Eötvös Loránd Tudományegyetem

Természettudományi Kar

Horvat Anna

Konvergencia-sebesség az extrémérték modellekben

BSc Elemző Matematikus Szakdolgozat

Témavezető:

Dr. Zempléni András

Valószínűségelméleti és Statisztika Tanszék

Budapest, 2016

Tartalomjegyzék

Bevezetés **4**

1. Elméleti háttér **5**
1.1. Motiváció 5
1.2. Az n elemű minta maximumának határeloszlása 6
1.3. Maximum vonzási tartományok 8

2. Konvergencia-sebesség **11**
2.1. Egyenletes konvergencia-sebesség 12
2.2. A becslés pontosítása 15
2.3. Fordított eredmények 17
2.4. Konvergencia Ψ_α-hoz 20

3. Kitekintés **22**
3.1. Küszöbmeghaladás 22
3.2. Becslés GPD-vel 24

4. Szimulációk **26**
4.1. n elemű, exponenciális eloszlású minta maximumának eloszlása 27
4.2. n elemű, egyenletes eloszlású minta maximumának eloszlása 28
4.3. n elemű, standard normális eloszlású minta maximumának eloszlása 30

Összegzés **34**

Köszönetnyilvánítás

Köszönöm témavezetőmnek, Zempléni Andrásnak, hogy szakdolgozatom elkészítéséhez hasznos tanácsokat adott és iránymutatásul szolgált. Hálás vagyok türelméért és megértéséért. Valamint köszönöm családomnak az egyetemi éveim alatt nyújtott kitartó támogatást.

Bevezetés

Extrém események az élet minden területén előfordulhatnak. Ezen események közös jellemzője, hogy nagyon ritkán következnek be, viszont hatásuk annál jelentősebb. Ilyen eseménynek tekinthetők a nagy árvizek, földrengések; a pénzügyi világban extrém esemény például egy tőzsdekrach, vagy egy nagy katasztrófa esetén a biztosítók legnagyobb kifizetései. Az extrémérték elmélet fő feladata, hogy a kevés ismert megfigyelés, adatok alapján olyan modelleket készítsen, amelyekkel elég jó előrejelzések adhatók ezen események bekövetkezésére.

Dolgozatom célja az egyes valószínűségi változó sorozatok maximumainak konvergenciájának vizsgálata a megfelelő extrémérték eloszlásokhoz, illetve a velük kapcsolatos elméleti eredmények összefoglalása. Ezzel összhangban bevezetem az ún. g-sv függvények fogalmát, amely segítségünkre szolgál a konvergencia rendjének megállapításában.

Az első fejezetben röviden ismertetem a klasszikus extrémérték elméleti eredményeket, köztük a Fisher-Tipett-Gnedenko tételt, amely az extrémérték elmélet alapját képezi: egy minta maximumának vagy minimumának háromféle határeloszlása lehet; valamint a GEV eloszláscsaládot, amellyel a három extrémérték eloszlás egy formulába írható (Jenkinson, 1955). Ezt követően a második fejezetben Richard L. Smith: Uniform Rates of Convergence in Extreme Value Theory [7] c. cikke alapján részletesebben kitérek a mintamaximumok közelítésére az extrémérték eloszlásokhoz: ismertetem a korábban említett g-sv függvényeket, melyeket alkalmazok is a konvergencia vizsgálatánál. A három határeloszlás közül Fréchet eloszláshoz vizsgálom részletesen a mintamaximumok konvergenciáját, röviden pedig a Weibull határeloszláshoz. A harmadik fejezetben egy kis kitérő következik: mintamaximumok helyett egy minta u küszöb feletti, "elég nagy" értékeit vizsgálom, és bemutatom, miként alkalmazhatóak a második fejezetben bemutatott g-sv függvények a küszöbmeghaladáson alapuló modellezésben. Végül a negyedik fejezetben szimulációkon keresztül szemléltetem az egyes extrémérték eloszlásokhoz a konvergenciát.

1. fejezet

Elméleti háttér

1.1. Motiváció

A független valószínűségi változók maximumának határeloszlásáról szóló elmélet a centrális határeloszlás tétellel (CHT) számos hasonlóságot mutat. Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók. Míg a CHT az $X_1 + \ldots + X_n$ összeg határeloszlását adja meg $n \to \infty$ esetén, addig az extrém érték elmélet a minta extrémumainak ($\max(X_1, X_2, \ldots, X_n)$ vagy $\min(X_1, X_2, \ldots, X_n)$) határeloszlását vizsgálja, ha $n \to \infty$. Tehát a CHT az extrém érték elmélet motivációjaként is tekinthető.

Legyen $X_1+X_2+\ldots+X_n = S_n$. A CHT kimondja, hogy $X_1, X_2, \ldots$ független, azonos eloszlású, véges szórású valószínűségi változók összege alkalmas konstansokkal normalizálva a standard normális eloszlásfüggvényhez konvergál eloszlásban, ha n tart a végtelenbe. A konstansok: $a_n = n \cdot EX_1$ és $b_n = \sqrt{n} \cdot DX_1$. A tétel formalizálva:

$$P\left(\frac{S_n - a_n}{b_n} \leq x\right) \to \Phi(x), \quad n \to \infty. \tag{1.1}$$

1.1. Definíció. Legyen $X_1, X_2, \ldots$ független, azonos F eloszlású valószínűségi változók. Ekkor az F a G *vonzási tartományába* tartozik, ha megadható a_n, b_n normáló sorozat, hogy

$$\frac{X_1 + \ldots + X_n - a_n}{b_n} \to G.$$

Tehát (1.1) esetén az F eloszlásfüggvény a normális eloszlás vonzási tartományába tartozik, azaz $F \in DA(\Phi)$. Ám ha a szórásnégyzet végtelen, a független, azonos eloszlású valószínűségi változók normalizált összege α-stabilis eloszláshoz tart.

1.2. Definíció. Egy X valószínűségi változó *stabilis*, ha X_1 és X_2 függetlenek és azonos eloszlásúak X-szel, és bármely $a > 0$-ra és $b > 0$-ra teljesül, hogy $aX_1 + bX_2 \stackrel{d}{=} cX + d$, ahol $c > 0$ és d valós.

1.2. Az n elemű minta maximumának határeloszlása

Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók sorozata F közös eloszlásfüggvénnyel, és $M_n = \max(X_1, \ldots, X_n)$. Az M_n maximum eloszlásfüggvénye:

$$P(M_n \leq x) = P(X_1 \leq x, \ldots, X_n \leq x) = F^n(x).$$

Ekkor az $M_n \to x_F$, ha $n \to \infty$ 1 valószínűséggel, ahol x_F az F jobboldali végpontja, vagyis $x_F = \sup\{x \in \mathbb{R} \colon F(x) < 1\}$. Ez nem igazán hordoz magában információt, így felmerül a kérdés, hogy van-e nem elfajult határeloszlása a lineárisan transzformált M_n maximumnak. A CHT-hez hasonlóan próbáljuk a határeloszlást meghatározni, keressük, mely eloszlás vonzási tartományába fog tartozni.

1.3. Tétel. (Fisher-Tippett-Gnedenko tétel) *Legyen $X_1, X_2 \ldots$ független, azonos eloszlású valószínűségi változók sorozata F közös eloszlásfüggvénnyel. Ekkor, hogyha léteznek olyan valós $a_n > 0$, b_n normalizáló sorozatok, és*

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \to G(x), \quad n \to \infty, \tag{1.2}$$

ahol G nem elfajult eloszlásfüggvény, akkor G az alábbi három eloszlásfüggvény-típus valamelyikébe tartozik:

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0, \alpha > 0 \end{cases}$$

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

$$\Lambda_\alpha(x) = \exp\{-\exp\{-x\}\}, \quad -\infty < x < \infty.$$

1.4. Definíció. A $\Phi_\alpha(x)$ Fréchet, a $\Psi_\alpha(x)$ Weibull és $\Lambda_\alpha(x)$ Gumbel eloszlásfüggvényeket *standard extrémérték eloszlásfüggvényeknek* hívjuk.

1.5. Definíció. Ha $X_1, \ldots, X_n$ független, az X változóval azonos eloszlású valószínűségi változók, valamint $a_n > 0$ és b_n alkalmas konstansok úgy, hogy

$$M_n \stackrel{d}{=} a_n X + b_n, \tag{1.3}$$

akkor az X valószínűségi változót *max-stabilisnak* nevezzük.

Legyen $X_1, X_2 \ldots$ független, azonos eloszlású valószínűségi változók. Ekkor az (1.3) átírható az alábbi alakba:

$$\frac{M_n - b_n}{a_n} \stackrel{d}{=} X. \tag{1.4}$$

Megállapíthatjuk, hogy minden max-stabilis eloszlás határeloszlása a lineárisan transzformált független, azonos eloszlású valószínűségi változók maximumának. Sőt a max-stabilis eloszlások az egyedüli határeloszlásai a normalizált maximumoknak.

1.6. Tétel. *Max-stabilis eloszlások halmaza egybeesik az alkalmasan normált maximumok határeloszlásával.*

Ennélfogva, hogyha az X olyan extrém valószínűségi változó, amely eleget tesz (1.4)-nek, a Fisher-Tipett-Gnedenko tétel 3 esete megegyezik a következőkkel:

- Fréchet: $M_n \stackrel{d}{=} n^{\frac{1}{\alpha}} X$, ugyanis ha $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók $\Phi_\alpha(x)$ eloszlásfüggvénnyel, akkor

$$P(M_n < n^{\frac{1}{\alpha}} x) = \Phi_\alpha^n(n^{\frac{1}{\alpha}} x) = \exp\{-n(n^{\frac{1}{\alpha}} x)^{-\alpha}\} = \Phi_\alpha(x)$$

- Weibull: $M_n \stackrel{d}{=} n^{-\frac{1}{\alpha}} X$, ugyanis ha $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók $\Psi_\alpha(x)$ eloszlásfüggvénnyel, akkor

$$P(M_n < n^{-\frac{1}{\alpha}} x) = \Psi_\alpha^n(n^{-\frac{1}{\alpha}} x) = \exp\{-n(-n^{-\frac{1}{\alpha}} x)^{\alpha}\} = \Psi_\alpha(x)$$

- Gumbel: $M_n \stackrel{d}{=} X + \ln n$, ugyanis ha $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók $\Lambda_\alpha(x)$ eloszlásfüggvénnyel, akkor

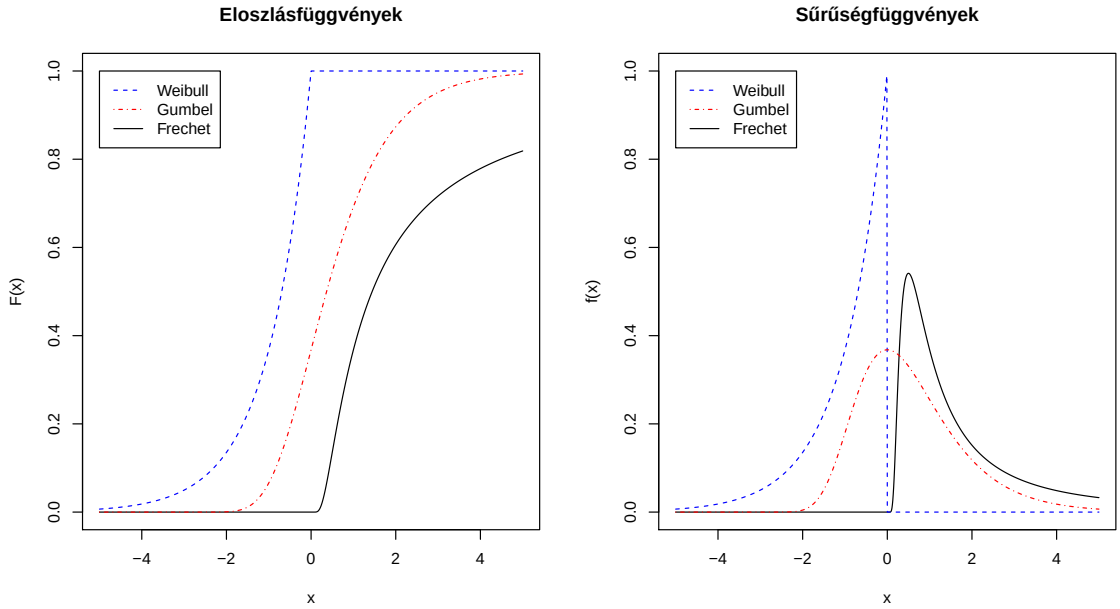
$$P(M_n < x + \ln n) = \Lambda_\alpha^n(x + \ln n) = \exp\{-n \exp\{-x - \ln n\}\} = \Lambda_\alpha(x).$$

Az előző három eloszláscsaládot magába foglalja a következő nagy eloszláscsalád, a GEV (Generalized Extreme Value) eloszlás:

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \tag{1.5}$$

ahol $1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0$. A képletben a μ a hely-, a σ a skálaparaméter. A ξ az alakparaméter, amely megadja az eloszlás szélének lecsengésének sebességét. Ha

- $\xi > 0$: lassan lecsengő eloszlás (Fréchet)
- $\xi = 0$: gyorsan lecsengő eloszlás (Gumbel)
- $\xi < 0$: rövid szélű eloszlás (negatív Weibull).



1.1. ábra. Standard extrém érték eloszlások eloszlásfüggvénye és sűrűségfüggvénye

1.3. Maximum vonzási tartományok

A továbbiakban vizsgáljuk meg, hogy milyen feltételeknek kell érvényesülniük F-re, hogy az F eloszlású független valószínűségi változók standardizált maximuma a megfelelő max-stabilis eloszláshoz konvergáljanak eloszlásban. De előbb, ahogy a valószínűségi változók összegénél is tettük, vezessük be az extrém érték eloszlások vonzási tartományát is.

1.7. Definíció. Abban az esetben, ha (1.2) teljesül valamilyen nem elfajult G eloszlásfüggvénnyel, akkor F a G *max-vonzási tartományában* (maximum domain of attraction, röviden MDA) van. Jelölése: $F \in MDA(G)$.

Láthatjuk, hogy mivel az extrém érték eloszlások folytonosak, így az $\dfrac{M_n - b_n}{a_n} \to G$ kifejezés felírható a következőképpen is:

$$\lim_{n\to\infty} P\left(M_n < a_n x + b_n\right) = \lim_{n\to\infty} F^n\left(a_n x + b_n\right) = G(x) \qquad \text{ahol } x \in \mathbb{R}.$$

Ezt ismerve könnyen belátható a következő állítás.

1.8. Állítás. *Az F eloszlás a G extrém érték eloszlás max-vonzási tartományába tartozik az $a_n > 0$, b_n konstansokkal akkor és csak akkor, ha*

$$n\overline{F}(a_n x + b_n) \to -\ln G(x), \quad n \to \infty, x \in \mathbb{R}$$

Ha $G(x) = 0$, a határérték ∞-nek vehető.

A konvergencia vizsgálathoz szükséges még ismertetnünk a reguláris- és lassú változású függvények fogalmát.

1.9. Definíció. (Reguláris változású függvény) A $h\colon (0,\infty) \mapsto (0,\infty)$ függvény *reguláris változású* a végtelenben $p \in \mathbb{R}$ indexszel, ha $x > 0$ esetén

$$\lim_{t\to\infty} \frac{h(tx)}{h(t)} = x^p.$$

1.10. Definíció. (Lassú változású függvény) Legyen $h\colon (0,\infty) \mapsto (0,\infty)$ regurális változású függvény. Ha $p = 0$, vagyis

$$\lim_{t\to\infty} \frac{h(tx)}{h(t)} = 1,$$

akkor a h függvényt *lassú változásúnak* mondjuk.

Ha egy $h(x)$ függvény reguláris változású p indexszel, akkor mindig felírható $h(x) = x^p L(x)$ alakban, ahol L egy lassú változású függvény.

A továbbiakban vizsgáljuk meg az egyes extrémérték eloszlások maximum vonzási tartományainak karakterizációját. Elsőként tekintsük a Fréchet-eloszlás esetét.

1.11. Tétel. (Fréchet-eloszlás max-vonzási tartománya) *Az F eloszlásfüggvény a $\Phi_\alpha(x)$ vonzási tartományába tartozik akkor és csak akkor, ha*

$$\overline{F}(x) = x^{-\alpha} L(x), \tag{1.6}$$

ahol L egy lassú változású függvény, és $\overline{F} = 1 - F$.

Az $\mathrm{MDA}(\Phi_\alpha)$ összes eloszlásának jobboldali végpontja végtelen, vagyis $x_F = \infty$.

A Weibull-eloszlásból egyszerű transzformációval megkaphatjuk a Fréchet eloszlást, a kapcsolatuk a következőképpen írható le:

$$\Psi_\alpha\left(-\tfrac{1}{x}\right) = \Phi_\alpha(x), \quad x > 0.$$

Ennek ismeretében várható, hogy a vonzási tartományaik is szoros kapcsolatban vannak. Továbbá ismert, hogy a Weibull max-vonzási tartományába tartozó eloszlásoknak van x_F véges jobboldali végpontjuk.

1.12. Tétel. (Weibull-eloszlás max-vonzási tartománya) *Az F eloszlásfüggvény a $\Psi_\alpha(x)$ vonzási tartományába tartozik akkor és csak akkor, ha x_F véges és*

$$\overline{F}\left(x_F - \tfrac{1}{x}\right) = x^{-\alpha} L(x), \tag{1.7}$$

ahol L egy lassú változású függvény.

A Gumbel eloszlás max-vonzási tartománya tartalmaz számos véges- és végtelen jobb végponttal rendelkező eloszlást. Ezen eloszlás vonzási tartományának jellemzése jóval bonyolultabb az előző kettőnél, összességben az exponenciális lecsengésű eloszlások tartoznak ide. Karakterizáció helyett nézzünk egy példát.

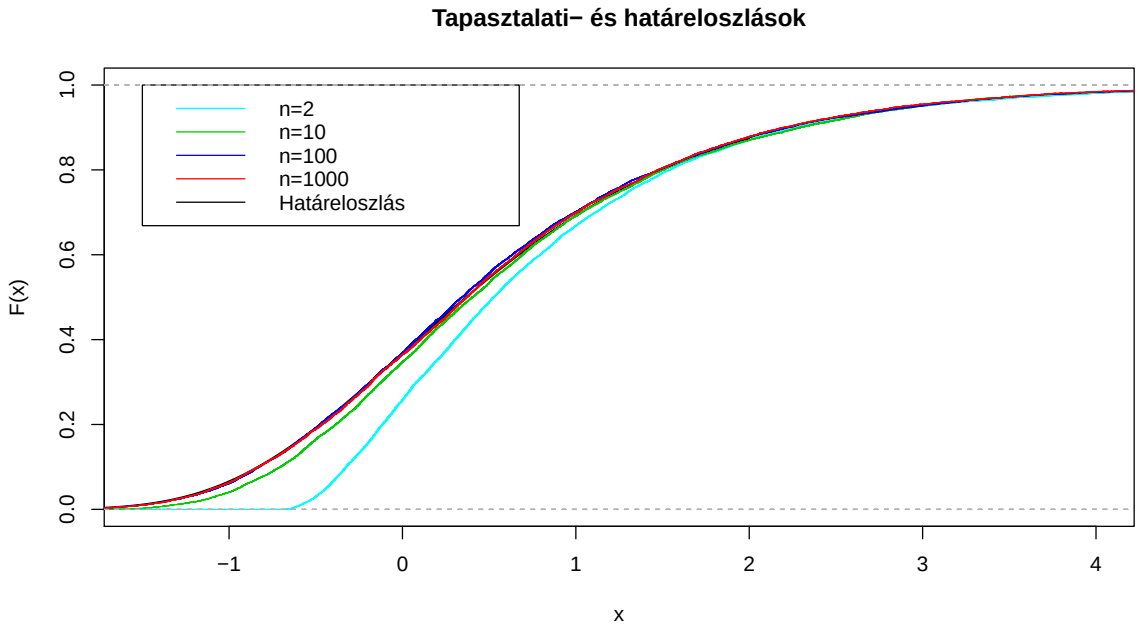
1.13. Példa. Legyen F λ paraméterű exponenciális eloszlás. Ekkor

$$P\left(\frac{M_n - b_n}{a_n} < x\right) = F^n(a_n x + b_n) = (1 - \exp\{-\lambda(a_n x + b_n)\})^n$$

Válasszuk a konstansokat úgy, hogy $a_n = \dfrac{1}{\lambda}$ és $b_n = \dfrac{\log n}{\lambda}$. Így

$$\left(1 - \exp\{-x\}\frac{1}{n}\right)^n \to \exp\left\{-\exp\{-x\}\right\},$$

tehát $F \in MDA(\Lambda)$. Az 1.2 ábrán látható, hogy véletlenszerűen generált exponenciális eloszlású n elemű minta maximumai $\lambda = 2$ paraméter mellett, a példában alkalmazott konstansokkal valóban a Gumbel eloszláshoz tartanak.



1.2. ábra. $\lambda = 2$ paraméterű exponenciális eloszlású n elemű minta maximumának konvergenciája a Gumbel határeloszláshoz. A konvergencia viszonylag gyors.

2. fejezet

Konvergencia-sebesség

Ebben a fejezetben a Fisher-Tipett-Gnedenko tételben megfogalmazott

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \to G(x), \quad n \to \infty$$

konvergencia sebességét fogjuk vizsgálni. A témát Richard L. Smith [7] cikke alapján közelítjük meg.

A konvergencia-sebesség fontos a határeloszlás tételek esetében, mert segítségével megadható egy n küszöb, úgy, hogy a legalább n elemű mintákra már elegendően jól simul a tapasztalati eloszlás a határeloszlás görbéjére. A fejezetben megnézzük ennek tulajdonságait, vele kapcsolatos tételeket, példát. Az extrémérték eloszlások konvergencia-sebességére nem születtek olyan eredmények, mint a CHT esetében a Berry-Esséen tétel, amely felső becslést ad a normális közelítés hibájára általános feltételek esetén.

2.1. Tétel. (Berry-Esséen) *Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók, továbbá $E|X_1|^3$ véges és $T_n := \left(\sum_{k=1}^n X_k - n \cdot m\right) / \left(\sigma \cdot \sqrt{n}\right)$. Ekkor*

$$\sup_x |F_{T_n}(x) - \Phi(x)| \leq C \cdot \frac{E|X_1 - m|^3}{\sigma \cdot \sqrt{n}},$$

ahol C egy olyan konstans, amelyet időről időre pontosítanak (2010-ben $C = 0,4785$).

Az extrémérték eloszlások konvergencia-sebességének vizsgálatánál az adott F eloszlás szélének viselkedését kell vizsgálnunk, hiszen a legkisebb és legnagyobb értékek az eloszlások szélén helyezkednek el. Tekintsük a $G = \Phi_\alpha$ esetet. Az 1.11 tételben láttuk, hogy az a_n, b_n normalizáló konstansok létezésének szükséges és elégséges feltétele az F szélének reguláris viselkedése $-\alpha$ indexszel. Ezzel ekvivalens felírás

$$\lim_{t \to \infty} \frac{\log F(tx)}{\log F(t)} = x^{-\alpha}, \quad x > 0. \tag{2.1}$$

Rendezzük át (1.6)-ot, amelyből az L lassú változású függvény a következőképpen felírható:

$$L(t) = -t^\alpha \log F(t), \quad t > 0. \tag{2.2}$$

A következő fogalom rendkívül fontos lesz a továbbiakban.

2.2. Definíció. Azt mondjuk, hogy az L *lassú változású függvény a g maradékfüggvénnyel*, ha

$$\frac{L(tx)}{L(t)} - 1 = O(g(t)), \quad x > 0, t \to \infty, \tag{2.3}$$

ahol g olyan függvény, hogy $g(t) \to 0$ $t \to \infty$ esetén.

A továbbiakban jelöljük a fentebb definiált fogalmat g-sv-vel, mely az angol megfelelőjének rövidítéséből származik (slowly varying with remainder function g). A konvergencia rendjéről viszonylag pontos eredményeket kapunk általa. A tételeinkben rendre megjelenik az elmélet, eredményeinkkel szoros párhuzamban fog állni.

A következő rövid példán keresztül láthatjuk, hogyha L g-sv függvény, akkor hogyan kapjuk meg a konvergencia-sebességet. Legyen $b_n = 0$ és a_n olyan sorozat, hogy

$$-\log F(a_n) \leq \frac{1}{n} \leq -\log F(a_n-). \tag{2.4}$$

Továbbá legyen

$$-n \log F(a_n) = 1 + O(r(n)), \tag{2.5}$$

ahol $r_n = g(a_n)$. Legyen $x > 0$. Ekkor felhasználva, hogy F reguláris változású:

$$\begin{aligned} -n \log F(a_n x) &= \frac{\log F(a_n x)}{\log F(a_n)} (1 + O(r_n)) \\ &= x^{-\alpha} \frac{L(a_n x)}{L(a_n)} (1 + O(r_n)) \\ &= x^{-\alpha} (1 + O(r_n)) . \end{aligned}$$

Ebből következik, hogy

$$F^n(a_n x) - \Phi_\alpha(x) = O(r_n). \tag{2.6}$$

Ez az összefüggés felvet több érdekes kérdést is, a továbbiakban ezekkel foglalkozunk.

2.1. Egyenletes konvergencia-sebesség

A kérdések közül elsőként azt vizsgáljuk meg, hogy a (2.6) hogyan viselkedik $0 < x < \infty$ esetén, ehhez először tekintsük a következő lemmát.

2.3. Lemma. *Legyen L g-sv függvény valamely pozitív g függvénnyel, amelyre*

$$\frac{g(tx)}{g(t)} < C, \quad x > 1, t \geq t_0, \tag{2.7}$$

C és t_0 pozitív konstansokkal. Ekkor (2.3) *egyenletesen konvergál* $(0, \infty)$*-n és léteznek olyan K, t_0 konstansok, amelyekre*

$$|\log L(tx) - \log L(t)| \leq K g(t)(1 + \log x), \quad x \geq 1, t \geq t_0. \tag{2.8}$$

Látni fogjuk, hogy ha az L g-sv függvény, és teljesíti a (2.7)-et, valamint

$$\frac{g(tx)}{g(t)} \geq Bx^{-\theta} \text{ minden } t \geq t_0 \text{ és } x \geq 1 \text{ esetén, ahol } \theta > 0, B > 0, \tag{2.9}$$

akkor az F egyenletesen tart az extrémérték határeloszláshoz $0 < x < \infty$ esetén. A g-re vonatkozó (2.7) és (2.9) feltételek enyhék, hiszen az első feltétel automatikusan teljesül g-re, ha az csökkenő függvény, a második kikötés pedig csupán annyit jelent, hogy g nem csökkenhet túl gyorsan. A következő konvergencia-sebességre vonatkozó tétel az egyenletességre vonatkozó feltételeket összefoglalja, amit be is bizonyítunk.

2.4. Tétel. *Tegyük fel, hogy $F(x) < 1$ minden véges x-re, és $-\log F$ reguláris változású függvény $-\alpha$ indexszel, ahol $\alpha > 0$. Továbbá a (2.2) szerint definiált L függvény legyen g-sv függvény valamely g pozitív függvénnyel, amely megfelel a (2.7) és (2.9) feltételeknek. Legyen $b_n = 0$ és a_n (2.4) szerint definiálva, valamint $r_n = g(a_n)$. Ekkor*

$$\sup_x |F^n(a_n x) - \Phi_\alpha(x)| = O(r_n), \quad n \to \infty. \tag{2.10}$$

Vagyis a hiba nagyságrendjét úgy határozzuk meg, hogy vesszük az eloszlásgörbék különbségének supremumát.

Bizonyítás. Ha F folytonos, $-n \log F(a_n) = 1$ minden n-re. Különben legyen ε_n pozitív számokból álló sorozat úgy, hogy $\varepsilon_n = O(r_n)$ $n \to \infty$ esetén. Ekkor

$$\begin{aligned} -n \log F(a_n) \leq 1 \leq -n \log F(a_n(1-\varepsilon_n)) \\ &= [-n \log F(a_n)](1-\varepsilon_n)^{-\alpha} \frac{L(a_n(1-\varepsilon_n))}{L(a_n)} \\ &= [-n \log F(a_n)](1 + O(r_n)). \end{aligned}$$

Tehát

$$-n \log F(a_n) = 1 + O(r_n). \tag{2.11}$$

Bizonyításunkhoz felhasználjuk (2.8)-t, valamint megfelelőjét $0 < x < 1$ esetén, ami a következő:

$$|\log L(tx) - \log L(t)| \leq Kg(tx)(1 + |\log x|), \quad tx \geq t_0. \tag{2.12}$$

Adott $\delta \geq 0$-hoz létezik $t_\delta > t_0$, úgy, hogy ha $x > 0$ és $t > t_\delta$, $tx > t_\delta$, akkor

$$|\log L(tx) - \log L(t)| \leq \delta(1 + |\log x|). \tag{2.13}$$

A $0 < x < 1$ esetben (2.9)-ből kifejeztük $g(tx)$-et és behelyettesítettük (2.12)-be, és $tx \geq t_0$ feltétel mellett

$$|\log L(tx) - \log L(t)| \leq \frac{K}{B} g(t) x^{-\theta}(1 + |\log x|) \tag{2.14}$$

teljesül.

Ebből újradefiniáljuk K-t, ami legyen K^*:

$$|\log L(tx) - \log L(t)| \leq K^* g(t) x^{-2\theta}. \tag{2.15}$$

A bizonyítás további részében tegyük fel, hogy $\delta < \dfrac{\alpha}{4}$.

Külön fogjuk igazolni az $x > 1$ és $0 < x < 1$ eseteket. Kezdjük az előbbivel, tehát legyen $x > 1$. A (2.8) és (2.13)

$$\begin{aligned} &|-\log(-\log F(a_n x)) + \log(-\log F(a_n)) - \alpha \log x| = \\ &|-\log L(a_n x) + \log L(a_n)| < (1 + \log x) \min(K r_n, \delta) \end{aligned}$$

feltéve, hogy $a_n > t_\delta$.

Felhasználva (2.11)-t és K^*-t, megkapjuk, hogy minden elég nagy n-re:

$$|-\log(-\log F(a_n x)) - \log n - \alpha \log x| \leq (1 + \log x) \min(K r_n, 2\delta). \tag{2.16}$$

Vezessük be a H függvényt, amelyet a következőképpen definiálunk:

$$H(y) = \exp\{-e^{-y}\}.$$

Ha $y_1 = -\log(-n \log F(a_n x))$ és $y_2 = \alpha \log x$, akkor ezeket H-ba helyettesítve:

$$\begin{aligned} H(y_1) &= F^n(a_n x), \\ H(y_2) &= \Phi_\alpha(x), \\ H(y_1) - H(y_2) &= (y_1 - y_2) H'(\xi), \end{aligned}$$

ahol $y_1 < \xi < y_2$ és $H' = \exp\{-y - \exp\{-y\}\}$ H első deriváltja. Legyen $x > e$. Ekkor (2.16) révén

$$\begin{aligned} \xi &> \alpha \log x - 2\delta(2 \log x) = (\alpha - 4\delta) \log x, \\ H'(\xi) &< H'((\alpha - 4\delta) \log x) \sim x^{-(\alpha - 4\delta)}, \quad x \to \infty. \end{aligned}$$

Így visszahelyettesítve $H(y_1) - H(y_2) = (y_1 - y_2) H'(\xi)$-be, $x \to \infty$ és $n \to \infty$ esetén

$$F^n(a_n x) - \Phi_\alpha(x) < |y_1 - y_2| x^{-(\alpha - 4\delta)} < K r_n (1 + \log x) x^{-(\alpha - 4\delta)}.$$

Így a jobb oldal $O(r_n)$, hiszen az x-től függő tényező korlátos, aminek következtében $|F^n(a_n x) - \Phi_\alpha(x)|$ konvergenciájának sebessége egyenletesen $O(r_n)$ minden $x > 1$ esetén.

Folytassuk a bizonyítást $0 < x < 1$-re. A (2.16)-hez hasonlóan

$$|-\log(-\log F(a_n x)) - \log n - \alpha \log x| < \min\left(2\delta(1 + |\log x|), K r_n x^{-2\theta}\right), \tag{2.17}$$

ahol $a_n x \geq t_\delta$.

Legyen ismét y_1, y_2 és ξ az $x > 1$ esethez hasonlóan definiálva. Ha $x < \dfrac{1}{e}$, akkor $\xi < -(\alpha - 4\delta)|\log x|$, és

$$H'(\xi) < H'(-(\alpha - 4\delta)) = x^{-(\alpha - 4\delta)} \exp\left\{-x^{-(\alpha - 4\delta)}\right\},$$

és így $x \to \infty$ és $n \to \infty$ esetén, $a_n x \geq t_\delta$ mellett

$$|F^n(a_n x) - \Phi_\alpha(x)| \leq K r_n x^{-2\theta-\alpha+4\delta} \exp\{-x^{-(\alpha-4\delta)}\}. \tag{2.18}$$

Ez, ha $x \to 0$ valójában $O(r_n)$, tehát az állítást beláttuk $x > \frac{t_\delta}{a_n}$-ra. Elegendő igazolnunk, hogy $F^n(t_\delta)$ és $\Phi_\alpha(\frac{t_\delta}{a_n})$ is $o(r_n)$. A fentebbi feltételek F-re és g-re biztosítják, hogy r_n legfeljebb olyan gyorsan csökkenhet, mint n valahányadik hatványa. Mivel $F(t_\delta) < 1$, $F^n(t_\delta) = o(r_n)$ nyilvánvaló (hiszen feltettük, hogy r_n nem csökkenhet túl gyorsan). Másrészt, $a_n \geq n^{\frac{1}{2}\alpha}$, így

$$\Phi_\alpha\left(\frac{t_\delta}{a_n}\right) \leq \Phi_\alpha\left(\frac{t_\delta}{n^{\frac{1}{2}\alpha}}\right) = \exp\left\{-\frac{n^{\frac{1}{2}}}{t_\delta^\alpha}\right\},$$

ami szintén $o(r_n)$, tehát beláttuk az állítást. □

Érdekes kérdés lehet még, hogy az a_n, melyet (2.4) szerint definiáltunk, milyen széles intervallumon választható, hogy a tétel érvényes maradjon, vagyis teljesüljön az egyenletes konvergencia $0 < x < \infty$ intervallumon. Ahhoz, hogy a konvergencia-sebesség $O(r_n)$ legyen, a_n-re a (2.5) szükséges és elégséges feltétel.

2.2. A becslés pontosítása

Ebben a részben megvizsgáljuk, hogy meg tudunk-e határozni egy fő hibatagot a közelítésben. A 2.4 tételben a hiba nagyságának rendjéről kaptunk információt, ami $O(r_n)$, a gyakorlatban viszont sokszor van szükség annak függvényes formájára. Legyen az L és g függvény a $(0, \infty)$ intervallumon értelmezve, és g legyen olyan függvény, hogy $g(t) \to 0$, ha $t \to \infty$. Ekkor

$$\frac{L(tx)}{L(t)} - 1 \sim v(x)g(t), \quad x > 0, t \to \infty. \tag{2.19}$$

A cikk szerint, hogy ha (2.19) fennáll, és v teljesíti a feltételt, hogy létezik olyan x, hogy $v(x) \neq 0$, és minden y-ra teljesül, hogy $v(xy) - v(y) \neq 0$, akkor g reguláris változású függvény p indexszel, ahol $p \leq 0$ és $v(x) = ch_p(x)$ valamely c konstansra és h_p-re, ahol $h_p(x) = \int_1^x u^{p-1} du$.

2.5. Tétel. *Legyen L, F, a_n és b_n a (2.4) tételben látottakhoz hasonlóan definiálva. Tegyük fel, hogy L (2.19) szerinti g-sv függvény, ahol g reguláris változású függvény p indexszel, és $v(x) = ch_p(x)$ $p \leq 0$ esetén. Ekkor $0 < x < \infty$ esetén*

$$F^n(a_n x) - \Phi_\alpha(x) = -c r_n h_p(x) x^{-\alpha} \Phi_\alpha(x) + o(r_n). \tag{2.20}$$

2.6. Következmény. Ha $p < 0$, akkor

$$F^n(a_n x) - \Phi_\alpha(x) = -c r_n x^{-\alpha} \Phi_\alpha(x)(h_p(x) + o(1))$$

egyenletesen $x > 1$, ahogy $n \to \infty$.

Ez az eredmény jelentős a gyakorlatban, tipikusan a legfontosabb kérdés a konvergencia az eloszlások felső szélén.

2.7. Következmény. Tegyük fel, hogy létezik $f(x) = F'(x)$ és legyen

$$k(x) = -\frac{f(x)}{F(x)\log F(x)}.$$

Tegyül fel még, hogy

$$\lim_{x\to\infty} xk(x) = \alpha,$$

amely ekvivalens von Mises szükséges, de nem elégséges feltételével az $F^n(a_n(x)) \to \Phi_\alpha(x)$-ra vonatkozólag. Továbbá teljesüljön, hogy

$$\lim_{y\to\infty} \frac{xyk(xy) - \alpha}{yk(y) - \alpha} = x^p, \quad x > 0.$$

Legyen a_n olyan, hogy $F(a_n) = \exp\{-\frac{1}{n}\}$. Ekkor

$$F^n(a_n x) = \Phi_\alpha(x) + \frac{1}{\alpha}(a_n k(a_n) - \alpha)xh_p(x)\phi_\alpha(x) + o(a_n k(a_n) - \alpha)$$

konvergencia $0 < x < \infty$ esetén egyenletes lesz, ahol $\phi_\alpha(x) = \Phi'_\alpha(x)$.

2.8. Példa. Most nézzünk meg egy példát. Tekintsük az F eloszlásfüggvényt olyannak, hogy

$$F(x) = 1 - cx^{-\alpha} - dx^{-\alpha-\beta} + o(x^{-\alpha-\beta}), \quad x \to \infty,$$

ahol $c > 0$ és $d \neq 0$ konstansok, és $0 < \beta < \alpha$. Legyen $L(x)$ a (2.2)-nél látottakhoz hasonlóan definiált lassú változású függvény. Tudjuk, hogy az $1-F(x) = \overline{F}(x)$ jó közelítése a $\log F(x)$, tehát $1 - F(x) \sim \log F(x)$. Ezt felhasználva

$$\begin{aligned}
\frac{L(tx)}{L(t)} &= \frac{-(tx)^\alpha\big(1 - \big(1 - c(tx)^{-\alpha} - d(tx)^{-\alpha-\beta} + o((tx)^{-\alpha-\beta})\big)\big)}{-(t)^\alpha\big(1 - (1 - c(t)^{-\alpha} - d(t)^{-\alpha-\beta} + o((t)^{-\alpha-\beta}))\big)} \\
&= \frac{1 + d(tx)^{-\beta}/c + o((tx)^{-\beta})}{1 + dt^{-\beta}/c + o(t^{-\beta})} \\
&= \frac{1 + dt^{-\beta}/c + 1 + d(tx)^{-\beta}/c - (1 + dt^{-\beta}/c) + o((tx)^{-\alpha-\beta})}{1 + dt^{-\beta}/c + o(t^{-\beta})} \\
&= 1 - dt^{-\beta}\frac{(1 - x^{-\beta})}{c} + o(t^{-\beta}),
\end{aligned}$$

és ez éppen a (2.19)-ben fennálló eset, ahol

$$g(t) = t^{-\beta} \quad \text{és} \quad v(x) = -d(1 - x^{-\beta})/c.$$

Az a_n sorozat pontos értékét ugyan nehéz megadni, de láttuk, hogy $\Phi_\alpha(x)$ esetén $a_n = n^{\frac{1}{\alpha}}$ és $b_n = 0$. Tehát legyen $a_n \sim (nc)^{\frac{1}{\alpha}}$, és így $g(a_n) \sim (nc)^{-\frac{\beta}{\alpha}}$ $n \to \infty$ esetén. Alkalmazzuk a 2.5 tételt, és megkapjuk, hogy

$$F^n(a_n(x)) = \Phi_\alpha(x) + n^{-\frac{\beta}{\alpha}}c^{-1-\frac{\beta}{\alpha}}dx^{-\alpha}\left(1 - x^{-\beta}\right)\Phi_\alpha(x) + o\left(n^{-\frac{\beta}{\alpha}}\right), \quad 0 < x < \infty.$$

A példa, illetve az ott látott módszer alkalmazható a Pareto- és a stabilis eloszlások maximuma konvergencia-sebességének becslésére (a Paretonál $d = 0$).

2.3. Fordított eredmények

Korábban már láttuk, hogy az a_n konstansoknak milyen intervallumba kell esniük, hogy az egyenletes sebesség érvényes maradjon az extrémérték eloszlásokhoz, most pedig ennél erősebb feltételeket keresünk. Feltehetjük a kérdést, hogy vajon az a_n és b_n konstansok optimálisak-e, azaz ezen normalizáló sorozatok mellett a leggyorsabb-e a konvergencia sebessége az extrém érték eloszlásokhoz. Amint látni fogjuk, ez nem mindig igaz.

Másik kérdés is feltehető, méghozzá az, hogy mi történik akkor, ha a 2.4 tételünkben módosított normalizáló sorozatokat vizsgálunk, vajon akkor is érvényes-e az egyenletes sebesség, szükséges-e a g-sv tulajdonság. Mivel ez a két kérdés szorosan összefügg, így ezeket a részeket együtt vizsgáljuk meg.

Legyenek $a'_n > 0$, b'_n és $r'_n > 0$ olyan módosított sorozatok, hogy $r'_n \to 0$ és

$$F^n(a'_n x + b'_n) = \Phi_\alpha(x)(1 + O(r'_n)) \tag{2.21}$$

egyenletesen $x \in (0, \infty)$ valamely véges halmazán.

L legyen továbbra is (2.2) szerinti lassú változású függvény, és

$$\frac{a'_n}{a_n} \to 1 \text{ és } \frac{b'_n}{a_n} \to 0, \tag{2.22}$$

ahol a_n (2.4) szerint van definiálva.

2.9. Tétel. *L (2.3) szerinti g-sv függvény olyan g függvénnyel, hogy $g(a'_n) = r_n$, ha $r_n = \max(r'_n, |b'_n/a'_n|)$.*

Bizonyítás. F reguláris változású tulajdonsága miatt (2.21) átírható

$$n\left(a'_n + \frac{b'_n}{x}\right)^{-\alpha} L(a'_n x + b'_n) = 1 + O(r'_n). \tag{2.23}$$

Ekkor a konvergencia-sebesség az extrém érték eloszlásokhoz egyenletes véges x intervallumokon. Rögzítsük $x_0 > 1$-et, ekkor a konvergencia-sebesség egyenletes lesz a $\frac{1}{2x_0^2} \le x \le 2x_0$ esetén. Adott $t > 0$-hoz legyen $n(t)$ a legkisebb olyan egész szám, amelyre $a_{n(t)} \ge t$. Az $n(t)$ jól definiált, mert tudjuk, hogy az $a'_n \to \infty$, ami következik (2.22)-ból, és $\dfrac{a'_{n+1}}{a'_n} \to 1$. Továbbá igaz $n(t)$-re, hogy teljesíti az

$$a'_{n(t)} < x_0 t, \text{ minden elég nagy } t \text{ esetén.} \tag{2.24}$$

Ehhez hasonlóan (2.22)-ból és (2.24)-ból $\dfrac{b'_{n(t)}}{t} \to 0$ $t \to \infty$ esetén is teljesül.

Rögzítsük t-t, és teljesüljön rá a (2.24) és $|\frac{b'_{n(t)}}{t}| < \frac{1}{2x_0}$. Adott y legyen olyan, hogy $\frac{1}{x_0} < y < x_0$, és legyen $x(y)$ olyan, hogy $a'_{n(t)}x(y) + b'_{n(t)} = ty$. Hogyha az $x(y) < \frac{1}{2x_0^2}$, akkor behelyettesítve az előzőket azt az ellentmondást kapjuk, hogy

$$y < \frac{x_0}{2x_0^2} + \frac{1}{2x_0} = \frac{1}{x_0},$$

hiszen $\frac{1}{x_0} < y < x_0$ nem teljesül. Így innentől fogva az $\frac{1}{2x_0^2} \leq x \leq 2x_0$ feltétel mellett folytatjuk az igazolást. Alkalmazzuk a (2.23)-t $x(y)$-ra és $x(1)$-re:

$$\begin{aligned}
\frac{L(ty)}{L(t)} &= \left(a'_{n(t)} + \frac{b'_{n(t)}}{x(y)}\right)^{\alpha} \left(a'_{n(t)} + \frac{b'_{n(t)}}{x(1)}\right)^{-\alpha} \left(1 + O(r'_{n(t)})\right) \\
&= \left(\frac{yx(1)}{x(y)}\right)^{\alpha} \left(1 + O(r'_{n(t)})\right) \\
&= \left(\frac{ty - b'_{n(t)}y}{ty - b'_{n(t)}}\right)^{\alpha} \left(1 + O(r'_{n(t)})\right) \\
&= 1 - \alpha \left(\frac{b'_{n(t)}}{t}\right)\left(1 - \frac{1}{y}\right) + o\left(\frac{b'_{n(t)}}{t}\right) + O\left(r'_{n(t)}\right).
\end{aligned}$$

Látható, hogy a konvergencia-sebesség egyenletes $\frac{1}{x_0} < y < x_0$ esetén, és mivel x_0 tetszőlegesen választható, így gyakorlatilag bármely véges y intervallumon egyenletes a sebesség. Írjuk át a kapott összefüggést egy vele ekvivalens alakba $g_1(t) = -\alpha\frac{b_{n(t)}}{t}$ és $g_2(t) = r'_{n(t)}$ definiálásával:

$$\frac{L(ty)}{L(t)} - 1 = g_1(t)(1 - \tfrac{1}{y}) + o(g_1(t)) + O(g_2(t)). \tag{2.25}$$

Legyen $g(t) = r_n(t)$; az $r_n = \max(r'_n, b'_n/a'_n)$ definíciójából és (2.24)-ből következik, hogy $g_1(t)$ és $g_2(t)$ nagyságrendje $O(g(t))$, $t \to \infty$ esetén, így innentől L teljesíti (2.3)-t g-vel.

Hogyha $\frac{b'_n}{a'_n r'_n}$ korlátos, akkor a 2.9 tételből következik, hogy L g-sv függvény olyan g-vel, hogy $g(a'_n) = r'_n$. Ha g-re teljesülnek a 2.4 tétel feltételei, vagyis hogy g csökkenő függvény, de nem gyorsan csökkenő, akkor láthatjuk, hogy ha a'_n és b'_n helyébe a_n-t és $b_n = 0$-át helyettesítünk, akkor a konvergencia sebessége $O(g(a_n))$. Az előbb említett g-re vonatkozó (2.7), (2.9) és (2.22) feltételek mellett ez megegyezik az $O(r_n)$-nel vagy $O(r'_n)$-vel, vagyis a konvergencia sebessége nem javul az új, módosított sorozatok használatával.

Ám felmerül a kérdés, mi történik, ha $\frac{b'_n}{a'_n r'_n}$ nem korlátos. Akkor látni fogjuk, hogy az új sorozatok javítani fognak a_n és b_n konstansokon, csak speciálisabb feltételek mellett.

Tegyük fel, hogy $\frac{r'_n}{r_n} \to 0$. Ekkor $\frac{|b'_n|}{a'_n r'_n} \to \infty$, így $|\frac{g_2(t)}{g_1(t)}| \to 0$ ha $t \to \infty$, és így (2.25) ekvivalens

$$\frac{L(ty)}{L(t)} - 1 = g_1(t)(1 - \tfrac{1}{y}) + o(g_1(t)), \quad y > 0, \tag{2.26}$$

ami gyakorlatilag a (2.19), vagyis a g_1 reguláris változású függvény -1 indexszel. □

2.10. Tétel. *Az olyan a'_n és b'_n konstansok létezésének, amelyekkel a 2.9 tétel teljesül az $\frac{r'_n}{r_n} \to 0$ feltétel mellett, szükséges és elégséges feltétele, hogy L g-sv függvény* (2.19) *szerinti alakban, ahol g valamilyen reguláris változású függvény -1 indexszel. Továbbá*

$$v(y) = c(1 - \tfrac{1}{y}), \quad y > 0, c \neq 0.$$

Hogyha ez teljesül, az a'_n és b'_n konstansokat választhatjuk a következőképpen:

$$a'_n + b'_n = a_n,$$
$$b'_n = -\frac{1}{\alpha} c a_n g(a_n).$$

Azaz ezen feltételek mellett a konvergenciánk sebessége javulni fog.
Bizonyítás. Először lássuk be a szükségességet. Ha $g = g_1$ és $c = 1$, akkor éppen (2.26)-t kapjuk, tehát az állítást beláttuk. Most pedig nézzük az elégségességet. Használjuk a g-sv függvények második, (2.19)-ben definiált felírását, és legyen $r_n = g(a_n)$. Felhasználva (2.11)-et, a következőt kapjuk:

$$\begin{aligned}
-n \log F(a'_n x + b'_n) &= \frac{\log F(a'_n x + b'_n)}{\log F(a'_n + b'_n)} (1 + o(r_n)) \\
&= \left(\frac{a'_n x + b'_n}{a'_n + b'_n} \right)^{-\alpha} \frac{L(a'_n x + b'_n)}{L(a'_n + b'_n)} (1 + o(r_n)) \\
&= x^{-\alpha} \left(1 + \left(\frac{\alpha b'_n}{a'_n} + r_n c \right) (1 - 1/x) + o(r_n) \right) \\
&= x^{-\alpha} \left(1 + r_n c (1 - 1/x)(1 - \frac{a_n}{a'_n}) + o(r_n) \right) \\
&= x^{-\alpha} (1 + o(r_n)).
\end{aligned}$$

Így

$$n(a'_n x + b'_n)^{-\alpha} L(a'_n x + b'_n) = 1 + o(r'_n),$$

amely pontosan (2.23) $O(r_n)$ helyett $o(r'_n)$-nel, amiből látható, hogy javult a konvergencia. □

Anderson (1971) némileg erősebb feltételek mellett szintén belátta a szükségességet.

$$\begin{aligned}
F^n(a'_n x - b'_n) - \Phi_\alpha(x) = &\left(\alpha \left(\frac{a'_n}{a_n} - 1 \right) + \alpha \frac{b'_n}{a_n x} - c r_n h_p x \right) x^{-\alpha} \Phi_\alpha(x) \\
&+ o\left(\max \left| \frac{a'_n}{a_n} - 1 \right|, \left| \frac{b'_n}{a_n} \right|, |r_n| \right).
\end{aligned}$$

Ez levezethető a 2.5 tételből, és egyenletes x-ben, ha $x_n = a'_n x_n + b'_n$ és a 2.5 tételt alkalmazzuk x_n-re.

Az egyenlet jobb oldala akkor és csak akkor $o(r_n)$, ha $p = -1$, ekkor ha a tételben kimondott módon választjuk az a'_n és b'_n konstansokat, a fő tag eltűnik. Azonban ezen konstansok alkalmazásával a fő tag bámely p esetén lokálisan minimalizálva van $x = 1$ környékén. Így, ha (2.19) fentáll, lehetséges az a_n és b_n sorozatok helyett ezek módosításait használni még akkor is, ha $p \neq 1$.

2.4. Konvergencia Ψ_α-hoz

Az 1.12 tételben láttuk, hogy milyen feltételek mellett tart egy eloszlásfüggvény a Weibull-határeloszláshoz. Egyik feltétele az volt, hogy az F eloszlásfüggvénynek egy véges x_F végponttal kell rendelkeznie, a másik pedig az, hogy az $F_1(x) = F(x_F - \frac{1}{x})$ szerint definiált $F_1(x)$-nek a $\Phi_\alpha(x)$ vonzási tartományában kell lennie (lsd. (1.7)). Ebben a részben vizsgáljuk meg röviden egy F eloszlásfüggvény konvergenciájának sebességét a Ψ_α-hoz, hasonlóan az előző részekhez.

Legyenek a_n és b_n konstansok olyanok, amely mellett

$$\lim_{n\to\infty} F^n(a_n x + b_n) = \Psi_\alpha(x).$$

Legyen $c_n = \frac{1}{a_n}$ és $\varepsilon_n = c_n(x_F - b_n)$. A vonzási tartományok karakterizációjánál szintén láttuk, hogy a Weibull és Fréchet határeloszlások között szoros kapcsolat van: legyen x és y olyan, hogy $y = -\frac{1}{x}$. Ekkor

$$F(a_n x + b_n) = F_1\left(\frac{c_n y}{1 + \varepsilon_n y}\right),$$

sőt $\Psi_\alpha(x) = \Phi_\alpha(y)$. Ha $b_n = x_F$, azaz $\varepsilon_n = 0$, akkor a Fréchetnél látott 2.4 és 2.5 tételt közvetlenül alkalmazhatjuk mostani esetünkben is.

2.11. Tétel. *Tegyül fel, hogy F_1 teljesíti a 2.4 tétel hipotézisét. Legyen $b_n = x_F$ és a_n olyan, amelyre teljesül, hogy*

$$-\log F(x_F - a_n) \le \frac{1}{n} \le -\log F(x_F - a_n-).$$

Legyen $r_n = g(\frac{1}{a_n})$. Ekkor

$$\sup_x |F^n(a_n x + b_n) - \Psi_\alpha(x)| = O(r_n).$$

2.12. Tétel. *Tegyük fel, hogy F_1 teljesíti a 2.5 tétel hipotézisét. Ekkor*

$$F^n(a_n x + b_n) - \Psi_\alpha(x) = c r_n h_{-p}(-x)(-x)^\alpha \Psi_\alpha(x) + o(r_n)$$

egyenletesen $-\infty < x < 0$ esetén.

A 2.12 tételből látható a $h_p(\frac{1}{y}) = -h_{-p}(y)$ összefüggés.

2.13. Példa. Az egyenletes eloszlás a Weibull eloszlás vonzási tartományába tartozik, így tekintsük az $F(x) = \dfrac{x-a}{b-a}$ eloszlásfüggvényt. Értelmezzük F-et a $(0,1)$ intervallumon, tehát $a = 0$, $b = x_F = 1$. Az (1.7)-ből $L(t) = -x^\alpha \overline{F}(x_F - \frac{1}{x})$. Így

$$\begin{aligned}
\frac{L(tx)}{L(t)} &= \frac{-(tx)^\alpha(1 - F(x_F - \frac{1}{tx})) + o((tx)^{-1})}{-t^\alpha(1 - F(x_F - \frac{1}{t})) + o(t^{-1})} \\
&= x^\alpha \frac{(tx)^{-1} + o((tx)^{-1})}{t^{-1} + o(t^{-1})} \\
&= x^\alpha \frac{t^{-1} + (tx)^{-1} - t^{-1} + o((tx)^{-1})}{t^{-1} + o(t^{-1})} \\
&= 1 - t^{-1}(1 - x^{-1}) + o(t^{-1}).
\end{aligned}$$

Így legyen

$$g(t) = t^{-1}, \quad v(x) = -(1 - x^{-1}).$$

Az $r_n = g(\frac{1}{a_n})$. Válasszuk $a_n = n^{-1}$, így $g(\frac{1}{a_n}) = n^{-1}$ és $b_n = 1$. Így

$$F^n(a_n x + b_n) = \Psi_\alpha(x) + n^{-1}(-x)^\alpha(1 - x^{-1})\Psi_\alpha(x) + o(n^{-1}).$$

A Weibull határeloszlás esetén az a'_n és b'_n módosító sorozatok esetén nem lesz gyorsabb a konvergencia, így azokat most nem ismertetjük.

3. fejezet

Kitekintés

3.1. Küszöbmeghaladás

A konvergencia-sebességről szóló elméleti összefoglalónkban nagy hasznát láttuk a lassú változású függvények g-sv tulajdonságának, tételeinkben és azok bizonyításában rendre fontos szerepet kaptak. Most vizsgáljuk meg, miként alkalmazható mindez a gyakorlatban, jelen esetben az eloszlások szélének vizsgálatában. Ezen fejezet alapjául [6] cikk szolgál, amely az eloszlásfüggvények szélének becslésével, azok aszimptotikus tulajdonságával foglalkozik, egy küszöb meghaladásán alapulva. Amiért számunkra érdekes, hogy több eredményt is a g-sv tulajdonság alkalmazásával kapott meg. Másrészt az eredmények lefedik az extrém érték elmélet mindhárom határeloszlás típusát is, amelyek közül a Fréchet esetet nézzük meg. Először tekintsük a vizsgált rész elméleti hátterét röviden.

Előző fejezetünk határeloszlás tételét a gyakorlatban úgy lehet használni, hogy az egyes blokkokból (ilyen blokkok lehetnek jellemzően fél éves vagy éves blokkok) kiválasztjuk a maximumot, és azt már extrém-érték eloszlásúnak tételezzük fel. Ezt a módszert blokk-maximum módszernek hívjuk. Azonban ebben a részben már nem a blokk-maximum módszert alkalmazzuk, az F eloszlásfüggvény szélének vizsgálata egy küszöb túllépésén fog alapulni. A blokk-maximum módszer egyik hátránya, hogy az - amúgy is kevés rendelkezésre álló - extrém adatok közül csak a maximumot használja fel, így viszont sok adat elvész. A küszöbmeghaladáson alapuló modellek során előre megadunk egy magas küszöböt, és amely események ezt túllépik, azokat fogjuk extrém értékeknek hívni. Ez formálisan a következő: legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók közös F eloszlásfüggvénnyel. Tekintsük extrém eseményeknek azokat, amelyek egy magas, tetszőlegesen megválasztott u küszöbértéket túllépnek, és legyen $x_F < \infty$ F felső végpontja. Ekkor Pickands (1975) eredményei alapján

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad u < x_F,\ 0 < y < x_F - u \tag{3.1}$$

a feltételes eloszlásfüggvénye az $X - u$-nak, ha $X > u$. Definiáljuk az általánosított Pareto

eloszlást (Generalized Pareto Distribution, GPD):

$$G(y;\sigma,k) = \begin{cases} 1-(1-\frac{ky}{\sigma})^{\frac{1}{k}}, & k \neq 0, \sigma > 0, \\ 1-\exp\{-\frac{y}{\sigma}\}, & k = 0, \sigma > 0, \end{cases} \tag{3.2}$$

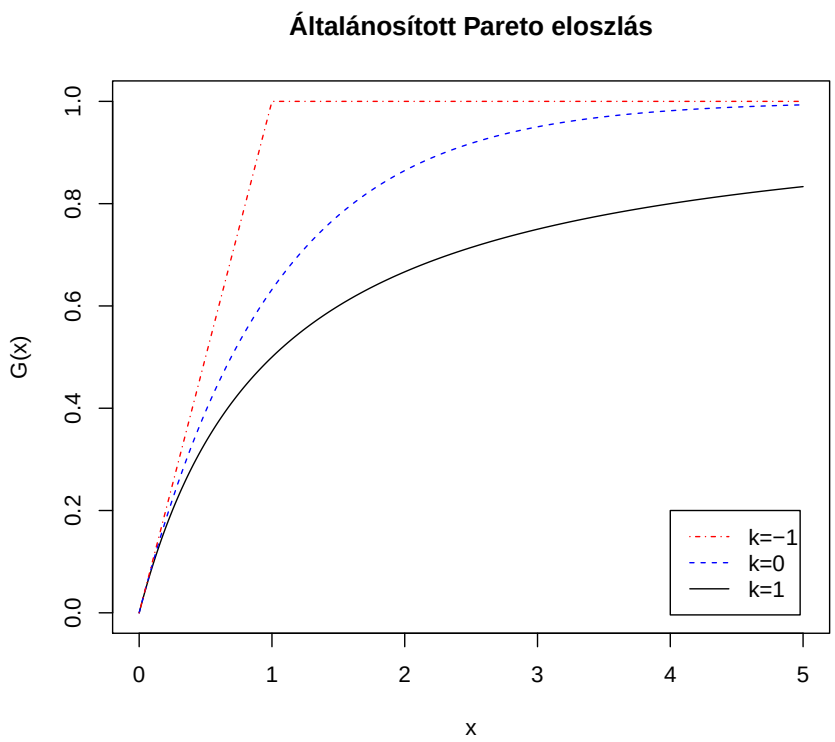
ahol

$$\begin{aligned} 0 < y < \infty, &\quad \text{ha } k \leq 0, \\ 0 < y < \frac{\sigma}{k}, &\quad \text{ha } k > 0. \end{aligned}$$

Ha az u küszöbszint tart az eloszlásfüggvény jobb oldali végpontjához, akkor a GPD jó közelítése lesz az u szint feletti maximumok eloszlásának, vagyis

$$\lim_{u \to x_F} \sup_{0<y<x_F-u} |F_u(y) - G(y;\sigma,k)| = 0 \tag{3.3}$$

rögzített k-ra és $\sigma(u)$-ra akkor és csak akkor, ha az F valamely H extrémérték eloszlás maximum vonzási tartományában van. Hogyha $H = \Phi_\alpha$, akkor $k = -\frac{1}{\alpha}$, ha $H = \Psi_\alpha$, akkor $k = \frac{1}{\alpha}$, és $H = \Lambda_\alpha$ esetén a $k = 0$. Így megkaphatunk egy becslést az F eloszlás felső széléről, ami konzisztens mindhárom maximum vonzási tartománnyal.



3.1. ábra. Általánosított Pareto eloszlás különböző alakparaméterekkel

3.2. Becslés GPD-vel

Legyen $X_1, \ldots, X_n$ független, azonos eloszlású valószínűségi változók közös F eloszlásfüggvénnyel. Rögzítsünk egy magas u küszöböt. Legyen $Y_1, \ldots, Y_N$ az extrém értékek, vagyis azok a változók, amelyek túlléptek u-n, és N jelzi a küszöbmeghaladások számát. Ekkor $Y_i = X_j - u$, ahol az i. túllépés indexe j. Ekkor a küszöbmeghaladások független, azonos eloszlású valószínűségi változók közös F_u eloszlásfüggvénnyel. Az F_u-t szeretnénk becsülni a GPD-vel. Egy u küszöb esetén maximum likelihood paraméterbecslést végzünk $G(y; \sigma, k)$-ra, és ezekkel a $\hat{\sigma}$ és $\hat{k}$ becsült paraméterekkel közelítjük F_u-t.

Legyen

$$g(y; \sigma, k) = \frac{\partial}{\partial y} G(y; \sigma, k)$$

a GPD sűrűségfüggvénye. Tekintsünk $Y_1, \ldots, Y_N$ független, azonos eloszlású valószínűségi változókat a pontos GPD sűrűséggel. Jelölje

$$L_N(\sigma, k) = \sum_{i=1}^{N} \log g(Y_i, \sigma, k)$$

a log-likelihood függvényt. Legyen (σ, k)-ra $U_N(\sigma, k) = \left[\frac{-\sigma \partial L_N}{\partial \sigma}, \frac{-\partial L_N}{\partial k}\right]$ negatív gradiens vektor és legyen $I_N(\sigma, k)$ 2×2 információs mátrix olyan, hogy diagonális elemei: $\frac{-\sigma^2 \partial L_N}{\partial \sigma^2}$, $\frac{-\partial^2 L_N}{\partial k^2}$, és az off-diagonális elemei: $\frac{-\sigma \partial L_N}{\partial \sigma \partial k}$. Ekkor $k < \frac{1}{2}$ esetén $EU_N = 0$, és

$$\frac{EU_N U_N^T}{N} = \frac{EI_N}{N} = M,$$

ahol

$$M = \frac{1}{(1-2k)(1-k)} \begin{bmatrix} 1-k & -1 \\ -1 & 2 \end{bmatrix}, \quad M^{-1} = (1-k) \begin{bmatrix} 2 & 1 \\ 1 & 1-k \end{bmatrix}.$$

Ekkor a $(\hat{\sigma}_N, \hat{k}_N)$ lokális maximumokra teljesül, hogy

$$\sqrt{N} \begin{bmatrix} \frac{\hat{\sigma}_N}{\sigma} - 1 \\ \frac{\hat{k}_N}{k} - 1 \end{bmatrix} \to \mathcal{N}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, M^{-1}\right], \quad k < \frac{1}{2}. \tag{3.4}$$

Az $\mathcal{N}$ a kétváltozós normális eloszlást jelöli.

Az $Y_1, \ldots, Y_N$ legyen most az F_u-ból meghatározva. Ha (3.3) tétel teljesül, akkor meglehet határozni egy olyan ϕ maradékfüggvényt, amelyre

$$\lim_{u \to x_F} \sup_{0 < y < x_F - u} |F_u(y) - G(y; \sigma, k)| = O(\phi(u)), \tag{3.5}$$

ahol $\phi(u) \to 0$ $u \to x_F$ esetén.

Precízebb eredményekért tegyük fel, hogy $N \to \infty$, $u = u_N \to x_F$ és $\sigma = \sigma_N = \sigma(u_N)$. Ekkor

$$\sqrt{N} \begin{bmatrix} \frac{\hat{\sigma}_N}{\sigma_N} - 1 \\ \hat{k}_N - k \end{bmatrix} \to \mathcal{N}[-M^{-1}\mathbf{b}, M^{-1}],$$

ahol

$$\frac{1}{\sqrt{N}} EU_N(\sigma_N, k) \to \mathbf{b}, \quad \text{ahol } \mathbf{b} \text{ véges vektor },$$
$$NI_N^{-1}(\sigma', k') \xrightarrow{p} M^{-1}, \quad \text{egyenletesen}(\sigma', k')\text{-n}.$$

Nézzük meg a Fréchet-határeloszlásra vonatkozó eredményeket részletesebben. Legyen F eloszlásfüggvény a Φ_α vonzási tartományában. Ekkor az (1.6)-ból $L(t) = t^\alpha \overline{F}(t)$.

3.1. Állítás. *Legyen L lassú változású függvény ϕ maradékfüggvénnyel. Tegyük fel, hogy L teljesíti a (2.3) vagy a (2.19) feltételt, g helyett most ϕ-re. Legyen w egy valós értékű függvény $(1, \infty)$-n. Tegyük fel valamely $p \leq 0$-val, hogy vagy (2.19) teljesül ϕ reguláris változású függvénnyel, melynek indexe p, vagy (2.3) teljesül nem növekvő ϕ-vel, és*

$$\frac{\phi(tx)}{\phi(t)} \leq Cx^p, \quad t \geq t_0 \geq 0, x > 1, C < \infty.$$

Ha $p = 0$ és $\int_1^\infty |w(x)| x^\varepsilon dx$ véges valamely $\varepsilon > 0$ esetén, vagy $p < 0$ és w integrálható, akkor

$$\int_1^\infty w(x) \frac{L(tx)}{L(t)} dx = \int_1^\infty w(x) dx + O(\phi(t)).$$

Ha 2.19 teljesül, akkor

$$\int_1^\infty w(x) \frac{L(tx)}{L(t)} dx = \int_1^\infty w(x) dx + \phi(t) \int_1^\infty w(x) v(x) dx + o(\phi(t)),$$

ahol v a 2.19-ben meghatározott függvény. Most már megfogalmazhatjuk fő eredményünket a $\Phi_\alpha(x)$ esetén.

3.2. Tétel. *Legyen L (2.19) szerint meghatározott lassú változású függvény ϕ maradékfüggvénnyel. Legyen $Y_1, \ldots, Y_N$ független, azonos eloszlású valószínűségi változók közös F_{u_N} eloszlásfüggvénnyel, ahol $N \to \infty$, $u_N \to \infty$ olyan, hogy*

$$\frac{\sqrt{N} c \phi(u_N)}{\alpha - p} \to \mu, \quad -\infty < \mu < \infty.$$

Legyen $k = -\frac{1}{\alpha}$, $\sigma_N = \frac{u_N}{\alpha}$. Ekkor a GPD paramétereire felírt likelihood függvénynek létezik egy $(\hat{\sigma}_N, \hat{k}_N)$ lokális maximuma, amelyre 1 valószínűséggel

$$\sqrt{N} \begin{bmatrix} \frac{\hat{\sigma}_N}{\sigma_N} - 1 \\ \hat{k}_N - k \end{bmatrix} \to \mathcal{N} \left[\begin{bmatrix} \dfrac{\mu(1-k)(1+2kp)}{1-k+kp} \\ \dfrac{\mu(1-k)k(1+p)}{1-k+kp} \end{bmatrix}, \quad M^{-1} \right].$$

Ha L (2.3) szerint meghatározott lassú változású függvény ϕ nem növő maradékfüggvénnyel, és $\sqrt{N}\phi(u_N) \to 0$, akkor az eredmény ugyanaz marad, de $\mu = 0$ lesz, azaz teljesül az aszimptotikus torzítatlanság.

4. fejezet

Szimulációk

Érdekes lehet szimulációkon keresztül megvizsgálni a mintamaximumok konvergenciáját a határeloszlásokhoz. A szimulációkat az R statisztikai programcsomaggal készítettem, 3 különböző eloszlásra: normális, exponenciális illetve egyenletes eloszlásokra. A szimulációk készítésének menete a következőképpen zajlott: először generáltam egy adott eloszlású, n elemű véletlen mintát, melyből az elemek sorbarendezése után kiválasztottam a legnagyobb elemet, a maximumot. Ezt a lépést kellően sokszor megismételtem, és a kiválasztott legnagyobb elemek adták az n elemű minta maximumainak eloszlását, M_n-t. Ezt az M_n-t a megfelelő a_n és b_n sorozatokkal normalizáltam, és különböző módszerekkel szemléltettem a konvergencia sebességét.

Először ábrázoltam a megfelelően normalizált, adott eloszlású minta maximumainak tapasztalati eloszlásfüggvényeit az adott n mintanagyságokra a határeloszlásukkal. Ez rendkívül szemléletes, hiszen látható, hogy a görbék a különböző n-ek esetén milyen szépen simulnak a határeloszlás görbéjére.

Második lépésként megvizsgáltam a határeloszlás és a normalizált maximumok QQ-plotját, valamint Michael-féle szórás stabilizált PP-plotját [4]. A QQ- és PP-plotot a modellek diagnosztizálására szokták alkalmazni, képet ad arról, hogy az általunk használt modell mennyire jó úgy, hogy összehasonlítja a tapasztalati- és illesztett eloszlásfüggvényt. A pontok minél inkább illeszkednek az $x = y$ egyenesre, annál jobb a modellünk. Amellett, hogy a QQ- és PP-plottal a modell milyenségét vizsgáljuk, a konvergencia gyorsaságáról is képet ad úgy, hogy a mintanagyságokra hogyan javul az illeszkedés.

Végül Kolmogorov-Szmirnov próbával vizsgáltam a tapasztalati- és határeloszlás görbék eltérésének supremumát, amely által számszerű eredményt kaptam a konvergencia sebességére. A Kolmogorov-Szmirnov próbát illeszkedés ellenőrzésére alkalmazzák, egymintás esetben megadja, hogy egy n elemű minta eloszlása csakugyan az az eloszlás-e, amit feltettünk, két mintás esetben pedig, hogy a két minta azonos eloszlásból való-e. Lényege (az egymintás esetre), hogy a döntést a tapasztalati és az elméleti eloszlásfüggvény abszolút eltérésének supremuma alapján hozzuk meg. Glivenko tétele szerint a tapasztalati eloszlásfüggvény egyenletesen tart a valódi eloszlásfüggvényhez, azaz:

$$d_n = \sup_x |\hat{F}_n(x) - F(x)| \to 0.$$

Így a nullhipotézisünk az, hogy a minta a feltételezett elméleti eloszlásból származik, az

ellenhipotézisünk pedig az, hogy nem. A minta minden elemére kiszámoljuk a különbséget, és ha a d_{max}, azaz a legnagyobb különbség nagyobb, mint egy előre meghatározott α-ra d_α, akkor a nullhipotézisünk nem teljesül. A supremumot a sorbarendezett mintaértékeken számoljuk. Számunkra a d_n-ek változása lesz érdekes a különböző n-ekre, mely megmutatja, milyen gyorsan csökken a két eloszlásgörbe eltérése.

A szimulációk során az exponenciális- és egyenletes minta maximumának esetében k-t 10^4-nek, normális eloszlású minta esetén $5 \cdot 10^3$-nak választottam technikai okokból. Mégis, ezen ismétlésszámok mellett is a minták maximumainak eloszlása elegendően hasonlítottak a különböző extrémérték eloszlásokhoz.

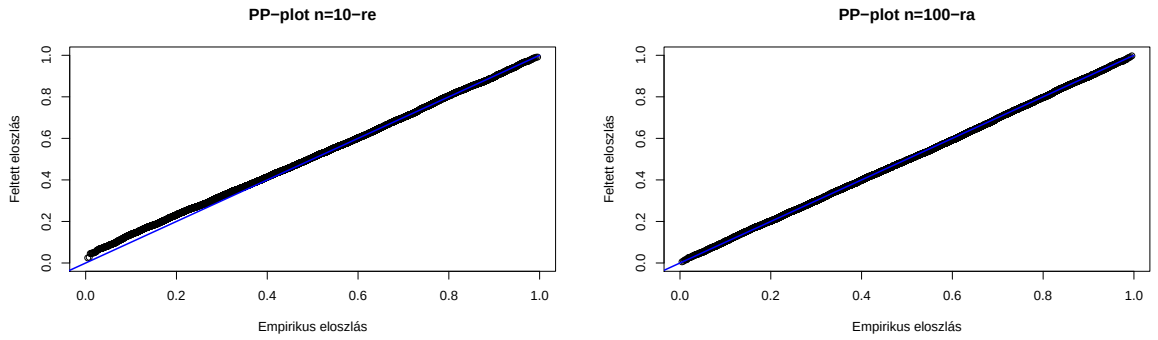
Végül a szimulációk bemutatása előtt megjegyzem, hogy ugyan a részletesen bemutatott elmélet nem alkalmazható a Gumbel esetre, de gyakorlati fontosságuk miatt fontos őket megvizsgálni. Ismert, hogy az exponenciális- és normális eloszlású minta maximumeloszlása is a Gumbel határeloszláshoz tart.

4.1. n elemű, exponenciális eloszlású minta maximumának eloszlása

Az 1.13 példában bevezettük, hogy az exponenciális eloszlású minta maximumának eloszlása a Gumbel eloszláshoz tart. A normáló konstansok maradjanak az 1.13-ban látottakhoz hasonlóak.

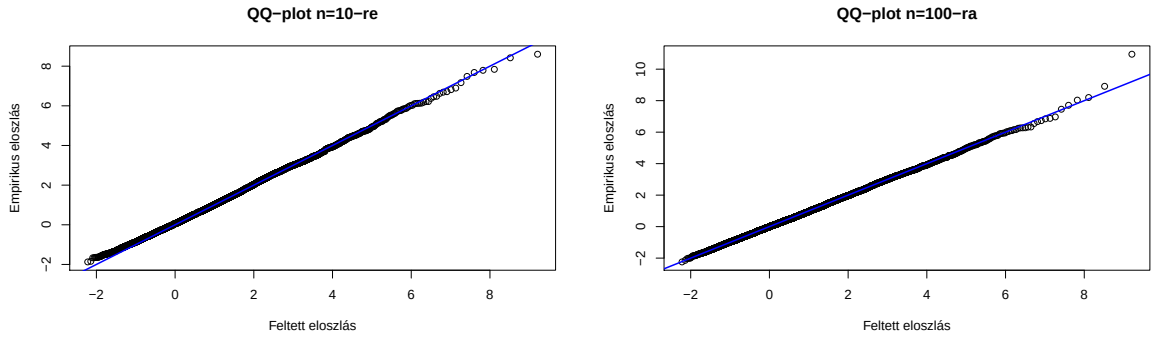
Az 1.2 ábrán szemmel látható, hogy már $n = 10$ esetén is csak kis eltérés van az empirikus- és a Gumbel eloszlásgörbék között. A görbéket n szerint növekvő sorrendben illesztettem az ábrára, így az, hogy csak a piros görbe látható azt jelenti, hogy teljesen lefedi az $n = 100$-as esetet és a Gumbel eloszlás görbéjét. Tehát az exponenciális eloszlású minta maximumának eloszlása gyorsan konvergál a határeloszláshoz.

Tekintsük a PP-plot ábrát az n elemű minta maximumával és Gumbel eloszlással. A konvergencia gyorsaságát támasztja alá a 4.1 ábra is. A képen látható, hogy gyakorlatilag már $n = 100$-tól illeszkednek a pontpárok az egyenesre.



4.1. ábra. A standard Gumbel eloszlás összehasonlítva a normalizált exponenciális maximumeloszlással

A QQ-plot ábrán is jól illeszkednek a pontok az $x = y$ egyenesre. A nagyobb eltérést adó pontok a véletlennek az eredményei: nagyobb ismétlésszám esetében szinte tökéletes az illeszkedés.



4.2. ábra. A standard Gumbel eloszlás összehasonlítva a normalizált exponenciális maximumeloszlással

A Kolmogorov-Szmirnov próba eredményei a 4.1 táblázatban vannak összefoglalva. A két eloszlás különbségének maximuma sebesen csökken, a p-érték eredményei pedig gyorsan nőnek. Ugrásszerű növekedést látunk $n = 1000$ és $n = 5000$ p-értékei között.

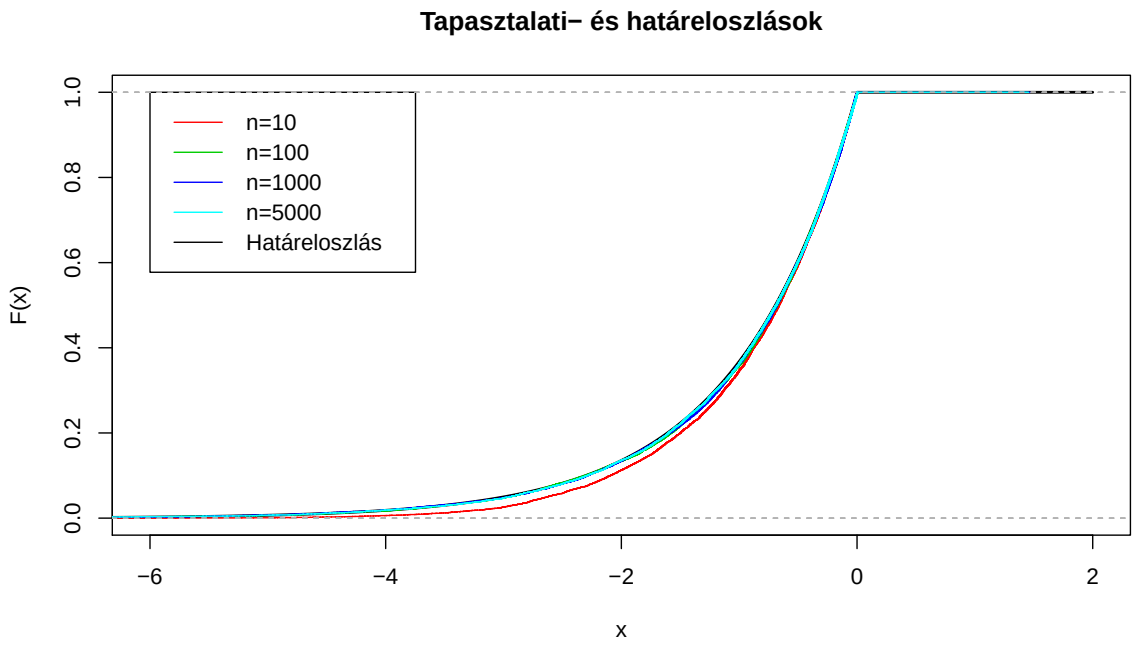
mintanagyság (n)	$d_{\max}$	p-érték
$n = 2$	0.1636	$< 10{-}16$
$n = 10$	0.0336	$< 10^{-10}$
$n = 100$	0.014	0.04064
$n = 1\,000$	0.0083	0.4994
$n = 5\,000$	0.0061	0.8456

4.1. táblázat. Kolmogorov-Szmirnov próba exponenciális eloszlású minta maximumára a Gumbel eloszlással

A következtetésünk tehát, hogy egy exponenciális eloszlású minta maximumának eloszlása hamar felveszi a Gumbel-eloszlás görbe alakját.

4.2. n elemű, egyenletes eloszlású minta maximumának eloszlása

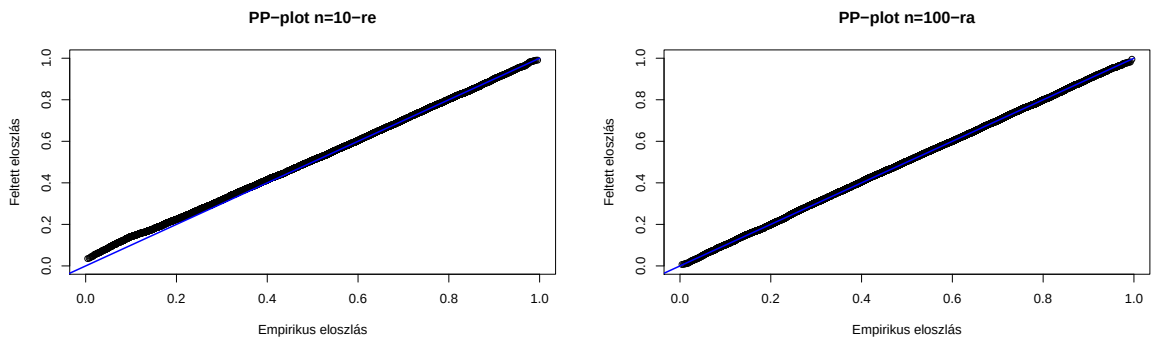
Az egyenletes eloszlású mintamaximumok eloszlása a Weibull maximum vonzási tartományába tartozik. Mi most tekintsük az $U(0, 1)$ standard egyenletes eloszlású minta maximumát. Az M_n normalizálásához válasszuk az $a_n = \dfrac{1}{n}$ és $b_n = 1$ konstansokat.



4.3. ábra. Az a_n, b_n konstansokkal normált standard egyenletes eloszlású minta maximumának közelítése a Weibull eloszláshoz

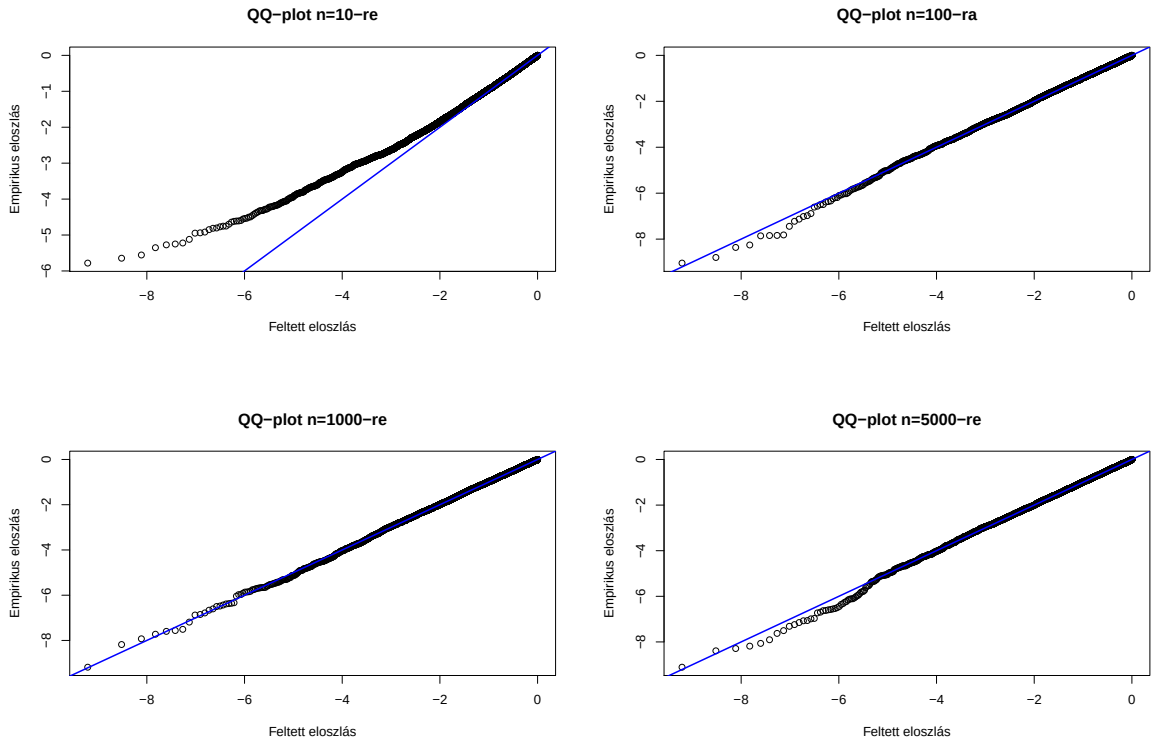
Az exponenciális mintamaximumhoz hasonló helyzet áll fent, ugyanis olyan gyorsan konvergál az egyenletes mintamaximum eloszlása a Weibull eloszláshoz, hogy már 100 mintaelem esetén is illeszkedik az empirikus görbe az elméleti görbére. Ezt mutatja, hogy csak világoskék, $n = 5000$ mintához tartozó görbe látszik, amely letakarja az $n = 100$ és $n = 1000$ eseteket is.

A PP-plotok az egyenletes mintamaximumoknál is gyors illeszkedést mutatnak.



4.4. ábra. A Weibull eloszlás összehasonlítva a normalizált egyenletes maximumeloszlással

A 4.5 ábrán $n = 10$ esetén ugyan a kvantilis pontpárok nem illeszkednek az egyenesre, $n = 100$-tól már az eloszlásgörbék alapján várt értékeket látunk.



4.5. ábra. A Weibull eloszlás összehasonlítva a normalizált egyenletes maximumeloszlással

A Kolmogorov-Szmirnov próba is a várt eredményeinket igazolja. A görbék eltérése már $n = 10$ esetén is alacsony volt, n növekedtével közel 0 a különbség.

mintanagyság (n)	$d_{\max}$	p-érték
$n = 10$	0.0264	10^{-6}
$n = 100$	0.0106	0.2108
$n = 1\,000$	0.0092	0.3624
$n = 5\,000$	0.0059	0.8722

4.2. táblázat. Kolmogorov-Szmirnov próba n elemű egyenletes eloszlású minta maximumára Weibull eloszlással

4.3. n elemű, standard normális eloszlású minta maximumának eloszlása

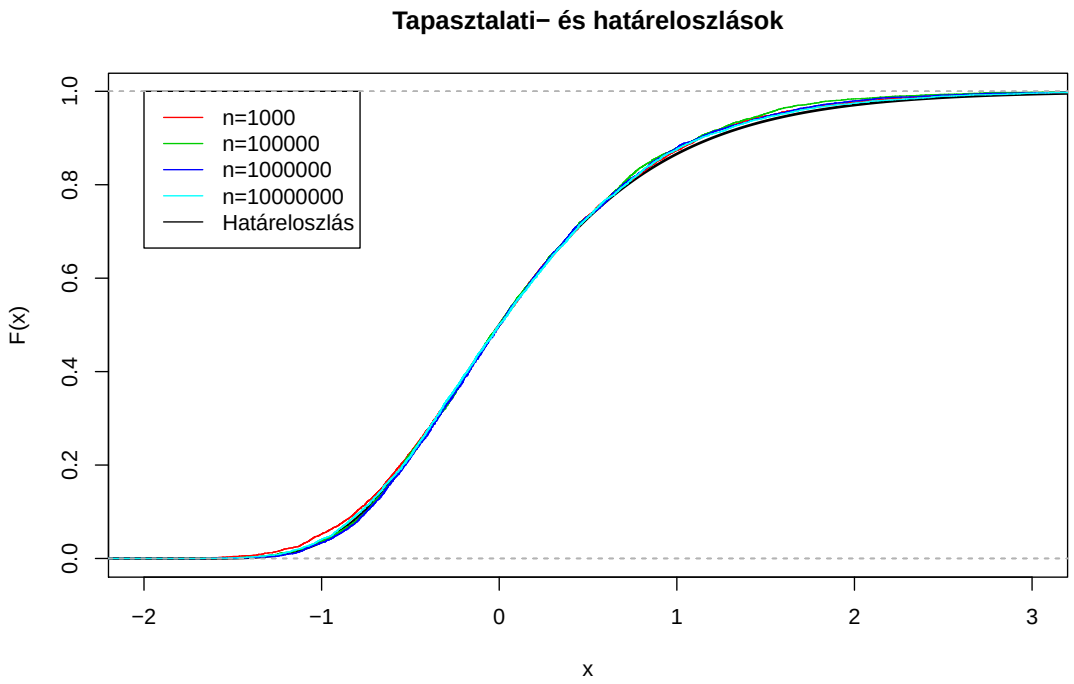
A különböző határeloszlások maximum vonzási tartományainál már láttuk, hogy az exponenciális lecsengésű eloszlások, tehát a normális eloszlású minták maximuma is a Gumbel eloszláshoz konvergál.

Keressük a megfelelő a_n és b_n normalizáló sorozatokat a kvantilisek segítségével. Legyen $0 < q < 1$. Ekkor F_n q-kvantilise az az x_q, amelyre teljesül, hogy $F_n(x_q) = q$. Azonos valószínűségi változók maximumának eloszlásfüggvénye $F_n(x) = F^n(x)$, ezért $x_{q,n} = F^{-1}(q^{\frac{1}{n}})$. Választhatjuk b_n-nek a mediánt, vagyis $b_n = x_{\frac{1}{2};n}$ és a_n-t pedig a két kvartilis különbségének, azaz $a_n = x_{\frac{3}{4};n} - x_{\frac{1}{4};n}$.

Mivel a becslésünk mediánja 0, szórása 1, így a határeloszlásunknak is hasonlóan kell viselkednie. Legyen β a skálaparaméter és α a helyparaméter. A medián $\alpha + \beta \log\log 2$, ami tehát egyenlő 0-val, és a szórás $\beta(\log\log 4 - \log\log\frac{4}{3})$, ami egyenlő 1-gyel. A két egyenletből megkapjuk a paramétereket:

$$\beta = \frac{1}{\log\log 4 - \log\log(\frac{4}{3})}, \quad \alpha = -\frac{\log\log 2}{\log\log\frac{4}{3} - \log\log 4}.$$

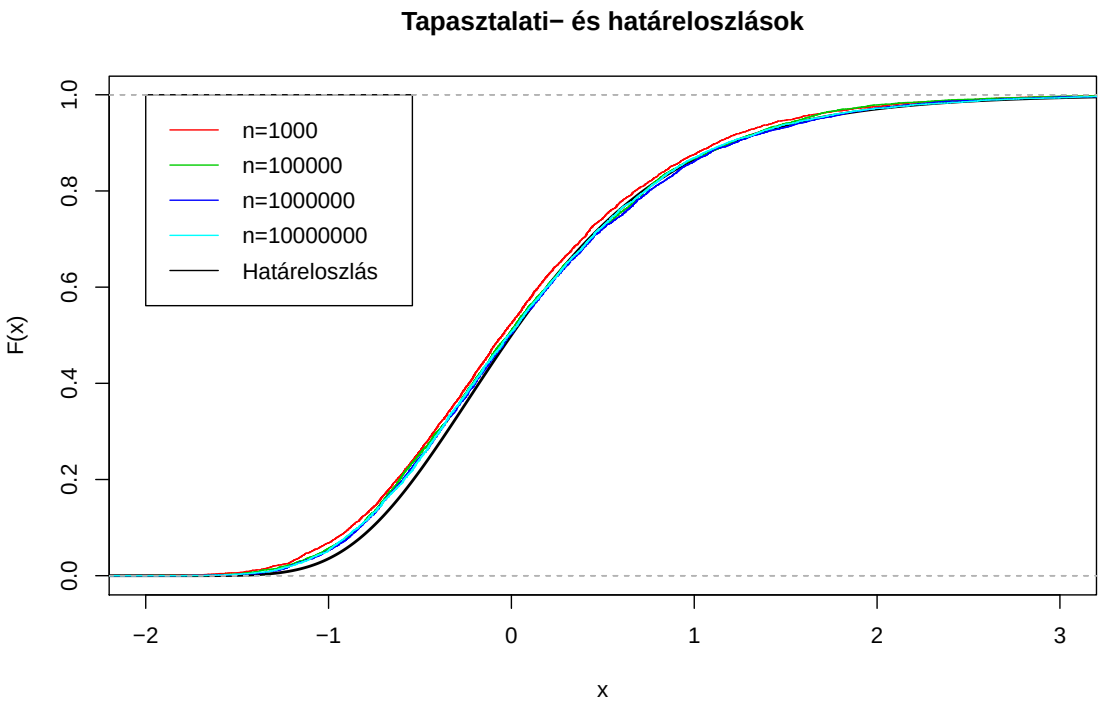
A Gumbel eloszlást ezekkel standardizáljuk.



4.6. ábra. Kvantilisekkel normált standard normális eloszlású minta maximumának közelítése az α, β paraméterű Gumbel eloszláshoz. A legnagyobb mintanagysághoz tartozó görbe majdnem lefedi a határeloszlást.

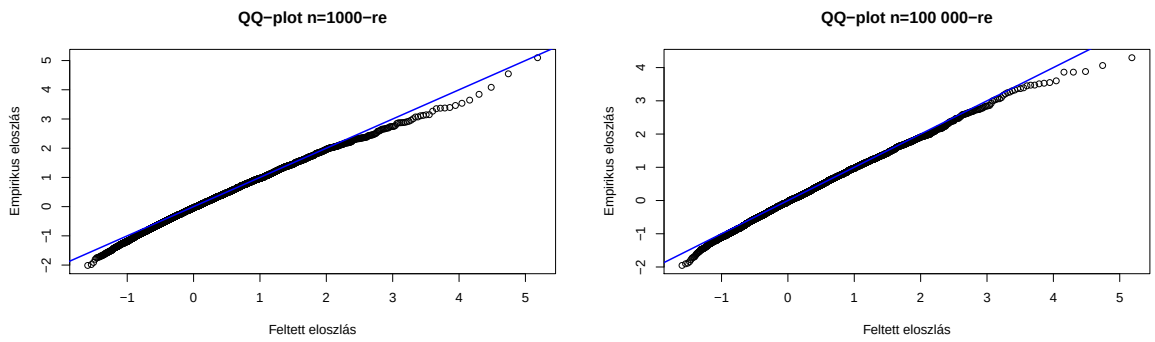
A normalizáló konstansoknak természetesen nem muszáj pontosan ezeket az értékekek felvenniük, az is elegendő, ha csupán közelítik ezeket. Nézzünk egy konkrét példát, ahol a normalizáló konstansok [9] alapján a következők:

$$a'_n = \frac{\log\left(\frac{4\log^2 2}{\log^2 4/3}\right)}{2\sqrt{2\log n}}, \quad b'_n = \sqrt{2\log n} - \frac{\log(\log n) + \log(4\pi\log^2 2)}{2\sqrt{2\log n}}.$$

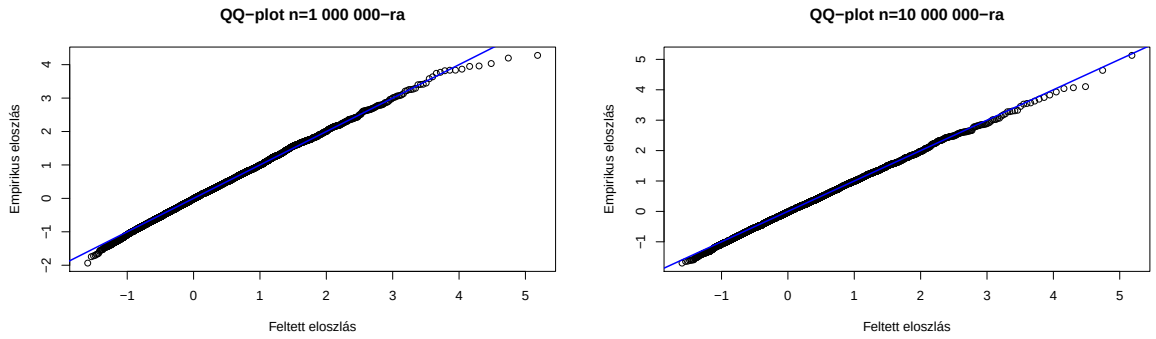


4.7. ábra. Az a'_n, b'_n konstansokkal normált standard normális eloszlású minta maximumának közelítése az α, β paraméterű Gumbel eloszláshoz

Most nézzük meg, mennyire illeszkednek a pontpárok az $x = y$ egyenesre.



4.8. ábra. A Gumbel eloszlás összehasonlítva a standard normális eloszlású minta normalizált maximumeloszlásával



4.9. ábra. A Gumbel eloszlás összehasonlítva a standard normális eloszlású minta normalizált maximumeloszlásával

A 4.8 és 4.9 ábrákon látható, hogy van konvergencia, ám a lassú közelítést igazolja az is, hogy 10^6-os mintanagyságra sincs tökéletes illeszkedés. Nézzük meg, milyen tendenciát mutat az eloszlásgörbék távolságának supremuma.

mintanagyság (n)	d_{max}	p-érték
$n = 1\,000$	0.0447	$< 10^{-9}$
$n = 100\,000$	0.0404	$< 10^{-7}$
$n = 1\,000\,000$	0.0316	$< 10^{-5}$
$n = 10\,000\,000$	0.0308	0.0001484

4.3. táblázat. Kolmogorov-Szmirnov próba normális eloszlású minta maximumára Gumbel eloszlással

A 4.3 táblázat d_{max} elemei noha lassan, de csökkenek, vagyis valóban közelíti az a'_n, b'_n konstansok által normalizált maximumok eloszlása a Gumbel eloszlást.

Összegzés

Szakdolgozatomban a független, azonos eloszlású valószínűségi változók maximumának konvergenciáját vizsgáltam a Fisher-Tipett-Gnedenko tételben megfogalmazott határeloszlásokhoz. Az elméleti eredményeket összefoglaló részben a Fréchet-eloszlásra tértem ki részletesen. Láthattuk, hogy milyen fontos a normalizáló konstansok megválasztása a konvergencia sebességének szempontjából, és azt is, hogyan javíthatunk rajtuk. A Weibull eloszlással lévő szoros kapcsolata miatt ezek a Fréchet határeloszlásra vonatkozó eredmények - egy kis transzformációval - alkalmazhatóak voltak a Ψ_α esetre is, melyet rövidebben ismertettem. A Gumbel határeloszlás esete jóval bonyolultabb az előző kettőnél, és meghaladja eme dolgozat kereteit, így arra nem tértem ki, jóllehet ennek vizsgálata érdekes lehet.

A dolgozatban bemutattam a g-sv függvények jelentőségét, a konvergencia rendjéről kaphatunk információt általuk. Mind a tételekben és bizonyításokban alkalmaztuk őket.

Az extrém-érték elmélet egy másik megközelítése a küszöb feletti értékek vizsgálata. Megnéztem, miképpen alkalmazhatók a g-sv függvények a küszöbmeghaladáson alapuló modellek esetén. Láthattuk, hogy az Általánosított Pareto eloszlás (GPD) jó közelítése az u szint feletti maximumok eloszlásának. Ezen GPD paramétereinek becslésére kaptunk eredményeket.

Végül szimulációk segítségével vizsgáltam és szemléltettem a mintamaximumok konvergenciáját a határeloszlásokhoz. Láthattunk két gyors, és egy rendkívül lassú konvergencia-sebességet. Ugyan a Gumbel eloszlás elméleti összefoglalónkban nem jelent meg, mégis érdekes volt a mintamaximumok ezen eloszláshoz tartó közelítését szemlélni. Az egyenletes eloszlású mintamaximumok Ψ_α-hoz és exponenciális eloszlású mintamaximumok Λ_α-hoz tartó konvergenciája gyors, $n = 100$ esetén is felvették a határeloszlásuk görbéjének alakját. A normális eloszlású mintamaximumoknál láthattuk, ha kvantilisekkel normáltuk az M_n-t, lényegesen gyorsabban közelített a Gumbel eloszláshoz, mint amikor konstansokkal normalizáltunk. Utóbbi esetben még $n = 10^7$ mintanagyságra sem simult a határeloszlásgörbére. A szimulációk során az R program evd csomagját használtam, a GPD ábra készítéséhez pedig az fExtremes csomagot alkalmaztam.

Irodalomjegyzék

[1] Arató Miklós, Prokaj Vilmos és Zempléni András: *Bevezetés a valószínűségszámításba és alkalmazásaiba: példákkal, szimulációkkal*, 2013

[2] Dr. Barabás Béla: *Extrémérték-elmélet*, 2010

[3] Stuart Coles: *An Introduction to Statistical Modeling of Extreme Values*, 2001

[4] John R. Michael: *The Stabilized Probability Plot*, 1982

[5] T. Mikosch: *Regular Variation, Subexponentiality and Their Applications in Probability Theory*

[6] Richard L. Smith: *Estimating Tails of Probability Distributions*, 1987

[7] Richard L. Smith: *Uniform Rates Of Convergence In Extreme-Value Theory*, 1982

[8] Dr. Zempléni András: *Valószínűségszámítási modellek előadás*

[9] http://stats.stackexchange.com/questions/105745/extreme-value-theory-show-normal-to-gumbel