A Valószínűségszámítás I. előadássorozat ötödik témája.

Szórásnégyzet, kovariancia, generátorfüggvények.

Először a szórásnégyzet fogalmát vezetem be, és annak legfontosabb tulajdonságait tárgyalom.

Az (egyelőre csak diszkrét eloszlású valószínűségi változók esetében definiált) várható érték azt méri, hogy valószínűségi változók körülbelül mekkora értéket vesznek fel. Mérni akarjuk azt is, hogy mekkora a valószínűségi változó ingadozása a várható értéke körül. Ennek érdekében vezették be a szórásnégyzet (angolul variance) fogalmát. Ismertetem ennek definicióját.

A szórásnégyzet definiciója: *Legyen ξ olyan valószínűségi változó, amelyre $E\xi^2 < \infty$. Ekkor a ξ valószínűségi változó szórásnégyzetét a*

$$\operatorname{Var}\xi = E\left(\xi - E\xi\right)^2$$

képlet definiálja. Ha $E\xi^2 = \infty$ akkor a $\operatorname{Var}\xi$ szórásnégyzetet nem definiáljuk vagy azt mondjuk, hogy $\operatorname{Var}\xi = \infty$.

Lemma a szórásnégyzet kiszámolásáról.

$$\operatorname{Var}\xi = E\xi^2 - (E\xi)^2.$$

Továbbá, $(E\xi)^2 \le E\xi^2$.

Bizonyítás:

$$\begin{aligned} 0 \le \operatorname{Var}\xi = E\left(\xi - E\xi\right)^2 &= E\left(\xi^2 - 2\xi E\xi + (E\xi)^2\right) \\ &= E\xi^2 - 2E\xi E\xi + (E\xi)^2 = E\xi^2 - (E\xi)^2 \end{aligned}$$

a várható érték tulajdonságai miatt. Összehasonlítva a bal és jobboldalt kapjuk a megfogalmazott egyenlőtlenséget.

1. megjegyzés. Ha precízek vagyunk, akkor felmerülhet a kérdés, hogy teljesen kifogástalan-e a fenti bizonyítás. Az okozhat problémát, hogy egy valószínűségi változó várható értéket csak akkor definiáltuk ha a definiáló összeg abszolut konvergens. Ezt viszont csak az $E\xi^2$ kifejezést definiáló összegről tettük fel. Ugyanakkor az $E\xi$ várható értékkel is szabadon számoltunk. Nem okoz-e ez gondot?

Az, hogy a fenti számolás jogos, például a következő módon látható. Vegyen fel a ξ valószínűségi változó $x_1, x_2, \dots$ értékeket, és definiáljuk a ξ valószínűségi változó ξ_n közelítéseit, $n = 1, 2, \dots,$ a következő módon: $\xi_n(\omega) = \xi(\omega)$, ha $\xi(\omega) = x_k$, $1 \le k \le n$, $\xi_n(\omega) = 0$, ha $\xi(\omega) = x_k$, és $k > n$. Ekkor a $\xi_n(\omega)$ illetve ennek abszolut értékével a $|\xi_n(\omega)|$ valószínűségi változókkal szabadon végrehajthatjuk a fenti számolásokat. Innen következik, hogy $(E|\xi_n(\omega)|)^2 \le E\xi_n^2(\omega) \le E\xi^2(\omega)$. Innen viszont $n \to \infty$ határátmenettel kapjuk, hogy $(E|\xi|)^2 = \lim_{n\to\infty} (E|\xi_n(\omega)|)^2 \le E\xi^2(\omega)$. Ez azt jelenti, hogy az $E\xi$ várható értéket meghatározó összeg is abszolut konvergens, és a fenti számolásokat elvégezhetjük.

2. megjegyzés. A szórásnégyzet egy viszonylag egyszerű mérőszáma a valószínűségi változó ingadozásának annak várható értékétől. Az, hogy valóban ez az ingadozás legjobb mérőszáma később ismertetendő eredményekből következik.

Feladat:

1.) Minden m valós számra

$$E(\xi - m)^2 = E(\xi - E\xi)^2 + (E\xi - m)^2.$$

Következésképpen

$$\operatorname{Var}\xi = \inf_{-\infty<m<\infty} E\left[(\xi - m)^2\right].$$

Megoldás:

$$\begin{aligned} E(\xi - m)^2 &= E\left[(\xi - E\xi) + (E\xi - m)\right]^2 \\ &= E(\xi - E\xi)^2 - 2(E\xi - m)E(\xi - E\xi) + (E(\xi - m))^2 \\ &= E(\xi - E\xi)^2 + (E\xi - m)^2, \end{aligned}$$

mert $2E(\xi - m)E(\xi - E\xi) = 2E(\xi - m)(E\xi - E\xi) = 0$. Ebből az azonosságból adódik, hogy $E(\xi - m)^2 \ge E(\xi - E\xi)^2$ minden m valós számra. Viszont $m = E\xi$ esetén az egyenlőtlenség két oldala megegyezik.

Lemma a szórásnégyzet tulajdonságairól. Minden a és b valós számra

$$\operatorname{Var}(a\xi + b) = a^2 \operatorname{Var}\xi.$$

Bizonyítás:

$$\begin{aligned} \operatorname{Var}(a\xi + b) &= E\left[(a\xi + b) - E(a\xi + b)\right]^2 = E\left[(a\xi + b) - aE\xi - b\right]^2 \\ &= E\left[a^2(\xi - E\xi)^2\right] = a^2 E(\xi - E\xi)^2 = a^2 \operatorname{Var}\xi. \end{aligned}$$

A következő eredmény rendkívül hasznos összefüggést ad, ha független valószínűségi változók összegének szórásnégyzetét akarjuk kiszámítani. Hangsúlyozom, hogy ebben az eredményben fontos a tekintett valószínűségi változók függetlensége. Később megfogalmazom ennek az eredménynek egy olyan általánosítását, amely akkor is érvényes, ha az összeadandók nem feltétlenül függetlenek. Olyan feladatokat is fogunk vizsgálni, amelyek megoldásában ezt az általánosabb eredményt érdemes alkalmazni.

Tétel független valószínűségi változók összegeinek szórásnégyzetéről. *Ha $\xi_1, \xi_2, \dots, \xi_n$ független valószínűségi változók, akkor*

$$\mathrm{Var}\,(\xi_1 + \xi_2 + \cdots + \xi_n) = \mathrm{Var}\,\xi_1 + \mathrm{Var}\,\xi_2 + \cdots + \mathrm{Var}\,\xi_n.$$

Következmény. Ha $\xi_1, \xi_2, \dots, \xi_n$ *független valószínűségi változók,* $c_1, c_2, \dots, c_n$ *valós számok, akkor*

$$\mathrm{Var}\,(c_1\xi_1 + c_2\xi_2 + \cdots + c_n\xi_n) = c_1^2 \mathrm{Var}\,\xi_1 + c_2^2 \mathrm{Var}\,\xi_2 + \cdots + c_n^2 \mathrm{Var}\,\xi_n.$$

A Tétel bizonyítása. Vezessük be a $\bar{\xi}_j = \xi_j - E\xi_j$ valószínűségi változókat. Ekkor

$$\begin{aligned}\mathrm{Var}\,(\xi_1 + \xi_2 + \cdots + \xi_n) &= E(\bar{\xi}_1 + \bar{\xi}_2 + \cdots + \bar{\xi}_n)^2 \\ &= E\left(\bar{\xi}_1^2 + \bar{\xi}_2^2 + \cdots + \bar{\xi}_n^2 + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \bar{\xi}_j\bar{\xi}_k\right) \\ &= E\bar{\xi}_1^2 + E\bar{\xi}_2^2 + \cdots + E\bar{\xi}_n^2 + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} E\bar{\xi}_j\bar{\xi}_k \\ &= E\bar{\xi}_1^2 + E\bar{\xi}_2^2 + \cdots + E\bar{\xi}_n^2 = \mathrm{Var}\,\xi_1 + \mathrm{Var}\,\xi_2 + \cdots + \mathrm{Var}\,\xi_n,\end{aligned}$$

mert $E\bar{\xi}_j\bar{\xi}_k = E\bar{\xi}_j E\bar{\xi}_k = 0$ minden $1 \le j < k \le n$ számra a ξ_j illetve a $\bar{\xi}_j$, $1 \le j \le n$, valószínűségi változók függetlensége miatt, és $E\bar{\xi}_j^2 = \mathrm{Var}\,\xi_j$ minden $1 \le j \le n$ indexre.

Feladat:

2.) Egy szabályos pénzdarabot feldobunk 100-szor egymás után. Számoljuk ki a fejdobások számának a várható értékét és szórásnégyzetét. Számítsuk ki ezt egyrészt direkt módon felírva és kiszámolva a várható értéket és szórásnégyzetet kifejező összegeket, másrészt egyszerűbben az előző tétel segítségével.

Megoldás: A fejdobások száma 0 és 100 között van. Annak valószínűsége, hogy k fejdobás következik be, $0 \le k \le 100$, $p_k = \binom{100}{k}2^{-100}$. Ezért a dobások számának várható értéke $E\xi = \sum\limits_{k=0}^{100} k\binom{100}{k}2^{-100}$, ahol ξ jelöli a fejdobások számát megadó valószínűségi változót. Továbbá $E\xi^2 = \sum\limits_{k=0}^{100} k^2\binom{100}{k}2^{-100}$, a szórásnégyzet pedig $\mathrm{Var}\,\xi = E\xi^2 - (E\xi)^2$. Ezeket az összegeket közvetlenül kiszámíthatjuk. Valóban, $k\binom{100}{k} = 100\binom{99}{k-1}$, $E\xi = \sum\limits_{k=0}^{100} k\binom{100}{k}2^{-100} = 50\sum\limits_{k=0}^{99} k\binom{99}{k}2^{-99} = 50\left(\frac{1}{2} + \frac{1}{2}\right)^{99} = 50$. Továbbá, $k^2\binom{100}{k} = [k(k-1)+k]\binom{100}{k} = 100 \cdot 99 \cdot \binom{98}{k-2} + 100 \cdot \binom{99}{k-1}$, ezért $E\xi^2 = \sum\limits_{k=0}^{100} k^2\binom{100}{k}2^{-100} = 2^{-100} \cdot 100 \cdot 99 \cdot \sum\limits_{k=0}^{98}\binom{98}{k} + 2^{-100} \cdot 100 \cdot \sum\limits_{k=0}^{99}\binom{99}{k} =$

$2^{-100}\left(100 \cdot 99 \cdot (1+1)^{98} + 100 \cdot (1+1)^{99}\right) = 25 \cdot 99 + 50$. Innen $\mathrm{Var}\,\xi = E\xi^2 - (E\xi^2) = 25$.

Valójában a vizsgált várható értéket és szórásnégyzetet egyszerűbben is kiszámíthatjuk. Vezessük be a $\xi_j(\omega)$ valószínűségi változókat, $\xi_j(\omega) = 1$, ha a j-ik dobás eredménye fej $\xi_j(\omega) = 0$, ha a j-ik dobás eredménye írás, $1 \le j \le 100$. Ekkor a fejdobások száma $\xi(\omega) = \sum\limits_{j=1}^{100} \xi_j(\omega)$, $E\xi = \sum\limits_{j=1}^{100} E\xi_j$, $\mathrm{Var}\,\xi = \sum\limits_{j=1}^{100} \mathrm{Var}\,\xi_j$. Mivel $E\xi_j = \frac{1}{2}$, $E\xi_j^2 = \frac{1}{2}$, $\mathrm{Var}\,\xi_j = \frac{1}{4}$, $1 \le j \le 100$, ezért $E\xi = \frac{1}{2} \cdot 100 = 50$, $\mathrm{Var}\,\xi = \frac{1}{4} \cdot 100 = 25$.

3.) Egy szabályos dobókockát feldobunk 100-szor egymás után. Számoljuk ki a dobáseredmények összegének várható értékét és szórásnégyzetét.

Megoldás: Legyenek $\eta_1, \dots, \eta_{100}$ *független* valószínűségi változók, amelyekre $P(\eta_j = k) = \frac{1}{6}$, $1 \le j \le 100$, $1 \le k \le 6$. Ekkor a kiszámítandó várható érték és szórásnégyzet $E \sum\limits_{j=1}^{100} \eta_j = \sum\limits_{j=1}^{100} E\eta_j$, $\mathrm{Var} \sum\limits_{j=1}^{100} \eta_j = \sum\limits_{j=1}^{100} \mathrm{Var}\,\eta_j$. (A második reláció felhasználja a tekintett valószínűségi változók függetlenségét.) $E\eta_j = \frac{1}{6}(1+2+3+4+5+6) = 3.5$, $E\eta_j^2 = \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}$, $\mathrm{Var}\,\eta_j = \frac{35}{12}$, $E\xi = 350$, $\mathrm{Var}\,\xi = \frac{3500}{12}$.

4.) Egy szabályos dobókockát feldobunk 100-szor egymás után. Tekintsük a páros értékű dobások eredményeinek az összegét, és számítsuk ki ennek várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 100$ valószínűségi változókat: $\xi_j = 2$, ha a j-ik dobás eredménye 2, $\xi_j = 4$, ha a j-ik dobás eredménye 4, $\xi_j = 6$, ha a j-ik dobás eredménye 6, és $\xi_j = 0$, ha a j-ik dobás eredménye 1, 3 vagy 5. Ekkor a $S = \sum\limits_{j=1}^{100} \xi_j$ valószínűségi változó várható értékét és szórásnégyzetét akarjuk kiszámolni. $E\xi_j = \frac{1}{6}(2+4+6) = 2$, $E\xi_j^2 = \frac{1}{6}(2^2+4^2+6^2) = \frac{28}{3}$, $\mathrm{Var}\,\xi_j = E\xi_j^2 - (E\xi_j)^2 = \frac{16}{3}$, ahonnan a ξ_j valószínűségi változók függetlensége miatt $ES = 100E\xi_1 = 200$, $\mathrm{Var}\,S = 100\mathrm{Var}\,\xi_1 = \frac{1600}{3}$.

Megfogalmazom az előző tétel egy olyan általánosítását, amely lehetővé teszi, hogy kiszámoljuk nem feltétlenül független valószínűségi változók összegének a szórásnégyzetét. Ennek érdekében először bevezetek egy új fogalmat, két valószínűségi változó kovarianciafüggvényét. Ez azt méri hogy milyen erős a kapcsolat két valószínűségi változó között.

A kovarianciafüggvény definiciója. *Legyen ξ és η két valószínűségi változó, amelyek mindegyikének létezik szórásnégyzete, azaz $E\xi^2 < \infty$ és $E\eta^2 < \infty$. Ekkor a ξ és η valószínűségi változók kovarianciafüggvényét* $\mathrm{Cov}\,(\xi, \eta)$*-t a*

$$\mathrm{Cov}\,(\xi, \eta) = E\left[(\xi - E\xi)(\eta - E\eta)\right]$$

kifejezés definiálja. Ha a $E\xi^2 < \infty$ és $E\eta^2 < \infty$ feltételek valamelyike nem teljesül, akkor nem definiáljuk a $\mathrm{Cov}\,(\xi, \eta)$ *kovarianciafüggvényt.*

Megjegyzés: $|(\xi - E\xi)||(\eta - E\eta)| \le \frac{|\xi - E\xi|^2}{2} + \frac{|\eta - E\eta|^2}{2}$, ahonnan $E|(\xi - E\xi)||(\eta - E\eta)| \le \frac{E|\xi - E\xi|^2}{2} + \frac{E|\eta - E\eta|^2}{2} < \infty$. Innen következik, hogy az $E\xi^2 < \infty$ és $E\eta^2 < \infty$ feltételek teljesülése esetén a $\mathrm{Cov}\,(\xi, \eta)$-t definiáló várható érték valóban létezik.

Korrelálatlan valószínűségi változók definiciója. *Azt mondjuk, hogy két ξ és η valószínűségi változó korrelálatlan, ha* $\mathrm{Cov}\,(\xi, \eta) = 0$.

Lemma a kovariancia kiszámolásáról.

$$\mathrm{Cov}\,(\xi, \eta) = E\xi\eta - E\xi E\eta.$$

Következmény. *Ha ξ és η két független valószínűségi változó, akkor* $\mathrm{Cov}\,(\xi, \eta) = 0$, *azaz e két valószínűségi változó korrelálatlan.*

A lemma bizonyítása.

$$\mathrm{Cov}\,(\xi, \eta) = E[\xi\eta - \eta E\xi - \xi E\eta + E\xi E\eta] = E\xi\eta - E\xi E\eta - E\xi E\eta + E\xi E\eta = E\xi\eta - E\xi E\eta.$$

Most megfogalmazom az alábbi eredményt valószínűségi változók összegének a szórásnégyzetéről, amely akkor is érvényes, ha az összeadandók nem függetlenek.

Tétel valószínűségi változók összegének a szórásnégyzetéről. *Ha $\xi_1, \dots, \xi_n$ valószínűségi változók ugyanazon a valószínűségi mezőn, amelyek mindegyikének létezik szórásnégyzete, akkor*

$$\begin{aligned}
\mathrm{Var}\left(\sum_{j=1}^n \xi_j\right) &= \sum_{j=1}^n \mathrm{Var}\,\xi_j + 2 \sum_{1 \le j < k \le n} (E\xi_j\xi_k - E\xi_j E\xi_k) \\
&= \sum_{j=1}^n \mathrm{Var}\,\xi_j + 2 \sum_{1 \le j < k \le n} \mathrm{Cov}\,(\xi_j, \xi_k) \\
&= \sum_{j=1}^n \mathrm{Var}\,\xi_j + \sum_{1 \le j,k \le n,\ j \ne k} \mathrm{Cov}\,(\xi_j, \xi_k).
\end{aligned}$$

Bizonyítás.

$$\begin{aligned}
\mathrm{Var}\left(\sum_{j=1}^n \xi_j\right) &= E\left(\sum_{j=1}^n \xi_j\right)^2 - \left(E\sum_{j=1}^n \xi_j\right)^2 = E\left(\sum_{j=1}^m \sum_{k=1}^n \xi_j\xi_k\right) - \left(\sum_{j=1}^m \sum_{k=1}^n E\xi_j E\xi_k\right) \\
&= \left(\sum_{j=1}^m \sum_{k=1}^n E\xi_j\xi_k\right) - \left(\sum_{j=1}^m \sum_{k=1}^n E\xi_j E\xi_k\right) \\
&= \sum_{j=1}^n \left(E\xi_j^2 - (E\xi_j)^2\right) + 2 \sum_{1 \le j < k \le n} (E\xi_j\xi_k - E\xi_j E\xi_k),
\end{aligned}$$

és $E\xi_j^2 - (E\xi_j)^2 = \text{Var}\,\xi_j$, $E\xi_j\xi_k - E\xi_j E\xi_k = \text{Cov}\,(\xi_j, \xi_k)$.

Házi feladat:

Mutassuk meg, hogy a független valószínűségi változók összegeinek szórásnégyzetéről szóló tétel következik a valószínűségi változók összegének a szórásnégyzetéről szóló tételből.

A következő két feladatban példát mutatok arra, hogy a fenti eredmény segítségével bizonyos esetekben viszonylag egyszerűen ki tudjuk számolni valószínűségi változók összegének a szórásnégyzetét.

Feladatok:

5.) Legyen egy urnában 20 piros és 30 fehér golyó. Húzzunk ki 20 golyót visszatevés nélkül. Számoljuk ki a kihúzott piros golyók számának várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 20$, valószínűségi változókat: $\xi_j(\omega) = 1$, ha a j-ik húzás eredménye piros, $\xi_j(\omega) = 0$, ha a j-ik húzás eredménye fehér. Ekkor a $\xi = \sum\limits_{j=1}^{20} \xi_j$ összeg várható értékét és szórásnégyzetét kell kiszámolnunk. Továbbá $E\xi_j = E\xi_1 = \frac{2}{5}$, $\text{Var}\,\xi_j = \text{Var}\,\xi_1 = \frac{2}{5} - \frac{4}{25} = \frac{6}{25}$, $E\xi_j\xi_k - E\xi_j E\xi_k = E\xi_1\xi_2 - E\xi_1 E\xi_2 = \frac{2}{5} \cdot \frac{19}{49} - \frac{4}{25} = -\frac{6}{1245}$, ha $j \ne k$. Itt felhasználtuk, hogy annak valószínűsége, hogy a j-ik húzásban piros golyót húzunk minden $1 \le j \le 20$ indexre ugyanannyi, mint annak a valószínűsége, hogy az első húzásban piros golyót húzunk. Hasonlóan, annak a valószínűsége, hogy a j-ik és k-ik húzásban piros golyót húzunk, $1 \le j, k \le 20$, $j \ne k$, ugyanannyi, mint annak a valószínűsége, hogy az első két húzásban piros golyót húzunk.

Innen a 20 dobásban kihúzott piros golyók számának várható értéke $20 \cdot \frac{2}{5} = 8$ és szórásnégyzete $20 \cdot \frac{6}{25} - 380 \cdot \frac{6}{1245} = \frac{144}{49}$.

6.) Feldobunk egy szabályos pénzérmét 100-szor egymás után. Tekintsük az egymást követő fej-fej dobássorozatok számát, és számítsuk ki ennek várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 99$, valószínűségi változókat: $\xi_j = 1$, ha a j-ik és $j+1$-ik dobások mindegyikének eredménye fej, $\xi_j = 0$ egyébként. Vegyük észre, hogy minket az $S = \sum\limits_{j=1}^{99} \xi_j$ valószínűségi változó várható értéke érdekel. Ezért $ES = E\left(\sum\limits_{j=1}^{99} E\xi_j\right) = \frac{99}{4}$, mivel $E\xi_j = \frac{1}{4}$. (Érdemes megjegyezni, hogy az ebben a feladatban tekintett ξ_j valószínűségi változók nem függetlenek, de a függetlenségre nincs szükség a várható érték additivításához.)

A szórásnégyzet kiszámításában viszont figyelembe kell vennünk azt, hogy nem csupa független valószínűségi változó összegét vizsgáljuk. Használjuk a szórásnégyzet kiszámolásánál a következő formulát.

$$\operatorname{Var} S = \operatorname{Var}\left(\sum_{j=1}^{99} \xi_j\right) = \sum_{j=1}^{99} \operatorname{Var} \xi_j + 2 \sum_{1\le j<k\le 99} \operatorname{Cov}(\xi_j, \xi_k).$$

Továbbá $\operatorname{Cov}(\xi_j, \xi_k) = 0$, ha $k \ge j+2$, mert ebben az esetben ξ_j és ξ_k függetlenek, és $\operatorname{Cov}(\xi_j, \xi_{j+1}) = \frac{1}{8} - \frac{1}{16} = \frac{1}{16}$ minden $1 \le j \le 98$ számra. Ugyanis $E\xi_j\xi_{j+1} = \frac{1}{8}$, mivel $\xi_j\xi_{j+1} = 1$, ha a j-ik, $j+1$-ik és $j+2$-ik dobások mindegyike fej, aminek valószínűsége $\frac{1}{8}$, és $\xi_j\xi_{j+1} = 0$ egyébként. Továbbá $E\xi_j E\xi_{j+1} = \frac{1}{16}$. Ezenkívül $\operatorname{Var} \xi_j = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$. Innen $\operatorname{Var} S = 99 \cdot \frac{3}{16} + 2 \cdot \frac{98}{16} = \frac{493}{16}$.

Független valószínűségi változók. A konvolució és generátorfüggvény.

Következő témánk a legfontosabb diszkrét eloszlások és azok tulajdonságainak tárgyalása lesz. Ki fogom számolni az ilyen valószínűségi változók várható értékét és szórásnégyzetét, és adott eloszlású független valószínűségi változók összegének az eloszlását. A kérdések egyszerűbb tárgyalása érdekében bevezetek néhány fogalmat, és felidézek néhány eredményt az analízisből. Először eloszlások konvolucióját majd generátorfüggvényét definiálom, és tárgyalom azt, hogy ezek a fogalmak milyen kapcsolatban vannak a minket érdeklő problémákkal.

Az első vizsgálandó kérdés az, hogy hogyan lehet kiszámolni független, diszkrét eloszlású valószínűségi változók összegének az eloszlását. Ennek kapcsán bevezetem diszkrét eloszlású valószínűségi változók eloszlásainak a konvolucióját, és teszek néhány megjegyzést.

Legyen ξ és η két független, diszkrét eloszlású valószínűségi változó, amelyek valamilyen $x_1, x_2, \dots$ értékeket vesznek fel. Ekkor ezek összege, a $\xi+\eta$ valószínűségi változó szintén diszkrét, és eloszlását ki tudjuk számolni a következő módon:

$$P(\xi+\eta = x) = \sum_{(x_j,x_k)\colon\, x_j+x_k=x} P(\xi = x_j, \eta = x_k) = \sum_{(x_j,x_k)\colon\, x_j+x_k=x} P(\xi = x_j)P(\eta = x_k)$$

Ez a képlet egyszerűsödik abban a fontos speciális esetben, ha a független ξ és η valószínűségi változók csak egész értékeket vesznek fel, és még egyszerűbb akkor, ha csak nem negatív egész értékeket vesznek fel. Azt az eredményt, amelyet ebben a speciális esetben kapunk, megfogalmazom tétel formájában is.

Tétel független, diszkrét eloszlású független valószínűségi összegének a kiszámolásáról. *Legyen ξ és η két független, egész értékeket felvevő valószínűségi változó. Ekkor a $\xi+\eta$ valószínűségi változó is csak egész értékeket vesz fel, és*

$$P(\xi+\eta = n) = \sum_{k=-\infty}^{\infty} P(\xi = k)P(\eta = n-k), \quad n = 0, \pm1, \pm2, \dots.$$

Ez a formula tovább egyszerűsödik, ha a független ξ és η valószínűségi változók csak nem negatív egész értékeket vesznek fel. Ekkor

$$P(\xi + \eta = n) = \sum_{k=0}^{n} P(\xi = k)P(\eta = n - k), \quad \text{ha } n \geq 0,$$

és

$$P(\xi + \eta = n) = 0, \quad \text{ha } n < 0.$$

Bizonyítás: A tétel első képlete azonnal adódik a tétel előtt tárgyalt általános esetből abban az esetben, ha a valószínűségi változók $x_1, x_2, \dots$ értékei egész számok. A második képlet az első képlet következménye, ha észrevesszük, hogy elég csak azon $(k, n-k)$ párokra összegezni, amelyekre $P(\xi = k)P(\eta = n-k) \neq 0$. Ez azt jelenti, hogy abban az esetben, ha ξ és η nem negatív egész értékeket vesznek fel, akkor az összegezésből kihagyhatjuk azokat a tagokat, amelyekre $k < 0$ vagy $n - k < 0$, azaz $k > n$.

A fenti eredmény miatt érdemes bevezetni a következő definiciót.

Diszkrét, egész értékű eloszlásások konvoluciójának a definiciója. *Legyen*

$$\mathcal{P} = \{p_k\colon\ k = 0, \pm 1, \pm 2, \dots\} \quad \text{és} \quad \mathcal{Q} = \{q_k\colon\ k = 0, \pm 1, \pm 2, \dots\}$$

két az egész számokon értelmezett diszkrét eloszás, azaz tegyük fel, hogy $p_k \geq 0$, $q_k \geq 0$, $k = 0, \pm 1, \pm 2, \dots$, $\sum\limits_{k=-\infty}^{\infty} p_k = 1$ és $\sum\limits_{k=-\infty}^{\infty} q_k = 1$. Ekkor a $\mathcal{P}$ és $\mathcal{Q}$ eloszlások konvoluciója, az a $\mathcal{R} = \mathcal{P} * \mathcal{Q} = \{r_n\colon\ n = 0, \pm 1, \pm 2, \dots\}$ eloszlás, amelyre

$$r_n = \sum_{k=-\infty}^{\infty} p_k q_{n-k}, \quad n = 0, \pm 1, \pm 2, \dots.$$

Abban a speciális esetben, ha $p_k = 0$ és $q_k = 0$ minden $k < 0$ számra, az $r_n = 0$ azonosság érvényes minden $n < 0$ indexre, és

$$r_n = \sum_{k=0}^{n} p_k q_{n-k}, \quad n = 0, 1, 2, \dots.$$

A generátor függvények fogalma.

Érdemes a következő észrevételt tenni. Legyen $f(x)$ és $g(x)$ két függvény, amelyek hatványsorba fejthetőek, azaz,

$$f(x) = \sum_{k=0}^{\infty} a_k x^k, \quad g(x) = \sum_{k=0}^{\infty} b_k x^k,$$

ha $-A < x < A$ valamilyen $A > 0$ számmal. Ekkor az $f(x)g(x)$ szorzat is sorba fejthető, azaz

$$f(x)g(x) = \sum_{n=0}^{\infty} c_n x^n$$

(ugyanabban a $-A < x < A$ intervallumban), és

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

Vegyük észre, hogy a fenti hatványsorokban szereplő a_k, b_k és c_n együtthatók között hasonló reláció érvényes mint a nem negatív egész értéket felvevő eloszlások konvoluciójának a definiciójában. Ezenkívül érdemes ismerni az analízis alábbi eredményeit.

a.) Ha az $f(x)$ függvény hatványsorba fejthető, azaz $f(x) = \sum\limits_{k=0}^{\infty} a_k x^k$ valamilyen $-A < x < A$ intervallumon, $A > 0$, akkor az $f(x)$ függvény értékei a $-A < x < A$ intervallumon meghatározzák az a_n együtthatókat. (Sőt, azok explicit módon kiszámíthatóak az $n!a_n = \left.\frac{d^n f(x)}{dx^n}\right|_{x=0}$ képlet segítségével.)

b.) Legyen $f_n(x)$, $n = 1, 2, \dots$, olyan függvények sorozata, amelyek mindegyike előállítható $f_n(x) = \sum\limits_{k=0}^{\infty} a_{k,n} x^k$ hatványsor alakban valamilyen $-A < x < A$, $A > 0$, intervallumon, ahol az $A > 0$ szám nem függ az n indextől, továbbá létezzen az $f(x) = \lim\limits_{n\to\infty} f_n(x)$ határérték minden $-A < x < A$ számra, sőt ez a konvergencia teljesüljön minden $|x| < A$ számra a komplex számsíkon, és az legyen egyenletes ezen a halmazon. Ekkor az $f(x)$ határfüggvény is hatványsorba fejthető $f(x) = \sum\limits_{k=0}^{\infty} a_k x^k$ alakban a $-A < x < A$ intervallumon alkalmas a_k együtthatókkal, és $a_k = \lim\limits_{n\to\infty} a_{k,n}$ minden $k = 0, 1, 2, \dots$ együtthatóra.

Ez az eredmény következik Weierstrass egy fontos komplex függvénytani tételéből. E szerint a tétel szerint (ld. például Szőkefalvi Nagy Béla komplex fügvénytan jegyzetében a III. rész 8. paragrafusát), ha $f_1(z), f_2(z), \dots$ holomorf függvények egy sorozata egy T tartományban, és e függvények egyenletesen konvergálnak a T tartomány minden kompakt részhalmazában egy $f(z)$ függvényhez, akkor ez az $f(z)$ függvény is holomorf a T tartományban, és minden $k = 1, 2, \dots$ számra az $f_n^{(k)}(z)$ függvénysorozat egyenletesen konvergál az $f^{(k)}(z)$ függvényhez a T halmaz minden kompakt részhalmazában. (Ebben az állításban a k felső index k-ik deriváltat jelöl.)

c.) Felmerülhet a kérdés: Ha azt akarjuk biztosítani, hogy valamely hatványsorok segítségével definiált $f_n(x) = \sum\limits_{k=0}^{\infty} a_{k,n} x^k$, $n = 1, 2, \dots$, függvények Taylor sorainak együtthatói úgy viselkedjenek, mint ahogy azt a b) részben előírtuk, azaz legyen egy olyan $f(x)$ függvény, amely $f(x) = \sum\limits_{k=0}^{\infty} a_k x^k$ alakú, és teljesüljön az

$a_k = \lim\limits_{n\to\infty} a_{k,n}$ reláció minden $k = 0, 1, 2, \dots$ indexre, akkor ehhez nem elég-e csak annyit megkövetelni, hogy az $f_n(x)$ függvények konvergáljanak az $f(x)$ függvényhez valamely $-A \le x \le A$ intervallumban. (Egy ilyen feltételt könnyebb ellenőrizni, mint a $\lim\limits_{n\to\infty} f_n(z) = f(z)$ relációt, és a konvergencia egyenletességét a komplex számsík egy origó középpontú körében.) Az általános esetben ez a gyengített feltétel túl kevés, de egy a valószínűségi alkalmazásokban nagyon fontos speciális esetben már ez is elegendő a számunkra.

Ha az $f_n(x) = \sum\limits_{k=0}^{\infty} a_{k,n} x^k$ hatványsorok együtthatói teljesítik az $a_{k,n} \ge 0$ feltételt minden $k = 0, 1, 2, \dots,$ $n = 1, 2, \dots,$ indexpárra, akkor annak érdekében, hogy az $f_n(x)$ függvények Taylor sor együtthatói teljesítsék a b) részben megfogalmazott konvergencia relációkat elég azt megkövetelni, hogy az $f_n(x)$ függvények konvergáljanak egy $f(x)$ függvényhez valamely $-A \le x \le A$ intervallumon. Ekkor ugyanis az $f_n(z)$ függvények egyenletesen korlátosak a komplex számsík $\{z\colon |z| \le A\}$ körében. Innen viszont a komplex függvénytan egy másik fontos eredményéből, a Vitali tételből (ld. például Szőkefalvi Nagy Béla komplex fügvénytan jegyzetében a VII. rész 4. paragrafusát) következik, hogy az (analitikus) $f_n(z)$ függvények egyenletesen konvergálnak a komplex számsík $\{z\colon |z| \le \frac{A}{2}\}$ origó sugarú körében is egy függvényhez, amely nem más, mint az $f(x)$ függvény (létező) analitikus kiterjesztése. Ez azt jelenti, hogy ebben az esetben az $f_n(x)$ függvények teljesítik a b) pontban megfogalmazott feltételeket, és alkalmazhatjuk rájuk az ott megfogalmazott állítást.

Megjegyzés. Konvergáljanak valamely $f_n(x) = \sum\limits_{k=0}^{\infty} a_{k,n} x^k$, $n = 1, 2, \dots,$ Taylor sorral megadott függvények (akár egyenletesen) egy $f(x)$ függvényhez valamely $-A \le x \le A$ intervallumban. Ha nem tesszük fel, hogy az $a_{k,n}$ Taylor együtthatók nem negatívak, akkor ebből nem következik, hogy az $f(x)$ függvény Taylor sorba fejthető, és az $a_{k,n}$ együtthatók az $f(x)$ függvény Taylor sorának megfelelő együtthatójához konvergálnak, ha $n \to \infty$. Valóban, Weierstrass első approximációs tétele szerint minden a $-A \le x \le A$ intervallumban folytonos (akár nem differenciálható) függvényhez található polinomok olyan sorozata, amelyek egyenletesen konvergálnak ehhez a függvényhez a $-A \le x \le A$ intervallumban, és ezen észrevétel alapján nem nehéz ellenpéldát találni.

Később látni fogjuk, hogy a hatványsorok fent említett tulajdonságai jól használhatóak bizonyos valószínűségi vizsgálatokban. Ugyanis egy nem-negatív egész értékékeket felvevő ξ valószínűségi változó $p_k = P(\xi = k)$, $k = 0, 1, 2, \dots,$ eloszlásához hozzárendelhetjük az $f(x) = \sum\limits_{k=0}^{\infty} p_k x_k$ függvényt, amelyet ezen eloszlás generátorfüggvényének neveznek. Megfogalmazom e fogalom definicióját pontosabban is.

Diszkrét, nem negatív, egész értékű eloszlás generátorfüggvényének a definiciója: *Legyen adva egy $\mathcal{P} = \{p_0, p_1, p_2, \dots\}$ eloszlás a nem-negatív egész számokon, azaz olyan számsorozat, amelynek tagjaira $p_k \ge 0$ minden $0 \le k \le \infty$ számra, és $\sum\limits_{k=1}^{\infty} p_k =$*

1. *(Az elnevezés oka az, hogy létezik olyan ξ nem-negatív egész értékű valószínűségi változó, amelyre $P(\xi = k) = p_k$ minden $k = 0, 1, 2, \dots$ számra.) Ennek az eloszlásnak a generátorfüggvénye az*

$$F(x) = \sum_{k=0}^{\infty} p_k x^k$$

függvény.

A fent mondottak alapján nemcsak az igaz, hogy egy $\mathcal{P}$ eloszlás meghatározza $F(x)$ generátorfüggvényét, hanem megfordítva az $F(x)$ generátorfüggvény segítségével ki lehet számolni azt az eloszlást, amelynek ez a generátorfüggvénye, mivel

$$k!p_k = \left.\frac{dF^k(x)}{dx^k}\right|_{x=0}, \qquad k = 0, 1, 2, \dots .$$

Egy eloszlás generátorfüggvénye mindig konvergens a $-1 \le x \le 1$ intervallumban. Ha egy $\mathcal{P}$ eloszlás generátorfüggvénye $F(x)$ egy $\mathcal{Q}$ eloszlás generátorfüggvénye $G(x)$, akkor $\mathcal{P} * \mathcal{Q}$ konvoluciójuk generátorfüggvénye az $F(x)G(x)$ függvény. Továbbá, eloszlások generátorfüggvényeinek konvergenciájából maguknak az eloszlásoknak a konvergenciájára is következtetni tudunk. Részletesebben kifejtve a következőről van szó.

Legyen adva minden $n = 1, 2, \dots$ számra egy $p_{n,k}$, $k = 0, 1, 2, \dots$, eloszlás a nem negatív egész számokon, azaz teljesüljenek a $p_{n,k} \ge 0$ és $\sum\limits_{k=0}^{\infty} p_{n,k} = 1$ tulajdonságok. Feleltessük meg ezeknek az eloszlásoknak az $F_n(x) = \sum\limits_{k=0}^{\infty} p_{n,k} x^k$ generátorfüggvényeket (hatványsorokat). Ha az $F_n(x)$ generátorfüggvények pontonként konvergálnak egy $F(x)$ függvényhez valamely $-A \le x \le A$, $0 < A \le 1$ intervallumban, akkor ez az $F(x)$ függvény felírható $F(x) = \sum\limits_{k=0}^{\infty} p_k x^k$, $-1 < x < 1$, hatványsor alakban, és $\lim\limits_{n\to\infty} p_{n,k} = p_k$ minden $k = 0, 1, 2, \dots$ számra.

A fenti eredmény hasznos módszert biztosít annak ellenőrzésére, hogy bizonyos eloszlások konvergálnak-e. Megjegyzem, hogy a megfogalmazott eredményben nem állítottam, hogy a p_k határértékek teljesítik a $\sum\limits_{k=0}^{\infty} p_k = 1$ azonosságot, azaz valószínűségi eloszlást határoznak meg. A legtöbb konkrét feladatban ez a tulajdonság teljesül, de azt külön ellenőrizni kell.

A fenti eredmények lehetővé teszik, hogy bizonyos valószínűségi vizsgálatokat viszonylag egyszerűen végrehajthassunk generátorfüggvények segítségével.

További feladatok a várható érték és szórásnégyzet kiszámolásáról

1.) Egy urnában 10 piros és 20 fehér golyó van. Kihúzunk visszatevés nélkül 10 golyót. A páros sorszámú húzások esetén fehér golyó húzás esetén nyerünk 3 forintot, piros golyó húzás esetén pedig nem nyerünk és nem veszítünk semmit. A páratlan sorszámú húzások esetén piros húzás esetén 2 forintot nyerünk, fehér golyó húzás esetén nem nyerünk és nem veszítünk semmit. Számoljuk ki a nyereményünk várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 10$, valószínűségi változókat: Páros j számokra $\xi_j = 3$, ha a j-ik húzás eredménye piros, $\xi_j = 0$, ha a j-ik húzás eredménye fehér golyó. Páratlan j számokra $\xi_j = 2$, ha a j-ik húzás eredménye piros, $\xi_j = 0$, ha a j-ik húzás eredménye fehér golyó. Ekkor az $S = \sum\limits_{j=1}^{10} \xi_j$ összeg várható értékét és szórásnégyzetét kell kiszámolnunk. Ennek érdekében számoljuk ki az $E\xi_j$ várható értékeket, $\operatorname{Var}\xi_j$ szórásnégyzeteket és $\operatorname{Cov}(\xi_j, \xi_k)$ kovarianciákat. Annak valószínűsége, hogy a j-ik húzás eredménye piros $\frac{1}{3}$, annak valószínűsége, hogy j-ik húzás eredménye fehér $\frac{2}{3}$. Ezért $E\xi_j = 2$, ha j páros, $E\xi_j = \frac{2}{3}$, ha j páratlan, és $ES = 5\left(2 + \frac{2}{3}\right) = 13\frac{1}{3}$.

A szórásnégyzet kiszámítása érdekében vegyük észre, hogy $E\xi_j^2 = 6$, $\operatorname{Var}\xi_j = 2$, ha j páros, és $E\xi_j^2 = \frac{4}{3}$, $\operatorname{Var}\xi_j = \frac{8}{9}$. Továbbá annak valószínűsége, hogy két j és k indexre $j \ne k$, a j-ik és k-ik húzás mindegyike fehér $\frac{2\cdot 19}{3\cdot 29} = \frac{38}{87}$, annak, hogy mindkét húzás piros $\frac{9}{3\cdot 29} = \frac{3}{29}$, annak, hogy az egyik húzás fehér, a másik húzás piros $\frac{2\cdot 10}{3\cdot 29} = \frac{20}{57}$. Ezért $E\xi_j\xi_k = 9 \cdot \frac{2\cdot 19}{3\cdot 29} = \frac{114}{29}$, $\operatorname{Cov}(\xi_j, \xi_k) = \frac{114}{29} - 4 = -\frac{2}{29}$, ha j és k páros, $E\xi_j\xi_k = 4 \cdot \frac{3}{29} = \frac{12}{29}$, $\operatorname{Cov}(\xi_j, \xi_k) = \frac{12}{29} - \frac{4}{9} = -\frac{8}{263}$, ha j és k páratlan, $E\xi_j\xi_k = 6 \cdot \frac{20}{57} = \frac{40}{29}$, $\operatorname{Cov}(\xi_j, \xi_k) = \frac{40}{29} - \frac{4}{3} = \frac{4}{87}$, ha j és k közül az egyik páros, a másik páratlan. Olyan (j,k) pár, $1 \le j,k \le 10$, $j \ne k$, melyre j és k mindegyike páros vagy mindegyike páratlan, összesen 20 van, és olyan (j,k) pár, amelyekre az egyik páros, a másik páratlan $2 \cdot 5 \cdot 5 = 50$ van, ezért

$$\sum_{1\le j,k\le 10,\, j\ne k} \operatorname{Cov}(\xi_j, \xi_k) = 20\left(-\frac{2}{29} - \frac{8}{263}\right) + 50 \cdot \frac{4}{87} = \frac{80}{263}.$$

Innen

$$\begin{aligned}\operatorname{Var} S &= \sum_{j=1}^{10} \operatorname{Var}\xi_j + \sum_{1\le j,k\le 10,\, j\ne k} \operatorname{Cov}(\xi_j, \xi_k) \\ &= 5\left(2 + \frac{8}{9}\right) + \frac{80}{263} = \frac{130}{9} + \frac{80}{263} = \frac{3850}{263}.\end{aligned}$$

2.) Egy estélyen megjelenik n házaspár. Egy táncmester, aki nem tudja, hogy kik házastársak és kik nem, véletlen módon párba rendezi a férfiakat és nőket a tánc előtt. Számoljuk ki az egymással táncoló házaspárok számának a várható értékét és szórásnégyzetét. Mi ezek határértéke, ha $n \to \infty$?

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le n$, valószínűségi változókat: $\xi_j = 1$, ha a j-ik házaspár együtt táncol, és $\xi_j = 0$, ha nem táncolnak együtt. Ekkor minket az $S_n = \sum\limits_{j=1}^{n} \xi_j$ valószínűségi változó várható értéke és szórásnégyzete érdekel. Ennek kiszámítása érdekében vegyük észre, hogy $P(\xi_j = 1) = 1 - P(\xi_j = 0) = \frac{1}{n}$ minden $1 \le j \le n$ indexre, és $P(\xi_j = 1, \xi_k = 1) = \frac{1}{n(n-1)}$ minden $1 \le j, k \le n$, $j \ne k$ indexre. Innen $E\xi_j = E\xi_j^2 = \frac{1}{n}$, $1 \le j \le n$, $E\xi_j\xi_k = \frac{1}{n(n-1)}$, ha $j \ne k$, $\operatorname{Var}\xi_j = \frac{1}{n} - \frac{1}{n^2}$, és $\operatorname{Cov}(\xi_j, \xi_k) = E\xi_j\xi_k - E\xi_j E\xi_k = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$, ha $j \ne k$. Ezért $ES_n = \sum\limits_{j=1}^{n} E\xi_j = 1$, $\operatorname{Var} S_n = \sum\limits_{j=1}^{n} \operatorname{Var}\xi_j + 2 \sum\limits_{1 \le j < k \le n} \operatorname{Cov}(\xi_j, \xi_k) = n\left(\frac{1}{n} - \frac{1}{n^2}\right) + n(n-1)\frac{1}{n^2(n-1)} = 1$.

3.) Feldobunk egy dobókockát, majd ezután egy szabályos pénzdarabot annyi alkalommal, amennyi a kockadobás eredménye volt. (Az egyes dobások eredményei függetlenek egymástól.) Számoljuk ki a fejdobások számának várható értékét és szórásnégyzetét.

Megoldás: Vezessük be a következő $\xi_{j,k}$, $1 \le j \le 6$, $1 \le k \le j$, valószínűségi változókat:

$$\xi_{j,k}(\omega) = \begin{cases} 1 & \text{ha a kockadobás eredménye } j \text{ és a } k\text{-ik pénzdobásé fej} \\ 0 & \text{ha a kockadobás eredménye } j \text{ és a } k\text{-ik pénzdobásé írás} \\ 0 & \text{ha a kockadobás eredménye nem } j \end{cases}$$

Ekkor minket az $S = \sum\limits_{j=1}^{6} \sum\limits_{k=1}^{j} \xi_{j,k}$ valószínűségi változó várható értéke és szórásnégyzete érdekel. Vegyük észre, hogy

$$E\xi_{j,k} = P(\text{a kockadobás eredménye } j, \text{ a } k\text{-ik pénzdobásé fej}) = \frac{1}{12},$$

$E\xi_{j,k}^2 = \frac{1}{12}$, $\operatorname{Var}\xi_{j,k}^2 = \frac{11}{144}$, $E\xi_{j,k}\xi_{j,k'} = \frac{1}{24}$, ha $k \ne k'$, ahonnan $\operatorname{Cov}(\xi_{j,k}, \xi_{j,k'}) = \frac{5}{144}$ ebben az esetben. Továbbá $E\xi_{j,k}\xi_{j',k'} = 0$, $\operatorname{Cov}(\xi_{j,k}, \xi_{j',k'}) = -\frac{1}{144}$, ha $j \ne j'$ (függetlenül a k és k' számok viszonyától. Innen $ES = \sum\limits_{j=1}^{6} \sum\limits_{k=1}^{j} E\xi_{j,k} = \frac{1+\cdots+6}{12} = \frac{21}{12} = \frac{7}{4}$, és

$$\operatorname{Var} S = \sum_{j=1}^{6} \sum_{k=1}^{j} \operatorname{Var}\xi_{j,k} + \sum_{(j,k),(j',k')\colon\, j \ne j' \text{ vagy } k \ne k'} \operatorname{Cov}(\xi_{j,k}, \xi_{j',k'}).$$

Ezért $\operatorname{Var} S = \frac{11}{144}(1+\cdots+6) - \frac{1\cdot(21-1)+2\cdot(21-2)+\cdots 6\cdot(21-6)}{144} + \frac{5}{144}(2\cdot 1+3\cdot 2+\cdots 6\cdot 5)$. Ezt a képletet úgy kapjuk, hogy külön számoljuk azon $\operatorname{Cov}(\xi_{j,k}, \xi_{j',k'})$ mennyiségek hozadékát amelyekre $j \ne j'$ és azután azokét, amelyekre $j = j'$, de $k \ne k'$. Innen $\operatorname{Var} S = \frac{231-350+350}{144} = \frac{77}{48}$.

4.) Csodaország munka törvénykönyve szerint egy cég minden munkása fizetett szabadságot kap azokon a napokon, amikor legalább az egyiküknek születésnapja van. Ezen napok kivételével azonban az év minden napján mindenkinek dolgoznia kell. Minden munkás 1 TV-készüléket készít egy nap alatt. Mi a várható értéke és szórásnégyzete az egy évben gyártott TV készülékeknek, ha n munkás dolgozik a cégben? Hány alkalmazottat vegyen fel a cégtulajdonos, ha azt akarja, hogy a gyártott TV-készülékek számának a várható értéke a lehető legnagyobb legyen?

Megoldás. Jelölje ξ_j, $1 \le j \le 365$, a j-ik nap gyártott TV-készülékek számát. Ekkor $\xi_j = n$, ha az n munkás egyikének sincs születésnapja a j-ik napon, és $\xi_j = 0$ egyébként. Ezért $P(\xi_j = n) = \left(\frac{364}{365}\right)^n$ és $P(\xi_j = 0) = 1 - \left(\frac{364}{365}\right)^n$. A gyártott TV-készülékek száma $Y = Y(n) = \sum\limits_{j=1}^{365} \xi_j$, ahonnan $EY = EY(n) = \sum\limits_{j=1}^{365} E\xi_j = 365n\left(\frac{364}{365}\right)^n$.

A szórásnégyzet kiszámolása érdekében számoljuk ki először a $\operatorname{Cov}(\xi_i, \xi_j) = E\xi_i\xi_j - E\xi_i E\xi_j$ kovarianciákat $1 \le i, j \le 365$, $i \ne j$ esetén. $E\xi_i\xi_j = n^2 P(\xi_i = n, \xi_j = n) = n^2\left(\frac{363}{365}\right)^n$, mert $\left(\frac{363}{365}\right)^n$ a valószínűsége annak, hogy egyik munkásnak sincs születésnapja sem az i-ik sem a j-ik napon. Innen

$$\operatorname{Cov}(\xi_i, \xi_j) = n^2\left(\left(\frac{363}{365}\right)^n - \left(\frac{364}{365}\right)^{2n}\right),$$

és mivel $\operatorname{Var} Y = \sum\limits_{j=1}^{365}(E\xi_j^2 - (E\xi_j)^2) + 2\sum\limits_{1\le i<j\le 365}\operatorname{Cov}(\xi_i, \xi_j)$

$$\operatorname{Var} Y = 365n^2\left(\left(\frac{364}{365}\right)^n - \left(\frac{364}{365}\right)^{2n}\right) + 365\cdot 364n^2\left(\left(\frac{363}{365}\right)^n - \left(\frac{364}{365}\right)^{2n}\right).$$

Számoljuk ki az $EY(n)$ maximumát az n változó szerint. Annak érdekében, hogy meghatározzuk mely n számra vétetik fel ez a maximum, számoljuk ki az $\frac{EY(n+1)}{EY(n)}$ hányadost minden pozitív egész n-re, és határozzuk meg, hogy az mely n-ekre kisebb, és mely n-ekre nagyobb, mint 1. $\frac{EY(n+1)}{EY(n)} = \frac{n+1}{n}\frac{364}{365}$, ahonnan $\frac{EY(n+1)}{EY(n)} > 1$, ha $n \le 363$, $\frac{EY(n+1)}{EY(n)} = 1$, ha $n = 364$, és $\frac{EY(n+1)}{EY(n)} < 1$, ha $n \le 365$. Ezért 364 vagy 365 munkást érdemes felvenni, és ekkor az évente gyártott TV készülékek száma $\frac{364^{365}}{365^{363}} = 364^2\left(1 - \frac{1}{365}\right)^{363} \sim \frac{364^2}{e}$.

5.) Egy kártyacsomag 75 kártyalapot tartalmaz, amelyek mindegyike az 1 és 75 közötti számok valamelyikével meg van számozva. Kihúzunk 40 kártyát visszatevéssel, és jelölje X az ily módon kapott különböző kártyák számát. Számoljuk ki az EX várható értéket.

Megoldás: Számozzuk meg a kártyákat 1-től 75-ig, és vezessük be a következő ξ_j, $1 \le j \le 75$, valószínűségi változókat. $\xi_j = 1$, ha a j-ik kártyát kiválasztjuk,

$\xi_j = 0$, ha a j-ik kártyát nem választjuk ki a 40 húzás során. Ekkor $X = \sum_{j=1}^{75} \xi_j$.

Ezért $EX = \sum_{j=1}^{75} E\xi_j = \sum_{j=1}^{75} P(\xi_j = 1)$. Annak a valószínűsége, hogy a j-ik kártyát nem húzzuk ki 40 húzás során $(\frac{74}{75})^{40}$. Innen $P(\xi_j = 1) = 1 - (\frac{74}{75})^{40}$, $EX = 75\left(1 - (\frac{74}{75})^{40}\right)$.

6.) Számoljuk ki az előző feladatban definiált X valószínűségi változó $\operatorname{Var} X$ szórásnégyzetet.

Jelölje Y a ki nem húzott kártyák számát, és vezessük be az $\eta_j = 1 - \xi_j$, $1 \le j \le 75$, valószínűségi változókat, amelyekre $y_j = 1$, ha a j-ik kártyát nem választjuk ki, és $\eta_j = 0$, ha a j-ik kártyát kiválasztjuk a 40 húzás során. Ekkor $Y = \sum_{j=1}^{75} \eta_j$, és $Y = 75 - X$, ahonnan $\operatorname{Var} Y = \operatorname{Var} X$ az előző feladatban definiált X valószínűségi változóval.

$\operatorname{Var} X = \operatorname{Var} Y = \sum_{j=1}^{75} \operatorname{Var} \eta_j + 2 \sum_{1 \le i < j \le 75} \operatorname{Cov}(\eta_i, \eta_j)$. Másrészt $\operatorname{Var} \eta_j = P(\eta_j = 1) - P(\eta_j = 1)^2$, és $\operatorname{Cov}(\eta_i, \eta_j) = P(\eta_i = 1, \eta_j = 1) - P(\eta_i = 1)P(\eta_j = 1)$, ha $i \ne j$. Továbbá $P(\eta_i = 1, \eta_j = 1) = \left(\frac{73}{75}\right)^{40}$, $P(\eta_i = 1) = P(\eta_j = 1) = \left(\frac{74}{75}\right)^{40}$. Innen $\operatorname{Cov}(\eta_i, \eta_j) = \left(\frac{73}{75}\right)^{40} - \left(\frac{74}{75}\right)^{80}$, és $\operatorname{Var} \eta_j = \left(\frac{74}{75}\right)^{40} - \left(\frac{74}{75}\right)^{80}$. Ezért

$$\operatorname{Var} X = \operatorname{Var} Y = 75 \left(\frac{74}{75}\right)^{40} \left(1 - \left(\frac{74}{75}\right)^{40}\right) + 74 \cdot 75 \left(\left(\frac{73}{75}\right)^{40} - \left(\frac{74}{75}\right)^{80}\right).$$

Megjegyzés: Az 5. feladatot a vizsgált X valószínűségi változó $X = \sum_{j=1}^{75} \xi_j$ alakú felbontásának segítségével oldottuk meg egyszerűbb valószínűségi változók összegeként. Más felbontással is kisérletezhetünk volna. Például írhattuk volna, hogy $X = \sum_{j=1}^{40} \zeta_j$, ahol $\zeta_j = 1$, ha a j-ik húzáskor új kártyát húzunk, és $\zeta_j = 0$, ha nem. Ekkor is felírhatjuk az $EX = \sum_{j=1}^{40} E\zeta_j$, azonosságot. Ez a módszer mégsem olyan hasznos, mint a feladat megoldásában alkalmazott eljárás, mert az $E\zeta_j = P(\zeta_j = 1)$ várható értékeket nem egyszerű kiszámolni. Ez azt mutatja, hogy érdemes a minket érdeklő valószínűségi változó számunkra hasznos felbontását keresni, és ez nem mindig magától értetődő.

7.) Egy kártyacsomag 75 kártyalapot tartalmaz, amelyek mindegyike az 1 és 75 közötti számok valamelyikével meg van számozva. Kihúzunk 40 kártyát úgy, hogy húzás után visszatesszük, és ezenkívül a kártyacsomagba teszünk még egy új, minden korábbi kártyától különböző kártyalapot. Jelölje X azt, hogy hány különböző kártyalapot húztunk ki. Számoljuk ki az EX várható értéket.

Megoldás: Számozzuk meg a kártyacsomagban kezdettől fogva tartalmazott kártyalapokat 1-től 75-ig, és vezessük be a következő ξ_j, $1 \le j \le 75$, valószínűségi változókat. $\xi_j = 1$, ha a j-ik kártyát kihúztuk, $\xi_j = 0$, ha a j-ik kártyát nem hútuk ki a 40 húzás során. Ezenkívül vezessük be a következő η_j, $1 \le j \le 40$, valószínűségi változókat is. $\eta_j = 1$, ha a j-ik húzás után a kártyacsomagba tett lapot később kihúztuk, és $\eta_j = 0$, ha nem húztuk ki. Ekkor $X = \sum_{j=1}^{75} \xi_j + \sum_{j=1}^{40} \eta_j$. Ezért $EX = \sum_{j=1}^{75} E\xi_j + \sum_{j=1}^{40} E\eta_j = \sum_{j=1}^{75} P(\xi_j = 1) + \sum_{j=1}^{40} P(\eta_j = 1)$. Annak a valószínűsége, hogy a j-ik (a csomagban eredetileg bennelevő) kártyát nem húzzuk ki 40 húzás során $\prod_{j=1}^{40} \frac{74+j-1}{75+j-1}$. Innen $P(\xi_j = 1) = 1 - \prod_{j=1}^{40} \frac{74+j-1}{75+j-1}$, $1 \le j \le 75$. Annak a valószínűsége, hogy a j-ik húzás után a kártyacsomagba tett kártyát nem húzzuk ki, $\prod_{l=j+1}^{40} \frac{74+l-1}{75+l-1}$. Ezért $E\eta_j = 1 - \prod_{l=j+1}^{40} \frac{74+l-1}{75+l-1}$, ha $1 \le j \le 39$, és $E\eta_{40} = 0$. Innen kapjuk, hogy

$$EX = 75\left(1 - \prod_{j=1}^{40} \frac{74+j-1}{75+j-1}\right) + \sum_{l=1}^{39}\left[1 - \prod_{l=j+1}^{40} \frac{74+l-1}{75+l-1}\right].$$

8.) Egy kártyacsomag 75 kártyalapot tartalmaz, amelyek mindegyike az 1 és 75 közötti számok valamelyikével meg van számozva. Kihúzunk 40 kártyát úgy, hogy a páratlan indexű húzások után a kártyát visszatesszük, a páros indexű húzások után pedig nem tesszük vissza a csomagba. Jelölje X azt, hogy hány különböző kártyalapot húztunk ki. Számoljuk ki az EX várható értéket.

Megoldás: Számozzuk meg a kártyacsomag lapjait 1-től 75-ig, és vezessük be a következő ξ_j, $1 \le j \le 75$, valószínűségi változókat. $\xi_j = 1$, ha a j-ik kártyát kihúztuk, $\xi_j = 0$, ha a j-ik kártyát nem húztuk ki a 40 húzás során. Ekkor $X = \sum_{j=1}^{75} \xi_j$, ezért $EX = \sum_{j=1}^{75} E\xi_j = \sum_{j=1}^{75} P(\xi_j = 1)$. Annak a valószínűsége, hogy a j-ik kártyát nem húzzuk ki 40 húzás során $\prod_{j=1}^{20} \left(\frac{74-j+1}{75-j+1}\right)^2$. Ezt a valószínűséget a következő észrevétel segítségével számoltuk ki. Jelölje $p_l = p_{l,j}$ annak a feltételes valószínűségét, hogy a j-ik lapot nem húztuk ki az l-ik húzásban, feltéve, hogy az első $l-1$ húzásban sem húztuk ki azt. (Legyen p_1 annak a valószínűsége, hogy a j-ik lapot az első húzásban nem húztuk ki.) Akkor annak valószínűsége, hogy a j-ik lapot 40 húzásban nem húztuk ki $\prod_{l=1}^{40} p_l = \prod_{j=1}^{20} \left(\frac{74-j+1}{75-j+1}\right)^2$. Innen $P(\xi_j = 1) = 1 - \prod_{j=1}^{20} \left(\frac{74-j+1}{75-j+1}\right)^2$, és $EX = 75\left(1 - \prod_{j=1}^{20} \left(\frac{74-j+1}{75-j+1}\right)^2\right)$.

9.) 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 húzunk. 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, piros) vagy (piros, fehér) húzás történik. Hasonlóan $P(\xi_j = 2) = \frac{1}{2}$. Ezért $E\xi_j = 2\frac{1}{2} - \frac{1}{2} = \frac{1}{2}$, $E\xi_j^2 = \frac{1}{2}(4+1) = \frac{5}{2}$, $\operatorname{Var}\xi_j = \frac{9}{4}$.

Hasonlóan, $P(\xi_j = 2, \xi_k = 2) = P(\xi_1 = 2, \xi_2 = 2) = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{9}{19} + 2 \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{10}{19} = \frac{1}{4}$, $P(\xi_j = 2, \xi_k = -1) = P(\xi_1 = 2, \xi_2 = -1) = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} \left(\frac{10}{19} + \frac{9}{19}\right) = \frac{1}{4}$, $P(\xi_j = -1, \xi_k = -1) = P(\xi_1 = -1, \xi_2 = -1) = \frac{1}{4}$ (itt felsoroltuk, hogy például a $\xi_1 = 2, \xi_2 = 2$ azt jelenti, hogy az ((F,F),(F,F)), ((F,F),(P,P)), ((P,P),(F,F)) vagy ((P,P),(P,P)) húzássorozatok valamelyike következik be. Ezért $E\xi_j\xi_k = P(\xi_j = -1, \xi_k = -1) + 4P(\xi_j = 2, \xi_k = 2) - 2P(\xi_j = -1, \xi_k = 2) - 2P(\xi_j = 2, \xi_k = -1) = \frac{1}{4}(1+4-4) = \frac{1}{4}$, 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 $ES = 8E\xi_1 = 2$, $\operatorname{Var} S = 8\operatorname{Var}\xi_1 = 18$.