Az október 1.-i előadáshoz kapcsolódó gyakorlat feladatai

1. Legyen ξ és η két független, a $\left[-\dfrac{1}{2},\dfrac{1}{2}\right]$ intervallumban egyenletes eloszlású valószínűségi változó, azaz legyen ξ és η sűrűségfüggvénye $f(x)=1$, ha $-\dfrac{1}{2}\le x\le\dfrac{1}{2}$, és $f(x)=0$ egyébként. Számoljuk ki $\xi+\eta$ sűrűségfüggvényét.
Megoldás: A $\xi+\eta$ valószínűségi változó sűrűségfüggvénye a $g(x)=\int f(y)f(x-y)\,dy$ függvény, ahol $f(x)$ a $\left[-\dfrac{1}{2},\dfrac{1}{2}\right]$ intervallumban egyenletes eloszlás sűrűségfüggvénye. Ezért $f(y)f(x-y)=1$, ha $-\dfrac{1}{2}\le y\le\dfrac{1}{2}$, és $-\dfrac{1}{2}\le x-y\le\dfrac{1}{2}$, azaz $-\dfrac{1}{2}+x\le y\le \dfrac{1}{2}+x$, és nulla egyébként. Ez azt jelenti, hogy a $\xi+\eta$ összeg $g(x)$ sűrűségfüggvénye az x pontban megegyezik a $\left[-\dfrac{1}{2},\dfrac{1}{2}\right]\cap\left[-\dfrac{1}{2}+x,\le\dfrac{1}{2}+x\right]$ intervallum hosszával. Ha $|x|>1$, akkor a fenti metszet üres, ezért ebben az esetben $g(x)=$. Ha $0\le x\le 1$, akkor ez a metszet a $\left[-\dfrac{1}{2}+x,\dfrac{1}{2}\right]$ intervallum, és ennek hossza $1-x$, azaz ebben az esetben $g(x)=1-x$. Ha $-1\le x\le 0$, akkor ez a metszet a $\left[-\dfrac{1}{2},\dfrac{1}{2}+x\right]$ intervallum melynek hossza $1+x=1-|x|$, azaz $g(x)=1+x=1-|x|$ ebben az esetben. Ez azt jelenti, hogy $g(x)=1-|x|$, ha $|x|\le 1$, és $g(x)=0$, ha $x>1$.

   Megadunk egy másik geometriai érvelésen alapuló megoldást is, amelyik geometriai érvelésen alapul.

   Számítsuk ki először a $\xi+\eta$ valószínűségi változó $G(x)$ eloszlásfüggvényét. Definiáljuk a $K=\left[-\dfrac{1}{2},\dfrac{1}{2}\right]\times\left[-\dfrac{1}{2},\dfrac{1}{2}\right]$ négyzetet, és jelölje λ a Lebesgue mértéket, azaz a területet a síkon. Ekkor a sík tetszőleges $A\subset R^2$ mérhető részhalmazára igaz az, hogy $P((\xi,\eta)\in A)=\lambda(A\cap K)$. Speciálisan, $G(x)=P(\xi+\eta<x)=\lambda(K\cap\{(u,v)\colon u+v<x\})$. Ha $x\le -1$, akkor $G(x)=0$, ha $-1\le x\le 0$, akkor $G(x)$ a $\left(-\frac{1}{2},-\frac{1}{2}\right)$, $\left(-\frac{1}{2},\frac{1}{2}+x\right)$ és $\left(\frac{1}{2}+x,-\frac{1}{2}\right)$ pontok által meghatározott háromszög területe $\frac{1}{2}(1+x)^2$. Hasonlóan, ha $x\ge 1$, akkor $G(x)=1$. Ha $0\le x\le 1$, akkor a $G(x)$ eloszlásfüggvény megegyezik annak a poligonnak területével, melyet úgy kapunk, hogy a K négyzetből kihagyjuk a $\left(\frac{1}{2},\frac{1}{2}\right)$, $\left(\frac{1}{2},-\frac{1}{2}+x\right)$ és $\left(-\frac{1}{2}+x,\frac{1}{2}\right)$ pontok által meghatározott háromszöget. Ezért $G(x)=1-\frac{1}{2}(1-x)^2$ ebben az esetben. A $G(x)$ függvényt deriválva kapjuk, hogy $g(x)=0$, ha $|x|\le 1$, $g(x)=1+x$, ha $-1\le x\le 0$, és $g(x)=1-x$, ha $0\le x\le 1$.

2. Legyen két urna, mind a kettőben 10 piros és 10 fehér golyó. Egymás után kihúzunk nyolc golyót mind a két urnából, az elsőből visszatevés nélkül, a másodikból visszatevéssel. Ha a j-ik húzásnál a két urnából kihúzott golyó egyforma színű, akkor két forintot nyerünk, ha különböző színűek, akkor egy forintot veszítünk, $1\le j\le 8$. Számítsuk ki nyereményünk várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 8$ valószínűségi változókat: $\xi_j = 2$, ha a j-ik húzásnál mind két urnából piros vagy mind a két urnából fehér golyót húzunk, és $\xi_j = -1$, ha az egyik urnából fehér és a másik urnából piros golyót hízunk. Akkor a minket érdeklő mennyiségek a $S = \sum\limits_{j=1}^{8} \xi_j$ valószínűségi változó várható értéke és szórásnégyzete. Ennek kiszámítása érdekében számítsuk ki az $E\xi_j$, $\operatorname{Var}\xi_j$ és $\operatorname{Cov}(\xi_j, \xi_k)$ mennyiségeket. Vegyük észre, hogy $P(\xi_j = 1) = P(\xi_1 = 1) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$, mert kifejezve külön annak a valószínűségét, hogy (fehér, fehér) vagy (piros, piros) húzás történik. Ezért $E\xi_j = 2\frac{1}{2} - \frac{1}{2} = \frac{1}{2}$, $E\xi_j^2 = \frac{1}{2}(4+1) = \frac{5}{2}$, $\operatorname{Var}\xi = \frac{9}{4}$.

Hasonlóan, $P(\xi_j = 2, \xi_k = 2) = P(\xi_1 = 2, \xi_2 = 2) = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} \frac{9}{19} + 2 \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{9}{19} = \frac{1}{4}$, $P(\xi_j = 2, \xi_k = -1) = P(\xi_1 = 2, \xi_2 = -1) = 2 \cdot \frac{1}{4} \cdot \frac{1}{2}\left(\frac{10}{19} + \frac{9}{19}\right) = \frac{1}{4}$, $P(\xi_j = -1, \xi_k = -1) = P(\xi_1 = -1, \xi_2 = -1) = \frac{1}{4}$ (itt felsoroltuk, hogy például a $\xi_1 = 2, \xi_2 = 2$ azt jelenti, hogy ((F,F), (F,P)) ((F,F), (P,F)), ((P,P), (F,P)), ((P,P), (P,F)) hízássorozatok valamelyike következik be. Ezért $E\xi_j\xi_k = \frac{1}{4}(1+4-4) = \frac{1}{4}$, és $\operatorname{Cov}(\xi_j, \xi_k) = E\xi_j\xi_k - E\xi_j E\xi_k = \frac{1}{4} - \frac{1}{4} = 0$. Innen következik, hogy a ξ_j valószínűségi változók korrelálatlanok, és $ES = 8E\xi_1 = 2$, $\operatorname{Var} S = 8\operatorname{Var}\xi_1 = 18$.

3. Mutassuk meg, hogy egy nem-nagatív értékű ξ valószínűségi változó (azaz $P(\xi \ge 0) = 1$) várható értéke akkor és csak akkor véges, ha $\sum\limits_{j=1}^{\infty} jP(j-1 \le \xi \le j) < \infty$, ami ekvivalens azzal, hogy $\sum\limits_{j=1}^{\infty} P(\xi > j) < \infty$.

Megoldás: Vezessük be a következő $\tilde{\xi}$ valószínűségi változót: $\tilde{\xi} = j$, ha $j - 1 < xi \le j$, $j = 1, 2, \dots$. Ekkor $P(0 \le \xi - \tilde{\xi}) \le 1$, ezért a ξ és $\tilde{\xi}$ valószínűségi változók várható értéke egyszerre véges vagy végtelen. Viszont $E\tilde{\xi} = \sum\limits_{j=1}^{\infty} jP(\tilde{\xi} = j) = \sum\limits_{j=1}^{\infty} jP(j-1 < \xi \le j)$, ahonnan következik a feladat első állítása.

Másrészt Felírhatjuk, hogy $\sum\limits_{j=1}^{\infty} P(j-1 < \xi \le j) = \sum\limits_{j=1}^{\infty} j(P(\xi > j-1) - P(\xi > j))$. Rendezzük át a fenti összeget a következő módon. Tetszőleges N számra

$$\sum_{j=1}^{N} jP(j-1 < \xi \le j) = \sum_{j=1}^{N} j(P(\xi > j-1) - P(\xi > j))$$
$$= \sum_{j=1}^{N-1} P(\xi > j)((j+1) - j) - NP(\xi > N). = \sum_{j=1}^{N-1} P(\xi > j) - NP(\xi > N).$$

Megmutatjuk, hogy $N \to \infty$ határátmenettel a fenti relációból következik a feladat második állítása. Valóban, ha $E\xi < \infty$, akkor létezik olyan $K < \infty$ szám, melyre $NP(\xi > N) \le \sum\limits_{j=N+1}^{\infty} jP(j-1 < \xi \le j) \le \sum\limits_{j=1}^{\infty} jP(j-1 < \xi \le j) \le K$, ahol a K konstans független az N számtól. Így ebben az esetben az előző becslés alapján léteznek olyan univerzáslis L és K számok, melyekre

$$L \ge \sum_{j=1}^{N-1} P(\xi > j) - K \quad \text{minden } L = 1, 2, \dots \text{ számra},$$

ahonnan $\sum\limits_{N=1}^{\infty} P(\xi > j) < \infty)$.

Ha $E\xi = \infty$, akkor $\sum\limits_{N=1}^{\infty} P(\xi > j) = \infty)$, mert ekkor $\sum\limits_{j=1}^{N-1} P(\xi > j) \ge \sum\limits_{j=1}^{N} jP(j-1 < \xi \le j)$, és $\lim\limits_{N\to\infty} \sum\limits_{j=1}^{N} jP(j-1 < \xi \le j) = \infty$.

Megjegyzés: Az összegnek a fenti számolásban történt átrendezését Abel féle átrendezésnek nevezik, és ez sokszor hasznos. Az Abel féle átrendezés egyébként az integrálszámításban alkalmazott parciális integrálás diszkrét megfelelője.

4. A Stirling formula felhasználásával (Ez azt mondja ki, hogy $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$) adjunk jó becslést arra, hogy amennyiben egymástól függetlenül n kisérletet végzünk, melyek mindegyike p valószínűséggel sikeres, $1-p$ valószínűséggel sikertelen, akkor pontosan k sikeres kisérlet következik be. Ha $k \sim np$ akkor ennek valószínűsége jól közelíthető egy alkalmas normális sűrűségfüggvénnyel.

Megoldás: Jelölje S_n a sikeres kisérletek számát. Ekkor $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Tekintsük ezt a valószínűséget abban az esetben, ha $\varepsilon n \le k \le (1-\varepsilon)n$. Ekkor

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \sim \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \sqrt{2\pi(n-k)}\left(\frac{n-k}{e}\right)^{n-k}}$$
$$= \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k},$$

$$P(S_n = k) \sim \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k}.$$

Ha $k = k(n) = np + x(n)$, és $\lim\limits_{n\to\infty} \frac{x(n)}{n} = 0$, azaz a sikeres kisérletek k száma közel van annak np várható értékéhez, akkor

$$\log\left(\frac{np}{k}\right)^k\left(\frac{n(1-p)}{n-k}\right)^{n-k}$$
$$=\left(k\log\left(1+\frac{np-k}{k}\right)+(n-k)\log\left(1+\frac{n(1-p)+(n-k)}{(n-k)}\right)\right)$$

Alkalmazzuk a fenti kifejezés mindkét tagjára a $\log(1+u)\sim u-\dfrac{u^2}{2}$ Taylor sorfejtés közelítést kis u számokra. A lineáris tagok kiesnek, és azt kapjuk, hogy

$$n\left(\frac{k}{n}\log\left(1+\frac{np-k}{k}\right)-\frac{n-k}{n}\log\left(1+\frac{n(1-p)+(n-k)}{(n-k)}\right)\right)$$
$$\sim-\frac{1}{2}\left(\frac{(np-k)^2}{k}+\frac{n(1-p)-(n-k))^2}{n-k}\right)=-\frac{n(k-np)^2}{2k(n-k)} \tag{$*$}$$
$$\sim-\frac{(k-np)^2}{2np(1-p)}.$$

Az utolsó közelítésben azt használtuk ki, hogy $k(n-k)\sim n^2p(1-p)$ a vizsgált esetben. Az előző számolásokban a $P(S_n=k)$ valószínűséget megadó kifejezés fő részének a logaritmusát számoltuk ki, ha $k\sim np$. A másik tényezőre

$$\sqrt{\frac{n}{2\pi k(n-k)}}\sim\frac{1}{\sqrt{2\pi np(1-p)}}$$

E becslések azt adják, hogy

$$P(S_n=k)\sim\frac{1}{\sqrt{2\pi np(1-p)}}\exp\left\{-\frac{(k-np)^2}{2np(1-p)}\right\},\quad \text{ha } \frac{kn-p}{n}\to 0.$$

Jegyezzük meg, hogy a becslő függvény a következő módon is írható:

$$\frac{1}{\sqrt{2\pi np(1-p)}}\exp\left\{-\frac{(k-np)^2}{2np(1-p)}\right\}=\frac{1}{\sqrt{2\pi\operatorname{Var}S_n}}\exp\left\{-\frac{(k-ES_m)^2}{2\operatorname{Var}S_n}\right\}.$$

Valójában a fenti vizsgálat nem teljes értékű, mert nem derült ki belőle, hogy a fenti közelítés milyen pontos. Próbáljuk finomítani, és megérteni, hogy milyen $k(n)=pn+x(n)$ számsorozatra mondhatjuk, hogy

$$\lim_{n\to\infty}\frac{P(S_n=k(n))}{\frac{1}{\sqrt{2\pi\operatorname{Var}S_n}}\exp\left\{-\frac{(k-ES_m)^2}{2\operatorname{Var}S_n}\right\}}=1. \tag{$**$}$$

Ha végiggondoljuk a becsléseket (azt kell meggondolni, hogy egy szorzatban az egyes tényezőket olyan tagokkal helyettesíthetjük, melyben az eredeti kifejezés és a becslésként felhasznált kifejezés hányadosa 1-hez tart $n \to \infty$ esetén, valamint $\lim\limits_{n\to\infty} \dfrac{a(n)}{b(n)} = 1$ akkor és csak akkor, ha $\lim\limits_{n\to\infty} (\log a(n) - \log b(n)) = 0$.) ahhoz, hogy a $(**)$ reláció teljesüljün az kell, hogy a $(*)$ formulában leírt becslésben mikor mondhatjuk, hogy

$$n\left(\frac{k}{n}\log\left(1+\frac{np-k(n)}{k(n)}\right) - \frac{n-k(n)}{n}\log\left(1+\frac{n(1-p)+(n-k(n))}{(n-k(n))}\right)\right) \\ - \frac{(k(n)-np)^2}{2np(1-p)} \to 0, \qquad \text{ha } n \to \infty.$$

Nem nehéz belátni, hogy ez érvényes akkor, ha $\dfrac{k(n)-np}{n^{1/6}} \to 0$ $n \to \infty$ esetén. Ugyanis amikor, a $\log(1+u)$ függvényt az $u - \frac{u^2}{2}$ kifejezéssel közelítettük, azaz a függvény Taylor sorának első két tagjának az összegével, akkor az elkövetet hiba kisebb, mint $\text{const.}\, u^3$ például $|u| \le \frac{1}{2}$ esetén (azaz a hiba nagyságrendje annyi, mint a Taylor sor harmadik tagja. Esetünkben ezt a becslést $u = \dfrac{np-k}{k}$ és $u = \dfrac{n(1-p)+(n-k)}{(n-k)}$ választással alkalmaztuk, és mindkét kifejezés nagyságrendje $\dfrac{x(n)}{n} = \dfrac{k(n)-np}{n}$. Ezért az elkövetett hiba nagyságrendje e kifejezés harmadik hatványa beszorozva n-nel. Az elkövetett hiba akkor elhanyagolható, a $(**)$ formula akkor érvényes, ha a $\dfrac{k(n)-np}{n^{1/6}} \to 0$ $n \to \infty$ reláció teljesül. Ezt a becslést követve, és a Taylor sorfejtést a harmadik tagig elvégezve kapnánk, hogy $\dfrac{k(n)-np}{n^{1/6}} \sim \text{const.}$ esetén egy harmadfokú tagot is figyelembe kell venni, ha a $P(S_n = k(n))$ mennyiséget a $(**)$ formulához hasonló pontossággal akarjuk becsülni.