Az idősorelemzés alapjai

Ferenci Tamás
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 9.

Tartalom

Tartalomjegyzék

1. Az idősor fogalma, jelentősége, idősorelemzés

1.1. Az idősor fogalma

Idősor sokaságban és mintában I.

- Mi a sokaság és a minta? – ismétlés stat 2-ből
- („Sokaságban valszám kell, mintánál statisztika")
- **Idősor** minta értelemben: időben rendezett megfigyelések

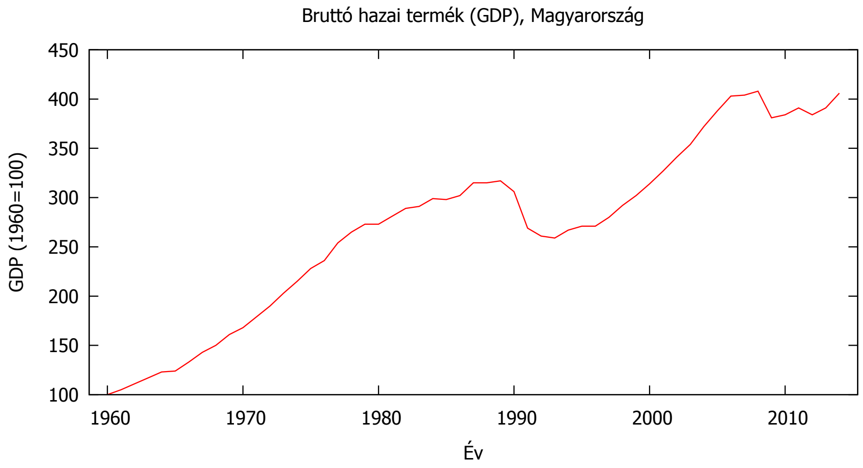
  $$\{5492, 5640, \ldots, 7317\},$$

  általános jelöléssel

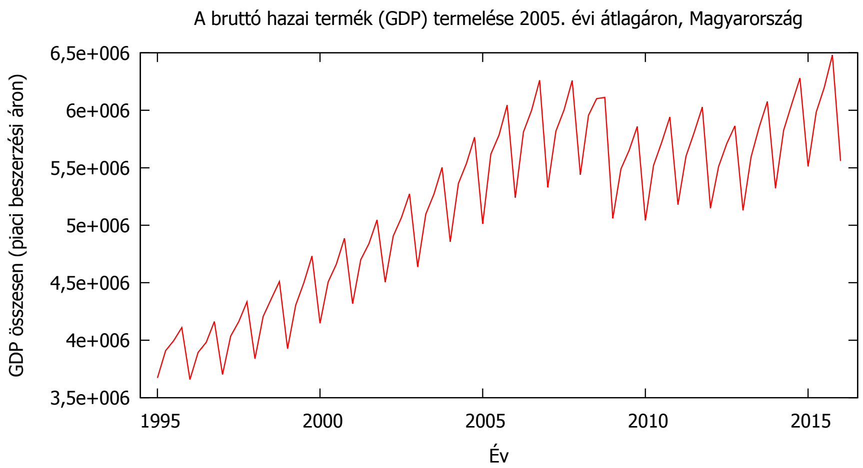
  $$\{y_1, y_2, \ldots, y_T\}$$

- Az egyetlen eltérés a keresztmetszethez képest: *sorrendjük* van (hatalmas jelentősége lesz majd)

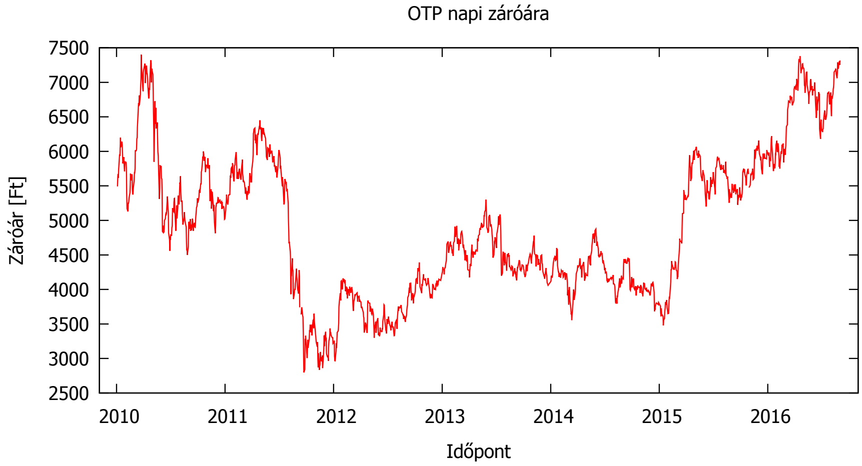
Példa: magyar GDP (hosszú éves)



Példa: magyar GDP (negyedéves)



Példa: OTP napi záróár



Idősor sokaságban és mintában II.

- Sokasági definíció: valószínűségi változók indexelt halmaza (ahol az indexet „idő”-nek hívjuk):
$$\{Y_t, t \in N\}$$
- Az idősor egy időpontban tehát egy valószínűségi változó
- Itt az N tehát egy *rendezett* halmaz
- Valszámos szó: **sztochasztikus folyamat**
- Közgázban jellemzően N diszkrét, sőt véges: $N = \{1, 2, \ldots, T\}$
- A minta tehát ennek egy realizációja (természetesen), itt a neve: **trajektória**

Mivel a valószínűségi változó pedig egy függvény ($X : \omega \to \mathbb{R}$), ezért a sztochasztikus folyamat igazából egy *kétváltozós* függvény: $Y : \omega \times N \to \mathbb{R}$. Ennek megfelelően két metszete van – úgy értve, hogy melyik paramétert rögzítjük – az egyik a trajektória, a másik egy adott időpillanathoz tartozó valószínűségi változó.

Egyetlen realizáció (reprodukálhatatlanság) problémája I.

- Mi a különbség 1406 lakás ára és az OTP 1406 napi záróárfolyama között?
- Mi van emögött?
- A függetlenség feltevés (tarthatósága)!

Egyetlen realizáció (reprodukálhatatlanság) problémája II.

- Elég, ha ismerjük az *egyes időpontok* eloszlásait? (F_{Y_1}, F_{Y_2} stb.)
- Ami teljes mértékben leírja az idősort: az *összes* időpont *együttes* eloszlása:

$$F_{Y_1,Y_2,\ldots,Y_T}$$

- (Igazából keresztmetszetnél is ez volt, csak a függetlenségi feltevés miatt esett szét egyváltozósokra)

Egyetlen realizáció (reprodukálhatatlanság) problémája III.

- Azaz: 1 dimenziós adatra vett 100 megfigyelés vs. 100 dimenziósra vett 1 megfigyelés!
- (Ezért nem segít az sem, ha hosszabbítjuk az idősorunkat!)
- Hogyan lehet egyáltalán így bármi becsülni…?
- 1 megfigyelésből? Az érdekes lesz… → további feltevésekre lesz majd szükség!

1.2. Az idősorelemzés közgazdasági jelentősége, története

Az idősorelemzés közgazdasági jelentősége

- A legtöbb közgazdasági adat *igazából* idősorként érhető el!
- Számos feladatnál időbeli a fókusz, gondoljunk a – szó szoros értelmében vett – előrejelzési kérdésekre

Idősorelemzés a közgazdaságtanban

- Eleinte: egyszerű determinisztikus módszerek (pl. dekompozíciós modellek már a XIX. században)
- Később regresszió is, de még tekintet nélkül az idősoros jellegre
- Cochrane és Orcutt mutatott rá először 1949-ben, hogy ez nem jó ötlet
- Megindult a kutatás ennek figyelembevételére
- Box és Jenkins könyve 1970-ben fordulópont: sztochasztikus módszerek
- Korszerű eljárások és aktuális kérdések (nemstacionaritás, nemlinearitás, többváltozós módszerek, ARCH, …)

1.3. Idősorelemzési iskolák, a módszerek felosztása

Az idősorelemzési módszerek csoportosítása

- Időtartomány vs. frekvenciatartomány (csak az előbbivel fogunk most foglalkozni)
- Determinisztikus vs. sztochasztikus (definíciós kavarodások, most: a véletlennek van-e *folyamatépítő* szerepe; ld. később részletesen)
- Egyváltozós vs. többváltozós (+panel)

2. Idősorok jellemzői a sokaságban

Várhatóérték- és szórásnégyzet-függvény

Emlékezzünk rá, hogy egy adott időpontban az idősor egyszerűen egy valószínűségi változó, így definiálható várható értéke, szórásnégyzete, két ilyennek a kovarianciája; ez alapján:

- Várhatóérték-függvény ($\mu : N \to \mathbb{R}$):

$$\mu_t := \mathbb{E}Y_t$$

- Szórásnégyzet-függvény ($\sigma^2 : N \to \mathbb{R}_+$):

$$\sigma_t^2 := \mathbb{D}^2 Y_t = \mathbb{E}(Y_t - \mathbb{E}Y_t)^2 = \mathbb{E}(Y_t - \mu_t)^2 = \mathbb{E}Y_t^2 - \mu_t^2$$

Autokovariancia- és autokorreláció-függvény

- Autokovariancia-függvény (ACVF, $\gamma : N \times N \to \mathbb{R}$):

$$\gamma_{t,s} := \operatorname{cov}(Y_t, Y_s) = \mathbb{E}\left[(Y_t - \mathbb{E}Y_t)(Y_s - \mathbb{E}Y_s)\right] = \\ \mathbb{E}\left[(Y_t - \mu_t)(Y_s - \mu_s)\right] = \mathbb{E}(Y_t Y_s) - \mu_t \mu_s$$

- Nyilván $\gamma_{t,t} = \sigma_t^2$
- Autokorrelációs függvény (ACF, $\rho : N \times N \to [-1, 1]$):

$$\rho_{t,s} := \operatorname{corr}(Y_t, Y_s) = \frac{\gamma_{t,s}}{\sigma_t \sigma_s}$$

Ne feledjük: mindezekben semmi sztochasztikus nincs, teljesen közönséges – determinisztikus – függvények!

Parciális autokorrelációs függvény (PACF)

- Úgy viszonyul az ACF-hez mint a sima (keresztmetszeti) parciális korreláció a korrelációhoz: bizonyos változókon keresztül terjedő hatásokat szűrjük (lineárisan)
- De melyikeket?
- Ami közbeesik: t és s közti korreláció ($s > t$), szűrve a $t+1, t+2, \ldots, s-2, s-1$ időpontokon keresztül terjedő hatásokat
- Kiszámítható az ACF-ek ismeretében egyszerű mátrixműveletekkel
- (Avagy: a korreláció kijön egy olyan regresszióból, aminek egyetlen magyarázó változója van, a parciális korrelációhoz pedig hozzá kell adni a közbenső időpontokat is)

Korrelogram

- Korrelogram: ACF és PACF együtt ábrázolva
- (Egyelőre úgy tűnik, hogy ez egy kétdimenziós függvény, ezt majd később árnyalni fogjuk – és ezért nem is ábrázoljuk most még ténylegesen)
- Jelentősége: ha ismerjük nevezetes folyamatok – elméleti – korrelogramját, akkor egy minta mögötti, azt adó folyamatra következtethetünk az alapján, hogy a minta – empirikus – korrelogramja hogyan néz ki
- (Sajnos a gyakorlatban sokszor csak hozzávetőleges lehet)