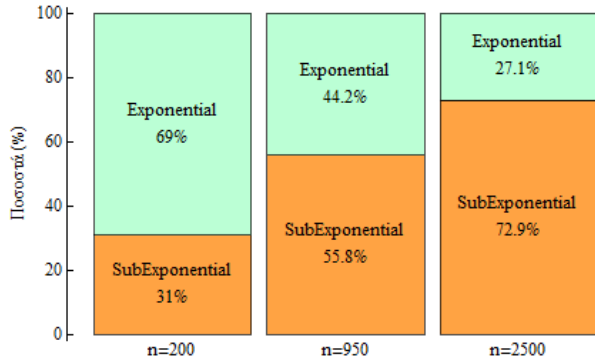


$\gamma = 0.2$

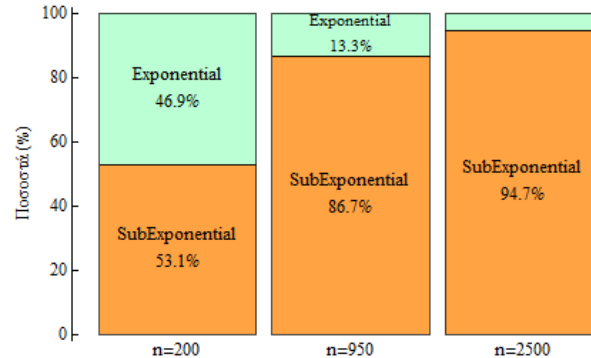


- Λογαριθμοκανονική

$\gamma = 0.6$



$\gamma = 0.8$



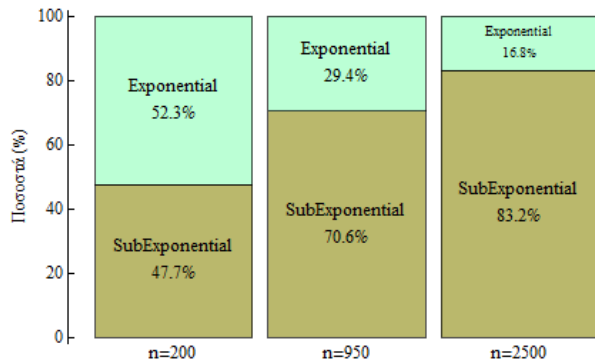
Για μικρή παράμετρο σχήματος το τεστ είναι αποτελεσματικό για μεγάλο μήκος δείγματος. Όσο η παράμετρος αυξάνεται τα αποτελέσματα είναι καλά ακόμα και για μικρό δείγμα

Mean Excess Function

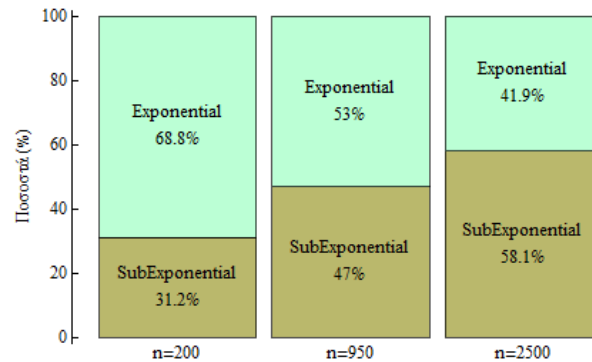
Προσομοίωση Monte Carlo (5)

- Weibull

$\gamma = 0.6$



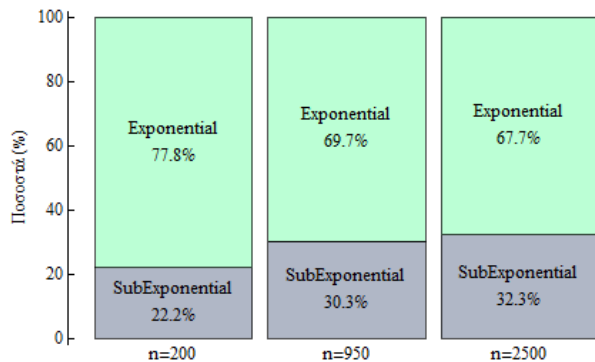
$\gamma = 0.8$



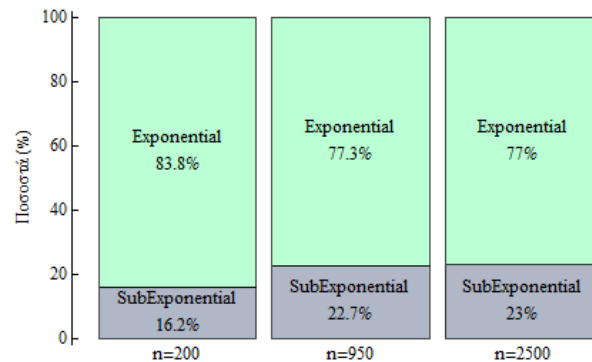
Για τη Weibull η αύξηση της παραμέτρου σχήματος συνεπάγεται «αύξηση» εκθετικότητας. Για $\gamma = 0.8$ η κατάσταση είναι ενδιάμεση.

- Γάμμα

$\gamma = 0.2$



$\gamma = 0.3$



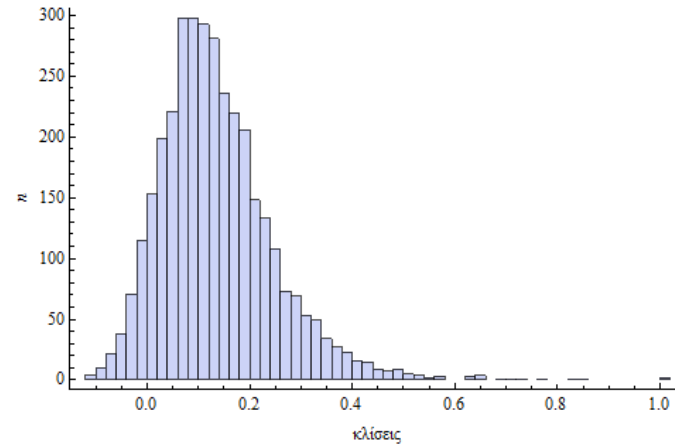
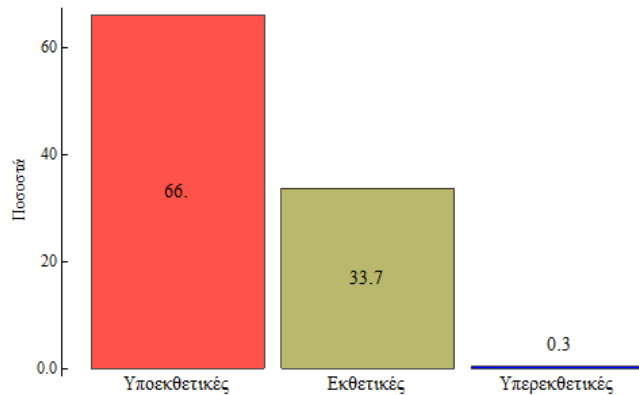
Η Γάμμα έχει υψηλά ποσοστά «εκθετικότητας» αφού έχει λεπτή ουρά.

Mean Excess Function

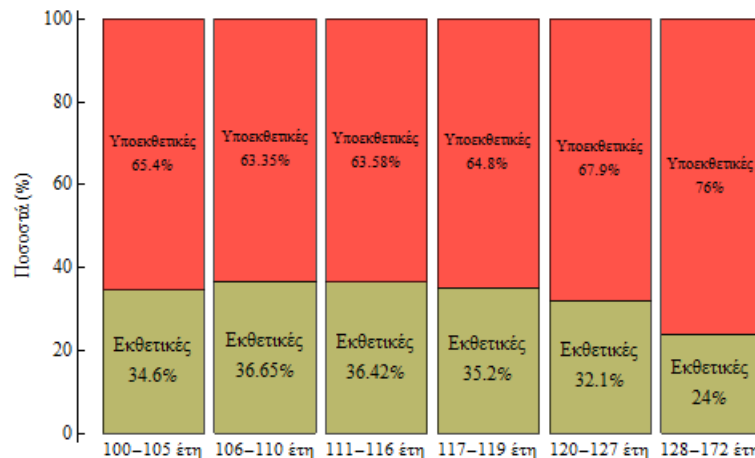
Εφαρμογή στα δεδομένα - Αποτελέσματα (1)

- Ως ουρά θεωρήθηκε το μεγαλύτερο 10% των παρατηρήσεων – Ανάγκη μεγάλου μήκους για καλύτερη απόδοση μεθόδου
- Χαρακτηριστικά κλίσεων σταθμών
- Αποτελέσματα ελέγχου

Min	Q ₁	Διάμεσος	Q ₃	Max
-0.852	0.062	0.121	0.194	2.402

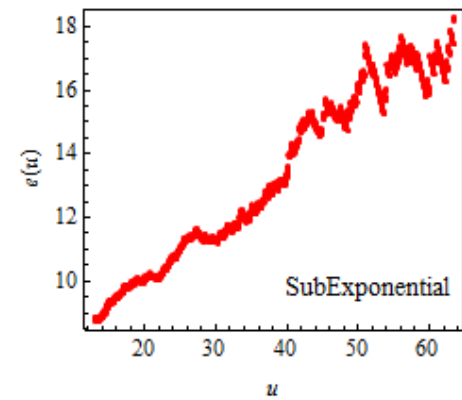
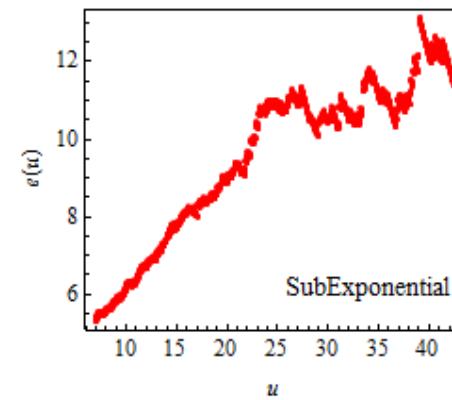
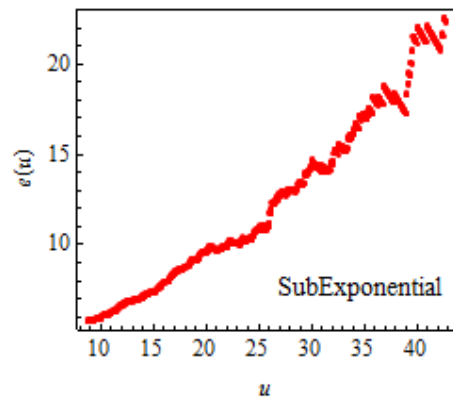
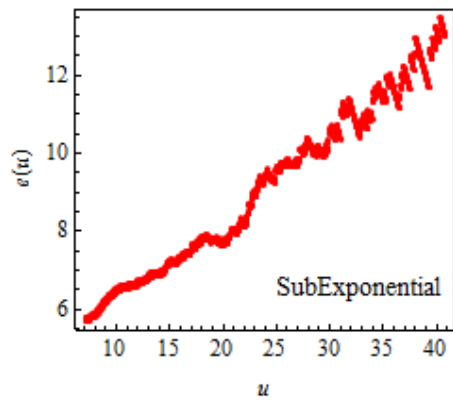
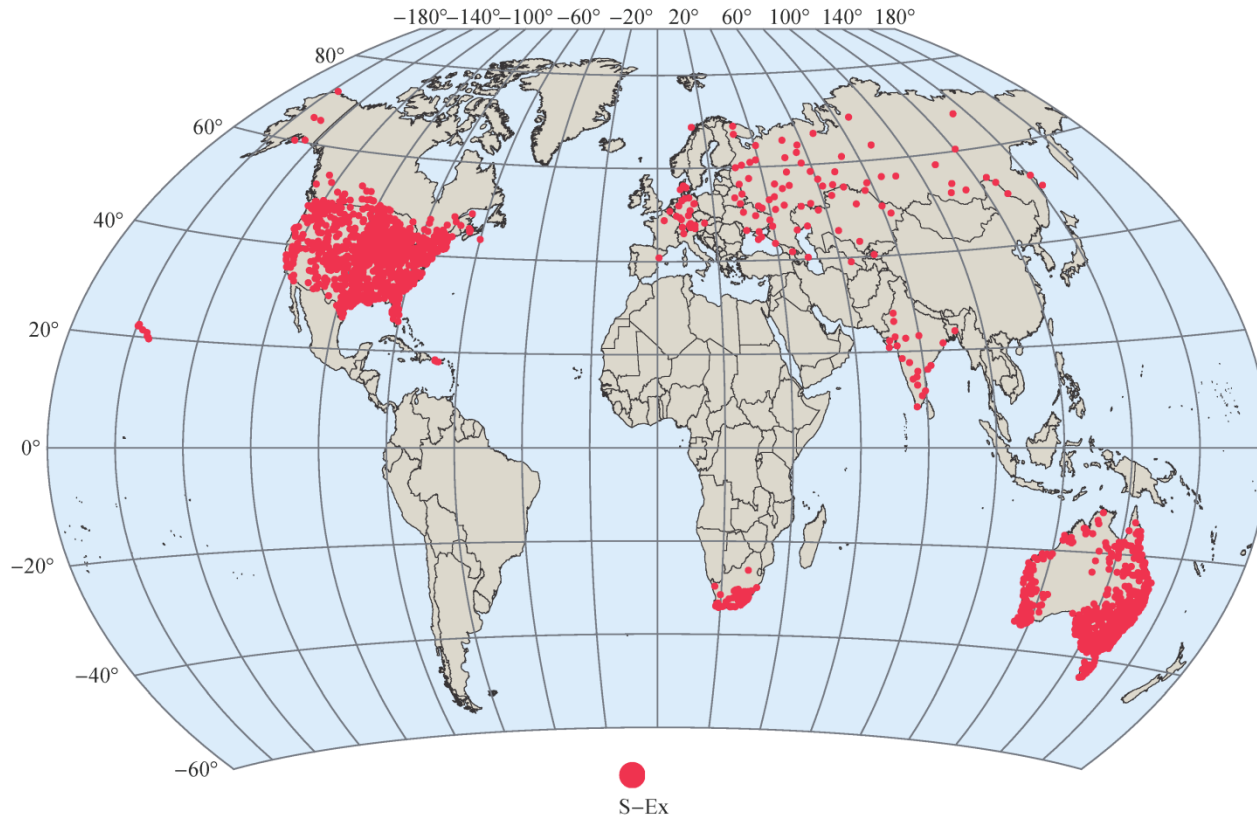


Στην πλειοψηφία
τους υποεκθετικές
ουρές

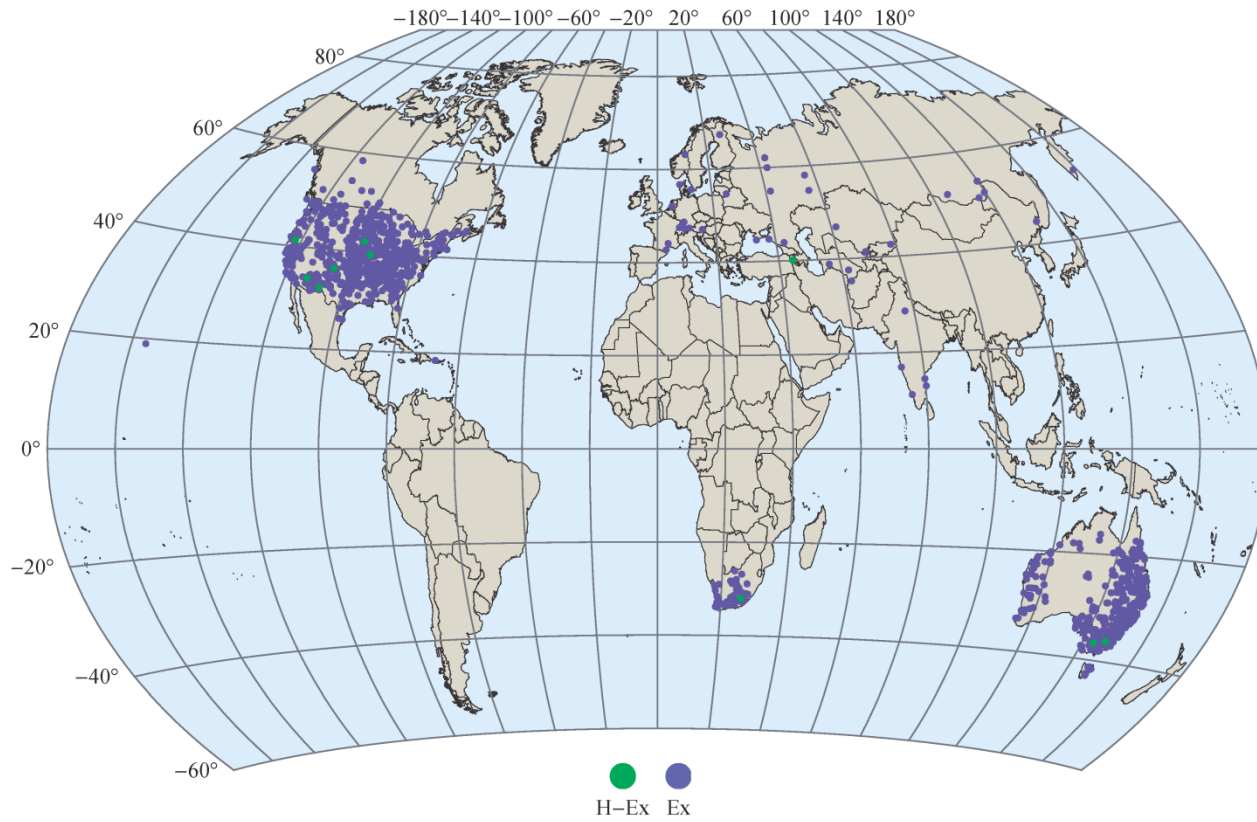


Για πλήθος ετών 111
και άνω το ποσοστό
των εκθετικών
ουρών μειώνεται με
την αύξηση των
ετών

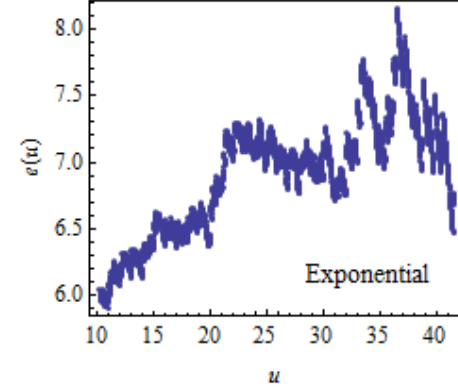
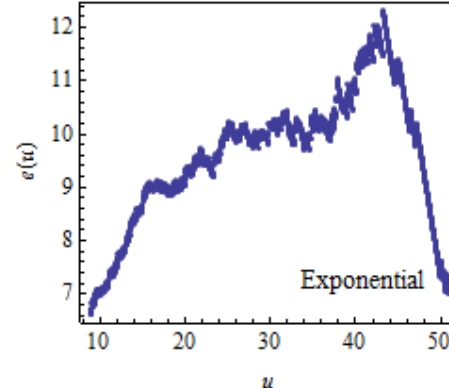
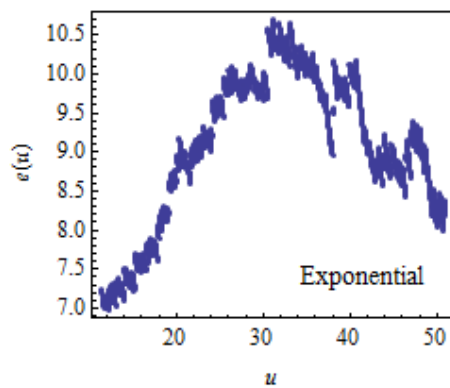
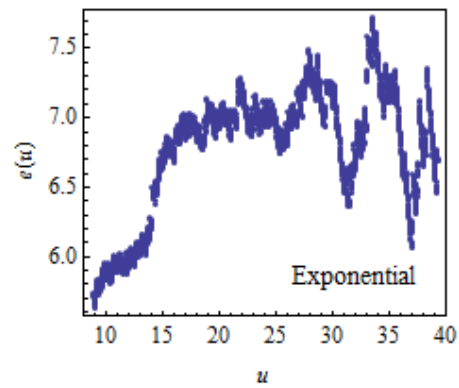
Γεωγραφική Απεικόνιση Υποεκθετικών κατανομών



Γεωγραφική Απεικόνιση Εκθετικών κατανομών



● H-Ex ● Ex



Μέθοδος ελαχιστοποίησης σφάλματος

Σύγκριση τύπων (1)

- Συγκρίθηκαν τέσσερις νόρμες ελαχιστοποίησης σφάλματος:

$$\sum_{i=n-N+1}^n [x'_i - x_i]^2 \quad (1) \qquad \sum_{i=n-N+1}^n \left[\frac{x'_i}{x_i} - 1 \right]^2 \quad (2)$$

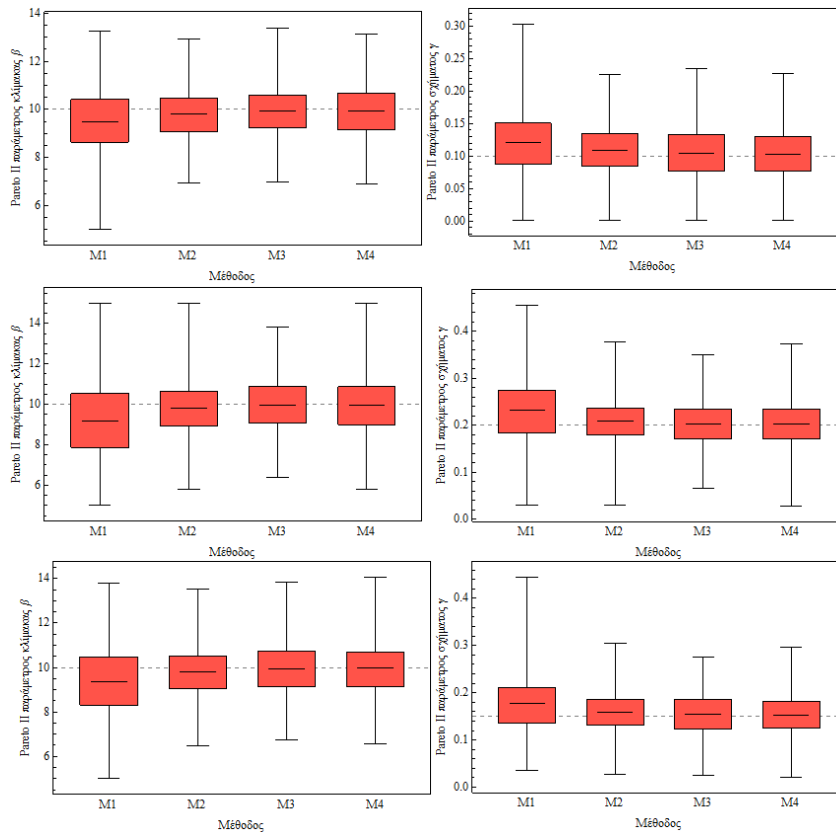
$$\sum_{i=n-N+1}^n [\bar{F}(x'_i) - \bar{F}(x_i)]^2 \quad (3) \qquad \sum_{i=n-N+1}^n \left[\frac{\bar{F}(x_i)}{\bar{F}(x'_i)} - 1 \right]^2 \quad (4)$$

- Μήκος δείγματος $m = 9500$ ίσο με το μέσο όρο των θετικών τιμών των δεδομένων των σταθμών
- Μέγεθος ουράς $N = 116$ ίσο με το μέσο όρο των ετών των καταγραφών
- Οι νόρμες εφαρμόζονται στις θεωρητικές κατανομές και ελαχιστοποιούνται με στόχο την εξίσωση των παραμέτρων

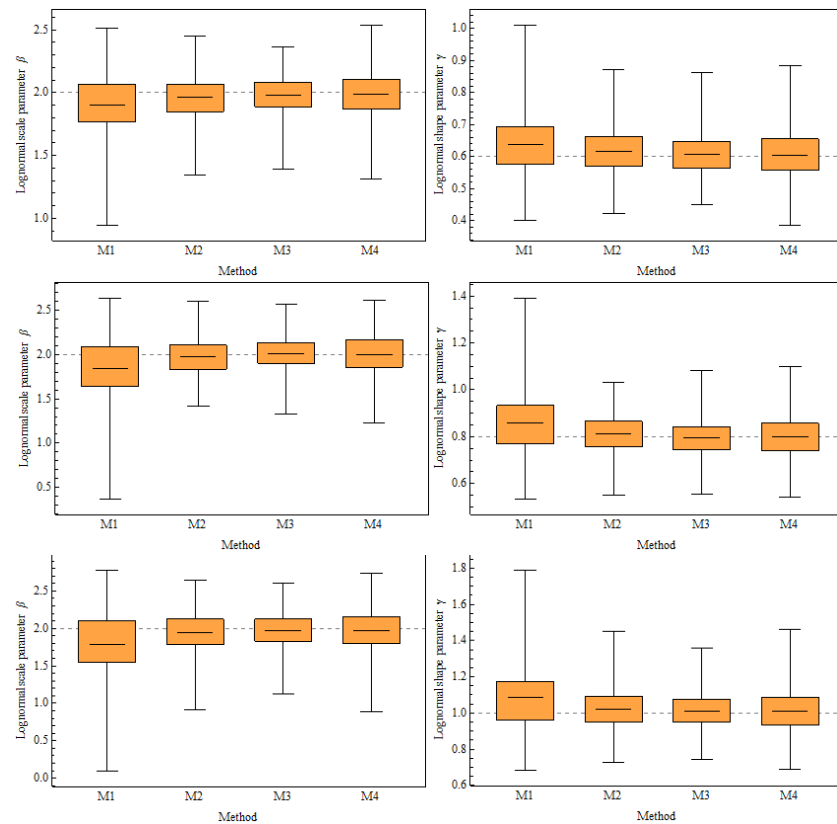
Μέθοδος ελαχιστοποίησης σφάλματος

Σύγκριση τύπων (2)

Pareto



Λογαριθμοκανονική

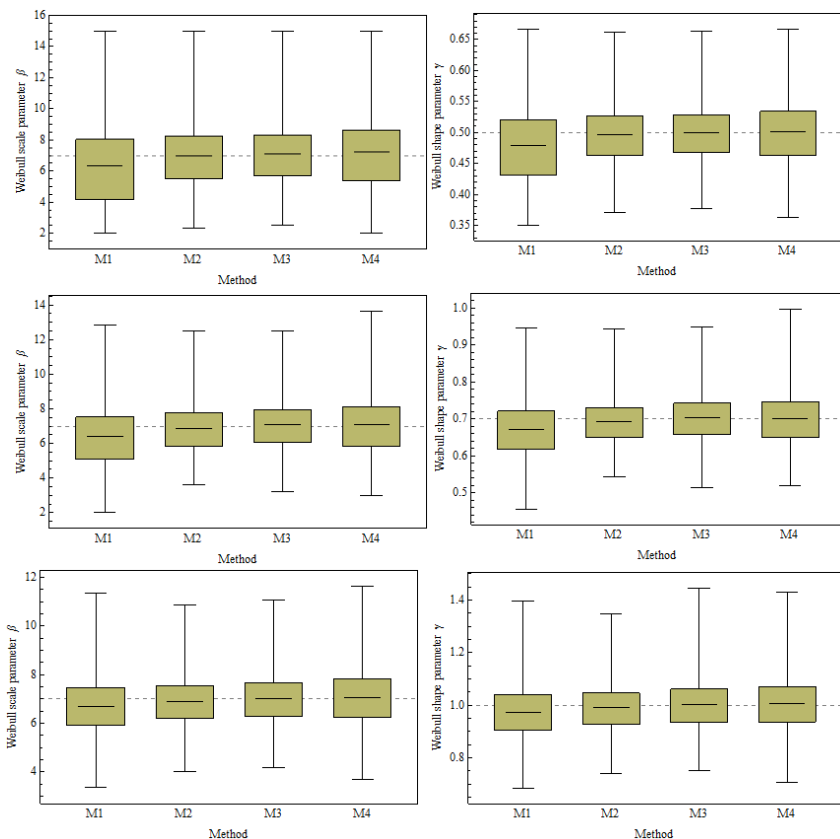


Και στις δύο περιπτώσεις κατανομών καλύτερη επίδοση έχει η νόρμα (4) και για τις δύο παραμέτρους. Ακολουθούν η (3), η (2) και η (1), που είναι η πιο συχνά χρησιμοποιούμενη, έχει τη μεγαλύτερη απόκλιση!

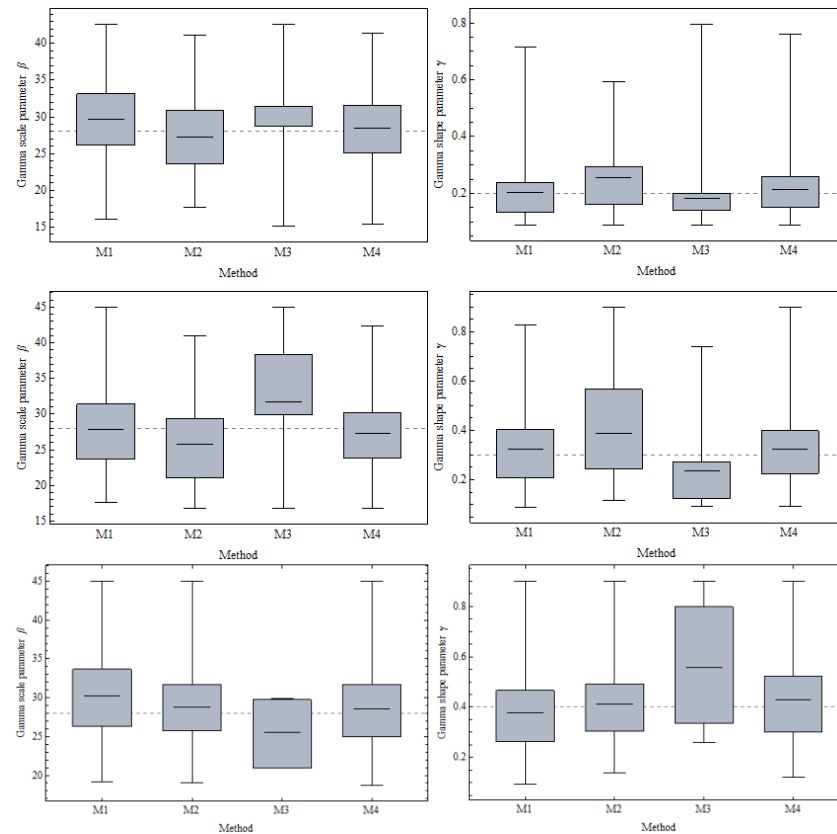
Μέθοδος ελαχιστοποίησης σφάλματος

Σύγκριση τύπων (3)

Weibull



Γάμμα



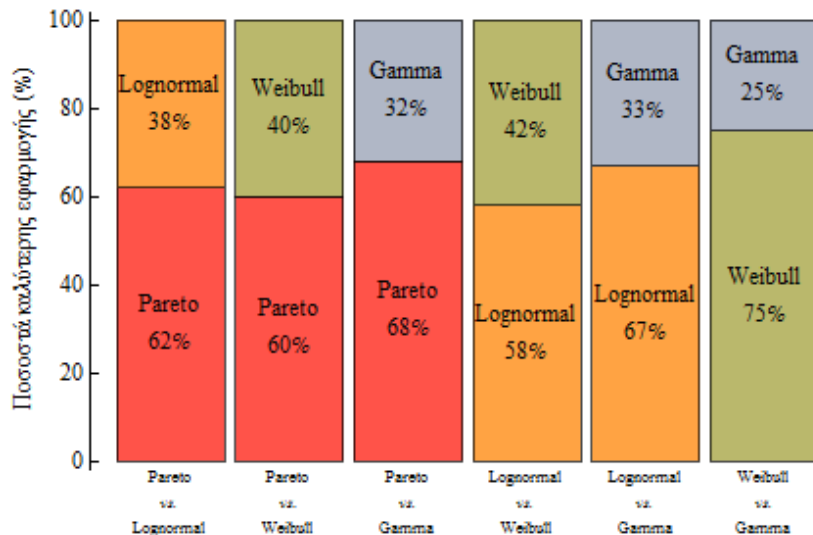
- Με τη μείωση του γ της Weibull, μειώνεται και η απόδοση της νόρμας (1)
- Η νόρμα (3) δίνει τη χειρότερη εκτίμηση για τη Γάμμα

Μέθοδος ελαχιστοποίησης σφάλματος

Εφαρμογή στα δεδομένα – Αποτελέσματα (1)

- Χρησιμοποιείται η νόρμα (4) η οποία ελαχιστοποιείται για την εύρεση της κατάλληλότερης κατανομής. Πραγματοποιούνται τρεις συγκρίσεις:

(α) Σύγκριση ανά δύο

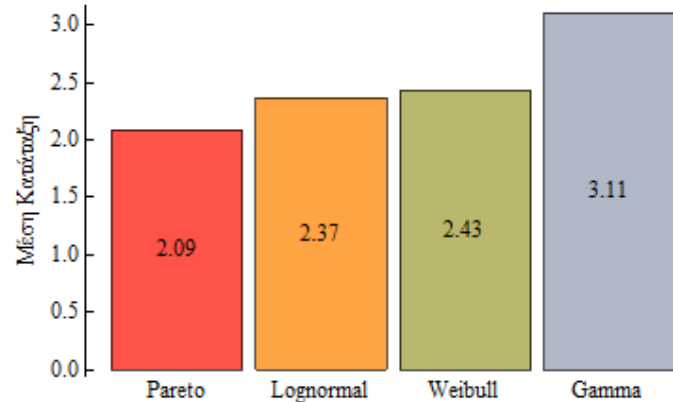


- Η Pareto προσαρμόστηκε καλύτερα στο 60% των σταθμών
- Η Γάμμα, πάντα «ερχόταν δεύτερη» με εφαρμογή στο 30% των σταθμών
- Σε κάθε σύγκριση προέκυπτε πάντα καταλληλότερη στο μεγαλύτερο ποσοστό των σταθμών η πιο βαριά ουρά

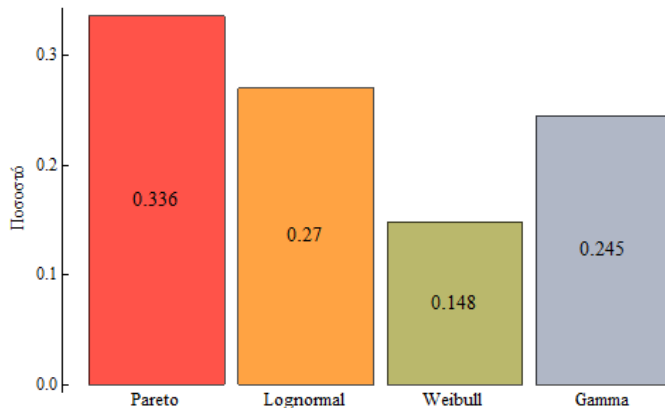
Μέθοδος ελαχιστοποίησης σφάλματος

Εφαρμογή στα δεδομένα – Αποτελέσματα (2)

(β) Κατάταξη επίδοσης
Βαθμονόμηση βάσει σφάλματος:
Μικρότερο σφάλμα → 1
Μεγαλύτερο σφάλμα → 4



(γ) Τελικά ποσοστά

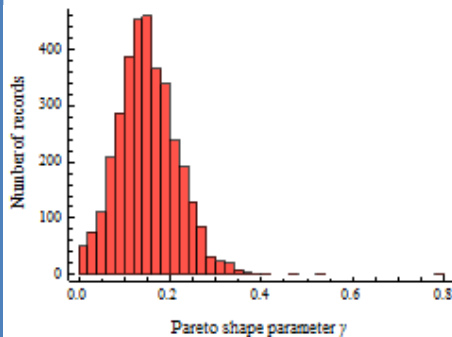


- Μέση κατάταξη των κατανομών στους σταθμούς
- Η Pareto είχε την καλύτερη επίδοση - η Γάμμα και πάλι τη χειρότερη
- Η επίδοση μειώνεται όσο «λεπταίνουν» οι ουρές

Μέθοδος ελαχιστοποίησης σφάλματος

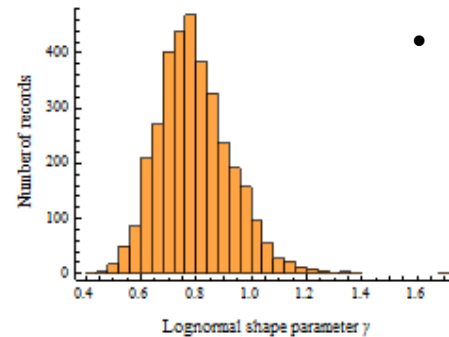
Εφαρμογή στα δεδομένα – Αποτελέσματα (3)

Pareto



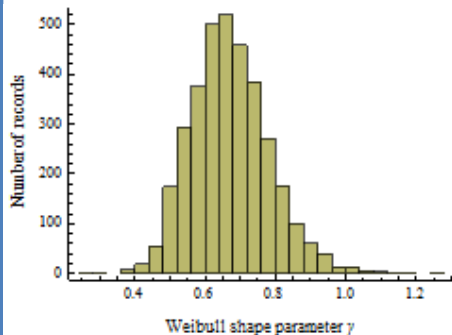
- Πιθανότερη τιμή για την κατανομή Pareto $\gamma = 0.142$

Λογαριθμοκανονική



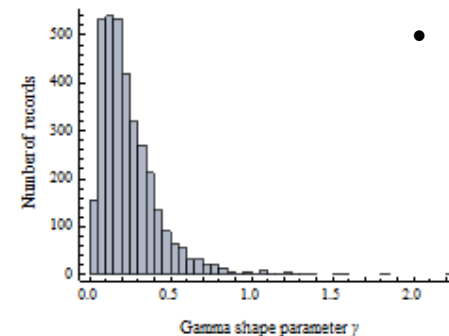
- Πιθανότερη τιμή για τη Λογαριθμοκανονική $\gamma = 0.77$

Weibull



- Πιθανότερη τιμή για την κατανομή Weibull $\gamma = 0.654 < 1$
- Υποεκθετικές ουρές

Γάμμα

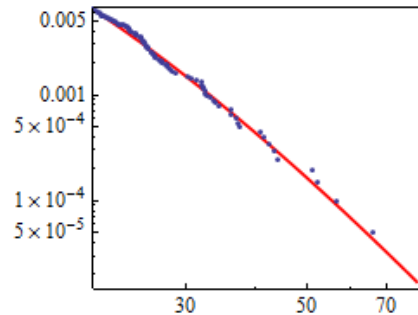


- Πιθανότερη τιμή για τη Γάμμα $\gamma = 0.144$

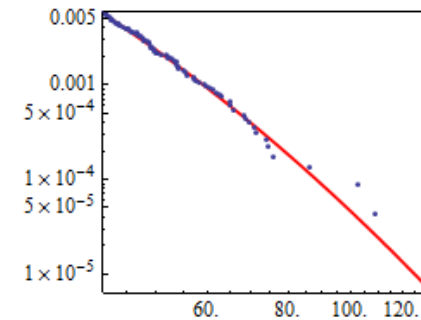
Μέθοδος ελαχιστοποίησης σφάλματος

Εφαρμογή στα δεδομένα – Αποτελέσματα (4)

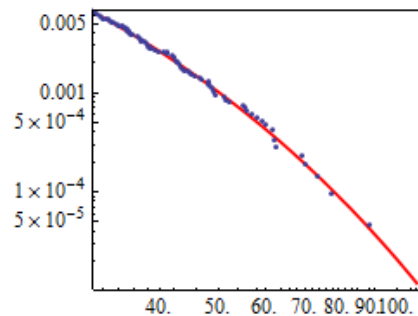
Pareto



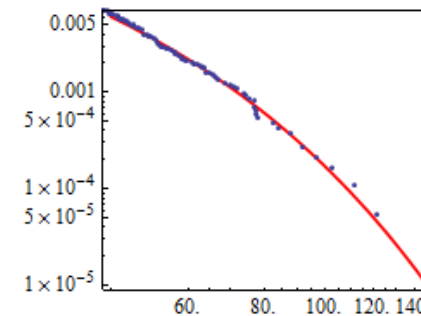
Λογαριθμοκανονική



Weibull



Γάμμα



- Η Γάμμα είχε καλύτερη εφαρμογή στο 24.5% των σταθμών.
- Η Weibull με παράμετρο σχήματος μεγαλύτερη της μονάδας σε 20 σταθμούς.
- Άρα συνολικά στο 25% των σταθμών καλύτερη εφαρμογή είχαν κατανομές με λεπτές ουρές

Συμπεράσματα

Από την εργασία προέκυψε:

- Η μέθοδος απεικόνισης ουράς σε λογαριθμικούς άξονες δεν αποτελεί αξιόπιστο εργαλείο για τη μελέτη ακραίων τιμών όπως προέκυψε από την εφαρμογή της σε γνωστές θεωρητικές κατανομές
- Έγινε επαλήθευση της μεθόδου Σ.Μ.Υ. η οποία μετά την εφαρμογή της στα δεδομένα έδειξε ότι το 33.7% των σταθμών περιγράφονται από εκθετικές ουρές
- Η Μέθοδος ελαχιστοποίησης σφάλματος προτείνει ότι το 25% των σταθμών έχουν λεπτές ουρές.
- Οι δύο μέθοδοι συμφωνούν σε ικανοποιητικό βαθμό και το γεγονός αυτό δίνει μεγαλύτερη αξιοπιστία στα αποτελέσματα
- Η ανάλυση αποδεικνύει ότι οι κατανομές με βαριές ουρές συμφωνούν καλύτερα με τα μέγιστα των βροχοπτώσεων σε σχέση με τις συχνά χρησιμοποιούμενες κατανομές με ελαφριές ουρές

Ευχαριστώ πολύ!

