



**"תוכנית רוטשילד-ויצמן למצוינות בהוראת המדעים"
במימונה של קרן קיסריה אדמונד בנימין דה רוטשילד**

Extreme Value Theory

התפלגות הערך המקסימאלי של

הרבה משתנים מקריים

מגישה: מיטל ישי

מנחה: דר' בועז נדלר

מועד הגשה: 24/4/12

תוכן עניינים:

מבוא.....	2
פרק 1 - דוגמאות של התפלגויות שונות.....	4
דוגמא 1 - התפלגות אקספוננציאלית.....	4
דוגמא 2 - התפלגות אקספוננציאלית כללית.....	7
דוגמא 3 - התפלגות דמוית גאוסית (נורמאלית).....	8
דוגמא 4 – התפלגות אחידה.....	10
דוגמא 5 - התפלגות עם דעיכה POWER-LOW.....	11
פרק 2 - משפטים מרכזיים בהתפלגות ערך המקסימום.....	13
א. ניסוח אחר בשאיפה לגבול.....	13
ב. התפלגות ערך הקיצון.....	15
פרק 3 - מהם מקרי קיצון? עד כמה זה חשוב לנו? וכיצד מנתחים אותם?.....	20
פרק 4 – סיכום.....	23
ביבילוגרפיה.....	23

מבוא

אחד המשפטים הידועים והחשובים ביותר בהסתברות וסטטיסטיקה הוא משפט הגבול המרכזי.

להלן אחד הניסוחים שלו.

משפט הגבול המרכזי: התפלגות גאוסית (נורמלית).

יהיו X_1, X_2, \dots, X_n משתנים מקריים, בלתי תלויים ושווי התפלגות, בעלי תוחלת μ וסטיית תקן σ סופיים.

נסמן את הממוצע שלהם ב- \bar{X}_n .

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

משפט הגבול המרכזי טוען כי כאשר $n \rightarrow \infty$, ההתפלגות של \bar{X}_n שואפת להתפלגות נורמלית עם

תוחלת μ וסטיית תקן $\frac{\sigma}{\sqrt{n}}$.

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

תזכורת: עבור משתנה מקרי Z עם התפלגות נורמלית $Z \sim N(0,1)$, מתקיים כי

$$\Pr(Z < z) = \Phi(z)$$

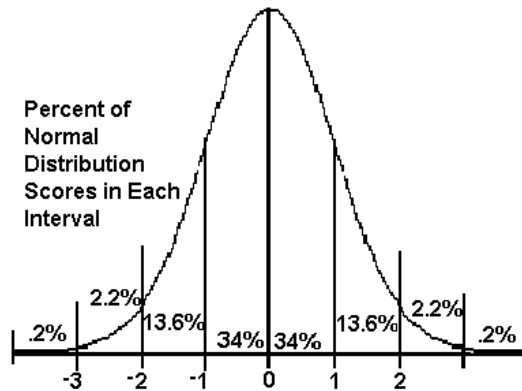
$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

כאשר

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

גרף 1 מתאר את הצפיפות של משתנה מקרי $Z \sim N(0,1)$.

הערה: נשים לב כי לאור משפט הגבול המרכזי, ההתפלגות הנורמלית היא ההתפלגות הגבולית היחידה, עד כדי מרכז ונרמול (centering & scaling), ללא תלות בהתפלגותם של המשתנים המקרים X_i .



גרף 1: הצפיפות של משתנה מקרי $Z \sim N(0,1)$ הידועה גם בשם פעמון גאוס.

בעבודה זו נדון במשפט גבול אחר, ידוע הרבה פחות, אך בעל חשיבות עצומה גם כן.

אנו נעסוק במשפטי גבול הקשורים להתפלגות של ערכי קיצון.

הגדרה מתמטית: ערך מקסימאלי.

תהי X_1, X_2, \dots, X_n סדרת משתנים מקריים בלתי תלויים ושווי התפלגות.

נגדיר משתנה מקרי חדש Y_n כערך המקסימאלי שלהם:

$$Y_n = \max(X_1, X_2, \dots, X_n)$$

בעבודה זו נחקור את ההתפלגות הגבולית של Y_n כאשר $n \rightarrow \infty$. באופן דומה ניתן גם לחקור את

התפלגות הערך המינימאלי $\min(X_1, X_2, \dots, X_n)$ ומשפטי גבול דומים יתקיימו עבורו.

שאלת המחקר

בעבודה זו אשאל מספר שאלות מתמטיות.

1. האם קיים משפט גבול עבור התפלגות ערך הקיצון?

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

2. נניח כי אכן קיימת התפלגות גבולית לערך קיצון. האם, בדומה למשפט הגבול המרכזי עבור הממוצע, התפלגות זו היא יחידה?

הנושא של ערכי קיצון הוא בעל חשיבות רבה בחיים האמיתיים. מעבר לסקרנות המתמטית נוכל לראות כי המשפטים והכלים הנלווים להם מסייעים רבות בהערכות של מקרי קיצון בתחומים רבים. על כך נדון בפרק 3.

במהלך העבודה נעזר בתכונות הבאות של התוחלת והשונות.

יהי X משתנה מקרי. אזי לכל $\alpha, \beta \in R$ מתקיים כי

$$E(\alpha X + \beta) = \alpha E(X) + \beta$$

$$Var(\alpha x + \beta) = Var(\alpha x) = \alpha^2 Var(x)$$

פרק 1 - דוגמאות של התפלגויות שונות

לפני שנתאר את התיאוריה הכללית, ננתח מספר דוגמאות פשוטות.

דוגמא 1 - התפלגות אקספוננציאלית.

תהי X_1, X_2, \dots, X_n סדרת משתנים מקריים בלתי תלויים, בעלי התפלגות אקפוננציאלית עם פרמטר

$$\Pr[X \geq t] = e^{-t} \quad \text{כלומר, } \lambda = 1$$

נגדיר משתנה מקרי חדש Y_n שהוא הערך המקסימאלי שלהם:

$$Y_n = \max_{1 \leq i \leq n} X_i$$

המטרה היא לחשב מהי ההתפלגות של Y_n , או לפחות לקרב אותה עבור ערכי n גדולים.

מכיוון שהמשתנים המקריים X_i הינם בלתי תלויים ושווי התפלגות מתקיימת המשוואה הבאה:

$$\Pr[Y_n \leq t] = \Pr[X_1 \leq t] \cdot \Pr[X_2 \leq t] \cdot \dots \cdot \Pr[X_n \leq t] = (\Pr[X \leq t])^n = (1 - \Pr[X > t])^n$$

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת. כלשהי של קובץ זה או חלק ממנו.

נציב את הביטוי עבור $\Pr[X > t]$ ונקבל

$$(1.1) \quad \Pr[Y_n \leq t] = (1 - e^{-t})^n$$

המשוואה דלעיל היא ההתפלגות המדויקת של Y_n .

אנו מעוניינים להבין את ההתנהגות שלה עבור n גדול, בפרט מהי התוחלת $E[Y_n]$ ומהי השונות

$$.Var[Y_n]$$

נזכור כי לכל a קבוע, $\lim_{n \rightarrow \infty} \left(1 - \frac{a}{n}\right)^n = e^{-a}$. על כן נבצע החלפת משתנים

$$e^{-t} = \frac{c(t)}{n}, \quad -t = \log c(t) - \log n$$

נציב במשוואה (1.1) ונקבל

$$\Pr[Y_n \leq t] = \left(1 - \frac{c(t)}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-c(t)}$$

$$\Pr[Y_n \leq \log n - \log c] \rightarrow e^{-c} \quad \text{כלומר}$$

נסמן $c = e^{-x}$, $x = -\log c$ ונקבל

$$(1.2) \quad \Pr[Y_n - \log n \leq x] \rightarrow e^{-e^{-x}}$$

לסיכום, קיבלנו שבגבול $n \rightarrow \infty$ המשתנה המקרי $Z_n = Y_n - \log n$ שואף להתפלגות גבולית שאינה תלויה ב- n .

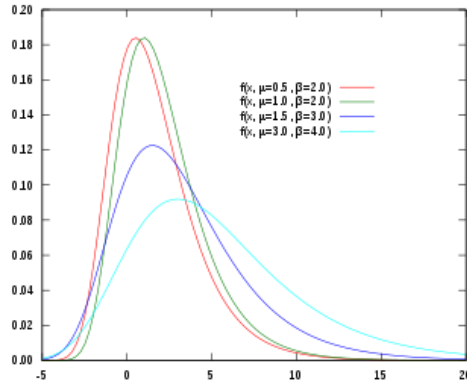
התפלגות גבולית זו ידועה בשם התפלגות Gumbel סטנדרטית. מעתה ואילך תסומן ב- $G(0,1)$.

למעשה קיימת משפחה שלמה של התפלגויות מסוג Gumbel, $G(\mu, \beta)$, המוגדרת באופן הבא:

אם $Z \sim G(\mu, \beta)$, אזי

$$(1.3) \quad \Pr[Z \leq x] = e^{-e^{-(x-\mu)/\beta}}$$

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.



גרף 2: צפיפות משתנה מקרי G עבור ערכי μ ו- β שונים.

גרף 2 מתאר התפלגויות שונות מסוג Gumbel עבור ערכי μ ו- β שונים.

❖ **Emil Julius Gumbel** (1891-1966) היה מתמטיקאי גרמני וסופר פוליטי. Gumbel היה פרופ'

לסטטיסטיקה מתמטית באוניברסיטת היילדברג. יחד עם Leonard Tippett ו-Ronald Fisher

הוא נחשב לחלוץ מתמטי בתחום ערך קיצון וההתפלגות $e^{-e^{-(x-\mu)/\beta}}$ נקראת על שמו.

טענה

יהי Z משתנה מקרי עם התפלגות Gumbel סטנדרטית.

אזי התוחלת של Z היא $E[Z] = \gamma$, כאשר γ הוא קבוע אוילר מסנצרוני $\gamma \approx 0.577$,

$$Var[Z] = \frac{\pi^2}{6} \quad \gamma = -\int_0^{\infty} \ln x e^{-x} dx \quad \text{והשונות היא}$$

הוכחה

לשם פשטות נוכיח רק עבור התוחלת.

לפי ההגדרה, הצפיפות של Z היא $g(z) = e^{-z} e^{-e^{-z}}$

$$E[Z] = \int_{-\infty}^{\infty} z g(x) dz \quad \text{התוחלת מוגדרת ע"י:}$$

נבצע החלפת משתנים $z = -\ln u$, $du = -e^{-z} dz$, $u = e^{-z}$

$$E[Z] = -\int_0^{\infty} \ln u e^{-u} du = \gamma \quad \text{גבולות האינטגרל ישתנו בהתאם}$$

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

הערה: עבור משתנה מקרי $Z \sim G(\mu, \beta)$, מתקיים באופן דומה כי

$$(1.6) \quad E[Z] = \mu + \beta\gamma$$

$$(1.7) \quad \text{Var}[Z] = \frac{\pi^2}{6} \beta^2$$

נחזור למשתנה שלנו $Y_n = \max X_i$.

$$(1.8) \quad Z_n = Y_n - \log n \xrightarrow{d} G(0,1) \quad \text{מצאנו כי}$$

לפי תכונת הליניאריות של התוחלת מתקיים כי עבור n גדול

$$(1.9) \quad E(Y_n) \approx \log n + \gamma$$

$$(1.10) \quad \text{Var}(Y_n) \approx \text{Var}(G(0,1)) = \frac{\pi^2}{6}, \quad \text{וכן,}$$

דוגמא 2 - התפלגות אקספוננציאלית כללית

יהיו X_i משתנים מקריים בעלי התפלגות אקספוננציאלית, $X_i \sim \exp(\lambda)$, $\left(\Pr[X \geq t] = e^{-\frac{t}{\lambda}} \right)$.

$$W_i = \frac{X_i}{\lambda} \sim \exp(1)$$

בדומה לדוגמא הקודמת Y_n יהיה הערך המקסימאלי של X_i . נסמן ב- \tilde{Y}_n את הערך המקסימאלי של המשתנים המקריים W_i .

$$\tilde{Y}_n = \max_{1 \leq i \leq n} W_i = \max_{1 \leq i \leq n} \frac{X_i}{\lambda}$$

בדוגמא הקודמת קיבלנו שהערך המקסימאלי של משתנים מקריים המתפלגים אקספוננציאלית עם פרמטר $\lambda = 1$ מקיים את המשוואה (1.8).

$$\tilde{Y}_n - \log n \rightarrow G(0,1) \quad \text{באופן דומה, כאשר } n \rightarrow \infty$$

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

$$\frac{Y_n}{\lambda} - \log n \rightarrow G(0,1) \quad \text{ולכן}$$

$$\Pr\left[\frac{Y_n}{\lambda} - \log n \leq \frac{x}{\lambda}\right] \rightarrow e^{-e^{-\frac{x}{\lambda}}} \quad \text{כלומר}$$

$$\Pr[Y_n - \lambda \log n \leq x] \rightarrow e^{-e^{-\frac{x}{\lambda}}} \quad \text{או}$$

ההתפלגות הגבולית שהתקבלה היא גם מסוג Gumbel, אבל עם פרמטר λ , $G(0, \lambda)$, כלומר

$$Y_n \approx \lambda \log n + Z$$

כאשר $Z \sim G(0, \lambda)$.

עזר ב- (1.6) ו- (1.7) ובתכונות התוחלת והשונות ונקבל

$$(1.11) \quad E(Y_n) \approx \lambda(\log n + \gamma)$$

$$(1.12) \quad \text{Var}(Y_n) \approx V(\lambda Z_i) = \lambda^2 \frac{\pi^2}{6}$$

בשתי הדוגמאות שהובאו קיבלנו כי התוחלת של הערך המקסימאלי תלויה ב- n ואילו סטיית התקן, בשני המקרים, אינה תלויה ב- n .

לשם השוואה, במשפט הגבול המרכזי עבור הממוצע, סטיית התקן $\rightarrow 0$ $\frac{\sigma}{\sqrt{n}} \xrightarrow{n \rightarrow \infty}$.

דוגמא 3 - התפלגות דמוית גאוסית (נורמאלית)

כעת נסתכל על משתנים מקריים X_i בלתי תלויים ואי שליליים המתפלגים באופן הבא:

$$t \in [0, \infty), \quad \Pr[X \geq t] = e^{-t^2}$$

התפלגות מעין זו נקראת double exponential, אשר דועכת הרבה יותר מהר מההתפלגות single exponential כפי שניתחנו בדוגמאות הקודמות.

כמו קודם נסמן ב- Y_n את הערך המקסימאלי שלהם.

כדי להבין מהי התוחלת והשונות של Y_n במקרה זה, נבצע החלפת משתנים $X = \sqrt{W}$.

נציב במשוואה ונקבל

$$\Pr[W \geq t^2] = e^{-t^2}$$

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת. כלשהי של קובץ זה או חלק ממנו.

כלומר עבור $t^2 = s$

$$\Pr[W \geq s] = e^{-s}$$

לכן

$$Y_n = \max_{1 \leq i \leq n} X_i = \max_{1 \leq i \leq n} \sqrt{W_i} = \sqrt{\max_{1 \leq i \leq n} W_i}$$

נשים לב כי $W_i = X_i^2$ הם משתנים מקריים עם זנב אקספוננציאלי כמו בדוגמא 1 שכבר ניתחנו דלעיל. על כן נשתמש בתוצאה (1.8) ונקבל

$$Y_n \approx \sqrt{\log n + Z} = \sqrt{\log n} \cdot \sqrt{1 + \frac{Z}{\log n}}$$

עבור $Z \sim G(0,1)$

$$\sqrt{1 + \varepsilon} = 1 + \frac{1}{2} \varepsilon + O(\varepsilon^2) \dots \quad \text{נעזר בפיתוח לטור טיילור}$$

$$Y_n \approx \sqrt{\log n} \left(1 + \frac{1}{2} \cdot \frac{Z}{\log n} + O\left(\frac{Z^2}{(\log n)^2}\right) \right) = \sqrt{\log n} + \frac{Z}{2\sqrt{\log n}} + O_p\left(\frac{1}{(\log n)^{3/2}}\right) + \dots$$

עבור $n \gg 1$ ניתן להזניח את האיבר השלישי דלעיל ונקבל כי

$$(1.13) \quad Y_n \approx \sqrt{\log n} + \frac{Z}{2\sqrt{\log n}}$$

$$(1.14) \quad E(Y_n) \approx \sqrt{\log n} + \frac{\gamma}{2\sqrt{\log n}} \quad \text{לכן}$$

$$(1.15) \quad \text{Var}(Y_n) \approx \frac{1}{4 \log n} \cdot \frac{\pi^2}{6}$$

מעניין להשוות את (1.14) ו-(1.15) עם דוגמא 1, שבה הדעיכה היא אקספוננציאלית בודדת. שם קיבלנו

$$\text{כי } E(Y_n) \approx \log n + \gamma \text{ ואילו כאן } E(Y_n) \approx \sqrt{\log n} + \frac{\gamma}{2\sqrt{\log n}} \text{ כלומר, בדוגמא 1 המקסימום גדל}$$

משמעותית מהר יותר מאשר בדוגמא 3. הסיבה לכך היא שדעיכה אקספוננציאלית בודדת היא הרבה יותר איטית מדעיכה אקספוננציאלית כפולה. יתר על כן, קיים הבדל מהותי בשונות. בעוד שהשונות

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

בדוגמא 3, $Var(Y_n) \approx \frac{1}{4 \log n} \cdot \frac{\pi^2}{6}$, שואפת לאפס כאשר $n \rightarrow \infty$, השונות בדוגמא 1,

$$Var(Y_n) = \frac{\pi^2}{6}, \text{ היא קבועה ואינה תלויה ב-}n.$$

הבדל זה הוא משמעותי ביישום של התיאוריה של ערכי הקיצון עבור מצבים מהחיים, על כך נדון בהמשך.

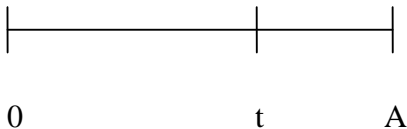
דוגמא 4 – התפלגות אחידה

עד כה ניתחנו דוגמאות בהן המשתנים המקריים X_i יכלו לקבל כל ערך ממשי. כעת נתבונן במקרים בהם המשתנים המקריים X_i חסומים, זאת אומרת בעלי תומך קומפקטי (compact support). בעולם האמיתי יש לא מעט משתנים מקריים חסומים. למשל: גובה של בני אדם, ציוני מבחנים, טמפרטורה וגדלים פיסיקליים נוספים. בניגוד למקרים הקודמים, בהם $Y_n \rightarrow \infty$, ברור לנו שבמקרה זה הערך המקסימאלי יהיה תמיד קטן שווה מהחסם העליון.

ברור כי Y_n , במקרים בעלי תומך קומפקטי, לא יכול לשאוף להתפלגות Gumbel שיש לה תומך אינסופי. נשאלת השאלה, האם קיימת התפלגות גבולית אחרת?

לשם פשטות נתבונן במקרה הפרטי הבא.

יהיו X_i משתנים מקריים בעלי התפלגות אחידה $X_i \sim [0, A]$, $\Pr(X = x) = \frac{1}{A}$,



כמו קודם נסמן ב- Y_n את ערך המקסימום של המשתנים המקריים X_i . לפי ההגדרה,

$$\Pr[Y_n \leq t] = \Pr[X \leq t]^n = \left(\frac{t}{A}\right)^n$$

נבצע החלפת משתנים $s = A - t$

$$(1.16) \quad \Pr[Y_n \leq A - s] = \left(1 - \frac{s}{A}\right)^n$$

נגדיר $\frac{s}{A} = \frac{w}{n}$ נציב במשוואה (1.16) ונקבל

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

$$\Pr\left[Y_n \leq A - \frac{Aw}{n}\right] = \left(1 - \frac{w}{n}\right)^n \rightarrow e^{-w}$$

(1.17) או במילים אחרות $\Pr\left[\frac{A - Y_n}{(A/n)} \geq w\right] \rightarrow e^{-w}$

כלומר המשתנה המקרי $\frac{A - Y_n}{(A/n)}$ שואף להתפלגות אקספוננציאלית סטנדרטית, שאינה תלויה ב- n .

דוגמא זו מראה לנו כי במשפטי קיצון, בניגוד למשפט הגבול המרכזי, ההתפלגות הגבולית אינה יחידה.

קיבלנו מקרה פרטי של אחת מהתפלגויות הגבול הנקראת על שם Weibull.

עבור $\alpha > 0$

$$H(x) = \begin{cases} \exp(-(-X)^\alpha) & x < 0 \\ 1 & x \geq 0 \end{cases}$$

❖ **Ernst Hjalmar Waloddi Weibull** (1887-1979), היה מהנדס שוודי, מדען ומתמטיקאי.

Weibull היה פרופ' באוניברסיטת אופסלה. פרסם מאמרים רבים בנושא מוצקים, כוחו של חומר, עייפות וכמובן התפלגות Weibull.

דוגמא 5 - התפלגות עם דעיכה power-law

יהיו X_i משתנים מקריים בעלי התפלגות עם דעיכה איטית מסוג power-law הבאה:

$$\Pr[X > t] = t^{-\alpha}$$

נגדיר ב- Y_n את הערך המקסימאלי של X_i .

(1.18) כמו בדוגמאות הקודמות, $\Pr[Y_n \leq t] = (1 - \Pr[X > t])^n = (1 - t^{-\alpha})^n$

נבצע החלפת משתנים $t^{-\alpha} = \frac{c(t)}{n}$, מכאן ש $t = \left(\frac{n}{c}\right)^{1/\alpha}$

כל קובץ המועלה לאתר נועד אך ורק לשימושם האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת. כלשהי של קובץ זה או חלק ממנו.

נציב במשוואה (1.18) ונקבל

$$\Pr\left(Y_n \leq \left(\frac{n}{c}\right)^{1/\alpha}\right) = \left(1 - \frac{c}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-c}$$

נגדיר $x = \frac{1}{c^{1/\alpha}}$ מכאן ש $c = x^{-\alpha}$.

$$(1.19) \quad \Pr\left(\frac{Y_n}{\sqrt[n]{n}} \leq x\right) \rightarrow e^{-x^{-\alpha}} \quad \text{עבור } n \rightarrow \infty \text{ נקבל}$$

בדוגמא זו קיבלנו מקרה פרטי של התפלגות גבולית נוספת בשם התפלגות Fréchet.

$$H(x) = e^{-x^{-\alpha}} \quad , \quad \text{עבור } x > 0$$

❖ **Maurice René Fréchet** (1878 - 1973), היה מתמטיקאי צרפתי. הוא תרם רבות בתחום

הטופולוגיה והציג תפיסה של מטריקה כמו כן הוא תרם תרומות חשובות לתחום הסטטיסטיקה, ההסתברות והחשבון הדיפרנציאלי.

מהדוגמאות דלעיל ראינו שערך הקיצון Y_n יכול לשאוף לאחת משלוש ההתפלגויות הגבוליות הבאות:

$$\Pr\left\{\frac{Y_n - b_n}{a_n} \leq x\right\} = F(a_n x + b_n)^n \rightarrow H(x)$$

התפלגות Gumbel, $H(x) = \exp(-e^{-x})$, לכל x ממשי.

התפלגות Fréchet, $H(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases}$, עבור $\alpha > 0$.

התפלגות Weibull, $H(x) = \begin{cases} \exp(-(-X)^\alpha) & x < 0 \\ 1 & x \geq 0 \end{cases}$, עבור $\alpha > 0$.

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

בפרק הבא נראה כי אלה הם שלוש ההתפלגויות ה**יחידות** עבור ערך הקיצון.

פרק 2 - משפטים מרכזיים בהתפלגות ערך המקסימום.

תהי X_1, X_2, \dots, X_n סדרת משתנים מקריים, בלתי תלויים, בעלי התפלגות $F(x)$.
נגדיר משתנה מקרי Y_n שהוא הערך המקסימאלי של X_i .

$$Y_n = \max_{1 \leq i \leq n} X_i$$

לפי ההגדרה לכל סדרת מספרים a_n, b_n מתקיים כי

$$\Pr[Y_n \leq a_n x + b_n] = \Pr\left[\frac{Y_n - b_n}{a_n} \leq x\right] = F^n(a_n x + b_n)$$

אנו נחפש מרכז b_n ונרמול a_n כך שבגבול נקבל התפלגות שאינה תלויה ב- n :

$$\Pr[Y_n \leq a_n x + b_n] \xrightarrow{n \rightarrow \infty} G(x)$$

(2.1) $\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x)$ כלומר

לשם הנוחות נעבוד עם הזנב של F , לכן נרשום $F = 1 - (1 - F)$ ונקבל

$$\lim_{n \rightarrow \infty} [1 - (1 - F(a_n x + b_n))]^n = G(x)$$

א. ניסוח אחר בשאיפה לגבול

עבור $0 < G(x) < 1$, ניקח לוגריתם מן המשוואה הקודמת ונקבל

$$\lim_{n \rightarrow \infty} n \cdot \log[1 - (1 - F(a_n x + b_n))] = \log G(x)$$

מכיוון שכאשר $(1 - F(a_n x + b_n)) \rightarrow 0, n \rightarrow \infty$ נעזר בפיתוח לטור $\log(1 - \varepsilon) \approx -\varepsilon + O(\varepsilon^2)$

$$\lim_{n \rightarrow \infty} [-n(1 - F(a_n x + b_n))] = \log G(x)$$
 ונקבל

(2.2) או
$$\lim_{n \rightarrow \infty} \frac{1}{[n(1 - F(a_n x + b_n))]} = -\frac{1}{\log G(x)}$$

(2.3) נגדיר U , הפונקציה ההפיכה של $\frac{1}{1-F}$ נסמן אותה כך $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$.

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

למציאת a_n ו- b_n נעזר בלמה הבאה. ניתן למצוא את ההוכחה שלה בספר הבא:

Extreme Value Theory: An Introduction by Laurens de Haan & Ana Ferreira.

למה

תהי f_n סדרה של פונקציות מונוטוניות עולות ו- g פונקציה מונוטונית עולה. f_n^{\leftarrow} ו- g^{\leftarrow} יהיו הפונקציות ההפיכות שלהן.

נניח כי לכל x בקטע כלשהו (a, b) היא נקודה רציפה ב- g . $\lim_{n \rightarrow \infty} f_n(x) = g(x)$.

אזי לכל x באינטרוול $(g(a), g(b))$ עבורו הפונקציה g^{\leftarrow} רציפה יתקיים: $\lim_{n \rightarrow \infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x)$.

במקרה שלנו אנו נתייחס רק לפונקציות רציפות.

נעזר בלמה על מנת לפשט את המשוואה (2.2).

נסתכל על אגף ימין של משוואה (2.2) ונסמנו באופן הבא:

$$-\frac{1}{\log G(x)} = \omega(x)$$

נרצה להגיע לפונקציה ההפיכה של הביטוי בצד שמאל

$$G(x) = e^{-1/\omega(x)}$$

$$(2.4) \quad x = G^{\leftarrow}\left(e^{-1/\omega(x)}\right)$$

נסתכל על אגף שמאל של משוואה (2.2) ובאופן דומה נגיע לפונקציה ההפיכה.

$$\lim_{n \rightarrow \infty} \frac{1}{(1 - F(a_n x + b_n))} \cdot \frac{1}{n} = \omega(x)$$

$$\frac{1}{(1 - F(a_n x + b_n))} = \omega n$$

נעזר בהגדרה של U ונקבל

$$(2.5) \quad x = \frac{U(n\omega) - b_n}{a_n}$$

$$(2.6) \quad \lim_{n \rightarrow \infty} \frac{U(n\omega) - b_n}{a_n} = G^{\leftarrow}\left(e^{-1/\omega}\right) := D(\omega) \quad (2.4) \text{ ו-} (2.5) \text{ ממשוואות}$$

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת. כלשהי של קובץ זה או חלק ממנו.

משפט (1)

יהי $a_n > 0$ ו- b_n סדרת קבועים ממשיים. תהי G פונקציית התפלגות רציפה. הטענות הבאות שקולות:

$$(2.7) \quad \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \quad \text{מתקיים} \quad 0 < G(x) < 1 \quad \text{עבור}$$

$$(2.8) \quad \lim_{t \rightarrow \infty} t(1 - F(a(t)x + b(t))) = -\log G(x) \quad (2)$$

$a(t) := a_{[t]}$ ו- $b(t) := b_{[t]}$, הוא החלק השלם של t .

$$(2.9) \quad \lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x) \quad (3)$$

$$. D(x) = G^{\leftarrow} \left(e^{-1/x} \right)$$

משפט (1) הינו בעל חשיבות רבה להמשך ההוכחה, המשמעות היא מעבר מניתוח של סדרות a_n, b_n לניתוח של פונקציות $a(t), b(t)$, המוגדרות לכל $t > 1$.

הוכחה (1)

השקילות בין טענה (2) ל- (3) נובעת מלמה (1). ראינו כי טענה (1) שקולה למשוואה (2.6), לכן מספיק להוכיח שמשוואה (2.6) מובילה לטענה (3). עבור $t \geq 1$,

$$(2.10) \quad \frac{U([t]x) - b_{[t]}}{a_{[t]}} \leq \frac{U(tx) - b_{[t]}}{a_{[t]}} \leq \frac{U([t]x(1 + 1/[t])) - b_{[t]}}{a_{[t]}}$$

מכאן שמתקיים

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b_{[t]}}{a_{[t]}} = D(x)$$

וזוהי טענה (3).

מ.ש.ל (1)

ב. התפלגות ערך הקיצון

כעת ננסה לזהות משפחה של התפלגויות רציפות שמתקבלות בגבול של המשוואה הבסיסית (2.1), משפחה זו נקראת משפחת התפלגויות הקיצון.

משפט (2)

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת. כלשהי של קובץ זה או חלק ממנו.

תהי G התפלגות קיצון גבולית המתאימה למשוואה (2.1) אשר שייכת למשפחת התפלגות ערכי קיצון

עבור $G_\gamma(ax+b)$ עבור $a > 0$ ו- b ממשי, המקיימת עבור $\gamma \neq 0$

$$(2.11) \quad G_\gamma(x) = \exp\left(-\left(1+\gamma x\right)^{-1/\gamma}\right), \quad 1+\gamma x > 0$$

ואילו עבור $\gamma = 0$, הביטוי מתפרש כ- $G_0(x) = \exp(-e^{-x})$.

הגדרה: הפרמטר γ ב- (2.11) נקרא אינדקס ערך הקיצון (extreme value index).

הוכחה (2)

נסתכל על משפחת פונקציות הגבול D ב- (2.9).

תחילה נניח כי 1 היא נקודת רציפות של D , עבור נקודות רציפות $x > 0$ מתקיים,

$$(2.12) \quad \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D(x) - D(1) =: E(x)$$

ניקח $y > 0$,

$$(2.13) \quad \frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}$$

טענה: $\lim_{t \rightarrow \infty} \frac{U(ty) - U(t)}{a(t)}$ ו- $\lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$ קיימים.

הוכחה: נניח שלא קיימים. אז ישנם A_1, A_2, B_1, B_2 כאשר $A_1 \neq A_2$ או $B_1 \neq B_2$ ו- B_i הן נקודות

הגבול של $U(ty) - U(t)/a(t)$ ו- A_i הן נקודות הגבול של $a(ty)/a(t)$, עבור $i=1,2$ עבור $t \rightarrow \infty$.

מ- (2.13) מתקיים

$$(2.14) \quad E(xy) = E(x)A_i + B_i$$

עבור x שרירותי, ניקח רצף של נקודות רציפות x_n , כאשר $x_n \uparrow x$ ($n \rightarrow \infty$).

מתקיים $E(x_n y) \rightarrow E(xy)$ ו- $E(x_n) \rightarrow E(x)$.

לכן (2.14) נכון עבור כל x ו- y חיוביים.

נשכתב את המשוואה (2.14) עבור $i=1,2$.

$$E(xy) = E(x)A_1 + B_1 = E(x)A_2 + B_2$$

$$E(x)(A_1 - A_2) = B_2 - B_1$$

לכל $x > 0$.

מכיוון ש- E לא מספר קבוע (נזכור ש- G רציפה), אז מתקיים $A_1 = A_2$ ו- $B_1 = B_2$ וזאת בסתירה

להנחה.

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

$$A(y) := \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)} \quad \text{מסקנה: עבור } y > 0 \text{ מתקיים}$$

$$(2.15) \quad E(xy) = E(x)A(y) + E(y) \quad \text{ועבור } x, y > 0 \text{ מתקיים}$$

$$H(x) := E(e^x) - 1, \quad (x, y \neq 1) \quad t := \ln y, \quad s := \ln x$$

נסמן
נשכתב את (2.15)

$$(2.16) \quad H(t+s) = H(s)A(e^t) + H(t)$$

אפשר לכתוב את (2.16) כך

$$(2.17) \quad H(0) = 0, \quad \frac{H(t+s) - H(t)}{s} = \frac{H(s) - H(0)}{s} A(e^t)$$

קיים בוודאות t מסוים עבורו H גזירה (מכיוון ש- H מונוטונית), לכן H גזירה בכל מקום. נשתמש בהגדרת הנגזרת עבור (2.17) ונקבל

$$(2.18) \quad H'(t) = H'(0)A(e^t)$$

נסמן

$$H'(0) \neq 0, \quad Q(t) = H(t)/H'(0) \\ Q(0) = 0, \quad Q'(0) = 1$$

לפי (2.16)

$$Q(t+s) - Q(t) = Q(s)A(e^t)$$

לפי (2.18)

$$\begin{cases} Q(t+s) - Q(t) = Q(s)Q'(t) \\ Q(t+s) - Q(s) = Q(t)Q'(s) \end{cases}$$

$$(2.19) \quad Q(t) \frac{Q'(s) - 1}{s} = \frac{Q(s)}{s} (Q'(t) - 1) \quad \text{נקבל}$$

נציב ב- (2.19) $Q'(0) = 1$ ו- $Q(0) = 0$.

$$Q(t)Q''(0) = Q'(0)(Q'(t) - 1) \quad \text{עבור } s \rightarrow 0 \text{ נקבל}$$

$$Q(t)Q''(0) = Q'(t) - 1 \quad \text{כאמור } Q'(0) = 1$$

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

$$Q'(t)Q''(0) = Q''(t) \quad \text{נגזור ונקבל}$$

$$(2.20) \quad Q''(0) = \frac{Q''(t)}{Q'(t)} = \frac{d}{dt} [\log Q'(t)] := \gamma \in \mathbb{R}$$

לכל t .

$$\gamma = \int_0^t \left(\frac{d}{dt} [\log Q'(t)] \right) dt = \log Q'(t) - \log Q'(0) \quad \text{נבצע אינטגרל}$$

$$\gamma = \log \frac{Q'(t)}{Q'(0)}$$

$$(2.21) \quad Q'(t) = Q'(0)e^{\gamma t}$$

נבצע אינטגרל נוסף

$$Q(t) - Q(0) = \int_0^t e^{\gamma s} ds$$

$$(2.22) \quad \frac{H(t)}{H'(0)} = Q(t) = \frac{1}{\gamma} (e^{\gamma t} - 1)$$

$$E(x) = D(x) - D(1) \quad , \quad H(x) = E(e^x) \quad \text{נזכור כי}$$

$$D(t) = D(1) + E(t) = D(1) + H(\ln t)$$

ניעזר ב- (2.22)

$$D(1) + H'(0) \frac{t^\gamma - 1}{\gamma} = y$$

$$t^\gamma - 1 = \frac{\gamma}{H'(0)} [y - D(1)]$$

$$(2.23) \quad t = \left(1 + \frac{\gamma}{H'(0)} [y - D(1)] \right)^{1/\gamma}$$

$$D(t) = G^{\leftarrow} \left(e^{-1/t} \right) \quad (2.6) \quad \text{נזכור כי לפי (2.6)}$$

$$D^{\leftarrow}(x) = \frac{-1}{\log G(x)} \quad \text{הפונקציה ההפיכה של D היא}$$

$$\log G(x) = \frac{-1}{D^{\leftarrow}(x)} \quad \text{לכן}$$

כל קובץ המועלה לאתר נועד אך ורק לשימושם האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

ידוע כי $D^{\leftarrow}(t) = t$ נציב את t ממשוואה (2.23) ונקבל

$$\log G(x) = \frac{-1}{\left(1 + \frac{\gamma}{H'(0)}[y - D(1)]\right)^{1/\gamma}}$$

$$G(x) = \exp\left(-\left(1 + \gamma\left[\frac{y}{H'(0)} - \frac{D(1)}{H'(0)}\right]\right)^{-1/\gamma}\right)$$

נסמן $a = \frac{y}{H'(0)}$ ו- $b = \frac{D(1)}{H'(0)}$

קיבלנו הצגה פרמטרית אחת מהצורה

$$(2.24) \quad G_\gamma(x) = \exp\left(-\left(1 + \gamma x\right)^{-1/\gamma}\right), \quad 1 + \gamma x > 0$$

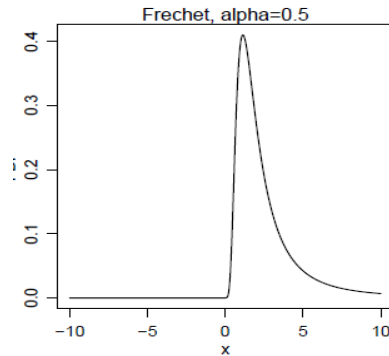
מ.ש.ל (2)

משפט זה הוא בעל חשיבות רבה. אנו רואים שקיימת משפחה **אחת** של התפלגות גבולית המייצגת שלוש קבוצות של התפלגות גבולית בעלי אופי שונה. להלן שלוש ההתפלגויות המתקבלות ממשוואה (2.24).

1. עבור $\gamma > 0$ ו- $\alpha = 1/\gamma > 0$, נסמן $\Phi_\alpha(x) = G_\gamma((x-1)/\gamma)$.

$$\Phi_\alpha(x) := \begin{cases} 0 & , x \leq 0 \\ \exp(-x^{-\alpha}) & , x > 0 \end{cases} \quad \text{ניתן לראות כי}$$

קבוצה זו נקראת **קבוצת התפלגות Fréchet**, אשר צפיפותה מתוארת בגרף 3, עבור $\alpha = 0.5$.



גרף 3

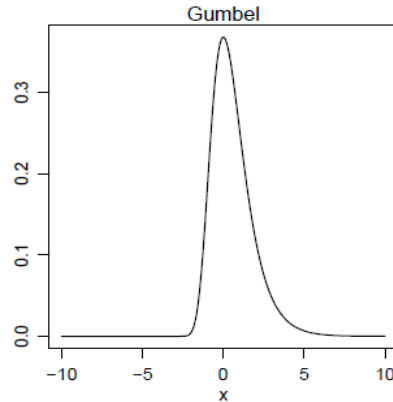
כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת. כלשהי של קובץ זה או חלק ממנו.

2. כאשר $\gamma = 0$, פונקצית ההתפלגות היא:

$$G_0(x) = \exp(-e^{-x})$$

לכל x ממשי,

זוהי התפלגות **Gumbel**, אשר צפיפותה מתוארת בגרף 4.

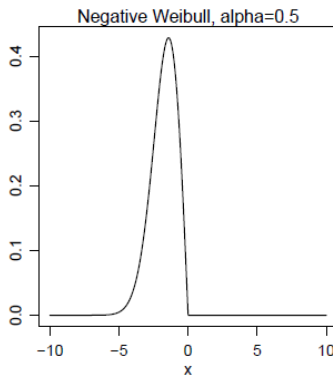


גרף 4

3. עבור $\gamma < 0$ ו- $\alpha = -1/\gamma > 0$, נסמן $\Psi_\alpha = G_\gamma(- (x-1)/\gamma)$

$$\Psi_\alpha(x) := \begin{cases} \exp(-(-x)^\alpha) & , x < 0 \\ 1 & , x \geq 0 \end{cases} \quad \text{ניתן לראות כי}$$

התפלגות זו נקראת **קבוצת התפלגות Weibull**, אשר צפיפותה מתוארת בגרף 5, עבור $\alpha = 0.5$.



גרף 5

פרק 3 - מהם מקרי קיצון? עד כמה זה חשוב לנו? וכיצד מנתחים אותם?

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

בתחומים רבים כגון: כלכלה, מדעי החברה והמדעים המדויקים נאספים נתונים רבים. למשל, ערכי מניות, ערכי תביעות ביטוח, כמות משקעים שנתית, הכנסה של משפחה ועוד. ביישומים רבים המדדים המעניינים הם הממוצע והשונות. למשל, ממוצע הכנסה למשפחה הוא נתון כלכלי חשוב. על כן במספר רב של יישומים, ערכים חריגים נחשבים כיוצאי דופן (outliers) ובדרך כלל מתעלמים מהם בניתוח הסטטיסטי. לדוגמא: בדרך כלל מחושב חציון ההכנסה למשפחה ממוצעת, מדד זה מתעלם ממעט המשפחות עם הכנסות גבוהות מאוד. עם זאת יש מקרים שבהם הגודל החשוב הוא דווקא ערכי הקיצון.

ניקח לדוגמא את הולנד. מחצית מהולנד נמצאת מתחת לפני הים. ברגע שסערה גדולה תפרוץ רוטרדם, הנמל העמוס ביותר באירופה, יוצף. הולנד עומדת בפני שני אתגרים למניעת הצפתה במים: (1) התמודדות עם משקעים. (2) התמודדות עם הגאות של הים. בשנת 1953 פגע בהולנד מבול קטלני, שהשתרע על יותר מ-600 קילומטרים רבועים והרג יותר מ-1800 אנשים. הולנד ידועה בתור אחת המדינות המפותחות ביותר בהגנה מפני הצפות מסוג זה, במיוחד לאור העובדה שקיימת תרומה גדולה של התחממות הגלובלית לכמות המים הגואה בעולם. גובה הסכר הדרוש, במקרה של הולנד ומדינות בעלי אתגרים דומים, היא דוגמא לבעיה מתמטית בה נעשה שימוש בהערכות על פי משפט הגבול של ערך הקיצון.

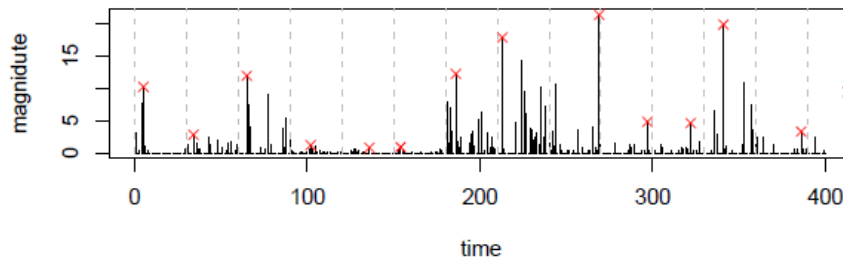
ערכי קיצון הם בעלי חשיבות אדירה גם בתעשייה הפיננסית, לרבות בנקאות וביטוח. לדוגמא, חברת ביטוח לוקחת פרמיות קטנות ומבטיחה להחזיר סכומים גדולים בהרבה במקרה של אסון. על מנת שחברת הביטוח לא תפשוט את הרגל ויהיו בחשבונה מספיק כסף להחזר התביעה, עליה לחשב את הסיכוי לסך תביעות שנתי גדול במיוחד. שתי הדוגמאות שהובאו לעיל הם רק מקרים פרטיים בהם נדרש להבין התפלגות של ערכי קיצון. כפי שראינו בפרק 2, התפלגות ערך הקיצון תלויה בהתפלגות הזנב של המשתנים המקריים המקוריים. לרוב התפלגות זו אינה ידועה.

ישנם שתי שיטות עיקריות בניתוח סטטיסטי של ערכי קיצון. הראשונה היא להתאים את הנתונים למודל בעזרת טכניקות סטטיסטיות מסורתיות, ולהסתכל על ערכי הקיצון בעזרת סימולציה על המודל. זאת אומרת מנסים להתאים את התפלגות הזנב להתפלגויות ידועות שונות. השיטה השנייה היא להתאים את הנתונים להתפלגות הגבולית של ערך הקיצון (משוואה 2.11), המסומנת ב-EVD (Extreme Value Distribution). לשיטה הזאת יש שתי גישות: Block maxima ו-Peaks over threshold (POT).

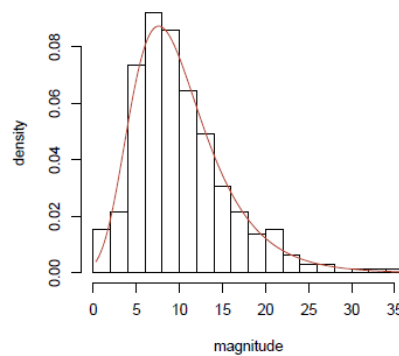
Block maxima

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

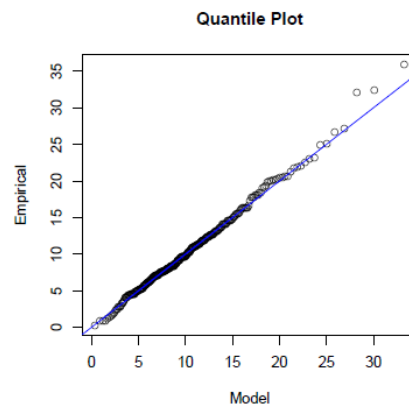
בגישה זו, לוקחים את הנתונים לאורך זמן ומחלקים אותם ל"בלוקים" של יחידות זמן (כמו בגרף 6). בכל "בלוק" בוחרים את הערך הגבוה ביותר, ויוצרים גרף בעזרת נתוני המקסימום האלה (ראה גרף 7). את הנתונים מנסים להתאים ל-EVD עם פרמטרים שונים, על ידי גרף QQ (Quantil-Quantil plot). היא שיטת התאמה גראפית של התפלגות הנתונים הגולמיים להתפלגות EVD בעזרת הישר $y = x$ (ראה גרף 8).



גרף 6 - Block maxima



גרף 7 - התפלגות הערכים הגבוהים ב"בלוקים".

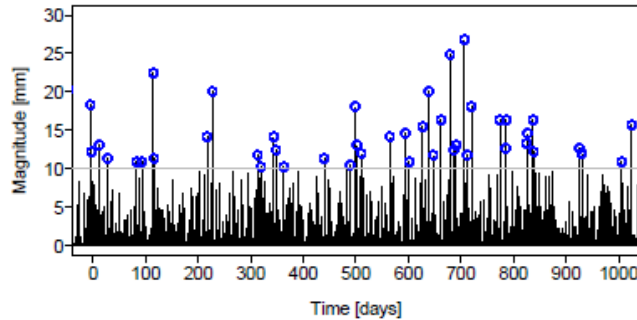


גרף 8 - התאמת התפלגות ערכי הקיצון למודל אסימפטוטי

(POT) Peaks over threshold

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת. כלשהי של קובץ זה או חלק ממנו.

על פי גישה זו בוחרים סף מסוים (threshold) ודוגמים את הערכים הגבוהים ממנו (ראה גרף 9). בגישה הזאת אין השפעה של חלוקת הזמן, אבל צריך לדעת לבחור את אותו סף בחוכמה. סף גבוה מדי עלול למחוק נתונים חיוניים ולגרום לכך שהערכה שלנו תהיה גבוהה מדי, סף נמוך מדי עלול לגרום להטיה, בגלל שמשוואה (2.11) היא פתרון אסימפטוטי בה נדרש סף גבוה דיו.



גרף 9 - POT Peaks over threshold

גם בגישה זו כמו בקודמת, מתאימים את התפלגות הממצאים המקסימאליים להתפלגות EVD, על מנת למצוא אינדקס ערך קיצון γ מתאים, כדי לקבל הערכה טובה. ישנם אפליקציות ועזרים שונים שיכולים להיעזר בהם על מנת להעריך פרמטרים שונים במקרי קיצון. בניהם כלים גרפיים המצויים ברשת כגון: Toolkit, exRemes.

פרק 4 – סיכום

בעבודה זו ראינו כי קיימת תיאוריה מתמטית שלמה של מקרי קיצון. להתפלגות הערך המקסימאלי יש הצגה פרמטרית **יחידה המייצגת שלוש** התפלגויות שונות: התפלגות Gumbel, התפלגות Fréchet והתפלגות Weibull. **Exteremes**, ערכי קיצון הם בעלי חשיבות רבה בתחומים רבים בחיינו. מושגים כמו, ברבור שחור, אל ניניו וכו', אינם אלא מקרי קיצון, להם נוכל להיערך או לפחות להבין בעזרת ניתוח סטטיסטי המסתמך על משפט הגבול שהוצג בעבודה זו, כפי שהובא בפרק 3.

ביבלוגרפיה

Chavez-Demoulin, V., & Roehrl, A. (2004). Extreme Value Theory can save your neck. [Electronic version]. Retrieved from

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.

http://www.approximity.com/papers/evt_wp.pdf

De Haan, L. & Ferreira, A. (2006). *Extreme Value Theory*. Springer.

Embrechts, P., Resnick, S., & Samorodnitsky, G. (1999). Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal*, Volume 3, Number 2. [Electronic version]. Retrieved from http://www.casact.org/library/studynotes/Embrechts_ExtremeValue.pdf

Gilleland, E., & Katz, R. (2006). Analyzing Seasonal to Interannual Extreme Weather and Climate Variability with the Extremes Toolkit. [Electronic version]. Retrieved from <http://www.assessment.ucar.edu/pdf/Gilleland2006revised.pdf>

Mersic, M. (2007). An Introduction to Extreme Value Theory. [Electronic version]. Retrieved from http://mersic.org/evt/evt1_paper.pdf

Wainer, N. (2007). The Most Dangerous Equation. [Electronic version]. Retrieved from <http://nsm.uh.edu/~dgraur/niv/TheMostDangerousEquation.pdf>

כל קובץ המועלה לאתר נועד אך ורק לשימוש האישי של מורים למתמטיקה, פיזיקה, כימיה וביולוגיה ולהוראה בכיתותיהם. אין לעשות שימוש כלשהו בקובץ זה לכל מטרה אחרת, ובכלל זה: שימוש מסחרי, פרסום באתר אחר (למעט אתר בית הספר בו מלמד המורה), העמדה לרשות הציבור או הפצה בדרך אחרת כלשהי של קובץ זה או חלק ממנו.