

# Beadandó feladatok

Idősorelemzés gyakorlat  
2023/2024. tavaszi félév

---

## 1. beadandó

### Technikai tudnivalók

Nézd végig a honlapomon található **R** bevezetőt és az alapján oldd meg **R**-es **szimuláció** segítségével a következő feladatot! A futtatásokhoz a forráskódot **R** nyelven kell megírni. Az értelmezéseket, válaszokat, illetve mindent, amit még szükségesnek érzel közölni velem, külön szövegfájlba írd vagy a forráskódba kommentként!

**Beadási határidő: február 26., 18:00**

Késve beadott megoldás nulla pontot ér.

---

### Feladat [5 pont]

Egy magyarkártya-csomagból visszatevéssel húzunk 7 lapot. Szimulációval számold ki, hogy milyen eséllyel húzunk pontosan 4 tők színű lapot! Legalább mennyi ismétlésszámot ajánlanál, hogy a valódi valószínűséget legalább 1%-os pontossággal közelítsük? Válaszodat a konvergencia sebességét bemutató alkalmas ábrával támaszd alá!

---

## 2. beadandó

### Technikai tudnivalók

Nézd végig a honlapomon található **R** bevezetőt és az alapján oldd meg **R** segítségével a következő feladatot! A futtatásokhoz a forráskódot **R** nyelven kell megírni. Az értelmezéseket, válaszokat, illetve mindent, amit még szükségesnek érzel közölni velem, külön szövegfájlba írd vagy a forráskódba kommentként!

**Beadási határidő: február 26., 18:00**

Késve beadott megoldás nulla pontot ér.

---

### Feladat [5 pont]

Szimulálj  $N = 10000$  darab  $n = 1000$  elemű mintát exponenciális eloszlás szerint  $\lambda = 2$  paraméterrel! Határozd meg az összes mintára az ML-beclést (jel.  $\lambda_{ML}$ )! Tanultuk bevezető statisztikából, hogy bizonyos regularitási feltételek esetén az ML-beclés aszimptotikusan normális. Ennek fényében tekints az ML-beclésekre mint mintára normális eloszlásból és becsüld meg a normális eloszlás ismeretlen paramétereit! Készíts hisztogramot 50 törésponttal a  $\lambda_{ML}$  értékekből, majd ábrázold a hisztogrammal együtt a (megfelelően becsült) normális eloszlás sűrűségfüggvényét! Végezd el ugyanezt  $n = 100$  és  $n = 10$  esetén is, majd értékeld a látottakat!

---

## 3. beadandó

### Technikai tudnivalók

A futtatásokhoz a forráskódot **R** nyelven kell megírni. A megoldást E-mail-ben küldjétek el, csatoljátok a forráskódot a levélhez! Az értelmezéseket, válaszokat, illetve mindent, amit még szükségesnek érezték

közölni velem, a levél szövegtörzsébe íjátok vagy a forráskódba kommentként az aktuális helyre!

**Beadási határidő: március 25., 18:00**

Késve beadott megoldás nulla pontot ér.

---

### Háttérinformációk a feladathoz

Az innovációkról gyakran kényelmes feltenni, hogy normális eloszlásúak, a gyakorlatban azonban egyáltalán nem ritka, hogy valamilyen vastag(abb) szélű eloszlást követnek. Az ML-beclsés során ennek ellenére azzal a feltételezéssel szokás élni a könnyebb kiszámítás érdekében, hogy az innovációk normális eloszlásúak – az ilyen becslést QML, azaz quasi ML-beclsésnek is hívják.

---

### Feladat

Tekintsük az AR(2) folyamatot  $\alpha_1 = 0,2$  és  $\alpha_2 = 0,5$  paraméterek esetén.

a.) Generálj 1000 elemű mintát a folyamatból, ha az innovációk eloszlása:

- standard normális;
- 3 szabadságfokú Student- $t$ ;
- Cauchy.

Kezdeti értéknek válasszuk az  $X_{-100} = X_{-99} = 1$ -et, majd generáljunk 1100 értéket, ezek közül pedig csak az utolsó 1000-et tartsuk meg. Ez gyakori módszer idősoroknál, hogy a kezdeti értékek megválasztása már ne legyen hatással az eredményeinkre.

Ábrázold a folyamatot, a tapasztalati autokorreláció és a parciális autokorreláció függvényt! Értékelj a látottakat! Stacionárius a folyamat?

b.) Vizsgáld meg, hogy a QML-beclsés aszimptotikusan torzítatlan becslése-e az eredeti paramétereknek, ha az innovációk 3 szabadságfokú  $t$ -eloszlást követnek!

*Útmutatás:* Adott  $n$  mintaméret esetén generálj 100 különböző mintát  $t_2$ -es innovációkkal, majd becsüld vissza a paramétereket, és vedd a 100–100 becsült paraméter átlagát. Tehát egy ilyen táblázatot kell kapnod, amiből már lehet látni, aszimptotikusan torzítatlan-e a QML-beclsés:

Mintaméret	$\hat{\alpha}_1$ értékek átlaga	$\hat{\alpha}_2$ értékek átlaga
100	...	...
150	...	...
200	...	...
400	...	...
700	...	...
1000	...	...
2000	...	...
5000	...	...

c.) Az `nlm` vagy tetszőleges másik minimalizáló függvény segítségével számítsd ki az innovációk tényleges eloszlásának megfelelő, "valódi" ML-beclsést! Vizsgáld meg újra a torzítatlanságot, ha a minta 100, 200, 400, 1000 és 2000 elemű, az innovációk pedig 3 szabadságfokú  $t$ -eloszlást követnek! Vesd össze a QML-beclsésnél kapottakkal! Értelmezd/véleményezd az eredményeket!

*Útmutatás:* megmutatható (nem elvárás a bizonyítása), hogy  $X_t|X_{t-1}, X_{t-2}$  sűrűségfüggvénye  $f_{X_t|X_{t-1}, X_{t-2}}(x_t|x_{t-1}, x_{t-2}) = f_{t_2}(x_t - \alpha_1 x_{t-1} - \alpha_2 x_{t-2})$ , ahol  $f_{t_2}$  a 3 szabadságfokú  $t$ -eloszlás sűrűségfüggvénye. A likelihood-függvény számításához használjuk a láncszabályt, majd minimalizáljuk a logaritmus  $(-1)$ -szeresét.

Pontozás:

Feladatok	a.)	b.)	c.)	Összesen
Maximális pontszám	6	6	6	18

---

## 4. beadandó

### Technikai tudnivalók

A megoldás elvárt formátumai:

- .r kiterjesztésű **R**-es forráskód (nem Word, txt stb.);
- a szövegfájl pdf, Word vagy html formátumban. Ha Rmd-ben készíted, akkor az Rmd fájlt és a lefordított pdf vagy html fájlt is csatold!

A fájlok nevében ne használjatok magyar ékezetes betűket!

**Beadási határidő: 2020. április 29., 18:00.**

Késve beadott megoldás arányosan kevesebb pontot ér: ha  $k$  nappal később küldöd be, akkor a beadandóért kapott pontszám a kiértékelés utáni pontszám  $\max\left(\frac{7-[k]}{7}, 0\right)$  része.

---

### A feladat leírása

A honlapomon található *gdp\_bead.xlsx* fájl az alábbi adatokat tartalmazza:

- *GDP* munkalap: negyedéves magyar reál GDP (millió Ft), lengyel reál GDP (millió zloty), mindkettő 2010-es áron.  $\rightsquigarrow$  **modellezendő idősorok**
- *Population* munkalap: magyar és lengyel lakosságszám (fő), az utolsó két adat becslült értékek
- *FX* munkalap: magyar forint és lengyel zloty átváltási árfolyama

Adatforrások: <https://fred.stlouisfed.org> és az MNB honlapja.

- a.) Válaszd ki a két modellezendő idősor egyikét és az a.)-b.) feladatrészeket arra vonatkozóan dolgozd ki! Ábrázold az eredeti idősort, a tapasztalati autokorreláció és parciális autokorreláció függvényeket, majd értékeld szövegesen a látottakat! Vizsgáld meg előzetesen modellező szemüveggel (vannak-e hibagyánús értékek, stacionaritás, dekompozíció szükségessége)!
- b.) Modellválasztás: adatok transzformációja, trend és szezonális komponens kiszűrése a szituáció szempontjából legalkalmasabbnak bizonyuló módszerrel, majd alkalmas idősor modell illesztése. Elvárás több módszert is megnézni, nem csak az **auto.arima** függvényt használni. Az optimális modell kiválasztása és az illeszkedés tesztelése. Jónak mondható az illeszkedés?
- c.) Előrejelzés a 2024-es és 2025-ös évekre az előző feladatban bemutatott legjobbnak tartott idősor modell alapján, intervallumbecslés az előrejelzett értékekre, végül ezek ábrázolása az eredeti idősorral együtt 2 scenárió esetén:
  - c1.) BAU (business as usual): az orosz–ukrán és az izraeli–palesztin háborúk 2024-ben befejeződnek és a világgazdaság lassan elkezd visszaállni a korábbi kerékvágásba;
  - c2.) SA (severely adverse): Oroszország legyőzi Ukrajnát és még 2024-ben lerohanja Litvániát a Suwalky-folyosó megszerzéséért. Izrael megtámadja Libanont, amibe aztán a Közel-Kelet számos további állama is bekapcsolódik.
- d.) A megadott inputok alapján számold ki az egy főre jutó magyar és lengyel GDP-t az egyes negyedévekre. Készíts egy olyan dataframe-et, ami tartalmazza az eredeti GDP adatokat és az egy főre eső értékeket! Feladat, hogy ne Excel-beli manipulációkkal hajtsd ezt végre, hanem **R**-ben programozással. Vigyázz, mert a *Population* és *FX* munkalapokon lévő idősorok nem mindig adottak azokra a napokra, amire a GDP adatok, ezért néhol szükséges lesz valamilyen interpolációra/becslésre. Javaslat: használd az *approx fun* függvényt. Ha megvannak az egy főre jutó adatok, akkor ábrázold őket és adj előrejelzést 1, 2 és 5 évre előre, majd értékeld a látottakat!
- e.) Formai követelmények/helyesírás: **a beadandót olyan szellemben írd át, mintha egy pár oldalas jelentés lenne.** Tipikus hossza ábrákkal együtt 5–8 oldal, de nincsenek terjedelmi megkötések. Hagyományos fogalmazás formája legyen, tehát ne vázletszerű, hanem összefüggő. Az itteni elemzés szempontjaiban lévő a.), b.), c.) bontást ne tüntessétek fel, egyik bekezdés kövesse a másikat, esetleg kicsit nagyobb helyet lehet kihagyni, ha úgy érzitek. Nem a mennyiség számít, hanem a minőség, igyekezz tömören, lényegre törően, ugyanakkor precízen, a statisztika/matematika

nyelvezetének elvárásai szerint fogalmazni. Nincs szükség barokkos körmondatokra, az elméleti tudnivalók általános leírására. A szemet zavaró helyesírási hibák, tömeges formai és vesszőhibák pontlevonással járnak. Az ábrákról/táblázatokról derüljön ki, mit tartalmaznak (tengelyfeliratok, színezések, vonaltípusok). A szövegbe írt számított adatértékeket megfelelő számú tizedesjegyre kerekítsétek – ne írjatok például 0,542342426-t, hanem ez helyett mondjuk 0,54-et.

Pontozás:

Feladatok	a.)	b.)	c1.)	c2.)	d.)	e.)	Összesen
Maximális pontszám	3	5	3	3	5	3	22