Extrémumokra vonatkozó határeloszlások

Szakdolgozat

Írta: Németh László

Alkalmazott matematikus BSc

Témavezető:

Dr. Zempléni András, egyetemi docens

Valószínűségelméleti és Statisztika Tanszék

Eötvös Loránd Tudományegyetem, Természettudományi Kar

Eötvös Loránd Tudományegyetem

Természettudományi Kar

2014

Tartalomjegyzék

Köszönetnyilvánítás

Ezúton is szeretnék köszönetet mondani témavezetőmnek, Zempléni Andrásnak, akihez mindig fordulhattam kérdéseimmel. Hasznos tanácsokkal látott el a szakdolgozattal kapcsolatban és nagyon sokat segített az R program megismerésében.

Továbbá köszönöm a családomnak és a barátaimnak a támogatást, és a dolgozat formai tökéletesítésében nyújtott segítséget.

Bevezetés

A valószínűségszámítás és a statisztika rendkívül fontos a mai világban. Rengeteg adat áll rendelkezésünkre, melyek segítségével meg lehet becsülni az életünket befolyásoló tényezők alakulását. Könnyen számolunk várható értékeket, szórást és azt is meg tudjuk határozni, hogy mennyire megbízható az előrejelzésünk. Azonban vannak olyan extrém esetek, amikről még mindig nincs elég adatunk, mert rendkívül ritkán fordulnak elő. Ilyenek például a nagy árvizek, vagy a nagy anyagi kárral járó balesetek. Azonban ezek eloszlását is meg kell valahogy becsülni, hiszen az embereknek fel kell készülniük az árvizekre és a biztosítótársaságok legnagyobb kifizetései a nagy anyagi károkból keletkeznek. Dolgozatom célja, hogy az ilyen speciális esetekkel foglalkozó extrém érték elmélet valószínűségszámítási alapjait ismertessem. Megbecsüljem valószínűségi változók sorozatának maximumát, adjak rá egy megfelelő eloszlást és megvizsgáljak más, különösen magas értékekkel foglalkozó modellt is. A dolgozatot Filip Lindskog - *The mathematics and fundamental ideas of extreme value theory* [1] című cikke alapján írom, kiegészítve más szerzők műveiből kiemelt részekkel és saját tesztelésekkel, szimulációkkal valamint példákkal.

A dolgozatban azt tárgyalom, hogy hogyan lehet megbecsülni egy adott eloszlásból generált mintában a ténylegesen legnagyobb előforduló elemet, illetve, hogy a különösen magas értékek előfordulása milyen eloszlást követ.

A maximumok vizsgálatát az egyértelműséggel kezdem, majd a Fisher-Tippet tétel bemutatásával folytatom, amely szerint, ha létezik a maximumoknak határeloszlása, akkor az a Fréchet, Weibull vagy Gumbel eloszláscsaládok egyikébe tartozik. E három eloszláscsaládot összefoglalja az általánosított extrém érték eloszlás, melynek a három paramétere az eloszlástípust, a várható értéket és a skálaparamétert határozza meg. Ennek a segítségével tudom jellemezni az egyes típusok max-vonzási tartományát, azaz, hogy egy adott eloszlásból vett minta maximuma vajon melyik típus szerint keresendő.

A küszöb feletti értékek módszerével nem csak a legmagasabb értéket szeretném megvizsgálni, hanem kicsit általánosabban az "elég magas" értékek eloszlását. A küszöb feletti értékek eloszlását az általánosított Pareto eloszlás írja le megfelelően.

Ezt követően több szimulációt is végzek az R program segítségével, hogy ábrázoljam a különféle eloszlásokból vett értékek maximumának alakulását. Nem célom mélyebb statisztikai vizsgálatokba kezdeni így a hibahatárt vagy a határeloszláshoz tartás sebességét nem is tüntettem fel. A szimulációk célja az, hogy adjon egy szemléletet a minták maximumának eloszlásáról és megmutassa az olvasó számára, hogy rendkívül hasonlít egy extrém érték eloszláshoz. Hasonló szimulációkat végzek a küszöb feletti módszerrel kapcsolatban is, külön kitérve a helyesen választott küszöbérték fontosságára.

1. Matematikai alapok

1.1. Definíciók

A dolgozatban feltételezem olyan alapvető definíciók ismeretét a valószínűségszámítás témaköréből, mint valószínűségi változó, eloszlás, eloszlásfüggvény. A nem feltétlenül közismert, de felhasznált fogalmakat itt szeretném előre definiálni.

1.1. Definíció (Gyenge konvergencia)**.** *Legyenek az $X_1, \dots, X_n$, X valószínűségi változók, a hozzájuk tartozó eloszlásfüggvények pedig $F_1, \dots, F_n$ és F. Azt mondjuk, hogy az $X_n \to X$ gyengén (vagy eloszlásban), ha $n \to \infty$ esetén $F_n \to F$ az utóbbi minden folytonossági pontjában.*

A későbbiekben ha konvergenciáról beszélünk, mindig gyenge konvergenciát értünk alatta.

1.2. Definíció. *Legyen az X_1 és az X_2 valószínűségi változók. Azt mondjuk, hogy X_1 és X_2 azonos eloszlású, ha eloszlásfüggvényeik megegyeznek. Jelölésben $X_1 \sim X_2$ vagy $X_1 \stackrel{d}{=} X_2$.*

1.3. Definíció. *Az X valószínűségi változót elfajult eloszlásúnak mondjuk, ha létezik olyan $c \in \mathbb{R}$, hogy $\mathbb{P}(X = c) = 1$.*

1.4. Definíció. *Legyen az X egy valószínűségi változó, melynek eloszlásfüggvénye F. Az F függvény jobb oldali végpontjának nevezzük azt az $x_F \in \overline{\mathbb{R}}$ számot, melyre $x_F = \inf\{x \in \overline{\mathbb{R}} : F(x) = 1\}$.*

1.5. Definíció (Túlélésfüggvény)**.** *Legyen az X valószínűségi változó. Ekkor az $\overline{F}(x) = \mathbb{P}(X > x)$ függvényt az X túlélési függvényének nevezzük.*

Megjegyezzük, hogy folytonos eloszlású X esetén $\mathbb{P}(X = x) = 0$, tehát $\overline{F}(x) = \mathbb{P}(X > x) = \mathbb{P}(X \geq x)$ egyenlőség is érvényes. Ebből az észrevételből ekkor következik, hogy $1 - F(x) = \overline{F}(x)$, ahol F az eloszlásfüggvényt jelöli. Az angolszász szakirodalommal összhangban az eloszlásfüggvényt $F(x) = \mathbb{P}(X \leq x)$ alakban definiáljuk, így $1 - F(x) = \overline{F}(x)$ minden esetben teljesül. A továbbiakban ezt a tulajdonságot többször fel fogom használni. Ahol diszkrét eloszlásokkal kapcsolatban használok túlélési függvényt, ott erre külön ki fogok térni.

1.6. Definíció (Általánosított inverz)**.** *Legyen h egy $\mathbb{R}$-en értelmezett monoton növő függvény. Ekkor*

$$h^{\leftarrow}(u) = \inf\{x \in \mathbb{R} : h(x) \geq u\}$$

függvényt h általánosított inverzének nevezzük. Üres halmaz infimumát ∞-ként definiáljuk.

Amennyiben az F egy eloszlásfüggvény, az $F^{\leftarrow}(u)$ általánosított inverzet az F függvény u-kvantilisének nevezzük.

1.2. Motiváció

Vegyük az X_k, $(k = 1, ...n)$ független, azonos eloszlású valószínűségi változók sorozatát. Legyen F ezeknek a közös eloszlásfüggvénye. Minden $n \geq 1$-re készítsük el az $S_n = X_1 + \cdots + X_n$ összeget. Mi lehet az S_n eloszlása, elég nagy n-ek esetén? Ez a várható értéktől és a szórástól függ. Ha a valószínűségi változó nem azonosan 0, akkor tudjuk, hogy $|S_n| \to \infty$, ha $n \to \infty$. Ezért érdekesebb az, hogy mi lehet a határeloszlása a normált és centralizált összegnek.

Ahhoz, hogy ezt megvizsgálhassuk, be kell vezetnünk a_n és b_n számsorozatokat, úgy, hogy $a_n > 0$ valamint $b_n \in \mathbb{R}$. Ekkor keressük azokat a nemelfajult W valószínűségi változókat, melyekhez létezik a fenti módon megadott a_n és b_n, valamint $(S_n - b_n)/a_n \to W$.

1.7. Definíció. *Egy adott X valószínűségi változót stabilisnak mondunk, ha bármely $c_1, c_2 \in \mathbb{R}_+$ számra és $X_1, X_2 \sim X$ független valószínűségi változókra léteznek olyan $a(c_1, c_2) > 0$ és $b(c_1, c_2) \in \mathbb{R}$ konstansok, melyekre*

$$c_1X_1 + c_2X_2 \stackrel{d}{=} a(c_1, c_2)X + b(c_1, c_2). \tag{1}$$

Ha az X_k, $(k = 1, ...n)$ egy független, stabilis, azonos eloszlású valószínűségi változókból álló sorozat, akkor az (1) egyenlet szerint létezik olyan $a_n > 0, b_n \in \mathbb{R}$, hogy minden $n \geq 1$-re teljesüljön az $S_n = X_1 + \cdots + X_n \stackrel{d}{=} a_nX + b_n$ egyenlőség. Azaz stabilis változók esetén létezik az összegnek határeloszlása. Az egyenlet átírható $(S_n - b_n)/a_n \stackrel{d}{=} X$ alakba is. A későbbiekben ezt az alakot fogom használni. Megmutatható tehát, hogy a stabilis eloszlások lehetséges nemelfajuló határeloszlások a megfelelően normált és centralizált összegek esetén. Véges szórású ilyen eloszlás csak a standard normális eloszlás, azaz ebben az esetben az összeg ehhez fog tartani, ami a centrális határeloszlás tétele miatt ismert. Ehhez hasonló módon próbáljuk a későbbiekben megközelíteni az n valószínűségi változó maximumának eloszlására vonatkozó kérdést.

2. A normált maximumok határeloszlása

2.1. Egyértelműség

Legyen az X_k, $(k = 1, ...n)$ független azonos eloszlású valószínűségi változók sorozata, melyeknek az eloszlásfüggvénye F. Vizsgáljuk ezek

$$M_n = \max\{X_1, \ldots, X_n\}$$

maximumát. Adott n esetén könnyen meg tudjuk határozni az M_n eloszlását, hiszen $F_{M_n}(x) = F(x)^n$. Legyenek $a_n > 0$, $b_n \in \mathbb{R}$ számsorozatok. Keressük azt az eloszlást, amihez az a_n, b_n sorozatokkal normált M_n tart, ha $n \to \infty$. Ez $\mathbb{P}((M_n - b_n)/a_n \leq x)$, vagy másképp $\mathbb{P}(M_n \leq u_n)$ ahol $u_n = a_n x + b_n$. Először is fontos tudni, hogy milyen feltételek mellett létezik egyáltalán nemtriviális határeloszlása a $\mathbb{P}(M_n \leq u_n)$, $n \to \infty$ sorozatnak. Erre a kérdésre a következő tétel ad választ.

2.1. Tétel (Poisson approximáció)**.** *Legyen adott $\tau \in [0, \infty]$ és egy $u_n \in \mathbb{R}$ sorozat, továbbá X_k, $(k = 1, ...n)$ független azonos eloszlású valószínűségi változók sorozata, melyek részsorozatainak maximuma M_n. Ekkor az alábbi állítások ekvivalensek:*

$$\text{i)} \qquad n\overline{F}(u_n) \to \tau, \tag{2}$$

$$\text{ii)} \qquad \mathbb{P}(M_n \leq u_n) \to e^{-\tau}. \tag{3}$$

Bizonyítás. i) $\Rightarrow$ ii)

Először legyen a $\tau < \infty$. Ha a (2) állítás teljesül, akkor $\mathbb{P}(M_n \leq u_n) = F^n(u_n) = (1 - \overline{F}(u_n))^n = (1 - n\overline{F}(u_n)/n)^n \to e^{-\tau}$, mivel $n\overline{F}(u_n) \to \tau$.

Most tegyük fel, hogy (3) teljesül. Először belátjuk, hogy $\overline{F}(u_n) \to 0$. Indirekt tegyük fel, hogy nem 0 a határértéke, ekkor létezik egy olyan (n_k) sorozat, melyre $\overline{F}(u_{n_k})$ eltávolodik 0-tól elég nagy k esetén. Ekkor $\mathbb{P}(M_n \leq u_{n_k}) = (1 - \overline{F}(u_{n_k}))^{n_k} = (1 - n_k\overline{F}(u_{n_k})/n_k)^{n_k} \to 0$, ami ellentmond a (3) összefüggésnek. Tehát $\overline{F}(u_n) \to 0$. Vegyük a (3) összefüggés logaritmusát. Ekkor azt kapjuk, hogy

$$\ln \mathbb{P}(M_n \leq u_n) = \ln(1 - \overline{F}(u_n))^n = n \ln(1 - \overline{F}(u_n)) \to -\tau.$$

Mivel $\ln(1 - x) \sim -x$ ha $x \to 0$, következik, hogy $n\overline{F}(u_n) \to \tau$ azaz (2) teljesül.

Most lássuk a $\tau = \infty$ esetet. Tegyük fel, hogy (2) fennáll, de (3) nem. Ekkor létezik olyan (n_k) részsorozat, melyre $P(M_{n_k} \leq u_{n_k}) \to e^{-\tau'}$, ha $k \to \infty$ valamely $\tau' < \infty$ esetén. De a fenti megfontolások alapján $n_k\overline{F}(u_n) \to \tau' < \infty$ ami ellentmondás, tehát (3) teljesül.

A ii) $\Rightarrow$ i) irány belátásához feltesszük, hogy $\mathbb{P}(M_n \leq u_n) \to 0$. Ezt átírva kapjuk, hogy $(1 - \overline{F}(u_n))^n \to 0$, ami akkor lehetséges, ha $\overline{F}(u_n) \to 1$, azaz $n\overline{F}(u_n) \to \infty$. $\square$

2.2. Definíció (max-stabilitás). *Egy X nemelfajult valószínűségi változót max-stabilisnak mondunk, ha minden $2 \leq n$ esetén léteznek $a_n > 0$ és $b_n \in R$ konstansok, hogy*

$$\max\{X_1, \dots, X_n\} \stackrel{d}{=} a_n X + b_n,$$

ahol $X_1, \dots, X_n$ függetlenek, azonos eloszlású valószínűségi változók és $X_i \sim X$.

Ha egy adott X_k, $(k = 1, ...n)$ valószínűségi változók független sorozata max-stabilis, akkor léteznek hozzá megfelelő a_n és b_n konstansok, hogy minden n esetén $M_n \stackrel{d}{=} a_n X_1 + b_n$. Tehát ha egy eloszlás max-stabilis, akkor lehet határeloszlás. Más lehetséges határeloszlás nem létezik, ezért csak a max-stabilis esettel foglalkozunk.

2.3. Tétel (Típusokhoz konvergálás). *Legyenek az X_n, U és V valószínűségi változók, melyek közül sem U, sem V nem elfajuló. Legyenek továbbá $a_n, \alpha_n > 0$, valamint $b_n, \beta_n \in \mathbb{R}$ konstansok. Ha*

$$\frac{X_n - b_n}{a_n} \to U \quad \text{és} \quad \frac{X_n - \beta_n}{\alpha_n} \to V, \tag{4}$$

akkor léteznek olyan $A > 0$ és $B \in \mathbb{R}$ számok, melyekre

$$\frac{\alpha_n}{a_n} \to A \quad \text{és} \quad \frac{\beta_n - b_n}{a_n} \to B, \tag{5}$$

valamint

$$V = \frac{U - B}{A} \tag{6}$$

is teljesül. Továbbá, ha (5) fennáll, akkor (4) bármelyikéből következik a másik, valamint (6) is igaz lesz.

A tétel bizonyítása megtalálható Resnick [2] könyvében.

Ez a tétel azt állítja, hogy adott, független, azonos eloszlású valószínűségi változók egy sorozata esetén, ha a maximumokhoz két határeloszlást is találtunk, csak más-más konstansokkal, akkor ez a két határeloszlás lényegében megegyezik, lineáris transzformációval egymásba vihető.

2.2. Lehetséges határeloszlások

Már látjuk, hogy ha létezik határeloszlás, az egy eloszláscsalád erejéig egyértelmű. A következő tétel választ ad arra a kérdésre, hogy milyen típusúak lehetnek az eloszláscsaládok.

2.4. Tétel (Fisher-Tippet)**.** *Legyen az X_k, $(k = 1, ...n)$ független azonos eloszlású valószínűségi változók sorozata. Ha létezik olyan a_n és b_n konstans, melyekre $a_n > 0$, $b_n \in \mathbb{R}$ valamint létezik egy olyan H úgy, hogy*

$$\frac{M_n - b_n}{a_n} \to H, \tag{7}$$

akkor H az alábbi három eloszlásfüggvény egyikének típusába tartozik:

$$\begin{aligned}
\textit{Fréchet}: \quad \Phi_\alpha(x) &= \begin{cases} 0, & x \leq 0, \\ e^{-x^{-\alpha}}, & x > 0, \end{cases} \qquad \alpha > 0. \\
\textit{Weibull}: \quad \Psi_\alpha(x) &= \begin{cases} e^{-(-x)^{\alpha}}, & x \leq 0, \\ 1, & x > 0, \end{cases} \qquad \alpha > 0. \\
\textit{Gumbel}: \quad \Lambda(x) &= e^{-e^{-x}}, \qquad x \in \mathbb{R}.
\end{aligned}$$

Bizonyítás. (7)-ből adódik, hogy minden $t > 0$-ra

$$F^{[nt]}(a_{[nt]}x + b_{[nt]}) \to H(x),$$

hiszen $[nt]$ egész számok sorozata, és $[nt] \to \infty$. Továbbá

$$F^{[nt]}(a_n x + b_n) = (F^n(a_n x + b_n))^{[nt]/n} \to (H(x))^{[nt]/n} \to H^t(x)$$

A típushoz konvergálás tétele szerint, ha $t > 0$, akkor léteznek olyan $f(t) > 0$ és $g(t) \in \mathbb{R}$ függvények, melyekre

$$\lim_{n\to\infty} \frac{a_n}{a_{[nt]}} = f(t), \quad \lim_{n\to\infty} \frac{(b_n - b_{[nt]})}{a_{[nt]}} = g(t)$$

és teljesül, hogy

$$H^t(x) = H(f(t)x + g(t)). \tag{8}$$

Mivel f és g mérhető függvények határfüggvényei, ők maguk is mérhetők. A (8) egyenlőségből következik, hogy $t > 0$-ra és $s > 0$-ra

$$H^{ts} = H\left(f(ts)x + g(ts)\right),$$

melyet tovább alakítva azt kapjuk, hogy

$$\begin{aligned}
H^{ts} = (H^s)^t &= H\left(f(s)x + g(s)\right)^t = H\left(f(t)[f(s)x + g(s)] + g(t)\right) \\
&= H(f(t)f(s)x + f(t)g(s) + g(t)).
\end{aligned}$$

Ha összevetjük az eredményeket, azt kapjuk, hogy

$$f(ts) = f(t)f(s), \tag{9}$$

$$g(ts) = f(t)g(s) + g(t) = f(s)g(t) + g(s). \tag{10}$$

Mivel $s, t > 0$, a (9) egyenlőségbe helyettesítsünk be a $t = e^x$, $s = e^y$ értékeket. Ekkor kapjuk:

$$\begin{aligned}
f(e^x e^y) &= f(e^x)f(e^y) \\
f(e^{x+y}) &= f(e^x)f(e^y) \\
\log(f(\exp(x+y))) &= \log(f(\exp(x))f(\exp(y))) \\
\log(f(\exp(x+y))) &= \log(f(\exp(x))) + \log(f(\exp(y)))
\end{aligned}$$

Ez éppen a Cauchy-féle függvényegyenlet $\log \circ f \circ \exp$ függvényre. Ismert, hogy x változó esetén ennek a megoldásai cx alakúak, ha $c \in \mathbb{R}$. Írjuk át a konstansot $c = -\kappa$ módon.

$$\begin{aligned}
\log(f(\exp(x))) &= -\kappa x, \qquad t = \exp(x) \\
\log(f(t)) &= -\kappa \log(t) \\
f(t) &= e^{-\kappa \log t} = t^{-\kappa}
\end{aligned}$$

Azaz a (9) egyenletnek minden lehetséges megoldása $f(t) = t^{-\kappa}$ alakú lesz.

A továbbiakban három eset lehetséges, $\kappa = 0$, $\kappa > 0$, vagy $\kappa < 0$. Vizsgáljuk először a $\kappa = 0$ esetet. Ekkor $f(t) = 1$ és a (10) egyenletből $g(ts) = g(t) + g(s)$ összefüggés adódik. Vegyük az egyenlet exponenciálisát. Ekkor $e^{g(ts)} = e^{g(t)}e^{g(s)}$ alakban szintén visszavezethető a Cauchy-egyenlőségre, melynek megoldását $e^{g(t)} = t^{-d}$ alakban adjuk meg, másképpen $g(t) = -d\ln(t)$, ahol $t > 0$, $d \in \mathbb{R}$. Ezek alapján a (8) egyenlet felírható a következő alakban:

$$H^t(x) = (x - d\ln t). \tag{11}$$

Látható, hogy d nem lehet 0, hiszen feltettük, hogy H nemelfajuló. Rögzített x-re $H^t(x)$ nem növekvő függvény t-ben, tehát $d > 0$. Tegyük fel, hogy valamely x_0-ra $H(x_0) = 1$, ekkor $H(x_0 - d\ln t) = 1$ egyenletet kapjuk minden t-re. Ebből következik, hogy $H(u) = 1$ minden u-ra, ami ellentmondás.

Tehát $H(x) < 1$ minden x-re. Hasonlóan megmutatható, hogy $H(x) > 0$ minden x-re. A (11) egyenletből így következik, hogy

$$H^t(0) = H(-d\ln t), \qquad t > 0.$$

Válasszunk egy p-t úgy, hogy $e^{-e^{-p}} = H(0) \in (0,1)$ teljesüljön, valamint legyen $u = -d\ln t$. Ekkor a fenti egyenlet a következőképpen alakul

$$H(u) = e^{-e^{-p}t} = e^{-e^{-(u/d+p)}} = \Lambda(u/d + p),$$

ami a Gumbel eloszlás típusába tartozik.

Most vizsgáljuk a $\kappa > 0$ esetet. A (10) egyenlet átalakításával az kapjuk, hogy

$$\frac{g(s)}{1-f(s)} = \frac{g(t)}{1-f(t)}.$$

Tehát $g(s)/(1-f(s)) = c$ bármely $f(s) \neq 1$, azaz $s \neq 1$ esetén, amiből kapjuk, hogy

$$g(t) = g(s)\frac{1-f(t)}{1-f(s)} = c(1-t^{-\kappa}).$$

Ebből következik, hogy

$$H^t(x) = H(t^{-\kappa}x + c(1-t^{-\kappa})) = H(t^{-\kappa}(x-c)+c).$$

Legyen $J(x) = H(x+c)$ függvény. Az így definiált függvény szintén nemelfajuló, ugyanolyan típusú, mint H és teljesül rá a következő egyenlőség:

$$J^t(x) = J(t^{-\kappa}x).$$

Legyen $x = 0$ és vegyük az egyenlőség logaritmusát. Ekkor $t \log J(0) = J(0)$ egyenletben $\log J(0)$ csak 0 vagy $-\infty$ lehet, azaz $J(0)$ a 0 és 1 értékek egyikét veszi fel. Nem lehet 1 a $J(0)$ értéke, mert $x < 0$ esetén a bal oldal csökkenő, a jobb oldal pedig növekvő, ha a t változót növeljük. Tehát $J(0) = 0$.

Most vizsgáljuk az $x = 1$ esetet, amely szerint $J^t(1) = J(t^{-\kappa})$. Ha $J(1)$ a 0 vagy 1 értéket veszi fel, akkor J konstans, ami ellentmond annak, hogy nemelfajuló. Vagyis $J(1) \in (0,1)$. Legyen $\kappa^{-1} = \alpha$. $J(1) = \exp(-p^{-\alpha})$, ahol p paraméter, továbbá $u = t^{-\kappa}$. Átalakítva $u^{-\alpha} = t$. Ekkor azt kapjuk, hogy

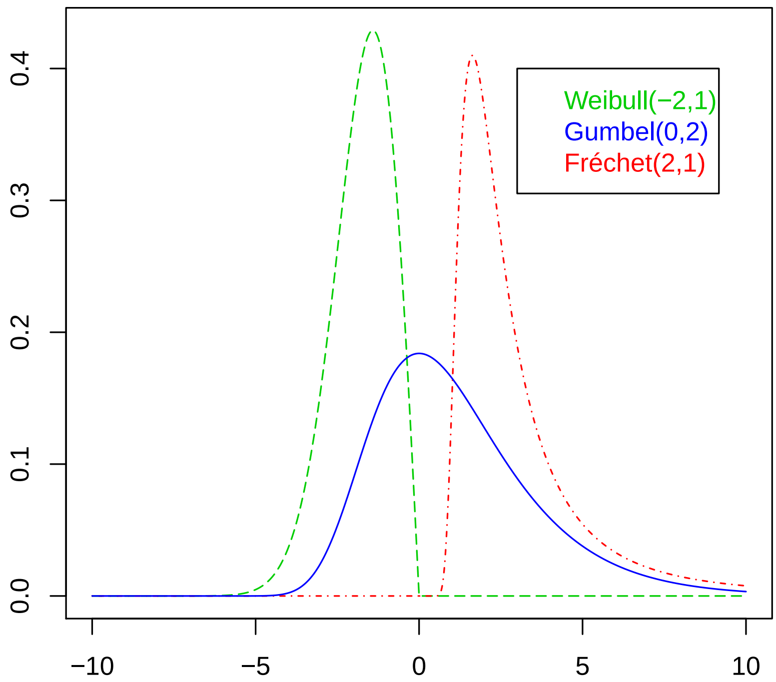
$$J(u) = J^t(1) = (\exp(-p^{-\alpha}))^t = \exp(-p^{-\alpha}t) = \exp(-(pu)^{-\alpha}) = \Phi_\alpha(pu).$$

Ez éppen a Fréchet eloszlás. A Weibull eloszlást hasonló módon kaphatjuk meg. Az összes esetet megvizsgáltuk, tehát valóban nem lehet más a határeloszlás. $\square$

Ez a tétel leszűkíti a lehetséges határeloszlásokat három esetre. E három lehetséges határeloszlás sűrűségfüggvénye között a legnagyobb különbséget az jelenti, hogy milyen számokon vesz fel pozitív értéket. A Fréchet eloszlás tartója ugyanis a pozitív számok, a Weibull típusúé a negatív számok, a Gumbel eloszlás tartója pedig az egész $\mathbb{R}$. A továbbiakban a jelölések megkönnyítése végett, következzen egy definíció.

2.5. Definíció. *A Φ_α, Ψ_α és Λ eloszlásfüggvényeket standard extrém érték eloszlásfüggvényeknek hívjuk.*

Érdekesség, hogy habár ez a három eloszlás eléggé eltérő, matematikailag mégis szoros kapcsolat van köztük. Legyen X egy Φ_α típusú eloszlás. Ekkor teljesül, hogy $-X^{-1} \sim \Psi_\alpha$, valamint $\ln X^\alpha \sim \Lambda$.



1. ábra. Az extrém érték eloszlások sűrűségfüggvényei, balról jobbra, egyre nagyobb mediánokkal

3. Vonzási tartomány

3.1. Általánosított extrém érték eloszlás

Az előző fejezetben beláttuk, hogy ha létezik a maximumoknak határeloszlása, akkor az milyen típusú lehet. Most tegyük fel egy tetszőleges eloszlásról, hogy belőle képzett valószínűségi változók sorozatának a maximumához létezik nemelfajuló határeloszlás. A következőkben azt fogjuk vizsgálni, hogy hogyan dönthető el a határeloszlás típusa.

A következő tétel egy szükséges és elégséges feltételt ad a határeloszlás létezésére.

3.1. Tétel. *Legyen F eloszlásfüggvény és ennek jobb oldali végpontja $x_F \leq \infty$ és legyen $\tau \in (0, \infty)$. Akkor, és csak akkor létezik olyan (u_n) sorozat, amely kielégíti $n\overline{F}(u_n) \to \tau$ összefüggést, ha*

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x)}{\overline{F}(x-0)} = 1$$

és $F(x_F) = 1$. Ez ekvivalens azzal, hogy

$$\lim_{x \uparrow x_F} \frac{p(x)}{\overline{F}(x-0)} = 0,$$

ahol $p(x) = F(x) - F(x-0)$.

Bizonyítás. A bizonyítás az [5] cikkben található. Legyen $0 < \tau < \infty$. Az első irányt indirekt bizonyítjuk, azaz tegyük fel, hogy létezik olyan $\epsilon > 0$, hogy bármely x_n sorozat esetén, ha $x_n \to x_F$, akkor

$$p(x_n) \geq 2\epsilon(1 - F(x_n - 0)).$$

Vegyünk egy olyan n_j sorozatot, amelyre $F(x_j) = 1 - \tau/n_j$ nagyságrendű, azaz:

$$1 - \frac{\tau}{n_j} \leq \frac{F(x_j - 0) + F(x_j)}{2} \leq 1 - \frac{\tau}{n_j + 1}.$$

Két eset lehetséges: $u_{n_j} < x_j$ teljesül végtelen sok j esetén, vagy $u_{n_j} \geq x_j$. Most először az első esetet vizsgáljuk meg.

$$n_j(1 - F(u_{n_j})) \geq n_j(1 - F(x_j - 0)).$$

Erre fel tudunk írni egy alsó becslést is

$$n_j(1 - F(x_j - 0)) = \tau + n_j \left(1 - \frac{\tau}{n_j} - \frac{F(x_j - 0) + F(x_j)}{2} + \frac{p(x_j)}{2}\right) \geq$$
$$\tau + \frac{n_j p(x_j)}{2} - n_j \left(\frac{\tau}{n_j} - \frac{\tau}{n_j - 1}\right) \geq \tau + \epsilon n_j (1 - F(x_j - 0)) - \frac{\tau}{n_j + 1}.$$

Ebből átrendezve az következik, hogy

$$(1-\epsilon)n_j(1-F(x_j-0)) \geq \tau - \frac{\tau}{n_j+1}.$$

Ha az $1-\epsilon$ szorzót elhagyjuk, az egy felső becslést ad. Most nézzük a határértékét, ha $n_j \to \infty$.

$$\limsup n_j(1-F(x_j-0)) > \tau.$$

Mivel $x_{n_j} < x_j$ az alábbi egyenlőtlenség is fennáll:

$$\limsup n_j(1-F(x_{u_j})) > \tau.$$

Ellentmondást kaptunk, hiszen feltettük, hogy $n\overline{F(u_n)} \sim \tau$. A $x_{n_j} \leq x_j$ esetet hasonlóan lehet bizonyítani.

Most jöjjön a másik irány. $F(u_n-0) \leq 1-\tau/n \leq F(u_n)$ ahol $u_n = F^{-1}(1-\tau/n)$. Ebből kapjuk, hogy

$$\frac{\overline{F}(x)}{\overline{F}(x-0)}\tau \leq n(1-F(u_n)) \leq \tau,$$

mivel $u_n \to x_F$, ha $n \to \infty$. $\square$

3.2. Definíció (Max-vonzási tartomány)**.** *Legyen az X valószínűségi változó. Azt mondjuk, hogy X a H extrém érték eloszlás vonzási tartományába tartozik, ha léteznek $a_n > 0$, $b_n \in \mathbb{R}$ konstansok úgy, hogy*

$$\frac{M_n - b_n}{a_n} \to H$$

teljesüljön. Ez a továbbiakban $X \in MDA(H)$ módon jelöljük (MDA $\sim$ Maximum Domain of Attraction).

Gyakran mondjuk egy F eloszlásfüggvényre is, hogy a H vonzási tartományába tartozik. Ekkor a fenti definíció érvényes arra a valószínűségi változóra, melynek F az eloszlásfüggvénye.

A max-vonzási tartomány tehát azon eloszlásokat foglalja magába, melyeknek az alkalmasan standardizált maximuma egy adott eloszláshoz fog tartani. Ezt a kicsit nehezen megfogható tulajdonságot a következő tétel segítségével könnyebb meghatározni.

3.3. Tétel (Max-vonzási tartomány jellemzése)**.** *Legyen az X egy valószínűségi változó F eloszlásfüggvénnyel. Ez egy H extrém érték eloszlás vonzási tartományába tartozik akkor és csak akkor, ha léteznek olyan $a_n > 0$ és $b_n \in \mathbb{R}$ normáló konstansok, hogy*

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

$H(x) = 0$ esetén a határértéket ∞-ként értelmezzük.

Bizonyítás. Először legyen $\lim n\overline{F}(a_nx+b_n) = -\ln H(x)$, ha $n \to \infty$ és $x \in \mathbb{R}$. A Poisson approximációs tétel (2.1.) miatt $\mathbb{P}(M_n \leq a_nx+b_n) \to H(x)$ minden $x \in \mathbb{R}$, azaz megkaptuk, hogy $F \in MDA(H)$.

Most legyen $F \in MDA(H)$. Ez azt jelenti, hogy $\mathbb{P}(M_n \leq a_nx+b_n) \to H(x)$ a H függvény minden folytonossági pontjában. Viszont mivel H extrém érték eloszlás, minden $x \in \mathbb{R}$ pontban folytonos. Rögzített x esetén vizsgáljuk $F(a_nx+b_n)^n \to H(x)$ összefüggést.

Ha $H(x) \neq 0$, akkor logaritmust véve azt kapjuk, hogy $-n\ln(1-\overline{F}(a_nx+b_n)) \to -\ln H(x)$, ahol $H(x) \in (0,\infty)$.Ebből következik, hogy $\overline{F}(a_nx+b_n) \to 0$. Mivel $\ln(1-y) \sim -y$, ha $y \to 0$ összefüggés teljesül, $n\overline{F}(a_nx+b_n) \sim -n\ln(1-\overline{F}(a_nx+b_n)) \to -\ln H(x)$. Azaz az állítás fennáll.

Ha $H(x) = 0$, akkor az $F(a_nx+b_n)^n = (1-(n\overline{F}(a_nx+b_n))/n)^n \to 0$, valamint az $(1-L/n)^n \to e^{-L}$ összefüggésekből következik, hogy minden $L > 0$ esetén létezik olyan n_0 szám, hogy bármely $n > n_0$ esetén $n\overline{F}(a_nx+b_n) > L$. Ekkor a sorozat bármely valós számnál nagyobb lehet, tehát $n\overline{F}(a_nx+b_n) \to \infty$, ami $-\ln H(x)$ határértéke. □

Ha ismerjük a normáló konstansokat, akkor már el tudjuk dönteni egy eloszlásról, hogy melyik extrém érték eloszlás vonzási tartományába esik.

Az extrém érték eloszlásokat általánosan is lehet kezelni. Ez az általánosabb fogalom segíthet a későbbiekben, hogy a megfelelő normáló konstansok ismerete nélkül is meg lehessen határozni egy határeloszlás típusát.

3.4. Definíció (Általánosított extrém érték eloszlás)**.** *A Gumbel, Fréchet és Weibull eloszlásokat általánosítja a következő, három paramétert tartalmazó eloszlásfüggvény:*

$$F_{\mu,\sigma,\gamma}(x) = \begin{cases} \exp\{-\{1+\gamma(\frac{x-\mu}{\sigma})\}^{-1/\gamma}\}, & ha \quad \gamma \neq 0, \\ \exp\{-\exp\{-\frac{x-\mu}{\sigma}\}\} & ha \quad \gamma = 0. \end{cases}$$

Az ilyen eloszlásfüggvénnyel rendelkező eloszlásokat nevezzük általánosított extrém érték eloszlásnak (GEV $\sim$ Generalized Extreme Value distribution).

3.5. Definíció (Standard általánosított extrém érték eloszlás)**.** *Standardizált esetben a GEV függvény paramétereire igaz, hogy $\mu = 0$ és $\sigma = 1$. Ekkor a $H_\gamma(x)$ eloszlásfüggvény a következő módon alakul:*

$$H_\gamma(x) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}}, & ha \quad \gamma \neq 0, \\ e^{-e^{-x}}, & ha \quad \gamma = 0, \end{cases}$$

ahol $1 + \gamma x \geq 0$. *A* H_γ *sűrűségfüggvényének tartója határozza meg, hogy milyen családba tartozik az eloszlás. Ez a tartó* γ *függvényében az alábbi módon alakul*

$$\begin{aligned} x > -1/\gamma, &\qquad ha \qquad \gamma > 0, \qquad \textit{ekkor Fréchet,} \\ x < -1/\gamma, &\qquad ha \qquad \gamma < 0, \qquad \textit{ekkor Weibull,} \\ x \in \mathbb{R}, &\qquad ha \qquad \gamma = 0 \qquad \textit{esetben pedig Gumbel.} \end{aligned}$$

Az így definiált eloszlást standard általánosított extrém érték eloszlásnak hívjuk.

Az így definiált eloszlás valóban alkalmas az extrém érték eloszlások általános leírására. Megfelelő paramétereket választva a Weibull, Fréchet és Gumbel eloszlások egyikét kapjuk. Ezt bizonyítja a következő tétel.

3.6. Tétel (MDA(H_γ) jellemzése)**.**

i.) $X \in MDA(H_\gamma)$, *ahol* $\gamma > 0$, *akkor és csak akkor, ha* $X \in MDA(\Phi_\alpha)$, *ahol* $\alpha = 1/\gamma > 0$.

ii.) $X \in MDA(H_\gamma)$, *ahol* $\gamma = 0$, *akkor és csak akkor, ha* $X \in MDA(\Lambda)$.

iii.) $X \in MDA(H_\gamma)$, *ahol* $\gamma < 0$, *akkor és csak akkor, ha* $X \in MDA(\Psi_\alpha)$, *ahol* $\alpha = -1/\gamma > 0$.

Bizonyítás. Először legyen $\gamma = 0$. Ekkor $H_0 = e^{-e^{-x}} = \Lambda$ a definíció alapján, tehát $\gamma = 0$ esetén a Gumbel eloszlást kapjuk.

Igazoljuk az első állítást. Tegyük fel, hogy $\gamma > 0$ rögzített és $X \in MDA(H_\gamma)$, azaz

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq a_n x + b_n) = \begin{cases} 0, & ha \quad x \leq -\frac{1}{\gamma}, \\ e^{-(1+\gamma x)^{-1/\gamma}}, & ha \quad x > -\frac{1}{\gamma}. \end{cases}$$

A bizonyítás során úgy választunk új konstansokat, hogy a 2.3. tétel értelmében a határeloszlás ugyanolyan típusú maradjon. Alakítsuk át tehát a konstansokat az alábbi módon:

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq (a_n/\gamma)x + b_n - (a_n/\gamma)) = \lim_{n\to\infty} \mathbb{P}(M_n \leq (a_n/\gamma)(x-1) + b_n) =$$
$$= \begin{cases} 0, & ha \;\; x \leq 0, \\ e^{-(1+\gamma(x-1)/\gamma)^{-1/\gamma}} = e^{-x^{-1/\gamma}}, & ha \;\; x > 0. \end{cases}$$

Megkaptuk a Fréchet eloszlást $1/\gamma$ paraméterrel, azaz $X \in MDA(\Phi_{1/\gamma}) = MDA(\Phi_\alpha)$. A másik irány belátásához tegyük fel, hogy $\alpha > 0$ és $X \in MDA(\Phi_\alpha)$. Ekkor

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq a_n x + b_n) = \begin{cases} 0, & ha \quad x \leq 0, \\ e^{-x^{-\alpha}}, & ha \quad x > 0. \end{cases}$$

Ismét válasszunk új konstansokat, melyek legyenek $a_n\gamma$ és $a_n + b_n$. Ekkor

$$\lim_{n\to\infty} \mathbb{P}(M_n \le a_n(1+\gamma x) + b_n) = \begin{cases} 0, & ha \quad x \le -\gamma^{-1}, \\ e^{-(1+\gamma x)^{-1/\gamma}}, & ha \quad x > -\gamma^{-1}. \end{cases}$$

Tehát $F \in MDA(H_\gamma)$.

A harmadik állítás igazolása hasonlóan történik. Legyen $\gamma < 0$ rögzített és tegyük fel, hogy $X \in MDA(H_\gamma)$.

$$\lim_{n\to\infty} \mathbb{P}(M_n \le a_n x + b_n) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}}, & ha \quad x \le -\frac{1}{\gamma}, \\ 1, & ha \quad x > -\frac{1}{\gamma}, \end{cases}$$

egyenlőségből kiindulva, vegyünk új konstansokat, a_n helyett $(-a_n)/\gamma$ valamint b_n helyett $(b_n - a_n)/\gamma$ szerepeljen. Ekkor

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(M_n \le (-a_n/\gamma)x + b_n - (a_n/\gamma)) &= \lim_{n\to\infty} \mathbb{P}(M_n \le (a_n/\gamma)(-x-1) + b_n) \\ &= \begin{cases} e^{-(1+\gamma(-x-1)/\gamma)^{-1/\gamma}} = e^{-(-x)^{-1/\gamma}}, & ha \ x \le 0, \\ 1, & ha \ x > 0, \end{cases} \end{aligned}$$

ami a Weibull eloszlás $-1/\gamma$ paraméterrel, tehát $X \in MDA(\Psi_{-1/\gamma}) = MDA(\Psi_\alpha)$.

A fordított irány belátásához tegyük fel, hogy $\alpha < 0$ és $X \in MDA(\Psi_\alpha)$. Ekkor

$$\lim_{n\to\infty} \mathbb{P}(M_n \le a_n x + b_n) = \begin{cases} e^{-(-x)^\alpha}, & ha \quad x \le 0 \\ 1, & ha \quad x > 0. \end{cases}$$

Most is válasszunk új konstansokat a következő módon. Az a_n helyett legyen $-a_n\gamma$ és b_n helyett $b_n - a_n$. Ekkor

$$\lim_{n\to\infty} \mathbb{P}(M_n \le a_n(-1-\gamma x) + b_n) = \begin{cases} e^{-(1+\gamma x)^{-1/\gamma}}, & ha \quad x \le -\gamma^{-1} \\ 1, & ha \quad x > -\gamma^{-1}. \end{cases}$$

Tehát $F \in MDA(H_\gamma)$. □

3.2. Extrém érték eloszlások vonzási tartománya

Az eloszlásfüggvény viselkedése a jobb oldali végpont közelében meghatározza a maximumok viselkedését, ezért érdemes ezek vizsgálatával folytatni. Vezessünk be három definíciót.

3.7. Definíció (Reguláris változású függvény)**.** *Legyen* $h : (0, \infty) \to (0, \infty)$ *mérhető függvény. Ezt reguláris változásúnak nevezzük* ∞*-ben az* $\eta \in \mathbb{R}$ *index-szel, ha*

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

minden $x > 0$ *számra. Továbbiakban* $h \in RV_{\eta}$ *módon jelöljük.*

1. Példa. *Legyen* $F(x) = 1 - x^{-\alpha}$*, ahol* $x \geq 1$ *és* $\alpha > 0$*. Ekkor*

$$\frac{\overline{F}(tx)}{\overline{F}(t)} = \frac{(tx)^{-\alpha}}{t^{-\alpha}} = x^{-\alpha}$$

azaz $\overline{F} \in RV_{-\alpha}$*.*

3.8. Definíció (Lassú változású függvény)**.** *Legyen* $h : (0, \infty) \to (0, \infty)$ *reguláris változású függvény. Ha* $\eta = 0$*, azaz*

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

akkor a h *függvényt lassú változásúnak mondjuk. Továbbiakban* $h \in RV_0$ *módon jelöljük.*

2. Példa. *Legyen adott* $b \in \mathbb{R}$*, ekkor* $L(x) = \log^b(x)$ *függvény lassú változású, hiszen*

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} = \lim_{t \to \infty} \frac{\log^b(tx)}{\log^b(t)} = \lim_{t \to \infty} \left(\frac{\log(t) + \log(x)}{\log(t)} \right)^b = \lim_{t \to \infty} \left(1 + \frac{\log(x)}{\log(t)} \right)^b = 1.$$

Lassú változású még minden olyan függvény is, amelynek létezik $c \in (0, \infty)$ határértéke a ∞-ben.

3.9. Definíció (Gyors változású függvény)**.** *Legyen* $h : (0, \infty) \to (0, \infty)$ *mérhető függvény. Ha minden* $x \in \mathbb{R}_+$ *esetén*

$$\lim_{t \to \infty} \frac{h(tx)}{h(t)} = \begin{cases} 0, & ha \quad x > 1, \\ 1, & ha \quad x = 1, \\ \infty, & ha \quad 0 < x < 1, \end{cases}$$

akkor h *függvényt gyors változásúnak mondjuk.*

3. Példa. *Gyors változású függvényre egy példa az* $f(x) = e^{-x}$*, ahol* $x \in (0, \infty)$*. Leellenőrizve a határértéket*

$$\lim_{t \to \infty} \frac{f(tx)}{f(t)} = \lim_{t \to \infty} \frac{e^{-tx}}{e^{-t}} = \lim_{t \to \infty} e^{t(1-x)} = \begin{cases} 0, & ha \quad x > 1, \\ 1, & ha \quad x = 1, \\ \infty, & ha \quad 0 < x < 1. \end{cases}$$

Megjegyezzük, hogy ha $h \in RV_\eta$, akkor $h(x)/x^\eta \in RV_0$, $x > 0$ esetén. Ha $L(x)$-nek nevezzük az így kapott lassú változású függvényt, átalakítva $h(x) = x^\eta L(x)$ egyenlőséget kapjuk. Mivel h és x tetszőleges volt, bármely reguláris változású függvény előáll egy lassú változású függvény, valamint egy hatvány szorzataként.

3.10. Tétel (Lassú változású függvények reprezentációja)**.** *Egy L függvény esetén $L \in RV_0$ akkor és csak akkor teljesül, ha léteznek $c, \delta : (0,\infty) \to (0,\infty)$, $\lim_{x\to\infty} c(x) = c$, $\lim_{x\to\infty} \delta(x) = 0$ függvények, melyekre teljesül, hogy*

$$L(x) = c(x)\exp\left(\int_1^x \frac{\delta(t)}{t}dt\right), \quad x > 0.$$

Bizonyítás. Az első irány megtalálható Soulier [6] cikkének 4-6. oldalán. A másik irányt bizonyítom.

$\Leftarrow$ Vizsgáljuk meg a $\lim_{s\to\infty}(L(sx)/L(s))$ határértéket, legyen $x \geq 1$.

$$\begin{aligned}\lim_{s\to\infty}\frac{L(sx)}{L(s)} &= \lim_{s\to\infty}\frac{c(sx)\exp\left(\int_1^{sx}\frac{\delta(t)}{t}dt\right)}{c(s)\exp\left(\int_1^{s}\frac{\delta(t)}{t}dt\right)} = \lim_{s\to\infty}\frac{c(sx)\exp\left(\int_s^{sx}\frac{\delta(t)}{t}dt\right)}{c(s)} \\ &= \lim_{s\to\infty}\frac{c(sx)}{c(s)} = \frac{\lim_{s\to\infty}c(sx)}{\lim_{s\to\infty}c(s)} = 1.\end{aligned}$$

Ha $x < 1$, akkor,

$$\lim_{s\to\infty}\frac{L(sx)}{L(s)} = \lim_{s\to\infty}\frac{c(sx)\exp\left(\int_{sx}^{s}\frac{\delta(t)}{t}dt\right)}{c(s)} = \lim_{s\to\infty}\frac{c(sx)}{c(s)} = \frac{\lim_{s\to\infty}c(sx)}{\lim_{s\to\infty}c(s)} = 1.$$

Azaz $h(x) \in RV_0$.

□

3.11. Tétel (Reguláris változású függvények reprezentációja)**.** *Egy h függvény esetén $h \in RV_\eta$ akkor és csak akkor, ha léteznek $c, \Omega : (0,\infty) \to (0,\infty)$, $\lim_{x\to\infty} c(x) = c$, $\lim_{x\to\infty} \Omega(x) = \eta$ függvények, melyekre teljesül, hogy*

$$h(x) = c(x)\exp\left(\int_1^x \frac{\Omega(t)}{t}dt\right), \quad x > 0.$$

Bizonyítás. Mivel $h(x) = L(x)x^\eta$ valamely L lassú változású függvényre, a fenti egyenlet átírható

$$L(x)x^\eta = c(x)\exp\left(\int_1^x \frac{\Omega(t)}{t}dt\right)$$

alakba. Ha leosztunk az $x^\eta \neq 0$ tényezővel, akkor

$$\begin{aligned} L(x) = c(x)x^{-\eta}\exp\left(\int_1^x \frac{\Omega(t)}{t}dt\right) &= c(x)\exp(-\eta\ln x)\exp\left(\int_1^x \frac{\Omega(t)}{t}dt\right) \\ &= c(x)\exp\left(-\eta\ln x + \int_1^x \frac{\Omega(t)}{t}dt\right) \\ &= c(x)\exp\left(\int_1^x \frac{-\eta}{t}dt + \int_1^x \frac{\Omega(t)}{t}dt\right) \\ &= c(x)\exp\left(\int_1^x \frac{\Omega(t)-\eta}{t}dt\right) \end{aligned}$$

egyenlőséget kapjuk, ahol $\lim_{x\to\infty}(\Omega(t)-\eta) = 0$, azaz éppen visszakaptuk a lassú változású esetre vonatkozó tételt. $\square$

3.12. Definíció. *Egy nemnegatív valószínűségi változót reguláris változásúnak hívunk, ha a túlélésfüggvényére teljesül, hogy $\overline{F} \in RV_{-\alpha}$ valamely $\alpha > 0$ számra. Ha $\alpha = 0$, akkor lassú változású.*

3.13. Tétel (Φ_α max-vonzási tartománya)**.** *Egy X valószínűségi változó Φ_α $(\alpha > 0)$ vonzási tartományába tartozik, akkor és csak akkor, ha a hozzá tartozó eloszlásfüggvényre $\overline{F} \in RV_{-\alpha}$. Ha $F \in MDA(\Phi_\alpha)$, és $a_n = F^{\leftarrow}(1-1/n)$, akkor az*

$$\frac{M_n}{a_n} \to \Phi_\alpha.$$

Bizonyítás. Az $\overline{F} \in RV_{-\alpha} \Rightarrow F \in MDA(\Phi_\alpha)$ irányt bizonyítom, a másik irány megtalálható Resnick [2] könyvében az 55-57. oldalakon. Legyen $\overline{F} \in RV_{-\alpha}$ és $a_n = F^{\leftarrow}(1-1/n)$. Ekkor $\overline{F}(a_n) \sim 1/n$, ha $n \to \infty$ és $x > 0$, valamint

$$n\overline{F}(a_n x) \sim \frac{\overline{F}(a_n x)}{\overline{F}(a_n)} \to x^{-\alpha},$$

ha $n \to \infty$ és $a_n \to \infty$. Mivel $x > 0$,

$$n\ln(1-\overline{F}(a_n x)) \sim -n\overline{F}(a_n x) \to -x^{-\alpha}.$$

Tehát $x > 0$ esetben

$$\mathbb{P}(M_n \leq a_n x) = F(a_n x)^n = \exp\{n\ln(1-\overline{F}(a_n x))\} \to e^{-x^{-\alpha}} = \Phi_\alpha(x).$$

Ha pedig $x \leq 0$, akkor az $\overline{F} \in RV_{-\alpha} \Rightarrow F(0) < 1$ összefüggést kihasználva $F(a_n x)^n \leq F(0)^n \to 0 = \Phi_\alpha(x)$. Megkaptuk a Fréchet eloszlást. $\square$

3.14. Tétel (Ψ_α max-vonzási tartománya)**.** *Egy X valószínűségi változó Ψ_α $(\alpha > 0)$ vonzási tartományába tartozik, akkor és csak akkor, ha $x_F < \infty$ és a hozzá tartozó eloszlásfüggvényre $\overline{F}(x_F - 1/x) \in RV_{-\alpha}$. Továbbá, ha $F \in MDA(\Psi_\alpha)$, és $a_n = x_F - F^{\leftarrow}(1 - 1/n)$, akkor*

$$\frac{M_n - x_F}{a_n} \to \Psi_\alpha.$$

Bizonyítás. Az $\overline{F}(x_F - 1/x) \in RV_{-\alpha} \Rightarrow F \in MDA(\Psi_\alpha)$ irányt bizonyítom, a másik irány megtalálható Resnick [2] könyvében a 60-62. oldalon. Legyen F_* a következő függvény:

$$F_*(x) = \begin{cases} 0, & ha \quad x < 0, \\ F(x_F - 1/x), & ha \quad x \geq 0. \end{cases}$$

Ekkor $\overline{F}_* \in RV_{-\alpha}$ és a 3.13. tétel miatt $a_{*n} = F_*^{\leftarrow}(1 - 1/n)$ konstansokkal teljesül $F_*(a_{*n}x)^n \to \Phi_\alpha(x)$, ha $x > 0$. Átalakítva ezt a függvénysorozatot kapjuk, hogy $F(x_F - 1/(a_{*n}x))^n \to \exp\{-x^{-\alpha}\}$, ha x >0. Használjuk az $x = -1/y$ jelölést. Ekkor $F(x_F + 1/a_{*n}x)^n \to \exp\{-(-y)^\alpha\}$, ha $y < 0$ összefüggést kapjuk. Tudjuk, hogy

$$\begin{aligned} a_{*n} &= F_*^{\leftarrow}(1 - 1/n) = \inf\{u : F_*(u) \geq 1 - 1/n\} \\ &= \inf\{u : F(x_F - 1/u) \geq 1 - 1/n\} = (\sup\{s : F(x_F - s) \geq 1 - 1/n\})^{-1} \\ &= (x_F - \inf\{u : F(u) \geq 1 - 1/n\})^{-1} = (x_F - F^{\leftarrow}(1 - 1/n))^{-1} \end{aligned}$$

Ebből adódik, hogy $a_n = 1/a_{*n} = x_F - F^{\leftarrow}(1 - 1/n)$, valamint

$$F^n(x_F + a_n y) \to \Psi_\alpha(y), \quad y < 0,$$

vagy másképp

$$\mathbb{P}\left(\frac{M_n - x_F}{a_n} \leq y\right) \to \Psi_\alpha(y), \quad y < 0,$$

ami éppen a bizonyítandó állítás. □

Most már el tudjuk dönteni egy eloszlásról, hogy egy belőle vett minta maximumainak van-e határeloszlása. Ha tudjuk, hogy létezik, az előbbi tételek segítségével megvizsgálható, hogy a Weibull vagy a Fréchet-féle eloszláshoz tart, sőt, a normáló konstans sorozatokat is megkaphatjuk. A Gumbel típushoz tartó eloszlásokat nem lehet egy reguláris változású függvény segítségével jellemezni, mert sokszor egy gyors változású függvényről van szó. Nagyon sokféle függvény lehet, és a jobb végpont felvehet valós értéket, vagy tarthat a végtelenhez. Zárt forma helyett vizsgáljuk a szükséges és elégséges feltételeket.

3.15. Tétel ($MDA(\Lambda)$ jellemzése I.)**.** *Egy X valószínűségi változó Λ max-vonzási tartományába esik, akkor és csak akkor, ha létezik olyan pozitív, mérhető $a(.)$ függvény, melyre*

$$\lim_{n\to\infty} \frac{\overline{F}(x+ta(x))}{\overline{F}(x)} = e^{-t}, \quad t \in \mathbb{R}.$$

3.16. Tétel ($MDA(\Lambda)$ jellemzése II.)**.** *Legyen az X valószínűségi változó és F a hozzá tartozó eloszlásfüggvény tetszőleges jobb végponttal. Ekkor az F függvény a Λ max-vonzási tartományába tartozik akkor és csak akkor, ha létezik egy $w < x_F$ szám, melyre a túlélésfüggvény felírható*

$$\overline{F}(x) = c(x) \exp\left(\int_w^{x_F} \frac{g(t)}{a(t)} dt \right)$$

alakban, ahol c és g olyan mérhető függvények, melyekre $\lim c(x) = c > 0, c \in \mathbb{R}$ *valamint* $\lim g(x) = 1$*, ha $x \to x_F$, továbbá a egy olyan pozitív, abszolút folytonos függvény, melyre igaz, hogy* $\lim_{x\to x_F} a'(x) = 0$*. Egy ilyen lehetséges a függvény a feltételes várhatóérték, azaz*

$$a(x) = \mathbb{E}(X - x | X > x) = \int_x^{x_F} \frac{\overline{F}(x)}{\overline{F}(t)} dt, \quad x < x_F.$$

A [4] cikk a 23. oldalon bővebben tárgyalja $MDA(\Lambda)$ lehetséges jellemzéseit és több példát is említ az egyes vonzási tartományokba tartozó eloszlásokra.

4. Küszöb feletti értékek

Ebben a fejezetben kicsit más oldalról próbáljuk megközelíteni a problémát. Az éves vízállások maximuma csak a legnagyobb árvizeket tartalmazza. Lehet, hogy egy évben több árvíz is volt, lehet, hogy valamikor egy sem. Másik közelítése a magas értékek vizsgálatának, ha egy küszöb felett vizsgáljuk a vízállások értékeit. Azt, hogy mi a jó küszöb, matematikailag nem lehet jól megfogalmazni, de ettől függetlenül fel lehet rá állítani egy működőképes modellt. Előbb azonban be kell vezetünk az általánosított Pareto-eloszlást, mint a modell egyik technikai hátterét.

4.1. Definíció (Általánosított Pareto-eloszlás)**.** *Definiáljuk a $G_{\xi,\mu,\sigma}$, $\xi \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$, függvényt az alábbiak szerint:*

$$G_{\xi,\mu,\sigma}(x) = \begin{cases} 1 - (1 + \frac{\xi(x-\mu)}{\sigma})^{-1/\xi}, & ha \quad \xi \neq 0, \\ 1 - e^{-\frac{(x-\mu)}{\sigma}}, & ha \quad \xi = 0, \end{cases}$$

ahol az értelmezési tartomány

$$D(G) : \begin{cases} \mu \leq x < \infty, & ha \quad \xi \geq 0, \\ \mu \leq x \leq (\mu - \sigma)/\xi, & ha \quad \xi < 0. \end{cases}$$

Az így definiált függvényhez tartozó eloszlást általánosított Pareto-eloszlásnak hívjuk.

4.2. Definíció (Standard általánosított Pareto-eloszlás)**.** *Legyen az általánosított Pareto-eloszlásban szereplő paraméterek közül $\mu = 0$ és $\sigma = 1$. Az így kapott G_ξ függvény tehát*

$$G_\xi = \begin{cases} 1 - (1 + \xi x)^{-1/\xi}, & ha \quad \xi \neq 0, \\ 1 - e^{-x}, & ha \quad \xi = 0, \end{cases}$$

alakú, ahol

$$D(G) : \begin{cases} 0 \leq x < \infty, & ha \quad \xi \geq 0, \\ 0 \leq x \leq -1/\xi, & ha \quad \xi < 0. \end{cases}$$

Az így definiált függvényt standard általánosított Pareto-eloszlásnak hívjuk.

4.3. Tétel (MDA(H_γ) jellemzése Pareto-eloszlás segítségével)**.** *Legyen $\gamma \in \mathbb{R}$, ekkor az alábbi állítások ekvivalensek.*
i) $F \in MDA(H_\gamma)$
ii) Létezik egy olyan a pozitív, mérhető függvény, melyre

$$\lim_{u \uparrow x_F} \frac{\overline{F}(u + xa(u))}{\overline{F}(u)} = \overline{G}_\gamma(x),$$

ha x a G_γ standard általánosított Pareto eloszlásfüggvény értelmezési tartományából való.

Bizonyítás. Csak a $\gamma > 0$ esetet vizsgáljuk, mert $\gamma < 0$ hasonlóan bizonyítható, $\gamma = 0$ pedig triviális.

$(i) \Rightarrow (ii)$. Ha $x > -\gamma$ vezessük be a $\alpha = 1/\gamma$ jelölést, ekkor azt kapjuk, hogy

$$H_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\} = \exp\{-(1+x/\alpha)^{-\alpha}\} = \Phi_\alpha(1-x/\alpha).$$

Ha $x \leq -\gamma$, akkor $H_\gamma(x) = 0 = \Phi_\alpha(x)$. Ezekből a 3.14 tétel alapján következik, hogy $\overline{F} \in RV_{-\alpha}$. A 3.11 tétel ekvivalens átfogalmazást ad a reguláris változásra. Legyen $a(t) = -t/\Omega(t)$, ahol ahol $\lim_{x\to\infty} \Omega(x) = -\alpha$. Ekkor

$$\overline{F}(x) = h(x) = c(x)\exp\left(\int_1^x \frac{\Omega(t)}{t}dt\right) = c(x)\exp\left(-\int_1^x \frac{1}{a(t)}dt\right), \quad x > 0.$$

Alakítsuk át *(ii)* feltétel bal oldalát az alábbi módon:

$$\lim_{u\uparrow x_F} \frac{\overline{F}(u+xa(u))}{\overline{F}(u)} = \lim_{u\uparrow x_F} \frac{\overline{F}((1+xa(u)/u)u)}{\overline{F}(u)}.$$

Mivel $\overline{F}$ reguláris változású,

$$\lim_{u\to\infty} \frac{\overline{F}((1+xa(u)/u)u)}{\overline{F}(u)} = (1+xa(u)/u)^{-\alpha}.$$

Az $a(.)$ függvény definíciója miatt $a(u)/u = -1/\Omega(u)$ és $\lim_{u\to\infty} -1/\Omega(u) = -(1/-\alpha) = 1/\alpha$, tehát $\lim_{u\to\infty} a(u)/u = 1/\alpha$. Mivel $\gamma > 0$, a határeloszlás Fréchet típusú, tehát F jobb oldali végpontja ∞ lesz. Azaz fennáll

$$\lim_{u\to\infty} \frac{\overline{F}(u+xa(u))}{\overline{F}(u)} = (1+x/\alpha)^{-\alpha} = G_{1/\alpha} = G_\gamma.$$

$(ii) \Rightarrow (i)$. Válasszuk $u_n = F^{\leftarrow}(1-1/n)$ konstans sorozatot. Ekkor $\overline{F}(u_n) \sim 1/n$ és

$$(1+x/\alpha)^{-\alpha} = \lim_{n\to\infty} \frac{\overline{F}(u_n+xa(u_n))}{\overline{F}(u_n)} = \lim_{n\to\infty} n\overline{F}(u_n+xa(u_n))$$

egyenlőségeket kapjuk. A 3.3 tételből következik, hogy $F \in MDA(H_\gamma)$, ahol a H_γ eloszlásfüggvénye $H_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$. □

A tétel jelentősége, hogy $X \in MDA(H_\gamma)$, akkor és csak akkor lehetséges, ha létezik egy szigorúan pozitív mérhető $a(.)$ függvény, melyre

$$\lim_{u\uparrow x_F} \mathbb{P}((X-u)/a(u) > x | X > u) = \overline{G}_\gamma(x).$$

A statisztikai alkalmazásoknál ezt nem mindig könnyű kimutatni, hiszen $u \to x_F$ esetén egyre kevesebb mintaelem áll a rendelkezésünkre.

4.4. Definíció (Meghaladási eloszlásfüggvény)**.** *Legyen az X valószínűségi változó, melynek eloszlásfüggvénye F, jobb végpontja x_F. Egy $u < x_F$ számra*

$$F_u(x) = \mathbb{P}(X - u \leq x | X > u), \quad x \geq 0,$$

feltételes eloszlásfüggvényt nevezzük az X eloszlás u szint feletti meghaladási eloszlásfüggvényének.

4.5. Tétel. *Adott $\gamma \in \mathbb{R}$ esetén az alábbi állítások ekvivalensek.*
i) $F \in MDA(H_\gamma)$
ii) Létezik egy olyan $b : \mathbb{R} \to \mathbb{R}$ pozitív, mérhető függvény, melyre

$$\lim_{u \uparrow x_F} \sup_{x \in (0, x_F - u)} |F_u(x) - G_{\gamma,0,b(u)}(x)| = 0.$$

(Azaz, ha egy eloszlásból vett minta maximumaihoz létezik határeloszlás, akkor a szint feletti értékek eloszlását vizsgálva azt tapasztaljuk, hogy egy általánosított Pareto-eloszláshoz tart, ha a szint közelít az eloszlás felső végpontjához, és fordítva.)

Bizonyítás. A 4.3. tétel ii) állítása átírható

$$\lim_{u \uparrow x_F} |F_u(a(u)x) - G_{\gamma,0,0}(x)| = 0, \quad x \in D(G)$$

alakba. Mivel $G_{\gamma,0,0}$ folytonos, az előbbi kifejezés egyenletesen konvergens a $D(G)$ tartományon, azaz

$$\lim_{u \uparrow x_F} \sup_{x \in (0, x_F - u)} |F_u(a(u)x) - G_{\gamma,0,0}(x)| = 0.$$

Ebből következik $y = a(u)x$ helyettesítéssel, hogy

$$0 = \lim_{u \uparrow x_F} \sup_{x \in (0, x_F - u)} |F_u(y) - G_{\gamma,0,0}(y/a(u))| = \lim_{u \uparrow x_F} \sup_{x \in (0, x_F - u)} |F_u(y) - G_{\gamma,0,a(u)}(y)|,$$

ami éppen az ii) állítás. Vagyis a 4.3 tétel átalakításából kapjuk a bizonyítást. $\square$

5. Szimulációk

A szimulációkat az R program segítségével készítettem. Minden esetben vettem egy adott elemszámú - jelöljük n-nel - véletlen mintát a megfigyelt eloszlásból, és megkerestem a maximumát és megjegyeztettem. Ezt megismételtem elég sokszor és a kapott maximumok eloszlását kirajzoltattam egy hisztogramként a 'hist' paranccsal. Elég nagy ismétlésszám esetén ez már jól közelíti a maximumok eloszlását. Ez azonban csak az n elemre vonatkozó maximumok határeloszlását közelítette, a dolgozatban pedig $n \to \infty$ esetére adtunk képleteket. Ilyen eseteket sajnos nehéz lett volna vizsgálni és bonyolultabb statisztikai összefüggéseket igényelt volna, ezért én az n számot igyekeztem úgy megválasztani, hogy a kapott eloszlás már kellően hasonlítson egy extrém érték eloszláshoz.

Az adatsorhoz tartozó extrém érték eloszlás meghatározása volt a következő lépés. Az {ismev} csomagban található 'gev.fit' parancs megadta az adatokhoz leginkább illeszkedő extrém érték eloszlás paramétereit. Először a hely, majd a skála, végül az alakparamétert. Mivel statisztikai adatokról van szó, a Gumbel típushoz tartozó 0 értékű alakparamétert ténylegesen soha nem kaptam, ám ha az alak elég közel volt 0-hoz, akkor az eloszlást Gumbel típusúnak tekintettem. A kapott paramétereket felhasználva az {evir} csomagban található 'dgev' parancs segítségével az ábrákra rajzoltam (kék vonallal) az illesztett extrém érték eloszlást.

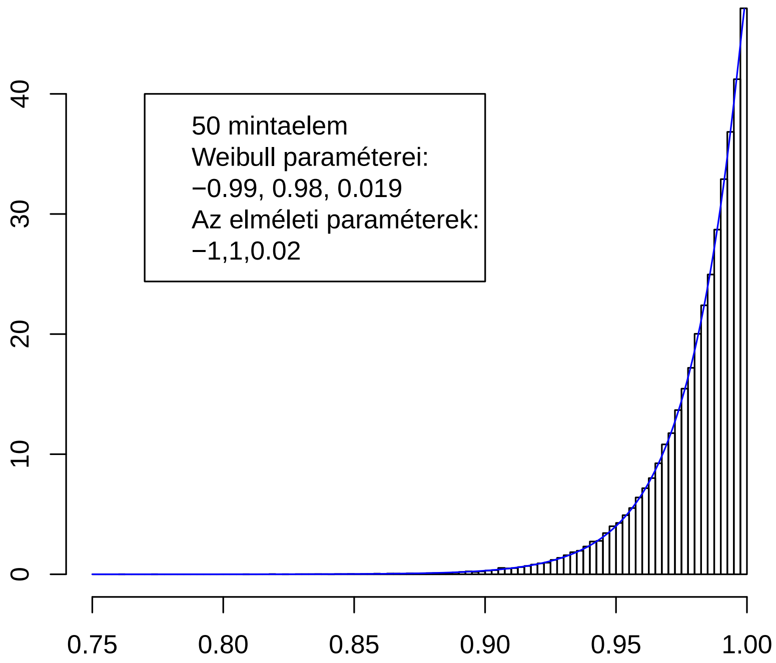
Hasonló módon jártam el a küszöb feletti szimulációk esetén is, csak ott Pareto eloszlást illesztettem az adatsorhoz a 'gpd.fit' és 'dgpd' parancsok segítségével.

Most pedig nézzünk néhány konkrét szimulációt.

5.1. Egyenletes eloszlás (0,1) intervallumon

Legyen az $X \sim E(0,1)$ valószínűségi változó. $X_1, \dots, X_{50}$ független, X változóval azonos eloszlásúak. Vegyük $\max\{X_1, \dots, X_{50}\}$ értékét.

A szimulációban 50 darab mintaelem maximumát vettem. Ahhoz, hogy megkapjam az eloszlásukat, 10^5 alkalommal ismételtem ezt meg és ezeket az adatokat ábrázoltam az alábbi hisztogramon. A mintaelemek generálása körülbelül egy percet vett igénybe. A Fisher-Tippet tétel szerint a minta eloszlása egy extrém érték eloszláshoz fog tartani. A sűrűségfüggvényhez egy Weibull típusú eloszlás sűrűségfüggvénye illeszthető. A 'runif' parancs segítségével generáltam véletlen mintát a megadott intervallumon. Az illesztett extrém érték eloszlás a paraméterek alapján egy Weibull típusú eloszlás volt, melyet kék vonallal ábrázoltam is a hisztogramon.



2. ábra. Egyenletes eloszlásból vett 50 mintaelem maximumához tartozó sűrűségfüggvény és az illesztett Weibull eloszlás

5.2. Poisson eloszlás

A Poisson eloszlás vizsgálata során néhány ötletet a [3] cikkből merítettem. Vegyünk $X_1, \ldots, X_n$ valószínűségi változókat λ paraméterű Poisson-eloszlással és vizsgáljuk ezek maximumát. Első lépésben ellenőrizzük a szükséges és elégséges feltétel teljesülését. Tudjuk, hogy $\mathbb{P}(X = k) = e^{-\lambda}\lambda^k/k!, \quad k \in \mathbb{R}_0, \quad \lambda > 0$. Ebből kapjuk, hogy

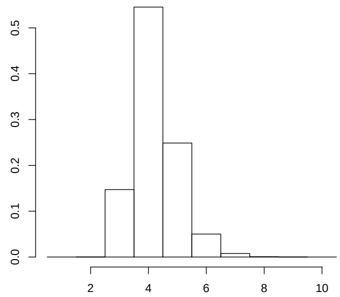
$$\begin{aligned}\frac{\overline{F}(k+0)}{\overline{F}(k)} &= \frac{\overline{F}(k) - \mathbb{P}(X=k)}{\overline{F}(k)} = 1 - \frac{\mathbb{P}(X=k)}{\overline{F}(k)} = 1 - \frac{F(k+1)-F(k)}{\overline{F}(k)} \\ &= 1 - \frac{\lambda^k}{k!}\left(\sum_{r=k}^{\infty}\frac{\lambda^r}{r!}\right)^{-1} = 1 - \left(1 + \sum_{r=k+1}^{\infty}\frac{k!}{r!}\lambda^{r-k}\right)^{-1}.\end{aligned}$$

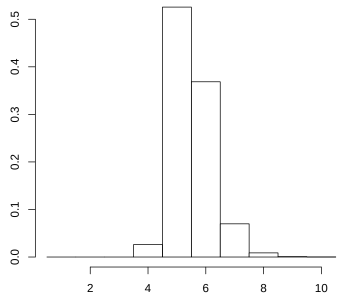
Elég nagy k esetén $(k > \lambda)$

$$\sum_{r=k+1}^{\infty} \frac{k!}{r!}\lambda^{r-k} = \sum_{s=1}^{\infty} \frac{\lambda^s}{(k+1)\cdots(k+s)} \leq \sum_{s=1}^{\infty} \left(\frac{\lambda}{k}\right)^s = \frac{\lambda/k}{1-\lambda/k} \to 0, \quad \text{ha } k \to \infty.$$

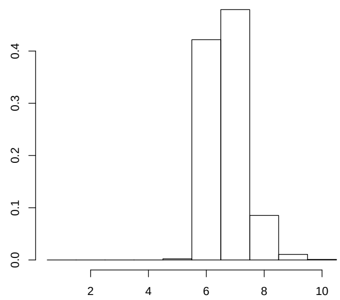
Ebből következik, hogy $\overline{F}(k+0)/\overline{F}(k) \to 0$, ha $k \to 0$, azaz nincs olyan (u_n) sorozat, melyre $n\overline{F}(k) \to \tau$. A Poisson-approximációs tétel alapján tehát nincs határeloszlása a maximumoknak.

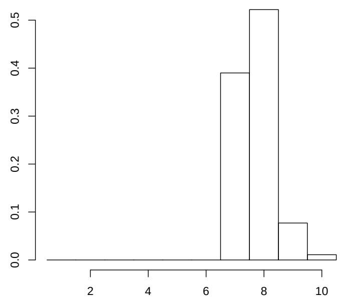
Viszont ha nincs határeloszlás, akkor mégis hogyan viselkedik a maximum? Erre az alábbi szimulációkkal kerestem a választ:





(a) 100 mintaelem maximuma (b) 1000 mintaelem maximuma





(c) 10^4 mintaelem maximuma (d) 10^5 mintaelem maximuma

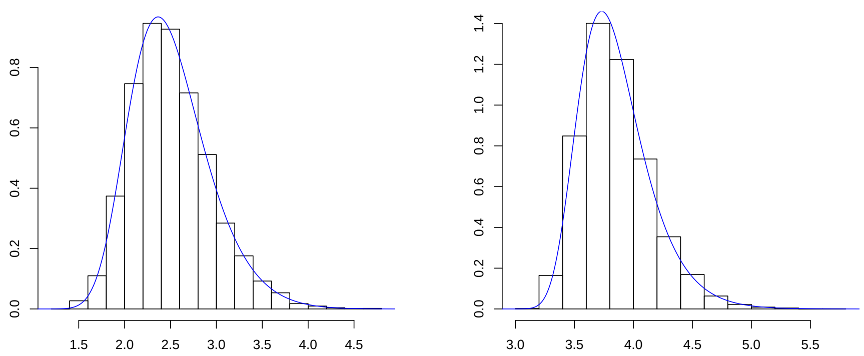
3. ábra. Poisson eloszlásból vett minták maximumainak eloszlása. 10^4 vizsgált elem alapján nem konvergálnak, a két leggyakoribb érték aránya egymáshoz képest folyamatosan változik

A szimulációkkal 1 paraméterű Poisson eloszlást vizsgáltam, egyre több kísérletnek véve a maximumát. Kellően sok maximum már elég jól közelíti ezek eloszlását. Az ábrákon látszik, hogy a maximumok a legtöbb esetben két érték valamelyikét veszik fel. Ezt a két értéket modális értéknek hívják. Azonban ha növelem a vizsgált elemek számát, a valószínűségek nem közelednek egy konkrét értékhez, egyszer egyik, egyszer másik valószínűsége nagyobb.

Érdekesség, hogy a modális értékek együttes valószínűsége ugyan nem mutat szigorú növekedést, de oszcillálva mégis tart 1-hez. A [3] cikk részletesebben tárgyalja ennek az okait.

5.3. Néhány további eloszlás maximumaihoz tartozó szimulációk

A normális eloszlásból generált, adott számú mintaelemnek vettem a maximumát. Ezekre általánosított extrém érték eloszlást illesztve azt kaptam, hogy az alakparaméter közel 0, azaz Gumbel típusú eloszlásról beszélhetünk. A hely és skálaparaméterek a következőképpen alakultak: 100 mintaelem esetén 2.325 és 0.382, valamint 10^4 mintaelem esetén 3.718 és 0.252 voltak.

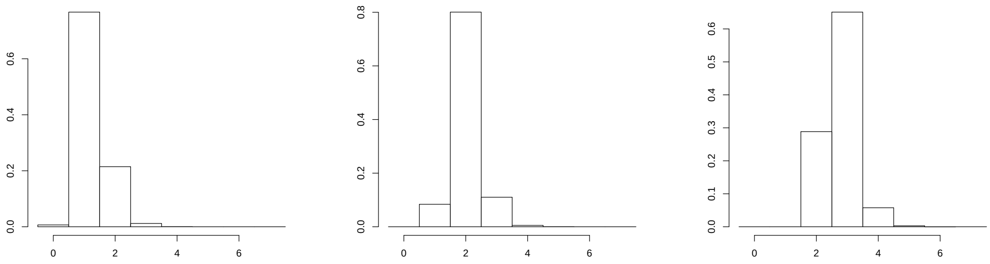


(a) 100 mintaelem maximuma

(b) 10^4 mintaelem maximuma

4. ábra. Két ábra a standard normális eloszlású minta maximumáról 10^4 szimuláció alapján és a hozzájuk illesztett Gumbel eloszlás

Geometriai, 0.95 paraméterű eloszlásból vettem mintákat. Azért választottam ilyen magasra a paramétert, hogy csak nagyon ritkán vehessen fel a valószínűségi változó magas értékeket és könnyebb legyen ábrázolni. A geometriai eloszlásnál is a Poissonhoz hasonlóan azt kell tapasztalnunk, hogy ha a vizsgált elemek számával tartunk végtelenhez a maximumok eloszlása nem konvergál.

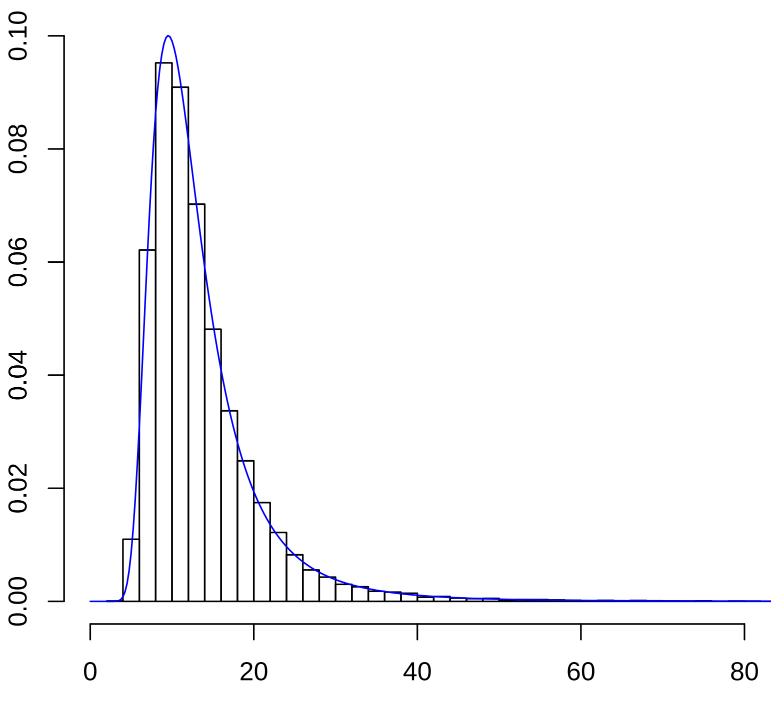


(a) 100 mintaelem

(b) 10^3 mintaelem

(c) 10^4 mintaelem

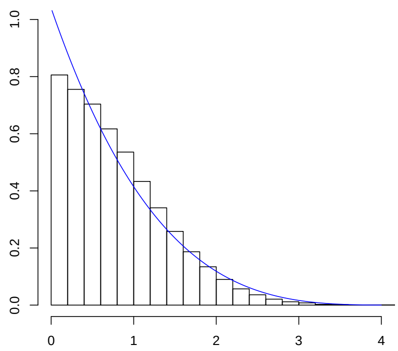
5. ábra. Három ábra a 0.95 paraméterű geometriai eloszlású minta maximumáról, melyek nem mutatnak konvergens viselkedést



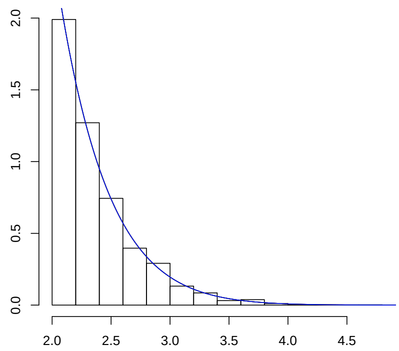
6. ábra. Lognormális eloszlásból vett minta maximumához tartozó eloszlás 30000 szimuláció alapján láthatóan jól közelít egy Fréchet típusú eloszlást

5.4. Normális eloszlás küszöb feletti módszerrel

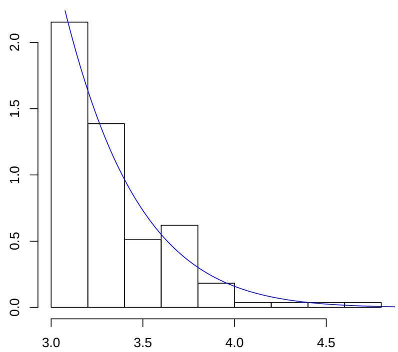
Vegyünk egy standard normális eloszlásokból álló független mintát. Válasszunk ki egy u küszöbszámot és vizsgáljuk az ennél nagyobb mintaelemek eloszlását. Először vizsgáljunk egy standard normális eloszlást. A minta 10^5 elemből áll, nézzük meg a 0, 2 és 3 küszöbök feletti elemek viselkedését. A megfelelő mintaelemek eloszlása és a 4.5. tétel értelmében hozzájuk leginkább illeszkedő általánosított Pareto-eloszlás az alábbi ábrákon látható.



(a) A küszöb $u = 0$



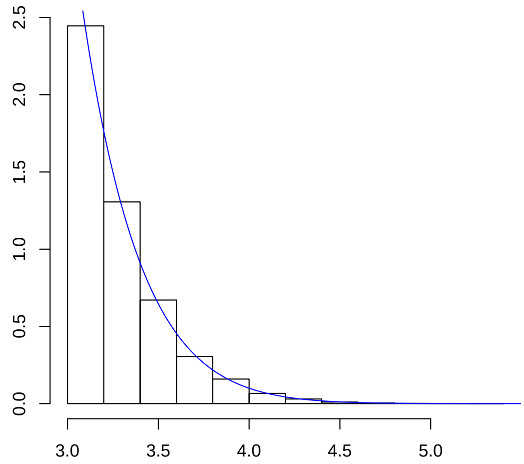
(b) A küszöb $u = 2$



(c) A küszöb $u = 3$

7. ábra. 10^5 mintaelemből kapott normális eloszlás u küszöb feletti sűrűségfüggvénye és a hozzá legjobban illeszkedő Pareto eloszlás

Megfigyelhető, hogy a 0 és a 3 küszöb esetén komolyabb eltérések vannak a minta eloszlása és a Pareto eloszlás között. Ez amiatt fordulhatott el, mert ezek nem megfelelő küszöbszámok. A 0 esetén túl sok mintaelemet vizsgáltunk, ezért az eloszlást nem csak az extrém értékek befolyásolják, a 3 esetén pedig túl kevés mintaelem esett a vizsgált területbe és ez torzította a kapott eloszlást. A 2 esetén a modell megfelelően működik. Az állítások igazolása végett vizsgáljuk meg, hogy 10^7 pont esetén megfelelő küszöb lesz-e a 3.



8. ábra. A küszöb itt is u=3, de a normális eloszlást 10^7 pontból állítottuk elő. Itt már látszik a Pareto-eloszláshoz való illeszkedés

Az ábra alátámasztja, hogy több mintaelem esetén már illeszkedik az általánosított Pareto eloszlás a magasabb küszöbszám feletti értékekre is. A küszöbszintet gyakran szokták egy felső kvantilishez tartozó értéknek választani. Ez egyrészről garantálja, hogy csak magas értékeik legyenek, másrészről ad egy arányt a minta elemszáma és a szint feletti adatok száma között. Így meghatározható, hogy mekkora minta esetén marad elég adtunk a megfelelő Pareto eloszlás illesztéséhez.

5.5. Pareto eloszlás u küszöb felett

Mint már tudjuk, ha egy tetszőleges eloszlás u küszöb feletti értékeit vizsgáljuk, azok Pareto eloszlásba fognak tartozni. Ha egy Pateto eloszlás esetén végezzük el, szintén Pareto eloszlást kapunk. Érdekes megvizsgálni, hogy hogyan hat a választott u küszöb az eloszlás paramétereire. Ehhez először írjuk fel a meghaladási eloszlásfüggvényt X valószínűségi változó és $u \in D(F(x))$ esetén. A teljes valószínűség tétele miatt kapjuk, hogy

$$
\begin{aligned}
F_u(x) = \mathbb{P}(X-u \leq x | X > u) &= \frac{\mathbb{P}(X-u \leq x) - \mathbb{P}(X-u \leq x | X \leq u)\mathbb{P}(X \leq u)}{\mathbb{P}(X > u)} \\
&= \frac{F(x+u) - 1F(u)}{1-F(u)}.
\end{aligned}
$$

Az általánosított Pareto eloszlás

$$
G_{\xi,\mu,\sigma}(x) = \begin{cases} 1-(1+\frac{\xi(x-\mu)}{\sigma})^{-1/\xi}, & ha \quad \xi \neq 0, \\ 1-e^{-\frac{(x-\mu)}{\sigma}}, & ha \quad \xi = 0, \end{cases}
$$

alakú. Vizsgáljuk először $\xi \neq 0$ esetben a feltételes valószínűséget.

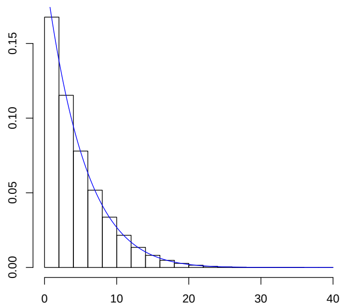
$$
\begin{aligned}
F_u(x) &= \frac{1-(1+\frac{\xi(x+u-\mu)}{\sigma})^{-1/\xi} - (1-(1+\frac{\xi(u-\mu)}{\sigma})^{-1/\xi})}{1+\frac{\xi(u-\mu)}{\sigma})^{-1/\xi}} \\
&= 1-\left(\frac{\sigma+\xi(x+u-\mu)}{\sigma+\xi(u-\mu)}\right)^{-1/\xi} \\
&= 1-\left(1+\frac{\xi(x)}{\sigma+\xi(u-\mu)}\right)^{-1/\xi} = P_{\xi,0,\sigma+\xi(u-\mu)}(x).
\end{aligned}
$$

Azt kaptuk, hogy a feltétel szerinti eloszlásban a Pareto eloszlás paraméterei részben megváltoztak az eredetihez képest. Az alakparaméter ugyanúgy ξ maradt, de a hely-paraméter 0 lett. A skálaparaméter változott a legtöbbet, méghozzá hozzáadódott $\xi(u-\mu)$, a paraméterektől és u küszöbtől függő érték.

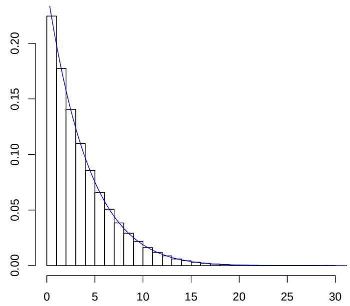
Nézzük meg a $\xi = 0$ esetet is.

$$
\begin{aligned}
F_u(x) &= \frac{1-\exp(-\frac{(x+u-\mu)}{\sigma}) - (1-\exp(-\frac{(u-\mu)}{\sigma}))}{\exp((-\frac{(u-\mu)}{\sigma})} \\
&= 1-\frac{\exp(-\frac{(x+u-\mu)}{\sigma})}{\exp(-\frac{(x-\mu)}{\sigma})} \\
&= 1-\exp(\frac{-x-u+\mu+u-\mu}{\sigma}) = 1-e^{-\frac{-x}{\sigma}} = P_{0,0,\sigma}(x).
\end{aligned}
$$

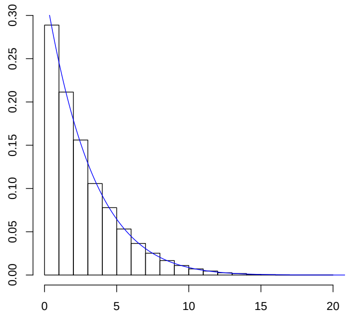
$P_{\xi,0,\sigma+\xi(u-\mu)}(x)$ esetén $\xi = 0$ behelyettesítéssel éppen visszakapjuk a fenti paraméterezést. Tehát egy Pareto eloszlás esetén, tetszőleges értelmezési tartománybeli u küszöbhöz a kapott eloszlás $P_{\xi,0,\sigma+\xi(u-\mu)}(x)$ lesz. Nézzünk erre egy példát $\xi = -0.1$, $\mu = 0$, $\sigma = 5$ paraméterekkel. A minta legyen 10^7 elemű és vizsgáljuk meg $0, 10, 20, 30$ küszöbökhöz tartozó eloszlásokat. A képlet szerint $5, 4, 3, 2$ kell, hogy legyen a skálaparaméter, ehhez képest az illesztett eloszlások 5, 4, 2.99, valamint 1.91 értékeket adtak eredményül. Ez közelítőleg megfelel az elvárásainknak.



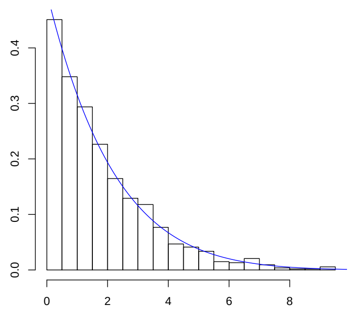
(a) A küszöb 0, azaz nincs megkötés



(b) A küszöb: $u = 10$



(c) A küszöb: $u = 20$



(d) A küszöb: $u = 30$

9. ábra. 10^7 mintaelemből előállított Pareto eloszlások u küszöb felett valamint a hozzá illesztett Pareto eloszlás

6. Összefoglalás

A dolgozatban bemutattam az extrém érték elmélet alapjait, a Fisher-Tippet tételt, a vonzási tartományok meghatározását, valamint a szint feletti értékek módszerét. Az extrém értékek témaköre azonban itt még nem ér véget. Két fontos kérdés merülhet fel ezek után, amelyekkel kapcsolatban még általánosabb állításokat tudunk megfogalmazni. Hogyan alakulnak a maximumok többváltozós esetben? Mi lehet a határeloszlás, ha a mintaelemek nem függetlenek?

A többváltozós esettel foglalkozik bővebben a [7] cikk. Legyen G egy többváltozós valószínűségi változó. Ha G minden peremeloszlása Fréchet eloszlást követ és az összefüggőséget egy megfelelő mérték írja le, akkor az ilyen típusú eloszlást többváltozós extrém érték eloszlásnak hívjuk. Az ilyen eloszlásokat nem lehet néhány paraméterrel megadni, szükséges egy mérték definiálása hozzá. Emiatt nincs is megfelelő karakterizációja. Egy F eloszlás akkor és csak akkor tartozik a Fréchet típusú többváltozós extrém érték eloszlás vonzási tartományába, ha

$$\lim_{t\to\infty} \frac{-\log F(t\overline{x})}{-\log F(t\overline{1})} = \frac{-\log G(\overline{x})}{-\log G(\overline{1})}.$$

A vonzási tartományhoz további ekvivalens átalakítások is léteznek. Amennyiben a vektorváltozó minden dimenziója azonos eloszlású, a független vektorok maximuma a hagyományos extrém érték eloszlással jellemezhető.

Az összefüggő mintaelemek esetén nem tudunk minden esetben a Fisher-Tippet tételhez hasonló állítást megfogalmazni. Azonban ha a mintaelemek csak gyengén összefüggőek, akkor a dolgozatban kifejtett tételek fennállnak. Ezzel a témakörrel a [8] cikk foglalkozik.

Ábrák jegyzéke

Hivatkozások

[1] Filip Lindskog, *The mathematics and fundamental ideas of extreme value theory*, 2004.

[2] Sidney Ira Resnick, *Extreme Values, Regular Variation, and Point Processes*, Springer, New York, 1987.

[3] Keith Briggs, Linlin Song, Thomas Prellberg, *A note on the distribution of the maximum of a set of Poisson random variables*, 2009.

[4] T. Mikosch, *Regular Variation Subexponentiality and Their Applications in Probability Theory*, 1999.

[5] M.R. Leadbetter, G. Lindgren, H. Rootzén *Extremes and Related Properties of Random Sequences and Process*, Springer, Berlin, 1983.

[6] Philippe Soulier, *Some applications of regular variation in probability and statistics*, Caracas, 2009.

[7] Anne-Laure Fougéres, *Multivariate extremes*, Extreme Values in Finance, 2004.

[8] M.R.Leadbetter, *On extreme values in stationary sequences*, Springer, 1974.