A december 14.-i gyakorlat feladatai

Tárgyaljuk azt a kérdést, hogyan lehet két független valószínűségi változó összegének az eloszlását kiszámítani. Először felidézem a Poisson eloszlás definicióját, és tárgyaljuk azt a feladatot, hogy hogyan számoljuk ki két független Poisson eloszlású valószínűségi változó összegének az eloszlását. Ez tulajdonképpen egy példa arra, hogy általános esetben, hogyan tudunk ilyen feladatot megoldani. Ezután foglalkozunk ennek a feladatnak az analogjával, amikor két független sűrűségfüggvénnyel rendelkező valószínűségi változó összegének a (létező) sűrűségfüggvényét akarjuk kiszámolni. Ennek érdekében bevezetjük a konvolució fogalmát.

Poisson eloszlás definiciója. *Egy ξ valószínűségi változó λ paraméterű Poisson eloszlású, ha ξ nem negatív egész értékeket vesz fel, és*

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

1.) Lássuk be, hogy ez valóban valószínűség eloszlás.
Megoldás: Azt kell ellenőrizni, hogy $P(\xi = k) \geq 0$ minden k értékre, ami nyílvánvaló, és $\sum\limits_{k=0}^{\infty} P(\xi = k) = 1$. Viszont

$$\sum_{k=0}^{\infty} P(\xi = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

2.) Ha ξ és η két független, λ illetve μ paraméterű Poisson eloszlású valószínűségi változó, akkor $\xi + \eta$ egy $\lambda + \mu$ paraméterű Poisson eloszlású valószínűségi változó.
Megoldás:

$$\begin{aligned} P(\xi + \eta = k) &= \sum_{j=0}^{k} P(\xi = j, \eta = k - j) = \sum_{j=0}^{k} P(\xi = j) P(\eta = k - j) \\ &= \sum_{j=0}^{k} \frac{\lambda^j}{j!} e^{-\lambda} \frac{\mu^{k-j}}{(k-j)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} \lambda^j \mu^{(k-j)} \\ &= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k. \end{aligned}$$

Tárgyaljuk a következő feladatot. Legyen ξ és η két független valószínűségi változó, amelyeknek létezik $f(\cdot)$ és $g(\cdot)$ sűrűségfüggvényük. Be lehet látni, hogy a $\xi + \eta$ összegnek is létezik $h(\cdot)$ sűrűségfüggvénye, és meg lehet adni azt a formulát, amelynek segítségével ki lehet számolni a $\xi + \eta$ összeg sűrűségfüggvényét. Ennek érdekében bevezetjük két (sűrűség)függvény konvoluciójának a fogalmát.

(Sűrűség)függvények konvoluciójának a definiciója. *Legyen $f(\cdot)$ és $g(\cdot)$ két sűrűségfüggvény a számegyenesen, általánosabban integrálható függvények, azaz tegyük fel,* *hogy* $\int_{-\infty}^{\infty} |f(u)|\,du < \infty$ *és* $\int_{-\infty}^{\infty} |g(u)|\,du < \infty$. *Az* $f(\cdot)$ *és* $g(\cdot)$ *függvények* $f * g(\cdot)$ *konvoluciója az*

$$f * g(x) = \int_{-\infty}^{\infty} f(u)g(x-u)\,du, \quad -\infty < x < \infty,$$

függvény.

Tétel független valószínűségi változók összegének a sűrűségfüggvényéről. *Legyen* ξ *és* η *két független valószínűségi változó* $f(\cdot)$ *és* $g(\cdot)$ *sűrűségfüggvénnyel. Ekkor a* $\xi+\eta$ *összegnek is létezik sűrűségfüggvénye, és az az*

$$f * g(x) = \int_{-\infty}^{\infty} f(u)g(x-u)\,du, \quad -\infty < x < \infty$$

függvény.

3.) Legyen ξ és η két független, a $\left[-\frac{1}{2}, \frac{1}{2}\right]$ intervallumban egyenletes eloszlású valószínűségi változó, azaz legyen ξ és η sűrűségfüggvénye $f(x) = 1$, ha $-\frac{1}{2} \le x \le \frac{1}{2}$, és $f(x) = 0$ egyébként. Számoljuk ki $\xi + \eta$ sűrűségfüggvényét.

Megoldás: A $\xi + \eta$ valószínűségi változó sűrűségfüggvénye a

$$g(x) = \int_{-\infty}^{\infty} f(y)f(x-y)\,dy$$

függvény, ahol $f(x)$ a $\left[-\frac{1}{2}, \frac{1}{2}\right]$ intervallumban egyenletes eloszlás sűrűségfüggvénye. Ezért $f(y)f(x-y) = 1$, ha $-\frac{1}{2} \le y \le \frac{1}{2}$, és $-\frac{1}{2} \le x-y \le \frac{1}{2}$, azaz $-\frac{1}{2}+x \le y \le \frac{1}{2}+x$, és nulla egyébként. Ez azt jelenti, hogy a $\xi+\eta$ összeg $g(x)$ sűrűségfüggvénye az x pontban megegyezik a $\left[-\frac{1}{2}, \frac{1}{2}\right] \cap \left[-\frac{1}{2}+x, \frac{1}{2}+x\right]$ intervallum hosszával. Ha $|x| > 1$, akkor a fenti metszet üres, ezért ebben az esetben $g(x) = 0$. Ha $0 \le x \le 1$, akkor ez a metszet a $\left[-\frac{1}{2}+x, \frac{1}{2}\right]$ 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 $\left[-\frac{1}{2}, \frac{1}{2}+x\right]$ intervallum amelynek 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 $|x| > 1$.

4.) Legyenek ξ_1 és ξ_2 független exponenciális eloszlású valószínűségi változók, azaz legyen sűrűségfüggvényük $f(x) = \lambda e^{-\lambda x}$ ha $x \ge 0$, és $f(x) = 0$, ha $x < 0$. Számítsuk ki $\xi_1 + \xi_2$ sűrűségfüggvényét.

Általánosabban, legyenek $\xi_1, \dots \xi_m$ független exponenciális eloszlású valószínűségi változók $\lambda > 0$ paraméterrel. Számítsuk ki $\xi_1 + \cdots + \xi_m$ sűrűségfüggvényét.

Megoldás: Ki kell számolnunk az $f * f(x)$ illetve $\underbrace{f * \cdots * f}_{m-\text{szer}}(x)$ konvoluciókat a fenti $f(x)$ sűrűségfüggvénnyel. Mivel $f(x) = 0$, ha $x \le 0$, a konvoluciót meghatározó integrálban szereplő $f(y)f(x-y)$ integrandus nulla, ha $y \le 0$ vagy $x - y \le 0$. Innen a konvoluciót definiáló integrál csak $x \ge 0$ esetén lehet nulla, az $x \le 0$ esetben $f(y)f(x-y) > 0$ minden y-ra nulla, és $x \ge 0$ esetén az $f(y)f(x-y) > 0$

integrandus csak $0 \le y \le x$ esetén nem nulla. Innen a $\xi_1 + \xi_2$ valószínűségi változó sűrűségfüggvénye $f_2(x) = f * f(x)$ $x < 0$-ra $f_2(x) = 0$, és

$$f_2(x) = f * f(x) = \int_{-\infty}^{\infty} f(y) f(x-y)\, dy = \int_0^x \lambda e^{-\lambda y} \lambda e^{-\lambda(x-y)}\, dy$$
$$= \int_0^x \lambda^2 e^{-\lambda x}\, dy = \lambda^2 x e^{-\lambda x}, \quad \text{ha } x \ge 0.$$

Hasonlóan, ha $f_m(x) = \underbrace{f * \cdots * f}_{m-\text{szer}}(x)$ jelöli $\xi_1 + \cdots \xi_m$ sűrűségfüggvényét, akkor $f_m(x) = 0$ minden $m \ge 1$ számra, ha $x < 0$. Azt állítjuk, hogy $f_m(x) = \frac{\lambda^m x^{m-1}}{(m-1)!} e^{-\lambda x}$, ha $x \ge 0$. Ezen állítás bizonyításához elég belátni teljes indukcióval azt, hogy $f_{m-1} * f(x) = f_m(x)$ a fent definiált f_m függvényekkel. Viszont

$$f_{m-1} * f(x) = \int_{-\infty}^{\infty} f_{m-1}(y) f(x-y)\, dy = \int_0^x \lambda^{m-1} \frac{y^{m-2}}{(m-2)!} \lambda e^{-\lambda y} e^{-\lambda(x-y)}\, dy$$
$$= \lambda^m e^{-\lambda x} \int_0^x \frac{y^{m-2}}{(m-2)!}\, dy = e^{-\lambda x} \frac{\lambda^m x^{m-1}}{(m-1)!}, \quad \text{ha } x \ge 0.$$

Másrészt $f_m(x) = 0$, ha $x \le 0$.

5.) Legyen ξ és η két független valószínűségi változó, mind a kettő $f(x) = \frac{1}{2} e^{-|x|}$, $-\infty < x < \infty$, sűrűségfüggvénnyel. Lássuk be először, hogy $f(x)$ valóban sűrűségfüggvény. Számítsuk ki a $\xi + \eta$ valószínűségi változó $g(x)$ sűrűségfüggvényét.
Megoldás: Az $f(x)$ függvény minden pontban nem negatív. Annak ellenőrzéséhez, hogy $f(x)$ sűrűségfüggvény azt kell megmutatnunk, hogy $\int_{-\infty}^{\infty} f(x)\, dx = 1$. Viszont hogy $\int_{-\infty}^{\infty} f(x)\, dx = \frac{1}{2} \int_{-\infty}^{0} e^x\, dx + \frac{1}{2} \int_0^{\infty} e^{-x}\, dx = \frac{1}{2} [e^x]_{-\infty}^0 + \frac{1}{2} [-e^{-x}]_0^{\infty} = 1$.
A $\xi + \eta$ valószínűségi változó $g(x)$ sűrűségfüggvényét a $g(x) = \int_{-\infty}^{\infty} f(y) f(x-y)\, dy$ formula segítségével számíthatjuk ki. Számítsuk ki ezt az integrált. Tekintsük először azt az esetet, amikor $x \ge 0$. Az integrált számítsuk ki úgy, hogy nézzük mind a négy (elvileg) lehetséges esetet, amikor a) $y \ge 0$ és $x - y \ge 0$, b) $y \ge 0$ és $x - y < 0$, c) $y < 0$, $x - y \ge 0$, d) $y < 0$, $x - y < 0$. Számítsuk ki mind a négy esetben azt, hogy milyen tartományban veszi fel értékét az y változó, és mi az integrandus illetve az integrál értéke ebben a tartományban. Az a) esetben $0 \le y \le x$, az integrandus $f(y) f(x-y) = \frac{1}{4} e^{-y} e^{-(x-y)} = \frac{e^{-x}}{4}$, az integrál pedig $\frac{x e^{-x}}{4}$ az a) tartományban. A b) esetben $y > x$ és $f(y) f(x-y) = \frac{e^{-y} e^{x-y}}{4} = \frac{e^{x-2y}}{4}$ az integrál pedig $\frac{1}{4} \int_x^{\infty} e^{x-2y}\, dy = \frac{e^{-x}}{8}$, a c) esetben $y < 0$ és $f(y) f(x-y) = \frac{1}{4} e^y e^{-(x-y)} = \frac{e^{2y-x}}{4}$, az integrál pedig $\frac{1}{4} \int_{-\infty}^{0} e^{2y-x}\, dy = \frac{e^{-x}}{8}$, a d) eset nem lehetséges, mert ekkor egyrészt az $y < 0$ másrészt az $y > x \ge 0$ feltételeknek kellene teljesülniük. Innen azt kapjuk, hogy $g(x) = \frac{(x+1)e^{-x}}{4}$, ha $x > 0$. Mivel f szimmetrikus függvény, ezért mint nem nehéz megmutatni, $f(x)$ is az. Tehát $g(-x) = g(x)$, és $g(x) = \frac{(|x|+1)e^{-|x|}}{4}$.

6.) Egy pénzdarabról ellenőrizni akarjuk, hogy igaz-e az a hipotézis, amely szerint ez az érme legalább $\frac{3}{4}$ valószínűséggel esik a fej és legfeljebb $\frac{1}{4}$ valószínűséggel az írás oldalára. Ennek érdekében feldobjuk a pénzdarabot 30 000 alkalommal, és a következő döntési szabályt hozzuk. Választunk egy k számot, és akkor fogadjuk el a hipotézist helyesnek, ha legalább k fejdobás történt. Legalább mekkorának kell választanunk ezt a k számot, ha azt akarjuk, hogy egy a hipotézist teljesítő pénzdarab esetén legalább 0.9 valószínűséggel döntsünk úgy, hogy a hipotézis teljesül?

Megoldás: Vezessük be a következő valószínűségi változókat: $\xi_j = 1$, ha a j-ik dobás eredménye fej, $\xi_j = 0$, ha a j-ik dobás eredménye írás, $1 \le j \le 30\,000$, $S = S_{30000} = \sum_{j=1}^{30\,000} \xi_j$. Ha a fejdobás eredményének valószínűsége pontosan $\frac{3}{4}$, $E\xi_j = \frac{3}{4}$, $E\xi_j^2 = \frac{3}{4}$, $\operatorname{Var}\xi_j = E\xi_j^2 - (E\xi_j)^2 = \frac{3}{16}$, $ES = 30\,000 E\xi_j = 22\,500$, $\operatorname{Var} S = 30\,000 \operatorname{Var}\xi_j = 5625 = 75^2$. Innen és a centrális határeloszlástételből,

$$
\begin{aligned}
P(S > k) &= P\left(\frac{S-ES}{\sqrt{\operatorname{Var} S}} > \frac{k-22\,500}{75}\right) = 1 - P\left(\frac{S-ES}{\sqrt{\operatorname{Var} S}} \le \frac{k-22\,500}{75}\right) \\
&\sim 1 - \Phi\left(\frac{k-22\,500}{75}\right).
\end{aligned}
$$

Válasszuk a k számot úgy, hogy a fenti valószínűség körülbelül 0.9 legyen. Ekkor a $\Phi\left(\frac{k-22\,500}{75}\right) = 0.1$ vagy ami ezzel ekvivalens, a $\Phi\left(\frac{22\,500-k}{75}\right) = 0.9$ egyenletet kell kielégítenünk. A normális eloszlás-táblázat alapján $\frac{22\,500-k}{75} \sim 1.28$, ami azt jelenti, hogy $k = 22\,500 - 75 \times 1.28$ és $p = \frac{3}{4}$ esetén annak valószínűsége, hogy a fejdobások száma nagyobb mint $k = 22\,500 - 75 \times 1.28 = 22\,212$ és $p = \frac{3}{4}$ esetében annak valószínűsége, hogy legalább ennyi fejdobás történik körülbelül 0.9. Ha $p \ge \frac{3}{4}$, akkor ez a valószínűség nagyobb. Ezért a $k = 22\,212$ helyes választás.