HATÁRELOSZLÁSTÉTEL POISSON HATÁRELOSZLÁSSAL

A centrális határeloszlástételhez hasonlóan létezik határeloszlástétel Poisson határeloszlással. Ennek általános alakját szintén szériasorozatokra érdemes megfogalmazni. Ez a következőképpen szól.

Határeloszlástétel Poisson határeloszlással. *Legyen adva egy*

$$\begin{array}{l} \xi_{1,1} \dots, \xi_{1,n_1} \\ \vdots \qquad \vdots \\ \xi_{k,1} \dots, \xi_{k,n_k} \\ \vdots \qquad \vdots \end{array}$$

szériasorozat, azaz $\xi_{k,j}$, $k = 1, 2, \dots,$ $1 \le j \le n_k$ valószínűségi változók egy olyan rendszere, amelyben az egy sorban álló (azaz ugyanolyan k index-szel rendelkező) valószínűségi változók függetlenek. Tegyük fel továbbá, hogy e valószínűségi változók teljesítik a következő feltételeket is:

1.) A $\xi_{k,j}$ valószínűségi változók nem negatív egész értékeket vesznek fel.

2.) $P(\xi_{k,j} = 1) = \lambda_{k,j}$, $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} \lambda_{k,j} = \lambda > 0$.

3.) $\sup\limits_{1\le j\le n_k} \lambda_{k,j} \to 0$, ha $k \to \infty$, és $\sum\limits_{j=1}^{n_k} P(\xi_{k,j} \ge 2) \to 0$, ha $k \to \infty$.

Ekkor az $S_k = \sum\limits_{j=1}^{n_k} \xi_{k,j}$ véletlen összegek eloszlásban konvergálnak a λ paraméterű Poisson eloszláshoz, ha $k \to \infty$, azaz $\lim\limits_{k\to\infty} P(S_k = l) = \frac{\lambda^l}{l!} e^{-\lambda}$ minden $l = 0, 1, 2, \dots$ számra.

Megjegyzés: Értsük meg a feladat feltételeit. Nem negatív egész értékű valószínűségi változókból álló szériasorozatot vettünk, amely teljesít egy az egyenletes kicsiséghez hasonló feltételt. A 3.) feltétel alapján $\lim\limits_{k\to\infty} \sup\limits_{1\le j\le n_k} P(\xi_{k,j} > 0) = 0$, ami hasonlít az egyenletes kicsiség feltételéhez, bár megengedi, hogy nagyon kis valószínűséggel a $\xi_{k,j}$ valószínűségi változók olyan nagy értékeket vegyenek fel, ami miatt a szórásnégyzetük nagy.

Ugyanakkor, mint azt a centrális határeloszlástétel egyik példájának tárgyalásában láttuk a (normalizált) $\xi_{k,j} - E\xi_{k,j}$ valószínűségi változók a 2.) tulajdonság miatt nem teljesítik a Lindeberg feltételt. Ezért rájuk nem érvényes a centrális határeloszlástétel. A 3.) feltétel alapján mind a $P(\xi_{k,j} = 1)$ mind a $P(\xi_{k,j} \ge 2)$ valószínűségek nagyon kicsik nagy k indexekre. De míg a $P(\xi_{k,j} \ge 2)$ valószínűségek 'másodrendben' kicsik a $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} P(\xi_{k,j} \ge 2) = 0$ feltétel alapján a $P(\xi_{k,j} = 1)$ valószínűségek csak 'elsőrendben' kicsik a $\lim_{k\to\infty} \sum_{j=1}^{n_k} \lambda_{k,j} = \lambda > 0$ feltétel miatt. A tétel azt mondja ki, hogy ilyen feltételek teljesülése esetén az $S_k = \sum_{j=1}^{n_k} \xi_{k,j}$ véletlen összegek egy λ paraméterű Poisson eloszlású valószínűségi változóhoz konvergálnak eloszlásban.

Jegyezzük meg, hogy bár formális értelemben az eloszlások konvergenciáját másképpen definiáltuk, abban a speciális esetben, ha mindegyik S_k, $k = 0, 1, 2, \dots$, valószínűségi változó egész értéket vesz fel az S_k valószínűségi változók eloszlásban való konvergenciája az S_0 valószínűségi változóhoz $k \to \infty$ esetben ekvivalens azzal, hogy a $\lim_{k\to\infty} P(S_k = n) = P(S_0 = n)$ reláció teljesül minden n egész számra. Ezért a kimondott határeloszlástétel bizonyításához elég azt megmutatni, hogy az S_k véletlen összegek karakterisztikus függvényei konvergálnak egy λ paraméterű Poisson eloszlású S_0 valószínűségi változó karakterisztikus függvényéhez. Ennek bizonyítása azon alapul, hogy jó aszimptotikus formulát adunk az $Ee^{it\xi_{k,j}}$ karakterisztikus függvények értékeire.

A tétel bizonyítása. Egy λ paraméterű Poisson eloszlású S_0 valószínűségi karakterisztikus függvénye

$$Ee^{itS_0} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} e^{ijt} = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda e^{it})^j}{j!} = e^{\lambda(e^{it}-1)}.$$

Másrészt $Ee^{it\xi_{k,j}} = P(\xi_{k,j} = 0) + e^{it} P(\xi_{k,j} = 1) + \varepsilon_{k,j}(t)$, ahol $\varepsilon_{k,j}(t) = \sum_{l=2}^{\infty} P(\xi_{k,j} = l)e^{ilt}$. Vezessük be a $\delta_{k,j} = P(\xi_{k,j} \ge 2)$ jelölést. Mivel $|\varepsilon_{k,j}(t)| \le P(\xi_{k,j} \ge 2) = \delta_{k,j}$, $P(\xi_{k,j} = 0) = 1 - P(\xi_{k,j} = 1) - P(\xi_{k,j} \ge 2) = 1 - \lambda_{k,j} - \delta_{k,j}$, ezért $Ee^{it\xi_{k,j}} = 1 - \lambda_{k,j} + e^{it}\lambda_{k,j} + \eta_{k,j}(t)$ valamely $\eta_{k,j}(t)$ mennyiséggel, amelyre $|\eta_{k,j}(t)| \le 2\delta_{k,j}$. Mivel $|\log(1+z) - z| \le 2|z|^2$, ha $|z| \le \frac{1}{10}$ (tetszőleges z, $|z| \le \frac{1}{10}$ *komplex* számra), és $\lim_{k\to\infty} \sup_{1\le j\le n_k} \lambda_{k,j} = 0$, $\lim_{k\to\infty} \sup_{1\le j\le n_k} \delta_{k,j} = 0$ az előző becslések alapján felírhatjuk, hogy

$$|\log Ee^{it\xi_{k,j}} - (Ee^{it\xi_{k,j}} - 1)| \le 2|Ee^{it\xi_{k,j}} - 1|^2,$$

és mivel $|(Ee^{it\xi_{k,j}} - 1) - \lambda_{k,j}(e^{it} - 1)| \le 2\delta_{k,j}$

$$|\log Ee^{it\xi_{k,j}} - \lambda_{k,j}(e^{it} - 1)| \le 2\delta_{k,j} + 2|Ee^{it\xi_{k,j}} - 1|^2 \le 8\lambda_{k,j}^2 + 4\delta_{k,j} \tag{1}$$

minden $1 \le j \le n_k$ indexre, ha $k \ge k_0$ valamilyen alkalmas k_0 index-szel. Az utolsó egyenlőtlenségben azt használtuk ki, hogy

$$|Ee^{it\xi_{k,j}} - 1|^2 \le 2|(Ee^{it\xi_{k,j}} - 1) - \lambda_{k,j}(e^{it} - 1)|^2 + 2|\lambda_{k,j}(e^{it} - 1)|^2 \le 4\delta_{k,j}^2 + 8\lambda_{k,j}^2.$$

Az (1) egyenlőtlenséget összeadva minden $1 \le j \le n_k$ indexre azt kapjuk, hogy nagy k indexekre

$$\left| \sum_{j=1}^{n_k} \log Ee^{it\xi_{k,j}} - (e^{it} - 1) \sum_{j=1}^{n_k} \lambda_{k,j} \right| \le \sum_{j=1}^{n_k} (4\lambda_{k,j}^2 + 8\delta_{k,j}).$$

Mivel $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}\delta_{k,j}=0$ és $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}\lambda_{k,j}^2\le\lim\limits_{k\to\infty}\sup\limits_{1\le j\le n_k}\lambda_{k,j}\cdot\sum\limits_{j=1}^{n_k}\lambda_{k,j}=0$, ezért

$$\left|\sum_{j=1}^{n_k}\log Ee^{it\xi_{k,j}}-(e^{it}-1)\sum_{j=1}^{n_k}\lambda_{k,j}\right|\to 0,\quad \text{ha } k\to\infty.$$

Továbbá, mivel $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}\lambda_{k,j}=\lambda>0$, ezért

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}\log Ee^{it\xi_{k,j}}=\lambda(e^{it}-1),\quad \text{ha } k\to\infty.$$

Ez azt jelenti, hogy $\lim\limits_{k\to\infty}Ee^{itS_k}=\lim\limits_{k\to\infty}\prod\limits_{j=1}^{n_k}Ee^{it\xi_{k,j}}=e^{\lambda(e^{it}-1)}=Ee^{itS_0}$, és ezt kellett belátni.

Megjegyzés: A bevezető valószínűségszámítás előadásban a fenti tétel egy gyengített változatát a generátorfüggvény módszer segítségével bizonyítottuk be. A karakterisztikus függvény és generátor függvény módszer alkalmazása között (nem negatív egész értékű valószínűségi változók vizsgálatában) szoros kapcsolat van. Valóban, legyen ξ nem negatív egész értékű valószínűségi változó, $P(\xi=k)=p_k$, $0\le k<\infty$, $\sum\limits_{k=0}^{\infty}p_k=1$. Ekkor a ξ valószínűségi változó generátorfüggvénye $P(x)=\sum\limits_{k=0}^{\infty}p_kx^k$, karakterisztikus függvénye pedig $\varphi(t)=\sum\limits_{k=0}^{\infty}p_ke^{ikt}$. Innen $\varphi(t)=P(e^{it})$. Tehát nem negatív egész értékű valószínűségi változók karakterisztikus függvénye megegyezik e valószínűségi változók generátorfüggvényének az értékével a komplex egységkörön. (A $\varphi(t)$ szám a generátorfüggvény értékével egyenlő az abszcissza tengellyel t szöget bezáró komplex egységvektoron.)

A centrális és többdimenziós centrális határeloszlástétellel kapcsolatos feladatok.

A többdimenziós normális eloszlás definiciója szerint egy n-dimenziós $\xi=(\xi_1,\dots,\xi_n)$ standard normális eloszlású vektorra és egy $n\times n$ méretű A mátrixra ξA (nulla várható értékű, n-változós) normális eloszlású vektor. Meg akarjuk mutatni, hogy igaz a következő általánosabb állítás.

Állítás. *Tetszőleges n-dimenziós $\xi=(\xi_1,\dots,\xi_n)$ normális eloszlású vektorra és $m\times n$ méretű A mátrixra ξA (nulla várható értékű, m-változós) normális eloszlású vektor.*

Ezt meg lehet mutatni a ξA karakterisztikus függvényének a kiszámolásával, de lássuk be ezt az állítást közvetlenül lineáris algebrai módszerekkel.

Az állítás bizonyítása. Feltehetjük, hogy ξ n-dimenziós standard normális eloszlású valószínűségi változó, mert eloszlása megegyezik egy $\bar{\xi} = \eta B$ véletlen vektoréval, ahol η standard normális, és B $n \times n$-es mátrix. Ezért ξA eloszlása megegyezik $\bar{\xi} A = \eta BA$ eloszlásával, és elég az utóbbi véletlen vektor normalitását belátni.

Ha $m = n$, akkor ez az állítás a normális eloszlás definiciójának a közvetlen következménye. Az $m > n$ eset szintén egyszerű. Ekkor ugyanis a ξ véletlen vektort kiegészíthetjük egy $\tilde{\xi} = (\xi_1, \dots, \xi_m)$ m-dimenziós standard normális eloszlású véletlen vektorrá, A-t pedig egy $m \times m$ méretű $\tilde{A}$ mátrixra az utolsó $m - n$ sor minden elemét nullának választva. Ekkor $\xi A = \tilde{\xi} \tilde{A}$ m-dimenziós normális eloszlású vektor. Az $m < n$ eset vizsgálata érdekében igazolunk egy önmagában is érdekes és hasznos állítást. Ennek érdekében definiáljuk a következő Euklideszi teret.

Legyen ξ_1, ξ_2, ... ,ξ_n n független, standard normális eloszlású valószínűségi változó. Tekintsük az összes $\eta = \sum\limits_{j=1}^{n} a_j \xi_j$ valószínűségi változót, ahol $a_1, \dots, a_n$ tetszőleges valós számok. Az így definiált valószínűségi változók egy $\mathcal{N}$ Euklideszi teret alkotnak, ha e tér két $\eta = \sum\limits_{j=1}^{n} a_j \xi_j \in \mathcal{N}$, $\bar{\eta} = \sum\limits_{j=1}^{n} \bar{a}_j \xi_j \in \mathcal{N}$, elemének az $A\eta + B\bar{\eta}$ lineáris kombinációját az $A\eta + B\bar{\eta} = \sum\limits_{j=1}^{n} (Aa_j + B\bar{a}_j)\xi_j)$ képlettel definiáljuk, skalárszorzatukat pedig az $(\eta, \bar{\eta}) = E\eta\bar{\eta} = \sum\limits_{j=1}^{n} a_j \bar{a}_j$ formulával.

Lemma. *Az előbb definiált $\mathcal{N}$ Euklideszi tér elemei nulla várható értékű normális eloszlású valószínűségi változók. Továbbá, ha adva van k darab $\eta_j \in \mathcal{N}$, $1 \le j \le k$, korrelálatlan, azaz olyan valószínűségi változó, amelyekre $E\eta_i\eta_j = 0$, ha $1 \le i, j \le k$, és $i \ne j$, akkor ezek az $\eta_1, \dots, \eta_k$ valószínűségi változók függetlenek (és normális eloszlásúak).*

Bizonyítás. Tanultuk, hogy független normális eloszlású valószínűségi változók összege is normális eloszlású. Ezért minden $\eta = \sum\limits_{j=1}^{n} a_j \xi_j$ alakú, azaz $\eta \in \mathcal{N}$ valószínűségi változó normális eloszlású és nulla várható értékű. Amikor be akarjuk látni, hogy az $\eta_j \in \mathcal{N}$, $1 \le j \le k$, és $E\eta_i\eta_j = 0$, $1 \le i, j \le k$, $i \ne j$ feltételeket teljesítő valószínűségi változók függetlenek azt is feltehetjük (az η_j vektorokat alkalmas konstanssal megszorozva), hogy $E\eta_j^2 = 1$, $1 \le j \le k$. Ekkor az $\eta_j \in \mathcal{N}$, $1 \le j \le k$, vektorokat kiegészíthetjük egy ortonormált $\eta = (\eta_1, \dots, \eta_n)$ bázissá a $\mathcal{N}$ térben. Ekkor viszont $\eta = \xi U$ egy unitér U mátrix-szal, és mint tanultuk, ebből következik, hogy η standard normális eloszlású véletlen vektor, tehát koordinátái függetlenek.

Az állítás bizonyítása az $m < n$ esetben. Tekintsük a $\zeta = \xi A$ vektor $(\zeta_1, \dots, \zeta_m)$ koordinátáit, mint az előbb definiált $\mathcal{N}$ Euklideszi tér elemeit. Tekintsük az $\mathcal{N}$ tér $\zeta_1, \dots, \zeta_m$ vektorok által kifeszített lineáris alterét, és abban egy $\eta_1, \dots, \eta_{m'}$, $m' \le m$ ortonormált bázist. Ekkor $\eta = (\eta_1, \dots, \eta_{m'})$ egy standard normális eloszlású vektor, és $\zeta = \eta A'$ alkalmas A' $m' \times m$ méretű mátrix-szal. Innen következik, hogy a ζ vektor is normális eloszlású.

Feladatok:

1.) Legyen $\xi_1,\dots,\xi_n$ n-változós normális eloszlású véletlen vektor, és k, $1\le k\le n$, egész szám. Mutassuk meg, hogy $(\xi_1,\dots,\xi_k)$ k-változós normális eloszlású véletlen vektor.

Megoldás: $(\xi_1,\dots,\xi_n)$ azonos eloszlású egy $(\bar\xi_1,\dots,\bar\xi_n)=\zeta A+(m_1,\dots,m_n)$ vektorral, ahol $\zeta=(\zeta_1,\dots,\zeta_n)$ n-változós standard normális eloszlású vektor, A alkalmas $n\times n$ méretű mátrix, és $(m_1,\dots,m_n)$ alkalmas n-dimenziós (determinisztikus) vektor. Jelölje B azt a $k\times n$ méretű mátrixot, amelyet úgy kapunk, hogy az A mátrixnak csak az első k oszlopát őrizzük meg. Ekkor a $(\xi_1,\dots,\xi_k)$ vektor azonos eloszlású a $(\zeta_1,\dots,\zeta_n)B+(m_1,\dots,m_k)$ vektorral. Innen, illetve a feladatsor előtt megfogalmazott Állításból következik a feladat állítása.

2.) Legyenek $\xi_1,\xi_2,\dots,\xi_n$ független normális eloszlású valószínűségi változók m várható értékkel és σ^2 szórással. Ekkor a $\bar\xi=\frac1n\sum\limits_{j=1}^n\xi_j$ és az $S_n=\sum\limits_{j=1}^n(\xi_j-\bar\xi)^2$ valószínűségi változók függetlenek egymástól. Továbbá $\frac{\sqrt n}{\sigma}(\bar\xi-m)$ standard normális eloszlású, $\frac1\sigma S_n$ pedig $n-1$ szabadságfokú χ^2 eloszlású valószínűségi változó. (Ez az eredmény magyarázza meg, hogy bizonyos statisztikai feladatokban miért jelenik meg az U-eloszlás, amely egy standard normális és egy tőle független χ^2 eloszlású valószínűségi változó hányadosának az eloszlása.)

Megoldás: Tekintsük a $(\bar\xi,\xi_1-\bar\xi,\dots,\xi_n-\bar\xi)$ véletlen vektort. Ez az előbb tárgyalt eredmény alapján egy $n+1$ dimenziós normális eloszlású véletlen vektor. Továbbá egyszerű számolás mutatja, hogy $\mathrm{Cov}\,(\bar\xi,\xi_j-\bar\xi)=\mathrm{Cov}\,(\bar\xi,\xi_j)-\mathrm{Var}\,\bar\xi=0$ minden $1\le j\le n$ indexre. Ezért a $(\bar\xi,\xi_1-\bar\xi,\dots,\xi_n-\bar\xi)$ véletlen vektor $D=(d_{i,j})$, $1\le i,j\le n+1$, covariancia mátrixa felbomlik egy 1×1-es és $n\times n$-es mátrix direkt szorzatára, azaz $d_{1,j}=d_{j,1}=0$, $2\le j\le n+1$. Ezért a normális eloszlású vektorok tulajdonságaiból, (abból, hogy egy normális eloszlású véletlen vektor eloszlását meghatározza a véletlen vektor kovariancia mátrixa és várható érték vektora) következik, hogy $\bar\xi$ és $(\xi_1-\bar\xi,\dots,\xi_n-\bar\xi)$ függetlenek, továbbá normális eloszlásúak. Ezért a $\bar\xi$ és az $S_n=\sum\limits_{j=1}^n(\xi_j-\bar\xi)^2$ valószínűségi változók függetlenek egymástól. A χ^2 statisztikáról tanultak alapján S_n eloszlása megegyezik egy $\chi^2(n-1)$ eloszlású valószínűségi változó eloszlásával megszorozva a σ számmal. Ez az eredmény megkapható, a χ^2 statisztikáról szóló eredmény speciális eseteként, ha felírjuk a $(\xi_1-\bar\xi,\dots,\xi_n-\bar\xi)$ vektor kovariancia mátrixát. E vektor kovarianciamátrixa ugyanis $\mathrm{Cov}\,(\xi_i-\bar\xi,\xi_j-\bar\xi)=\frac{n-2}{n^2}\mathrm{Var}\,\xi_1-2\frac{n-1}{n^2}\mathrm{Var}\,\xi_1=-\frac1n\mathrm{Var}\,\xi_1$, ha $i\ne j$, és $\mathrm{Var}\,(\xi_i-\bar\xi)=\frac{n-1}{n^2}\mathrm{Var}\,\xi_1+(1-\frac1n)^2\mathrm{Var}\,\xi_1=(1-\frac1n)\mathrm{Var}\,\xi_1$.

A következő az előadásban bebebizonyított eredményt alkalmazzuk. Egy k-változós, nulla várható értékű és $(d(,j))$, kovariancia mátrixú normális vektor koordinátáinak négyzetösszege χ-négyzet eloszlású $k-1$ szabadságfokkal, ha $d(i,j)=-\sqrt{p_i}\sqrt{p_j}$, ha $i\ne j$, és $d(i,i)=1-p_i$, továbbá $1\le p_i\le1$, $1\le i\le k$, és $\sum\limits_{i=1}^k p_i=1$, Ezt az eredményt alkalmazzuk $n=k$, $p_i=\frac1n$ választással.

3.) Legyenek $\xi_1, \dots, \xi_n$ független, a $\left[-\frac{1}{2}, \frac{1}{2}\right]$ intervallumban egyenletes eloszlású valószínűségi változók. Mutassuk meg, hogy a $\sum_{j=1}^n \xi_j$ és $\sum_{j=1}^n \xi_j^2$ összegek normalizáltjainak, azaz a $\sqrt{\frac{12}{n}} \sum_{j=1}^n \xi_j$ és $\sqrt{\frac{180}{n}} \sum_{j=1}^n \left(\xi_j^2 - \frac{1}{12}\right)$ valószínűségi változóknak az együttes eloszlása a két-dimenziós standard normális eloszláshoz konvergál, ha $n \to \infty$.

Megoldás: $E\xi = 0$, $E\xi^2 = \frac{1}{12}$, $\operatorname{Var}\xi = \frac{1}{12}$, $\operatorname{Var}\xi^2 = E\xi^4 - (E\xi^2)^2 = \frac{1}{80} - \frac{1}{144} = \frac{1}{180}$. Továbbá $\operatorname{Cov}(\xi, \xi^2) = E\xi^3 - E\xi E\xi^2 = 0$. Ezért a $\left(\sqrt{12}\xi_j, \sqrt{180}\left(\xi_j^2 - \frac{1}{12}\right)\right)$, $j = 1, 2, \dots$, véletlen vektorok függetlenek, nulla várható értékkel és az identitás kovariancia mátrix-szal. Innen, és a több-dimenziós centrális határeloszlástételből következik a feladat állítása.

4.) Egy népszavazási kérdést a választáson akkor fogadnak el, ha egyrészt többen szavaztak rá igennel, mint nemmel, másrészt az igen szavazatok száma meghaladja az összes választópolgárok számának a 20%-át. Legyen mondjuk, $n = 5000000$ választópolgár, mindenki egymástól függetlenül 40% valószínűséggel megy el szavazni, és 50% valószínűséggel szavaz igennel. Mi a valószínűsége annak, hogy a népszavazás eredményeként elfogadják a népszavazási kérdést?

Megoldás: Vezessük be a következő (kétváltozós) vektorértékű (ξ_j, η_j) valószínűségi változókat. $\xi_j = 1$, ha a j-ik szavazó igennel szavaz, $\xi_j = -1$, ha nemmel szavaz, és $\xi_j = 0$, ha nem megy el szavazni. $\eta_j = 1$, ha a j-ik szavazó igennel szavaz, $\eta_j = 0$ egyébként, tehát ha nemmel szavaz vagy ha nem megy el szavazni. Legyen $S_n = \sum_{j=1}^n \xi_j$, $T_n = \sum_{j=1}^n \eta_j$. Ekkor minket annak a valószínűsége érdekel, hogy mind az $S_n > 0$ mind a $T_n > 0.2n$ esemény bekövetkezik. Erre a kérdésre a többváltozós centrális határeloszlástétel segítségével tudunk válaszolni, ha azt a (ξ_j, η_j) véletlen vektorok összegére alkalmazzuk. Ekkor $E\xi_j = 0$, $E\eta_j = 0.2$, $\operatorname{Var}\xi_j = E\xi_j^2 = 0.4$, $\operatorname{Var}\eta_j = E\eta_j^2 - (E\eta_j)^2 = 0.2 - 0.2^2 = 0.16$, $\operatorname{Cov}(\xi_j, \eta_j) = E\xi_j\eta_j - E\xi_j E\eta_j = E\xi_j\eta_j = P(\xi_j = 1, \eta_j = 1) = 0.2$. (A számolás utolsó lépésében kihasználtuk, hogy η_j két értéket vesz fel. Továbbá, ha $\eta_j = 0$, akkor $\xi_j\eta_j = 0$, és ha $\eta_j = 1$, akkor a ξ_j és η_j definiciója szerint $\xi_j = 1$.) Innen $\lim_{n\to\infty} P(S_n > 0, T_n > 0.2n) = \lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}} S_n > 0, \frac{1}{\sqrt{n}}(T_n - ET_n) > 0\right)$, és ez azon (X, Y) normális eloszlású véletlen vektor által meghatározott $P(X > 0, Y > 0$ valószínűséghez tart, amelyre $EX = EY = 0$, kovariancia mátrixát pedig a $\operatorname{Var}X = 0.4$, $\operatorname{Var}Y = 0.16$, $\operatorname{Cov}(X, Y) = 0.2$ képletek határozzák meg.

Megjegyzés: Az ebben a feladatban kapott valószínűség valamivel nagyobb, mint 0.25, azaz az az érték, ami akkor jelenne meg, ha a határeloszlásban megjelenő normális eloszású véletlen vektor X és Y koordinátái korrelálatlanok, ezért függetlenek lennének. 0.25 annak az (aszimptotikus) valószínűsége, hogy legalább a szavazatok 40%-át leadták, és az igen szavazatok voltak többségben. De pozitív valószínűséggel az is bekövetkezhet, hogy az össz-szavazatok száma valamivel kevesebb, mint a szavazásra jogosultak számának 40%-a, de ezen belül a többségben levő igennel szavazók száma meghaladja a szavazásra jogosultak számának a 20%-át.

5.) Számoljuk ki az előző feladat megoldásában megjelenő határeloszlás sűrűségfüggvényét, azaz egy olyan normális eloszlású (ξ, η) véletlen vektor sűrűségfüggvényét, amelyre $E\xi = E\eta = 0$, $\operatorname{Var} X = 0.4$, $\operatorname{Var} Y = 0.16$, $\operatorname{Cov}(X, Y) = 0.2$.

Megoldás: Számoljuk ki a tekintett véletlen vektor kovariancia mátrixának a determinánsát és inverzét. A determináns értéke $\operatorname{Var}\xi \operatorname{Var}\eta - (\operatorname{Cov}(\xi,\eta))^2 = 0.4 \cdot 0.16 - 0.2^2 = 0.024$, az inverz mátrix olyan $D = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$ mátrix amelyre $A = \frac{0.16}{0.024} = \frac{160}{24} = \frac{20}{3}$, $C = \frac{0.4}{0.024} = \frac{50}{3}$, $B = -\frac{0.2}{0.024} = -\frac{25}{3}$. Innen a keresett sűrűségfüggvény értéke

$$f(x,y) = \frac{1}{2\pi\sqrt{\det D}} e^{-(Ax^2+Cy^2+2Bxy)/2} = \frac{\sqrt{15}}{50\pi} \exp\left\{-\frac{5}{3}(2x^2+5y^2-5xy)\right\}.$$

Innen az is következik, hogy az előző feladat megoldását megadó valószínűség közelítőleg $\int_0^\infty \int_0^\infty f(x,y)\,dx\,dy$ a fenti $f(x,y)$ sűrűségfüggvénnyel.

6.) Egy termék megvásárlásakor kapunk egy kupont. Összesen n különböző kupon létezik, és az egyes vásárlások során egymástól függetlenül $\frac{1}{n}$ valószínűséggel kapjuk meg valamelyik kupont. Egymás után sok alkalommal vásároljuk meg ezt a terméket. Jelölje $S_1 = S_1^{(n)}$ azt a valószínűségi változót, hogy hanyadik vásárlásnál gyüjtöttük össze a kuponok felét, azaz $\frac{n}{2}$ különböző kupont, (legyen n páros szám), és $S_2 = S_2^{(n)}$ azt a valószínűségi változót, hogy hanyadik vásárlásnál gyüjtöttük össze az összes kupont. Számoljuk ki $S_1 = S_1^{(n)}$ és $S_2 = S_2^{(n)}$ várható értékét és szórásnégyzetét.

Megoldás: Jelölje U_1 azt, hogy hanyadik vásárlásnál kaptuk az első, U_2 azt, hogy hanyadik vásárlásnál kaptuk a második, és így tovább U_k azt, hogy hanyadik vásárlásnál kaptuk a k-ik kupont, $1 \le k \le n$. Ekkor a $\xi_k = U_k - U_{k-1}$, $1 \le k \le n$, (az $U_0 = 0$ jelöléssel) független geometriai eloszlású valószínűségi változók, és ξ_k eloszlásának a paramétere $p = \frac{n-k+1}{n}$. Továbbá $S_1^{(n)} = \sum\limits_{k=1}^{n/2} \xi_k$, $S_2^{(n)} = \sum\limits_{k=1}^{n} \xi_k$. Ezért a bevezető előadásban a geometriai eloszlás várható értékńek és szórásnégyzetének kiszámolásáról szóló eredmény alapján $ES_1^{(n)} = \sum\limits_{k=1}^{n/2} \frac{n}{n-k+1} = n \sum\limits_{k=n/2+1}^{n} \frac{1}{k}$, $ES_2^{(n)} = \sum\limits_{k=1}^{n} \frac{n}{n-k+1} = n \sum\limits_{k=1}^{n} \frac{1}{k}$, $\operatorname{Var} S_1^{(n)} = \sum\limits_{k=1}^{n/2} \frac{\frac{k-1}{n}}{\frac{(n-k+1)^2}{n^2}} = n \sum\limits_{k=1}^{n/2} \frac{k-1}{(n-k+1)^2}$, és

$$\operatorname{Var} S_2^{(n)} = \sum_{k=1}^{n} \frac{\frac{k-1}{n}}{\frac{(n-k+1)^2}{n^2}} = n \sum_{k=1}^{n} \frac{k-1}{(n-k+1)^2}.$$

Megjegyzés: Vegyük észre, hogy $ES_1^{(n)} \sim n \int_{n/2}^{n} \frac{1}{x}\,dx = n \log 2$, míg $ES_2^{(n)} \sim n \log n$. Hasonlóan, $\operatorname{Var} S_1^{(n)} \sim n \int_0^{n/2} \frac{x}{(n-x)^2}\,dx = n \int_0^{1/2} \frac{t}{(1-t)^2}\,dt = \text{const.}\, n$. Másrészt $\operatorname{Var} S_2^{(n)}$

nagyságrendje n^2. Valóban, tekintve a $\operatorname{Var} S_2^{(n)}$-et definiáló összegnek a $k=n$ paraméterhez tartozó tagját kapjuk, hogy $\operatorname{Var} S_2^{(n)} \ge n(n-1)$. Továbbá $\operatorname{Var} S_2^{(n)} = n\sum\limits_{k=1}^{n} \frac{k-1}{(n-k+1)^2} \le n\sum\limits_{k=1}^{n} \frac{n}{(n-k+1)^2} = n^2\sum\limits_{k=1}^{n} \frac{1}{k^2} \le \text{const.}\, n^2$.

7.) Hogyan tudunk a centrális határeloszlástétel segítségével egy olyan $[A_n, B_n]$ intervallumot definiálni, amelyre nagy n számra $P(S_1^{(n)} \in [A_n, B_n]) \sim 0.9$ a 6. feladatban definiált $S_1^{(n)}$ valószínűségi változóval? Mutassuk meg, hogy az $S_2^{(n)}$ valószínűségi változókra nem alkalmazható a centrális határeloszlástétel.

Megoldás: Láttuk a 6. feladat megoldásában, és az azt követő megjegyzésben, hogy $S_1^{(n)}$ felírható $\frac{n}{2}$ független valószínűségi változó összegeként, és a várható értéke $C_n = n\sum\limits_{k=n/2+1}^{n} \frac{1}{k}$, szórásnégyzete pedig

$$\begin{aligned}\operatorname{Var} S_1^{(n)} &= n\sum_{k=1}^{n/2} \frac{k-1}{(n-k+1)^2} = n\sum_{k=1}^{n/2} \frac{\frac{k-1}{n}}{(1-\frac{k+1}{n})^2} n \sim n\int_0^{1/2} \frac{t}{(1-t)^2}\,dt \\ &= n\left(\left[\frac{1}{1-t}\right]_0^{1/2} + [\log(1-t)]_0^{1/2}\right) = (1-\log 2)n.\end{aligned}$$

Ezért a centrális határeloszlástétel alapján

$$P\left(-t \le \frac{S_1^{(n)} - C_n}{\sqrt{n(1-\log 2)}} \le t\right) \sim 2\Phi(t) - 1.$$

Válasszuk t-t, mint azt a számot, amelyre $\Phi(t) = 0.95$. A normális eloszlás táblázata alapján $t \sim 1.645$. Legyen $A_n = C_n - 1.645\sqrt{(1-\log 2)n}$, $B_n = C_n + 1.645\sqrt{(1-\log 2)n}$, ahol $C_n = n\sum\limits_{k=n/2+1}^{n} \frac{1}{k}$.

Meg kell még indokolni, hogy a centrális tétel alkalmazható ebben az esetben. A szériasorozatokról szóló centrális határeloszlástételre van szükségünk a $\xi_k^{(n)}$, $n = 2, 4, 6, \dots$, $1 \le k \le n$, szériasorozatra. Be lehet látni az előadáson független valószínűségi változók összegére tárgyalt érveléssel, hogy a centrális határeloszlástétel teljesül, ha az összeadandók valamely 2-nél nagyobb (például negyedik) momentumai nem túl nagyok. A 42. feladat számolása alapján könnyen ellenőrizhető, hogy $E\xi_k^4 \le C_1$, ha $1 \le k \le \frac{n}{2}$ egy C_1 konstanssal, amely nem függ n-től. Másrészt $\operatorname{Var} \xi_k \ge C_2$ egy n-től független C_2 konstanssal, ha $\frac{n}{4} \le k \le \frac{n}{2}$. Ezek a relációk elegendőek a centrális határeloszástétel teljesüléséhez.

A 6. feladat utáni megjegyzésben láttuk, hogy $\operatorname{Var} \xi_n \ge \text{const.}\, n^2$, és $\operatorname{Var} S_2^{(n)} \le \text{const.}\, n^2$. Ez azt jelenti, hogy ebben az esetben a centrális határeloszlástételhez szükséges egyenletes kicsiség feltétele nem teljesül.

8.) Egy nagyvárosban népszavazást tartanak egy kérdésről. A város egy folyó két oldalán fekszik, és a folyó két oldalán lakóknál más mind a kérdés támogatottsága, mind a szavazási hajlandóság. A folyó baloldalán 85000 szavazópolgár lakik, az ottlakók $\frac{3}{4}$ valószínűséggel támogatják a javaslatot, és $\frac{4}{5}$ valószínűséggel mennek el szavazni. A folyó jobbpartján 50400 szavazópolgár lakik, az ottlakók $\frac{1}{2}$ valószínűséggel támgatják a javaslatot, és $\frac{2}{3}$ valószínűséggel mennek el szavazni. Az egyes lakosok véleménye és szavazási hajlandósága független egymástól. Mi annak a (közelítő) valószínűsége, hogy a leadott igen szavazatok száma nagyobb, mint a leadott nem szavazatok kétszerese plusz 1080?

Megoldás. Vezessük be a következő ξ_j, $1 \le j \le 85000$ és η_j, $1 \le j \le 50400$, valószínűségi változokat. $\xi_j = 1$, ha a folyó j-ik baloldali partján lakó szavazópolgár igennel szavaz, $\xi_j = -2$, ha nemmel szavaz, és $\xi_j = 0$, ha nem megy el szavazni, $1 \le j \le 85000$. Hasonlóan, $\eta_j = 1$, ha a folyó j-ik jobboldali partján lakó szavazópolgár igennel szavaz, $\eta_j = -2$, ha nemmel szavaz, és $\eta_j = 0$, ha nem megy el szavazni, $1 \le j \le 50400$. Legyen $S = \sum\limits_{j=1}^{85000} \xi_j + \sum\limits_{j=1}^{50400} \eta_j$. Vegyük észre, hogy minket a $P(S > 1080)$ valószínűség nagysága érdekel. (A ξ_j és η_j valószínűségi változókat hasonlóan definiáltuk. Azért tettünk közöttük különbséget, mert más az eloszlásuk.) Erre a valószínűségre a független, de nem feltétlenül egyforma eloszlású valószínűségi változók összegére vonatkozó centrális határeloszlástétel segítségével tudunk jó közelítést adni. Ennek érdekében számoljuk ki az S összeg várható értékét és szórásnégyzetét.

Mivel $P(\xi_j = 1) = \frac{3}{5}$, $P(\xi_j = -1) = \frac{1}{5}$, $P(\xi_j = 0) = \frac{1}{5}$, ezért $E\xi_j = \frac{1}{5}$, $E\xi_j^2 = \frac{7}{5}$, $\operatorname{Var}\xi_j = \frac{34}{25}$. Hasonlóan, $P(\eta_j = 1) = \frac{1}{3}$, $P(\eta_j = -1) = \frac{1}{3}$, $P(\eta_j = 0) = \frac{1}{3}$, ezért $E\eta_j = -\frac{1}{3}$, $E\eta_j^2 = \frac{5}{3}$, $\operatorname{Var}\xi_j = \frac{14}{9}$. Ezért $ES = 85000 \times \frac{1}{5} - 50400 \times \frac{1}{3} = 200$, $\operatorname{Var} S = 85000 \times \frac{34}{25} + 50400 \times \frac{14}{9} = (20)^2(17^2 + 14^2) \sim 440^2$.

Innen $P(S > 1080) = P\left(\frac{S-ES}{\operatorname{Var} S} > \frac{1080-ES}{\operatorname{Var} S}\right) = P\left(\frac{S-ES}{\operatorname{Var} S} > \frac{880}{440}\right) \sim 1 - \Phi(2) \sim 0.0228$.

További feladatok.

9.) Legyenek ξ és η független, standard normális eloszlású valószínűségi változók. Lássuk be, hogy $\xi^2 + \eta^2$ exponenciális eloszlású valószínűségi változó $\lambda = \frac{1}{2}$ paraméterrel.

Megoldás: $P(\xi^2 < x) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1$, ha $x \ge 0$. Írjuk fel ξ^2 sűrűségfüggvényét és konvolució segítségével a kívánt sűrűségfüggvényt. A ξ^2 valószínűségi változó sűrűségfüggvénye $g(x) = \frac{\varphi(\sqrt{x})}{\sqrt{x}} = \frac{1}{\sqrt{2\pi x}} e^{-x/2}$, ha $x \ge 0$, és $g(x) = 0$, ha $x < 0$, és $\xi^2 + \eta^2$ sűrűségfüggvénye

$$
\begin{aligned}
f(x) = g * g(x) &= \int_0^x \frac{1}{2\pi\sqrt{u(x-u)}} e^{-u/2} e^{-(x-u)/2}\, du \\
&= e^{-x/2} \int_0^1 \frac{1}{2\pi\sqrt{v(1-v)}}\, dv = \frac{1}{2} e^{-x/2}, \quad \text{ha } x \ge 0,
\end{aligned}
$$

és $f(x) = 0$, ha $x \le 0$.

Megjegyzés: Az x paramétertől nem függő $\int_0^1 \frac{1}{\sqrt{v(1-v)}}\,dv$ integrál értékét meghatározza az a tény, hogy a végeredményként kapott függvény sűrűségfüggvény, ezért integrálja a számegyenesen eggyel egyenlő. De ki is tudjuk számolni ezt az integrált. Vagyük észre, hogy

$$\frac{1}{\sqrt{v(1-v)}} = \frac{1}{\sqrt{\frac{1}{4}-(v-\frac{1}{2})^2}} = \frac{2}{\sqrt{1-(2v-1)^2}},$$

ezért $u = 2v-1$ helyettesítéssel

$$\begin{aligned}\int_0^x \frac{1}{2\pi\sqrt{v(1-v)}}\,dv &= \int_{-1}^{2x-1} \frac{1}{2\pi\sqrt{1-u^2}}\,du \\ &= \frac{1}{2\pi}\left[\arcsin x\right]_{-1}^{2x-1} = \frac{\arcsin(2x-1)+\frac{\pi}{2}}{2\pi}.\end{aligned}$$

Innen következik, hogy a tekintett integrál értéke $x=1$ esetén $\frac{1}{2}$, mivel $\arcsin 1 = \frac{\pi}{2}$.

10.) Legyen η_1 és η_2 két független normális eloszlású valószínűségi változó m_1 illetve m_2 várható értékkel, σ_1^2 és σ_2^2 szórásnégyzettel. Lássuk be, hogy az $\eta_1+\eta_2$ összeg m_1+m_2 várható értékű és $\sigma_1^2+\sigma_2^2$ szórásnégyzetű normális eloszlású valószínűségi változó.

A feladat megoldása előtt érdemes megjegyezni, hogy

$$\int_{-\infty}^{\infty} e^{-(x-A)^2/B}\,dx = \sqrt{B\pi}.$$

Ezt láthatjuk például az $y = \sqrt{2}\frac{x-A}{\sqrt{B}}$ helyettesítéssel, és abból a tényből, hogy a $\varphi(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}$ függvény sűrűségfüggvény. Ez az észrevétel azért hasznos, mert ez lehetővé teszi, hogy amennyiben olyan integrált kell kiszámolni, amelyben az integrandus exponensében egy kvadratikus alak szerepel, akkor az inegrandusban szereplő kifejezést teljes négyzetté alakítva ki tudjuk számolni az integrált. Ez a gondolata a jelen feladat megoldásának is.

Megoldás: Az $\eta_1+\eta_2$ valószínűségi változó sűrűségfüggvénye

$$\begin{aligned}f(x) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-(u-m_1)^2/2\sigma_1^2} e^{-(x-u-m_2)^2/2\sigma_2^2}\,du \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-u^2\left(\frac{1}{2\sigma_1^2}+\frac{1}{2\sigma_2^2}\right) + u\left(\frac{m_1}{\sigma_1^2}+\frac{x-m_2}{\sigma_2^2}\right)\right. \\ &\qquad\qquad \left. -\frac{m_1^2}{2\sigma_1^2} - \frac{(x-m_2)^2}{2\sigma_2^2}\right\}du \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(u - \frac{m_1\sigma_2^2+(x-m_2)\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2\right.\end{aligned}$$

$$+\frac{(m_1\sigma_2^2+(x-m_2)\sigma_1^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)}-\frac{m_1^2}{2\sigma_1^2}-\frac{(x-m_2)^2}{2\sigma_2^2}\Bigg\}\,du$$

$$=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}u^2\right\}du$$

$$\exp\left\{\frac{(m_1\sigma_2^2+(x-m_2)\sigma_1^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)}-\frac{m_1^2}{2\sigma_1^2}-\frac{(x-m_2)^2}{2\sigma_2^2}\right\}$$

$$=\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left\{\frac{(m_1\sigma_2^2+(x-m_2)\sigma_1^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)}-\frac{m_1^2}{2\sigma_1^2}-\frac{(x-m_2)^2}{2\sigma_2^2}\right\}$$

$$=\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left\{-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right\}.$$

Megjegyzés. A feladatot lehet egyszerűsíteni. Elég például arra az esetre koncentrálni a figyelmünket, amikor $m_1=0$ és $m_2=0$. Valóban, vezessük be az $\bar\eta_1=\eta_1-m_1$ és $\bar\eta_2=\eta_2-m_2$ valószínűségi változókat. Ekkor $\bar\eta_1$ és $\bar\eta_2$ független normális eloszlású valószínűségi változók 0 várható értékkel, σ_1^2 és σ_2^2 szórásnégyzettel. Másrészt $\eta_1+\eta_2=\bar\eta_1+\bar\eta_2+m_1+m_2$, és ha $\bar\eta_1+\bar\eta_2$ sűrűségfüggvénye $h(x)$, akkor $\eta_1+\eta_2$ sűrűségfüggvénye $h(x-m_1-m_2)$. Miért? Ez azt jelenti, hogy elég $\bar\eta_1+\bar\eta_2$ sűrűségfüggvényét kiszámolni.

11.) Legyen ξ standard normális eloszlású valószínűségi változó, azaz legyen ξ sűrűségfüggvénye $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, és legyen t valós szám. Számoljuk ki az $e^{t\xi}$ valószínűségi változó $Ee^{t\xi}$ várható értékét.

Megoldás:

$$Ee^{t\xi}=\int_{-\infty}^{\infty}e^{tu}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{tu-u^2/2-t^2/2+t^2/2}\,du$$
$$=e^{t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(t-u)^2/2}\,du=e^{t^2/2}.$$

12.) Legyen ξ normális eloszlású valószínűségi változó 2 várható értékkel és 3 szórásnégyzettel. Számoljuk ki minden valós t számra az $Ee^{t\xi}$ várható értéket.

Megoldás. A feladatot megoldhatjuk az előző feladat megoldásához hasonlóan. A ξ valószínűségi változó sűrűségfüggvényét is fel tudjuk írni, és ezután az előző feladathoz hasonló integrál segítségével fel tudjuk írni a keresett várható értéket, és azt ehhez a feladathoz hasonlóan ki tudjuk számolni. De egyszerűbben is célhoz érhetünk. A feladatot közvetlenül visszavezethetjük erre a feladatra. Felírhatjuk ugyanis a ξ valószínűségi változót $\xi=\sqrt{3}\eta+2$ alakban, ahol η standard normális eloszlású valószínűségi változó. Innen következik, hogy $Ee^{t\xi}=Ee^{t(\sqrt{3}\eta+2)}=e^{2t}Ee^{t\sqrt{3}\eta}=e^{2t}Ee^{(\sqrt{3}t)\eta}=e^{2t+3t^2/2}$ az előző feladat eredménye alapján.

13.) Legyen ξ standard normális eloszlású valószínűségi változó, azaz legyen a sűrűségfüggvénye $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $-\infty<x<\infty$. Számoljuk ki a ξ valószínűségi változó $E\xi^k$ k-ik momentumát minden $k=0,1,2,\dots$, nem negatív egész számra.

Megoldás: $E\xi^k = \int_{-\infty}^{\infty} x^k \varphi(x)\,dx$ minden $k = 0, 1, 2, \ldots$ számra. Mivel páratlan $\bar{k} = 2k+1$ számokra a fenti integrál $x^{2k+1}\varphi(x)$ magfüggvénye páratlan, innen adódik, hogy $E\xi^{2k+1} = 0$. Páros indexekre a következő számolást végezhetjük el parciális integrálás segítségével.

$$
\begin{aligned}
E\xi^{2k} &= \int_{-\infty}^{\infty} x^{2k}\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2k-1} x e^{-x^2/2}\,dx \\
&= \frac{-1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2k-1}\frac{d}{dx}\left(e^{-x^2/2}\right) dx \\
&= \frac{-1}{\sqrt{2\pi}}\left[x^{2k-1}e^{-x^2/2}\right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(2k-1)x^{2k-2}e^{-x^2/2}\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(2k-1)x^{2k-2}e^{-x^2/2} = (2k-1)E\xi^{2k-2}\,dx.
\end{aligned}
$$

Mivel $E\xi^0 = 1$ a fenti azonosságból kapjuk, hogy $E\xi^2 = 1$, $E\xi^4 = 3E\xi^2 = 3$, $E\xi^6 = 5E\xi^4 = 5\cdot 3$, és teljes indukcióval $E\xi^{2k} = (2k-1)\cdot(2k-3)\cdot(2k-5)\cdots 3\cdot 1$.

14.) Legyen ξ és η két független standard normális eloszlású valószínűségi változó. Számítsuk ki az $\frac{\eta}{\xi}$ hányados eloszlás és sűrűségfüggvényét.

A megoldás kidolgozása előtt tegyünk először egy általános megjegyzést. Ha adva van két valószínűségi változó ξ és η, amelyek (együttes) sűrűségfüggvénye egy ismert $f(u,v)$ sűrűségfüggvény, akkor az $\frac{\eta}{\xi}$ hányados eloszlásfüggvényét a következő módon számolhatjuk ki: Vezessük be a $g(u,v) = g_x(u,v)$ függvényt, amely a síkon az $\left\{(u,v)\colon\ \frac{v}{u} < x\right\}$ halmaz indikátorfüggvénye, azaz $g(u,v) = 1$, ha $\frac{v}{u} < x$, és $g(u,v) = 0$, ha $\frac{v}{u} \ge x$. Ekkor $P\left(\frac{\eta}{\xi} < x\right) = Eg(\xi,\eta) = \int\int g(u,v)f(u,v)\,du\,dv = \int\int_{\{(u,v)\colon\ \frac{v}{u}<x\}} f(u,v)\,du\,dv$. Látni fogjuk, hogy az ebben a feladatban vizsgálandó integrál viszonylag egyszerű, könnyebben kezelhető.

Megoldás. A feladatban vizsgálandó hányados eloszlásfüggvénye

$$
\begin{aligned}
F(x) = P\left(\frac{\eta}{\xi} < x\right) &= \iint_{\{(u,v)\colon\ v<xu\}} \frac{1}{2\pi}e^{-(u^2+v^2)/2}\,du\,dv \\
&= 2\frac{1}{2\pi}\int_{-\frac{\pi}{2}<\varphi<\arctan x}\int_0^{\infty} re^{-r^2/2}\,dr\,d\varphi = \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\arctan x} d\varphi = \frac{1}{2} + \frac{1}{\pi}\arctan x.
\end{aligned}
$$

Ebben a számolásban a felírt integrált átírtuk $u = r\cos\varphi$, $v = r\sin\varphi$ transzformációval polárkoordinátarendszerben. E számolás során az integrandusban megjelenik az r Jacobian mint szorzó faktor. Ezután azt vegyük észre, hogy a belső r változó szerinti integrál $\int_0^{\infty} re^{-r^2/2}\,dr = \left[-e^{-r^2/2}\right]_0^{\infty} = 1$.

Kiszámoltuk a keresett valószínűségi változó eloszlásfüggvényét. E valószínűségi változó sűrűségfüggvénye az eloszlásfüggvény deriváltja, azaz az $f(x) = \frac{dF(x)}{dx} = \frac{1}{\pi(1+x^2)}$ függvény.

Második megoldás. A (ξ,η) vektor sűrűségfüggvénye $\frac{1}{2\pi}e^{-(x^2+y^2)/2}$, ami forgatásinvariáns függvény. Innen következik, hogy annak valószínűsége, hogy a (ξ,η) vektor egy origóból kiinduló α szögű szögtartományba esik, $\frac{\alpha}{2\pi}$. Ezért

$$\begin{aligned}P\left(\frac{\eta}{\xi}<x\right)&=P\left((\xi,\eta)\in\left[-\frac{\pi}{2},\arctan x\right)\cup\left[\frac{\pi}{2},\arctan x+\pi\right)\text{ szögtartományban}\right)\\&=\frac{1}{2\pi}2\left(\arctan x+\frac{\pi}{2}\right)=\frac{1}{2}+\frac{1}{\pi}\arctan x.\end{aligned}$$

15.) Legyen (ξ,η) két-dimenziós valószínűségi változó $f(x,y)$ sűrűségfüggvénnyel. Lássuk be, hogy a $\frac{\eta}{\xi}$ valószínűségi változónak is létezik sűrűségfüggvénye, és az a $g(t)=\int_{-\infty}^{\infty}f(x,tx)|x|\,dx$ függvény. Adjunk ennek az eredménynek a segítségével új megoldást az előző feladatra.

Megoldás: Jelölje $G(t)$ a $\frac{\eta}{\xi}$ tört $G(t)=P\left(\frac{\eta}{\xi}<t\right)$ eloszlásfüggvényét. Ekkor

$$G(t)=\int_{(x,y)\colon\ \frac{y}{x}<t}f(x,y)\,dx\,dy.$$

Számítsuk ki ezt az integrált az $(\bar x,z)=\left(x,\frac{y}{x}\right)$ helyettesítéssel. Ekkor a $G(t)$ függvényt kifejező integrálban az új integrálási tartomány az $\{(\bar x,z)\colon -\infty<\bar x<\infty,\ -\infty<z<t\}$, $f(x,y)=f(\bar x,z\bar x)$, és az integráltranszformáció kiszámításához meg kell határoznunk a leképezés Jacobi transzformációját. Ez az $\bar x=h_1(x,y)=x$, $z=h_2(x,y)=\frac{y}{x}$ jelöléssel

$$J(\bar x,z)=\left|\begin{pmatrix}\frac{\partial\bar x}{\partial x}, & \frac{\partial\bar x)}{\partial y}\\ \frac{\partial z}{\partial x}, & \frac{z}{\partial y},\end{pmatrix}\right|=\left|\begin{pmatrix}\frac{\partial h_1(x,y)}{\partial x}, & \frac{\partial h_1(x,y)}{\partial y},\\ \frac{\partial h_2(x,y)}{\partial x}, & \frac{\partial h_2(x,y)}{\partial y},\end{pmatrix}\right|=\left|\begin{pmatrix}1, & 0\\ -\frac{y}{x^2}, & \frac{1}{x}\end{pmatrix}\right|=\left|\frac{1}{x}\right|,$$

és informálisan $d\bar x\,dz=J(x,z)\,dx\,dy$, ahonnan $dx\,dy=\frac{1}{J(x,z)}\,d\bar x\,dz=|\bar x|\,d\bar x\,dz$, ahonnan

$$G(t)=\int_{(x,y)\colon\ \frac{y}{x}<t}f(x,y)\,dx\,dy=\int\int_{(x,z):z<t}f(x,zx)|x|\,dx\,dz=\int_{-\infty}^{t}K(z)\,dz,$$

ahol

$$K(z)=\int_{-\infty}^{\infty}f(x,xz)|x|\,dx.$$

Innen látható, hogy a keresett sűrűségfüggvény $g(t)=K(t)$, amint állítottuk.

Ha ξ és η két független standard normális eloszlású valószínűségi változó, akkor (ξ,η) sűrűségfüggvénye $f(x,y)=\frac{1}{2\pi}e^{-(x^2+y^2)/2}$, ezért $\frac{\eta}{\xi}$ sűrűségfüggvénye

$$\begin{aligned}g(t)&=\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{-(x^2+t^2x^2)/2}|x|\,dx\\&=\int_{-\infty}^{\infty}\frac{1}{2\pi(t^2+1)}e^{-(\sqrt{(t^2+1)}x)^2/2}\left|\left(\sqrt{(t^2+1)}x\right)\right|\,d\left(\sqrt{(t^2+1)}x\right),\end{aligned}$$

ahonnan

$$\begin{aligned} g(t) &= \frac{1}{2\pi(t^2+1)} \int_{-\infty}^{\infty} e^{-x^2/2}|x|\,dx = \frac{1}{\pi(t^2+1)} \int_0^{\infty} e^{-x^2/2} x\,dx \\ &= \frac{1}{\pi(t^2+1)} \left[-e^{-x^2/2} \right]_0^{\infty} = \frac{1}{\pi(t^2+1)}. \end{aligned}$$

16.) Legyen adva k darab urna, és ezekbe dobjunk be véletlen ξ számú golyót, ahol ξ Poisson eloszlású valószínűségi változó $\lambda > 0$ paraméterrel. Legyenek az egyes dobások eredményei egymástól és a ξ valószínűségi változótól függetlenek. Tegyük fel továbbá, hogy minden egyes dobásnál a golyó $p_j \geq 0$ valószínűséggel esik a j-ik urnába, $j = 1, \dots, k$, $\sum\limits_{j=1}^{k} p_j = 1$. Jelölje η_j a j-ik urnába eső golyók számát. Ekkor az η_j, $j = 1, \dots, k$, valószínűségi változók függetlenek, és η_j Poisson eloszlású λp_j paraméterrel, $j = 1, \dots, k$.

Megoldás:

$$\begin{aligned} P(\eta_1 = l_1, \dots, \eta_k = l_k) &= P(\xi = l_1 + \cdots + l_k) \frac{(l_1 + \cdots + l_k)!}{l_1! \cdots l_k!} p_1^{l_1} \cdots p_k^{l_k} \\ &= \frac{\lambda^{(l_1 + \cdots + l_k)}}{l_1! \cdots l_k!} p_1^{l_1} \cdots p_k^{l_k} e^{-\lambda} \\ &= \frac{\lambda^{l_1} \cdots \lambda^{l_k}}{l_1! \cdots l_k!} p_1^{l_1} \cdots p_k^{l_k} e^{-\lambda(p_1 + \cdots + p_k)} = \prod_{j=1}^{k} \frac{(\lambda p_j)^{l_j}}{l_j!} e^{-\lambda p_j} \end{aligned}$$

tetszőleges $l_1 \geq 0$, ... , $l_k \geq 0$ egész számokra. Innen adódik az állítás.

Lássuk be az előző feladat segítségével a következő állítást:

17.) Legyen adva egy ξ Poisson eloszlású valószínűségi változó λ paraméterrel. Dobjunk le egymástól és a ξ valószínűségi változótól függetlenül véletlenül egyenletesen ξ darab pontot az egységintervallumra, azaz tegyük fel, hogy minden pont $b - a$ valószínűséggel esik valamely $[a, b] \subset [0, 1]$ intervallumba. Ekkor a $[0, 1]$ intervallum tetszőleges felbontására $[s_0, s_1]$, $[s_1, s_2]$, ... , $[s_{k-1}, s_k]$ diszjunkt intervallumokra, $0 = s_0 < s_1 < \cdots < s_k = 1$ igaz az, hogy az egyes intervallumokba eső pontok száma egymástól független Poisson eloszlású valószínűségi változó $s_j - s_{j-1}$, $1 \leq j \leq k$, paraméterrel.

Megoldás: Tekintsük a következő urnamodellt. Veszünk ξ számú golyót, tehát annyit, ahány ledobott pontot vettünk az előző feladatban. Tegyük az l-ik golyót a j-ik urnába, ha a l-ik pont az $[s_{j-1}, s_j]$ intervallumba esett, $1 \leq j \leq k$. Akkor az előző feladat eredménye alapján az egyes urnákba eső golyók száma egymástól függetetlen Poisson eloszlású valószínűségi változó $s_j - s_{j-1}$, $1 \leq j \leq k$, paraméterrel. Innen következik a feladat állítása.

18.) Legyen ξ és η két valószínűségi változó, amelyek együttes eloszlásának (létező) sűrűségfüggvénye $f(u, v) = \frac{2}{3}u + 2uv^2 + \frac{2}{3}v$ alakú, ha $0 \leq u, v \leq 1$, és $f(u, v) = 0$

egyébként. Lássuk be először, hogy $f(u,v)$ valóban sűrűségfüggvény, majd számoljuk ki a $\xi+\eta$ valószínűségi változó sűrűségfüggvényét.

Megoldás: Annak érdekében, hogy ellenőrizzük, hogy $f(u,v)$ sűrűségfüggvény azt kell megmutatni, hogy $f(u,v)\ge 0$ (majdnem) minden (u,v) számpárra, és

$$\int\int f(u,v)\,du\,dv=1.$$

Az nyilvánvaló, hogy $f(u,v)\ge 0$ minden (u,v) számpárra. Másrészt mivel

$$\int_0^1\int_0^1 uv^2\,du\,dv=\int_0^1 u\,du\int_0^1 v^2\,dv=\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{6},$$

$\int_0^1\int_0^1 u\,du\,dv=\frac{1}{2}$ és $\int_0^1\int_0^1 v\,du\,dv=\frac{1}{2}$, ezért $\int\int f(u,v)\,du\,dv=1$.
A $\xi+\eta$ valószínűségi változó eloszlásfüggvényét az

$$G(x)=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x-u} f(u,v)\,dv\right)du$$

képlet segítségével számíthatuk ki, hasonlóan $1-G(x)=\int_{-\infty}^{\infty}\left(\int_{x-u}^{\infty} f(u,v)\,dv\right)du$. Másrészt, a sűrűségfüggvény $g(x)=\frac{dG(x)}{dx}$ formula segítségével kiszámítható. Ez a mi esetünkben a következőt jelenti, $g(x)=0$, ha $x\le 0$, $g(x)=1$, ha $x\ge 2$, mert $G(x)=0$, ha $x\le 0$, $G(x)=1$, ha $x\ge 2$. Másrészt

$$g(x)=\frac{d}{dx}\left(\int_0^1\int_0^{x-u}\left(\frac{2}{3}u+2uv^2+\frac{2}{3}v\right)du\right)dv,$$

ha $0\le x\le 1$, $g(x)=-\frac{d}{dx}\left(\int_0^1\int_{x-u}^1\left(\frac{2}{3}u+2uv^2+\frac{2}{3}v\right)du\right)dv$, ha $1\le x\le 2$. Azért volt érdemes a $0\le x\le 1$ és $1\le x\le 2$ eseteket szétválasztani, mert a konkrét feladatban a $G(x)$ függvényt tudjuk kényelmesen kiszámítani $0\le x\le 1$ és az $1-G(x)$ függvényt az $1\le x\le 2$ intervallumban. Ez a szétbontás azonban nem kötelező.

A fent vázolt módon ki lehet számolni a sűrűségfüggvényt, de valójában ezt a számolást lehet egyszerűsíteni. Ez hasonló ahhoz, ahogy a konvolució formulát vezetik le független valószínűségi változók összegének a sűrűségfüggvényére. Valóban, a v változó $z=u+v$ helyettesítésével felírhatjuk, hogy

$$\begin{aligned}G(x)&=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x-u} f(u,v)\,dv\right)du=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x} f(u,z-u)\,dz\right)du\\&=\int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f(u,z-u)\,du\right)dz,\end{aligned}$$

ahonnan deriválással

$$g(x) = \frac{dG(x)}{dx} = \int_{-\infty}^{\infty} f(u, x-u)\, du.$$

E képlettel a számolásokat lehet egyszerűsíteni. Azt kapjuk, hogy jelen esetben $g(x) = 0$, ha $x < 0$ vagy $x > 2$, (ebben az esetben a $g(x)$ függvényt kifejező integrálban szereplő integrandus azonosan nulla. Ugyanis $x < 0$ esetében vagy $u < 0$ vagy $x - u < 0$, és $x > 2$ esetben vagy $u > 1$ vagy $x - u > 1$. Ha $0 \le x \le 1$, akkor

$$g(x) = \int_0^{1-x} \left(\frac{2}{3}u + 2u(x-u)^2 + \frac{2}{3}(x-u) \right) du,$$

ha $1 \le x \le 2$, akkor

$$g(x) = \int_{x-1}^{1} \left(\frac{2}{3}u + 2u(x-u)^2 + \frac{2}{3}(x-u) \right) du.$$

18a.) Mutassuk meg, hogy a fenti feladat megoldásában kapott részeredmények tartalmazzák speciálisan azt az eredményt is, hogy két független valószínűségi változó összegének a sűrűségfüggvényét konvolució segítségével lehet kiszámolni.

Megoldás. Láttuk, hogy amennyiben egy (ξ, η) véletlen vektor sűrűségfüggvénye $f(x, y)$, akkor a $\xi + \eta$ valószínűségi változónak is van sűrűségfüggvénye, és az $g(x) = \int_{-\infty}^{\infty} f(u, x-u)\, du$. Legyen ξ és η két független valószínűségi változó $h_1(x)$ és $h_2(x)$ sűrűségfüggvénnnyel. Ekkor a (ξ, η) vektornak is van sűrűségfüggvénye, és az az $f(x, y) = h_1(x)h_2(y)$ függvény. Így a $\xi + \eta$ összegnek is van sűrűségfüggvénye, és az az idézett eredmény szerint $g(x) = \int_{-\infty}^{\infty} f(u, x-u)\, du = \int_{-\infty}^{\infty} h_1(u)h_2(x-u)\, du$.

19.) Legyen a (ξ, η) vektor egyenletes eloszlású a $(0,0)$, $(1,0)$ és $(0,1)$ pontok által meghatározott háromszögön, azaz legyen sűrűségfüggvénye 2 azon a háromszögön, amelynek ezek a pontok a csúcspontjai, és legyen nulla ezen a háromszögön kívül. Számítsuk ki a ξ és η valószínűségi változók kovarianciáját.

Megoldás: $\operatorname{Cov}(\xi, \eta) = E\xi\eta - E\xi E\eta$, és $E\xi\eta = \int xyf(x,y)\, dx\, dy$, $E\xi = \int xf(x,y)\, dx\, dy$, $E\eta = \int yf(x,y)\, dx\, dy$, ahol $f(x, y)$ a (ξ, η) vektor sűrűségfüggvénye. Ezért

$$E\xi\eta = \int_0^1 \left(\int_0^{1-x} 2xy\, dy \right) dx = \int_0^1 2x \left[\frac{y^2}{2} \right]_0^{1-x} dx = \int_0^1 x(1-x)^2\, dx$$
$$= \left[\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4} \right]_0^1 = \frac{1}{12},$$

$$E\xi = \int_0^1 \left(\int_0^{1-x} 2x\, dy \right) dx = \int_0^1 2x(1-x)\, dx = \frac{1}{3},$$

és

$$E\eta = \int_0^1 \left(\int_0^{1-x} 2y\, dy \right) dx = \int_0^1 (1-x)^2\, dx = \frac{1}{3},$$

(Szimmetria meggondolások alapján is belátható, hogy $E\eta = E\xi$.) Innen $\operatorname{Cov}(\xi, \eta) = \frac{1}{12} - \frac{1}{9} = -\frac{1}{36}$.