

Eötvös Loránd Tudományegyetem
Természettudományi Kar

Vincze Gergely

**A centrális határeloszlás-tétel
konvergenciasebességének szimulációs
vizsgálata**

Bsc szakdolgozat

Témavezető:

Zempléni András, egyetemi docens

Valószínűségelméleti és Statisztika Tanszék



Budapest, 2010.

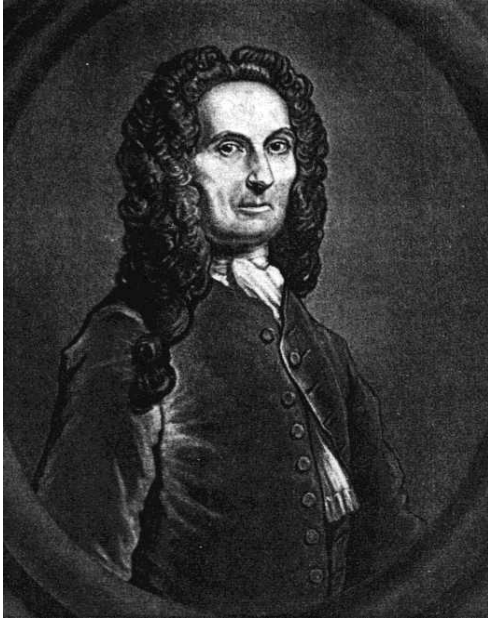
Előszó

Bár a szakdolgozatom címe arra enged következtetni, hogy a matematikának csak egy jól behatárolható területét érinti, mégsem jöhetett volna létre a megfelelő háttértudás nélkül. Éppen ezért először meg szeretném köszönni az ELTE oktatóinak színvonalas oktatását, ami lehetővé tette nekem, hogy az elmúlt három év alatt felhalmozott tudás tekintélyes hányadát felhasználjam a dolgozatomban. Továbbá köszönettel tartozom néhány csoporttársamnak is, akik segítettek a megfogalmazásomban lévő következetlenségeket felszámolni. A legnagyobb hálával viszont témavezetőmnek, Zempléni Andrásnak tartozom, akinek a türelme és útmutatásai nélkül a munkám színvonala minden bizonnyal jóval alulmúlná a meglévőt.

Tartalom

| | |
|--------------------------------------|----|
| Bevezetés..... | 1 |
| 1. Elméleti eredmények..... | 2 |
| 1.1 Berry-Esséen tétel..... | 3 |
| 2. Konvergenciasebességek vizsgálata | |
| 2.1 Gamma-eloszlás..... | 9 |
| 2.2 Béta-eloszlás..... | 15 |
| 2.3 Pareto-eloszlás..... | 23 |
| 3. Összefoglalás..... | 29 |

Bevezetés



Abraham de Moivre

A centrális határeloszlás-tétel története a XVIII. századra vezethető vissza. Az első változat a francia-születésű matematikus, *Abraham de Moivre* nevéhez köthető. 1733-ban egyik publikációjában a normális eloszlást használta, hogy közelítse az érme-dobálásoknál a fejek számának az eloszlását. Ezen észrevétel távol állt korának szellemiségétől és majdnem homályba is veszett. *Laplace* felhasználta *de Moivre* eredményeit a *Théorie Analytique des Probabilités* című művében, ahol a binomiális eloszlást közelítette normális eloszlással. *Laplace* ezen műve 1812-ben jelent meg. Sajnos ő sem keltette fel korának figyelmét, így az igazi áttörésre egészen 1901-ig kellett várnunk. Az orosz születésű *Alekszandr Ljapunov* ekkor fogalmazta meg a tételt. A centrális határeloszlás-tétel elnevezést először *Pólya György* használta az egyik 1920-ban íródott cikkében. *Pólya* a „centrális”- kifejezéssel lényegében a tétel valószínűségelméletben betöltött szerepének fontosságára utal. Nyilván a felsorolt matematikusokon kívül is voltak gondolkodók, akik nélkül nem jöhetett volna létre a tétel. A normális eloszlás főként *Gauss* nevéhez fűződik, ugyanakkor *Lindeberg* nélkül nem ismerhetnénk a tétel általános alakját.

Dolgozatom első, bevezető fejezetében bevezetem a centrális-határeloszlás tételt és az ebben a tételben központi szerepet betöltő normális eloszlás fogalmát, továbbá a második fejezetben vizsgálni kívánt három eloszlás-családot, végül az *Andrew C. Berry* és *Carl Gustav Esséen* által megfogalmazott egyenlőtlenség különböző alakjaival fogok foglalkozni. A második fejezet három alfejezete a három eloszláscsalád konvergenciájának a vizsgálatáról fog szólni. Végül az utolsó fejezetben összegezem az elért eredményeket.

1. Elméleti eredmények

1.1 Definíció (*normális eloszlás*): Egy X valószínűségi változó *normális eloszlást* követ pontosan akkor, ha a sűrűségfüggvénye:

$$f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2s^2}}$$

ahol s és m valós paraméterek, (szórás, várható érték), s pozitív. Speciálisan: $s=1$, $m=0$ paraméterek esetén *standard normális eloszlás*.

1.2 Tétel (*Centrális határeloszlás-tétel*): Legyenek X_1, X_2, \dots, X_n független, azonos eloszlású valószínűségi változók, melyeknek várható értéke m és szórása s , ami véges, ekkor minden valós x esetén:

$$P\left(\frac{\sum_{i=1}^n X_i - nm}{s\sqrt{n}} < x\right) \rightarrow \Phi(x) \quad \text{ha } n \rightarrow \infty$$

ahol Φ a standard-normális eloszlásfüggvény.

Ennek a tételnek valóban óriási elméleti és gyakorlati jelentősége van. Lényegében a természet bizonyos szintű kiismerhetőségéről szól. Szakdolgozatom célja a tételben szereplő konvergencia sebességének a szimulációs vizsgálata. Erre *Matlab* és *R* programot fogok használni három eloszlás család esetén. Ezek pedig az alábbiak:

1.3 Definíció (*Gamma-eloszlás*): Egy X valószínűségi változó P -edrendű λ paraméterű *Gamma-eloszlású*, ha a sűrűségfüggvénye:

$$f(x) = \frac{\lambda^P x^{P-1} e^{-\lambda x}}{\Gamma(P)}$$

ahol $\Gamma(P)$ valós függvény (*gamma-függvény*, amit tekinthetünk a faktoriális művelet általánosításának), továbbá P és λ pozitív paraméterek.

1.4 Definíció (Béta-eloszlás): Egy X valószínűségi változó *béta-eloszlású*, ha a sűrűségfüggvénye:

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad x \in [0, 1]$$

ahol $B(\alpha, \beta)$ a béta-függvény továbbá α, β pozitív paraméterek.

1.5 Definíció (Pareto-eloszlás): egy X valószínűségi változó *Pareto-eloszlású*, ha a sűrűségfüggvénye:

$$f(x) = \frac{\alpha}{\beta} \left(\frac{\beta}{x}\right)^{\alpha+1} \quad x \in (\beta, \infty)$$

ahol α és β egyaránt pozitív valós paraméterek.

1.1 Berry-Esséen tétel

Az olyan eloszlástételek esetében, melyeknél a konvergencia központi szerepet tölt be, mindig nagy jelentősége van az adott konvergencia sebességének. Felmerülhet a kérdés, hogy a centrális határeloszlás-tételben szereplő konvergencia milyen sebességű. Ezt átfogalmazhatjuk úgy is, hogy a tételben szereplő standardizált összeg-eloszlás görbék milyen sebesen simulnak rá a standard normális eloszlás görbére. Ezt akár két megközelítésből is vizsgálni tudjuk. Első esetben ha az eloszlások értelmezési tartományának összes lehetséges értékén vesszük a *supremumát* az eloszlás görbék távolságának. Második esetben, a két görbe által közrezárt *területet* vizsgáljuk meg. Az első esetre ad nekünk egy felső becslést a *Berry-Esséen tétel*. Ez a tétel a későbbiekben is nagy hasznunkra lesz, ugyanis ennek a tételnek a segítségével könnyedén ellenőrizni tudjuk a szimulációs eredményeinket.

1.6 Tétel (Berry-Esséen): Legyenek X_1, X_2, \dots, X_n független, azonos eloszlású valószínűségi változók, s szórással és m várható értékkel. Legyen

$\rho = E|X_1 - E(X_1)|^3$, azaz a harmadik *abszolút centrált momentum* továbbá teljesüljön: $E|X|^3 < \infty$.

Legyen $\frac{\sum_{i=1}^n X_i - m}{s\sqrt{n}}$ valószínűségi változó eloszlása H_n , ekkor minden valós x

esetén:

$$|H_n(x) - \Phi(x)| \leq C \frac{\rho}{s^3 \sqrt{n}}$$

1.7 Megjegyzés: Jellegzetessége ennek a tételnek, hogy a felső korlát csak a szórástól és a harmadik abszolút centrált momentumtól függ, de emellett jó valószínűségi becslés kapható az első két momentumra is. Tipikus becslése a tételben szereplő C -nek lehet 3. *P. van Beek* élesítette 1972-ben 0.7975-re, később *Shiganov* tovább pontosította 0.7655-re. *Shiganov* ezen eredménye feltehetően a legjobb, ugyanakkor fontos megjegyezni, hogy $n \geq 65$ esetén ennek a konstansnak az értéke még tovább csökkenthető, amennyiben az n -nel végtelenhez tartunk, akkor a konstans 0.7164-hez fog közelíteni.

A vizsgálataim során a *Shiganov* által meghatározott konstans fogom használni. Ezen konstans értékének a levezetése meghaladja a dolgozatom kereteit, így *Berry* által megfogalmazott eredeti állítást fogom kimondani és bebizonyítani, ami nem szolgál olyan éles becsléssel, mint a fentebb leírt alak, ám annál általánosabb, elvégre ez nem követeli meg, hogy a valószínűségi változók ugyanolyan eloszlásúak legyenek. Továbbá a tétel bizonyítása nagyon szellemes, és sok szempontból nagyon tanulságos.

1.8 Tétel: Legyenek X_1, X_2, \dots, X_n független valószínűségi változók rendre F_1, F_2, \dots, F_n eloszlással, továbbá legyenek τ_k, τ_k' olyanok, hogy $-\infty \leq -\tau_k < 0 < \tau_k' \leq \infty$, ekkor legyen $\bar{X}_k = X_k$ ha $-\tau_k < x_k < \tau_k'$, 0 egyébként, ekkor $X_k = \bar{X}_k + X_k'$, továbbá legyen $\alpha_k = E(\bar{X}_k)$, $\beta_k = E(\bar{X}_k^2)$ és $\gamma_k = E|X_k|^2$ ekkor legyen $a = \alpha_1 + \dots + \alpha_n$, $b = \beta_1 + \dots + \beta_n$ és $c = \gamma_1 + \dots + \gamma_n$. Tegyük fel, hogy $E(X) = 0$, és $E(X_k^2) = s_k^2$ ami véges minden X_k esetén. Legyen $S^2 = s_1^2 + \dots + s_n^2$ és $\beta_k' = E(X_k'^2)$, továbbá $b' = \beta_1' + \dots + \beta_n'$, ekkor

$$\Delta \leq 6 \left(\frac{c}{S^3} + \frac{b'}{S^2} \right)$$

ahol $\Delta = \sup |F(x) - \Phi(x)|$, Φ a standard normális eloszlás.

Hogy bebizonyítsuk ezt a tételt, szükségünk lesz két lemmára (ezek közül az első *Berry* nevéhez fűződik).

1.9 Lemma: Ha Φ a karakterisztikus függvénye az F eloszlásnak, akkor

$$\pi \Delta \leq \int_{-T}^T \left| \Phi_k \left(\frac{x}{S} \right) e^{-\frac{1}{2}x^2} \right| |x|^{-1} dx + \frac{24}{T \sqrt{2\pi}}$$

1.10 Lemma: Legyen $u_k = \Phi_k \left(\frac{x}{S} \right)$ és $v_k = e^{\frac{1}{2} \frac{s_k^2 x^2}{S^2}}$, ahol Φ_k a karakterisztikus függvénye F_k -nak, ekkor

$$u_1 \dots u_n - v_1 \dots v_n = \sum u_1 \dots u_{v-1} (u_v - v_v) v_{v+1} \dots v_n$$

1.8 Bizonyítása: Jelölje az 1.9 lemmában szereplő integrált I , első dolgunk megmutatni:

$$I \leq \int_{-\infty}^{\infty} e^{\frac{-x^2}{8}} \sum \left| \Phi_k \left(\frac{x}{S} \right) e^{-\left(\frac{s_k^2 x^2}{2S^2} \right)} \right| |x|^{-1} dx \quad (1)$$

Felhasználva a karakterisztikus függvény *Taylor-sorba* fejthetőségét adódik:

$$\left| \Phi_k \left(\frac{x}{S} \right) - 1 + \frac{s_k^2 x^2}{2S^2} \right| \leq \frac{\gamma_k |x|^3}{6S^3} + \frac{\beta'_k x^2}{S^2} \quad (2)$$

Mivel minden pozitív x -re fennáll $0 \leq e^{-x} - 1 + x < \frac{1}{2} x^2$, így tovább írva:

$$\left| \Phi_k \left(\frac{x}{S} \right) - e^{-\frac{s_k^2 x^2}{2S^2}} \right| \leq \frac{\gamma_k |x|^3}{6S^3} + \frac{\beta'_k x^2}{S^2} + \frac{x^4 s_k^4}{8S^4} \quad (3)$$

Ezt visszairhatjuk a kiinduló egyenlőtlenségbe:

$$I \leq \int_{-\infty}^{\infty} e^{-\frac{x^2}{8}} \left[\frac{\gamma_k |x|^3}{6S^3} + \frac{\beta'_k x^2}{S^2} + \frac{x^4}{8S^4} \sum s_k^4 \right] dx \leq \frac{32}{9} + 8 \frac{b'}{S^2} + 8 \sum \left(\frac{s_k}{S} \right)^4 \quad (4)$$

Felhasználva: $\sqrt{2\pi} < \frac{8}{3}$

Hogy igazoljuk az első állításunkat, felhasználjuk az 1.10 lemmát, célunk minden u_k -ra egy felső korlátot találni. Rögzítsük T paramétert, ekkor két osztályba osztjuk a lemmában szereplő u -khoz tartozó alsó indexeket, legyen A osztály az olyan indexek halmaza, melyekre:

$$\frac{\sqrt{\beta_k}}{S} \leq \frac{4}{3} \quad (5)$$

A (2)-es pontban felírtakat felhasználva:

$$\left| \Phi_k \left(\frac{x}{S} \right) \right| \leq \left| 1 - \frac{\beta_k x^2}{2S^2} \right| + \frac{\gamma_k |x|^3}{6S^3} + \frac{\beta'_k x^2}{2S^2} \quad (6)$$

Ugyanakkor a bal oldalt tudjuk becsülni máshogyan is, az olyan indexek esetén, melyek benne vannak A -ban, és $|x| < T$ teljesül, ekkor

$$(7) \quad \left| \Phi_k \left(\frac{x}{S} \right) \right| \leq e^{-\frac{x^2}{2} \left(\frac{\beta_k}{S^2} - \frac{\gamma_k T}{3S^3} - \frac{\beta'_k}{2S^2} \right)}, \text{ így } \prod_{k \in A} \left| \Phi_k \left(\frac{x}{S} \right) \right| \leq e^{-\frac{x^2}{2} \left(\frac{b_A}{S^2} - \frac{C_A T}{3S^3} - \frac{b'_A}{2S^2} \right)} \quad (8)$$

ahol $b_A = \sum_{k \in A} \beta_k$ és $C_A = \sum_{l \in A} \gamma_l$. Következő lépésben a jobb oldalon a kitevőben levő zárójelben szereplő kifejezésre szeretnénk találni egy kisebb becslést.

Tudjuk, hogy A komplementerére igaz lesz az alábbi észrevétel (felhasználva a momentumokra vonatkozó egyenlőtlenséget):

$$\frac{4S}{3T} \beta_k \leq \beta_k^{\frac{2}{3}} \leq \gamma_k, \text{ ha összegezzük } \beta_k \text{ és } \gamma_k \text{ kifejezéseket olyan } k\text{-kra, melyek}$$

nincsenek A -ban, akkor $\frac{4S}{3T} (b - b_A) \leq C - C_A$ adódik. Ezután tekintsük b_A -t és

c_A -t, melyek egyaránt pozitív paraméterek.

Belátható, hogy a legkisebb lehetséges b_A minimalizálja a már korábban említett zárójelben szereplő kifejezést, ami $b_A = b - \frac{3T}{4S} C$ és $C_A = 0$, továbbá mivel $b + b' = S^2$, adódik:

$$\prod_{k \in A} \left| \Phi_k \left(\frac{x}{S} \right) \right| \leq e^{\frac{-x^2}{2} \left(1 - \frac{3CT}{4S^3} - \frac{2b'}{S^2} \right)} \quad (9)$$

Most térjünk vissza a 1.10 lemmához. Az olyan k indexekre, melyek benne vannak

A -ban, (7)-ben szereplő becslése $\left| \Phi_k \left(\frac{x}{S} \right) \right|$ -nek meghaladja $e^{\left(\frac{-s_k^2 x^2}{2S^2} \right)}$ értékét,

tehát u_k -ra és v_k -ra egyaránt igaz. Az olyan k indexekre, melyek A -n kívülre esnek, felhasználhatjuk, hogy $|u_k| \leq 1$, $|v_k| \leq 1$. Ha v index nincs benne A -ban, akkor a 1.10 lemmában lévő $|u_v - v_v|$ tényezője kisebb, mint (9). Ugyanez igaz

akkor is, ha v benne van A -ban, és elhagyjuk $\left| \Phi_v \left(\frac{x}{S} \right) \right|$. Ugyanakkor látható, hogy

(8) és (5) miatt a v -edik tag elhagyása az exponenciális tagot kevesebbel növeli, mint

$$\frac{x^2}{2S^2} \leq \frac{8x^2}{9T^2}, \text{ ez alapján fennáll}$$

$$\left| \sum u_1 u_2 \dots u_{v-1} (u_v - v_v) v_{v+1} \dots v_n \right| < \sum |u_k - v_k| e^{\frac{-1}{2} x^2 \left(1 - \frac{3Tc}{4S^3} - \frac{2b'}{S^2} - \frac{16}{9T^2} \right)}$$

Ez alátámasztja (1)-et, feltéve, ha az alábbi teljesül:

$$T \frac{c}{S^3} + \frac{8}{3} \frac{b'}{S^2} + \frac{64}{27T^2} \leq 1$$

Ezzel a kikötéssel a (4)-ben szereplő egyenlőtlenség jobb oldalán szereplő összeget ki tudjuk fejezni. Tegyük fel, hogy az alábbi fennáll:

$$\frac{c}{S^3} + \frac{b'}{S^2} \leq \frac{1}{6} \quad (12)$$

Felhasználva a momentumokra vonatkozó egyenlőtlenséget: $\beta_k^3 \leq \gamma_k^2$, kapjuk:

$$s_k^4 \leq (\gamma_k^{\frac{2}{3}} + \beta_k')^2 \leq c^{\frac{1}{3}} \gamma_k + \beta_k (2c^{\frac{2}{3}} + b')$$

Mivel bal oldal a (10) miatt akkor veszi fel a maximumát, ha $c = \frac{s^3}{6}$ és $b' = 0$,

továbbá felhasználva: $6^{\frac{1}{3}} > \frac{9}{5}$ kapjuk: $\sum \left(\frac{s_k}{S} \right)^4 \leq \frac{5}{9} \frac{c}{s^3} + \frac{50}{81} \frac{b'}{s^2}$ így ezt, és a (4)-et

felhasználva adódik: $I < 8 \frac{c}{s^3} + 13 \frac{b'}{s^2}$.

Alkalmazzuk ezt a becslést az 1.9 lemmára, felhasználva azt a tényt, hogy:

$$\sqrt{2\pi} < \frac{2}{5}$$

A *Berry-Esséen* tétel igazolásához szükségünk van egy olyan T -re, amire fennáll:

$$8 \frac{c}{s^3} + 13 \frac{b'}{s^2} + \frac{48}{5T} \leq 6\pi \left(\frac{c}{s^3} + \frac{b'}{s^2} \right) \quad (13)$$

de mivel $6\pi > \frac{94}{5}$, és a (13) miatt $\frac{1}{T} = \frac{9}{8} \frac{c}{s^3} + \frac{5}{8} \frac{b'}{s^2}$, ezt felhasználva és (11)-be

visszaírva: $\frac{8}{9} - \left(\frac{5T}{9} - \frac{8}{3} \right) \frac{b'}{s^2} + \frac{64}{27T^2}$ (14) (egyenlőtlenség bal oldala). Ebből és

(12)-ből következik, hogy $T^{-1} < \frac{3}{16}$, emiatt igaz lesz $\frac{64}{27T^2} < \frac{1}{12}$, továbbá b'

együtthatója pozitív, így (14) biztos, hogy kisebb, mint 1 és ezt akartuk belátni.

□

Fontos megemlíteni, hogy a konvergencia sebességét a görbék által közbezárt területtel is mérhetjük. Ez a fajta mérőszám több statisztikai eljárásban is szerepel, melyeket elsőként Harald Cramér és Richard Edler von Mises javasoltak 1928 és 1930 között. Ez lényegében egy hipotézisvizsgálat, ami azt vizsgálja, hogy a tapasztalati eloszlás, F_n mennyire simul rá az elméleti eloszlásra F :

$$w^2 = \int_{-\infty}^{\infty} [F(x) - F_n(x)]^2 dF(x)$$

Gyakorlatilag ez képezi A *Cramér-von-Mises tesztek* magját.

2. Konvergenciasebességek vizsgálata

2.1. Gamma-eloszlás

Az első eloszláscsalád, amivel foglalkozni szeretnék, az a gamma-eloszlás, ennek elsősorban az az oka, hogy ha a valószínűségi változóink gamma-eloszlásúak, akkor az összeg is gamma-eloszlású lesz. Mindez nagyon könnyen belátható, ha tekintjük a gamma-eloszlás sűrűségfüggvényét (1.1 definíció). Az összeg eloszlásának vizsgálatához szükségünk lesz a konvolúciós formulára.

2.1.1 Állítás (konvolúciós formula): Ha X, Y független, abszolút folytonos eloszlású valószínűségi változók, akkor $X+Y$ is abszolút folytonos eloszlású, és sűrűségfüggvénye:

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(y) f_Y(x-y) dy = \int_{-\infty}^{\infty} f_X(x-y) f_Y(y) dy$$

Felhasználva a konvolúciós formulát két azonos, p rendű és λ paraméterű gamma eloszlású valószínűségi változóra:

$$f_{X+Y} = \frac{\int_0^{\infty} \lambda^p y^{p-1} e^{-\lambda y} \lambda^p (x-y)^{p-1} e^{-\lambda(x-y)} dy}{\Gamma^2(p)} = \frac{\lambda^{2p} e^{-\lambda x}}{\Gamma^2(p)} \int_0^x y^{p-1} (x-y)^{p-1} dy$$

Ekkor a jobb oldalon szereplő integrálnak meg tudjuk mondani a pontos értékét, mert az nem más, mint a Béta-eloszlás sűrűségfüggvényének egy $B(p,p)$ -vel való nyújtása. Mivel tudjuk, hogy egy abszolút folytonos Y valószínűségi változó sűrűségfüggvényét az Y értelmezési tartományán integrálva 1-et kapunk, ezért ennek az integrálnak az értéke:

$$B(p,p) = \frac{\Gamma(p) \Gamma(p)}{\Gamma(2p)}$$

ami az úgynevezett *Béta-függvény*. Ezt visszaírva az eredeti egyenletbe:

$$f_{x+y} = \frac{\lambda^{2p} e^{-\lambda x} x^{2p-1}}{\Gamma(2p)}$$

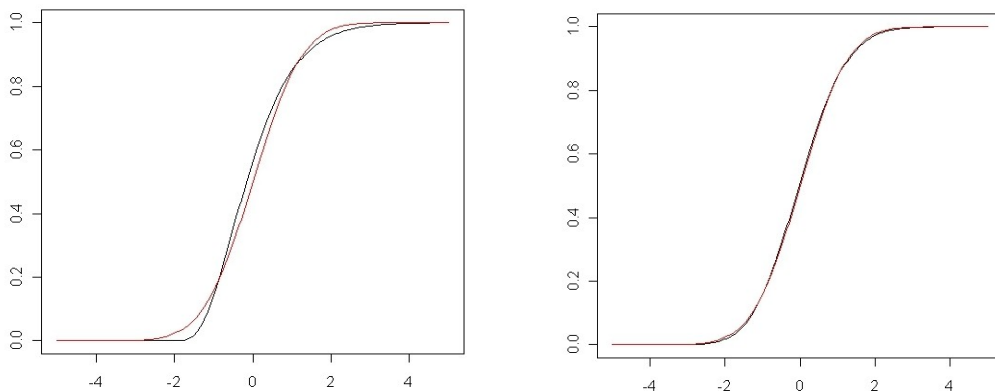
és pont ezt szeretnénk volna belátni.

Tehát a közös eloszlás λ paraméterű és $2p$ rendű, így indukcióval belátható, hogy n darab azonos rendű és paraméterű gamma-eloszlású valószínűségi változónak az összege is gamma-eloszlású, λ paraméterű, np rendű valószínűségi változó lesz. Ez lényegesen megkönnyíti a dolgunkat, elvégre nem is kell szimulálni számítógéppel az összeget. A vizsgálat megkezdése előtt ellenőriznünk kell, hogy a centrális határeloszlás-tételben szereplő feltételek teljesülnek-e, azaz a szórásnégyzet véges lesz-e.

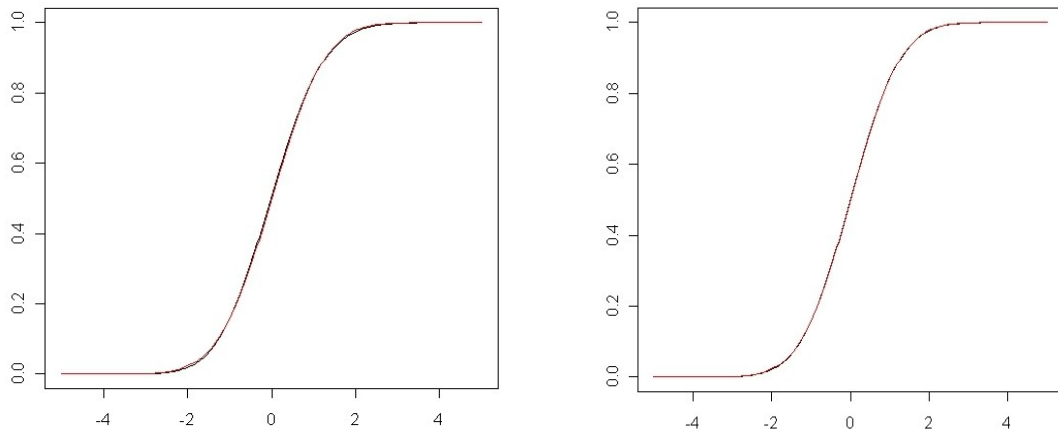
$$E^2(X) = \frac{p}{\lambda^2}$$

Ami nem lehet végtelen, mivel λ és p egyaránt pozitívak, így most már ténylegesen megkezdhetjük a vizsgálatokat.

A valószínűségi változók összegének a vizsgálatára egy egyszerű algoritmus adta magát. Vegyük egy adott intervallum ekvidisztans felosztását. A felosztáshoz tartozó értékekhez rendeljük a megadott paraméterekkel rendelkező gamma-valószínűséget (természetesen ezt standardizálva szórással és várható értékkel). *R*-ben programozva a megfogalmazott algoritmus az alábbi eredményeket adja különböző számú független, azonos paraméterű, standardizált gamma-eloszlású valószínűségi változók összegére (*fekete*), összehasonlítva a normális (*piros*) eloszlással:



2.1 ábra

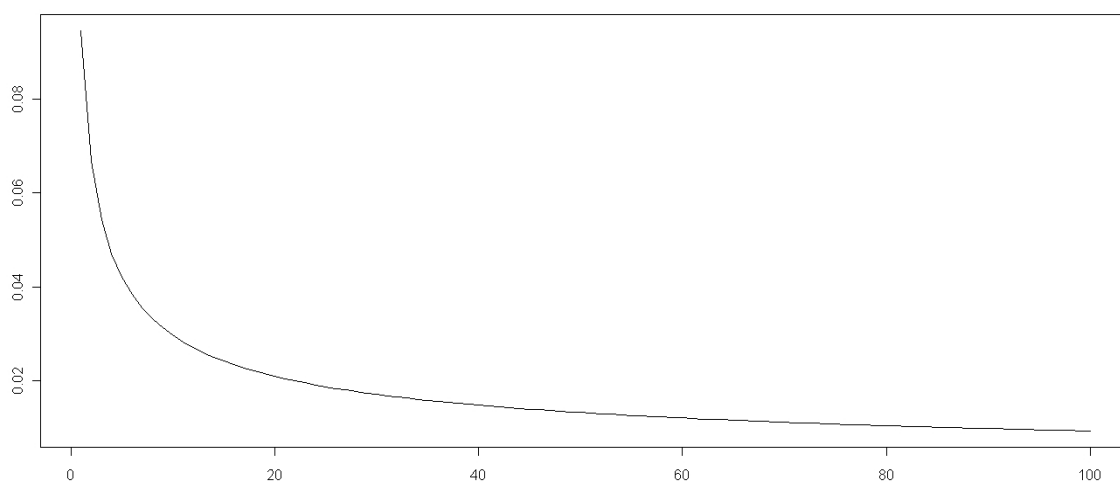


2.2 ábra

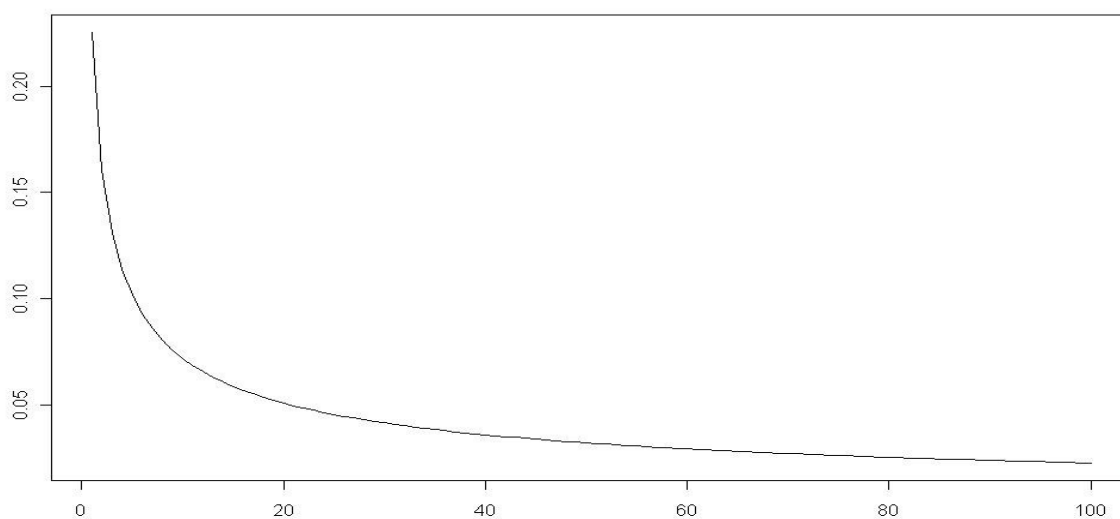
A 2.1-es ábrán 2, 50, a 2.2-es ábrán pedig 100, 1000 számú gamma-eloszlású valószínűségi változók összegét mutatja. Jól kivehető, hogy a 2.1-es ábra esetén még nagyjából elkülönül a két eloszlás, ugyanakkor a 2.2-es ábrán már szinte teljesen egymásra simulnak.

Tehát meg kell vizsgálnunk, hogy a fentebb említett konvergencia hogyan függ az n -től, azaz a független azonos eloszlású valószínűségi változók számától. Mint már a bevezetőben említettem, erre kettő darab kézenfekvő lehetőség van. Elsőként, ha vesszük a supremumát az ekvidisztans vektorban szereplő elemekhez tartozó normális és gamma valószínűségeket különbségének abszolút értékét. Másodsor, ha a két grafikon által közrezárt területet mérjük le. Ezt úgy tehetjük meg a legegyszerűbben, ha felhasználjuk azt a vektort, amibe beleírtuk a normális és a gamma-eloszlás különbségét. Továbbá felhasználhatjuk azt a tényt, hogy a pontok, amiket vizsgáltunk, egy intervallum ekvidisztans felosztása, így a közrezárt terület közelíteni tudjuk trapézokkal. Ekkor ezen eredményeket ábrázolhatjuk n függvényében és az alábbi ábrákat kapjuk $n=100$ -re (azaz 2 és 100 közötti független valószínűségi változók összegére vonatkozó eltérések):

Supremumok alakulása n függvényében:



Eloszlás-görbék által közrezárt terület alakulása n függvényében:



2.3 ábra

Szemmel láthatóan igazolódik a centrális határeloszlás-tétel, a két görbe n -et növelve valóban 0-hoz fog tartani. Továbbá úgy tűnik, hogy a két görbe alakja egészen hasonló, habár egyszerre ábrázolva őket a supremumok alakulását szemléltető ábra a területeket reprezentáló ábra alatt lesz.

Következő lépésben hasonlítsuk össze a kapott eredményt a *Berry-Esséen-egyenlőtlenség* által biztosított felső korláttal. Előtte azonban ellenőriznünk kell, hogy a harmadik abszolút momentum véges lesz-e.

2.1.2 Állítás: Legyen X gamma-eloszlású valószínűségi változó, ekkor:

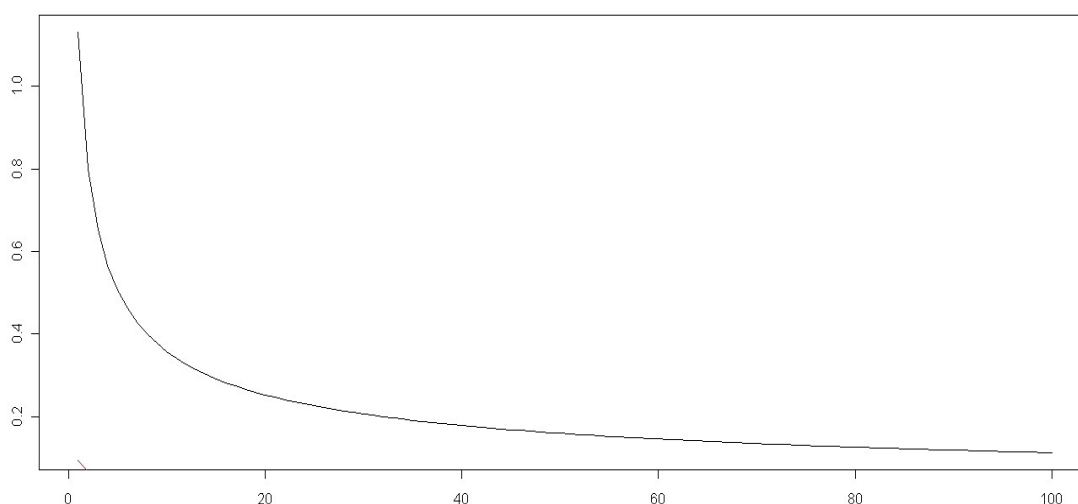
$$E|X^3| = \frac{\Gamma(p+k)}{\Gamma(p)\lambda^k}$$

Bizonyítás: Mivel gamma-eloszláshoz tartozó sűrűségfüggvény abszolút folytonos és pozitív, ezért:

$$E(X^k) = \int_0^{\infty} \frac{\lambda^p x^{p+k-1} e^{-\lambda x}}{\Gamma(p)} dx = \frac{\Gamma(p+k)}{\Gamma(p)\lambda^k} \int_0^{\infty} \frac{\lambda^p x^{p+k-1} e^{-\lambda x}}{\Gamma(p+k)} dx$$

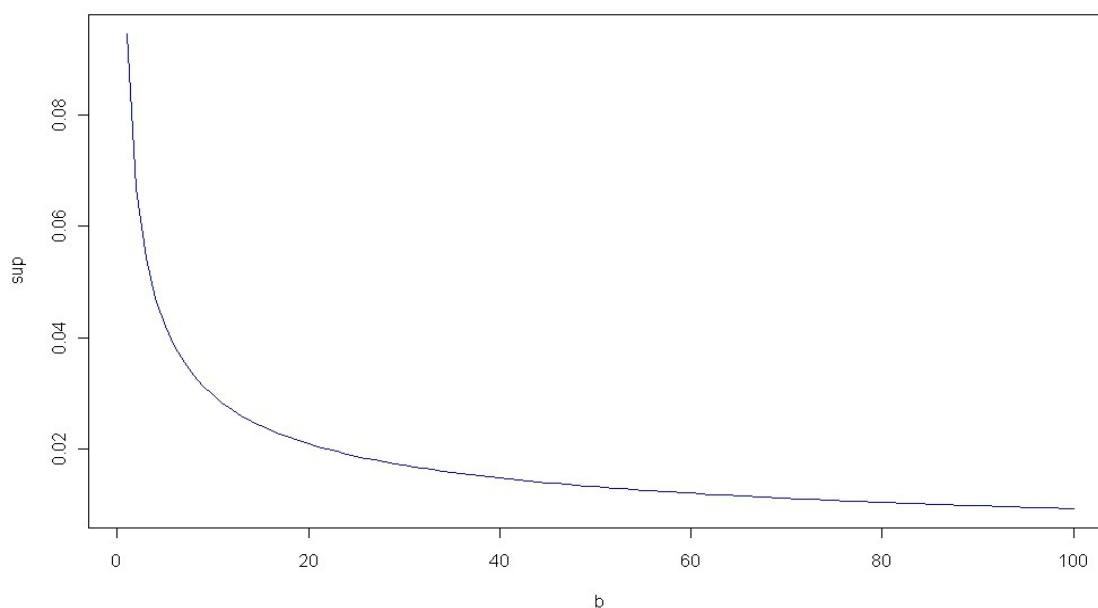
ahol a jobb oldalon szereplő integrálon belül egy sűrűségfüggvény van (λ és $p+k$ paraméterű, gamma eloszlásé), így annak értéke 1, így pont azt kapjuk eredményül, amire szükségünk volt.

A k helyére 3-at helyettesítve arra jutunk, hogy a harmadik abszolút momentum minden paraméter mellett pozitív lesz. Így nincs más dolgunk, mint a harmadik abszolút centrált momentumát kiszámolni a gamma-eloszlásnak, majd alkalmazni a Berry-Esséen-egyenlőtlenséget:



2.4 ábra

Amint az a 2.4-es ábráról kivehető, A *Berry-Esséen tétel* által biztosított felső korlát ebben az esetben jóval meghaladja a ténylegesen mért értékeket. Vizsgáljuk meg, hogy a paraméterek változtatásával a konvergencia hogyan is módosul. Első esetben legyen $p=2$, és válasszuk λ paraméter értékét 1.3; 10, 50, 100 -nak, ezeket az ábrán jelöljük rendre *fekete*, *piros*, *zöld* és *kék* színekkel. Ilyen módon kapott supremum-görbék:



2.5 ábra

Hogy csak egy kék színű görbét látunk, az azt jelenti, hogy a görbék egymást takarják le. Mivel a $\lambda=100$ esetet vizsgáltam utoljára, a neki megfelelő *kék* szín maradt fölül. Ebből azt a következtetést vonhatjuk le, hogy a standard normális eloszláshoz való konvergenciája a gamma-eloszlásnak nem függ a λ paraméter megválasztásától. Ezzel szemben ha rögzítjük λ értékét, akkor számítógép nélkül is belátható, hogy a p értékének növelésével arányosan nő a konvergencia sebessége is. Amint már korábban láttuk, (λ, p) paraméterű, gamma-eloszlású valószínűségi változók összege is gamma-eloszlású lesz $(\lambda, 2p)$ paraméterekkel. Így p értékének növelésével a kiinduló standardizált eloszlásunk egyre közelebb kerül a standard normális eloszláshoz. Nyilvánvalóan a *supremum-görbék* is egymás alá fognak esni.

2.2. Béta-eloszlás

Ezen eloszlás vizsgálata előtt is meg kell néznünk, hogy a szórásnégyzet milyen paraméterekre lesz véges (felhasználva az 1.4 definíciót):

$$D^2(X) = \left(\frac{1}{\alpha + \beta}\right) \frac{\alpha\beta}{\alpha + \beta + 1}$$

Mivel α, β egyaránt pozitív, ezért minden paraméterre alkalmazható a tétel.

Sajnos ebben az esetben nincs olyan nagy szerencsénk, mint a gamma eloszlás esetében. Ha veszünk n darab független, azonos, α, β paraméterű, béta-eloszlású valószínűségi változót, akkor annak az összegének a sűrűségfüggvényének sajnos nincs olyan szép zárt alakja, mint a gamma-eloszlásnak volt. Így kénytelenek vagyunk szimulációsán vizsgálni. Erre újra a *Matlab* programot hívjuk segítségül. Ezenkívül felhasználjuk a statisztika alaptételét:

2.2.1 Tétel (Statisztika alaptétele): Legyenek X_1, \dots, X_n független, azonos F eloszlásfüggvényű valószínűségi változók. Ekkor az F_n tapasztalati eloszlásfüggvény 1 valószínűséggel egyenletesen tart F -hez, azaz:

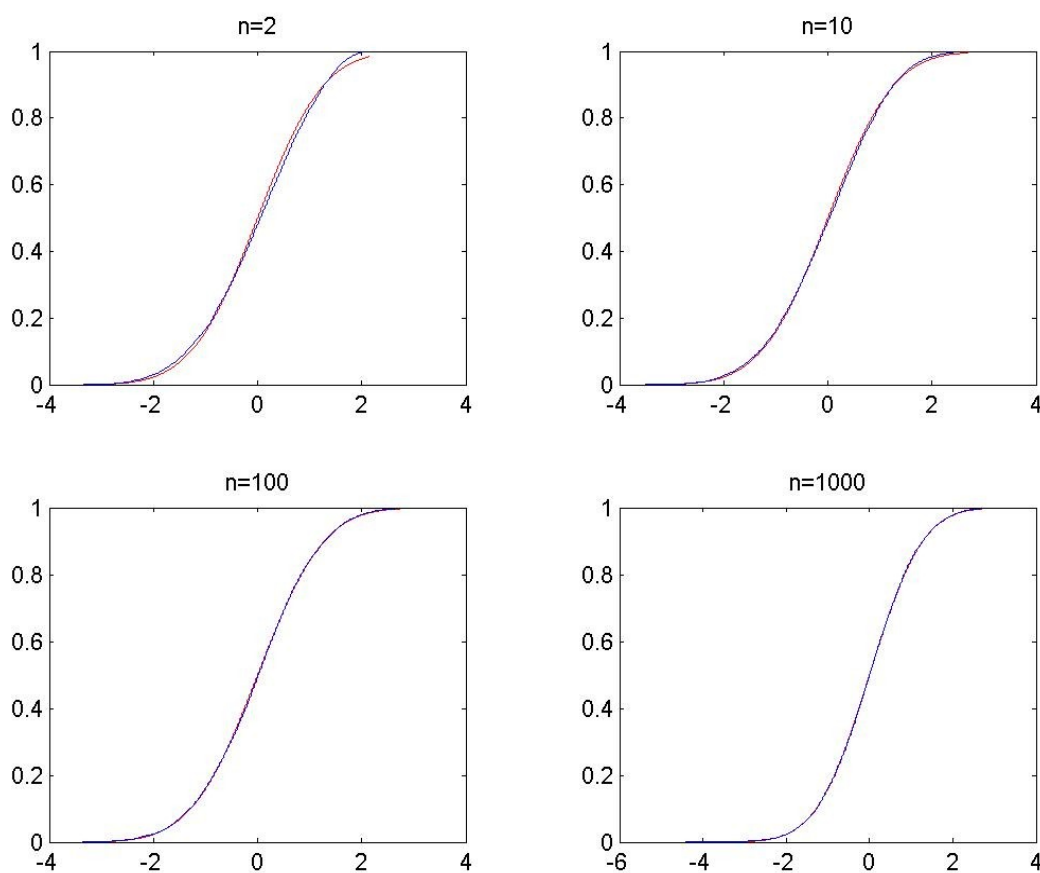
$$P\left(\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1$$

A tétel jelentése az, hogy ha elég sok megfigyelést végzünk, akkor tetszőleges pontossággal visszakapjuk a valódi eloszlást.

Ezt a gyakorlatban a következőképpen valósíthatjuk meg. Generálunk egy $n \times k$ -as véletlen mátrixot, melynek az elemei béta-eloszlásúak. Majd a k darab oszlopban szereplő elemeket összeadjuk. Az így kapott sorvektorban szereplő elemek közelítőleg normális eloszlásúak lesznek, tehát standardizálnunk kell. Mivel egy k elemű normális eloszlású véletlen számsorozatunk van, erre tekinthetünk úgy, mint egy k elemű normális eloszlású mintára. Ekkor a mintának a várható értéke torzítatlanul becsülhető a mintaátlaggal (*empirikus középpel*), a szórásnégyzet pedig a *varianciával*. Így ezen értékekkel standardizálhatjuk a mintánkat.

2.2.2 Megjegyzés: Számolhatunk elméleti várható értékkel és szórásnégyzettel is, ugyanis a szimuláció többszöri lefuttatása alkalmával az empirikus értékek várható értékeik maguk az elméleti paraméterek lesznek.

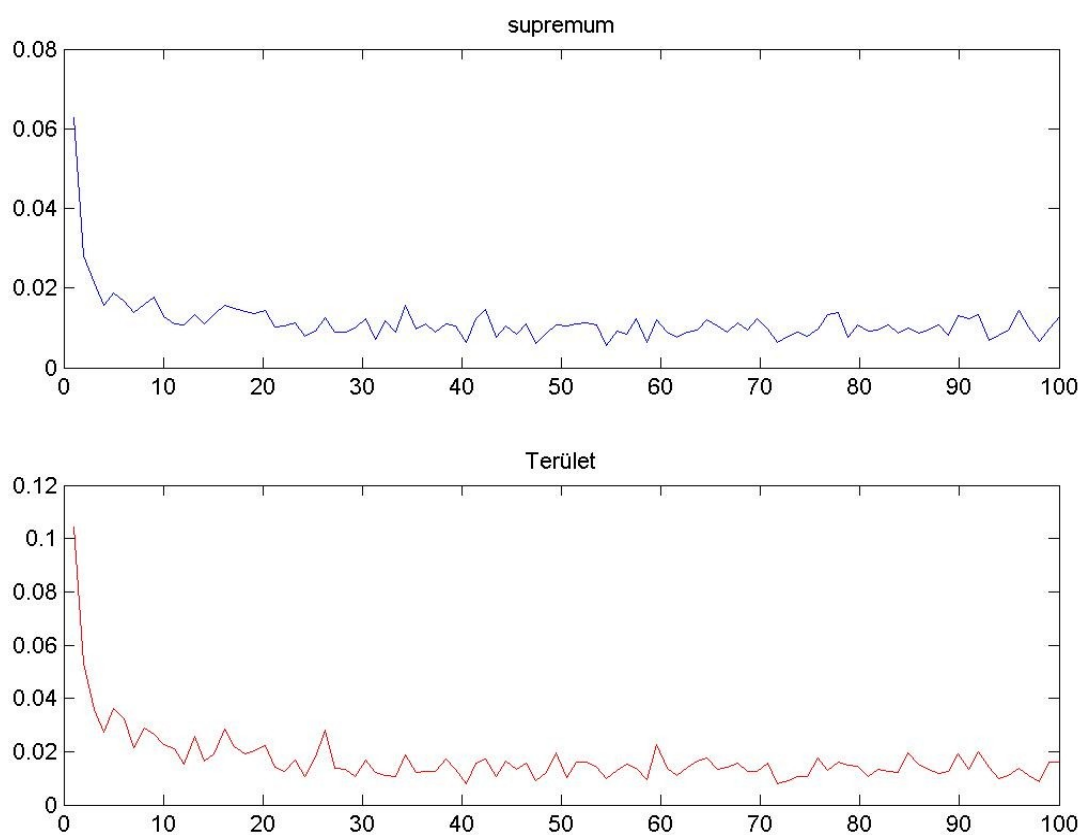
Standardizálás után a tapasztalati eloszlást könnyen meghatározhatjuk, majd ábrázolhatjuk is, felhasználva a tapasztalati eloszlás definícióját. Bevezetünk egy vektort, aminek a mérete megegyezik a minta elemszámával. Ez legyen k . Ekkor ez a vektor tartalmazza az $1/k$ és $(k-1)/k$ közötti ekvidisztans felosztás elemeit. Ezen vektor i . eleme pontosan az i . rendezett mintaelemhez tartozó tapasztalati valószínűség lesz. Így különböző számú (n db) béta-eloszlású valószínűségi változók összegére a 3.1-es ábrán látható eredményeket kapjuk. Ezeken az ábrákon kék színnel van feltüntetve a béta-eloszlású valószínűségi változók összegének standardizált eloszlása, pirossal pedig a standard normális eloszlás. A programot ötezer darabos mintára $\alpha=2$ és $\beta=1.2$ paraméterekre futtattam.



3.1.ábra

Jól kivehető, hogy 2 és 10 darab valószínűségi változó esetén még elkülönül a két görbe (habár egyre jobban rásimulnak a standard normális eloszlásra). Ezzel szemben 100 és 1000 darab valószínűségi változó esetén szinte teljesen egybeolvadnak. Gyakorlatilag ezen utóbbi két k -hoz tartozó ábra megkülönböztethetetlen. Ez szolgál némi információval a konvergencia sebességéről.

A konvergencia sebességet újra a már gamma-eloszlás vizsgálatánál használt módszerek segítségével végezzük, (supremum, területek különbsége).



3.2 ábra

Amint az jól látható, itt már nem olyan egyértelmű a helyzet, mint a gamma-eloszlás esetén, abból adódóan, hogy ebben az esetben kénytelenek voltunk szimulálni az összeget, ugyanakkor érdemes megjegyezni, hogy a statisztika alaptétele miatt, ha sokkal több mintaelemünk van (más szóval többször végezzük el a kísérletet) akkor az ábránk nyilván sokkal egyértelműbb lesz. Ezután elkezdhetjük a vizsgálódást a *Berry-Esséen tétel* segítségével is. Ahhoz hogy alkalmazhassuk a tételt, ellenőriznünk kell, hogy a harmadik abszolút momentum véges lesz-e. Könnyű dolgunk van, mivel a béta-eloszlású valószínűségi változók pozitívak, így elég a harmadik momentumot megvizsgálni:

A harmadik momentumot ki tudjuk számolni a sűrűségfüggvény felhasználásával:

$$E(X)^3 = \int_{\mathbb{R}} x^3 f(x) dx$$

ahol $f(x)$ az X valószínűségi változóhoz tartozó sűrűségfüggvény. Mindezt azért tehetjük meg, mert a béta-eloszlás abszolút folytonos. Alkalmazva a képletet:

$$E(X)^3 = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+2} (1-x)^{\beta-1} dx$$

Látszik, hogy a jobb oldalon szereplő integrál véges, és a már korábban látott béta-integrál (Mivel ha a jobboldalon szereplő integrált leosztjuk $B(\alpha+3, \beta)$ -val, akkor 1-et kapunk). Így a harmadik momentum:

$$\frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 3)}{\Gamma(\alpha)\Gamma(\alpha + \beta + 3)}$$

ami akkor lesz végtelen, ha az α -t 0 -nak, és ha $\alpha + \beta = -3$ -nak választjuk, de ilyen paramétereket nem is vehet fel a béta-eloszlás, ez is alátámasztja azt, hogy a béta-eloszlás korlátossága miatt minden momentum véges.

2.2.3 Megjegyzés: Bármilyen k esetén tetszőleges α, β párra véges lesz, mivel a béta-eloszlásfüggvény értelmezési tartománya korlátos.

Eszerint alkalmazhatjuk a *Berry-Esséen-egyenlőtlenséget*. A következő probléma, amibe beleütközünk, az a *harmadik centrált* abszolút momentum értékének meghatározása. Próbáltam a *Matlab* program segítségével megoldani, mégsem voltam sikeres. Úgy döntöttem, hogy approximálva fogom kiszámolni, mégpedig az összetett trapézformulát felhasználva:

2.2.4 Állítás (összetett trapézformula, és hibája): Legyen m az osztópontok száma

$$\int_a^b f(x) dx \sim h \frac{\sum_{j=0}^{m-1} f(x_j) + f(x_{j+1})}{2} \quad \text{ahol} \quad h = \frac{(a+b)}{m}$$

hibája:

$$\left| \int_a^b f(x) dx - h \frac{\sum_{j=0}^{m-1} f(x_j) + f(x_{j+1})}{2} \right| \leq (b-a) \|f\|_c \frac{h^2}{12}$$

Ez alapján becsülni tudjuk a pontatlanságát az integrálunknak, ami 5000 osztópont esetén. Ehhez szükségünk lesz $\|f\|_c$ -re, ami

$$f'(x) = \frac{x^{\alpha-2}(1-x)^{\beta-2}}{B(\alpha, \beta)} ((\alpha-1) - x(\alpha+\beta-2))$$

Melynek nyilvánvalóan gyökei az 1 és 0, viszont azok minimumok lesznek, tehát a maximum csak egy helyen vétetik fel:

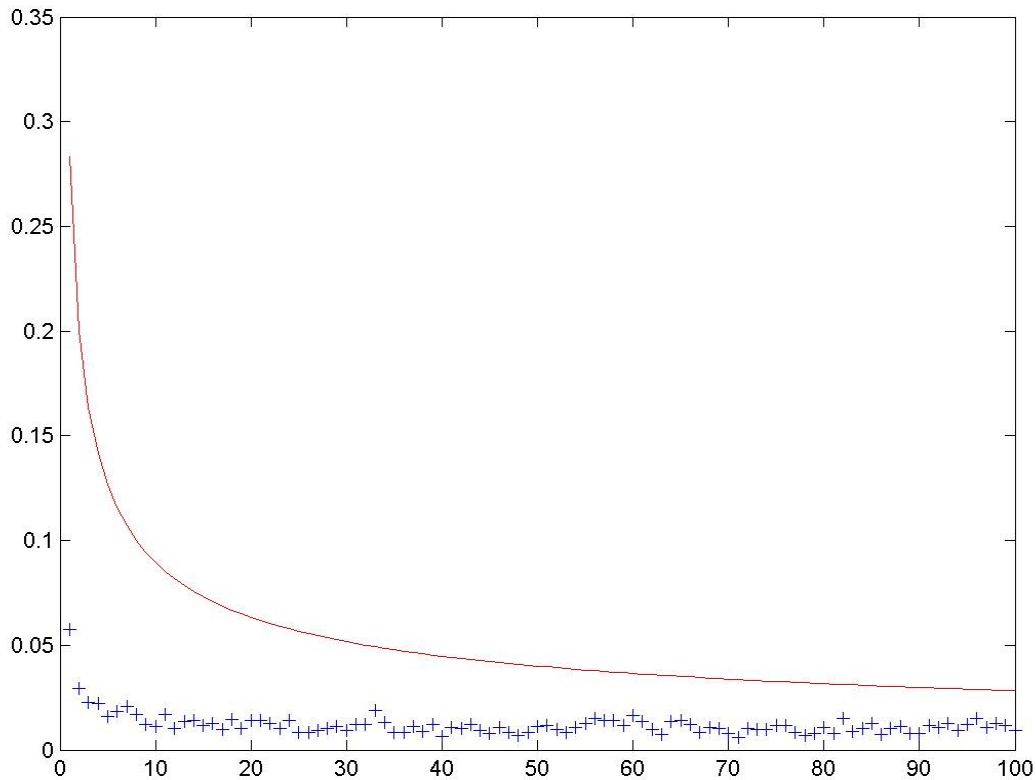
$$x = \frac{\alpha-1}{\beta+\alpha-2}$$

Így $\alpha=2.3$ és $\beta=1.4$ paraméterek mellett:

$$\|f\|_c = f(0.7647)$$

Aminek az értéke megközelítőleg 1.5963, így a hiba: 0.000016667, tehát az eredményekre nem lesz nagy hatással az approximációból adódó hiba.

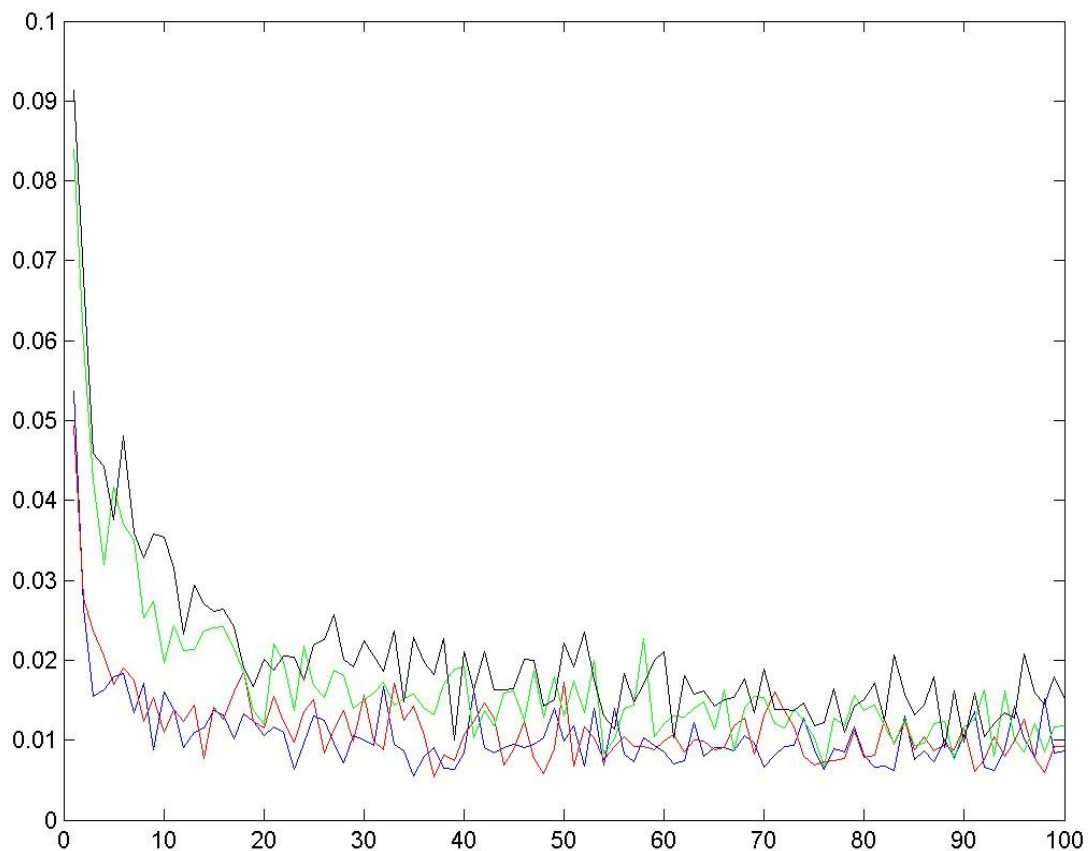
Így már minden adott, hogy lássuk, a szimulációnk eredménye mennyire kerül közel a *Berry-Esséen-egyenlőtlenség* által meghatározott felső korláthoz.



3.3 ábra

Ahol a *piros görbe* a *Berry-Esséen tétel* által megadott felső korlát n függvényében, a *kék kereszt* pedig az eltérések abszolút értékének supremuma. Ezt is 5000 mintaelemre és a már korábban említett $\alpha = 2.3$ és $\beta = 1.4$ paraméterekre van futtatva.

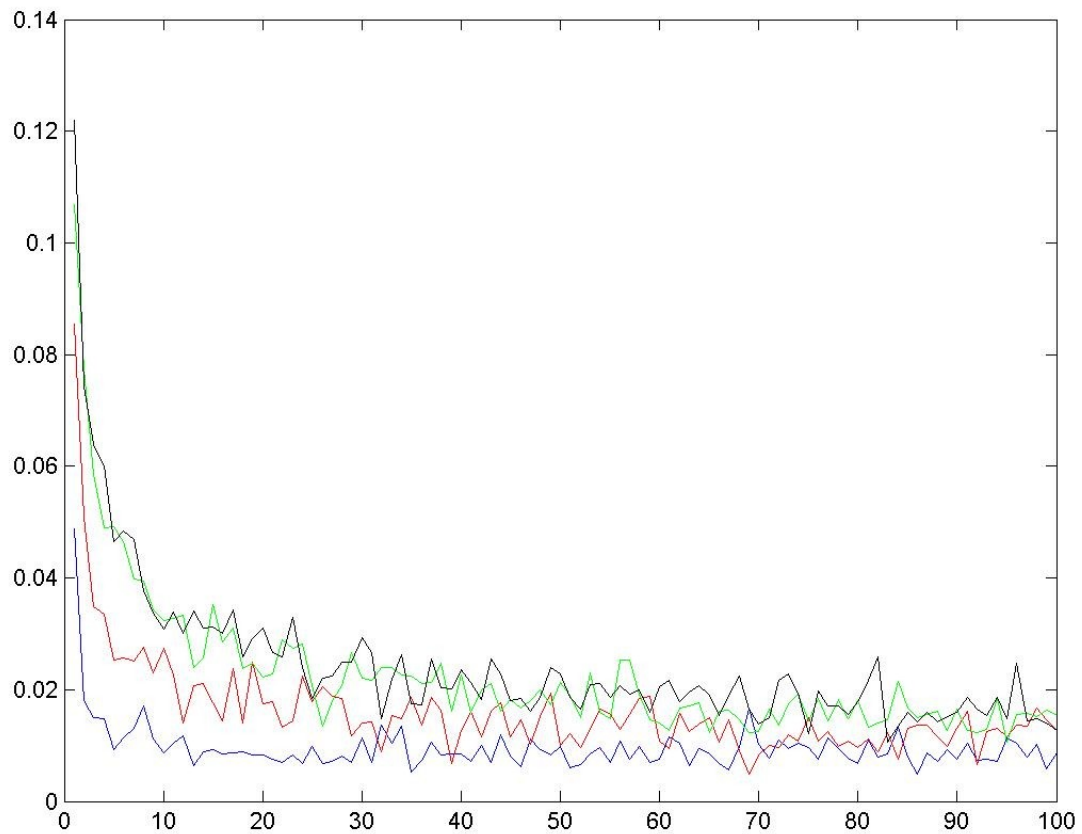
Most is megvizsgálhatjuk, hogy ez a konvergenciát reprezentáló ábra hogyan változik, ha a paramétert változtatjuk. Mivel ebben az esetben két paraméterünk van, érdemes úgy vizsgálni, hogy az egyik paramétert rögzítjük és a másikat változtatjuk, így mondjuk, ha α értékét rögzítjük 2.3-nak, illetve ha β -át 1.4 (*kék*); 2.5 (*piros*); 20 (*zöld*); 100 (*fekete*) értékeknek választjuk akkor az alábbi eredményeket kapjuk:



3.4 ábra

Amint az a 3.4-es ábrán is jól látható a β paraméter növelésével a két eloszlás függvény különbségének a supremuma egyre lassabb ütemben konvergál 0-hoz. Így a standardizált eloszlásfüggvény is egyre lassabban konvergál a standard normális eloszláshoz. (Megközelítőleg a görbék egymás felett vannak.)

Természetesen ugyanezt a vizsgálatot elvégezzük arra az esetre, ha a β paramétert rögzítjük (esetünkben 1.4 értéken), és az α paraméterhez a 2.3, 5, 20, 100 értékeket rendelünk, melyeket rendre, kék, piros, zöld fekete színekkel jelölünk az ábránkon.



3.5 ábra

A 3.5-ös ábrán is hasonló eredményt láthatunk, mint a 3.4-esen, tehát az α paraméter növelése ebben az esetben is a konvergencia lassulásához vezet. Így ebben az esetben is szemmel fel tudunk állítani egy nagyság szerinti sorrendet a görbék között (növekvő: *kék, piros, zöld, fekete*).

2.3. Pareto-eloszlás

Végül elérkeztünk az utolsó eloszláshoz, ami több szempontból is nagyon érdekes. Most is – ugyanúgy, mint korábban – ellenőriznünk kell, hogy a szóban forgó eloszlású valószínűségi változó milyen paraméterek mellett elégíti ki a *Ch*t alkalmazásához szükséges feltételt, azaz a szórás mikor lesz véges.

2.3.1 Állítás (*Pareto-eloszlású valószínűségi változó szórása*):

$$D(X) = \sqrt{\frac{\alpha\beta^2}{\alpha-2} - \left(\frac{\alpha\beta}{\alpha-1}\right)^2}$$

Ami akkor lesz véges, ha α paraméter értéke nagyobb, mint 2. Így megkezdhetjük a vizsgálatokat. Sajnos most is abba a problémába ütközünk, mint a gamma-eloszlás esetében. Azonos, Pareto-eloszlású független valószínűségi változóknak az összegének sűrűségfüggvényének nincs szép zárt alakja. Tovább nehezíti a helyzetet, hogy se a *Matlabban*, se az *R*-ben nincs Pareto-eloszlású véletlen számgenerátor. Ez utóbbi nehézséget könnyen fel tudjuk oldani, ugyanis, ha Pareto-eloszlásfüggvény *inverzébe* behelyettesítünk véletlen számokat, amik (0,1) intervallumon egyenletes eloszlásúak, akkor a kapott számok véletlen számok maradnak, csak Pareto-eloszlásban. Mindez általánosan is megfogalmazható:

2.3.2 Állítás: Legyen az $F(t)$ egydimenziós eloszlásfüggvény (*balról folytonos*) általánosított inverze:

$$F^{-1}(t) = \sup\{s : F(s) < t\} = \inf\{s : F(s) \geq t\}, 0 < t < 1$$

Ekkor, ha U a (0,1) intervallumban egyenletes eloszlású valószínűségi változó, akkor $X = F^{-1}(U)$ eloszlásfüggvénye éppen F lesz.

2.3.3 Megjegyzés: A 2.3.2 állítás, triviális abban az esetben, ha F abszolút folytonos, mivel ekkor az általánosított inverz maga az inverz lesz, így:

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x))$$

De mivel u abszolút folytonos valószínűségi változó, ezért adódik:

$$P(u \leq F(x)) = F(x)$$

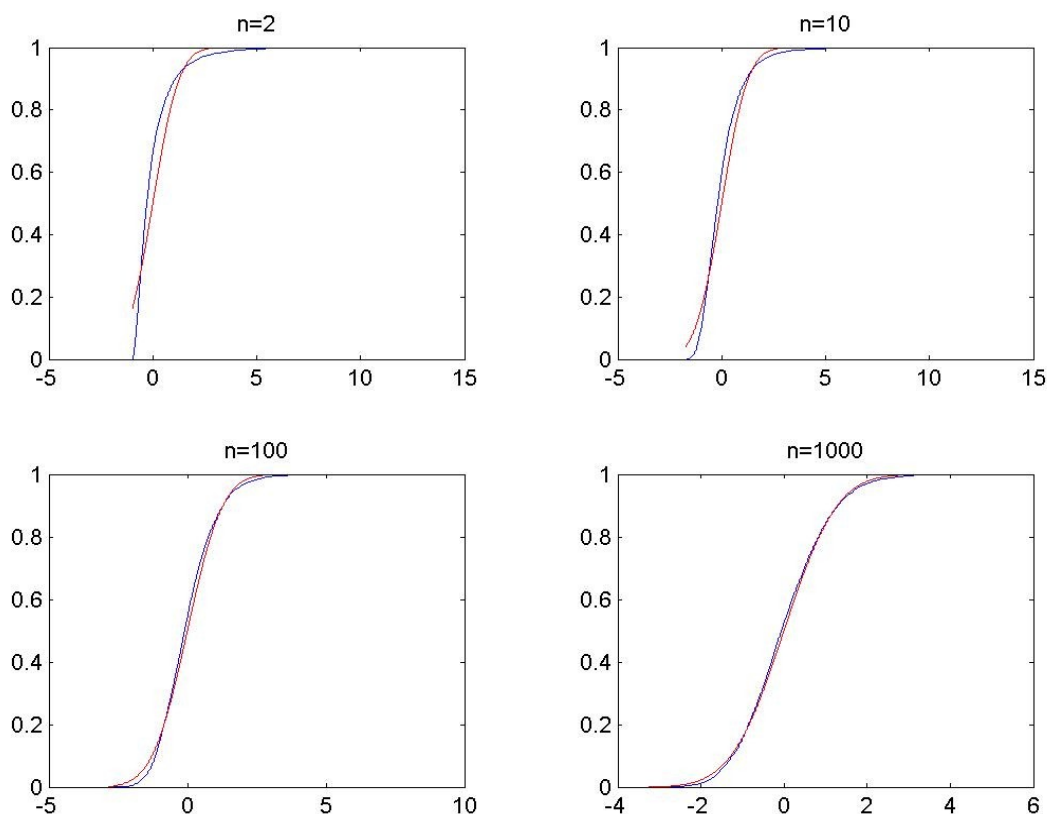
Tehát nincs más dolgunk, mint a Pareto-eloszlás eloszlásfüggvény-inverzének meghatározása. Ennek a kiszámolásához felhasználhatjuk azt, hogy a Pareto-eloszlás

abszolút folytonos, továbbá azt, hogy maga az eloszlás $F(t) = 1 - \left(\frac{\beta}{t}\right)^\alpha$, így az

általánosított inverz meg fog egyezni az inverzzel, ami:

$$F^{-1}(t) = \frac{\beta}{(1-t)^{\frac{1}{\alpha}}}$$

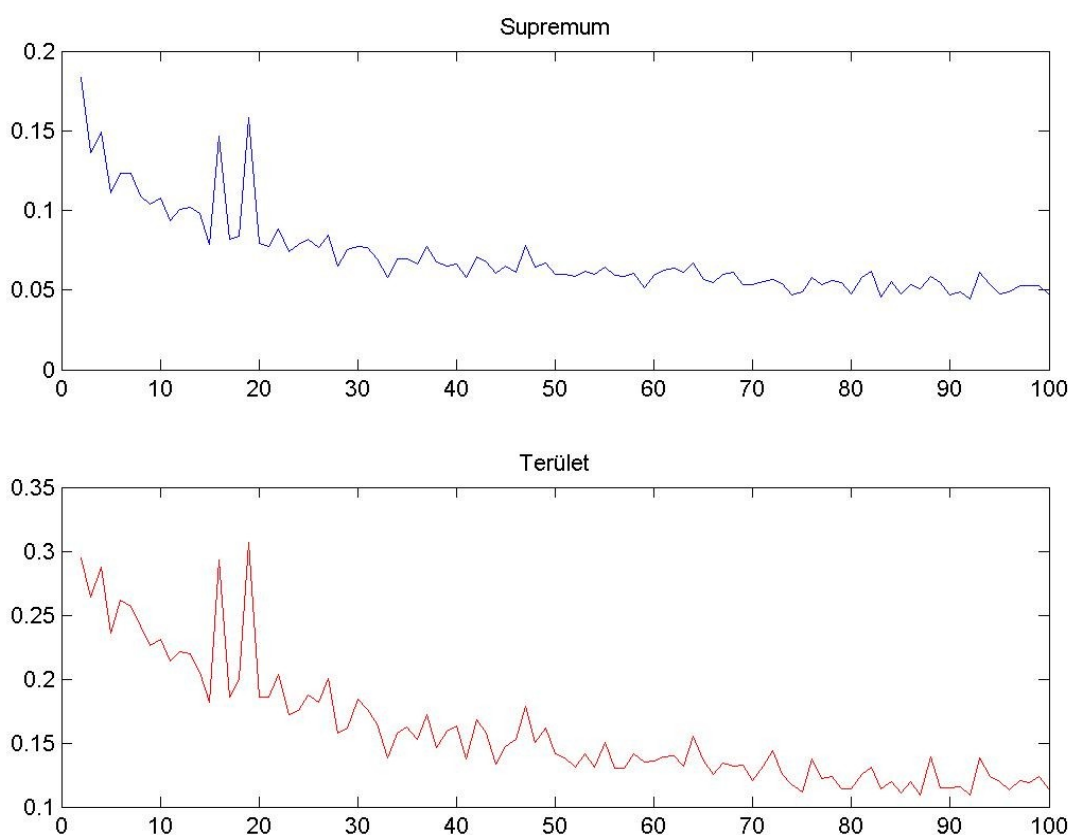
Miután ezzel megvagyunk, tudunk Pareto-eloszlású véletlen számokat generálni, hiszen egyenletes eloszlású véletlenszám generátort tartalmaz a *Matlab* program. Innentől a Béta-eloszlásnál használt algoritmust ugyanúgy alkalmazhatjuk erre az esetre is. Mindezek után az alábbi eredményeket kapjuk:



4.1ábra

Ezek az eredmények $\alpha=3.3$ és $\beta=0.4$ paraméterek és 5000 darabos mintaelem mellett születtek.

Az ábrákon megfigyelhetjük, hogy a Pareto-eloszlású valószínűségi változók konvergenciája jóval lassabb a korábban vizsgált két eloszláshoz képest, ugyanis ebben az esetben még az 1000 db független valószínűségi változó összegének az eloszlása is szemmel könnyen elkülöníthető a standard-normális eloszlás-görbétől. Következő lépésben megvizsgáljuk a konvergencia sebességét a már korábban is használt két módszerrel:



4.2 ábra

Ezek az eredmények szintén $\alpha=3.3$ és $\beta=0.4$ paraméterek és 5000 darabos mintavétel mellett születtek. A fentebb említett észrevételünket a konvergencia lassúságáról alá is támasztják a 4.2 ábrán látható eredmények, tényleg jóval lassabban konvergálnak az értékek 0-hoz, mint a gamma-, illetve a Béta-eloszlás esetén. Pareto-eloszlás

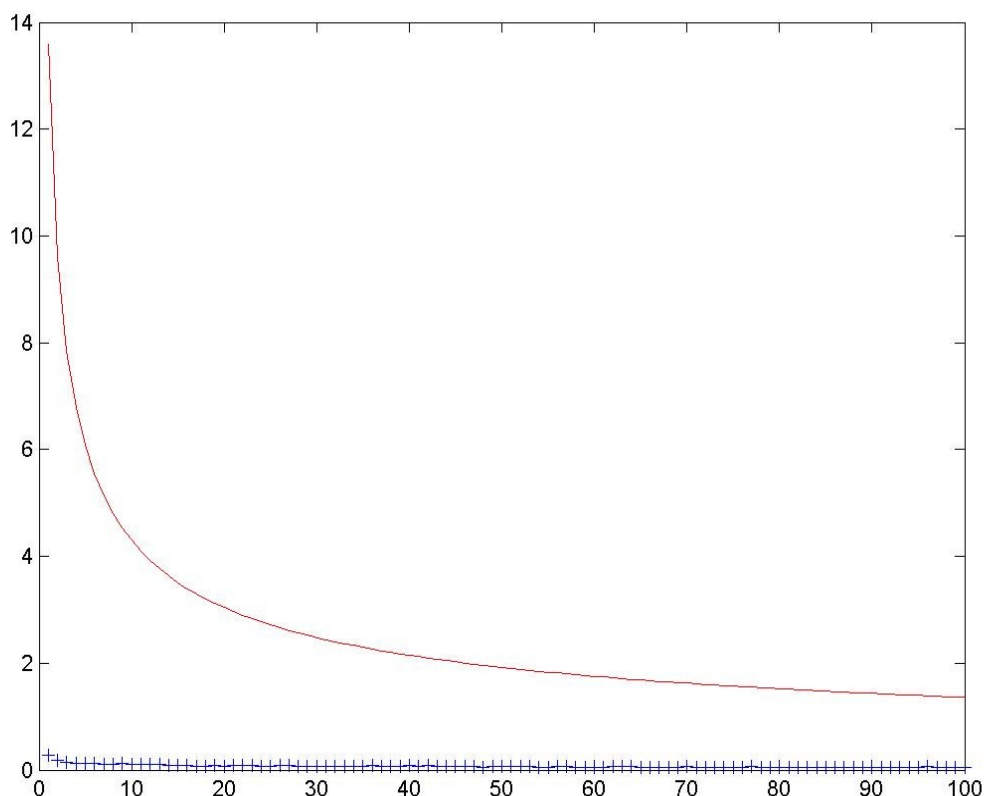
vizsgálatait ebben az esetben is a *Berry-Esséen tétel* által meghatározott felső korlát és a 4.1 ábrán is szereplő eredmények összehasonlításával zárom. Most is szükségünk van arra, hogy a Pareto-eloszlású valószínűségi változó harmadik abszolút momentuma milyen paraméterek mellett lesz véges.

2.3.4 Állítás: Pareto-eloszlású valószínűségi változó harmadik abszolút momentuma:

$$E|X|^3 = \frac{\alpha \beta^3}{\alpha - 3}$$

ami akkor lesz véges, ha α paraméter értéke meghaladja a 3-at.

Mivel a már korábban használt paraméterek ($\alpha=3.3$ és $\beta=0.4$) esetén a szórás véges lesz használhatjuk őket a Pareto-eloszlás paramétereinek. Az így kapott eredmény az alábbi:

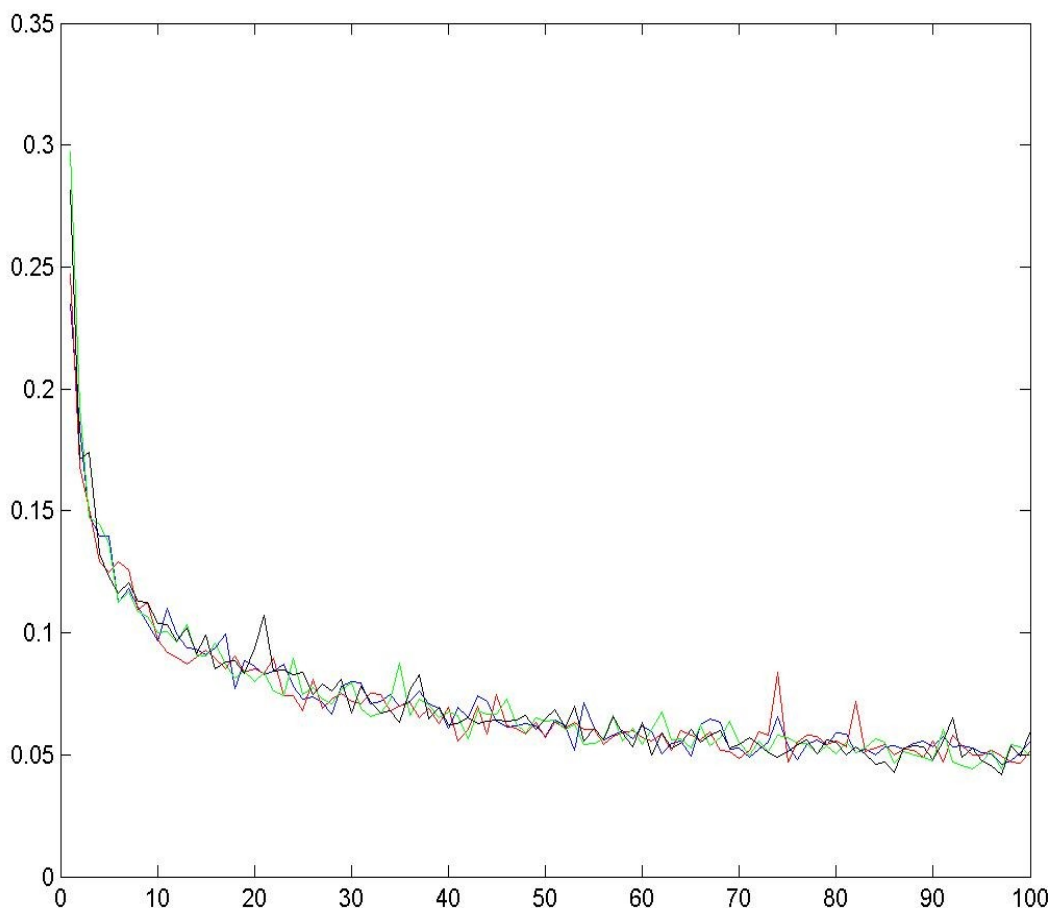


4.3 ábra

Ahol a piros görbe a *Berry-Esséen* tétel által megadott felső korlát a valószínűségi változók számának a függvényében ábrázolva, a kék kereszt jelenti a supremumot, szintén a valószínűségi változók számának a függvényében. Amint azt az ábra is mutatja, erre a paraméterpárra a felső korlát jóval meghaladja az *1.6 tételben* szereplő egyenlőtlenség bal oldalát.

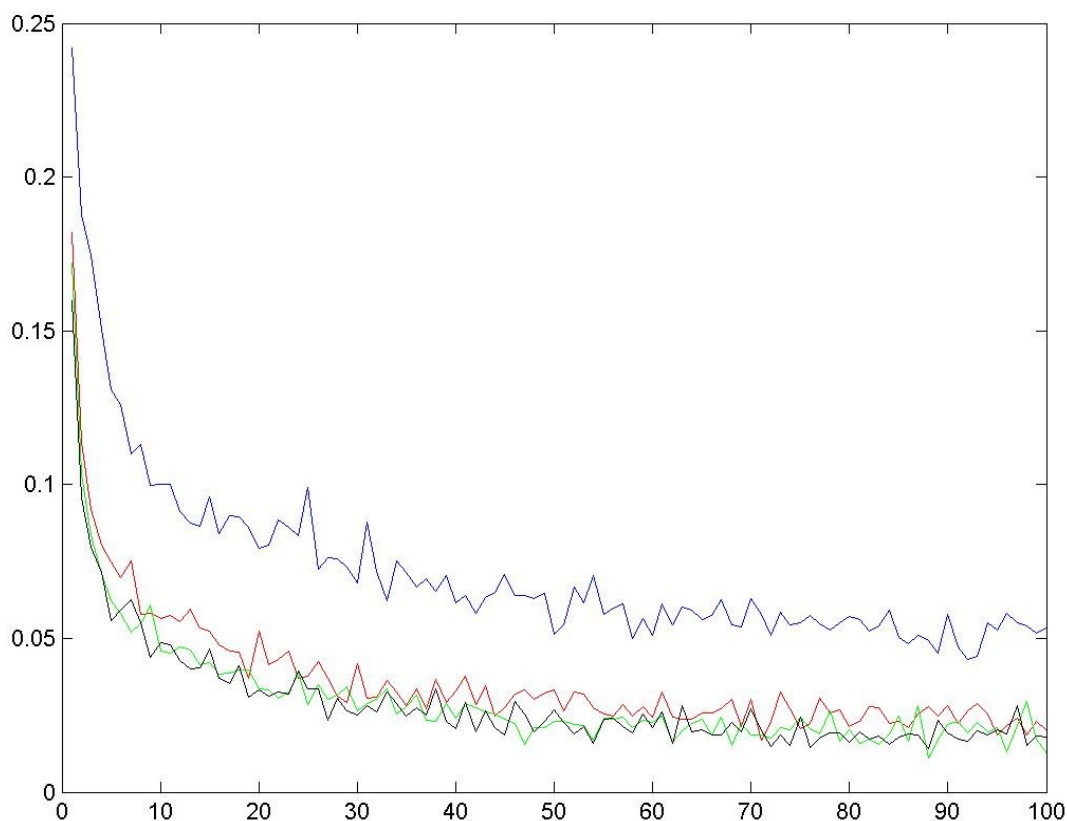
Végül vizsgáljuk meg a Pareto-eloszlás esetében is azt, hogy a paraméterek megváltoztatásával hogyan változik a konvergencia sebessége.

Első esetben rögzítsük α értékét 3.3-nak. Jelöljük rendre a $\beta=0.4$, $\beta=5$, $\beta=20$ és $\beta=100$ paraméterekhez tartozó szimulációs konvergenciagörbéket *kék, piros, zöld és fekete* színekkel. Ekkor az alábbi eredményeket kapjuk:



4.4 ábra

Amint az a 4.4-es ábrán jól látható, a β paraméter változtatásával (növelésével) szemmel láthatóan nem voltunk hatással a konvergenciára, elvégre a különböző színű görbék nem nagyon különíthetők el egymástól. A másik paraméter vizsgálata esetén már más lesz a helyzet. Ebben az esetben β paraméter értékét rögzítsük 0.4-nek, és tekintsük a konvergenciát reprezentáló ábrát $\alpha=3.3$, $\alpha=10$, $\alpha=50$, és $\alpha=100$ paraméterek mellett, melyeknek színei legyenek rendre: *kék, piros, zöld és fekete*.



4.5 ábra

A 4.5-ös ábrán egyértelműen látszik, hogy a konvergencia $\alpha=3.3$ estén a leglassabb, az utána következő konvergenciát $\alpha=10$ esetében figyelhetjük meg. A szimulációból adódó oszcilláció miatt $\alpha=50$ és $\alpha=100$ esetében már nem ilyen egyértelműen eldönthető a sorrend. Azért megközelítőleg látható, hogy $\alpha=50$ felette van az $\alpha=100$ -hoz tartozó görbének. Így megfigyeléseink alapján arra a következtetésre juthatunk, hogy az α paraméter növelésének arányában nő a konvergencia sebessége.

3. Összefoglalás

A harmadik fejezetben sikerült számítógéppel is a *Berry-Esséen tétel* által adott felső korlát helyességét igazolni, ugyanakkor sikerült következtetéseket levonni arról, hogy mennyire közelíti meg a tényleges konvergenciát. Láthattuk, hogy a béta-eloszlás esetében közelítette meg a legjobban a standardizált gamma-eloszlás és a standard normális eloszlás-görbék között mért távolságok supremumát, ezzel szemben Pareto-, illetve gamma-eloszlás esetén jóval meghaladta azt.

Megvizsgáltuk a centrális határeloszlás-tételben szereplő konvergenciát úgy is, hogy az adott eloszlású valószínűségi változók összegének standardizált és a standard normális eloszlás görbék által közrezárt terület változása hogyan függ az összeadott valószínűségi változók számától. Ekkor arra a következtetésre jutottunk, hogy a területek is 0-hoz fognak tartani. Továbbá ennek a konvergenciának a sebessége hasonló a supremumoknál megfigyelt konvergencia sebességéhez.

Végül megvizsgáltuk, hogy a supremumok konvergenciája a különböző eloszlások esetén hogyan változik a paraméterek változtatásával. Ekkor arra a következtetésre jutottunk, hogy p rendű, λ paraméterű gamma eloszlás esetén p paraméter növelésével gyorsítottuk a konvergenciát, viszont λ paraméter megváltoztatása nem volt hatással a konvergencia sebességére. α , β paraméterű Béta-eloszlás esetén úgy α , mint β paraméterek növelésével arányban nőtt a sebesség is. Végül a Pareto-eloszlás esetében azt tapasztaltuk, hogy a definícióban szereplő β paraméter megváltoztatásával a konvergencia sebessége nem változik, ezzel szemben α növelésével a konvergencia csökken.

Irodalomjegyzék

[1] W. Feller, An Introduction to Probability Theory and its Applications Vol. 2 –
Kiadó: Wiley, 1970

[2] Po-Ning Chen, Asymptotic Refinement of the Berry-Esséen Constant
2002, *Elérhetőség: <http://shannon.cm.nctu.edu.tw/html/paper/be02.pdf>*

[3] Arató Miklós, Valószínűségszámítás – *internetes jegyzet*,
2008 *Elérhetőség: http://www.cs.elte.hu/~arato/valoszinusegszamitas_2007.pdf*

[4] Stoyan Gisbert, *Matlab*
Kiadó: Typotex, 2008

[5] Abari Kálmán, Gyakori R parancsok – *Internetes oktatási segédlet* *Elérhetőség:*
http://psycho.unideb.hu/munkatarsak/abari_kalman/szamitastechnika_II/gyakori_r_parancsok.pdf