

ΚΕΦΑΛΑΙΟ 1, Εισαγωγή στη θεωρία ακραίων τιμών

- *Αντικείμενο της θεωρίας ακραίων τιμών αποτελεί:*

Η ανάπτυξη και μελέτη στοχαστικών μοντέλων με σκοπό την **επίλυση προβλημάτων** που σχετίζονται με την **εμφάνιση «πολύ μεγάλων» ή «πολύ μικρών» τιμών** σε τυχαία πειράματα.

- Η θεωρία αυτή μπορεί χοντρικά να χωριστεί σε δύο μέρη:
 - το **πιθανοθεωρητικό** μέρος που αφορά τη μελέτη της στοχαστικής συμπεριφοράς ακραίων παρατηρήσεων, και
 - το **στατιστικό** μέρος που αποτελείται από μοντέλα ερμηνείας, εκτιμήσεων και προβλέψεων ακραίων συμβάντων με βάση πραγματικά δεδομένα.

Εφαρμογές

*Σε πάρα πολλές ερευνητικές περιοχές και αφορούν την **μοντελοποίηση** και **μακρο-πρόθεσμη πρόβλεψη** εμφανίσεων ακραίων συμβάντων, π.χ. :*

- ***Ακραίων καιρικών φαινομένων** όπως τυφώνες, έντονες χιονοπτώσεις, μεγάλες πλημμύρες κ.ο.κ. (Μετεωρολογία)*
- ***Μεγάλων σεισμικών δονήσεων** σε συγκεκριμένες γεωγραφικές περιοχές (Σεισμολογία)*
- ***Πολύ υψηλών ή πολύ χαμηλών επιπέδων στάθμης των υδάτων** σε συγκεκριμένες λίμνες, ποτάμια, φράγματα κ.ο.κ. (Υδρολογία)*
- ***Καταστροφικών βλαβών ή αποτυχιών μηχανημάτων ή εξαρτημάτων**, δηλαδή αποτυχιών που γίνονται σε σύντομο χρονικό διάστημα από την έναρξη λειτουργίας τους, πολύ πριν το μέσο χρόνο λειτουργίας ή ζωής τους (Αξιοπιστία Συστημάτων)*
- ***Ηλιακών καταιγίδων** που επηρεάζουν τις τηλεπικοινωνίες ή εκρήξεων υπερκαινοφανών (Αστρονομία)*

Κ.Ο.Κ

Εφαρμογές στην Ασφάλιση και στα Χρηματοοικονομικά

Τα περισσότερα παραπάνω παραδείγματα μπορεί να οδηγήσουν

- σε απαιτήσεις πολύ *μεγάλων αποζημιώσεων*, ή
- σε πολύ *μεγάλες ζημιές σε χαρτοφυλάκια* που περιέχουν τίτλους των οποίων η αξία μπορεί να επηρεάζεται από ακραία ή γενικότερα «σπάνια» συμβάντα.

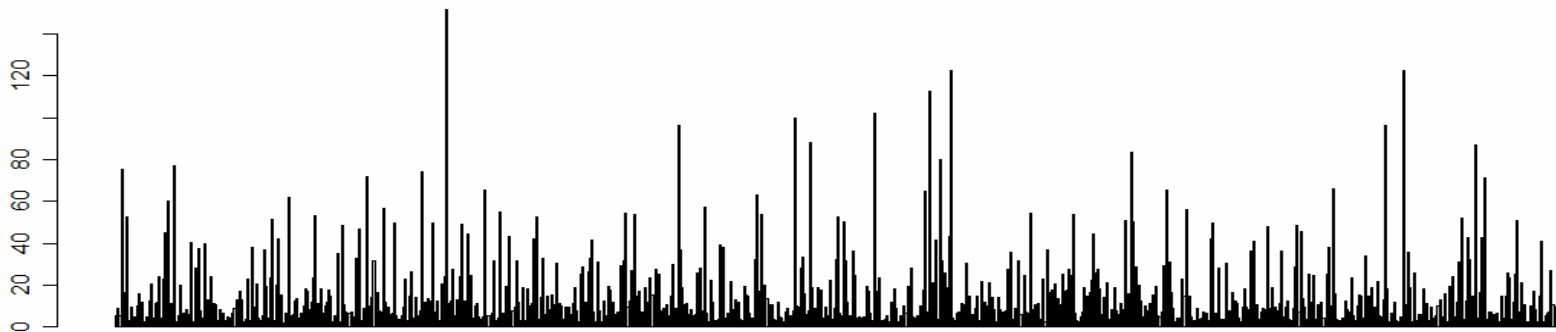
Τυπικές ερωτήσεις

- Ποια είναι η πιθανότητα η ταχύτητα του ανέμου σε μια συγκεκριμένη περιοχή, το επόμενο έτος, να υπερβεί ένα προκαθορισμένο υψηλό κατώφλι;
- Δεδομένων των μετρήσεων της στάθμης του ύδατος σε ένα φράγμα τα τελευταία 10 έτη, ποια είναι η πρόβλεψη για την υψηλότερη στάθμη στα επόμενα 100 χρόνια;
- Ποιο είναι το ύψος του ποσού το οποίο δεν θα υπερβεί καμία απαίτηση ζημιάς (για το επόμενο έτος) με πιθανότητα 0.1%;.

Κ.Ο.Κ.

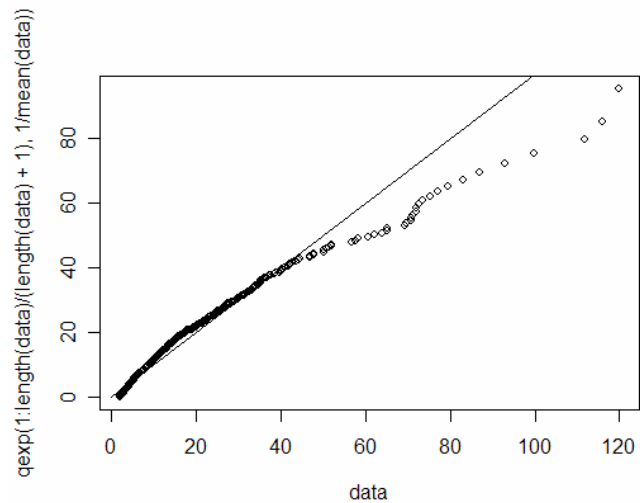
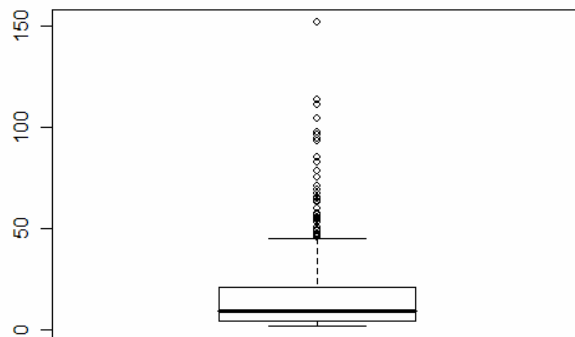
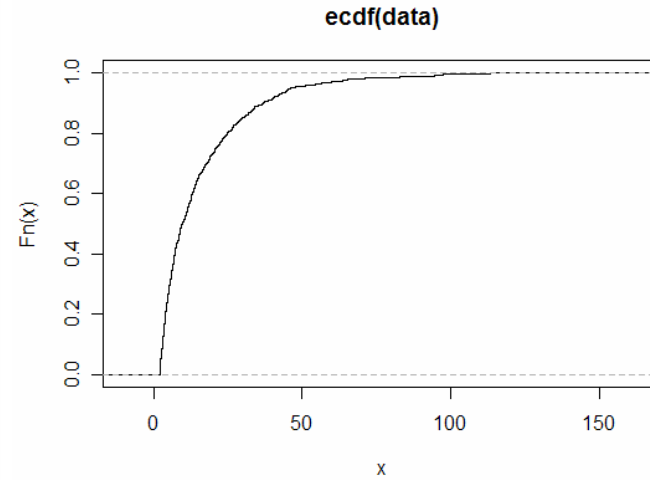
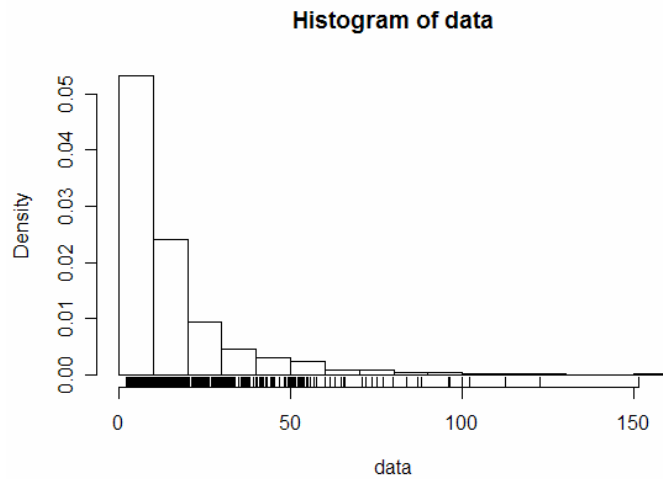
Παράδειγμα

- Μία μεγάλη ασφαλιστική εταιρία έχει καταγράψει τις **ημερήσιες απαιτήσεις ζημίας** που κλήθηκε να καλύψει τα προηγούμενα τρία έτη. (> 2.000 ευρώ).
- Θεωρούμε ότι *δεν εμφανίζουν κάποια περιοδικότητα* και είναι μεταξύ τους *ανεξάρτητες*. (πρόκειται για ημερήσιες απαιτήσεις, και επομένως, αν υπάρχει κάποιο ακραίο γεγονός που «γεννά» πολλές απαιτήσεις ταυτόχρονα, αυτές θα μετρηθούν ως μια απαίτηση).
- Τα δεδομένα καταγράφονται στο επόμενο γράφημα: στον οριζόντιο άξονα αναπαρίσταται *ο χρόνος*, ενώ στον κάθετο άξονα αναπαρίσταται *το ύψος της ημερήσιας απαίτησης ζημίας* (σε χιλιάδες ευρώ):



Min.	1st Quarter	Median	Mean	3rd Quarter	Max.
2.018	4.394	9.167	15.400	18.660	151.500

Επίσης το ιστόγραμμα, το γράφημα της εμπειρικής συνάρτησης κατανομής, το θηκόγραμμα, καθώς και το $Q-Q$ plot ως προς την εκθετική κατανομή αντίστοιχα είναι



- Διαφαίνεται ότι οι παρατηρήσεις ίσως προέρχονται από κάποια κατανομή **με βαριά δεξιά ουρά**. (Χοντρικά μπορούμε να πούμε ότι ορισμένες παρατηρήσεις μπορεί να είναι πάρα πολύ μεγάλες σε σχέση με τις περισσότερες).
- Παρατηρούμε ότι, ενώ η μέση παρατήρηση είναι 15.4, έχουμε πολλές παρατηρήσεις πάνω από 30, 40 ακόμη και 50, με μεγαλύτερη την 151.5.
- Επίσης από το παραπάνω Q-Q plot των παρατηρήσεων ως προς την εκθετική κατανομή, διαφαίνεται ότι η δεξιά ουρά της κατανομής των παρατηρήσεων είναι *πιο βαριά από την εκθετική*.

Ερωτήματα:

- ποια είναι η κατανομή των «πολύ υψηλών» απαιτήσεων ζημιάς;
- ποιο είναι το κατώφλι c_p το οποίο δεν θα υπερβεί καμία απαίτηση ζημιάς (π.χ. για τις επόμενες n ημέρες) με κάποια δεδομένη πιθανότητα $1 - p$ (π.χ. 99%);

Κ.Ο.Κ.

Μια πρώτη προσέγγιση του προβλήματος:

- Υποθέτουμε ότι οι παρατηρήσεις $X_1, \dots, X_n$ (π.χ. απαιτήσεις ζημιάς) προέρχονται από κάποια συγκεκριμένη κατανομή F_θ (π.χ. κανονική ή εκθετική),
- Εκτιμούμε τις άγνωστες παραμέτρους θ (π.χ. μ, σ^2) από τα δεδομένα
- Μελετούμε την συμπεριφορά της ουράς με βάση την $F_{\hat{\theta}}$

Για παράδειγμα:

Το κατώφλι c_p που αναζητείται παραπάνω προσδιορίζεται από την εξίσωση

$$\begin{aligned} P(\max\{X_1, \dots, X_n\} < c_p) &= 1 - p \Leftrightarrow P(X_1 < c_p, \dots, X_n < c_p) = 1 - p \\ &\Leftrightarrow P(X_1 < c_p) \dots P(X_n < c_p) = 1 - p \\ &\Leftrightarrow F_\theta(c_p)^n = 1 - p \\ &\Leftrightarrow c_p = F_\theta^{-1}((1 - p)^{1/n}) \end{aligned}$$

και επομένως

$$\hat{c}_p = F_{\hat{\theta}}^{-1}((1 - p)^{1/n})$$

Η παραπάνω διαδικασία **σπάνια** θα οδηγήσει σε ασφαλείς προβλέψεις διότι:

- $(1 - p)^{1/n} \rightarrow 1$ και επομένως η $\hat{c}_p = F_{\hat{\theta}}^{-1}((1 - p)^{1/n})$ βασίζεται αποκλειστικά στη μορφή της δεξιάς ουράς της πραγματικής κατανομής των δεδομένων

Παρατηρούμε όμως ότι:

(1) Η επιλογή της F_{θ} είναι **αυθαίρετη** και τίποτε δεν εγγυάται ότι η δεξιά ουρά της παρουσιάζει την ίδια συμπεριφορά με την πραγματική ουρά των δεδομένων.

(2) Οι εκτιμήσεις, $\hat{\theta}$, πιθανότατα θα γίνουν με βάση **το σύνολο των παρατηρήσεων** και επομένως ελάχιστα θα βασίζονται στις **ακραίες παρατηρήσεις** (που υποδηλώνουν τη μορφή της δεξιάς ουράς της πραγματικής κατανομής των δεδομένων).

(3) Ακόμη και να έχει γίνει σωστή επιλογή της F_{θ} , ελάχιστες αποκλίσεις των εκτιμήσεων $\hat{\theta}$ από τις πραγματικές τιμές θ , μπορεί να οδηγήσει **σε πολύ μεγάλες αποκλίσεις στις εκτιμήσεις** που γίνονται «κοντά» στο δεξί όριο της F_{θ} .

(4) Πολλές φορές έχουμε στη διάθεσή μας είτε **μέγιστες τιμές ανά χρονικές υποπεριόδους** (π.χ. μέγιστα ανά εβδομάδα, μήνα ή έτος) **ή περικομμένα δεδομένα** (π.χ. παρατηρήσεις μεγαλύτερες από μία συγκεκριμένη τιμή) με αποτέλεσμα να μην μπορούμε εύκολα να εκτιμήσουμε την F_{θ} .

- **Επομένως:**

Θα πρέπει αναγκαστικά να επικεντρωθούμε αποκλειστικά στην **μελέτη της συμπεριφοράς της δεξιάς ουράς** της κατανομής των παρατηρήσεων.

- **Ειδικότερα θα πρέπει να βασιστούμε:**

- Στην **κατανομή της μέγιστης παρατήρησης** (ανά χρονική περίοδο) ή
- Στην **κατανομή των παρατηρήσεων πάνω από ένα υψηλό κατώφλι**.

- Οι παραπάνω κατανομές προφανώς εξαρτώνται από την κατανομή F των παρατηρήσεων $X_1, \dots, X_n$. Από την στιγμή όμως που η F δεν είναι γνωστή ή δεν μπορεί να εύκολα να εκτιμηθεί, θα πρέπει με κάποιο τρόπο να βασιστούμε

- Στην **οριακή κατανομή της μέγιστης παρατήρησης** (αν υπάρχει) ή
- Στην **οριακή κατανομή των παρατηρήσεων πάνω από ένα υψηλό κατώφλι** (αν υπάρχει)

Η μορφή της κατανομής της μέγιστης παρατήρησης

- Έστω $X, X_1, X_2, \dots$ iid, μη ιδιάζουσες τ.μ. με κοινή σ.κ. F και

$$M_n = \max\{X_1, X_2, \dots, X_n\}.$$

- Από την μελέτη της M_n μπορούν εύκολα να προκύψουν συμπεράσματα και για την ελάχιστη τιμή διότι, $\min\{X_1, X_2, \dots, X_n\} = -\max\{-X_1, -X_2, \dots, -X_n\}$,

- Η ακριβής σ.κ. της M_n θα είναι

$$P(M_n \leq x) = P(\max\{X_1, X_2, \dots, X_n\} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = F(x)^n, x \in \mathbb{R}.$$

- Δεν πρέπει να βασιστούμε στην F , θα πρέπει να προσπαθήσουμε να προσεγγίσουμε την παραπάνω κατανομή, για μεγάλο n , ανεξαρτήτως της F

Η σύγκλιση της κατανομής της M_n όταν $n \rightarrow \infty$

- Συμβολίζουμε με x_F , το δεξιό άκρο του στηρίγματος της F ,

$$x_F = \sup \{x \in R : F(x) < 1\} ,$$

Το x_F μπορεί να είναι πεπερασμένο, αλλά μπορεί να ισούται και με άπειρο.

- Είναι διαισθητικά προφανές ότι, για $n \rightarrow \infty$,

$$M_n = \max \{X_1, X_2, \dots, X_n\} \rightarrow x_F$$

Πράγματι:

- αν $x_F = \infty$ τότε, για $n \rightarrow \infty$, $P(M_n \leq x) = F(x)^n \rightarrow 0, x \in R$,

- ενώ αν $x_F < \infty$, τότε $P(M_n \leq x) = F(x)^n \rightarrow \begin{cases} 0, & x < x_F \\ 1, & x > x_F \end{cases}$

Ομοιότητες με τον Νόμο των Μεγάλων Αριθμών και το ΚΟΘ

- Παρατηρούμε ότι η σχέση

$$M_n \rightarrow x_F$$

θυμίζει την (Νόμος Μεγάλων Αριθμών)

$$\bar{X}_n \rightarrow \mu$$

- Η ταχύτητα της σύγκλισης $\bar{X}_n \rightarrow \mu$ περιγράφεται από το ΚΟΘ (για $\sigma < \infty$)

$$\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \rightarrow_d N(0,1).$$

- Παρόμοια, θα μπορούσαμε, «τυποποιώντας» το M_n , να πάρουμε κάτι ανάλογο;

$$\frac{M_n - d_n}{c_n} \rightarrow_d G$$

Ερώτημα: κάτω από ποιες συνθήκες ισχύει το παραπάνω, και αν ισχύει, ποιες μπορεί να είναι οι ακολουθίες $d_n, c_n, n = 1, 2, \dots$ και η οριακή κατανομή G ;

Παράδειγμα 1.2. ($X_i \sim \text{Εκθετική}$)

- Θεωρούμε ότι οι iid τ.μ. $X_1, X_2, \dots, X_n \sim \text{Εκθετική}(\lambda)$.
- Αν πάρουμε $d_n = \lambda^{-1} \ln n$, $c_n = 1/\lambda$, τότε

$$\begin{aligned} P\left(\frac{M_n - d_n}{c_n} \leq x\right) &= P\left(\frac{M_n - \frac{\ln n}{\lambda}}{\frac{1}{\lambda}} \leq x\right) = P\left(M_n \leq \frac{\ln n + x}{\lambda}\right) = \left(1 - \overline{F}\left(\frac{\ln n + x}{\lambda}\right)\right)^n \\ &= \left(1 - e^{-\lambda \frac{\ln n + x}{\lambda}}\right)^n = \left(1 - \frac{e^{-x}}{n}\right)^n \rightarrow e^{-e^{-x}}, \quad x \in \mathbb{R}, \end{aligned}$$

Δηλαδή, το κανονικοποιημένο μέγιστο συγκλίνει στην σ.κ. $\Lambda(x) = e^{-e^{-x}}$, $x \in \mathbb{R}$.

Ερωτήματα:

- (α) το ίδιο ισχύει για όλες τις κατανομές;
- (β) η κατανομή Λ (Gumbel) είναι η μοναδική οριακή κατανομή οριακού μέγιστου;
- (γ) αν αλλάξουμε τα d_n , c_n μπορεί να προκύψει διαφορετική οριακή κατανομή;

Γενικότερα: Αν $X_1, X_2, \dots, X_n \sim F$ (άγνωστη) που μπορεί να συγκλίνει η M_n ;

- Για οποιεσδήποτε ακολουθίες d_n, c_n θα ισχύει ότι

$$\begin{aligned}
 P\left(\frac{M_n - d_n}{c_n} \leq x\right) &= P(M_n \leq d_n + c_n x) \\
 &= P(X_1 < d_n + c_n x, \dots, X_n < d_n + c_n x) \\
 &= (1 - \bar{F}(d_n + c_n x))^n = \left(1 - \frac{n\bar{F}(d_n + c_n x)}{n}\right)^n
 \end{aligned}$$

Επομένως, αποδεικνύεται η ακόλουθη πρόταση*:

Πρόταση 1.3. Για $n \rightarrow \infty, x \in \mathbb{R}$,

$$P\left(\frac{M_n - d_n}{c_n} \leq x\right) \rightarrow e^{-\lambda(x)} \quad \text{αν και μόνο αν} \quad n\bar{F}(d_n + c_n x) \rightarrow \lambda(x)$$

όπου $\lambda(x) \in [0, \infty]$.

*Το αποτέλεσμα αυτό ουσιαστικά οφείλεται στο νόμο των μικρών αριθμών (προσέγγιση της Διωνυμικής $Y = \sum_{i=1}^n I(X_i > d_n + c_n x)$ από την κατανομή Poisson).

Παραδείγματα μη ύπαρξης οριακής κατανομής του μέγιστου

- Υπάρχουν περιπτώσεις κατανομών F για τις οποίες δεν μπορούν να βρεθούν ακολουθίες c_n, d_n , ώστε, για κάποια μη-εκφυλισμένη κατανομή G , να ισχύει ότι

$$(M_n - d_n)/c_n \rightarrow_d G.$$

- Π.χ. όταν $x_F < \infty$ και $P(X_i = x_F) > 0$ (π.χ. X_i διακριτή με πεπερασμένο πλήθος δυνατών τιμών)

- Για να υπάρχει μια οριακή μη εκφυλισμένη G θα πρέπει $\lim_{x \uparrow x_F} \bar{F}(x)/\bar{F}(x-) = 1$

(Αν $X_i \sim$ διακριτή κατανομή στο $N = \{0, 1, \dots\}$, θα πρέπει $\bar{F}(n)/\bar{F}(n-1) \rightarrow 1$)

- Π.χ. για την κατανομή Poisson^{*} (λ) και τη Γεωμετρική^{*} (p) κατανομή ισχύει ότι $\bar{F}(n)/\bar{F}(n-1) \rightarrow 0$ και $1-p < 1$ αντίστοιχα και επομένως, για αυτές τις κατανομές, δεν υπάρχει οριακή κατανομή G .

*Σε αυτές τις περιπτώσεις, η κατανομή του $(M_n - d_n)/c_n$ «συγκεντρώνεται» σε μία περιοχή αλλά λόγω του διακριτού της κατανομής «ταλαντεύεται». Για **κατάλληλες υποακολουθίες** της $(M_n - d_n)/c_n$ θα υπάρχουν διαφορετικές οριακές κατανομές (πρόκειται για μετατοπίσεις της ίδιας G).

Κατανομές που μπορούν να προκύψουν ως όρια μεγίστων:

Max-stable κατανομές

- Συμβολίζουμε με **MD** το σύνολο των (μη-εκφυλισμένων) κατανομών που μπορούν να προκύψουν ως όριο κάποιου κανονικοποιημένου μέγιστου.

- Όταν $X_i \sim \text{Εκθετική}$ αποδείξαμε ότι $\frac{M_n - d_n}{c_n} \rightarrow \Lambda(x) = e^{-e^{-x}}, x \in \mathbb{R}$, άρα

$$\Lambda \in \text{MD}$$

Ερώτημα:

- Είναι η Λ (Gumbel) η μοναδική κατανομή που ανήκει στο MD, ή μπορούν να προκύψουν και άλλες κατανομές ως όρια κανονικοποιημένου μέγιστου τ.μ. ;

- Θα διερευνήσουμε από ποιές κατανομές F , το κανονικοποιημένο μέγιστο $(M_n - d_n)/c_n$ ακολουθεί και πάλι την ίδια F . Ονομάζουμε

max-stable (μέγιστο-ευσταθή)

οποιαδήποτε (μη-εκφυλισμένη) κατανομή F με την παραπάνω ιδιότητα.

- Προφανώς

max-stable $\subseteq$ MD

- Η κατανομή Λ (Gumbel) που είδαμε παραπάνω είναι *max-stable*: αν $X_i \sim \Lambda$, και επιλέξουμε $d_n = \ln n$ και $c_n = 1$,

$$P\left(\frac{M_n - d_n}{c_n} \leq x\right) = P(M_n \leq \ln n + x) = \Lambda(\ln n + x)^n = e^{-ne^{-(\ln n + x)}} = e^{-e^{-x}}.$$

Ερώτημα: Υπάρχουν κατανομές στην κλάση MD που δεν είναι *max-stable*;

Απάντηση (Θεώρημα 1.7):

max-stable = MD

Σχέδιο Απόδειξης: Αρκεί ν.δ.ο.

αν $G \in \mathbf{MD}$ τότε $G \in \mathbf{max-stable}$

- Αν $G \in \mathbf{MD}$, θα υπάρχουν ανεξ. τ.μ. $X_1, X_2, \dots \sim F$ ώστε $(M_n - d_n)/c_n \approx_d G$

- Θεωρούμε τα μέγιστα σε k το πλήθος ανεξάρτητες n -άδες των X_i :

$$M_n^{(i)} = \max \{X_{(i-1)n+1}, X_{(i-1)n+2}, \dots, X_{(i-1)n+n}\}, i = 1, 2, \dots, k.$$

- Θα είναι $Y_i = (M_n^{(i)} - d_n)/c_n \approx_d G$

- Παρατηρούμε ότι ένα κατάλληλα κανονικοποιημένο μέγιστο των $Y_1, Y_2, \dots, Y_n$,

$$\frac{\max \{Y_1, \dots, Y_k\} - \frac{d_{kn} - d_n}{c_n}}{c_{kn} / c_n} = \frac{\max \{M_n^{(1)}, \dots, M_n^{(k)}\} - d_{nk}}{c_{nk}} = \frac{M_{kn} - d_{kn}}{c_{kn}} \approx_d G$$

- Δηλαδή (αν $n \rightarrow \infty$), το κανονικοποιημένο μέγιστο των $Y_1, Y_2, \dots, Y_n$ που ακολουθούν την G , ακολουθεί και πάλι την G (διότι γράφεται και ως κανονικοποιημένο μέγιστο των X_i). Άρα η G θα είναι **max-stable**.

Κατανομές ιδίου τύπου

- Δύο X, Y (αντ. κατανομές) είναι *του ίδιου τύπου* αν $Y =_d cX + d$ για κάποια c, d .
- Παρατηρούμε ότι αν υπάρχουν d_n, c_n και G έτσι ώστε

$$\frac{M_n - d_n}{c_n} \rightarrow_d G,$$

Τότε για $\tilde{c}_n = \frac{c_n}{c}$, $\tilde{d}_n = d_n - d \frac{c_n}{c}$, για $d \in \mathbb{R}, c \neq 0$,

$$\frac{M_n - \tilde{d}_n}{\tilde{c}_n} = \frac{M_n - d_n + d \frac{c_n}{c}}{\frac{c_n}{c}} = c \frac{M_n - d_n}{c_n} + d \rightarrow_d G'$$

και η G' είναι *ιδίου τύπου* με την G .

- Επομένως, *οι ακολουθίες* c_n, d_n που εξασφαλίζουν ότι $(M_n - d_n)/c_n$ συγκλίνει σε μια οριακή κατανομή G *δεν είναι μοναδικές*.
- Μπορούμε δηλαδή να βρούμε και άλλες c_n, d_n και να καταλήξουμε σε μια άλλη οριακή κατανομή G' . *Αποδεικνύεται όμως ότι καταλήγουμε πάντοτε σε οριακές κατανομές ίδιου τύπου* ($G'(x) = G((x - d)/c)$).

Προσδιορισμός της κλάσης **MD**

- Επιθυμούμε να προσδιορίσουμε την κλάση **MD** (δηλ. ποιες κατανομές μπορούν να εμφανιστούν ως όρια κανονικοποιημένων μέγιστων)
- Αφού **MD** = **max-stable**, αρκεί να προσδιορίσουμε την κλάση των **max-stable** κατανομών
- Έχουμε ήδη βρεί ότι η Λ (Gumbel) είναι max - stable

Ερώτημα:

Ποιες κατανομές είναι max-stable; Υπάρχουν και άλλες max-stable κατανομές εκτος της Gumbel;

Το επόμενο θεώρημα δίνει την απάντηση σε αυτό το ερώτημα:

Το Θεμελιώδες Θεώρημα των *Fisher-Tippett*

Έστω $X_1, X_2, \dots$ μια ακολουθία από iid τ.μ. Αν υπάρχουν ακολουθίες $c_n > 0$, $d_n \in \mathbb{R}$ και μια μη-εκφυλισμένη κατανομή G ώστε

$$\frac{M_n - d_n}{c_n} = \frac{\max\{X_1, \dots, X_n\} - d_n}{c_n} \rightarrow_d G$$

τότε η G θα είναι του ιδίου τύπου με μια από τις τρεις ακόλουθες κατανομές:

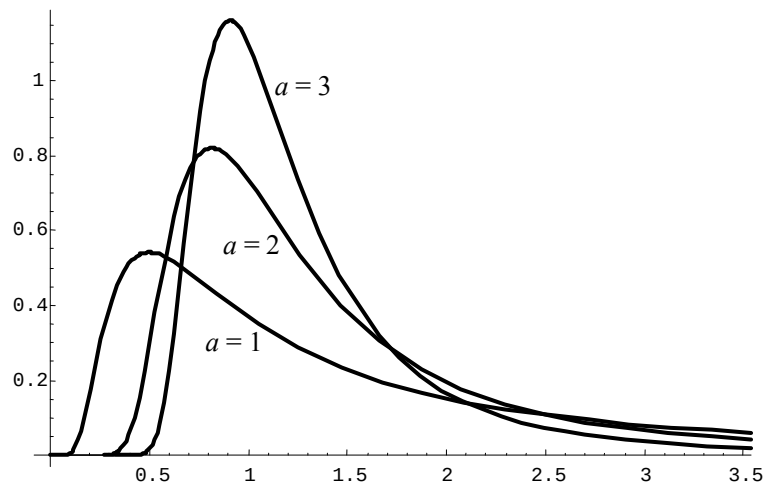
$$\text{Frechet: } \Phi_a(x) = \begin{cases} 0, & x \leq 0 \\ e^{-x^{-a}}, & x > 0 \end{cases}, \quad a > 0$$

$$\text{Weibull: } \Psi_a(x) = \begin{cases} e^{-(-x)^a}, & x \leq 0 \\ 1, & x > 0 \end{cases}, \quad a > 0$$

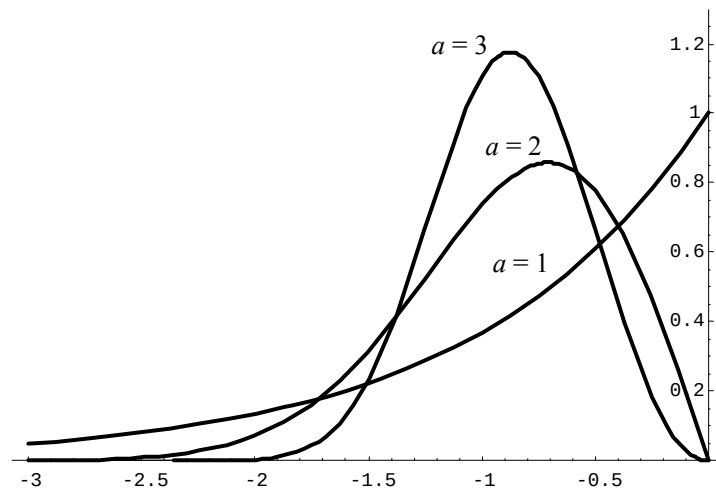
$$\text{Gumbel: } \Lambda(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}.$$

- Οι κατανομές Φ_a , Ψ_a , Λ καλούνται **τυπικές κατανομές ακροτάτων** ενώ κατανομές που είναι του ιδίου τύπου με αυτές καλούνται **κατανομές ακροτάτων**.

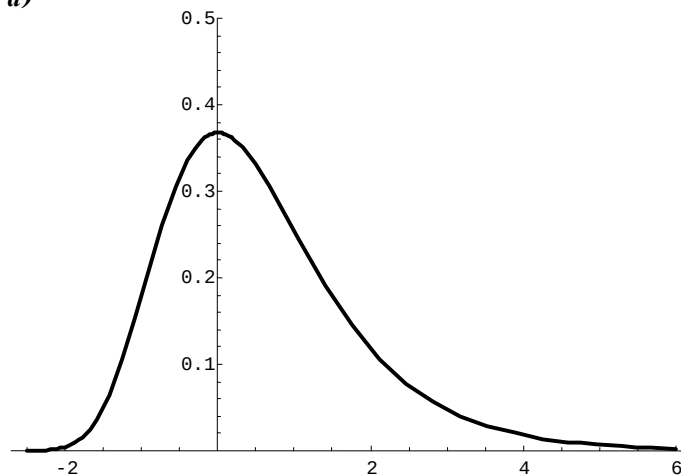
Σχέδιο Απόδειξης: Αποδεικνύεται ότι οι μόνες (μη-εκφυλισμένες) λύσεις της συναρτησιακής εξίσωσης που προκύπτει από τον ορισμό των max-stable κατανομών, είναι οι τρεις παραπάνω.



Frechet (Φ_a)



Weibull (Ψ_a)



Gumbel (Λ)

Οι κατανομές Φ_a και Λ έχουν $x_F = \infty$, ενώ η Ψ_a έχει $x_F = 0$.

Παρατηρήσεις

(α) Από το Θ. F-T, η ασυμπτωτική κατανομή, G , του $(M_n - d_n)/c_n$ των iid $X_1, \dots, X_n \sim F$ για οποιαδήποτε σ.κ. F , θα είναι (αν υπάρχει) ίση με

$$\Lambda((x - d)/c) \quad \text{ή} \quad \Phi_\alpha((x - d)/c) \quad \text{ή} \quad \Psi_\alpha((x - d)/c)$$

για κάποιες σταθερές d, c .

(β) Οι $\Lambda, \Phi_\alpha, \Psi_\alpha$ συνδέονται ως εξής:

$$X \sim \Lambda \Leftrightarrow e^{X/a} \sim \Phi_a \Leftrightarrow -e^{-X/a} \sim \Psi_a$$

(γ) Η κατανομή Ψ_α αναφέρεται παραπάνω ως κατανομή Weibull. Ουσιαστικά όμως πρόκειται για μια αντεστραμμένη (reversed) κατανομή Weibull.

(δ) Οι κατανομές *Frechet*, *Weibull* και *Gumbel* είναι (οι μόνες) max-stable κατανομές. Αποδεικνύεται εύκολα ότι

$$\text{- αν } X_1, X_2, \dots \sim \Phi_\alpha \text{ τότε } M_n/n^{1/a} \sim \Phi_\alpha,$$

$$\text{- αν } X_1, X_2, \dots \sim \Psi_\alpha \text{ τότε } M_n/n^{-1/a} \sim \Psi_\alpha$$

$$\text{- αν } X_1, X_2, \dots \sim \Lambda \text{ τότε } M_n - \ln n \sim \Lambda$$

Ροπογεννήτρια της Gumbel και Ροπές των κατανομών ακροτάτων

- Αν $X \sim \Lambda$, αποδεικνύεται ότι $E(e^{tX}) = \Gamma(1-t)$, $t < 1$ (Γ : συνάρτηση *Γάμμα*).
- Παραγωγίζοντας την παραπάνω ροπογεννήτρια ως προς t λαμβάνουμε:

$$E(X) = -\Gamma'(1) = \gamma$$

όπου $\gamma = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \ln n \right) \approx 0.5772156\dots$ είναι η *σταθερά Euler*.

- Επίσης, αποδεικνύεται ότι $V(X) = \pi^2 / 6$.
- Αν $X \sim \Lambda$, τότε η $Y = e^{X/a} \sim \Phi_a$ και $W = -e^{-X/a} \sim \Psi_a$ και άρα

$$E(Y) = E(e^{X/a}) = \Gamma(1-1/a), \quad a > 1, \quad E(W) = E(-e^{-X/a}) = -\Gamma(1+1/a), \quad a > 0.$$

ενώ αν $a \leq 1$, η μέση τιμή της Frechet θα είναι $E(Y) = \infty$.

Παράδειγμα με οριακή κατανομή κανον. μεγίστου την κατανομή Frechet

- Έστω iid $X_1, X_2, \dots \sim \text{Cauchy}$ ($x_F = \infty$) με σ.π.π.

$$f(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}$$

- Αν πάρουμε $d_n = 0$, $c_n = n/\pi$, τότε

$$P\left(\frac{M_n - d_n}{c_n} \leq x\right) = P\left(\frac{\pi M_n}{n} \leq x\right) = P\left(M_n \leq \frac{nx}{\pi}\right) = \left(1 - \bar{F}\left(\frac{nx}{\pi}\right)\right)^n, \quad x \in \mathbb{R},$$

και επειδή,

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{(\pi x)^{-1}} = \lim_{x \rightarrow \infty} \frac{f(x)}{\pi^{-1} x^{-2}} = \lim_{x \rightarrow \infty} \frac{x^2}{1+x^2} = 1,$$

προκύπτει τελικά ότι,

$$\lim_{n \rightarrow \infty} \left(1 - \bar{F}\left(\frac{nx}{\pi}\right)\right)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{\bar{F}\left(\frac{nx}{\pi}\right)}{nx(\pi \frac{nx}{\pi})^{-1}}\right)^n = \exp\left(-\lim_{n \rightarrow \infty} \left(\frac{\bar{F}\left(\frac{nx}{\pi}\right)}{x(\pi \frac{nx}{\pi})^{-1}}\right)\right) = e^{-\frac{1}{x}},$$

και επομένως εδώ η οριακή κατανομή είναι η Φ_1 .

Παράδειγμα με οριακή κατανομή κανον. μεγίστου την κατανομή Weibull

- Έστω iid $X_1, X_2, \dots \sim \text{Ομοιόμορφη στο } [0,1] \text{ (} x_F = 1 \text{)}$
- Επιλέγοντας $c_n = 1/n$ και $d_n = 1$, η κατανομή του κανονικοποιημένου μεγίστου θα είναι

$$\begin{aligned} P\left(\frac{M_n - d_n}{c_n} \leq x\right) &= P(nM_n - n \leq x) = P\left(M_n \leq 1 + \frac{x}{n}\right), \\ &= \left(F\left(1 + \frac{x}{n}\right)\right)^n = \left(1 + \frac{x}{n}\right)^n \rightarrow e^x, \quad x < 0, \end{aligned}$$

η οποία είναι η Ψ_1 .

Περιοχές έλξης μιας κατανομής ακροτάτων

- Είδαμε ότι η *οριακή κατανομή* ενός κανονικοποιημένου μεγίστου $(M_n - d_n)/c_n$ των X_i έχει τον ίδιο τύπο με μια από τις Λ , Φ_a , Ψ_a .
- **Επομένως:** Σε κάθε κατανομή F των X_i , θα *αντιστοιχεί μια μοναδική οριακή κατανομή*, μία από τις Λ , Φ_a , Ψ_a (αρκεί το $(M_n - d_n)/c_n$ να συγκλίνει),

Ερώτημα:

Σε ποιες κατανομές F αντιστοιχεί η Λ , σε ποιες η Φ_a και σε ποιες η Ψ_a ;

- **Ορισμός** Μια τ.μ. X (ή αντίστοιχα η κατανομή της, F) ανήκει «**στην περιοχή έλξης μιας κατανομής ακροτάτων G** » αν υπάρχουν ακολουθίες $c_n > 0$, $d_n \in \mathbb{R}$ έτσι ώστε η $(M_n - d_n)/c_n \rightarrow_d G$. (συμβολικά, X ή $F \in \text{MDA}(G)$).
- Είδαμε ότι:
 $\text{Εκθετική} \in \text{MDA}(\Lambda)$,
 $\text{Cauchy} \in \text{MDA}(\Phi_a)$,
 $\text{Ομοιόμορφη} \in \text{MDA}(\Psi_a)$.

Συναρτήσεις ομαλής κύμανσης

- **Ορισμός** Μια θετική συνάρτηση h στο $(0, \infty)$ καλείται **ομαλής κύμανσης** (*regularly varying*) στο ∞ με δείκτη a (συμβ. $h \in \mathcal{R}_a$) αν

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^a, \quad t \geq 0.$$

- Αν $a = 0$ τότε καλείται **βραδείας κύμανσης** (*slowly varying*).
- Αν $a = -\infty$ ($t^a = \infty$ αν $t \in (0, 1)$ ή 0 αν $t \in (1, \infty)$) καλείται **ταχείας κύμανσης** (*rapidly varying*).
- Κάθε συνάρτηση $h \in \mathcal{R}_a$, $a \in \mathbb{R}$, μπορεί να γραφεί ως $h(x) = x^a L(x)$, όπου $L \in \mathcal{R}_0$.

- **Παραδείγματα:**

- Συναρτήσεις **βραδείας κύμανσης**: οι σταθερές συναρτήσεις, οι συναρτήσεις που συγκλίνουν σε μια σταθερά (στο ∞), και συναρτήσεις που μεταβάλλονται «πολύ αργά» στο ∞ όπως η λογαριθμική συνάρτηση, κ.ο.κ.
- Συναρτήσεις **ομαλής κύμανσης** είναι π.χ.: οι συναρτήσεις x^a , $x^a \ln(1+x)$ κ.ο.κ.
- Συναρτήσεις **ταχείας κύμανσης** είναι π.χ. οι εκθετικές συναρτήσεις e^{-x} .

Κατανομές ισοδύναμης ουράς

- Δύο σ.κ. F, G μπορούν να έχουν διαφορετική μορφή στο R αλλά η δεξιά ουρά τους να έχει την ίδια συμπεριφορά. Θα καλούμε δύο τέτοιες κατανομές «ισοδύναμης ουράς» (tail-equivalent).

Ορισμός Δύο σ.κ. F, G έχουν ισοδύναμη ουρά αν $x_F = x_G$ και, για κάποιο $c > 0$,

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x)}{\overline{G}(x)} = c ,$$

- Αν δύο σ.κ. έχουν ισοδύναμη ουρά τότε ανήκουν στην ίδια περιοχή έλξης μιας κατανομής ακροτάτων (με τις ίδιες ακολουθίες c_n, d_n).

Μορφή των ακολουθιών c_n, d_n που κανονικοποιούν το M_n

- Αν $X_i \sim F$ (συνεχείς) τότε οι τ.μ. $Y_i = F(X_i)$, $\sim$ Ομοιόμορφη κατανομή το $[0,1]$

Επομένως:

$$\max \{Y_i\} = \max \{F(X_i)\} = F(\max \{X_i\}) = F(M_n) \approx 1 - n^{-1}$$

(διότι η $\max \{Y_i\}$ από n ομοιόμορφες τ.μ. έχει μέση τιμή $\approx 1 - n^{-1}$).

- Άρα, γενικά, $M_n \approx F^{-1}(1 - n^{-1})$ και είναι λογικό να εξετάσουμε αν μπορούμε να επιλέξουμε

$$c_n = F^{-1}(1 - n^{-1}) \quad \text{ή} \quad d_n = F^{-1}(1 - n^{-1}).$$

- Για να αντιμετωπίσουμε και την περίπτωση που η F δεν αντιστρέφεται ορίζουμε την **γενικευμένη αντίστροφη** μιας σ.κ. F :

$$F^{\leftarrow}(t) = \inf F^{-1}([t,1]) = \inf \{x \in R : F(x) \in [t,1]\}, \quad 0 < t < 1.$$

(αν η F είναι αντιστρέψιμη τότε $F^{\leftarrow} = F^{-1}$)

Η περιοχή έλξης της Φ_a (Frechet)

- Για την $\overline{\Phi}_a(x) = 1 - e^{-x^{-a}}$, $a > 0$, ισχύει ότι

$$\lim_{x \rightarrow \infty} \frac{\overline{\Phi}_a(tx)}{\overline{\Phi}_a(x)} = \lim_{x \rightarrow \infty} \frac{1 - e^{-(tx)^{-a}}}{1 - e^{-x^{-a}}} = \lim_{x \rightarrow \infty} \frac{(-a(tx)^{-a-1}t)e^{-(tx)^{-a}}}{(-ax^{-a-1})e^{-x^{-a}}} = \lim_{x \rightarrow \infty} \frac{t^{-a}e^{-(tx)^{-a}}}{e^{-x^{-a}}} = t^{-a},$$

και επομένως, $\overline{\Phi}_a \in \mathcal{R}_{-a}$ (ομαλής κύμανσης με δείκτη $-a$).

Ερώτηση: Οι κατανομές με την «ίδια» ουρά (δηλ. ομαλής κύμανσης με δείκτη $-a$) ανήκουν όλες στην περιοχή έλξης της Φ_a ;

Απάντηση: ΝΑΙ

- Μια σ.κ. $F \in \text{MDA}(\Phi_a)$ αν και μόνο αν $\overline{F} \in \mathcal{R}_{-a}$, $a > 0$.

- Αν $F \in \text{MDA}(\Phi_a)$ τότε $x_F = \infty$ και

$$M_n / c_n \rightarrow_d \Phi_a,$$

όπου $c_n = F^{\leftarrow}(1 - n^{-1})$ και $d_n = 0$.

- **Παραδείγματα:** Pareto, Cauchy, Loggamma $\in \text{MDA}(\Phi_a)$

Η περιοχή έλξης της Ψ_a (Weibull)

- Παρατηρούμε ότι $\overline{\Phi}_a(x) = \overline{\Psi}_a(0 - x^{-1})$ και επειδή $\overline{\Phi}_a(x) \in \mathfrak{R}_{-a}$ έπεται ότι και $\overline{\Psi}_a(0 - x^{-1}) = \overline{\Psi}_a(x_F - x^{-1}) \in \mathfrak{R}_{-a}$ (η Ψ_a έχει $x_F = 0$).

Ερώτηση: οι σ.κ. F με $\overline{F}(x_F - x^{-1}) \in \mathfrak{R}_{-a}$ θα ανήκουν πάντοτε στην $MDA(\Psi_a)$;

Απάντηση: ΝΑΙ

- Μια σ.κ. $F \in MDA(\Psi_a)$ αν και μόνο αν $x_F < \infty$ και $\overline{F}(x_F - x^{-1}) \in \mathfrak{R}_{-a}$, $a > 0$.
- Αν $F \in MDA(\Psi_a)$ τότε

$$\frac{M_n - d_n}{c_n} \rightarrow_d \Psi_a,$$

όπου $c_n = x_F - F^{\leftarrow}(1 - n^{-1})$, $d_n = x_F$.

- **Παραδείγματα:** Ομοιόμορφη, Βήτα $\in MDA(\Psi_a)$

Η περιοχή έλξης της Λ (Gumbel)

- Ισχύει ότι $\lim_{x \rightarrow \infty} \bar{\Lambda}(x) / e^{-x} = 1$ και άρα, $\bar{\Lambda}(x) = c(x)e^{-x}$, όπου $c(x) \rightarrow 1$ για $x \rightarrow \infty$. Δηλαδή, $\bar{\Lambda} \in \mathfrak{R}_{-\infty}$ (ταχείας κύμανσης).

Ερώτηση: Η κλάση $MDA(\Lambda)$ περιλαμβάνει μόνο κατανομές με $\bar{F} \in \mathfrak{R}_{-\infty}$;

Απάντηση: ΝΑΙ, αρκεί $x_F = \infty$. Αλλά υπάρχουν και F στην $MDA(\Lambda)$ με $x_F < \infty$.

- Γενικά αποδ. ότι: μια σ.κ. $F \in MDA(\Lambda)$ αν και μόνο αν μπορεί να αναπαρασταθεί

$$\bar{F}(x) = c(x)e^{-\int_z^x \frac{g(t)}{a(t)} dt}, \quad z < x < x_F$$

για κάποιο $z < x_F$, όπου c, g είναι συναρτήσεις με $c(x) \rightarrow c_0 > 0$, $g(x) \rightarrow 1$ όταν $x \uparrow x_F$, και $\lim_{x \uparrow x_F} a'(x) = 0$.

- Επίσης, αν μια κατανομή $F \in MDA(\Lambda)$ έχοντας την παραπάνω αναπαράσταση, τότε $(M_n - d_n) / c_n \rightarrow_d \Lambda$ όπου $d_n = F^{\leftarrow}(1 - n^{-1})$, $c_n = a(d_n)$.
- **Παραδείγματα:** Εκθετική, Weibull, Κανονική, Λογαριθμοκανονική κ.α.

Συνοψίζοντας, παρατίθεται ο ακόλουθος πίνακας:

Weibull (Reversed Weibull)	Gumbel	Frechet
$\Psi_a(x) = e^{-(-x)^a}, x \leq 0, a > 0$	$\Lambda(x) = e^{-e^{-x}}, x \in R$	$\Phi_a(x) = e^{-x^{-a}}, x > 0, a > 0$
MDA: $x_F < \infty,$ $\bar{F}(x_F - x^{-1}) = x^{-a} L(x) \in \mathfrak{R}_{-a}$ $(L \in \mathfrak{R}_0)$	MDA: $x_F \leq \infty,$ $\bar{F}(x) = c(x) e^{-\int_z^x \frac{g(t)}{a(t)} dt}, z < x < x_F$ $c(x) \rightarrow c_0 > 0, g(x) \rightarrow 1, a'(x) \rightarrow 0, x \uparrow x_F$	MDA: $x_F = \infty,$ $\bar{F}(x) = x^{-a} L(x) \in \mathfrak{R}_{-a}$ $(L \in \mathfrak{R}_0)$
$c_n = x_F - F^{\leftarrow}(1 - n^{-1}), d_n = x_F$	$c_n = a(d_n), d_n = F^{\leftarrow}(1 - n^{-1})$	$c_n = F^{\leftarrow}(1 - n^{-1}), d_n = 0$
Uniform, Beta,...	Exponential, Weibull (not reversed), Gamma, Normal, Lognormal	Cauchy, Pareto, Loggamma, Burr,...