Áringadozások gyakorlat

BPM, kvantitatív pénzügyek szakirány

2016/2017 tavaszi félév

Játékszabályok

- A gyakorlatokról maximum 3-szor lehet "büntetlenül" hiányozni. 3 hiányzás felett minden további hiányzás 5 pont levonással jár az év végi maximális 100 pontból.
- A félév végi vizsgajegyben 50% súllyal (50 pont) szerepel a gyakorlat:
  - 30 pont: beadandó feladat – egyéni adatsor elemzése, legalább 10 pontot el kell érni
  - 20 pont: házi feladatok
    - egyenként 5 pontosak, a 4 legmagasabb pontszámot veszem figyelembe
    - folyamatosan lesznek kihirdetve
    - a beadási határidőt a feladat sorszáma mellett találjátok
    - E-mail-ben küldjétek be, a forráskódokat/ábrákat mellékeljétek

Infók a gyakorlatvezetőről

Név	Varga László
Tanszék	Valószínűségelméleti és Statisztika Tanszék (ELTE TTK)
Szoba	D 3-309
E-mail	`vargal4@cs.elte.hu`
Honlap	`vargal4.elte.hu`

Szimulációkhoz használt szoftver/programnyelv: **R**

- Statisztikai modellezésre, adatok elemzésére kiválóan alkalmas programnyelv
- Nyílt forráskódú, ma már alig van probléma, feladat, aminek a megoldására ne lenne valamilyen package – akár több is
- Jelenleg a legelterjedtebb matematikai célú programnyelv
- A gyakorlaton és az előadáson is ezt használjuk, a házi feladatokhoz és a nagy beadandóhoz a számításokat/szimulációkat ebben kell megírni

- Letöltési helye: `https://cran.r-project.org/`
- Szövegszerkesztésre ajánlott szoftver: RStudio; letöltési helye: `https://www.rstudio.com/products/rstudio/download3/`

Ajánlott irodalom: Pröhle Tamás, Zempléni András: Statistical Problem Solving in R. $\rightsquigarrow$ jegyzet, **R**-es bevezető

Elérési helye: `http://zempleni.elte.hu/Stat_R_Prohle_Zempleni`

1.) Készítsünk egy ábrát ábramagyarázattal, amely a standard normális eloszlás, valamint az 1, 3, 8 és 30 szabadságfokú t-eloszlások sűrűségfüggvényét tartalmazza! Mit látunk?

2.) Szimuláljunk 1, 3, 8 és 30 szabadságfokú t-eloszlású mintákat, majd vizsgáljuk meg a normalitásukat (illeszkedésvizsgálat)

a.) grafikus módon: Q-Q plottal

b.) formális tesztekkel:

- χ^2-próbával (diszkretizálás);
- Kolmogorov-Szmirnov / Anderson-Darling próbákkal (folytonos illeszkedésvizsgálat).

Vizsgáljuk meg a próbák erejét különböző mintanagyságok esetén!

3.) Generálj egy 500 elemű mintát 3 szabadságfokú t-eloszlásból, majd becsüld vissza a szabadságfok mint paraméter értékét maximum likelihood módszerrel

- az 'optim'/'nlm' függvény;
- a 'stats4' library 'mle' függvénye;
- a 'fitdistrplus' library 'fitdist' függvénye

segítségével! Határozd meg az Akaike-féle információs kritériumot!

A 'fitdistrplus' package használatáról jó leírás:

Delignette-Müller, Dutang: *Fitdistrplus: An R Package for Fitting Distributions.* Journal of Statistical Software, 2015.

4.) Használjuk a 'fitdistrplus' package-et eloszlásillesztésre!

Generálj egy 500 elemű mintát gamma-eloszlásból, majd vizsgáld meg alkalmas illeszkedésvizsgálati módszerrel, hogy a minta vajon lehet-e egyenletes, normális, exponenciális, gamma, lognormális eloszlású! Válaszd ki a legjobban illeszkedő eloszlást az AIC, valamint a BIC segítségével!

5.) Használjuk a 'fitdistrplus' package-et eloszlásillesztésre!

Generálj egy 500 elemű mintát negatív binomiális eloszlásból, majd vizsgáld meg alkalmas illeszkedésvizsgálati módszerrel, hogy a minta vajon lehet-e geometriai, negatív binomiális, illetve Poisson-eloszlású! Válaszd ki a legjobban illeszkedő eloszlást az AIC, valamint a BIC segítségével!

HF1.) **[III.2.]** A honlapomon található 'EurUsd.Rdata' fájl az euró és az USA dollár közötti keresztárfolyamok loghozamának abszolút értékét tartalmazza (2011 – 2016, napi adatok). Próbálj eloszlásokat illeszteni erre a mintára a tanult módszerek segítségével (grafikusan és formális teszttel/tesztekkel is)! Melyik eloszlás bizonyult a legjobbnak?

6.) Az X valószínűségi változó $\mathbf{S}(\alpha,\beta,\gamma,\delta,0)$ eloszlású, ha karakterisztikus függvénye $\varphi(t) = \exp\left\{-\gamma^\alpha|t|^\alpha \left[1 + i\beta \mathrm{tg}\left(\frac{\pi\alpha}{2}\right)\mathrm{sgn}(t)\left(|\gamma t|^{1-\alpha} - 1\right)\right] + i\delta t\right\}$ alakú ($\alpha \neq 1$ esetén), ahol $0 \leq \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma > 0$, δ valós paraméterek.

a.) Milyen α, β, γ, δ paraméterekkel stabilis eloszlású az m várható értékű és σ szórású normális eloszlás?

b.) Milyen α, β, γ, δ paraméterekkel stabilis eloszlású az a eltolás- és b skálaparaméterű Cauchy-eloszlás? Használjuk fel, hogy a standard Cauchy-eloszlás karakterisztikus függvénye $\varphi(t) = e^{-|t|}$.

c.) Határozd meg a fent definiált $\mathbf{S}(\alpha,\beta,\gamma,\delta,0)$ stabilis eloszlás várható értékét

$\alpha > 1$ esetén!

7.) Ábrázoljuk stabilis eloszlású valószínűségi változók sűrűségfüggvényét a 'stabledist' package 'dstable' függvénye segítségével, vizsgáljuk meg a paraméterek hatását a sűrűségfüggvény alakjára!

8.) Lássuk be grafikusan, hogy a normális eloszlás egy speciális stabilis eloszlás: generáljunk mintákat az adott várható értékű és szórású normális eloszlásból, illesszünk rájuk stabilis eloszlást, majd vizsgáljuk meg, hogy az illesztett eloszlás sűrűségfüggvénye és az eredeti sűrűségfüggvény mennyire van közel egymáshoz! Ugyanezt csináljuk meg Cauchy-eloszlás esetén is!

9.) Illesszünk stabilis eloszlást a *gbp_ dm_ rate* és *nasdaq_ logreturn* adatsorokra ML-módszerrel, majd készítsük el a szórásstabilizált P-P plotot, és vizsgáljuk meg az illeszkedést!

10.) Illesszünk stabilis eloszlást a *nasdaq_ djia_ 2005_ 2010.txt* adatsor vetületeire kvantilis, illetve ML-módszerrel (az utóbbi nagyon lassú, az eredményeket javasolt az *ered2.Rdata* fájlból importálni), majd készítsük el a szórásstabilizált P-P plotot. Vizsgáljuk meg a paraméterfüggvényeket kvantilis és ML-módszer esetén is! Vessük össze a kettőt!

11.) Legyen az $X_1, X_2, \dots$ valószínűségi változó sorozat független

a.) $E(0;1)$;

b.) Frechet (az eloszlásfüggvény: $F(x) = e^{-1/x} I(x > 0)$)

eloszlású. Határozzuk meg az a_n és b_n valós számsorozatot, hogy az $M_n = \max(X_1, \dots, X_n)$ maximumok ezekkel eltolt/átskálázott $U_n = \frac{M_n - a_n}{b_n}$ sorozata $n \to \infty$ esetén egy GEV eloszláshoz konvergáljon! Határozd meg a határeloszlás paramétereit és az eloszlás tartóját!

12.) Szimuláljunk GEV eloszlásból!

a.) Vizsgáljuk meg a paraméterek hatását a sűrűségfüggvényre!

b.) Szimuláljunk 25, 100 és 400 elemű mintákat GEV eloszlásból, majd becsüljük vissza a paremétereket és a 95%-os kvantilist! Vizsgáljuk meg a paraméterbecslések tulajdonságait (torzítatlanság, gyenge konzisztencia)!

c.) Generáljunk különböző eloszlásokból különböző elemű mintákat, majd vizsgáljuk meg, hogy a blokkméret növelésével a blokkmaximumok eloszlása milyen gyorsan közelít a GEV eloszláshoz!

13.) *Profil likelihood konfidenciaintervallum a visszatérési szintre.*

a.) Ha l év során rendelkezésünkre áll n megfigyelés, akkor az m éves visszatérési időhöz tartozó visszatérési szint (röviden: m éves visszatérési szint) melyik kvantilishez tartozó ismérvértéknek felel meg?

b.) Paraméterezzük át a GEV eloszlás log-likelihood függvényét: a (μ, σ, ξ) helyett legyenek a paraméterek (q_y, σ, ξ), ahol q_y az y-kvantilis.

c.) Generáljunk 100 elemű mintát GEV eloszlásból, majd becsüljük vissza a paramétereket! Készítsünk profil likelihood konfidencia intervallumot a 20 éves visszatérési szintre!

14.) A *nasdaq_ logreturn* adatsor megfelelően választott blokkmaximumaira illesszünk GEV eloszlást, majd vizsgáljuk meg az illeszkedést!

15.) Vizsgáljuk meg a kétdimenziós normális eloszlást és a kétdimenziós t-eloszlást az aszimptotikus függetlenség szempontjából!

16.) Szimuláljunk 2 dimenziós GEV eloszlású változókat különböző eloszláscsaládokból és vizsgáljuk meg a paraméterek hatását a sűrűségfüggvényre! Mennyire pontosan lehet becsülni a paramétereket?

17.) A *nasdaq_ djia_ 2005_ 2010* adatsorra illesszünk 2 dimenziós GEV eloszlást!

HF2.) **[III.23.]** A honlapomon található 'hf2.Rdata' egy Magyarországon működő bank egyik kötvényalapjának árfolyamadataiból számolt loghozamokat tartalmazza 2013. március 9. és 2017. március 6. között. Illessz stabilis eloszlást az adatokra, illetve GEV-eloszlást a blokkmaximumokra alkalmas blokkméret-választással! Értékeld az illeszkedés jóságát!

18.) *Példa olyan becsléssorozatra, amely torzítatlan, de nem konzisztens.*
Legyen $X_1, X_2, \dots, X_n, \dots$ i.i.d. minta egy ϑ várható értékű és 1 szórású eloszlásból. Mutassuk meg, hogy ekkor a $T_n(\mathbf{X}) = X_1$ becsléssorozat torzítatlan, azonban nem konzisztens becslése ϑ-nak!

19.) *Példa konzisztens becsléssorozatra, melynek szórásnégyzete nem tart 0-hoz.*
Legyen $X_1, X_2, \dots, X_n, \dots$ i.i.d. minta egy ϑ várható értékű és 1 szórású eloszlásból; U_n az X-ektől független valószínűségi változó, amire $P(U_n = \pm n) = \frac{1}{n^2}$ és $P(U_n = 0) = 1 - \frac{2}{n^2}$. Mutassuk meg, hogy ekkor a $T_n(\mathbf{X}) = \overline{X} + U_n$ becsléssorozat erősen konzisztens és torzítatlan becslése a ϑ paraméternek, azonban $D^2_\vartheta T_n(\mathbf{X})$ nem tart 0-hoz!

20.) *Konzisztenciából mikor következik az aszimptotikus torzítatlanság.*
Legyen T_n konzisztens becslése ϑ-nak; $C > 0$ valós szám, amire $D^2 T_n < C$ minden i-re. Ekkor bizonyítsuk be, hogy T_n aszimptotikusan torzítatlan becslése ϑ-nak!

21.) Szimuláljunk GPD eloszlású valószínűségi változókat különböző paraméterekkel és vizsgáljuk meg a paraméterek hatását a sűrűségfüggvényre! Mennyire pontosan lehet becsülni a paramétereket?

22.) Generáljunk 100 elemű mintát GPD eloszlásból, majd készítsünk profil likelihood konfidencia intervallumot a 50 éves visszatérési szintre!

23.) Legyen az $X_1, X_2, \dots$ valószínűségi változó sorozat független

a.) $\text{Exp}(\lambda)$;

b.) $E(0;1)$

eloszlású. Határozzuk meg a küszöbmeghaladási szint eloszlását (a GPD eloszlás paramétereit)!

24.) Normális, illetve gamma eloszlású minták esetén vizsgáljuk meg a küszöbválasztás hatását a becslésre! Készíts átlagos meghaladási függvényt az egyes eloszlásokra (mean excess function), majd értékeld a látottakat!

25.) Illesszünk a nasdaq_logreturn adatsorra GPD eloszlást! Tegyük fel, hogy évi átlagban 250 adatunk van. Készítsünk 96%-os profil likelihood intervallumot a 0,5 éves visszatérési szintre!

HF3.) **[IV.6.]** A honlapomon található 'hf2.Rdata' egy Magyarországon működő bank egyik kötvényalapjának árfolyamadataiból számolt loghozamokat tartalmazza 2013. március 9. és 2017. március 6. között. Illessz GPD eloszlást a veszteségekre megfelelő küszöbszint választásával! Értékeld az illeszkedés jóságát az ismert tesztekkel (A-D próba, K-S próba, χ^2-próba) is! Adj profil likelihood intervallumbecslést a 2 éves visszatérési időhöz tartozó visszatérési szintre (az adatok körülbelül 4 évet ölelnek át, és évente átlagosan 312 megfigyelésünk van)!

26.) Tekintsünk egy n elemű i.i.d. mintát $E(0,\vartheta)$ eloszlásból, célunk az ismeretlen ϑ paraméter becslése. Használjuk a $T(\mathbf{X}) = \frac{n+1}{n}\max(X_1,\dots,X_n) = \frac{n+1}{n}X_n^{(n)}$ torzítatlanná tett ML-becslést ϑ becslésére!

a.) Generáljunk sokszor ebből az eloszlásból $n = 50$ elemszámú mintát, majd becsüljük $T(\mathbf{X})$ eloszlását – készítsünk hisztogramot, magfüggvényes sűrűségfüggvény-becslést! Ennek segítségével adjunk 95%-os megbízhatóságú konfidenciaintervallumot ϑ-ra!

b.) Generáljunk egy $n = 50$ elemszámú mintát, abból $m = 50$ elemű i.i.d. (nemparaméteres) bootstrap mintákat (jel. $\mathbf{X}^*$), majd ezek segítségével becsüljük a $T(\mathbf{X}^*)|\mathbf{X}$ bootstrap eloszlást! Mit tapasztalunk? Adjunk 95%-os megbízhatóságú konfidenciaintervallumot ϑ-ra!

c.) Találjunk elméleti magyarázatot az előző feladatrészben tapasztalt jelenségre – határozzuk meg a $P\left((X^*)_n^{(n)} = X_n^{(n)}\middle|\mathbf{X}\right)$ valószínűséget! Hova tart $n\to\infty$ esetén?

d.) Generáljunk egy $n = 50$ elemű mintát, majd abból paraméteres boostrap segítségével szimuláljunk $m = 50$ elemű mintákat! Ezek segítségével becsüljük $T(\mathbf{X}^*)|\mathbf{X}$ booststrap eloszlást! Mit tapasztalunk? Adjunk 95%-os megbízhatóságú konfidenciaintervallumot ϑ-ra!

segítségével!

27.) Becsüljük meg egy 100 elemű $N(m,2^2)$ eloszlású mintából a $T(\mathbf{X}) = \frac{X_{50}^{(100)}+X_{51}^{(100)}}{2}$ tapasztalati medián szórását

a.) i.i.d. (nemparaméteres) bootstrap-pel;

b.) paraméteres bootstrap-pel.

Vessük össze a kapottakat a "valódi" szórással (számoljuk szimulációval)!

28.) Legyen X_t AR(1) folyamat α paraméterrel és 1 szórású innovációval, Mutasd meg, hogy ekkor $D^2(\overline{X}_n) = \frac{\sigma^2}{n^2(1-\alpha^2)}\frac{n-n\alpha^2-2\alpha+2\alpha^{n+1}}{(1-\alpha)^2}$.

29.) Szimuláljunk különböző paraméterű AR(1) és MA(1) folyamatokat, majd vizsgáljuk meg a különböző bootstrap eljárásokat:

- i.i.d. bootstrap
- blokk bootstrap (használjuk a Politis-White algoritmust az optimális blokkméret meghatározására)
- reziduális bootstrap

az átlag eloszlásának, illetve az átlag szórásának vizsgálatára!

30.) Szimuláljunk különböző paraméterű AR(1) és MA(1) folyamatokat, célunk az átlag szórásának a becslése. Vizsgáljuk meg empirikus úton, hogy melyik blokkméretnél lesz a blokk bootstrap becslés a legközelebb az elméleti értékhez! Ez megegyezik a Politis-White algoritmus által számolt optimális blokkmérettel?

31.) Elemezzük a 'reszv_ar.txt' fájlban található két idősort!

a.) Vizsgáljuk meg, van-e szignifikáns trend, és amennyiben van, alkalmas transzformációval tegyük az idősorokat (közel) stacionáriussá!

b.) Becsüld az átlag szórását blokk bootstrap módszerrel!

c.) Adj 'bootstrap-t interval' és 'bootstrap percentile interval' intervallumbecslést az átlagra! Az egyes idősorokra melyiket használnád, várhatóan melyik lesz a jobb?

HF4.) **[IV.20.]** Tekintsük az $X_t = \varepsilon_t - \varepsilon_{t-1} + \varepsilon_{t-2} - \varepsilon_{t-3} + \varepsilon_{t-4}$ MA(4) folyamatot, ahol ε_t 0 várható értékű és 1 szórású fehér zaj. Célunk az átlag szórásának a becslése egy $n = 5000$ elemű mintából. A kiindulási minta generálása előtti sorban használjátok a set.seed(11111)-et, hogy rekonstruálni lehessen az eredményeket!

a.) Számold ki (vagy becsüljük szimulációval) az átlag szórását!

b.) Vizsgáld meg empirikus úton, hogy melyik blokkméretnél lesz a blokk bootstrap becslés a legközelebb az elméleti értékhez! Ez megegyezik a Politis-White algoritmus által számolt optimális blokkmérettel?

c.) Becsüld meg az átlag szórását reziduális bootstrap segítségével! Ez jobb becslést ad, mint a blokk bootstrap?

Ajánlott irodalom kopulákhoz:

P. Embrechts, F. Lindskog, A. McNeil: Modelling Dependence with Copulas and Applications to Risk Management. Department of Mathematics, ETHZ, Zürich, 2001.

32.) Szimuláljunk különböző paraméterű normális, t, Gumbel és Clayton kopulákat és ábrázoljuk is őket! Nézzük meg, mennyire különböznek egymástól! Becsüljük vissza a generált kopulák paramétereit inverz tau módszerrel, majd adjunk intervallumbecslést a paraméterekre!

33.) Vizsgáljuk a *nasdaq_djia_2005_2010.txt* kétdimenziós részvényindex-adatsorban lévő összefüggőségi struktúrát kopulákkal!

a.) Teszteljük az adatsor elliptikusságát!

b.) Illesszünk különböző kopulákat, vizsgáljuk az illeszkedést grafikusan és nézzük meg az illesztés stabilitását tolóablakos módszerrel!

c.) Teszteljük az illesztett kopulát Kendall-folyamat és Rosenblatt-transzformáció segítségével!

34.) A kopulás megközelítés erőssége a peremek és az összefüggőségi struktúra szétválasztása. Vizsgáljuk meg az alábbi példa alapján, hogy kopulák segítségével előállított együttes eloszlásokból visszabecsült kopula valóban olyan eloszlású-e, mint amiből generáltunk!

a.) Szimuláljunk 1000 elemű mintákat olyan kétdimenziós eloszlásból, melynek kopulája Clayton-kopula $\theta = 2$ paraméterrel, peremei pedig exponenciális eloszlásúak 4 paraméterrel, illetve normális eloszlásúak 2 várható értékkel és 3 szórással!

b.) A mintából becsüljük vissza a kopulát, illesszünk rá Clayton-kopulát és Gauss-kopulát, és végezzünk illeszkedésvizsgálatot Kendall-folyamattal!

HF5.) **[V.11.]** Illessz alkalmas kopulá(ka)t a honlapomon található 'hf5.Rdata' fájlban található adatsorra (az oszlopokban két magyar részvényalap 9 éves loghozamai szerepelnek)! Teszteld az illeszkedést grafikusan és formális tesztekkel egyaránt! A legjobban (vagy legkevésbé rosszul) illeszkedő kopula modell esetén vizsgáld az illeszkedés stabilitását tolóablakos módszerrel!

35.) A *reszvenyek.txt* napi hozam adatsorra oldjuk meg a portfólió-optimalizálás feladatát!

a.) Először csak a szórást minimalizáljuk, majd különféle elvárt hozamokat tűzzünk ki célul! Hogyan változnak ilyenkor az optimális súlyok? Mi történik, ha a shortolást is megengedjük?

b.) Szimuláljuk a hozamokat többdimenziós normális eloszlásból – vegyünk különböző elemszámú mintákat –, és vizsgáljuk az optimális portfólió szórását!

c.) Szimuláljuk a hozamokat olyan többdimenziós eloszlásból, ahol a peremek 4 szabadságfokú t-eloszlásúak, a kopulájuk pedig t-kopula!

d.) Milyen értéket érdemes választani az elméleti eredmények alapján a T szimulálási mintaelemszámnak, hogy a portfólió szórásának becslési hibája (q_0) 20 %-on belül legyen, amennyiben többdimenziós normális eloszlásból szimulálunk? Végezzünk ezzel a T-vel szimulációkat, majd vizsgáljuk meg, hogy mennyire sikerült a kitűzött 20%-os hibahatárhoz közel kerülni! Az adott mintára mi lenne az "optimális" T?

e.) Szimuláljuk a hozamokat többdimenziós normális eloszlásból és vizsgáljuk meg, hogy $T \to \infty$ esetén a portfólió szórásának becslési hibája (q_0) valóban 1-hez konvergál! Empirikus úton nézzük meg, milyen nagy T-re van szükség legalább, hogy a limeszt 1%-on belül megközelítsük?

36.) Alkalmazzuk a főkomponens-analízis módszerét a sok részvény hozamát tartalmazó *many_equities.txt* adatsorra! A szűrt kovarianciamátrixot úgy készítsük el, hogy az első k főkomponens a teljes szóródás legalább 80%-áért feleljen!

37.) Szimuláljunk nagy méretű Wigner-mátrixot és vizsgáljuk meg, hogy ha a sajátértékekre mint mintára tekintünk, vajon származhatnak-e Wigner-eloszlásból!
Szimuláljunk nagy méretű Wishart-mátrixot és vizsgáljuk meg, hogy ha a sajátértékekre mint mintára tekintünk, vajon származhatnak-e Marchenko-Pastur-eloszlásból!

38.) Az *50equity.txt* adatsorra vizsgáljuk meg empirikusan, hogy teljesül-e a Marchenko-Pastur-tétel. A legnagyobb sajátértékek közül hányat kell elhagyni, hogy a spektrum eloszlása közel Marchenko-Pastur-eloszlású legyen?

HF6.) **[V.11.]** Elemezd a honlapomon található *HF6.Rdata* fájlban található 8 éves magyar részvényadatokat (napi záróárfolyamok) portfólió-optimalizálási szemszögből. Mindenekelőtt készíts az adatokból napi hozamokat.

a.) Készíts ábrát arról, hogy különböző kitűzött – az adatok szempontjából értelmes – hozamcélok esetén hogyan alakul az egyes részvényekbe fektetett pénz részaránya!

b.) Milyen értéket érdemes választani az elméleti eredmények alapján, valamint tapasztalati úton a T szimulálási mintaelemszámnak, hogy a portfólió szórásának becslési hibája (q_0) 10 %-on belül legyen, amennyiben többdimenziós normális eloszlásból szimulálunk?

Alternatív feladat: Tűzz ki egy elérendő hozamcélt! Vizsgáld meg, hogy az első 7 éves adatokat használva hogyan alakul az egyes részvényekbe fektetett pénz részaránya, ha 5 éves tolóablakokra becsülöd meg az optimális portfóliót! Az ebből előálló idősor segítségével próbálj becslést adni arra, hogy az utolsó 5 éves adatokra mi lesz az optimális portfólió! Vesd össze a valódival!
