A sztochasztikus idősorelemzés alapjai

Ferenci Tamás
BCE, Statisztika Tanszék
tamas.ferenci@medstat.hu

2011. december 19.

Tartalomjegyzék

1. fejezet

Az idősorelemzés fogalma, megközelítései

Ebben a fejezetben elsőként definiáljuk az idősor fogalmát (1.1. pont), majd bemutatjuk az idősorok elemzésével foglalkozó két alapvető megközelítési módot (1.2. pont).

1.1. Az idősor fogalma

Az idősor fogalmát megragadhatjuk „sokasági" (elméleti idősor) és „minta" (empirikus idősor) szemléletben.

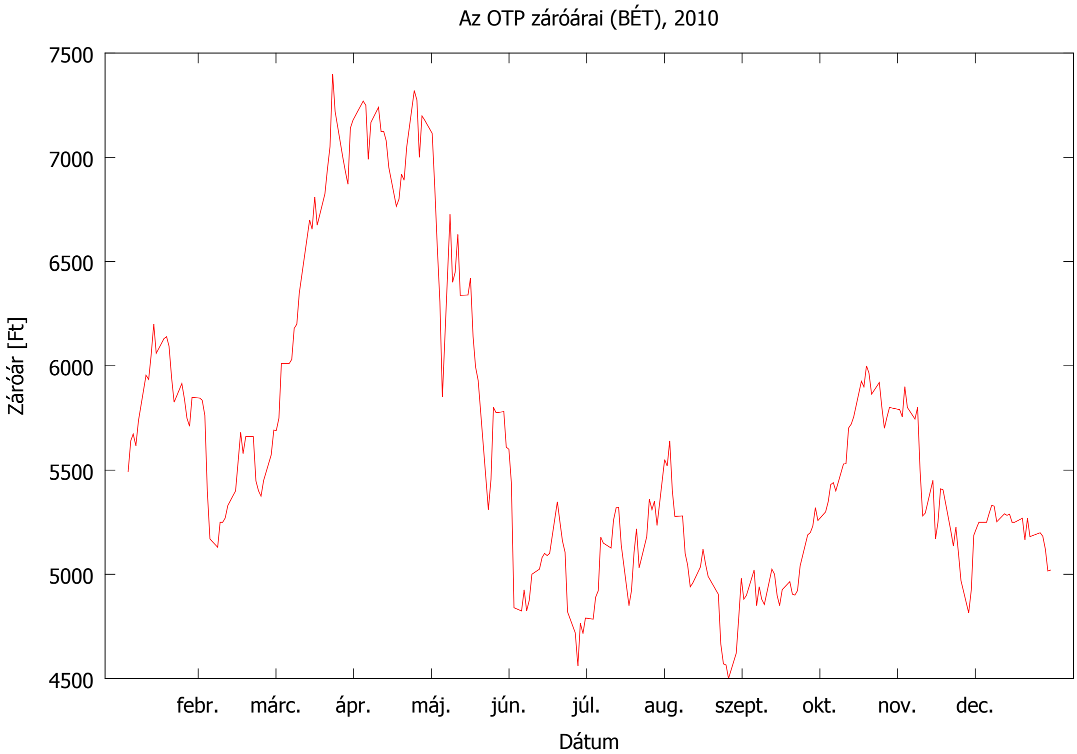
Sokasági értelemben idősornak nevezzük valószínűségi változók egy indexelt $\{Y_t, t \in N\}$ családját[1], ahol a t indexet „idő"-nek fogjuk nevezni. (Az indexelés, amire az egyes valószínűségi változóknál az alsó index utal, a valószínűségi változók között egy sorrendet állít fel: Y_1 „előbb" van, Y_2 „később".) Ilyen értelemben ez tehát lényegében valószínűségi változók egy *sorbarendezett* halmaza. Jelentheti például Y_t a t nap végi OTP záróárfolyamot, a t nap végi HUF/EUR árfolyamot, a t évbeli magyarországi gabonatermelést stb. Néha az is hasznos lesz, ha úgy gondolunk erre, mint egy többdimenziós valószínűségi változóra, melynek Y_t-k a komponensei (csak épp, szemben a szokásos többdimenziós eloszlásokkal, a sorrendjünk nem indifferens). N a lehetséges időpontok halmaza; ez a közgazdasági gyakorlatban legtöbbször diszkrét, nagyon gyakran egy véges halmaz, például $N = \{1, 2, \ldots, T\}$. Ekkor tehát T számú időpontunk van, melyeket 1-től T-ig indexeltünk; a valószínűségi változóink így tehát: $Y_1, Y_2, \ldots, Y_T$. (Műszaki és természettudományos gyakorlatban előfordulnak folytonos indexhalmazok is, pl. lehet $N = \mathbb{R}_+$. Mi ilyen, ún. folytonos idejű idősorokkal a továbbiakban nem foglalkozunk, csak diszkrét idejűekkel, és az egyszerűség kedvéért azon belül is az $N = \{1, 2, \ldots, T\}$ esettel.) Lehet például az indexelés jelentése az, hogy $t = 1$ a jövő hétfői OTP záróárfolyam, $t = 2$ a jövő keddi stb., és $T = 5$ (azaz a jövő hét öt munkanapjára vonatkozó OTP záróárfolyam az idősorunk).

Minta megközelítésben idősornak azt a $y_1, y_2, \ldots, y_T$ statisztikai adatsort (adatbázist) nevezzük, amelynek az megfigyelési egységei sorbarendezettek, valamilyen időponthoz kötöttek. (Itt már láthatóan alkalmaztuk azt a konvenciót, hogy az időpontjaink az $\{1, 2, \ldots, T\}$ halmaz elemei.) Míg tehát az előbbiekben valószínűségi változók (sorbarendezett) sorozatával van dolgunk, itt már konkrét számok (sorbarendezett) sorozatával. Nyilvánvaló, hogy ez utóbbi az előbbi egy realizációja. (Pontosan ugyanúgy, ahogy az 1406 eladásra kínált lakás is egy-egy, összesen 1406 realizáció a kínálati ár, alapterület stb. változók (ismeretlen) együttes eloszlásából.)

[1]Pontosan ugyanezt a fogalmat a valószínűségszámításban sztochasztikus folyamatnak szokás nevezni.

Ez utóbbi megjegyzés rögtön mutatja az idősorelemzés legnagyobb problémáját (és egyben persze kihívását): azt, hogy a gyakorlati feladatokban az egyes időpontokhoz tartozó valószínűségi változónkra csak egy realizációnk lesz! Nem hogy 1406 mintát nem vehetünk a valószínűségi változóból, de kettőt sem. (Aligha lehet a „holnapi OTP-záróárfolyam", mint valószínűségi változóból két mintát venni. . .) Szokás ezt, nagyon találóan, a *reprodukálhatatlanság* problémájának is nevezni.

Mindezeket példázza az 1.1. ábra[2], mely az OTP 2010 évi záróárait ábrázolja. 2010-ben összesen 254 kereskedési nap volt a BÉT-en, ez tehát egy $T = 254$ elemű idősor; január 4-től (az első kereskedési naptól) 1-gyel kezdve, minden kereskedési napon egyesével növekvően indexelhetjük. A fentiek fényében világos, hogy az OTP záróárfolyamainak alakulását elvileg egy 254-dimenziós valószínűségi változó írja le; az ábrán ennek egyetlen realizációja látható. . . ami egyúttal bizonyosan az egyetlen létező realizáció is erre az idősorra.



1.1. ábra. Az OTP 2010 évi záróárai a BÉT-en

Ennek megfelelően tehát soha ne felejtsük el, hogy az empirikus idősor, hiába is áll 254 számból, valójában egyetlen realizáció – csak épp egy 254-dimenziós valószínűségi változóból. (Ahogy például a kínálati ár, alapterület, szobaszám együttes eloszlásából vett egyetlen realizáció is három számból áll.) A fenti ábrázolás azt juttatja kifejezésre, hogy az idősoros adatok specialitása (szemben a lakásos példával), hogy a valószínűségi változó komponensei között (és így persze a realizált komponensei között is) sorrendezés van: az ábrázolásnak csak úgy van értelme, ha erre

[2]Az ábrázolás elvileg nem teljesen korrekt, hiszen ez az idősor ugyebár diszkrét, ezért az egyes pontokat nem köthetnénk össze, ám itt olyan sok időpontunk van, hogy ez lényeges hibát nem jelent (amúgyis „szabad szemmel" szinte megkülönböztethetetlenül közel lennének a pontok).

tekintettel vagyunk; nyilván az ábra is e sorbarendezés figyelembevételével készült.

A fentiek mind ún. *egyváltozós idősorok* voltak, hiszen skalárértékű valószínűségi változókat, ill. realizáltjaikat vizsgáltuk. Természetesen semmi akadálya annak, hogy ehelyett vektorértékű valószínűségi változókra térjünk át (pl. OTP árfolyam és HUF/EUR árfolyam *együttes* vizsgálata), ilyen *többváltozós idősorról* szokás beszélni. Ez egy még izgalmasabb, és persze bonyolultabb matematikai formalizmusú terület, hiszen ilyenkor nem csak a különböző időpontok közötti, hanem a különböző változók közötti kapcsolat kérdését is kezelni kell. Mi most egyváltozós idősorokkal fogunk foglalkozni.

1.2. Determinisztikus és sztochasztikus idősorelemzési megközelítések

Az idősorok elemzésének két alapvető megközelítése, módszertana alakult ki: a determisztikus (1.2.1. alpont) és a sztochasztikus (1.2.2. alpont) idősorelemzés.

Megjegyezzük, hogy valójában mindkettő az ún. *időtartományon történő elemzés* kategóriájába esik (mivel azon alapulnak, hogy az idősor külöböző időpontokhoz tartozó értékei között teremtenek kapcsolatot). Tágabban szemlélve, az idősorelemzés másik nagy kategóriája a *frekvenciatartományon történő elemzés*, ezzel azonban nem fogunk részleteiben foglalkozni. A frekvenciatartományon történő elemzés lényege (némiképp leegyszerűsítve), hogy az idősort Fourier-transzformációval szinuszhullámok összegére bontja. Belátható, hogy idősorok egy széles csoportja ekvivalensen reprezentálható úgy, hogy megadjuk, hogy az egyes frekvenciájú szinuszhullámokat milyen súllyal kell kombinálni, hogy megkapjuk az idősort. (Az ekvivalens reprezentáció alatt azt értjük, hogy oda-vissza át lehet térni a két leírás között.) Ez utóbbira szokták azt mondani, hogy időtartomány helyett frekvenciatartományon írtuk fel az idősort; ebből is sok hasznos és érdekes következtetést lehet levonni. Ezt a módszertant szokták spektrális elemzésnek is nevezni.

1.2.1. Determinisztikus idősorelemzés

Az ún. determinisztikus idősorelemzés azon a feltevésen nyugszik, hogy az idősort alakító tényezők, elvileg legalábbis, teljeskörűen számbavehetőek, és ez alapján az idősor alakulása, elvileg legalábbis, tökéletes pontossággal felírható lenne. A véletlennek „csak” annyiban jut szerep, hogy valami átkozott pech folytán a gyakorlati esetekben ez a teljeskörű felírás soha nem valósul meg (korlátozottak a mérési lehetőségek, korlátozott a tudásunk, hibával tudunk csak mérni stb.), emiatt a valóság mindig eltér a modell szerinti becslésünktől. (Vegyük észre, hogy ez az eltérés teljesen analóg a keresztmetszeti regresszió eltérésváltozójával, amiben szintén a fenti okok miatti hibát sűrítettük.) Azonban, és ez nagyon fontos, ebben a modellezési filozófiában a véletlen szerepe itt véget is ért: kialakítja az adott időszakbeli pontos értéket (eltérve valamennyivel a becslésünktől), ám ennyi, a későbbi időszakokra ennek már nincs hatása! (Természetesen a későbbi időpontokban is lesz eltérés, tehát ott is szerepet kap a véletlen, ám ez már az előző(ek)től függetlenül alakul – mintha minden időpillanatban pénzfeldobásszerűen döntenénk a tényleges érték becsülttől való eltérítéséről.)

Ez a filozófia a *dekompozíciós idősormodellek* felé mutat, melyek különböző, eltérő tartalmú komponensekre próbálják bontani az idősort (melyektől az idősor valamilyen függvényszerű módon függ). A legnépszerűbb modellben szokás beszélni pl. trendről (hosszú távú alapirányzat), ciklusról (éven túli ingadozás a trend körül) és szezonalitásról (évszakról-évszakra ingadozó, tehát éven belüli eltérés a trend és ciklus szerinti értéktől). Amennyiben feltesszük, hogy ezek additíve tevődnek össze, úgy az idősormodellünk a következőképp néz ki:

$$Y_t = R_t + C_t + S_t + u_t,$$

ahol R, C és S a trend, a ciklus és a szezonalitás rendre, u_t pedig a már említett eltérésváltozó.

Ezek a komponensek klasszikus statisztikai (jellegében deskriptív statisztikai) módszerekkel (pl. analitikus trendszámítás) becsülhetőek. A determinisztikus idősorelemzéssel (mellyel elsősorban hosszútávú előrejelzések adása a cél), a továbbiakban nem foglalkozunk.

1.2.2. Sztochasztikus idősorelemzés

A sztochasztikus idősorelemzés alapvető filozófiai eltérése, hogy bár ez a modell is adni fog egy becsült értéket az idősor adott időpontbeli értékére, és feltételezi, hogy a valós érték ettől véletlen módon eltér, ám abból indul ki, hogy ennek a véletlen eltérésnek *később* is hatása van: az idősor későbbi alakulását is befolyásolja. Úgy is szokták mondani, hogy az idősor fejlődésében öngeneráló hatások érvényesülnek: egy adott időpillanatbeli (véletlen) eltérés befolyásolja a későbbi értékeket is, tehát a véletlennek folyamatépítő szerepe van.

E megközelítést nagy sikerrel alkalmazták különböző közgazdasági (kiemelten: pénzügyi) idősorok modellezésére; elsősorban rövid távra.

1.3. Egy példa a két filozófiára: a determinisztikus és a sztochasztikus trend

Most megnézünk egy egyszerű példát, mely közvetlenül a két modellezési iskola feltevéseit szemlélteti, ám a későbbiek szempontjából is nagyon jól fog jönni.

Vegyük a következő két idősor-specifikációt:

$$\begin{aligned} Y_t^{(D)} &= \alpha t + u_t, \\ Y_t^{(S)} &= \alpha + Y_{t-1}^{(S)} + u_t, \quad Y_0 = 0. \end{aligned}$$

(Látható, hogy mindkét idősort a sokaságban[3] specifikáltuk.) Az egyszerűség kedvéért feltételezzük, hogy $u_t \sim \mathcal{N}\left(0, \sigma^2\right)$, mégpedig különböző t-kre függetlenül.

Első ránézésre több hasonlóság is felfedezhető a két specifikáció között. Mindkét idősor a 0-ból indul (a 0 időpillanatban) és érezhető, hogy mindkettőre igaz, hogy egy időszakkal később várhatóan α-val nagyobb értéket vesznek fel. (Az α természetesen lehet negatív is.) Hogy ezt az állítást precízebben is megfogalmazzuk, vegyük észre, hogy az $Y_t^{(S)}$ definíciójába rekurzíve behelyettesíthetünk (hiszen ha $Y_t^{(S)} = \alpha + Y_{t-1}^{(S)} + u_t$, akkor nyilván $Y_{t-1}^{(S)} = \alpha + Y_{t-2}^{(S)} + u_{t-1}$ stb.):

$$\begin{aligned} Y_t^{(S)} &= \alpha + Y_{t-1}^{(S)} + u_t = \alpha + \left(\alpha + Y_{t-2}^{(S)} + u_{t-1}\right) + u_t = \\ &= \alpha + \left[\alpha + \left(\alpha + Y_{t-3}^{(S)} + u_{t-2}\right) + u_{t-1}\right] + u_t = \ldots = \alpha t + \sum_{i=1}^{t} u_i. \end{aligned}$$

Az intuitív észrevételünket tehát úgy fogalmazhatjuk meg most már precízen, hogy $\mathbb{E}Y_t^{(D)} = \mathbb{E}\left(\alpha t + u_t\right) = \alpha t$ és $\mathbb{E}Y_t^{(S)} = \mathbb{E}\left(\alpha t + \sum_{i=1}^{t} u_i\right) = \alpha t$, tehát *várhatóértékben* minden időpillanatban ugyanaz a két idősor értéke. (Kihasználtuk, hogy összeg várhatóértéke a várhatóértékek összege, illetve, hogy konstans várhatóértéke saját maga és $\mathbb{E}u_i = 0$.)

[3]Pontosan innen látszik az is, hogy az $Y_0 = 0$ természetesen nem azt jelenti, hogy az Y_0 az a 0 (mint valós szám), hiszen az Y_t-k valószínűségi változók, ez tehát úgy értendő, hogy az Y_0 az az elfajult valószínűségi változó, mely 1 valószínűséggel a 0 értéket veszi fel.

Ez egy nagyon komoly megállapítás, ám van különbség is a két idősor fejlődése között. Ez rögtön világos lesz, ha felírjuk a szórásnégyzetet egy általános időpontra: $\mathbb{D}^2 Y_t^{(D)} = \mathbb{D}^2\left(\alpha t + u_t\right) = \sigma^2$ és $\mathbb{D}^2 Y_t^{(S)} = \mathbb{D}^2\left(\alpha t + \sum_{i=1}^t u_i\right) = t\sigma^2$. (Kihasználtuk, hogy konstans minden valószínűségi változótól független (és szórásnégyzete nulla), és hogy független valószínűségi változók összegének szórásnégyzete a szórásnégyzetek összege.) A két idősor tehát várhatóértékban ugyan azonos, ám $Y_t^{(S)}$ egyre nagyobb kilengésekkel ingadozik ezen várhatóérték körül, míg $Y_t^{(D)}$ állandóakkal.

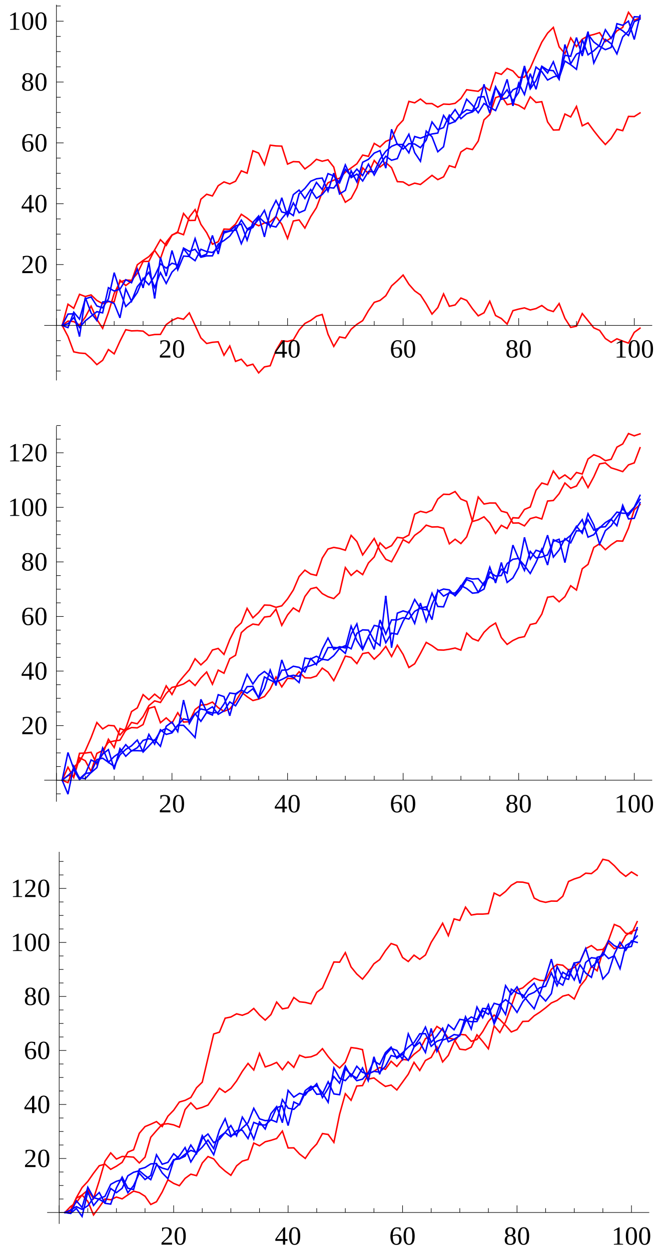
Megjegyezzük, hogy mivel nyilván mind $Y_t^{(D)}$, mind $Y_t^{(S)}$ normális eloszlású (hiszen a normális eloszláscsalád zárt a konvolúcióra és a konstanssal eltolásra), így a fenti két megállapítással teljesen le is írtuk az idősorokat. (Hiszen egy normális eloszlást egyértelműen meghatároz várhatóértéke és varianciája.)

A fentieket számítógépes szimulációval szemléltethetjük: véletlenszámgenerátorral előállítunk $u_t \sim \mathcal{N}\left(0, \sigma^2\right)$ számokat, majd így „lejátszhatjuk" egy lehetséges lefutását az idősornak. (Mind $Y_t^{(D)}$-t, mind $Y_t^{(S)}$-t szimulálhatjuk ilyen módon.) E megközelítés előnye, hogy míg valós szituációban nem ismerjük a sokasági eloszlást (épp ennek meghatározása lesz a feladat), csak egy realizáltat (sőt, idősorelemzésnél biztosan legfeljebb egy realizáltját láthatjuk az ismeretlen sokasági eloszlásnak), addig ebben az esetben ismerjük, sőt, mi határozzuk meg (a véletlenszámgenerátor beállításával) a sokasági eloszlást. Ennek megfelelően természetesen ebben az esetben vehetünk akárhány realizáltat a sokasági eloszlából (egyszerűen újra lefuttatjuk a szimulációt). Ezt szemlélteti az 1.2. ábra, melyen 9-9 realizáltat ábrázoltunk a $Y_t^{(D)}$ (kékkel) és $Y_t^{(S)}$ (pirossal) idősorokból $\alpha = 1, \sigma^2 = 9$ paraméterek mellett. (A jobb áttekinthetőség végett csak 3-3 realizáltat ábrázoltunk egy grafikonon.) Megjegyezzük, hogy ezeket a lefutásokat szokás trajektóriának is nevezni.

Az ábra tanulságosan igazolja vissza mindazt, amit eddig elméleti úton levezettünk. Egyrészt jól látszik, hogy várhatóértékben tényleg egyezik a két idősor (már ennyi ábrából is érezhető, hogy az $y = x$ egyenes körül ingadoznak); és az is tökéletesen látszik, hogy míg $Y_t^{(D)}$ állandó szórással ingadozik ezen egyenes körül, addig $Y_t^{(S)}$ egyre nagyobb szórással („legyezőszerűen" kitágul ezen egyenes körül).

Most már elárulhatjuk, hogy $Y_t^{(D)}$-t szokás determinisztikus trendnek, $Y_t^{(S)}$-et pedig sztochasztikus trendnek nevezni. (A „trend" elnevezés jogosságát épp a várhatóértékek alakulásáról tett megállapításunk indokolja.) Már a specifikációkból is látható, hogy a determinisztikus trend alakulása a determinisztikus idősorelemzési iskola premisszáit teljesíti, a sztochasztikus trend a sztochasztikus idősorelemzésiét.

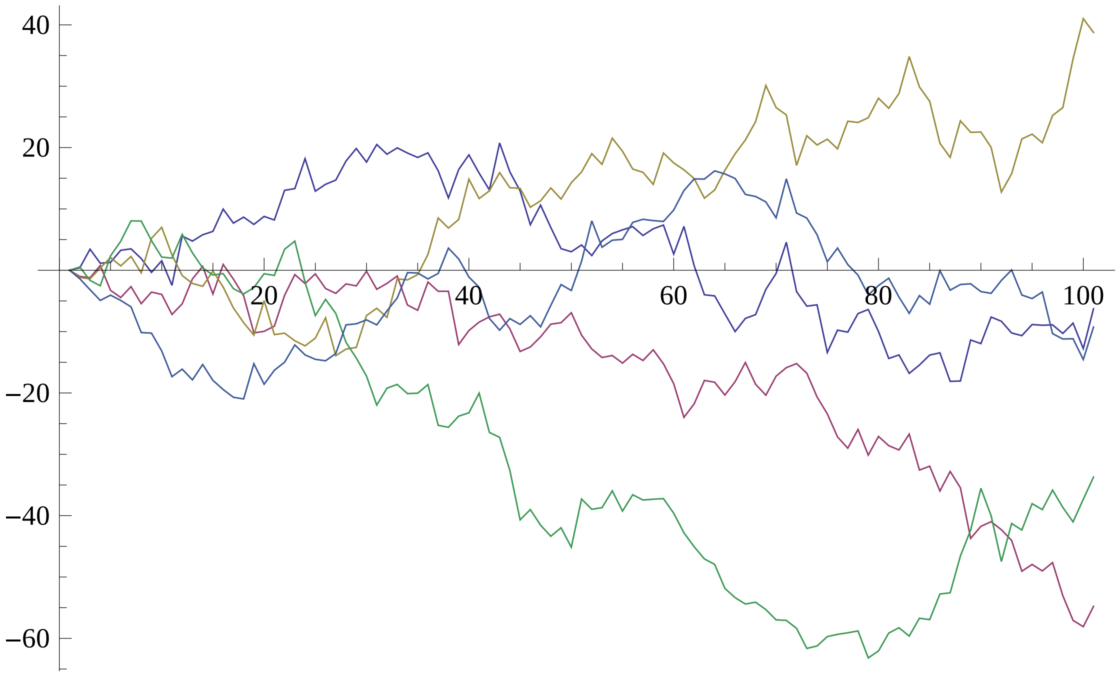
És ebből egy újabb fontos, tartalmi következtetést vonhatunk le az idősorok fejlődésének jellegzetességeire vonatkozóan. A determinisztikus trend alakulását úgy lehet elképzelni, hogy a t-edik időpontban fellépünk az αt pontba, majd egy $\mathcal{N}\left(0, \sigma^2\right)$ véletlenszám szerint perturbáljuk a pozíciónkat. A sztochasztikus trend esetén az előző pozícióból fellépünk α-t és utána térítjük el a pozíciónkat egy $\mathcal{N}\left(0, \sigma^2\right)$ véletlenszám szerint. (Mindezek a – most ismert – sokasági specifikációból világosak.) Hogy mi a különbség? Ha a véletlengenerátor pont egy nagyon nagy, vagy nagyon kicsi értéket dob ki, az ugyan kiugró pozíciót fog eredményezni, ám ennek determinisztikus trendnél semmilyen jelentősége nincs a későbbiek szempontjából (ettől függetlenül αt-be lépünk és generálunk újra véletlenszámot a következő időpontban), addig sztochasztikus trend esetén nagyon is van: az egész folyamat a kiugró pozíciótól folytatódik tovább! Ez legjobban talán a legalsó részábra legfelső trajektóriáján látszik: miután a 25. időpillanatban egy nagy pozitív véletlenszámot kaptunk, a nagy kiugrás nem egyszeri volt (ahogy determinisztikus trend esetén lett volna), hanem lényegében „eltolódott" az idősor, és onnan folytatódott az épülése. Mindez a specifikációból következik, természetesen. Azt is mondhatnánk, hogy a sztochasztikus trend esetén az idősorba „beépülnek a sokkok", míg a determinisztikus trendbe nem.



1.2. ábra. $Y_t^{(D)}$ (kék) és $Y_t^{(S)}$ (piros) szimulált lefutása 100 időegységre

Ez a példa jól szemlélteti, hogy mit kell az alatt érteni, hogy egy idősorban a véletlennek „folyamatépítő szerepe" van, hogy „öngeneráló" hatások érvényesülnek.

Megjegyezzük, hogy az $Y_t = Y_{t-1} + u_t$ specifikációt szokás *véletlen bolyongásnak* (RW, random walk) is nevezni, hiszen Y_t felfogható úgy is, mint egy objektum („a bolyongó") térbeli helyzete. Azaz: a bolyongó kezdetben az origóban áll, minden időpillanatban elővesz a véletlenszám-generátorát, és annyit lép felfelé, amennyit a véletlenszám-generátor mutat. (Ez természetesen negatív is lehet.) Ez tehát egy ún. egydimenziós bolyongás. Az $Y_t = \alpha + Y_{t-1} + u_t$ típusú folyamat neve *eltolásos véletlen bolyongás* (RWD, random walk with drift), hiszen ilyenkor a bolyongó először determinisztikusan felfelé lép α-t (eltolás, „sodródás") és utána nézi meg a véletlenszám-generátorát. RWD-folyamatokra a fentiekben már láttunk példákat, az 1.3. ábrán pedig 5 szimulált RW-folyamat trajektóriáját láthatjuk. Érdemes ellenőrizni a megbeszélt tulajdonságok teljesülését! Ezen folyamatoknak a valószínűségszámításban van nagy jelentőségük.



1.3. ábra. Véletlen bolyongás (RW) szimulált trajektóriái

2. fejezet

Stacionaritás

A stacionaritás az idősorelemzés egyik alapvető fogalma; lényegében egy megkötést jelent az idősor valószínűségi struktúrájára nézve. E megkötés azért szükséges, hogy az idősor statisztikai eszközökkel kézbentartható legyen. (És mert – épp emiatt – a későbbi ismertetendő módszertan is stacionárius idősorokat fog igényelni.)

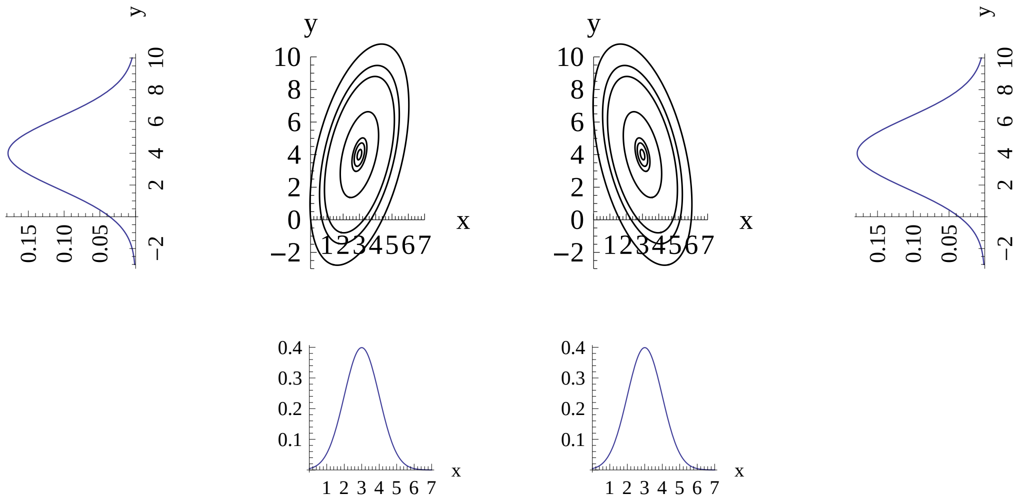
A 2.1. pontban megindokoljuk, hogy miért szükséges ez a fogalom, mi a bevezetésének a logikája. Ezt követően, a 2.2. pontban precízen is bevezetjük a két, gyakran használt stacionaritás fogalmat, majd a 2.3. pontban megmutatjuk pár nevezetes jellemzőjét a stacionárius idősoroknak. Ez már csak azért is fontos, mert a későbbiek szinte kizárólag ilyen idősorokkal fogunk dolgozni.

A 2.4. pontban megmutatjuk, hogy az idősorok stacionaritását hogyan tudjuk megvizsgálni. Ez azért különösen fontos, mert a későbbiekben, amint már utaltunk is rá, stacioner idősorokra lesz szükségünk a modellépítéshez. Emiatt itt tárgyaljuk azt a másik nagyon fontos kérdést is, hogy mi a teendő nem-stacioner idősorok esetén, hogyan tudjuk őket stacioner idősorrá transzformálni („stacionarizálás").

2.1. A stacionaritás fogalmának szükségessége

Hasonlóan a keresztmetszeti adatelemzéshez, idősoros esetben is az lehet a célunk, hogy a minta alapján rekonstruáljuk az ismeretlen háttéreloszlást, amiből a minta származik. Az idősor teljes leírását az adja, ha megadjuk a komponenseinek, tehát az egyes időponthoz tartozó valószínűségi változóknak az együttes eloszlását.

Jegyezzük meg, hogy az egyes időpontok önmagában vett eloszlása nyilván kevés, hiszen ezekből semmit nem tudunk az időpontok közötti kapcsolatokról. Gondoljunk akár csak a legegyszerűbb esetre: $T = 2$ és az idősor eloszlása kétdimenziós normális. A 2.1. ábra két ilyen esetet szemléltet, a kétdimenziós sűrűségfüggvényt szintvonalakkal megjelenítve. A két eset jellegzetessége, hogy mindkét vetületi eloszlásuk (tehát: mindkét időpontban az adott időpontra önmagában vett eloszlás) *pontosan* ugyanaz (ezt szemléltetendő az ábrán szintén feltüntettük ezeket a vetületi eloszlásokat, mégpedig pont úgy, ahogy a kétdimenziós eloszlást „le kéne vetíteni'), mégis a két idősor tartalma drámaian eltérő: nagyon nem mindegy, hogy ha a részvény adott napi árfolyama az átlagánál magasabb, akkor a következő napi várhatóan átlagánál alacsonyabb, vagy pont hogy magasabb lesz... Kicsit precízebb valószínűségszámítási terminológiával megfogalmazva: a két eloszlás várhatérték-vektora és szórásnégyzetei teljesen azonosak, ami eltér, az a kovariancia (és így persze a korreláció is). Az is világos mellesleg valószínűségszámításból, hogy kizárólag akkor mondhatjuk, hogy a vetületi eloszlások elégségesek, ha az egyes időpontok függetlenek (hiszen ekkor az együttes sűrűségfüggvény[1] előállítható a vetületi sűrűségfüggvényekből, egyszerű szorzással). Ez nyilván irreális feltevés a legtöbb gyakorlati esetben. Minden információt általánosságban csak az együttes eloszlás hordoz.



2.1. ábra. Két kétdimenziós normális eloszlás sűrűségfüggvénye (szintvonalakkal) és vetületi eloszlásaik (szemléletesen ott, ahová tényleg „vetíteni" lehetne a kétdimenziós felületet); úgy, hogy az eloszlások mindkét vetületi eloszlása pontosan egyezik, mégis drámaian eltérőek tartalmilag

Az 1. fejezetben mondottak szerint azonban az együttes eloszlás meghatározására – szemben a keresztmetszeti esettel – semmilyen reményünk nincs: míg 1406 minta alapján a 7 változónk együttes eloszlása jól rekonstruálható, addig itt, egyetlen minta lévén, lényegében semmit nem tudunk mondani az együttes eloszlásról.

Sőt, nem csak az együttes eloszlásról nem tudunk érdemben nyilatkozni, de még a vetületi eloszlásokról (itt: az idősornak az egyes időpontokban önmagában vett eloszlásairól) sem: egyetlen realizációból értelmesen még várhatóértéket sem tudunk mondani, nemhogy eloszlást rekonstruálni.

Ebből tehát világos, hogy az idősorelemzéssel csak úgy tudunk érdemben továbbhaladni, ha az együttes eloszlásra bizonyos megkötéseket teszünk. Ez fog elvezetni minket a stacionaritás fogalmához.

Először egy rávezető problémával kezdünk. Tegyük fel, hogy valaki megkér minket, hogy adjuk meg a 2010. január 4.-i OTP záróárfolyam várhatóértékét. (Ez az Y_1 az 1.1. ábra példáján.) Ahogy már mondtuk, további megkötés nélkül ez a feladat reménytelen, hiszen arra a valószínűségi változóra egyetlen mintánk van ($y_1 = 5492$), amiből, értelmesen legalábbis, nem lehet várhatóértéket becsülni. (Mint ahogy semmi mást sem.)

Tegyük fel azonban, hogy valaki „megsúgja", hogy az együttes eloszlás olyan, hogy minden vetületi eloszlásnak (tehát minden nap záróárának) *ugyanaz* a várhatóértéke. Ekkor már drasztikusan más a helyzet! Ha ugyanis ez a feltevés igaz, akkor természetesen a második (vagy bármelyik más) időpontbeli realizáció is ugyanúgy használható a várhatóérték becslésére, mint az első napi; azaz: a különböző naphoz tartozó értékeket „összeönthetjük" a várhatóérték becslé-

[1] Mi most csak olyan idősorokkal foglalkozunk, melyeknél a valószínűségi változó eloszlása folytonos, és így létezik sűrűségfüggvény.

séhez. Márpedig 254 értékből nagyon is lehet várhatóértéket becsülni! E feltevés híján azonban a külöböző időpontbeli értékeket nem használhattuk volna fel együtt.

Ez a példa rámutat arra, hogy ha bizonyos megszorításokat teszünk (a példában: hogy minden vetületi eloszlás várhatóértéke ugyanaz), akkor a kezdetben reménytelen feladatot kezelhetővé tesszük (legalábbis bizonyos szempontok szerint). E kikötések teljesülését persze valahogy ellenőrizni kell, de erre majd később, a 2.4. pontban térünk vissza.

2.2. Erős és gyenge stacionaritás

Most bevezetjük, a fentiek által motiválva, idősorok stacionaritásának fogalmát. (Jobban mondva fogalmait, mert több stacionaritási fogalmat is definiálni fogunk.) Előtte még emlékeztetünk arra, hogy egy többdimenziós eloszlás vetületének néhány kiválasztott komponensének együttes eloszlását nevezzük. Például az X, Y, Z, V valószínűségi változókból (mint komponensekből) álló többdimenziós (négydimenziós) valószínűségi változónak 4 darab egydimenziós vetületi eloszlása van (az X, az Y, a Z és a V változók (önmagában vett) eloszlása), 6 darab kétdimenziós vetületi eloszlása van (az X, Y, X, Z, X, V, Y, Z, Y, V és Z, V párok együttes eloszlása) és így tovább. (Természetesen, mint azt ez a példa is mutatja, a vetületi eloszlás is lehet többdimenziós, azaz egy együttes eloszlás vetületi eloszlása is jelenthet együttes eloszlást.)

2.2.1. Erős stacionaritás

Kezdjük egy nagyon erős megkötéssel. Egy idősort *erős értelemben stacionáriusnak* (vagy: erős értelemben stacionernek) nevezünk, ha *minden* véges dimenziós vetületének együttes eloszlása (tehát: *akárhány* elemű vetületről van szó, és ezeket az elemeket *akárhogy* választjuk ki az idősor komponensei (azaz időpontjai) közül) eltolásinvariáns, *minden* értelmes eltolásra. Például mondjuk azt, hogy háromdimenziós vetületek együttes eloszlására vagyunk kíváncsiak; egy ilyen lehetséges vetület például az Y_1, Y_3 és Y_7 együttes eloszlása. Ha az erős stacionaritás fennáll, akkor e három együttes eloszlásának ugyanannak kell lennie, mint Y_2, Y_4 és Y_8 együttes eloszlásának, vagy épp Y_{12}, Y_{14} és Y_{18} együttes eloszlásának, vagy épp Y_{212}, Y_{214} és Y_{218} együttes eloszlásának, és egyáltalán: minden más, eltolással kijelölt időpont együttes eloszlásának. (Eltolással kijelölés itt olyan, mintha egy ablakot mereven végigtolnánk az idősoron, tehát az egyes időpontok közti különbségeknek ugyanannyinak kell lenniük.) És ennek természetesen nem csak a háromdimenziós vetületekre kell teljesülnie, hanem az egydimenziósokra, a kétdimenziósokra, ... és az $n-1$ dimenziósokra is. (Azért kellett úgy fogalmaznunk, hogy minden „értelmes” eltolásra, mert nem véges idősoroknál ezt az ablakot nyilván nem tolhatjuk ki az idősoron „túlra”.)

Precízen megfogalmazva: $\forall k \geq 1$ esetén $\forall t_1, t_2, \ldots, t_k$-ra $Y_{t_1}, Y_{t_2}, \ldots, Y_{t_k}$ együttes eloszlása megegyezik $Y_{t_1+h}, Y_{t_2+h}, \ldots, Y_{t_k+h}$ együttes eloszlásával, $\forall h$-ra (ha az itt szereplő komponensek mind az idősor részei).

Ez a feltétel (már a $\forall$ kvantorok számából is érezhetően...) rendkívül sokat követel meg. Vegyük észre például, hogy ennek része az előző pont végének példa-megkötése: a definíciót $k = 1$-gyel alkalmazva azt kapjuk, hogy erős stacionaritás esetén minden egyes időpontban *pontosan ugyanannak* kell lennie az idősor adott időpontbeli (vetületi) eloszlásának. (Így nyilván a várhatóértékének, és egyáltalán, minden momentumának is ugyanannak kell lennie, minden időpontban.)

Ennek megfelelően, ha egy idősor erősen stacioner, az rendkívüli módon megkönnyíti az elemzését. Nem csak a várhatóértékének becsléséhez használható fel együtt az összes időpontbeli érték (tehát $y_1, y_2, \ldots, y_T$), de a szórásnégyzetének, ferdeségének stb., tehát általában, bármilyen momentumának becsléséhez, sőt: a definíció $k = 2$-re alkalmazásából látszik, hogy az egymást követő időpontok kétdimenziós eloszlása is ugyanaz, tehát minden időpont és a rákövetkező időpont közti korreláció is ugyanaz kell legyen (függetlenül attól, hogy melyik ez a két időpont, csak az számít, hogy egymás utániak legyenek), és ennek becsléséhez felhasználhatjuk az (y_1, y_2), (y_2, y_3), $\ldots$, (y_{T-1}, y_T) értékeket. Sőt, a kettő különbségű időpontok eloszlása is azonos, függetlenül attól, hogy melyek a konkrét időpontok (csak az számít, hogy időkülönbségük kettő legyen), így korrelációjuk is azonos, és emiatt e korreláció (szokták úgy is hívni: a kettő késleltetéshez tartozó korreláció) becsléséhez felhasználhatóak az (y_1, y_3), (y_2, y_4),$\ldots$,(y_{T-2}, y_T) párok értékei, és így tovább. (Tehát például minden háromdimenziós (stb.) vetületi eloszlás is azonos lesz, eltolástól függetlenül, csak ezek nem bírnak olyan nagy gyakorlati jelentőséggel, nincs is olyan közismert leírójuk, mint kétdimenziósoknál a korreláció.)

2.2.2. Gyenge stacionaritás

Az látható, hogy a felvázolt problémát az erős stacionaritás elfogadása megoldja, csak épp ezzel bizonyos értelemben „átesünk a ló túloldalára": az erős stacionaritás olyan komplex követelményrendszer, hogy ellenőrzése lényegében reménytelen mintából.

Éppen emiatt, a gyakorlati alkalmazásokban ehelyett inkább egy gyengített változatát szokás használni. A gyengítés motivációja, hogy csak azokat a követelményeket hagyjuk meg az erős stacionaritásból, amelyek kézzelfogható statisztikai jellemzőkhöz kapcsolódnak: például a kétdimenziós eloszlásoknak van gyakorlati jelentőségük (fontos leírójuk kovariancia/korreláció), ám a három (és több) dimenziós eloszlásoknak nincs ilyen jellemzőjük, ezért kettőnél nagyobb dimenziójú vetületekre egyáltalán nem teszünk kikötést. Sőt, az egy- és kétdimenziós eloszlásoknál is enyhítünk a feltételeken: nem az eloszlások teljes egyezőségét követeljük meg, csak az első és második momentumban történő egyezést. (Kétdimenziós esetben az első momentumnak nincs értelme, a második momentum pedig a kovariancia lesz.)

Mindezeket összefoglalva, és precízzé téve: egy idősort *gyenge értelemben stacionáriusnak* (vagy gyenge értelemben stacionernek) nevezünk, ha a következő három feltétel teljesül rá:

1. Minden időpontban ugyanaz az idősor várhatóértéke, tehát létezik a közös $m \equiv \mathbb{E}Y_i$ várhatóérték.
2. Minden időpontban ugyanaz az idősor szórásnégyzete, tehát létezik a közös $\sigma^2 \equiv \mathbb{D}^2 Y_i$ szórásnégyzet.
3. Két időpont közti kovariancia kizárólag a két időpont külöbségétől (a késleltetéstől) függ, tehát $\mathrm{cov}\left(Y_i, Y_{i+k}\right) = \mathbb{E}\left[\left(Y_i - \mathbb{E}Y_i\right)\left(Y_{i+k} - \mathbb{E}Y_{i+k}\right)\right] \equiv \gamma_k$, minden i-re.

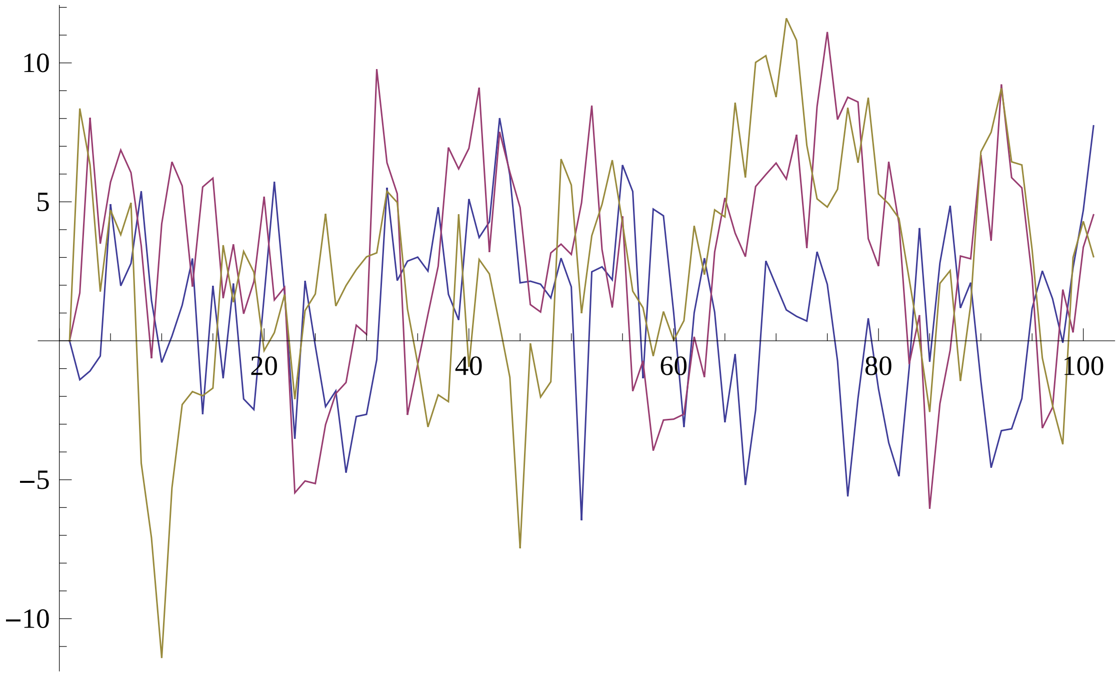
Azonnal látható, hogy egy-az-egyben az erős stacionaritás követelményeit ismételtük meg, csak épp mindössze egy- és kétdimenziós vetületekre, és mindössze első két momentumban történő egyezésre. (Szokás a gyenge stacionaritást kovariancia-stacionaritásnak is nevezni.)

Észrevehető mellesleg, hogy ez a gyengítés pontosan összhangban van azzal, amit akkor tennénk, ha tudnánk, hogy az idősor többdimenziós normális eloszlású (és így persze minden vetületi eloszlása is (többdimenziós) normális). Ekkor ugyanis egyrészt az első két momentum teljesen meghatározza az eloszlást, másrészt pl. négy időpont együttes eloszlása teljesen determinált, ha ismerjük a belőlük kiválasztható összes pár eloszlását. (Hiszen az egyes időpontok várhatóértékén túl csak a kovarianciamátrixra van szükségünk (e kettő teljeskörűen leír egy többdimenziós normális eloszlást), márpedig az páronként számolható.) Magyarán: többdimenziós normális eloszlásnál semmilyen „pluszt" nem jelent a kettőnél több elemű vetületek, illetve a kettőnél nagyobb momentumok ismerete.

A fentiekből tehát világos, hogy az erős stacionaritás fogalma valóban erősebb: az erősebb stacionaritás implikálja a gyengét. A fordított irány általában nem áll fenn (tehát a két fogalom nem ekvivalens), de az előző bekezdés fényében világos, hogy speciálisan többdimenziós normális eloszlású idősorra (ún. Gauss-folyamat) igen, ott tehát e két fogalom egybeesik.

A továbbiakban, ha nem mondunk mást, stacionaritás alatt gyenge stacionaritást fogunk érteni.

A 2.2. ábra példát mutat egy stacioner idősorra: adott, biztosan stacioner sokasági specifikációból[2] szimulációval generáltunk három trajektóriát. Érdemes megfigyelni a stacionaritási követelmények „ránézésre" történő teljesülését: nem úgy tűnik, mintha a trajektóriáknak lenne trendjük és szintén nem tűnik úgy, mintha a szórás változna. (Az autokorrelációk időfüggését nyilván kevésbé lehet szabad szemmel megítélni.)



2.2. ábra. Példa stacionárius idősorra: három szimulált trajektória az $Y_t = 0{,}5 + 0{,}7 \cdot Y_{t-1} + u_t, u_t \sim \mathcal{N}(0, 3)$, beláthatóan stacionárius specifikációból

2.3. Stacionárius idősorok nevezetes jellemzői

Gyenge stacionaritás fennállása esetén az egyes megkötések logikusan diktálnak bizonyos jellemzőket az idősorra vonatkozóan. A következőkben ezeket fogjuk sorra venni.

Ha egy idősor stacioner, úgy az előzőek alapján beszélhetünk az egyes komponenseinek közös $m \equiv \mathbb{E}Y_i$ várhatóértékéről. Ennek becslése minta alapján nyilván:

$$\widehat{m} = \frac{1}{T} \sum_{i=1}^{T} y_i.$$

[2]Ez a specifikáció egy ún. AR(1) folyamat, melyet később fogunk tárgyalni, itt most csak annyi fontos, hogy bizonyosan stacioner. Konkrét specifikációja: $Y_t = 0{,}5 + 0{,}7 \cdot Y_{t-1} + u_t, u_t \sim \mathcal{N}(0, 3)$.

Ha egy idősor stacioner, úgy az előzőek alapján beszélhetünk az egyes komponenseinek közös $\sigma^2 \equiv \mathbb{D}^2 Y_i$ szórásnégyzetéről. Ennek becslése minta alapján nyilván:

$$\widehat{\sigma^2} = \frac{1}{T} \sum_{i=1}^{T} (y_i - \widehat{m}) .$$

(Itt már kihasználtuk, hogy a gyenge stacionaritás miatt létezik közös várhatóérték.)

Sokkal izgalmasabb a korreláció kérdése. Először is jegyezzük meg, hogy a különböző időpontok (mint az idősor egyes komponens valószínűségi változói) közötti korrelációt *autokorrelációnak* szokás nevezni (az „auto" utal arra, hogy – az idősor szintjén – „önmagával vett" korrelációról van szó). A gyenge stacionaritás tehát azt köti ki, hogy az Y_t és Y_k közötti autokorreláció kizárólag a $k - t$ különbségtől (a késleltetéstől) függ, t-től nem. Kézenfekvő akkor, hogy erre a késleltetéstől függő mennyiségre külön elnevezést vezessünk be.

Egy idősor *autokovariancia-függvényének* nevezzük a $\gamma_k = \operatorname{cov}(Y_1, Y_{1+k})$ kifejezést, mely láthatóan k függvénye. Megismételjük, hogy ez a definíció gyenge stacionaritás esetén jogos, hiszen $\operatorname{cov}(Y_1, Y_{1+k}) = \operatorname{cov}(Y_2, Y_{2+k}) = \ldots = \operatorname{cov}(Y_{T-k}, Y_T)$, azaz általában $\operatorname{cov}(Y_1, Y_{1+k}) = \operatorname{cov}(Y_{1+h}, Y_{1+h+k})$ minden értelmes h eltolásra. Az autokovariancia függvény becslése mintából:

$$\widehat{\gamma_k} = \frac{1}{T-k} \sum_{i=1}^{T-k} (y_i - \widehat{m}) (y_{i+k} - \widehat{m}) .$$

(A gyenge stacionaritás feltételezése miatt természetesen támaszkodhattunk arra, hogy létezik közös várhatóérték.)

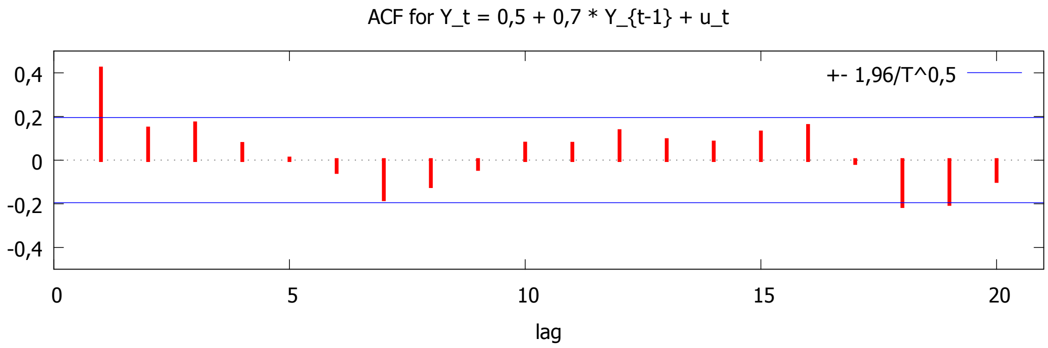
Az autokovarianciának ugyanaz a baja, mint keresztmetszeti esetnél a kovarianciának: a számértéke önmagában keveset mond. Emiatt, szintén a keresztmetszeti esethez hasonlóan, be szokás vezetni az autokorreláció fogalmát, ami gyenge stacionaritás esetén nyilván $r_k = \frac{\gamma_k}{\sigma^2}$, mintából becsült értéke előállítható az eddigiek felhasználásával:

$$\widehat{r_k} = \frac{\widehat{\gamma_k}}{\widehat{\sigma^2}}.$$

Ezt szokás *autokorreláció-függvénynek* (ACF, autocorrelation function) nevezni. (Emiatt néha r_k helyett az AC_k megnevezést is használják.)

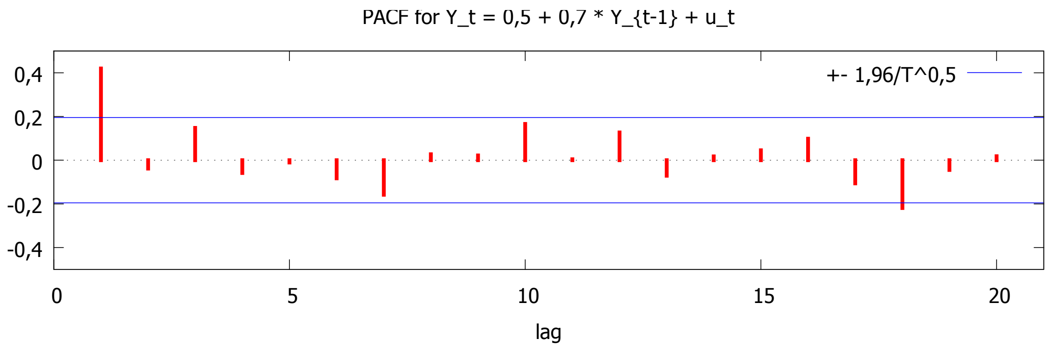
Az ACF kézenfekvően ábrázolható grafikusan, ha a vízszintes tengelyen k lehetséges értékeit, a függőleges tengelyen pedig az e késleltetésekhez tartozó autokorrelációt ábrázoljuk. ($r_0 = 1$ nyilván, ezért ezt nem szokás külön feltüntetni.) Ezt az ábrát, mely természetesen diszkrét függvény lesz, és így oszlopdiagrammal vagy hasonló módon jeleníthető meg, *korrelogramnak* szokás nevezni. A 2.3. ábra a 2.2. ábra kék színnel jelölt idősorának korrelogramját mutatja.

Szintén szoktak beszélni egy idősor *parciális autokorrelációs függvényéről* (PACF, partial autocorrelation function). Emlékezhetünk rá, hogy a parciális autokorrelációt nem két változó, hanem két változó és változók egy halmaza között értelmezzük; tartalma: a két változó közti kapcsolat erőssége és iránya *ha* a köztük a megadott változókon keresztül terjedő hatásokat kiszűrjük. (Ezúttal sem fontos, hogy ezt pontosan hogyan valósíthatjuk meg, mindenesetre megvalósíthatjuk. Itt természetesen csak lineáris hatásokról beszélünk.) A PACF-függvényt is két időpont között számoljuk, a kérdés már csak az, hogy mik a kiszűrt változók. A válasz kézenfekvő: a két időpont *közötti* időpontok (mint valószínűségi változók). Ennek megfelelően a 0 késleltetéshez tartozó PACF szükségképp 0, az 1 késleltetéshez tartozó PACF pedig r_1 (hiszen ilyenkor „nincs mit" kiszűrni). A PACF az ACF-hez hasonlóan becsülhető mintából, ezzel most nem foglalkozunk.



2.3. ábra. A 2.2. ábra kék színű idősorának korrelogramja (ACF-függvénye)

A 2.4. ábra mutat egy parciális autokorrelációs (PACF) függvényt, ezúttal is a 2.2. ábra kék idősorának példáján. (Tehát ez, és az előző korrelogram összevethető, olyan értelemben, hogy ugyanahhoz az idősorhoz tartoznak.) Érdemes ellenőrizni előbbi megjegyzésünket arról, hogy az 1 késleltetéshez tartozó ACF és PACF megegyezik.



2.4. ábra. A 2.2. ábra kék színű idősorának parciális autokorrelációs (PACF) függvénye

2.4. A stacionaritás ellenőrzése, stacionarizálás, trend- és differenciastacionárius idősorok

A következő feladat