



Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

Önhasonló Hálózati Forgalom Matematikai Leírása

Geffertth András

Tudományos vezetők
Dr. Molnár Sándor és Dr. Darryl N. Veitch

Budapest, 2005

1. Bevezetés

Távközlőhálózatok, számítógép-hálózatok, távközlési protokollok, illetve alkalmazások tervezéséhez fontos a hálózati forgalom viselkedésének ismerete. Ezt a forgalmat sztochasztikus leírókkal jellemezhetjük. A klasszikus telefonhálózat forgalmát jellemző sztochasztikus modellek a huszadik század elején születtek és mára jól bevált módszerek léteznek az ilyen hálózatok méretezésére, tervezésére. A számítógép-hálózatok forgalma azonban jelentősen különbözik a telefonhálózatokétól. Itt ugyanis a forgalomban igen erős korreláció van jelen, ami sokszor egy úgynevezett hosszútávú összefüggő és önhasonló folyamathoz vezet.

Hosszútávú összefüggő folyamatok különféle természetes, illetve mesterséges rendszerekben fordulnak elő. Első ízben Hurst fedezett fel ilyen folyamatokat 1950 táján, mikor a Nílus több száz évre visszamenőleges vízállását vizsgálta. Azóta számos más tudományterületen, pl. mezőgazdaságban, fizikában, illetve a távközlésben is fedeztek fel ilyen jellegű folyamatokat [1].

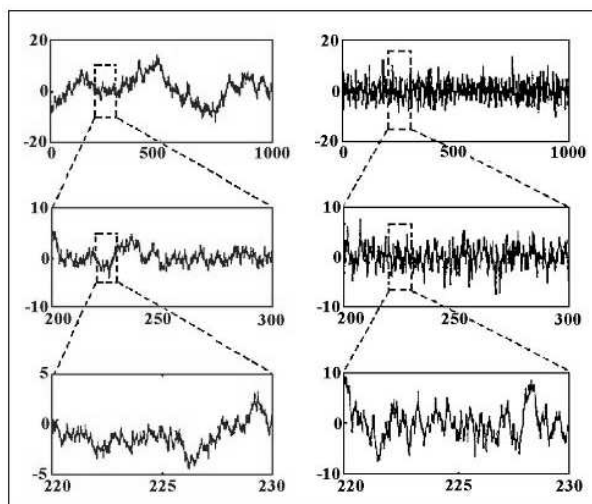
A hosszútávú összefüggőségnek többféle definíciója létezik. Ezek közül néhányat majd a 4.3.2. részben, egy bővebb listát pedig a disszertációban ismertetek. Itt ezeknek a folyamatoknak néhány olyan – minden definíció esetén érvényes – tulajdonságát sorolom fel, melyek rájuk a legjellemzőbbek és leginkább megkülönböztetik őket a megszokott, ún. rövid távú összefüggő folyamatoktól.

- A hosszútávú összefüggőséget általában diszkrét, stacionárius folyamatok esetében értelmezzük.
- A folyamat autokovariancia függvénye ($\gamma(k)$) lassan cseng le, így ennek végtelen összege, $\sum_{i=0}^{\infty} \gamma(k)$ végtelen.
- Az X folyamat mintaátlag, melyet a $\bar{X} := \frac{\sum_{i=1}^m X(i)}{m}$ képlettel számolunk, szórásnégyzete lassabban cseng le mint $O(m^{-1})$, ahol m a mintaméret.
- Ahogy a folyamatot egyre nagyobb időskálán tekintjük, az így kapott folyamatok autokorrelációs függvényeinek sorozata egy megadott függvényhez konvergál így a határérték folyamat sem lesz korrelálatlan. Ezt a tulajdonságot aszimptotikus önhasonlóságnak nevezzük. Egy ilyen folyamatot ábrázol az 1. ábra, melyen látható, hogy az intenzitás-görbe az egyes időskálákon vizuálisan hasonló.

A távbeszélő hálózatok esetén a kihasználtság növelése egyben az eredő forgalom kisimulásához vezet, vagyis a forgalom szórásnégyzete az átlaghoz képest csökken. Ilyen jelenség számítógép-hálózatok esetén nem figyelhető meg, vagyis a forgalom nagy kihasználtság esetén is erősen változékonny marad. Az egyszerre érkező nagy mennyiségű adat kezeléséhez nagyméretű tárolókra van szükség, melyek jelentős késleltetést okoznak. Habár a távbeszélő- illetve adathálózatok forgalmi méretezése között egyéb különbségek is vannak ez a példa szemlélteti, hogy adathálózatok tervezéséhez újfajta megközelítésre van szükségünk.

A hosszútávú függés hálózatra gyakorolt hatásai mellett annak okát is vizsgálták. A lehetséges okok között az átvitt fájlok méretének lassan lecsengő eloszlása, illetve a TCP¹ protokoll

¹Transmission Control Protocol



1. ábra. Azonos folyamatok intenzitásgörbéi különböző időskálákon. Amennyiben az időskála kicsi (alsó sor) mindkét folyamat szórása jelentős. Az időskála növelésével a jobb oldali folyamat görbéje kisimul, míg a baloldalié továbbra is erős változásokat mutat.

hatásai is szerepelnek [5, 10]. Szintén érdekes kutatási terület a hosszútávú függés jelenlétének illetve erősségének becslése [1].

Jelen munka a hosszútávon összefüggő, illetve az önhasonló folyamatok elméleti kérdéseinek tisztázásával járul hozzá a témakörben felmerülő problémák megértéséhez.

2. Kutatási célkitűzések

A kutatás háttérét az a feladat adta, hogy számítógép-hálózatok forgalmi viszonyait vizsgáltam, és olyan módszereket kerestem, melyek segítségével egy adott mérési ponton áthaladó forgalom átlaga becsülhető, illetve az átlag változása észlelhető.

Ehhez mindenképp szükséges volt a hálózati forgalmi modellek, illetve ezen modellekhez kidolgozott fontosabb tételek, módszerek megismerése. Itt kerültem szembe először azzal a felismeréssel, hogy annak ellenére, hogy a hálózati modellezés aktívan kutatott terület vannak még hiányosságok az idevonatkozó modellek elméletében. Ez indított abba az irányba, hogy a hosszútávon összefüggő folyamatok elméletébe mélyebben elmerüljek. Habár tisztában voltam vele, hogy a hosszútávon összefüggő folyamatok az aszimptotikusan önhasonló folyamatok részhalmazát képezik, ezen tágabb kör vizsgálata gyakorlati szempontból nem tűnt különösebben érdekesnek. Egészen addig, míg a kidolgozott elméleti módszereket nem kezdtem el egy egyszerű, jól ismert gyakorlati példán ellenőrizni. Ez a példa konkrétan egy FARIMA típusú folyamat volt, egy ismert hosszútávon összefüggő folyamat. A gyakorlati tapasztalatok azonban nem voltak az addig kidolgozott elmélet segítségével magyarázhatók. Ez vezetett végül ahhoz, hogy részletesen foglalkozzak az egzaktul, illetve aszimptotikusan önhasonló folyamatok teljes

osztályával².

Így tehát a meglévő elméletben fellelt hiányosságok késztettek arra, hogy a kutatómunka eredeti célkitűzését mindvégig szem előtt tartva első lépésként a gyakorlat szempontjából fontos elméleti alapok tisztázását tűzzem ki célul.

Az újonnan megfogalmazott cél így a diszkrét idejű, másodrendben egzaktul, illetve aszimptotikusan önhasználó, illetve hosszútávon összefüggő sztochasztikus folyamatok alapvető összefüggéseinek, tulajdonságainak vizsgálata lett. Ezen folyamatokra vonatkozóan a szakirodalomban többféle definíció található. Fontosnak tartottam ezen definíciók összegyűjtését, részletes összehasonlítását. Az irodalomban ezen felül számos eredmény is megtalálható, gyakran azonban elnagyolt, vagy teljesen hiányzó bizonyítással. Ezen eredmények összegyűjtése, rendszerezése, illetve a hiányzó, vagy hiányos bizonyítások pótlása, kiegészítése is kutatásom fontos célja volt. Ezen felül foglalkoztam minden olyan kérdéssel, mely a témakörhöz tartozik, és szükséges ahhoz, hogy megválaszoljuk a megválaszolatlan kérdéseket, illetve egységes rendszerbe foglalhassuk a meglévő eredményeket.

3. Kutatási módszerek

Az itt bemutatott eredményekhez analitikus bizonyítási módszerekkel jutottam. Munkám során rövid Matlab programokat készítettem, melyek kifejezetten hasznosnak bizonyultak a még nem bizonyított elméletek helyességének megerősítésére, illetve cáfolására, valamint váratlan eredmények esetén az újabb kutatási irányok kijelölésére. A diszkrét idejű függvények tárgyalásához egy operátor leírást vezettem be. Ennek segítségével külön tudtam választani a különböző korrelációs struktúrát leíró függvények (pl. autokovariancia fv., autokorrelációs fv.) közötti funkcionális összefüggések vizsgálatát a függvények által teljesítendő egyéb kritériumok (pl. pozitív szemidefinitség) vizsgálatától.

A diszkrét idejűség miatt számelméleti kérdések is felvetődtek, melyek megoldásait elemi úton vezettem le. A munka során diszkrét idejű regulárisan változó függvényeket is használtam. Ennek során rendszereztem és összegeztem ezek főbb tulajdonságait is.

4. Eredmények

Jelen munka diszkrét idejű, másodrendben önhasználó folyamatokkal foglalkozik.

Legyen $\{X(t), t \in \mathbf{Z}\}$ egy diszkrét idejű sztochasztikus folyamat. A folyamatot másodrendben stacionáriusnak mondjuk, ha $\mu := \mathbf{E}[X(0)] = \mathbf{E}[X(t)]$ és $\mathcal{V} := \mathbf{E}[(X(0) - \mu)^2] = \mathbf{E}[(X(t) - \mu)^2]$, $\forall t \in \mathbf{Z}$, vagyis sem $X(t)$ várható értéke, sem annak szórása (más néven varianciája) nem függ t -től, továbbá a folyamat *autokovariancia függvénye*, $\gamma(k) := E[(X(t) - \mu)(X(t+k) - \mu)]$, csak a $k \in \mathbf{Z}$ értéktől függ, továbbá $\gamma(k) = \gamma(-k)$. Vizsgálataink szempontjából egy folyamatot ismertnek tételezünk fel, ha annak autokovariancia függvénye adott. Az egyes értékek konkrét eloszlását nem vizsgáljuk, csupán a szórás létezését követeljük meg.

A folyamat *autokorrelációs függvényét* a $\rho(k) := \frac{\gamma(k)}{\gamma(0)} = \frac{\gamma(k)}{\mathcal{V}}$ összefüggéssel adjuk meg. Ezen ismert függvények mellett további két függvényt definiálunk a folyamat korrelációs,

²A tapasztalt jelenség precíz matematikai bizonyítása jelenleg kutatás tárgya, azonban az azóta kidolgozott eredmények ismeretében a megfigyelt jelenség megérthető és megmagyarázható.

illetve kovariancia struktúrájának jellemzésére. A *variancia idő függvényt* ω -val jelöljük, és a következőképp definiáljuk:

$$\omega(n) = \sum_{k=0}^{n-1} \sum_{i=-k}^k \gamma(i) = n\gamma(0) + 2 \sum_{i=1}^{n-1} i\gamma(n-i), n = 1, 2, 3, \dots \quad (1)$$

A *korrelációs idő függvény* ennek normált változata: $\phi(n) = \frac{\omega(n)}{\omega(1)} = \frac{\omega(n)}{\gamma}$. Ezen két függvény használata – mint később látni fogjuk – nagyban leegyszerűsíti a számításokat.

Az (1) összefüggés inverzeként az autokovariancia függvényt is kifejezhetjük a korrelációs idő függvény segítségével az alábbi módon:

$$\gamma = \mathbf{D}\{\omega\}, \quad (2)$$

ahol \mathbf{D} az ún. kettős differencia operátor, melyet a következőképp definiálunk:

$$\mathbf{D}_i\{f(i)\}(n) = \begin{cases} f(1) & : n = 0 \\ \frac{1}{2}(f(2) - 2f(1)) & : n = 1 \\ \frac{1}{2}(f(n+1) - 2f(n) + f(n-1)) & : n > 1. \end{cases} \quad (3)$$

Vizsgálataink szempontjából mindenképp meg kell ismerkednünk a pozitív szemidefinittség fogalmával. Egy nemnegatív egészeken értelmezett függvényt akkor nevezünk *pozitív szemidefinitnek*, ha $\forall n \in \mathbf{Z}^+$ (pozitív egészre) és minden n hosszú \mathbf{a} valós vektorra:

$$\sum_{1 \leq i, j \leq n} a_i f(|i-j|) a_j \geq 0.$$

Bizonyítható, hogy a pozitív szemidefinittség szükséges, de egyben elégséges feltétele is annak, hogy egy függvény valamilyen folyamat autokovariancia függvénye lehessen [3].

4.1. Önhasonló folyamatok jellemzése

Az önhasonlóságnak több definíciója létezik. Ezen definíciók a folyamat különböző időskálákon való viselkedésének vizsgálatán alapulnak. Akkor tekintjük a folyamatot önhasonlónak, ha egy időbeli aggregálást követő normálás után a számunkra lényeges tulajdonságok változatlanok maradnak.

Definíció 4.1 (Önhasonlóság #1 (SS1))

Adott $X(t)$ esetén definiáljuk $X^{(m)}$, illetve $X'^{(m)}$ folyamatokat a következőképp:

$$X^{(m)}(t) := \frac{1}{m} \sum_{j=m(t-1)+1}^{mt} X(j). \quad (4)$$

$$X'^{(m)}(t) := A_m \sum_{j=m(t-1)+1}^{mt} X(j) = mA_m X^{(m)}(t), \quad (5)$$

ahol az A_m egy előre definiált sorozat, vagy sorozathalmaz. Az X folyamatot önhasonlónak mondjuk, ha X és $X'^{(m)}$ autokovariancia függvényei minden egész m esetén egyeznek.

Ez a definíció valójában egy definíció család az A_m sorozattal (sorozat halmazzal) mint paraméterrel. A szakirodalomban ez, illetve az ehhez hasonló definíciók kivétel nélkül az $A_m := m^{-H}$, $H \in [0, 1]$ feltétellel szerepelnek. Ezek a definíciók tehát az X és $m^{1-H}X^{(m)}(t)$ folyamatok ekvivalenciáját követelik meg. A továbbiakban az $X^{(m)}$ folyamat autokovariancia, autokorreláció stb. függvényeit a $\gamma^{(m)}$, $\rho^{(m)}$, stb. jelölésekkel illetjük.

Meg kell említsük, hogy ez a definíció általában nem ilyen formában szerepel a szakirodalomban. Samorodnitsky és Taqqu [11] definíciója hasonló, ők azonban az X és $X^{(m)}$ folyamatoktól nem csak másodrendű, hanem teljes sztochasztikus ekvivalenciát követelnek.

Sinai [12] és Major [9] definíciói szintén teljes sztochasztikus ekvivalenciát követelnek meg, azonban foglalkoznak többdimenziós sztochasztikus folyamatokkal is. Sinai önhasonlósági definíciója nem stacionárius terekre is alkalmazható.

Az előbbi önhasonlósági definíciót a következővel vetjük össze:

Definíció 4.2 (Önhasonlóság #2 (SS2))

Az X folyamatot önhasonlónak nevezük ha X és $X^{(m)}$ autokorrelációs függvényei minden nemnegatív m -re egyeznek. ($\rho = \rho^{(m)}$, $\forall m \in \mathbf{Z}^+$.)

Itt ezen utóbbi (SS2) definíciót használom.

SS1 folyamatok esetén $X^{(m)}$ autokovariancia függvénye $\gamma^{(m)} \equiv C_m \gamma$, ahol $C_m = (mA_m)^{-2} = m^{2H-2}$. A C_m sorozatot tehát az A_m sorozat egyértelműen definiálja. Tekintsük most a 4.2. definíciót! Egy folyamat autokorrelációs függvénye az autokovariancia függvénytől csupán egy multiplikatív konstansban tér el. Így ha két folyamat autokorrelációs függvénye egyezik, akkor autokovariancia függvényük szintén legfeljebb egy multiplikatív konstans erejéig különbözhet. Vagyis az SS2 folyamatok esetén mindig létezik egy C_m sorozat, melyre $\gamma^{(m)} \equiv C_m \gamma$, továbbá az SS2 folyamatok halmaza tartalmazza az összes lehetséges C_m sorozathoz tartozó folyamatot.

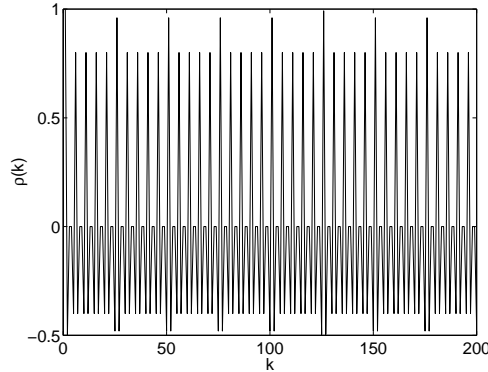
Így tehát láthatjuk, hogy az SS1 folyamatok az SS2 folyamatok részhalmazát képezik. Hogy valódi részhalmazról van-e szó, az attól függ, hogy hogyan definiáljuk az A_m sorozatok halmazát.

1. Téziscsoport Önhasonló folyamatok feltárása

Ebben a téziscsoportban a 4.2. definíció szerinti összes SS2 folyamat feltárásával és leírásával foglalkozom. Célom az összes ilyen folyamat autokorrelációs vagy korreláció idő függvényének explicit megadása.

Az SS2 folyamatok közül eddig csupán a frakcionális zaj volt ismeretes, én egy új típusú önhasonló folyamatot definiáltam.

A frakcionális zajt legegyszerűbben korrelációs idő függvényével definiálhatjuk, mely $\phi = n^{2H}$ alakú, ahol $H \in [0, 1]$. Rövid számítással igazolható, hogy az a sorozat kielégíti az SS2 definíciót, illetve az SS1 definíciót is a már fentebb említett $A_m = m^{-H}$ sorozattal. Amennyiben a folyamatról megköveteljük, hogy minden véges dimenziós eloszlása normális legyen, akkor frakcionális Gaussi zajnak hívjuk. Tekintve, hogy itt csak másodrendű leírással foglalkozunk a „Gaussi” jelzőtől eltekintünk. Ezeket a folyamatokat FN_H -val fogjuk jelölni.



2. ábra. Az $AP_{5,0.2}$ folyamat autokorrelációs függvényének első 200 értéke.

Tézis 1.1 (Az önhasonló kvázi-periodikus $AP_{q,c}$ osztály definiálása [P4])

A kvázi-periodikus (almost periodic, AP) osztályt a következőképp definiáljuk: Legyen q egy prímszám, és c egy 0 és 1 közötti konstans. Az $AP_{q,c}$ folyamat $\phi_{q,c}$ korrelációs idő függvénye ekkor p prímszám értékeknél a következőképp alakul:

$$\phi_{q,c}(p) = \begin{cases} 1, & \text{ha } p \neq q \\ c, & \text{ha } p = q \end{cases}$$

Továbbá $\phi_{q,c}(1) = 1$ és minden nem prím n esetén a függvény értéke:

$$\phi_{q,c}(n) = \prod_{i=1}^s \phi^{r_i}(p_i),$$

ahol p_1, p_2, \dots, p_s jelöli n prímtényezőit, r_i pedig p_i kitevőjét n prímfelbontásában.

Az $AP_{q,c}$ osztály egy tagjának autokorrelációs függvényét mutatja a 2. ábra. Az ábrán jól látszik, hogy miért kapta a folyamat a kvázi-periodikus nevet.

Tézis 1.2 (A kvázi-periodikus folyamatok önhasonlók [P4])

Megmutattam, hogy az 1.1. tézisben definiált $\phi_{q,c}$ függvény pozitív szemidefinit, így tehát a kvázi-periodikus osztály létezik. Ezen folyamatok önhasonlóságát is bebizonyítottam.

Tézis 1.3 (Az összes önhasonló osztály [P3])

Bebizonyítottam, hogy az itt bemutatott kvázi-önhasonló, és a korábban már ismert frakcionális zaj típusú osztályokon kívül más önhasonló folyamat nem létezik.

4.2. Aszimptotikusan önhasonló folyamatok

Ebben a részben ismertetem az aszimptotikusan önhasonló folyamatokat és bemutatom azok néhány jellemző részhalmazát.

2. Téziscsoport Aggregációs határértékek, az aszimptotikusan önhasonló folyamatok definíciója

Itt azon folyamatokkal foglalkozunk, melyek nem önhasonlóak, de a $\rho^*(k) := \lim_{m \rightarrow \infty} \rho^{(m)}(k)$ határérték létezik minden $k = 0, 1, 2, \dots$ értékre.

Nem nehéz példát konstruálni olyan esetre, mikor egy végtelen sorozat összes tagja rendelkezik egy adott tulajdonsággal, a sorozat határértéke azonban nem. Ezért bebizonyítottam, hogy

Tézis 2.1 (Az autokorrelációs függvény határértéke pozitív szemidefinit [P3])

Bebizonyítottam, hogy amennyiben ρ egy pozitív szemidefinit autokorrelációs függvény és $\rho^(k) = \lim_{m \rightarrow \infty} \rho^{(m)}(k)$ létezik minden $k = 0, 1, 2, \dots$ esetén, akkor $\rho^*(k)$ is pozitív szemidefinit, azaz létezik olyan folyamat melynek $\rho^*(k)$ az autokorrelációs függvénye.*

Tézis 2.2 (Az autokorrelációs függvény határértéke önhasonló [P3])

Bebizonyítottam, hogy amennyiben a $\rho^(k) := \lim_{m \rightarrow \infty} \rho^{(m)}(k)$ határérték minden $k = 0, 1, 2, \dots$ értékre létezik, akkor minden $n = 1, 2, 3 \dots$ esetén: $\rho^{*(n)} \equiv \rho^*$, azaz ρ^* egy önhasonló folyamat autokorrelációs függvénye.*

A 2.1. illetve 2.2. tézisek tudatában az aszimptotikusan önhasonló folyamatok definícióját a következőképp adhatjuk meg:

Definíció 4.3 (Aszimptotikusan önhasonló folyamat)

Aszimptotikusan önhasonlónak nevezünk egy folyamatot akkor, ha a $\rho^*(k) := \lim_{m \rightarrow \infty} \rho^{(m)}(k)$ határérték létezik.

Az aszimptotikus önhasonlóságnak létezik egyéb definíciója is, mely megköveteli, hogy $\rho^*(k)$ a frakcionális zaj autokorrelációs függvénye legyen [6]. Az így definiált folyamatok részhalmazát képezik a 4.3 által definiáltaknak.

3. Téziscsoport Aszimptotikusan önhasonló folyamatok jellemző osztályai

Az aszimptotikusan önhasonló folyamatoknak jellegzetes osztályait kerestem. Gyakorlati jelentőségük miatt azokkal a folyamatokkal foglalkoztam, melyek a frakcionális zajhoz konvergálnak. Az aszimptotikusan önhasonló folyamatok kereséséhez $\rho^{(m)}$ viselkedését kell vizsgálnunk. Ennek értékeit a ρ függvény értékeivel fejezzük ki, és megvizsgáljuk, hogy milyen kezdeti ρ értékeknél konvergál. Azonban $\rho^{(m)}$ kifejezése ρ segítségével meglehetősen bonyolult, nehézkes.

Ezzel ellentétben a $\rho^{(m)}$ -mel ekvivalens $\phi^{(m)}$ függvényt ϕ -ből kifejezni egyszerű:

$$\phi^{(m)}(n) = \frac{\phi(mn)}{\phi(m)}. \tag{6}$$

A (6) egyenlet egyszerűségét kihasználva egy új definíciót adtam az aszimptotikus önhasonlóságra. Erről bemutattam, hogy ekvivalens a 4.3. definícióval.

Tézis 3.1 (Asz. önhasonlóság definiálása a korrelációs idő függvénnyel [P4, P3])

Azon folyamatokat, melyekre $\phi^(k) := \lim_{m \rightarrow \infty} \phi^{(m)}(k)$ létezik, aszimptotikus önhasonlónak nevezzük.*

Nem nehéz megmutatni, hogy egy folyamat autokorrelációs (ρ) és korrelációs idő (ϕ) függvényei között egyértelmű megfeleltetés van. Megmutattam, hogy ez az egyértelmű megfeleltetés ezen függvények határértékei között is fennáll.

Tézis 3.2 (ρ és ϕ felcserélhetősége [P3])

Jelölje ρ illetve ϕ egyazon folyamat autokorrelációs, illetve korreláció idő függvényeit. Megmutattam, hogy $\rho^* := \lim_{m \rightarrow \infty} \rho^{(m)}$ pontosan abban az esetben létezik, amikor $\phi^* := \lim_{m \rightarrow \infty} \phi^{(m)}$ is létezik, továbbá ϕ^* és ρ^* között ugyanaz az egy-egy értelmű megfeleltetés van, mint ϕ és ρ között, vagyis az (1) illetve (2) egyenletek által definiált kettős összeg, illetve kettős különbség operátorokkal számolhatók ki egymásból.

A (6) egyenlet segítségével egyszerű feltételt kapunk arra, hogy az aggregáció során a korrelációs idő függvény a frakcionális zajhoz konvergáljon:

$$\lim_{m \rightarrow \infty} \phi^{(m)}(n) = \lim_{m \rightarrow \infty} \frac{\phi(mn)}{\phi(m)} = n^{2H}. \quad (7)$$

Azon autokorrelációs függvények halmazát, melyek egy adott autokorrelációs függvényhez konvergálnak, ezen utóbbi függvény *vonzáskörzetének* hívjuk. Hasonlóképp definiáljuk a korrelációs idő függvények, illetve az önhasonló folyamatok vonzáskörzetét is.

Annak ellenére, hogy a H paraméterű frakcionális zajhoz történő konvergencia kritériumát H -tól függetlenül ugyanazzal a (7) egyenlettel fejezhetjük ki, a különböző H értékekhez tartozó folyamatok viselkedése jelentősen különbözhet egymástól. Így a következő eseteket különböztetjük meg: $H \in \{0\}, (0, 0.5), \{0.5\}, (0.5, 1), \{1\}$.

Az itt bemutatott folyamatok halmaza nem fedti le egy adott H -hoz tartozó összes aszimptotikusan önhasonló folyamat halmazát, de jellemző példaként szolgálnak, melyek segítségével az osztályra jellemző tulajdonságok szemléltethetők.

Tézis 3.3 ($H = 0$ [P3])

Bebizonyítottam, hogy amennyiben egy X folyamat esetén $\lim_{k \rightarrow \infty} \rho(k)$ létezik és $\rho(1) < 1$, úgy az $Y(i) = X(i+1) - X(i)$ folyamat FN_0 vonzáskörzetében van.

Tézis 3.4 ($H \in (0, 0.5)$ [P3])

Amennyiben $\sum_{i=-\infty}^{\infty} \rho(i) = 0$ és $\rho(k) \sim ck^{2H-2}$, továbbá $H \in (0, 0.5)$, úgy a folyamat FN_H -hoz konvergál.

A „ \sim ” jel az aszimptotikus ekvivalenciát jelöli. Két végtelen sorozat aszimptotikusan ekvivalens, ha hányadosuk határértéke 1-hez tart.

Tézis 3.5 ($H = 0.5$ [P3])

Azon folyamatok, melyekre $\sum_{i=-\infty}^{\infty} \rho(i) \in (0, \infty)$, vagyis az autokorrelációs függvény összegezhető, de nem nulla értéket ad, az $FN_{0.5}$ osztályhoz konvergálnak.

Az $FN_{0.5}$ osztályt más néven fehér zajnak hívjuk, a hozzá konvergáló folyamatokat pedig *rövid-távon összefüggőnek*.

Tézis 3.6 ($H \in (0.5, 1)$ [P3])

Azon folyamatok, melyekre $\rho(k) \sim ck^{2H-2}$, ahol $H \in (0.5, 1)$ az FN_H folyamathoz konvergálnak.

Ezek a folyamatok az ún. *hosszútávon összefüggő* folyamatok tipikus a esetét adják. Legfontosabb jellemzőjük, hogy az autokorreláció annyira erős, hogy $\sum_{i=-\infty}^{\infty} \gamma(i) = \infty$. Az ilyen folyamatokkal bővebben a 4.3. fejezetben foglalkozunk.

Tézis 3.7 ($H = 1$ [P3])

Legyen Y egy nulla várható értékű, egységnyi szórású folyamat. Definiáljuk az X folyamatot az $X = \{Y, aY, Y, aY, Y \dots\}$ egyenlettel, ahol $a \in [-1, 1]$. A folyamatot azonos valószínűséggel indítjuk az Y , illetve aY értékektől, így biztosítva a stacionaritást. Ebben az esetben X az FN_1 vonzáskörzetében van.

4.3. Hosszútávú összefüggés

Az ún. hosszútávon összefüggő folyamatok az aszimptotikusan önazonosoknak fontos esetei, mivel gyakorlati helyzetekben jellemzően ezekkel találkozunk. Ilyen folyamattal modellezhetjük többek között a csomagkapcsolt hálózatok forgalmát is.

4. Téziscsoport Hosszútávú összefüggés definíciói

A szakirodalomban számos definíciót találunk a hosszútávú függés definíciójára [1, 4, 6]. Ezek egymáshoz hasonlóak, ugyanazt a jelenséget próbálják leírni. Egy ilyen rendszerező, áttekinthető munka megkövetelte ezen definíciók összehasonlítását. Ez az összehasonlítás azért is szükséges mert a folyamatok viselkedésének tanulmányozásakor fontos tudnunk, hogy az egyes állításoknak, tételeknek pontosan milyen szükséges kritériumai vannak. Erre vonatkozólag ugyanis gyakran csak elnagyolt utalások találhatók az irodalomban. A hosszútávú függés számos definíciója tartalmaz diszkrét idejű reguláris változást. Azonban ezen utóbbi definícióját, és fontosabb tulajdonságait is gyakran csak elnagyolva találjuk. Így a diszkrét reguláris változás megfelelő definíciója, illetve fontosabb tulajdonságainak feltárása is fontos a hosszútávon összefüggő folyamatok viselkedésének megértéséhez.

4.3.1. Diszkrét idejű regulárisan változó függvények

A folytonos idejű reguláris változás már kiforrott terület, jelentős szakirodalommal rendelkezik [2]. Annak érdekében, hogy a folytonos idejű eredményeket diszkrét időben is felhasználhassuk, olyan definíciót keresünk, mely a folytonos idejűnek a lehető legtöbb tulajdonságával rendelkezik. Ezek a később részletezett tulajdonságok azért is szükségesek, hogy a hosszútávon összefüggő folyamatokról bizonyos állításokat kimondhassunk. Olyan definíciót azonban, mely a folytonos idejűnek összes tulajdonságával rendelkezik, nem találunk.

Az általam használt definíció, habár nem ugyanabban a formában van kimondva, ekvivalens Galambos [8] definíciójával.

Definíció 4.4 (Folytonos reguláris változás)

Egy pozitív valós számokon értelmezett folytonos \tilde{f} függvény *regulárisan változó* (a végtelen-

ben) α indexszel, ha

$$\lim_{t \rightarrow \infty} \frac{\tilde{f}(tx)}{\tilde{f}(t)} = x^\alpha, \quad \alpha \in \mathbf{R} \quad (8)$$

minden $x \in \mathbf{R}^+$ esetén. (Szigorúan nézve az is elégséges ha (8) a pozitív valósak egy sűrű halmazán teljesül, [7] 275. oldal). Amennyiben $\alpha = 0$, úgy az \tilde{f} függvényt *lassan változónak* nevezzük. Az α indexű folytonos regulárisan változó függvények halmazát FRV_α -val jelöljük, míg a lassan változó függvények halmazát FLV jelöli.

Definíció 4.5 (Diszkrét reguláris változás (DRV))

A pozitív egészeken értelmezett f függvényt a (végtelenben) α indexű *regulárisan változónak* nevezzük, ha létezik egy $\tilde{f} \in FRV_\alpha$, ahol $f(n) = \tilde{f}(n)$ minden $n \in \mathbf{Z}^+$ esetén. Ezen függvények halmazát DRV_α -val jelöljük, míg a lassan változó függvényekét DLV-vel.

Tézis 4.1 (Diszkrét regulárisan változó függvények tulajdonságai [P3])

A diszkrét idejű reguláris változás definícióját a folytonos idejűre alapoztuk, így várható, hogy annak számos tulajdonságát megőrzi. Megmutattam, hogy ezek a tulajdonságok a következők:

- $f \in DRV_\alpha \Rightarrow \lim_{k \rightarrow \infty} f(kn)/f(k) = n^\alpha, \alpha \in \mathbf{R}, n \in \mathbf{Z}^+$
- $f \in DRV_\alpha \Leftrightarrow f(k) = s(k)k^\alpha, s(k) \in DLV$
- $f \in DRV_\alpha$ és $g \sim f$ akkor $g \in DRV_\alpha$
- $f \in DRV_\alpha \Rightarrow f(k) \sim f(k + k_0), \forall k_0$ konstans esetén.
- Legyen $K(n) \in DRV_\alpha$, és legyenek $L(t)$ valamint $U(t)$ az alábbi egyenletekkel definiálva:

$$L(m) := \sum_{n=0}^{m-1} K(n), \quad U(m) := \sum_{n=m}^{\infty} K(n).$$

(a) Ha $\alpha \geq -1$ akkor $\frac{mK(m)}{L(m)} \rightarrow (1 + \alpha)$, és $L \in DRV_{\alpha+1}$.

(b) Ha $\alpha < -1$ akkor $\frac{mK(m)}{U(m)} \rightarrow -(1 + \alpha)$, és $U \in DRV_{\alpha+1}$.

4.3.2. Hosszútávú összefüggőség definíciói

Ebben a részben összehasonlítom az irodalomban fellelhető definíciókat, valamint egy új definíciót javasolok, mely megőrzi az eddigi definíciók szellemét, azonban kiterjeszti a hosszútávon összefüggő folyamatok halmazát.

A 4.3.3. részben ezen definíciókat hasonlítom össze, illetve a fontosabb tulajdonságaikat vizsgálom.

Definíció 4.6 (LRD1)

Angol nyelvű betűszava után LRD1-nek nevezzük azokat a folyamatokat, melyek autokovariancia függvénye $\gamma(k) \sim c_\gamma k^{2H-2}$, ahol $H \in (0.5, 1)$, és $c_\gamma \in \mathbf{R}^+$ konstansok.

Ez a legelterjedtebb definíció, többek között ez szerepel Beran, illetve Cox műveiben is [1, 4].

Definíció 4.7 (LRD2)

LRD2-nek akkor nevezzük a folyamatot, ha autokovariancia függvénye

$$\gamma(k) = c_\gamma(k)k^{2H-2}, \quad (9)$$

alakú, ahol $H \in (0.5, 1)$ konstans, illetve $c_\gamma \in DLV$.

Ez a definíció az LRD1 általánosítása, hiszen megengedi, hogy a képlet jobb oldalán a c_γ konstans helyett - mely tulajdonképpen egy speciális lassan változó függvény - egy tetszőleges lassan változó függvény szerepeljen.

Definíció 4.8 (LRD3)

LRD3 a folyamat akkor, ha az autokovariancia függvény kumulatív összege végtelen, vagyis $\sum_{i=1}^{\infty} \gamma(i) = \infty$.

Ez a definíció, mely például Taqqu művében [11] szerepel, azt fejezi ki, hogy mennyire erős a távoli értékek egymásra való hatása.

Tézis 4.2 (Hosszútávú összefüggőség új definíciója [P3])

LRD osztályba soroljuk az összes olyan folyamatot, mely az FN_H , $H \in (0.5, 1)$ vonzáskörzetében van.

Mint azt később részletezzük ez a definíció kiterjeszti az LRD1, illetve LRD2 definíció szerinti folyamatok körét.

4.3.3. A Hosszútávú összefüggés fontosabb jellemzői

Tézis 4.3 (LRD2 folyamatok aggregált variancia függvénye lassan lecsengő [P3])

Bebizonyítottam, hogy LRD2 folyamatok esetén

$$\mathcal{V}^{(m)} \sim \frac{c_\gamma(m)m^{2H-2}}{H(2H-1)}. \quad (10)$$

Ez az állítás több helyen is szerepel, többek között Beran összefoglaló művében [1] is, azonban kellően körültekintő bizonyítását sem itt, sem másutt nem találtam. Ez az eredmény képezi az alapját az egyik legegyszerűbb és talán emiatt széles körben elterjedt hosszútávú összefüggés becslőnek, a variancia-idő diagramnak. Ez a módszer a hosszútávú összefüggés meglétére az aggregált variancia függvény $\mathcal{V}^{(m)}$ alakjából következtet.

Tézis 4.4 (Minden LRD2 folyamat egyben LRD is [P3])

Bebizonyítottam, hogy az LRD2 folyamatok az LRD részhalmazát képezik, vagyis az FN_H , $H \in (0.5, 1)$ vonzáskörzetében vannak.

Ez az állítás közvetlenül adódik a 4.3. tézisből, illetve (7) egyenletből. Ez a példa is szemlélteti, hogy a korrelációs idő függvény használata mennyire leegyszerűsíti a számításokat.

Tézis 4.5 (Az LRD2 valódi részhalmaza az LRD-nek [P3])

A szakirodalomban általában nem vetődik fel a kérdés, hogy vajon létezik-e az LRD2 folyamatokon kívül más is, mely az FN_H vonzaskörzetében van. Konstruáltam egy olyan folyamatot, mely LRD, azonban nem LRD2. Így konstruktívan bizonyítottam a két osztály különbözőségét.

5. Gyakorlati következmények

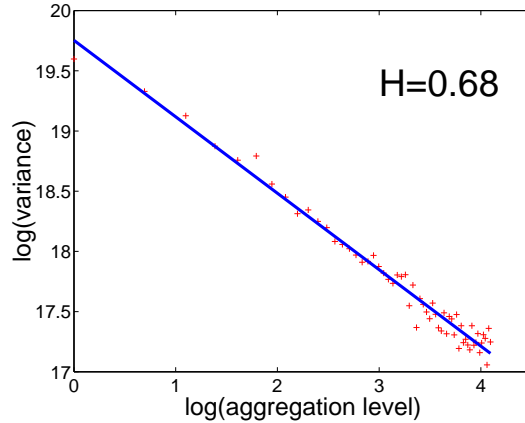
A disszertációban bemutatom, hogy a variancia idő függvény alkalmazása milyen előnyökkel jár folyamatok különböző időskálákon történő viselkedésének tanulmányozásakor. Itt egy példán keresztül azt szemléltetem, hogy mennyiben járulnak hozzá az itt bemutatott új eredmények a hosszútávú összefüggőség jelenlétét, illetve erősségét tesztelő eljárások korrekt alkalmazásához.

Tekintsük a variancia-idő diagramot, mely az LRD folyamatok aggregált varianciájának lassú lecsengésén alapszik (ld. 4.3. tézis).

Amennyiben a vizsgált folyamat LRD1 akkor a (9) képletben szereplő $c_\gamma(m)$ a konstans c_γ értékhez konvergál. Így a (10) egyenletet a következő alakba írhatjuk:

$$\mathcal{V}^{(m)} \sim \frac{c_\gamma m^{2H-2}}{H(2H-1)}. \quad (11)$$

Vagyis nagy m értékeknél az aggregált variancia ($\mathcal{V}^{(m)}$) közelítőleg hatványfüggvény alakú. Így ha $\mathcal{V}^{(m)}$ logaritmusának értéket mint m logaritmusának függvényét ábrázoljuk, akkor nagy m értékekre közelítőleg egy $2H-2$ meredekségű egyenest kell kapjunk. Ezt mutatja a 3. ábra.



3. ábra. A H paraméter becslése a variancia-idő diagram segítségével

Ezt a módszert használjuk hosszútávú összefüggőség jelenlétének ellenőrzésére, valamint a H paraméter becslésére.

A folyamat véletlen jellegéből és az adatsor végességéből adódóan mindig fellép kisebb-nagyobb mértékű pontatlanság. Szintén fontos kérdés, hogy mekkora m értéket tekintünk elegendően nagynak, vagyis honnantól kezdjük az egyenes illesztését. Ezekkel a fontos technikai kérdésekkel itt azonban nem foglalkozunk. Itt azt feltételezzük, hogy a mintaméret elegendően nagy, hogy az esetlegesen ebből adódó hibák hatását elhanyagolhassuk, vagyis a módszer elméleti alkalmazhatóságát vizsgáljuk.

Amennyiben a vizsgált folyamat LRD1 típusú akkor az ábrázolt függvény vége egy $2H - 2$ meredekségű egyenes, így H értéke becsülhető.

A becslés, illetve tesztelés algoritmikus lépései a következők:

1. Becsüljük meg $\mathcal{V}^{(m)}$ értékét $m = 1, 2, \dots, m_{\max}$ értékekre, ahol $m_{\max} \ll n$, illetve n a minták száma ³.
2. Ábrázoljuk $\log \mathcal{V}^{(m)}$ értéket $\log m$ függvényeként
3. Illesszünk egyenest a görbe végére
4. Sikeres az illesztés? Ha nem, akkor a folyamat nem LRD1
5. Ha igen, akkor a folyamat LRD1 (H értékétől függően) az egyenes meredekségéből becsüljük H értékét

Ez a módszer tehát működik LRD1 folyamatok esetén. Az LRD1 osztályba tartozás tesztelhető, H értéke becsülhető. Vizsgáljuk meg, hogy hogyan viselkedik a módszer nem LRD1 típusú folyamatok esetén!

Tegyük fel, hogy a vizsgált folyamat LRD2, de nem LRD1. Ebben az esetben a (10) egyenletben szereplő c_γ függvény nem konvergál konstans értékhez, vagyis akármilyen nagyra választhatjuk m -et, a felrajzolt görbe vége nem lesz egyenes. Így tehát a módszer elvileg nem alkalmas az LRD2 folyamatok detektálására.

Látjuk tehát, hogy az egyenesre illeszkedés szükséges feltétele annak, hogy a folyamat LRD1 osztályban legyen, ez azonban nem igaz az LRD2 osztály esetén. Most azt vizsgáljuk meg, hogy vajon ez a feltétel az LRD1 osztály esetén elégséges-e?

Amennyiben az adatok egy egyenesre illeszkednek, akkor arra következtethetünk, hogy a (11) egyenlet teljesül. Megmutatható, hogy ez szükséges, azonban nem elégséges feltétele annak, hogy a folyamat LRD1 legyen. Így tehát az adatok egyenesre illeszkedése nem elegendő annak bizonyítására, hogy a folyamat LRD1 típusú.

Habár több más tesztelő, illetve becslő eljárással kapcsolatban hasonló következményeket vonhatunk le, a variancia-idő diagram példáján jól látszik, hogy csak akkor tudjuk a kapott eredményt helyesen értelmezni, ha értjük az összefüggéseket, illetve különbségeket az egyes LRD definíciók között.

Köszönetnyilvánítás

Munkám során a legtöbb segítséget témavezetőimtől Molnár Sándortól illetve Darryl Veitch-től kaptam akiknek munkáját ezúton is köszönöm. Köszönet illeti szerzőtársaimat is, Maricza Istvánt és Ruzsa Imrét is, akik észrevételeikkel, munkájukkal hozzájárultak ahhoz, hogy az eredmények a mostani letisztult formájukat elnyerjék.

Köszönöm azoknak a személyeknek, illetve intézményeknek, melyek munkámhoz az anyagi, illetve infrastrukturális háttérrel biztosították:

³Ahogy m értéke növekszik az $X^{(m)}$ folyamatnak egyre kevesebb értéke lesz, így $\mathcal{V}^{(m)}$ becslése egyre pontatlanabb.

- Magyar Állam, állami ösztöndíj formájában
- Henk Tamás, és a Nagysebességű Hálózatok Laboratóriuma, a Budapesti Műszaki és Gazdaságtudomány Egyetem Távközlési és Médiainformatikai tanszékéről
- Ausztrál Ifjúsági és Oktatási Minisztérium
- Royal Melbourne Institute of Technology
- Darryl Veitch, EMUlab, University of Melbourne

Hivatkozások

- [1] J. Beran. *Statistics for Long-Memory Processes*. Chapman and Hall, New York, 1994.
- [2] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge England, 1987.
- [3] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, 1996.
- [4] D. R. Cox. *in Long-range dependence: a review*, H.A. David and H.T. David editors, chapter Statistics : an Appraisal, pages 55–74. Iowa State University Press, Ames (IA), 1984.
- [5] Mark E. Crovella and Azer Bestavros. Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. In *ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Networks*, May 1996.
- [6] D. L. Jagerman and B. Melamed and W. Willinger. Stochastic Modeling of Traffic Processes. In *Frontiers in Queueing: Models, Methods and Problems*. CRC Press, 1996.
- [7] William Feller. *An Introduction to Probability Theory and its Applications*, volume II. John Wiley and Sons, Brisbane, second edition, 1970.
- [8] J. Galambos and E. Seneta. Regularly Varying Sequences. *Proceedings of the American Mathematical Society*, 41(1):110–116, November 1973.
- [9] P. Major. *Multiple Wiener-Itô Integrals*, volume 849 of *Springer Lecture Notes in Mathematics*. Springer-Verlag, New York, 1981.
- [10] Kihong Park, Gi Tae Kin, and Mark E. Crovella. On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic. Technical report, Computer Science Department, Boston University, 1996. TR-96-016.
- [11] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman and Hall, 1994.
- [12] Ya. G. Sinai. Self-Similar Probability Distributions. *Theory of Probability and its Applications*, 21:64–80, 1976.

Publikációk

- [P1] Gefferth András. WWW információs rendszer az Austria Telecomnál (WWW Information system at Austria Telecom). *Magyar Távközlés (Hungarian Telecommunications)*, February 1996.
- [P2] A. Gefferth, S. Molnár, and D. Veitch. Discrete Self-Similarity. In *IFIP WG6.7 Workshop and EUNICE Summer School on Adaptable Networks and Teleservices*, Trondheim, Norway, September 2-4 2002,pp. 55-61.
- [P3] A. Gefferth, D. Veitch, I. Maricza, S. Molnár, and I. Ruzsa. The Nature of Discrete Second-Order Self-Similarity. *Advances in Applied Probability*, 35(2), June 2003,pp. 395-416.
- [P4] A. Gefferth, D. Veitch, I. Ruzsa, I. Maricza, and S. Molnár. A New Class of Second Order Self-Similar Processes. *Stochastic Models*, 20(3):381–389 September 2004,pp 381-389.
- [P5] S. Molnár and A. Gefferth. On the Scaling and Burst Structure of Data Traffic. In *8th International Conference on Telecommunication Systems, Modelling and Analysis*, Nashville, Tennessee, USA, March 2000.