Budapesti Műszaki és Gazdaságtudományi Egyetem
Természettudományi Kar

TÉMALABOR 2

Fluktuáció analízis

Impedimetriai vizsgálatok során regisztrált biológiai és háttér zajok elkülönítése

Készítette
Schrempf Dóra

Témavezető
Dr. Láng Orsolya
Semmelweis Egyetem,
Genetikai, Sejt- és Immunbiológiai Intézet

Konzulens
Farkas Lóránt
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Analízis Tanszék

2017.

Tartalomjegyzék

1. fejezet

Bevezetés

Az előző félévben kezdtem el foglalkozni sejtek kutatásával, vizsgálatával, melyben nagy segítségünkre volt a matematika tudománya. A Témalabor 1. tantárgy elvégzéséhez kitűzött célunk az volt, hogy megismerjük közelebbről a sejteket, megértsük az úgynevezett ECIS (Electric Cell-Substrate Impedance Sensing) készülék működését és saját eredményként rutint írjunk a gép által generált adatsorok egyes statisztikai paramétereinek (úgy, mint az átlag, variancia és inkrement) kiszámítására. Az idő függvényében kapott rezisztencia értékek tulajdonságai szerint kerestünk magyarázatot a grafikonban jól látható fluktuációra, a grafikon "szöszösségére", azaz az adatok zajjal való terhelésére, mely összefüggésben van az ECIS készülék elektródáját borító sejtek mozgásával, az ún. mikromotion-nel. Egy kutatócsoport kísérletei kimutatták, hogy I. Giaver és C. R. Keese által a XX. században feltalált ún. Electric Cell-Substrate Impedance Sensing (ECIS) készülék által generált eredmények nagy különbséget mutatnak daganatos és egészséges sejtek között. A mérés során a gépben található arany elektródok felületén kitapadó sejtek morfológiai változásai, adhéziója és migrációja is valós időben nyomon követhető. A készülékben külön mérhető az impedancia (Z), a rezisztencia (R) és a kapacitív ellenállás (C), mely értékeket az idő függvényében kapunk meg.

Ebben a félévben a feladatom, a Semmelweis Egyetem Genetikai, Sejt- és Immunbiológiai Intézet kutatócsoportja által végzett mérések során kapott idősorok jellemzésére alkalmas egyik fontos paraméternek, az ún. Hurst exponensnek a vizsgálata volt. Története a Nílus gátszabályozásáig nyúlik vissza, majd a matematika számos területén úgy, mint a fraktálok, a káosz elmélet, a hosszú távú folyamatok, illetve a spektrál analízis területén is elterjedt használata. Pénzügyi alkalmazása is nagyon ismert, azonban számunkra a biológiai , illetve orvostudományi alkalmazása a legfontosabb. Kíváncsiak voltunk arra, hogy alkalmas paraméter-e sejtes és sejtmentes minták elkülönítésére. Dolgozatomban a Hurst exponens kiszámítására alkalmas ötféle módszert is bemutatunk, majd ezek MATLAB-ban való beprogramozása után

EXCEL fájlban gyűjtjük össze a kapott eredményeket. A módszerek megismerésekor hátrányokba, akadályokba ütközünk, míg végül megtaláljuk a számunkra legmegfelelőbb módszert a Hurst exponens kiszámolására. Ez az úgynevezett GM2 módszer. Végül nincs más dolgunk, mint a matematikai statisztika segítségével megvizsgálni, hogy az utóbbi módszerrel meghatározott Hurst exponens ismeretében, vajon egy adatsorról szignifikánsan eldönthető-e, hogy az sejtes vagy sejtmentes mintából származik-e.

2. fejezet

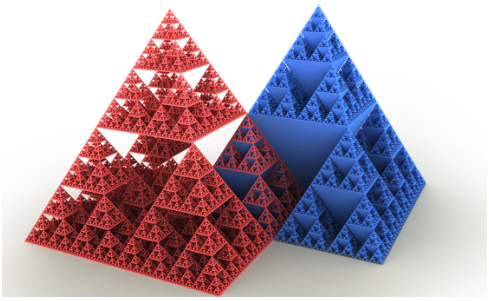
A Hurst exponens

2.1. Története

A Hurst exponens az alkalmazott matematika számos területén használt paraméter, úgy, mint a fraktálok, a káosz elmélet, a hosszú távú folyamatok, illetve a spektrál analízis. A Hurst exponens becslését a matematikusokon túl a biofizikusok, meteorológusok, pénzügyi szakemberek és az informatikusok is használják. Azonban eredete a hidrológiához kötődik, mégpedig az egyiptomi Nílus gátszabályozásához [1].

Edwin Harald Hurst britt hivatalnok 1906-ban került Kairóba, majd ott tartózkodása alatt beleszeretett a Nílusba, melynek tanulmányozásával, vizsgálatával teltek napjai. Látta, hogy az alsó gát nem tölti be a kívánt szerepét és szükség van egy víztárolóval kombinált nagyobb gátra. Számításaihoz szüksége volt az előző évi vízállásokra, ami hatalmas kutató munkát eredményezett. Megfigyelte, hogy a nagyobb áradásokat nagyobb valószínűséggel követnek nagyobb áradások és fordítva, kisebbeket pedig kisebb áradások. Tehát nem optimális a csapadék mennyiség átlagát venni és ez alapján kiszámítani a gát, illetve víztároló szükséges kapacitását, hanem nagyobb gátat kell építeni, mint ahogy azt a normális eloszlás adná. Ez a megfigyelés vezetett a róla lenevezett Hurst módszerhez [2].

Később egy újfajta geometriai objektumnál, a fraktáloknál játszott fontos szerepet a Hurst exponens modern technikákkal való becslése. Ezen objektumok népszerűsítése egy amerikai matematikus, Benoit Mandelbrot nevéhez fűződik. A fraktál geometria a tudomány szinte minden területén fellelhető, amely segít a körülöttünk lévő világ más szemszögből való szemlélésében. Fraktáloknak két csoportját különböztetjük meg, úgy, mint a szabályos és a véletlen fraktálok. Utóbbi csoportba tartoznak a természetben fellelhető fraktálok, melyekkel már mindenki találkozhatott. Ilyen fraktál például a páfrány levél és a Sierpinski piramis.



Ezek fő tulajdonsága az önhasonlóság, melynek jelentése, hogy az adott objektum egy része hasonló magához az egész objektumhoz. Egy új, úgynevezett fraktáldimenzió bevezetésére volt szükség, ami a következő összefüggésben áll az általunk vizsgált Hurst exponensel: $D = 2 - H$, ahol H jelöli a Hurst exponenst, D pedig a fraktáldimenziót. Miszerint egy nagyobb Hurst exponenssel rendelkező idősor kisebb fraktáldimenzióval, azaz simább felszínnel rendelkezik. Míg egy kisebb Hurst exponenssel rendelkező idősor nagyobb fraktáldimenzióval, azaz érdesebb felszínnel rendelkezik [3] [4].

2.2. Pénzügyi alkalmazása

A Hurst exponens pénzügyi, tőzsdei vonatkozásban alkalmas annak vizsgálatára, hogy a mérsékelt kockázatú befektetési alapok, pénzpiaci-, hosszú kötvény-, és ingatlanalapok árfolyamának változása rendelkezik-e hosszú emlékezettel egy adott időszakban.
Annak ismerete, hogy az adott alap rövid- vagy hosszú emlékezettel rendelkezik olyan információt biztosít a befektetők számára, mely segíthet a jó befektetési stratégia kialakításában, illetve a megfelelő belépési pont megtalálásában [7]. A Hurst exponens segítségével egyfajta mérőszámot kapunk a memóriáról: 1 jelöli, hogy teljesen emlékszik a múltra, $0,5$ a véletlen bolyongást, azaz csak arra emlékszik, hogy most hol van. 0 pedig, hogy még a mostani pozíciójára sem emlékszik.

Az idősorok statisztikai elemzésének egyik célja, hogy a múltbéli adatokból

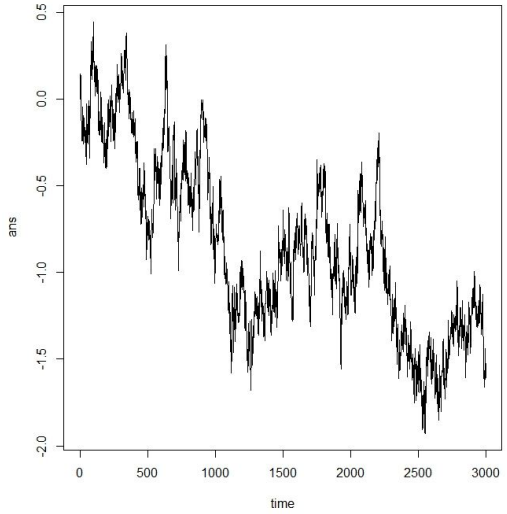
tudjunk következtetni az adatok jövőbeli alakulására. Így valamilyen mintázat beazonosításából, és e mintázatnak a jövőre történő kiterjesztéséből áll. Vagyis az a feltételezésünk, hogy a múltban felismert trend a jövőben is folytatódni fog.

A Hurst exponens becslésére az első ismert módszer az R/S-analízis volt. Az R/S-analízis különböző időperiódusokra kiszámolja a kumulált adatok átlag körüli ingadozásainak R terjedelmét, majd ezt az adatok S szórásával elosztva standardizálja, ezért is nevezik újraskálázott terjedelem-analízisnek (angolul: Rescaled Range Analysis, röviden R/S-analízis). Feltevésünk szerint $\frac{R}{S}(l) \sim l^H$ [5], ahol l jelöli az egyes időperiódusok hosszát.

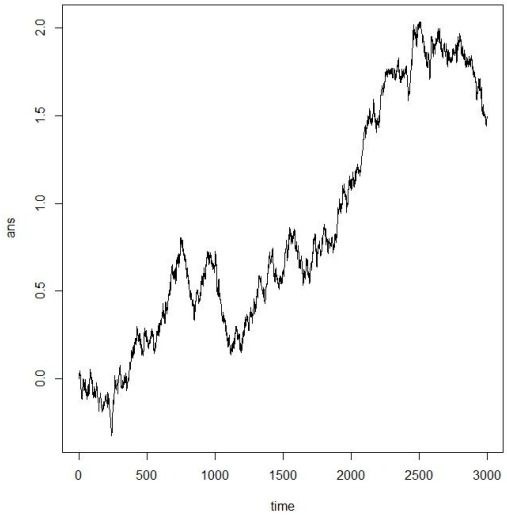
2.3. Interpretáció

A Hurst exponens egy numerikus becslést ad az idősor megjósolhatóságára. Az idősorok statisztikai elemzésének egyik célja, hogy a múltbeli adatokból következtetni tudjunk ezen adatok jövőbeli változására. A módszer alapja, hogy feltételezzük, a múltban tapasztalt trend a jövőben is folytatódni fog. Jelöljük a vizsgált paramétert H-val, melynek értéke 0 és 1 közé esik. Az értékek különböző jelentésekkel bírnak. Amennyiben $0 < H < 0.5$, akkor a görbe viselkedését antiperzisztens viselkedésnek nevezzük, másnéven negatív autokorrelációnak. Ami számunkra azt jelenti, hogy ha t_{i-1} és t_i időpillanatok között a görbén növekedés történt, akkor annak nagyobb a valószínűsége, hogy a következő időpillanatban t_i és t_{i+1} között csökkenés lesz, és fordítva, ha a múltban csökkent, akkor a jövőben növekedni fog. A $H = 0.5$ véletlen bolyongást jelöl. Ekkor nincs korreláció az idősor jelenlegi és jövőbeli értékei között, azaz nulla az autokorreláció. Amennyiben a paraméter értéke 0.5 és 1 közé esik, a görbe úgynevezett perzisztens viselkedést követ, másnéven ezt pozitív autokorrelációnak nevezzük. Ennek jelentése ellentétes a 0 és 0.5 közötti viselkedéssel, vagyis a görbe növekedését nagy valószínűséggel növekedés fogja követni, és csökkenés esetén hasonló viselkedés figyelhető meg [4] [5]. A Hurst exponens kiszámítására több módszer is ismert, melyek a következő alfejezetekben olvashatók időrendi sorrendben. Előtte lássunk példát három különböző Hurst exponenssel rendelkező adatsorra, melyeket az R program egy beépített függvénye segítségével generáltunk, megadva az elemszámot (3000) és a H értéket $(0.3, 0.5, 0.7)$.

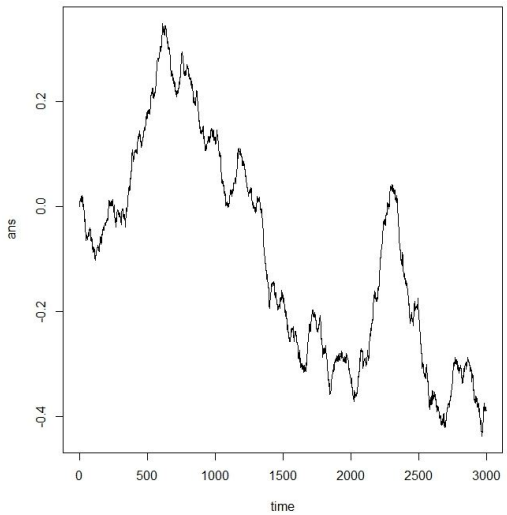
2.1. ábra. $H = 0.3$ - negatív autokorrelált folyamat



2.2. ábra. $H = 0.5$ - véletlen bolyongás



2.3. ábra. $H = 0.7$ - pozitív autokorrelált folyamat



2.3.1. R/S-analízis algoritmusa

Az R/S módszer különböző időperiódusokra kiszámolja a kumulált adatok átlag körüli ingadozásainak R terjedelmét, majd ezt az adatok S szórásával elosztva standardizálja, ezért is nevezik újraskálázott terjedelemnek (Rescaled Range, RR). Feltehető, hogy a feldolgozandó idősorunk n darab adatot tartalmaz. Ezt az adatsort szeretnénk felosztani különböző hosszúságú szakaszokra. Jelölje l a szakaszok hosszát és s a szakaszok számát. Egy lépésben a szakaszok hossza legyen azonos és legalább 10-20 adatot tartalmazzon egy szakasz, azaz a hosszuk legyen minimum 10-20. A j-edik szakaszon a következőképp számoljuk ki RR-t. Jelöljük E_i-vel az ezen a szakaszon mért ellenállások értékeit $(i = 1, 2, ..., l)$.

1. lépés: Szükségünk van a rezisztencia értékek átlagára: $m = \frac{E_1+E_2+...+E_l}{l}$.
2. lépés: Számoljuk ki az értékek eltérését a kapott átlagtól: $D_i = E_i - m$.
3. lépés: Képezzük a következő összegeket: $(t = 1, 2, ..., l)$

$$P_t = \sum_{i=1}^{t} D_i$$

4. lépés: Ezen P_t értékek maximumának és minimumának különbsége adja idősor R-rel jelölt terjedelmét: $R(l) = max(P_1, ..., P_l) - min(P_1, ..., P_l)$.
5. lépés: A standardizáláshoz határozzuk meg az S-sel jelölt E_i értékek szórását, más néven a standard deviációját:

$$S(l) = \sqrt{\frac{1}{l} \cdot \sum_{i=1}^{l} (E_i - m)^2}$$

6. lépés: Az úgynevezett Rescaled Range-t pedig úgy kapjuk meg, hogy az R-et elosztjuk a standard deviációval: $\frac{R}{S} = \frac{R(l)}{S(l)}$.
7. lépés: Számoljuk ki az összes szakaszra az R/S értéket és vegyük átlagukat.

Mivel a következő összefüggés teljesül: $(R/S)_l \approx c \cdot l^H$, ezért a H-val jelölt Hurst exponens lineáris regresszióval megkapható: $\log (R/S)_l = \log c + H \cdot \log l \Rightarrow H = \frac{\log (R/S)_l - \log c}{\log l}$ [4], [6].

A [6] és munkatársai az R/S analízist tesztelték több Brown mozgásból származó adatsorra, és azt figyelték meg, hogy az l érték, azaz a szakaszok hosszának a megválasztása nagy mértékben hatással van a kapott átlagok, illetve standard deviációk értékeire. Amennyiben l elég nagy, az átlag közelebb kerül a valós

$0,5$ értékhez. Kis l értékek esetén az átlag nő. Tehát ahhoz, hogy megfelelően használjuk az R/S analízist, nagy l értéket kell választanunk, ez azonban nem lehetséges rövid adatsorok esetén. Így az R/S analízis használatakor problémába ütközhetünk, ha nem elég hosszú az adatsorunk. Továbbá még egy probléma, hogy kvantált adatokkal dolgozunk, ezért kis l értékre előfordulhat, hogy a standard deviáció nulla, illetve a standard deviáció miatt a módszer érzékeny a kerekítésre.

2.3.2. Anis és Lloyd módszere

A Nílus gátszabályozásakor működő R/S módszer sajnos a gyakorlatban, más idősorok esetén nem adott mindig pontos, értelmezhető értékeket. Ezért más módszerek is születtek H kiszámítására. Például Anis és Lloyd egy új módszerrel állt elő.
Először a következő formulával kiszámítjuk az $\mathbb{E}(R/S)_l$ értéket:

$$\mathbb{E}(R/S)_l = \begin{cases} \frac{l-\frac{1}{2}}{l} \cdot \frac{\Gamma(\frac{l-1}{2})}{\sqrt{\pi}\cdot\Gamma(\frac{l}{2})} \cdot \sum_{i=1}^{l-1} \sqrt{\frac{l-i}{i}} & ha\ l \leq 340 \\ \frac{l-\frac{1}{2}}{l} \cdot \frac{1}{\sqrt{l\cdot\frac{\pi}{2}}} \cdot \sum_{i=1}^{l-1} \sqrt{\frac{l-i}{i}} & ha\ l \geq 340 \end{cases}$$

Majd a H-val jelölt Hurst exponenst úgy kapjuk, hogy $0,5$-höz hozzáadjuk az $(R/S)_l - \mathbb{E}(R/S)_l$ görbe meredekségét. Azonban ekkor a [6]-ban (és általunk is) kapott H érték negatív lehet, aminek nincs értelme. Ezt a problémát elkerülve $\mathbb{E}(R/S)_l$ meghatározása után a következő számítást javasolják [6]-ban:

$$\log H_l = \log (R/S)_l - \log (\mathbb{E}(R/S)_l) + \log (l)/2$$

Végül lineáris regresszióval meghatározzuk a Hurst exponenst:

$$\log H_l = \log c + H \cdot \log l$$

Ez a módszer abból indul ki, hogy normális eloszlású a zaj. Amennyiben ez nem teljesül (nálunk ez a kvantálás miatt vitatható), akkor nem garantált, hogy helyes H értéket kapunk.

2.3.3. GM1 módszer

A Hurst exponens becslésére a geometriában a következő formula ([6]) is ismert:

$$\overline{\Delta B} \quad \alpha \quad l^H$$

ahol $\overline{\Delta B} = \overline{|B(t+l) - B(t)|}$, l jelöli a szakaszok hosszát, H a Hurst exponenst, és α egy arányszám. Először felosztjuk az n hosszú adatsort s szakaszra, melyek hossza l és minden $m = 1, ..., s$ -re végezzük el a következő számításokat:

1. $S_m = X_{ml} - X_{(m-1)l+1}$
2. $H_l = mean\{S_m : m = 1, ..., s\}$.

Ekkor lineáris regresszióval a Hurst exponens kiszámítható:

$$\log H_l = \log c + H \log l$$

2.3.4. GM2 módszer

Az előző módszer már helyes H értéket ad eredményül, azonban a GM1 módosításával kapott GM2 módszer pontosabb megoldást adhat:

$$\overline{range(B)} \quad \alpha \quad l^H$$

ahol $\overline{range(B)} = \overline{max\{B(S) : t \leq S \leq t+l\}} - \overline{min\{B(S) : t \leq S \leq t+l\}}$.

Több véletlen bolyongásból származó adatsorra tesztelve a GM1 módszert és ennek módosítását, a GM2 módszert, [6] alapján azt állíthatjuk, hogy mindkettő megfelelően működik, hiszen Brown mozgásra mindkettő $0,5$-höz közeli Hurst exponenst ad eredményül. A [6] szerint a standard deviáció kisebb volt a GM2 módszernél, mint a GM1-nél. Ezért, amennyiben lehetőségünk van a GM2 módszert alkalmazni, ez a módszer pontosabb értéket ad.

3. fejezet

Célkitűzés

A Témalabor 2 munka során az alábbi célokat tűztük ki:

1. A Hurst exponens számítására alkalmas öt módszer beprogramozása, illetve összehasonlítása.
2. Minta adatsorok generálása különböző H értékekre és ezek tesztelése az öt módszerrel.
3. Biológiai mintából származó adatok vizsgálata módszerekkel.
4. Statisztikai elemzés.

4. fejezet

Módszer

4.1. Adatsorok előállítása

Először $H = 0.3, 0.5, 0.7$ értékekre generáltunk adatsorokat az R programban beépített függvény segítségével. Három esetet úgy, mint perzisztens, antiperzisztens és véletlen bolyongást vizsgáltuk (lásd 2.1, 2.2, 2.3 ábra). Mindhárom esetben a kapott öt különböző H értéket egy EXCEL táblában (7.3) gyűjtöttük össze, melyről leolvasható, hogy a GM2 módszer adja a legpontosabb értéket.

Majd a Semmelweis Egyetem Genetikai, Sejt- és Immunbiológiai Intézet kutató csoportja által, az ECIS készülék segítségével generált adatokat teszteltük (7.4, 7.5). Ezen adatok egy része sejtmentes mintából, míg másik része sejtes mintából származott. A statisztikai elemzés során legfőbb célunk, hogy a két minta a Hurs exponens ismeretében szignifikánsan megkülönböztethető legyen.

4.2. Hurst rutin

A korábbi fejezetekben bemutatott, a Hurst exponens kiszámítására alkalmas öt különböző módszert MATLAB-ban programoztuk be.

A rutin bemenetként kap egy $\underline{X}$ vektort, mely tartalmazza a mért rezisztencia adatokat, majd eredményül megadja az öt különböző módszerrel meghatározott H értékeket $(Hcl, Hal1, Hal2, Hgeo1, Hgeo2)$. Ezen adatokat EXCEL táblában foglaltuk össze (7.3, 7.4, 7.5).

```
rangek=200:floor(length(X)/5);
[rs,ers]=RSana(X,rangek,'Hurst');
Filter=~isnan(rs);
lrs=log10(rs(Filter));
lrangek=log10(rangek(Filter));
Hcl=polyfit(lrangek,lrs,1)
figure(1);
plot(lrangek,lrs,lrangek,lrangek*Hcl(1));
lH=lrs-log10(ers(Filter))+lrangek/2;
Hal2=polyfit(lrangek,lH,1)
figure(2);
plot(lrangek,lH,lrangek,lrangek*Hal2(1));
Hal1=polyfit(rangek(Filter),rs(Filter)-ers(Filter),1)
figure(3);
plot(lrangek,lH,lrangek,lrangek*Hal1(1));
rs=RSana2(X,rangek,'Geo1');
lH=log10(rs);
lrangek=log10(rangek);
Hgeo1=polyfit(lrangek,lH,1)
figure(4);
plot(lrangek,lH,lrangek,lrangek*Hgeo1(1));
rs=RSana2(X,rangek,'Geo2');
lH=log10(rs);
lrangek=log10(rangek);
Hgeo2=polyfit(lrangek,lH,1)
figure(5);
plot(lrangek,lH,lrangek,lrangek*Hgeo2(1));
```

MATLAB rutin

4.3. Eredmények

A generált minta adatsorok esetén kapott táblázatból (7.3) jól leolvasható, hogy a minták H értékét legjobban az időrendben utolsó GM2 módszerrel számolt értékek közelítik.

A [6] cikkben szintén ez a módszer bizonyult a legpontosabbnak. Habár ezen tanulmány a Hurst exponens gazdaságban betöltött szerepét vizsgálja, számunkra, az egészségügyben is teljesülnek az itt leírt eredmények.

A minta adatsornál kapott eredményeket, illetve a [6] cikket felhasználva a biológiai sejtmentes, illetve sejtes adatok esetén is ezen H értékre végezzük el a következőkben a statisztikai elemzést.

4.4. Statisztikai elemzés

Ahhoz, hogy vizsgáljuk, vajon a Hurst exponens ismeretében egy adatsorról eldönthető-e, hogy az sejtmentes vagy sejtes mintából származó adatsor, statisztikai elemzést végeztünk. A statisztikában szignifikáns eredményt akkor érhetünk el, ha minél több adat áll rendelkezésünkre. Sajnos, az idő rövidsége, illetve az ECIS műszer érzékenysége miatt az elemzést kevesebb adaton hajtottuk végre. Néhány esetben ki kellett zárnunk adatsorokat a vizsgálatból, hiszen ezek lineáris trendet mutattak. Ennek több oka is lehetett. Vagy, mivel az első mérések voltak, ezért a gép még nem melegedett be, vagy egy közeli számítógép megzavarta a mérés folyamatát. Végül a megmaradt adatsorokra először alkalmaztunk egy Shapiro-Wilk tesztet, melynek segítségével a normalitási feltételt vizsgáltuk. Ezután a feltételt teljesítő adatsorokra kétmintás t-próbát alkalmaztunk. Mindkét tesztet a következő fejezetekben részletesen is bemutatjuk.

5. fejezet

Statisztika

A matematikai statisztika segítségével megvizsgáljuk, hogy az előbbiekben meghatározott (GM2 módszerrel számolt) Hurst exponens ismeretében egy adatsorról szignifikánsan eldönthető-e, hogy az sejtmentes vagy sejtes mintából származik-e.

Feladatunk, hogy a H_0-lal jelölt nullhipotézis és a H_1-gyel jelölt alternatív hipotézis (ellenhipotézis) között döntsünk. Esetünkben a hipotézisek:

$$H_0 : \text{Hurst exponens nem megfelelő paraméter}$$
$$\text{versus}$$
$$H_1 : \text{Hurst exponens megfelelő paraméter}$$

ahol azt mondjuk, hogy a Hurst exponens akkor megfelelő, ha a sejtmentes adatok esetén számolt H érték nem azonos a sejtes adatok esetén számolt H értékkel. Ha megegyezik a kettő, akkor a Hurst exponens nem megfelelő.

Ezután csak egy számunkra megfelelő statisztikai próbát kell találnunk, melynek feltételei teljesülnek és így használni is tudjuk. Olyan statisztikára van szükségünk, amely azt vizsgálja, hogy két külön mintában egy-egy valószínűségi változó átlagai egymástól szignifikánsan különböznek-e. Ez nem más mint a kétmintás t-próba. Azonban mielőtt használnánk a tesztet, le kell ellenőriznünk, hogy feltételei teljesülnek-e a vizsgált adatsorokra, azaz, hogy normális eloszlásúak-e, illetve, hogy szórásuk megegyezik-e.

5.1. Kétmintás t-próba

A kétmintás t-próbát [8] kis mintákra szokták alkalmazni, melyek szükségképpen normális eloszlásúak, és szórásuk megegyezik. A próbát normális eloszlás várható értékének tesztelésére vagy két normális várható érték összehasonlítására hasz-

nálják. A kétmintás jelző arra utal, hogy két tetszőleges várható értékű, de azonos szórású háttérváltozót vizsgálunk. (Két normális eloszlású változó szórásának összehasonlítására az F-próbát használhatjuk.) Legyenek ezek: $X \sim \mathcal{N}(\mu_1, \sigma^2)$ és $Y \sim \mathcal{N}(\mu_2, \sigma^2)$. Ezen belül megkülönböztetünk egyoldali és kétoldali alternatívát. Nekünk most az utóbbira lesz szükségünk. A hipotézis vizsgálat ekkor:

$$H_0 : \mu_1 = \mu_2 \qquad \text{versus} \qquad H_1 : \mu_1 \neq \mu_2$$

Az n_1 elemű $X_1, X_2, \ldots, X_{n_1} \sim \mathcal{N}(\mu_1, \sigma^2)$ független, azonos eloszlású és az n_2 elemű $Y_1, Y_2, \ldots, Y_{n_2} \sim \mathcal{N}(\mu_2, \sigma^2)$ független, azonos eloszlású, egymástól is független mintákból konstruált próbastatisztika:

$$t(\mathbf{X}, \mathbf{Y}) = \frac{\overline{X} - \overline{Y}}{\sqrt{(n_1 - 1)(S_X^*)^2 + (n_2 - 1)(S_Y^*)^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

és az $1 - \epsilon$ szignifikanciaszinthez konstruált kritikus tartomány:

$$\mathcal{X}_k = \{(\mathbf{x}, \mathbf{y}) : |t(\mathbf{x}, \mathbf{y})| \geq t_{\epsilon/2}(n_1 + n_2 - 2)\}$$

ahol $\frac{(n_1-1)(S_X^*)^2 + (n_2-1)(S_Y^*)^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$.

Tehát a nullhipotézisünket $1 - \epsilon$ szinten elfogadjuk, ha a mintarealizációból számolt $|t(\mathbf{x}, \mathbf{y})| < t_{\epsilon/2}(n_1 + n_2 - 2)$, különben pedig elutasítjuk.

5.2. Normalitás teszt

A statisztikában a normalitás tesztet arra használják, hogy eldöntsék egy adatsorról, hogy az normális eloszlású-e. Több erre alkalmas tesztet ismerünk. Razali és Wah tanulmánya [10] szerint a legnagyobb ereje a Shapiro-Wilk tesztnek van.

5.2.1. Shapiro-Wilk teszt

Shapiro-Wilk teszt [9] esetén adott egy n hosszú adatsor, melyről el szeretnénk dönteni, hogy normális eloszlást követ-e.

1. lépés: Rendezzük növekvő sorrendbe az adatokat: $x_1 \leq x_2 \leq \cdots \leq x_n$
2. lépés: Számoljuk ki a következő összeget:

$$SS = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

ahol $\overline{x}$ jelöli az adatok átlagát.
3. lépés: Ha n páros , akkor $m = \frac{n}{2}$. Ha n páratlan, akkor $m = \frac{n-1}{2}$.
4. lépés: Számoljuk ki b-t a következőképp:

$$b = \sum_{i=1}^{m} a_i (x_{n+1-i} - x_i)$$

ahol az a_i értékeket a Shapiro-Wilk 1. táblából (7.1) olvashatjuk le. Megjegyezzük, hogy ha n páratlan, akkor a medián értékét nem használjuk b kiszámításában.
5. lépés: Számoljuk ki a teszt statisztikát:

$$W = b^2/SS$$

6. lépés: A kapott W érték és n ismeretében a Shapiro-Wilk 2. táblából (7.2) leolvasható a teszt p értéke.

Amennyiben a kapott p érték $> 0,05$, akkor a nullhipotézist fogadjuk el, miszerint az adatsor normális eloszlású.

5.3. Eredmények

Az R programban beépített függvények segítségével hajtottuk végre a statisztikai elemzést.

Elsőként a normalitás eldöntésére a sejtmentes és a sejtes adatsorok esetén kapott, GM2 módszerrel számolt Hurst exponensekre alkalmaztuk a Shapiro-Wilk tesztet(shapiro.test(x)). A sejtmentes esetben $0,825$ lett a p értéke, ami lényegesen nagyobb, mint $0,05$, így a nullhipotézist fogadjuk el, miszerint az adatsor normális eloszlást követ. A sejtes esetben pedig a kapott p érték $0,5061$, azaz ebben az esetben is normális eloszlású az adatsor.

Ezután F-próbával (var.test(x, y, alternative = "two.sided")) vizsgáltuk, hogy a két adatsor szórása megegyezik-e. Mivel a kapott p érték ebben az esetben is jóval meghaladta a $0,05$ értéket $(0,6334)$, ezért szórásuk ugyanannyinak tekinthető. A feltételek teljesülése után már csak a kétmintás t-próbát kellett alkalmaznunk.

A tesztet (t.test(x,y, var.equal=TRUE)) elvégezve eredményül $p = 1.63e-11$ értéket kaptunk, azaz a H_1 hipotézist fogadjuk el. Tehát a Hurst exponens 99.999999999%-os szignifikancia szint mellett jó paraméter annak eldöntésére, hogy az adatsor sejtes vagy sejtmentes mintából származik-e.

6. fejezet

Összefoglalás

A félév végére egy fontos megfigyelésre jutottunk, miszerint egy adatsorról eldönthető, hogy az sejtes vagy sejtmentes mintából származik-e. Ehhez szükségünk volt, hogy megismerjük a Hurst exponens tulajdonságait, illetve össze kellett hasonlítanunk az eddig felsorakozó módszereket a Hurst exponens kiszámítására, hogy kiválaszthassuk a legpontosabb értéket adót közülük. A módszerek beprogramozására a MATLAB volt segítségünkre, míg minta adatsorok generálására az R program. Mivel az irodalomban és a generált adatsoroknál is azt figyelhettük meg, hogy az időrendben utolsó, GM2 módszer bizonyult a legpontosabbnak, ezért a Semmelweis Egyetem laboratóriumában, az ECIS készülék segítségével előállított biológiai adatok H értékének kiszámítására is ezt a módszert használtuk. Azonban ahhoz, hogy eldöntsük, a vizsgált paraméter ismeretében a sejtes és sejtmentes minták megkülönböztethetőek-e, statisztikai elemzést végeztünk. Az elemzést is az R programban hajtottuk végre. A normalitási feltétel és a szórások egyezésének feltétele után kétoldali t-próbát használtunk az adatsorok tesztelésére. Mivel eredményül $0,05$-nél jóval kisebb p értéket kaptunk, ezért megállapítottuk, hogy a GM2 módszerrel számolt H érték ismeretében a sejtmentes és a sejtes minták szignifikánsan elkülöníthetők egymástól.

A következő féléves tervek között szerepel, hogy ne csak a sejtes és sejtmentes mintákat tudjuk megkülönböztetni, hanem egy sejtről eldönthető legyen, hogy az tumoros vagy egészséges-e.

7. fejezet

Függelék

7.1. ábra. Shapiro-Wilk 1. tábla - W értékek

n =	2	3	4	5	6	7	8	9	10	11	12	13	14
a1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	0.5601	0.5475	0.5359	0.5251
a2			0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	0.3315	0.3325	0.3325	0.3318
a3					0.0875	0.1401	0.1743	0.1976	0.2141	0.2260	0.2347	0.2412	0.2460
a4							0.0561	0.0947	0.1224	0.1429	0.1586	0.1707	0.1802
a5									0.0399	0.0695	0.0922	0.1099	0.1240
a6											0.0303	0.0539	0.0727
a7													0.0240

n =	15	16	17	18	19	20	21	22	23	24	25	26
a1	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407
a2	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043
a3	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533
a4	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151
a5	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836
a6	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563
a7	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316
a8		0.0196	0.0359	0.0496	0.0612	0.0711	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089
a9				0.0163	0.0303	0.0422	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876
a10						0.0140	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672
a11								0.0122	0.0228	0.0321	0.0403	0.0476
a12									0.0000	0.0107	0.0200	0.0284
a13											0.0000	0.0094

n =	15	16	17	18	19	20	21	22	23	24	25	26
a1	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407
a2	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043
a3	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533
a4	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151
a5	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836
a6	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563
a7	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316
a8		0.0196	0.0359	0.0496	0.0612	0.0711	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089
a9				0.0163	0.0303	0.0422	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876
a10						0.0140	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672
a11								0.0122	0.0228	0.0321	0.0403	0.0476
a12									0.0000	0.0107	0.0200	0.0284
a13											0.0000	0.0094

n =	27	28	29	30	31	32	33	34	35	36	37	38
a 1	0.4366	0.4328	0.4291	0.4254	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015
a 2	0.3018	0.2992	0.2968	0.2944	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774
a 3	0.2522	0.2510	0.2499	0.2487	0.2475	0.2463	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391
a 4	0.2152	0.2151	0.2150	0.2148	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110
a 5	0.1848	0.1857	0.1864	0.1870	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881
a 6	0.1584	0.1601	0.1616	0.1630	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686
a 7	0.1346	0.1372	0.1395	0.1415	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513
a 8	0.1128	0.1162	0.1192	0.1219	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356
a 9	0.0923	0.0965	0.1002	0.1036	0.1066	0.1093	0.1118	0.1140	0.1160	0.1179	0.1196	0.1211
a 10	0.0728	0.0778	0.0822	0.0862	0.0899	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075
a 11	0.0540	0.0598	0.0650	0.0697	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947
a 12	0.0358	0.0424	0.0483	0.0537	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824
a 13	0.0178	0.0253	0.0320	0.0381	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706
a 14	0.0000	0.0084	0.0159	0.0227	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592
a 15			0.0000	0.0076	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481
a 16					0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372
a 17							0.0000	0.0062	0.0119	0.0172	0.0220	0.0264
a 18									0.0000	0.0057	0.0110	0.0158
a 19											0.0000	0.0053

n =	39	40	41	42	43	44	45	46	47	48	49	50
a1	0.3989	0.3964	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
a2	0.2755	0.2737	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
a3	0.2380	0.2368	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
a4	0.2104	0.2098	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
a5	0.1880	0.1878	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
a6	0.1689	0.1691	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
a7	0.1520	0.1526	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
a8	0.1366	0.1376	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
a9	0.1225	0.1237	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
a10	0.1092	0.1108	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
a11	0.0967	0.0986	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
a12	0.0848	0.0870	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
a13	0.0733	0.0759	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.9190	0.0932
a14	0.0622	0.0651	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
a15	0.0515	0.0546	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
a16	0.0409	0.0444	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
a17	0.0305	0.0343	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
a18	0.0203	0.0244	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
a19	0.0101	0.0146	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
a20	0.0000	0.0049	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
a21			0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
a22					0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
a23							0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
a24									0.0000	0.0037	0.0071	0.0104
a25											0.0000	0.0035

7.2. ábra. Shapiro-Wilk 2. tábla - p értékek

n \ P	0.01	0.02	0.05	0.1	0.5	0.9	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979	0.986	0.991	0.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987
18	0.858	0.874	0.897	0.914	0.956	0.978	0.982	0.986	0.988
19	0.863	0.879	0.901	0.917	0.957	0.978	0.982	0.986	0.988
20	0.868	0.884	0.905	0.920	0.959	0.979	0.983	0.986	0.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	0.878	0.892	0.911	0.926	0.961	0.980	0.984	0.987	0.989
23	0.881	0.895	0.914	0.928	0.962	0.981	0.984	0.987	0.989
24	0.884	0.898	0.916	0.930	0.963	0.981	0.984	0.987	0.989
25	0.888	0.901	0.918	0.931	0.964	0.981	0.985	0.988	0.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	0.894	0.906	0.923	0.935	0.965	0.982	0.985	0.988	0.990
28	0.896	0.908	0.924	0.936	0.966	0.982	0.985	0.988	0.990
29	0.898	0.910	0.926	0.937	0.966	0.982	0.985	0.988	0.990
30	0.900	0.912	0.927	0.939	0.967	0.983	0.985	0.988	0.990
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	0.904	0.915	0.930	0.941	0.968	0.983	0.986	0.988	0.990
33	0.906	0.917	0.931	0.942	0.968	0.983	0.986	0.989	0.990
34	0.908	0.919	0.933	0.943	0.969	0.983	0.986	0.989	0.990
35	0.910	0.920	0.934	0.944	0.969	0.984	0.986	0.989	0.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	0.914	0.924	0.936	0.946	0.970	0.984	0.987	0.989	0.990
38	0.916	0.925	0.938	0.947	0.971	0.984	0.987	0.989	0.990
39	0.917	0.927	0.939	0.948	0.971	0.984	0.987	0.989	0.991
40	0.919	0.928	0.940	0.949	0.972	0.985	0.987	0.989	0.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	0.922	0.930	0.942	0.951	0.972	0.985	0.987	0.989	0.991
43	0.923	0.932	0.943	0.951	0.973	0.985	0.987	0.990	0.991
44	0.924	0.933	0.944	0.952	0.973	0.985	0.987	0.990	0.991
45	0.926	0.934	0.945	0.953	0.973	0.985	0.988	0.990	0.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	0.928	0.936	0.946	0.954	0.974	0.985	0.988	0.990	0.991
48	0.929	0.937	0.947	0.954	0.974	0.985	0.988	0.990	0.991
49	0.929	0.939	0.947	0.955	0.974	0.985	0.988	0.990	0.991
50	0.930	0.938	0.947	0.955	0.974	0.985	0.988	0.990	0.991

7.1. táblázat. Összehasonlító táblázat (R/S analízis, Anis és Lloyd 1., Anis és Lloyd 2.)

	R/S analízis	**Anis és Lloyd 1.**	**Anis és Lloyd 2.**
Kiszámítása	l szakaszok hossza, s szakaszok száma, E_i mért rezisztencia értékek, $m = \frac{E_1+E_2+...+E_l}{l}$, $D_i = E_i - m$, $P_t = \sum_{i=1}^{t} D_i$ $(t = 1, 2, ..., l)$, $R(l) = max(P_1, ..., P_l)$ $-min(P_1, ..., P_l)$, $S(l) = \sqrt{\frac{1}{l} \cdot \sum_{i=1}^{l}(E_i - m)^2}$, $H = \frac{\log (R/S)_l - \log c}{\log l}$	l szakaszok hossza, s szakaszok száma, E_i mért rezisztencia értékek, $m = \frac{E_1+E_2+...+E_l}{l}$, $D_i = E_i - m$, $P_t = \sum_{i=1}^{t} D_i$ $(t = 1, 2, ..., l)$, $R(l) = max(P_1, ..., P_l)$ $-min(P_1, ..., P_l)$, $S(l) = \sqrt{\frac{1}{l} \cdot \sum_{i=1}^{l}(E_i - m)^2}$, $\mathbb{E}(R/S)_l = \begin{cases} \frac{l-\frac{1}{2}}{l} \cdot \frac{\Gamma(\frac{l-1}{2})}{\sqrt{\pi}\cdot\Gamma(\frac{l}{2})} \cdot \sum_{i=1}^{l-1} \sqrt{\frac{l-i}{i}} & ha\ l \leq 340 \\ \frac{l-\frac{1}{2}}{l} \cdot \frac{1}{\sqrt{l \cdot \frac{\pi}{2}}} \cdot \sum_{i=1}^{l-1} \sqrt{\frac{l-i}{i}} & ha\ l \geq 340 \end{cases}$ $H = 0,5 + ((R/S)_l - \mathbb{E}(R/S)_l)$ görbe meredeksége	l szakaszok hossza, s szakaszok száma, E_i mért rezisztencia értékek, $m = \frac{E_1+E_2+...+E_l}{l}$, $D_i = E_i - m$, $P_t = \sum_{i=1}^{t} D_i$ $(t = 1, 2, ..., l)$, $R(l) = max(P_1, ..., P_l)$ $-min(P_1, ..., P_l)$, $S(l) = \sqrt{\frac{1}{l} \cdot \sum_{i=1}^{l}(E_i - m)^2}$, $\mathbb{E}(R/S)_l = \begin{cases} \frac{l-\frac{1}{2}}{l} \cdot \frac{\Gamma(\frac{l-1}{2})}{\sqrt{\pi}\cdot\Gamma(\frac{l}{2})} \cdot \sum_{i=1}^{l-1} \sqrt{\frac{l-i}{i}} & ha\ l \leq 340 \\ \frac{l-\frac{1}{2}}{l} \cdot \frac{1}{\sqrt{l \cdot \frac{\pi}{2}}} \cdot \sum_{i=1}^{l-1} \sqrt{\frac{l-i}{i}} & ha\ l \geq 340 \end{cases}$ $\log H_l = \log (R/S)_l - \log (\mathbb{E}(R/S)_l) + \log (l)/2$, $H = \frac{\log H_l - \log c}{\log l}$
Hátránya	1. Brown mozgás esetén kicsi l érték (szakaszhossz) esetén rossz H értéket ad. 2. Kis l értékre előfordulhat, hogy a standard deviáció nulla lesz. 3. Érzékeny a kerekítésre a standard deviáció miatt.	H értéke negatív lehet.	Ha a zaj nem normális eloszlású (nálunk nem normális eloszlású), akkor nem garantált, hogy helyes H értéket kapunk.

7.2. táblázat. Összehasonlító táblázat (GM1 és GM2 módszer)

	GM1 módszer	**GM2 módszer**
Kiszámítása	l szakaszok hossza, s szakaszok száma, $\overline{\Delta B}\ \alpha\ l^H$, $\overline{\Delta B} = \overline{\lvert B(t+l) - B(t)\rvert}$, α arányszám, $\forall\ m = 1, \cdots, s$-re $S_m = x_{ml} - x_{(m-1)l+1}$ $H_l = mean\{S_m : m = 1. \cdots, s\}$ $\log H_l = \log c + H \log l$ $H = \frac{\log H_l - \log c}{\log l}$	l szakaszok hossza, s szakaszok száma, $\overline{range(B)}\ \alpha\ l^H$, $\overline{range(B)} =$ $\overline{max\{B(S) : t \leq S \leq t+l\}}$ $-\overline{min\{B(S) : t \leq S \leq t+l\}}$, α arányszám
Hátránya	Helyes H értéket ad, de nem a legpontosabbat.	Ennél a módszernél a standard deviáció kisebb, mint a GM1 módszernél, ezért pontosabb eredményt ad.

7.3. táblázat. Minta adatsorok

	Szórás (4)	Szórás (16)	Szórás (64)	Szórás (256)	Hcl	Hal1	Hal2	Hgeo1	Hgeo2
0.3.csv	0.5043	0.4998	0.4924	0.4764	0.9738	-0.4643	0.4734	0.3372	0.3759
0.5.csv	0.67	0.6709	0.6735	0.6729	1.0247	-0.3891	0.5244	0.5099	0.5827
0.7.csv	0.2058	0.2059	0.2039	0.1991	0.9738	-0.3879	0.4736	0.5776	0.6987
átlag					0,9908	-0,4138	0,4905	0,4750	0,5525
szórás					0,0294	0,0438	0,0294	0,1239	0,1636

7.4. táblázat. Sejtmentes adatsorok

	Szórás (4)	Szórás (16)	Szórás (64)	Szórás (256)	Hcl	Hal1	Hal2	Hgeo1	Hgeo2
1.	0.4208	0.4123	0.4093	0.4104	1.1591	-0.5538	0.6583	0.4541	0.3126
2.	0.4768	0.4682	0.4682	0.4731	1.1720	-0.5533	0.6713	0.3991	0.3155
3.	0.6303	0.6270	0.6294	0.6403	1.2766	-0.3844	0.7758	0.5472	0.4577
4.	0.3726	0.3662	0.3660	0.3728	1.1918	-0.6119	0.6909	0.6867	0.2331
5.	0.8841	0.8833	0.8868	0.9002	1.262	-0.4491	0.7611	0.8168	0.4953
6.	0.7934	0.7897	0.7939	0.8112	1.2461	-0.4788	0.7452	0.7697	0.4651
7.	0.4028	0.3952	0.3945	0.3917	1.1566	-0.5138	0.6556	-0.2051	0.2497
8.	0.4124	0.4062	0.4057	0.4113	1.1353	-0.5343	0.6344	0.6230	0.3563
9.	0.2661	0.2599	0.2592	0.2612	1.2939	-0.5191	0.7931	0.4733	0.3967
10.	0.4353	0.4312	0.4309	0.4348	1.1549	-0.5543	0.6541	0.4642	0.3260
11.	0.3917	0.3829	0.3806	0.3821	1.1951	-0.5326	0.6943	0.5467	0.3866
12.	0.4051	0.4027	0.4033	0.4096	1.1711	-0.5117	0.6703	0.2398	0.3831
átlag					1,2012	-0,5165	0,7004	0,4847	0,3649
szórás					0,0540	0,0582	0,0540	0,2704	0,0825

7.5. táblázat. Sejtes adatsorok

	Szórás (4)	Szórás (16)	Szórás (64)	Szórás (256)	Hcl	Hal1	Hal2	Hgeo1	Hgeo2
1.	18.89891	18.9181	18.9670	18.6953	1.0063	-0.3926	0.5054	0.5402	0.7228
2.	120.5754	120.4479	121.166	123.8944	0.97	-0.4043	0.4692	0.8139	0.7806
3.	88.8376	89.0370	89.4301	90.9501	1.0411	-0.3775	0.5403	0.7304	0.6981
4.	73.5739	73.7369	73.7159	72.9556	0.9750	-0.4123	0.4743	0.3080	0.6021
5.	74.1366	74.2947	74.3439	71.6747	0.9556	-0.4194	0.4547	0.4410	0.6307
6.	56.6550	56.6131	56.3489	54.6691	0.9839	-0.3895	0.4829	0.3235	0.5281
7.	98.5827	98.7356	99.2151	100.8682	0.9497	-0.4131	0.4490	0.7883	0.7248
8.	79.4716	79.6779	79.9444	80.1905	0.9728	-0.4159	0.4721	0.5079	0.6236
9.	64.2205	64.1548	63.9898	62.7918	0.9911	-0.4027	0.4904	0.4108	0.5344
10.	24.2608	24.221	23.8379	22.1294	0.9903	-0.4078	0.4895	0.1736	0.5474
11.	121.1915	121.1657	121.6437	122.5998	0.9404	-0.4349	0.4395	0.5592	0.6366
12.	73.8623	73.9448	73.7741	72.5010	0.9771	-0.4123	0.4763	0.6574	0.6753
13.	176.0801	176.3190	177.2323	180.2062	1.0129	-0.375	0.5121	0.5225	0.7153
14.	40.4834	40.4562	40.6963	41.6473	0.9572	-0.4161	0.4563	0.8025	0.7306
15.	56.2034	55.9921	56.3023	57.4181	0.9802	-0.4139	0.4794	0.8938	0.7524
16.	43.1443	43.2618	43.4288	44.1041	0.9491	-0.4223	0.4482	0.7230	0.6329
17.	24.913	24.5753	24.6404	24.7331	0.9055	-0.4517	0.4047	0.6109	0.5858
18.	28.6842	28.6431	28.7009	28.5991	1.1069	-0.3371	0.6060	0.5655	0.7147
19.	34.3485	34.3868	34.4988	34.9844	0.9979	-0.3992	0.4970	0.5193	0.6491
20.	31.8283	31.9116	31.9697	31.8096	0.9980	-0.3869	0.4971	0.7699	0.6777
átlag					0,9831	-0,4043	0,4823	0,5831	0,6582
szórás					0,0414	0,0242	0,0414	0,1918	0,0735

Irodalomjegyzék

[1] *Estimating the Hurst Exponent*, http://www.bearcave.com/misl/misltech/wavelets/hurst/index.html

[2] *Herpai Nándor Nílusról szóló blogja*, https://herpainandor.com/2013/06/22/nilus-es-a-tozsde/

[3] *Fraktálok, a fraktáldimenzió*, http://www.geo.u-szeged.hu/ joe/pub/Tamop/Jegyzet/ch09s03.html

[4] Dávid Gergely, *Tumorsejtek elektromos jelének vizsgálata ECIS készülékben*, Szakdolgozat (2011.)

[5] Gombos Kitti Kata, *Fraktálok a tőzsdén*, Szakdolgozat, Szegedi Tudományegyetem (2010.)

[6] M. A. Sánchez Granero, J. E. Trinidad Segovia, J. García Pérez, *Some comments on Hurst exponent and the long memory processes on capital markets*

[7] Somogyi Balázs István *Mérsékelt kockázatú befektetési alapok vagyonának és árfolyamának stabilitása és hosszú emlékezete* OTDK, Debreceni Egyetem Gazdálkodástudományi és Vidékfejlesztési Kar

[8] Bolla Marianna, Krámli András *Statisztikai következtetések elmélete*

[9] Shapiro, Samuel Sanford, and Martin B. Wilk. *An analysis of variance test for normality (complete samples).* Biometrika 52.3/4 (1965): 591-611.

[10] Razali, Nornadiah Mohd, and Yap Bee Wah. *Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests.* Journal of statistical modeling and analytics 2.1 (2011): 21-33.