A május 2-i gyakorlat témája
Rövid összefoglaló

Először beszékjük meg röviden, miért konvolució segítségével lehet kiszámítani két ismert sűrűségfüggvénnyel rendelkező valószínűségi változó összegének sűrűségfüggvényét. Ha ξ és η két valószínűségi változó, melyek (együttes) sűrűségfüggvénye a $h(u,v)$ sűrűségfüggvény, akkor annak valószínűsége, hogy a (xi,η) véletlen vektor a sík valamely $A\subset R^2$ részhalmazába esik

$$P((\xi,\eta)\in A)=\int\int_{\{(u,v)\in A\}} h(u,v)\,du\,dv.$$

Speciálisan,

$$P(\xi+\eta<x)=\int\int_{\{(u,v)\colon\ u+v<x\}} h(u,v)\,du\,dv,$$

és ez a képlet adja meg a $\xi+\eta$ valószínűségi változó $F(x)=P(\xi+\eta<x)$ eloszlásfüggvényét. Ez az eredmény akkor is érvényes, ha ξ és η független valószínűségi változók $f(u)$ és $g(v)$ sűrűségfüggvénnyel. Azt is érdemes tudni, hogy ebben az esetben a (ξ,η) véletlen vektor sűrűségfüggvénye $h(u,v)=f(u)g(v)$ alakú. Ekkor $\xi+\eta$ sűrűségfüggvénye $f(x)=\dfrac{d}{dx}P(\xi+\eta<x)$, tehát

$$\begin{aligned}f(x)=\frac{d}{dx}P(\xi+\eta<x)&=\frac{d}{dx}\int\int_{\{(u,v)\colon\ u+v<x\}} f(u)g(v)\,du\,dv\\&=\frac{d}{dx}\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} f(u)g(v-u)\,du\right]dv=\int_{-\infty}^{\infty} f(u)g(x-u)\,du\end{aligned}$$

A fenti számolások első azonosságában egy integráltranszformációt alkalmaztunk $\bar v=u+v$, $\bar u=u$ helyettesítéssel (de $\bar u$ és $\bar v$ helyett u és v változót írtunk.) Felhasználtuk, hogy e transzformáció során az $\{(u,v)\colon u+v<x\}$ tartomány a $\{(\bar u,\bar v)\colon \bar v<x, -\infty<\bar u<\infty\}$ tartományba megy át, és a fenti (lineáris) transzformáció Jakobiánja azonosan 1. Az utolsó azonosságban a $\dfrac{d}{dx}\int_{-\infty}^{\infty} H(v)\,dv=H(x)$ azonosságot alkalmaztuk $H(v)=\int_{-\infty}^{\infty} f(u)g(v-u)\,du$ választással.

Ki tudjuk számolni tetszőleges független *egész értékű* valószínűségi változók összegének az eloszlását az adott valószínűségi változók eloszlásának a segítségével. E számolás megfogalmazásánának érdekében vezették be diszkrét eloszlások konvoluciójának a fogalmát.

Diszkrét eloszlások konvoluciójának a definiciója. *Legyen $\mathcal{P}=p(k)$ és $\mathcal{Q}=q(k)$, $k=0,\pm1,\pm2,\dots$, két eloszlás az egész számokon, azaz teljesüljenek a $p(k)\ge0$, $q(k)\ge0$, $k=0,\pm1,\pm2,\dots$, $\sum\limits_{k=-\infty}^{\infty}p(k)=1$, $\sum\limits_{k=-\infty}^{\infty}q(k)=1$ feltételek. Ekkor e két eloszlás konvoluciója a*

$$\mathcal{P}*\mathcal{Q}(k)=\sum_{j=-\infty}^{\infty}p(j)q(k-j),\qquad k=0,\pm1,\pm2,\dots$$

eloszlás.

Tétel. *Ha ξ és η két diszkrét egész értékeket felvevő független valószínűségi változó $\mathcal{P}$ és $\mathcal{Q}$ eloszlással akkor $\xi+\eta$ eloszlása a $\mathcal{P}*\mathcal{Q}$ eloszlás.*

(Tulajdonképpen a fenti eredményt már alkalmaztuk (a formulák közvetlen bizonyításával) a március 21-i gyakorlat feladatainak megbeszélésénél.)

Vizsgáljuk a következő problémakört. Egy szabályos pénzdarabot nagyon sokszor feldobunk egymás után, mondjuk 10 000 alkalommal. Azt várjuk, hogy nagy valószínűséggel körülbelül 5 000 fej és 5 000 írás dobás jelenik meg. De mit jelent ebben az állításban az, hogy körülbelül? Képzeljük el, hogy valaki egy játékot ajánl. Egy szerinte szabályos pénzdarabot egymás után 10 000 alkalommal feldob egymás után, ha fej az eredmény, akkor én fizetek neki 1 forintot, ha írás akkor ő fizet nekem 1 forintot. A játék során 10 000 fejdobás eredmény születik. Elhisszük-e ennek az embernek azt, hogy szabályos pénzdarabot dobott fel? Elhisszük-e akkor, ha 6 000 fej és 4 000 írásdobás jelenik meg? Akkor, ha 5 500 fej és 4 500 írás vagy ha 5 200 fej és 4 800 írásdobás jelenik meg? Az általános kérdés az, hogy igaz-e, hogy a fejdobások relatív gyakorisága közel van a dobások valószínűségéhez? Ha igen, akkor milyen közel? Mekkora eltérés tekinthető tipikusnak és mekkora eltérés atipikus, azaz rendkívül valószínűtlen? Természetesen hasonló kérdések vethetők fel más problémák vizsgálatában is. Például egy szabályos kocka sok egymás utáni független dobása esetén mi a valószínűsége van annak, hogy a hatos dobások relatív gyakorisága bizonyos pontossággal megközelíti az $\frac{1}{6}$ számot, azaz a hatos dobás valószínűségét? Megbeszéljük azt, hogy milyen becslést ad erre a valószínűségre a Chebishev egyenlőtlenség és a centrális (központi) határeloszlástétel, a valószínűségszámítás talán legfontosabb eredménye.

Chebishev egyenlótlenség. *Tekintsünk egy S valószínűségi változót, melyre $ES=M$, $\operatorname{Var} S=D^2$ valamilyen $D<\infty$ számmal. Ekkor minden $0\le x<\infty$ számra*

$$P(|S-M|>x)\le\frac{\operatorname{Var} S}{x^2}.$$

Speciálisan, legyenek ξ_j, $1\le j\le n$, független valószínűségi változók $E\xi_j=m_j$, $\operatorname{Var}\xi_j=\sigma_j^2$, $1\le j\le n$, $S=\sum\limits_{j=1}^n\xi_j$. Ekkor $ES=\sum\limits_{j=1}^n m_j$, $\operatorname{Var} S=\sum\limits_{j=1}^n\sigma_j^2$, ezért

$$P\left(\left|S-\sum_{j=1}^n m_j\right|>x\right)\le\frac{\sum\limits_{j=1}^n\sigma_j^2}{x^2}.$$

Centrális határeloszlástétel. *Legyenek ξ_j, $j=1,2,\dots$ független egyforma eloszlású valószínűségi változók sorozata, $E\xi_1=m$, $\operatorname{Var}\xi_1=\sigma^2$, $S_n=\sum\limits_{j=1}^n\xi_j$, $n=1,2,\dots$. Ekkor*

$$\lim_{n\to\infty}P\left(\frac{S_n-nm}{\sqrt{n}\sigma}<x\right)=\Phi(x)=\int_{-\infty}^x\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du.$$

minden $-\infty < x < \infty$ *számra.*

Megjegyezzük, hogy a fenti tétel általánosítható. Hasonló tétel mondható ki független nem feltétlenül egyforma eloszlású valószínűségi változók normalizált összegeinek eloszlásai (azaz az összegek olyan lineáris transzformáltjának eloszlásai, melynek a várható értéke nulla és szórásnégyzete egy) enyhe feltételek mellett konvergálnak a standard normális eloszláshoz. Ennek a problémának a részleteit azonban itt nem tárgyaljuk.

Vizsgáljuk a következő feladat megoldását:

1. *Tekintsünk egy szabályos pénzdarab 10 000 egymás utáni (független) feldobásából származó fej-írás sorozatot. Adjunk becslést a Chebishev egyenlőtlenség segítségével annak valószínűségére, hogy a fej-dobások számának eltérése a várt 5000 számtól legalább 100-zal, illetve legalább 200-zal eltér! Milyen becslést ad ezekre a valószínűségekre a centrális határeloszlástétel?*

Megoldás: Vezessük be a következő ξ_j, $\le j \le 10\,000$ valószínűségi változókat, melyekre $\xi_j = 1$, ha a j-ik dobás eredménye fej, $\xi_j = 0$, ha a j-ik dobás eredménye írás. Ekkor ξ_j, $1 \le j \le 10\,000$ független valószínűségi változők, $E\xi_j = \frac{1}{2}$, $\operatorname{Var}\xi_j = \frac{1}{4}$, és a $P\left(\left|\sum\limits_{j=1}^{10000}(\xi_j - E\xi_j)\right| > 100\right)$ és $P\left(\left|\sum\limits_{j=1}^{10000}(\xi_j - E\xi_j)\right| > 200\right)$ valószínűségekre kell becslét adnunk. A Chebishev egyenlőtlenség az első valószínűségre a $\dfrac{10000 \cdot \frac{1}{4}}{100^2} = \dfrac{1}{4}$, a második valószínűségre pedig a $\dfrac{10000 \cdot \frac{1}{4}}{200^2} = \dfrac{1}{16}$ felső becslést adja.

A centrális határeloszlástétel szerint $P\left(\dfrac{\sum\limits_{j=1}^{10000}(\xi_j - E\xi_j)}{\sqrt{10000 \cdot \frac{1}{4}}} > u\right) \sim 1 - \Phi(u)$. Innen kapjuk, hogy az első valószínűség kiszámításához az $u = \pm\frac{100}{\frac{1}{2} \cdot 100} = \pm 2$ értékeket kell tekinteni, és a vizsgált valószínűség közelítőleg $(1 - \Phi(2)) + \Phi(-2) = 2(1 - \Phi(2)) \sim 2(1 - 0.97720) = 0.0456$. A második valószínűség hasonlóan körülbelül $2(1 - \Phi(4)) \sim 0$, (az első 4 tizedesjegy 0).

Házi feladat.

Egy szabályos dobókockát feldobunk 2700 alkalommal egymástól függetlenül, és összeszámoljuk a páros értékű dobások eredményét. Adjunk jó közelítő becslést a centrális határeloszlástétel és egy normális eloszlástáblázat segítségével arra, hogy ez az összeg 420 és 720 közé esik.