

# Idősorok és többdimenziós statisztika gyakorlat

Matematika alapszak, matematikai elemző szakirány  
2022/2023 őszi félév

## Játékszabályok

- A gyakorlatokról maximum 3-szor lehet hiányozni. Aki többször hiányzik, nem kap gyakjegyet.
- 100 + x pontot lehet szerezni a félév során:
  - 28 pont: **B** jelű, "kis" beadandó feladatok (4 pont · 7)
  - 22 pont: "nagy" szimulációs/számítógépes beadandó feladat
  - 50 pont: ZH: dec. 8., Gépteremben Canvas-on keresztül
  - x pont: SZ jelű szorgalmi feladatok
- A ZH-n minimálisan teljesíteni kell 30 %-ot. Ha a ZH sikertelen, nem írod meg, vagy javítani szeretnél, akkor vizsgaidőszak első hetén lesz lehetőség a pótZH megírására vagy a javításra. A jobbik eredményt veszem figyelembe, azaz nem lehet rontani. Sikertelen vagy meg nem írt ZH esetén gyakUV-t kell írnod, a beadandókért kapott pontok megmaradnak.
- A ZH-n minden offline/online anyag használható, a kommunikáció bármely formája más személyekkel viszont tilos.
- **B** jelű beadandók: Mindegyik maximálisan 4 pontot ér, a legjobb 7-et veszem figyelembe. Több feladatot is kihirdetek, amik közül ízlés szerint válogathatsz. A beadandók célja, hogy folyamatosan tanuljatok, gyakoroljatok.
- "Nagy" beadandó: nagyobb lélegzetvételi, számítógép segítségével megoldandó feladat.
- Aki a beadandókból összesen nem éri el a 40%-ot (20 pont), nem kap gyakjegyet.
- Minden beadandónál elvárás az önálló munkavégzés. Amennyiben nyilvánvaló másolás gyanúja merül fel, az érintettek 0 pontban részesülnek.

elégtelen (1)	0	–	34,99
elégséges (2)	35	–	49,99
- Osztályozás:

közepes (3)	50	–	64,99
jó (4)	65	–	79,99
jeles (5)	80	–	$\infty^\pi$

## Infók a gyakorlatvezetőről

Név	Varga László, <i>óraadó</i>
Munkahely	Citi, MQA
Tanszék	Valószínűségelméleti és Statisztika Tanszék (ELTE TTK)
E-mail	vargala4@gmail.com
Honlap	vargala4.elte.hu

## Ajánlott irodalom (példatárak)

APZ Arató-Prokaj-Zempléni: Bevezetés a valószínűségszámításba és alkalmazá-

saiba: példákkal, szimulációkkal.

BP Barczy-Pap: Sztochasztikus folyamatok példatár, 1. rész.

BP2 Barczy-Pap: Sztochasztikus folyamatok példatár 2. rész.

VL Varga László: bevezető valószínűségszámítás kurzus példáinak megoldásai.

VL2 További példák megoldásokkal.

---

A félév során idősortmodellezésre/szimulációkhoz használt 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
- R nyelv/szoftver letöltési helye
- Szövegszerkesztésre ajánlott szoftver: RStudio; letöltési helye
- R Markdown: egyszerű szövegszerkesztő és kódot is futtató package, rövid bevezető

---

**Ismétlésre ajánlott:** VL-ből 81–85

- 1.) Legyenek  $Y_1, Y_2, \dots$  i.i.d. val. változók,  $P(Y_1 = 1) = 1 - P(Y_1 = -1) = p \in (0, 1)$ . Vizsgáljuk meg, hogy a következő sorozatok Markov-láncot alkotnak-e, és ha igen, vizsgáljuk meg, homogének-e, valamint írjuk fel az átmenetvalószínűség-mátrixukat!
  - a.)  $X_n = Y_1 \cdot \dots \cdot Y_n \quad n = 1, 2, \dots$
  - b.)  $X_n = Y_1 + \dots + Y_n \quad n = 1, 2, \dots$
  - c.)  $X_n = Y_n \cdot Y_{n+1} \quad n = 1, 2, \dots$
- 2.) A 0 időpontban 2000 Ft-om van. Az 1, 2, ... időpontokban 1000 Ft értékű fogadásokat kötök. A fogadást  $p$  valószínűséggel megnyerem,  $1 - p$  valószínűséggel pedig elveszitem. Célom az, hogy a tőkém 4000 Ft-ra növeljem, amint ez teljesül, befejezem a játékot. A játéknak akkor is vége van, ha tőkém 0 Ft-ra esik. Legyen  $X_t$  a  $t$ -edik fogadás után a tőkém nagysága.
  - a.) Mutasd meg, hogy  $X_t$  Markov-lánc, határozd meg az állapotteret és az átmenetvalószínűségi mátrixot! Ábrázold a Markov-lánc gráfját és osztályozd az állapotokat!
  - b.) Mi az esélye, hogy 4000 Ft-tal / 0 Ft-tal fejezem be a játékot, ha  $p = 0, 6$ ?
  - c.) Várhatóan hány lépésben fejeződik be a játék, ha  $p = 0, 6$ ?
  - d.) Nézzük meg, konvergál-e az átmenetmátrix  $n$ -edik hatványa  $n \rightarrow \infty$  esetén!
- 3.) Tegyük fel, hogy a kóla iparágban kétféle termék van: Coke és Poke. Ha egy személy legutóbb Coke-ot vett, akkor 90% eséllyel legközelebb is Coke-ot vesz. Amennyiben legutóbb Poke-ot ivott, akkor 80% valószínűséggel a következő vásárlás során is Poke-ot fog venni. Tegyük fel, hogy kezdetben a Coke gyártója a piac 60%-át uralja.
  - a.) Ábrázold a Markov-lánc gráfját, írd fel az átmenetvalószínűség-mátrixot és osz-

tályozd az állapotokat!

- b.) Ha egy vásárló Poke-ot iszik, akkor mennyi a valószínűsége, hogy a mostantól számított 2. vásárláskor Coke-ot fog inni?
  - c.) Ha egy vevő Coke-ot iszik, akkor mennyi a valószínűsége, hogy a mostantól számított 3. vásárláskor is Coke-ot fog inni?
  - d.) Számítsuk ki a két márka piaci részesedését (a fogyasztók hány %-a fogyasztja az egyiket és a másikat) a 3. időpont után!
  - e.) Vizsgáljuk meg szimulációval az  $n$  lépéses átmenetvalószínűség mátrixot és a két márka piaci részesedését  $n \rightarrow \infty$  esetén! Nézzük meg a piaci részesedés konvergenciáját akkor, ha kezdetben a Poke márka teljesen uralja a piacot!
  - f.) Számítsd ki a stacionárius valószínűségeket!
  - g.) Tegyük fel, hogy egy fél literes kóla ára 300 Ft, előállítási költsége pedig 200 Ft. Felmérésekből kiderítették, hogy kólát rendszeresen 120 ezer ember fogyaszt, ők hetente átlagosan 1-et isznak meg. Számoljunk 52 héttel egy évben. Számítsd ki, hogy a Coke és a Poke gyártója (hosszú távon) mekkora profitra tesznek szert 1 év alatt!
  - h.) Ha Peti ezen a héten Coke-ot fogyaszt, akkor várhatóan hány hét múlva fogyasztja el a következő Coke-ot?
- 4.) Modellezzük egy egyetem matematika BSc szakos hallgatóit Markov-lánccal! Az elsőévesek 10%-a évet ismétel, 10% hagyja ott végleg a matek alapszakot, a többiek a következő évtől másodévesek lesznek. A másodév a legnehezebb, a másodévesek 20%-a ismétel évet és 15%-uk adja fel végleg a szakját. A harmadévesek 10%-a ismétel évet, és 5%-uk dönt az alapszak elhagyása mellett, a többiek elvégzik a szakot.
- a.) Határozd meg a Markov-lánc állapotterét és az átmenetvalószínűségi mátrixot! Ábrázold a Markov-lánc gráfját és osztályozd az állapotokat!
  - b.) Milyen valószínűséggel lesz egy elsőéves hallgató 2 év múlva harmadéves?
  - c.) Milyen valószínűséggel lesz egy elsőéves hallgató 3 év múlva harmadéves?
  - d.) Milyen valószínűséggel hagyja el egy elsőéves hallgató végzettség nélkül a matematika alapszakot?
  - e.) Mi az esélye annak, hogy egy másodéves hallgató elvégzi a szakot?
  - f.) Várhatóan hány évig lesz egy másodéves hallgató másodéves?
  - g.) Várhatóan hány évet tölt egy elsőéves egyetemista az egyetemen?
  - h.) A visszajelzések szerint a munkaerőpiac évente 50 matematika alapszakos hallgatót képes felszívni. Ha a hallgatóknak várhatóan a 40%-a tovább szeretne tanulni valamilyen mesterszakra, akkor az egyetemnek hány golyót érdemes felvennie a szakra?
  - i.) Nézzük meg, konvergál-e az átmenetmátrix  $n$ -edik hatványa  $n \rightarrow \infty$  esetén!
  - j.) Tegyük fel, hogy "sok" éven keresztül, minden évben 60 hallgatót vesznek fel matematika alapszakra és más egyetemekről nem jönnek át másod-, illetve harmadévesek erre a szakra. "Sok-sok" év múlva várhatóan hány hallgató fog tanulni az egyes évfolyamokon? Ha egy hallgató képzése egy félévre 500 ezer Ft,

akkor a szak összesen mennyi pénzébe kerül 1 félév során az államnak?

- 5.) Egy országban a családok három csoportba oszthatók: városban, városok vonzáskörzetében vagy vidéken élők. Egy évben a városlakó családok 15%-a átköltözik a vonzáskörzetbe és 5%-a vidékre költözik; ugyanakkor a vonzáskörzetben élők 6%-a városba és 4%-a vidékre költözik; végül a vidéki lakosság 4%-a a városokba és 6%-a a városok vonzáskörzetébe települ át. Jelenleg a családok 40%-a városokban, 35%-a pedig a városok vonzáskörzetében él.
  - a.) Ha egy család városban lakik, akkor mi a valószínűsége, hogy 2 év múlva is városban fog lakni?
  - b.) Két év múlva a családok hány %-a fog városban élni?
  - c.) Ha ez a modell hosszú távon jól írja le a lakosok költözési viselkedését, akkor hogyan fog alakulni az ország lakóhely szerinti megoszlása?
- 6.) A Google által használt **PageRank** algoritmus a keresőben talált oldalakat sorrendbe állítja, egyszerűsített változata a következőképpen írható le. Tekintsünk egy Markov-láncot, melynek gráfját jelölje  $G = (V, E)$ , ennek csúcshalmaza az adott adatbázisban található oldalak, továbbá  $(u, v) \in E$  ha az  $u$  oldalról mutat link a  $v$  oldalra, az élre pedig az  $\frac{1}{d_{ki}(u)}$  számot írjuk, ahol  $d_{ki}(u)$  az  $u$  csúcsból más csúcsokba mutató élek száma (ki-fokszám). Amennyiben egy oldal semelyik másik oldalra se mutat, akkor egyenletes eloszlás szerint véletlenszerűen választunk egy oldalt és oda ugrunk, azaz az összes többi csúcshoz berajzolunk élet, és az azonos  $\frac{1}{|V|-1}$  számokat írunk rájuk. A Markov-lánc stacionárius eloszlása adja meg az oldalak rangszámait, minél nagyobb a szám, annál "fontosabb" az adott oldal. Tekintsünk egy olyan esetet, amikor összesen az  $(a, b, c, d)$  négy oldal van egy adatbázisban a következő linkekkel:  $a \rightarrow b, b \rightarrow c, d \rightarrow c, b \rightarrow d$ .
- a.) Írjuk fel a PageRank által meghatározott bolyongás állapotmátrixát és osztályozzuk az állapotokat!
  - b.) Határozzuk meg a stacionárius eloszlást, majd adjuk meg az oldalak sorrendjét!
- Gyakorlásra ajánlott:** BP2-ből 1.10–1.15, 1.16, 1.20, 1.21, 1.25, 2.1, 2.2, 2.16–2.19, 2.23, 2.27, 4.16–4.22, 4.25–4.31, 5.6–5.8, 5.13, 5.15, 5.17, 9.1
- B1.) [09.29.]** Egy urna a kiinduló helyzetben két festetlen golyót tartalmaz. Véletlenszerűen kiválasztunk egy golyót és feldobunk egy pénzérmét. Ha a választott golyó festetlen és fejet dobunk, akkor a kiválasztott festetlen golyót pirosra festjük; ha a választott golyó festetlen és írást dobunk, akkor a kiválasztott festetlen golyót feketére festjük. Ha a golyó már korábban meg lett festve, akkor átfestjük a másik színre. Jelöljük az urna állapotát az alábbi számhármassal:  $(i, j, k)$ , ahol  $i$  az urnában lévő festetlen,  $j$  a piros,  $k$  pedig a fekete golyók száma. Legyen  $X_t$  az urna állapota a  $t$ -edik lépés után. Mutasd meg, hogy  $X_t$  Markov-lánc és határozd meg az átmenetvalószínűségi mátrixot!
- Útmutatás:* a Markov-láncnak 6 állapota van, a kezdeti állapot a  $(2, 0, 0)$  számhármassal írható le.
- B2.) [10.06.]** Tekintsük a 3. feladat történetét. Egy reklámügynökség megkeresi a Coke tulajdonosát egy olyan ajánlattal, hogy a márkát rendkívül menővé fogják

tenni vagány TV-reklámokkal, ezáltal annak a valószínűsége, hogy egy fogyasztó a következő héten a másik márkát fogyasztja, 10%-ról 5%-ra fog lecsökkenni. A reklámkampányra 50 millió Ft-os ajánlatot adnak. Számításokkal alátámasztva adj tanácsot, a Coke tulajdonosának érdemes-e belekezdeni a reklámkampányba!

**B3.) [10.13.]** Egy vállalat telefonon értékesíti termékét. Tekintsünk egy potenciális vevőt, akit még nem hívott fel a cég képviselője – az 1. hívás után 60% az esélye, hogy az illető lanya érdeklődést fog tanúsítani a termék iránt, 30% eséllyel komolyan fog érdeklődni és 10% eséllyel a vállalat törölni fogja a potenciális vevők köréből. Az a vevő, aki korábban lanya érdeklődést mutatott, a következő hívás során 20% eséllyel továbbra is alacsony érdeklődést fog mutatni, 30% valószínűséggel komoly érdeklődő lesz, 30% eséllyel megveszi a terméket és 20% eséllyel törölni kell a listából. A korábban komolyan érdeklődő potenciális ügyfél a következő hívás során 10% eséllyel alacsony érdeklődést fog mutatni, 40% valószínűséggel továbbra is komoly érdeklődő marad, míg 50% eséllyel megvásárolja a terméket.

- Mi az esélye, hogy egy potenciális vevő megveszi a terméket?
- A vállalat átlagosan hányszor fog felhívni egy potenciális vevőt?

**SZ1.) [10.13.]** Tekintsük a szimmetrikus bolyongást az  $n \geq 5$  hosszú körön. Tegyük fel, hogy az éleken oda-vissza lehet közlekedni. Add meg az átmenetmátrixot és osztályozd az állapotokat! Határozd meg a stacionárius eloszlást! (1p)

**SZ2.) [10.13.]** Legyen  $\{X_k, k \geq 0\}$  homogén Markov-lánc  $S = \{0, 1, 2\}$  állapottérrel,  $X_0 = 1$ , az átmenetvalószínűség mátrix  $\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0.1 & 0.9 \end{pmatrix}$ . Mutasd

meg, hogy az  $Y_k = I(X_k \geq 1)$  folyamat nem Markov-lánc! (2p)

**SZ3.) [10.13.]** Egy gépjármű-biztosítással foglalkozó vállalat ügyfeleinek az alapján határozza meg a biztosítási díjat, hogy volt-e balesetük az előző két évben. Ha egy ügyfélnek az elmúlt két évben nem volt balesete, akkor a biztosítási díj évi 100 ezer Ft. Ha a partnernek az elmúlt két év mindegyikében volt balesete, akkor évi 200 ezer Ft-ot fizet. Amennyiben az ügyfélnek az elmúlt két év egyikében volt balesete, akkor a biztosítási díj 150 ezer Ft. Ha a partnernek balesete volt az elmúlt évben, akkor 10% a valószínűsége, hogy az aktuális évben is balesete lesz. Ha az ügyfélnek nem volt balesete az elmúlt évben, akkor csak 3% eséllyel lesz balesete az aktuális évben. Tegyük fel, hogy a fenti számok hosszú távon stabilnak tekinthetők. A biztosítónak mennyi az összes ügyfélre vonatkozó átlagos biztosítási díja? (2p)

**Ismétlésre ajánlott:** VL-ből 54–74

**7.)** Az előző két hónap során megfigyeltem, hogy a sarki zöldségesben a nektarin kilóját a minőségétől függően változó áron árulják. A megfigyelt értékesítési árak: 800, 1400, 1000, 1200, 1600, 1400, 1000, 1600, 1600, 1500, 1200, 1300, 1400, 1300. Az teljesen kiszámíthatatlan, hogy amikor betérek a zöldségesbe, épp milyen mi-

nőségű nektarint árulnak. Összesen 2000 forintot szánok nektarinra, jelölje  $Y$  azt a valószínűségi változót, hogy hány kiló nektarint tudok vásárolni.

- Hogyan modellezzük a nektarin árát? Alkalmass statisztikai próba végrehajtásával is mutasd meg, hogy a választott eloszlás megfelelő!
- Várhatóan hány kiló nektarint fogok tudni venni? Adj 95%-os konfidenciaintervallum-bebecslést is erre a mennyiségre!

**8.)** Szandi minden nap villamossal és busszal közlekedik, hogy eljusson az egyetemre. Az alábbi ábra tartalmazza a közlekedési időket:

<b>Ott-</b>	$\xrightarrow{\text{gyalog}}$	Villamos-	$\xrightarrow{\text{villamossal}}$	$\xrightarrow{\text{gyalog}}$	Busz-	$\xrightarrow{\text{busszal}}$	$\xrightarrow{\text{gyalog}}$	<b>Egye-</b>
<b>hon</b>	5 p	megálló	10 p	2 p	megálló	8 p	5 p	<b>tem</b>

Tapasztalatai alapján átlagosan 2 percet vár a villamos megállójában és 4 percet a busz megállójában. Szandi ma reggel később ébredt fel, ezért csak 7:43-kor indult el otthonról, az első órája 8:15-kor kezdődik. Becsüljük különböző, értelmes valószínűségi modellek segítségével annak a valószínűségét, hogy el fog késni!

**9.)** Mutassuk meg, hogy  $U = X + Y$  és  $V = \frac{Y}{X+Y}$  függetlenek, és határozzuk meg az eloszlásukat, amennyiben

- $X \sim \text{Exp}(\lambda)$ ,  $Y \sim \text{Exp}(\lambda)$  függetlenek;
- $X \sim \Gamma(\alpha, \lambda)$ ,  $Y \sim \Gamma(\beta, \lambda)$  függetlenek.

**10.)** Legyen  $X$  és  $Y$  együttes sűrűségfüggvénye  $f_{X,Y}(x, y) = \frac{e^{-\frac{x}{y}-y}}{y} I(x > 0, y > 0)$ .

- $X$  és  $Y$  független egymástól?
- $P(X < 5Y) = ?$
- Milyen eloszlású  $X$  az  $Y = y$  feltétel mellett? Számítsuk ki az  $E(X|Y)$  feltételes várható értéket!

**11.)** Mely  $c$  valós paraméter esetén lesz kétdimenziós sűrűségfüggvény

$$f_{X,Y}(x, y) = \begin{cases} c & \text{ha } 0 < y < 2x < 2 \\ 0 & \text{egyébként} \end{cases} ?$$

- Adjuk meg a perem-sűrűségfüggvényeket és a kovarianciamátrixot!
- Generáljunk véletlen számokat ebből az eloszlásból, majd becsüljük meg a  $P(X > \frac{1}{2}, Y < 1)$  valószínűséget! Számoljuk is ki a pontos értéket!
- Számítsuk ki és becsüljük szimulációval az  $E(X|Y)$  feltételes várható értéket és a  $D^2(X|Y)$  feltételes szórásnégyzetet!
- Legyen  $U = X + Y$  és  $V = X - Y$ . Határozd meg az  $(U, V)$  valószínűségi vektorváltozó (együttes) sűrűségfüggvényét!

**12.)** Legyen  $c$  alkalmas valós szám,  $X$  és  $Y$  együttes sűrűségfüggvénye  $f_{X,Y}(x, y) = (cx + 1) \cdot I(x > 0, y > 0, x + y < 1)$ .

- $c = ?$
- $P(Y < 2X^2) = ?$
- $E(Y|X) = ?$

**13.)** Három nyugdíjas barátnő, Kató néni, Luca néni és Margó néni együtt mennek a városi nagypostára, hogy feladják a gázszámlájukat. Mivel délelőtt fél 10-kor, munkaidőben mennek, a 6 pultból mindössze 2-nél van ügyfél, ezért mindannyian egyszerre oda tudnak állni egy-egy postai ügyintézőhöz feladni a csekket. Tegyük fel, hogy a csekkfeladás (kiszolgálási) ideje exponenciális eloszlású 2 perc várható értékkel. Határozd meg annak a valószínűségét, hogy

- a.) Kató néni előbb végez az ügyintézésel, mint Luca néni;  
 b.) Kató néni végez az ügyintézésel elsőként, Luca néni másodikként, Margó néni pedig utolsóként!

**Gyakorlásra ajánlott:** VL2-ből 1–4, VL-ből 99–102, APZ-ből 6.28, 6.29

**B4.) [10.27.]** Juliska néni a városi piacon értékesíti almáját. Mind az értékesítési mennyiség, mint az ár véletlennek tekinthető (lehet vele alkudozni). Az egy nap alatt eladott mennyiség (kg) Pareto-eloszlású  $\frac{5}{4}$  és 10 paraméterekkel, míg az eladott almák ára (Ft/kg) egyenletes eloszlású 180 és 220 paraméterekkel. Tegyük fel, hogy a mennyiség és az ár függetlenek egymástól. Költségei naponta 4000 forintot tesznek ki (benzinköltség).

- a.) Várhatóan mennyi profitra fog szert tenni egy nap alatt?  
 b.) Határozd meg annak a valószínűségét, hogy Juliska néni napi profitja meghaladja az 5000 forintot! ( $1+3=4p$ )

**B5.) [11.03.]** Legyen  $(X, Y)$  sűrűségfüggvénye  $f(x, y) = 2I(0 < x < y < 1)$ . Határozd meg  $U = \frac{X}{Y}$  eloszlását!

*Útmutatás:* ehhez tekintsük a  $V = Y$  valószínűségi változót, majd határozd meg  $U$  és  $V$  együttes sűrűségfüggvényét!

**B6.) [11.03.]** Legyen  $(X, Y)$  sűrűségfüggvénye  $f(x, y) = e^{-y}I(0 \leq x \leq y)$ .

Számítsd ki az  $E(Y|X)$  feltételes várható értéket és a  $D(Y|X)$  feltételes szórását!

**SZ4.) [11.03.]** Legyenek  $X$  és  $Y$  egymástól független valószínűségi változók. Mutasd meg, hogy  $U = XY$  sűrűségfüggvénye  $f_U(u) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_X\left(\frac{u}{y}\right) f_Y(y) dy$ . (2p)

**SZ5.) [11.03.]** Legyenek  $X, Y \sim N(0, 1)$  eloszlású, egymástól független valószínűségi változók. Határozd meg  $U = \frac{X}{Y}$  sűrűségfüggvényét és várható értékét! Milyen nevezetes eloszlást követ  $U$ ? (2p)

**SZ6.) [11.03.]** Az Északi és a Déli tömb közötti távolságot egymástól függetlenül 4-en lemérik. Tegyük fel, hogy az egyes mérések eredményei közös, a valódi távolsággal megegyező várható értékű és véges szórású normális eloszlást követnek. Mi az esélye, hogy az első 3 mérés átlaga pontosabban közelíti a valódi távolságot, mint a 4. mérés? (2 pont)

**SZ7.) [11.03.] Box-Müller transzformáció.** Legyenek  $X_1, X_2 \sim E(0, 1)$  függetlenek. Legyenek  $Y_1 = \sqrt{-2 \log X_1} \cos(2\pi X_2)$ ,  $Y_2 = \sqrt{-2 \log X_1} \sin(2\pi X_2)$ . Mutassuk meg, hogy  $Y_1, Y_2 \sim N(0, 1)$  függetlenek! (2p)

**SZ8.) [11.03.]** Legyen  $f_{X,Y}(x, y) = \begin{cases} c & \text{ha } 0 < x < 2 \text{ és } -2 + x < y < 2 - x \\ 0 & \text{különben} \end{cases}$

Határozd meg a  $c$  értékét, majd az  $E(X|Y)$  feltételes várható értékeket! (2p)

**14.)** Legyen  $X$  és  $Y$  független standard normális eloszlású,  $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 2X + Y \\ -X + Y \end{pmatrix}$ .

- a.) Milyen eloszlást követ a  $V$  valószínűségi változó?  
 b.) Határozzuk meg  $U$  és  $V$  együttes eloszlását! Mi a kovarianciamátrix és az

együttes sűrűségfüggvény? Rajzoljuk ki az együttes sűrűségfüggvény szintvonalait!

c.) Határozzuk meg a  $P(U < V)$  és  $P(-1 < U < 1, 0 < V < 3)$  valószínűségeket!  
**15.)** Legyen  $(X, Y)$  kétdimenziós normális eloszlású. Határozzuk meg az  $X|Y$  feltételes eloszlást!

**16.)** Tegyük fel, hogy a magyar férfiak magassága és testsúlya kétdimenziós normális eloszlású. A férfiak átlagmagassága 178 cm, 9 cm szórással; átlagos testsúlyuk pedig 85 kg, 10 kg szórással. A magasság és a testtömeg közötti korreláció 0,7. Számoljuk ki pontosan és R-es szimulációval egyaránt az alábbi mennyiségeket.

- a.) Mi az esélye, hogy egy férfi 180 cm magas és 80 kiló?  
 b.) Mi az esélye, hogy egy férfi magasabb 180 cm-nél és 80 kilogramm a testsúlya?  
 c.) Mi az esélye, hogy egy férfi magasabb 180 cm-nél és könnyebb 80 kg-nál?  
 d.) Feltéve, hogy egy férfi 80 kg, mi a valószínűsége, hogy magasabb 180 cm-nél?  
 e.) Átlagosan mekkora súlyú egy 190 cm magas férfi?  
 f.) Átlagosan milyen magas egy 94,33 kg-os férfi?

**17.)** Generálj  $10^4$  darab véletlen számot az

- a.) inverziós módszerrel az  $f(x) = \frac{3x^2}{56}I(2 < x < 4)$  sűrűségfüggvényű eloszlásból;  
 b.) elutasításos módszerrel az  $f(x) = \frac{3}{8} \left(1 + \frac{x^2}{4}\right)^{-5/2}$ ,  $x \in \mathbb{R}$  sűrűségfüggvényű eloszlásból.

Ábrázold a véletlen számokat pontdiagramon és készíts szintvonalas ábrá(ka)t!

**18.)** Tekintsük az alábbi kétdimenziós sűrűségfüggvényt:

$$f_{X,Y}(x, y) = \begin{cases} \frac{10}{49}y & \text{ha } 1 < x^2 < y < 4, x > 0 \\ 0 & \text{egyébként} \end{cases}$$

Generáljunk véletlen vektorokat ebből az eloszlásból, majd készítsünk többdimenziós sűrűséghisztogramot, valamint ábrázoljuk a magfüggvényes becslést és a szintvonalakat!

**19.)** Generálj

- a.)  $N(1, 2^2)$  és  $N(4, 3^2)$  eloszlású valószínűségi változókat 0,9 korrelációval;  
 b.)  $N(1, 3^2)$ ,  $N(2, 2^2)$ ,  $N(3, 1^2)$  eloszlású valószínűségi változókat úgy, hogy az első kettő közti korreláció 0,8, a második és a harmadik közti 0,5, valamint az első és a harmadik közötti 0,2.

A véletlen számokból készíts többdimenziós sűrűséghisztogramot, valamint ábrázold a magfüggvényes becslést és a szintvonalakat!

**20.)** Állítsuk elő a következő többdimenziós eloszlások kopuláját és kopulasűrűségfüggvényét, illetve ábrázoljuk a kopula-sűrűségfüggvény szintvonalait.

- a.)  $F_{X,Y}(x, y) = \begin{cases} \frac{(x+1)(e^y-1)}{x+2e^y-1} & \text{ha } (x, y) \in [-1; 1] \times (0, \infty) \\ 1 - e^{-y} & \text{ha } (x, y) \in (1; \infty) \times (0, \infty) \\ 0 & \text{egyébként} \end{cases}$   
 b.)  $f_{X,Y}(x, y) = I(0 < y < 2x < 2)$

**21.)** Generáljunk egy 1000 elemű kétdimenziós mintát úgy, hogy a peremek standard normálisak legyenek, a köztük lévő korreláció pedig 0.3 legyen. Vegyünk be a min-

tába néhány kiugró (outlier) értéket, majd nézzük meg, ennek hatására mennyire változtak meg a Pearson-, Kendall- és Spearman-féle korrelációk!

Magyarázzuk az egyik dimenziót a másik dimenzió segítségével lineáris modellel és itt is vizsgáljuk meg az outlierok hatását!

- 22.) Tekintsük azt az Archimédészi kopulát, aminek kopulageneráló függvénye  $\varphi(u) = (-\log u)^\theta$ , ahol  $\theta \geq 1$ . Állítsd elő a kopulát és számold ki a Kendall-féle  $\tau$  értékét.
- 23.) Generáljunk 1000 elemű kétdimenziós mintákat Gauss, 3 szabadságfokú Student  $t$ , Gumbel, Clayton kopulákból
- különböző paraméterek esetén;
  - úgy, hogy megegyezik a Kendall-féle  $\tau$  korrelációjuk!
- Vessük őket össze a minták pontdiagramját!
- 24.) Generáljunk 1000 elemű kétdimenziós mintákat úgy, hogy a peremek eloszlása 1 paraméterű exponenciális és 5 szabadságfokú  $t$ -eloszlású legyen, a kopulájuk pedig  $\theta$  paraméterű Clayton-kopula legyen. Ábrázoljuk a minták pontdiagramját különböző  $\theta$  paraméterértékek esetén!
- 25.) Legyen  $X$  és  $Y$  lognormális eloszlású 2 és 0,6 paraméterekkel, kopulájuk pedig legyen Gauss, 3 szabadságfokú Student  $t$ , Gumbel, Clayton kopulákból úgy, hogy megegyezik a Kendall-féle  $\tau$  korrelációjuk! Számítsuk ki mindegyik esetben a  $P(X < 4, Y < 4)$ ,  $P(X > 13, Y > 13)$  és  $P(X < 4, Y > 13)$  valószínűségeket!
- 26.) Generáljunk 1000 elemű mintát 2 szabadságfokú Student  $t$  kopulából 0,7 korreláció esetén. Illesszünk Gauss, Student  $t$ , Gumbel és Clayton kopulákat a mintára (becsüljük meg a kopula paramétereit), majd végezzünk illeszkedésvizsgálatot!
- 27.) Legyen  $X$  és  $Y$  független  $t_d$  eloszlású,  $r = -0,8$ . Legyen  $Z = rX + \sqrt{1-r^2}Y$ . Teszteljük különböző  $d$  értékek esetén alkalmas statisztikai próbával, hogy  $X$  és  $Z$  kopulája  $d$  szabadságfokú  $t$ -kopula-e.
- 28.) Tekintsük a MagyarReszvenyek.xlsx fájlban lévő mintát, amely Budapesti Értéktőzsdén nyíltan kereskedett néhány kiemelt részvény árfolyamadatait tartalmazza 2017-től kezdődően.
- Olvassuk be az adatokat, hozzuk őket az elemzéshez alkalmas formátumokra!
  - Tekintsük az OTP és a MOL adatokat 2017-től 2022-ig (5 évnyi adatmennyiség).
    - Vizsgáljuk meg az adatokat a tanult vizuális módszerekkel és mutatókkal egyaránt, szükség esetén végezzünk alkalmas transzformációt az eredeti adatokon!
    - Illesszünk különböző kopulákat, vizsgáljuk meg a paraméterbecslés stabilitását tolóablakos módszerrel! Adjunk intervallumbecslést is a paraméter(ek)re!  
Tolóablakos módszer: paramétert becslünk egy bizonyos intervallumra, majd ezt az intervallumot kicsit eltoljuk előre az időben és újra paramétert becslünk, majd újra megismételjük. A végén a becsült paraméterek stabilitását megvizsgáljuk, ábrázoljuk.
    - Végezzünk illeszkedésvizsgálatot – melyik kopula illeszkedik elfogadható-

an az adatainkra?

- IV.) Illesszünk megfelelően illeszkedő eloszlásokat a peremekre! Próbálkozunk például átskálázott  $t$ -eloszlással, azaz ha  $X$   $t$ -eloszlást követ  $d$  szabadságfokkal, legyen  $Y = \sigma \cdot X$ ,  $\sigma > 0$ , ezt az  $Y$ -t nevezzük skálázott  $t$ -eloszlásnak (paraméterei:  $\sigma, d$ ). A becslést például elvégezhetjük ML-becsléssel, a loglikelihood-függvény  $(-1)$ -szeresét minimalizálhatjuk az R-ben beépített `nlm` rutinnal.
- V.) Becsüljük meg annak a valószínűségét, hogy egy nap alatt mind az OTP, mind a MOL részvények  $l$  %-ot esnek, ahol  $l = 1, 2, \dots, 10$ . Vessük össze az értékeket azzal, amikor (tévesen) normális kopulával és normális peremekkel dolgozunk!
- c.) Oldjuk meg az előző pontban kérdéseit az OTP, MOL, Richter adatokra együttesen!
- B7.) [11.10.] Legyen  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3 \left( \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \right)$ .
- Milyen eloszlású az  $\begin{pmatrix} X_3 \\ X_1 \end{pmatrix}$  valószínűségi vektorváltozó?
  - Határozd meg  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$  eloszlását, ahol  $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$  és  $\mathbf{b} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ .
  - $P(X_1 > 2, 0 < X_2 < 1, X_3 > 0) = ?$
  - $E(X_1 | X_3 = -1) = ?$
- B8.) [11.10.] Az emberek IQ-ja 100 pont várható értékű 15 szórással. A Mensa nevű egyesületbe bárki jelentkezhet, aki egy hivatalos IQ-teszten magas, 130 feletti pontot ér el. Zsuzsi és Dóri testvérek. Ha tudjuk, hogy Zsuzsi Mensa-tag lett 135 ponttal, akkor mi az esélye, hogy Dóri is sikerrel jelentkezhet a Mensába, amennyiben a két testvér intelligenciahányadosa közti korreláció  $-1 \leq \rho \leq 1$  valós szám? Az eredményt számold ki számítógép segítségével, készíts belőle alkalmas grafikont (x tengely:  $\rho$ ; y tengely: a valószínűség)! Értékeld a látottakat – erre számítottál?
- B9.) [11.17.] Tekintsük a következő sűrűségfüggvényt:  

$$f(x) = \frac{1}{2x\sqrt{\pi}} \exp \left\{ -\frac{(\log x - m)^2}{4} \right\} I(x \geq 0)$$
, ahol  $m$  valós paraméter.
- Vezesd le  $n$  elemű minta esetén a maximum likelihood becslést!
  - Szimuláció segítségével készíts 95%-os konfidenciaintervallumot a 100 elemű mintából készített ML-becslésre! A véletlen számok generálását alkalmas tanult módszerrel végezd el!  
*Útmutatás:* konfidenciaintervallum például úgy készíthető, ha sokszor, mondjunk 1000-szer állítasz elő 100 elemű mintát, amiből 1000 becslés készíthető az  $m$  paraméterre, ezekből pedig a megfelelő tapasztalati kvantiliseket kiszámolva kapható intervallumbecslés.
- B10.) [11.17.] Töltsd le a honlapomról a B9.r fájlt és futtasd le R/RStudio segítségével.

- a.) Írd le nagy vonalakban, mi történik a kódban! Értékelj ki az eredményeket!  
 b.) Módosítsd a `konturdiag` és `hist_2d` függvényeket úgy, hogy a `pontdiag` függvényhez hasonlóan az ábrának legyen címe!

**B11.) [12.01.]** Legyen  $\theta \in [-1; 1]$  valós paraméter. Mutassuk meg, hogy az  $F_{X,Y}(x, y) = (1 - e^{-x} - e^{-y} + e^{x+y+\theta xy})^{-1}$  együttes eloszlásfüggvénnyel definiált eloszlás kopulája  $C(u, v) = \frac{uv}{1-\theta(1-u)(1-v)}$ ! Van olyan  $\theta$ , amelyre  $X$  és  $Y$  független egymástól?

**B12.) [12.08.]** Legyen  $X$  és  $Y$  normális eloszlású 10 és 2 paraméterekkel, kopulájuk pedig legyen Clayton kopula úgy, hogy a Kendall-féle  $\tau$  korreláció értéke 0,5 legyen! Számítsd ki a  $P(X < 7, Y < 7)$ ,  $P(X < 13, Y > 7)$  és  $P(X > 12, Y > 13)$  valószínűségeket!

**SZ9.) [11.24.]** Legyen  $(X, Y, Z)^T$  együttes sűrűségfüggvénye  $f_{X,Y,Z}(x, y, z) = c \cdot \exp\{-\frac{1}{8}x^2 - \frac{1}{2}y^2 - \frac{1}{2}z^2 + \frac{1}{4}x + \frac{1}{2}yz - \frac{1}{8}\}$ . Határozd meg a  $c$  értékét és számítsd ki az  $E(Y|X=0, Z=0)$  feltételes várható értéket! (2p)

**SZ10.) [11.24.]** Oldd meg a 15 feladatot a normális korreláció tételének felhasználása nélkül! (2p)

**SZ11.) [11.24.]** Tegyük fel, hogy egy gyorsúszó alvásideje normális eloszlásúnak tekinthető 8 óra várható értékkel és 1 óra szórással. Tegyük fel továbbá, hogy amennyiben  $x$  órát alszik, akkor a 100 méter gyorsúszáson elért ideje ugyancsak normális eloszlású  $58 - x$  mp várható értékkel és 1 mp szórással. Számold ki annak a valószínűségét, hogy a gyorsúszó megdönti a 100 méter gyorsúszás 46,86 mp-es világcsúcsát! (2p)

**SZ12.) [12.08.]** Legyen  $X_1$  és  $X_2$  független azonos, abszolút folytonos eloszlású  $F$  eloszlásfüggvénnyel. Mutassuk meg, hogy  $(X_1^*, X_2^*)^T$  kopulája az alábbi függvény:

$$C(u_1, u_2) = 2\sqrt{u_1}F(\min(F^{-1}(1 - \sqrt{1 - u_1}), F^{-1}(\sqrt{u_2}))) - [F(\min(F^{-1}(1 - \sqrt{1 - u_1}), F^{-1}(\sqrt{u_2})))]^2,$$

ahol  $X_1^* = \min(X_1, X_2)$ ,  $X_2^* = \max(X_1, X_2)$ . (3p)

**SZ13.) [12.08.]** Határozd meg  $(Y, 1 - Y)^T$  kopuláját, ahol  $Y \sim E(0; 10)$ . (2p)

**SZ14.) [12.08.]** Legyenek  $X, Y$  és  $Z$  normális eloszlásúak 9 és 3 paraméterekkel, kopulájuk pedig legyen 5 szabadságfokú  $t$  kopula úgy, hogy a Kendall-féle  $\tau$  korreláció értéke 0,7 legyen! Számítsd ki a  $P(X < 5, Y < 5, Z < 5)$ ,  $P(X < 5, 5 < Y < 13, Z > 13)$  és  $P(X > 13, Y > 13, Z > 13)$  valószínűségeket! (2p)

**29.)** Olvassuk be a `kerdoiv.txt` fájlt, ami egy 2017-es hallgató kérdőíves felmérés adatait tartalmazza. A következőkre válaszoltak: nem, testmagasság (cm), súly (kg), cipőméret, hányast szerzett valszámból a 2017-es vizsgán, hány percet utazik az egyetemre, szorgalmi időszakban átlagosan hány órát tanul egy héten.

a.) Nézzük meg pontdiagrammal néhány adatpár közti összefüggést (pl. magasság és súly, nem és cipőméret, stb.)!

b.) A továbbiakban célunk a testmagasság modellezése/magyarázása a többi változó segítségével. Tekintsük az alábbi regressziós modelleket:

I.) **Testmagasság = Testsúly + Hiba**, ami a **Testmagasság =  $a_0 + a_1 \cdot$  Testsúly + Hiba** modell rövidített változata

II.) **Testsúly = Testmagasság + Hiba**

III.) **Testmagasság = Testsúly + Lábméret + Hiba**

IV.) **Testmagasság = Nem + Hiba**

Vizsgáljuk meg a korrelációs mátrixot! Keressük meg a legjobban illeszkedő modellt!

c.) Adjunk előrejelzést a legjobbnak tűnő modell(ek) alapján egy olyan fiú hallgató testmagasságára, aki 70 kg-os, 45-ös a cipőmérete, 5-öse volt valszámból, 25 percet utazik az egyetemre és heti 12 órát tanul!

d.) Szignifikáns hatással van a hallgatók neve a testmagasságukra?

e.) Szignifikáns hatással van a hallgatók tavalvi vizsgajegye a testmagasságukra?

**30.)** Egy ügyvédi vállalkozásnak két irodája van: egy Budapesten és egy Miskolcon. 2018-ban Budapesten 4-en dolgoztak, a bruttó fizetésük 500 e Ft, 600 e Ft, 550 e Ft, 510 e Ft volt. A miskolci telephely dolgozói bruttó 450 e, 350 e, 400 e Ft-ot vittek haza.

a.) A telephely hány %-ban magyarázza a fizetések változékonyságát?

b.) A telephely szignifikáns hatással van a fizetésekre?

c.) Hogyan változik az előző két kérdésre adott válasz, ha Győrben is van egy telephely, az ott dolgozók fizetése pedig 450 e Ft és 550 e Ft?

**31.)** A távolsági autóbusz-közlekedés néhány jellemző adata egy év során:

Járat	Szállított utasok száma (fő)	Utazási távolság (km/fő)	
		átlaga	szórása
Menetrend szerinti	450	17	5
Szerződéses	30	23	15
Külön	3	151	70
Összesen/együtt	...	...	...

a.) Számítsd ki a táblázat kipontozott celláit!

b.) A járat fajtája szignifikáns hatással van az utazási távolságra?

**32.)** A következő táblázat a 2016. 1. félévben tanár szakos BSc-s hallgatóknak tartott 4 valszám gyakorlat év végi, 100-ra skálázott végső pontszámait tartalmazza:

Gyakvezér	Pontszámok											
Cs. V.	98	87	102	92	52	46	95	60	81	55	60	94
	81	58	80	93	70	66	49	94	50	88	74	
W. G.	77	46	54	57	50	45	39	63	26	107	75	
	66	52	109	91	35	65						
B. Á.	86	94	54	61	42	59	88	81	81	80	102	72
	88	96	58	90	110	58	80	90	84	80	94	
V. L.	66	60	72	49	52	54	80	56	36	91	68	
	60	51	40	38	54	62						

Vizsgáljuk meg, az év végi pontszám függ-e attól, hogy a hallgató melyik csoportba jár! Hány %-ban magyarázza a pontszámok változékonyságát az, hogy a hallgatók melyik csoportba járnak?

**33.)** Tekintsük a <https://stats.idre.ucla.edu/stat/data/binary.csv> linken elérhető amerikai egyetemi felvételi adatokat és próbáljuk meg modellezni a felvétel eredményét (valószínűségét) a rendelkezésre álló magyarázó változókkal! Értelmezzük a kapott eredményeket!

A modell alapján milyen eséllyel vesznek fel egy diákot, aki egy 1-es rangú iskolából jön, a jegyeinek átlaga 3,8 és a felvételin 790 pontot ért el?