A Valószínűségszámítás I. előadássorozat tizedik előadása.

2001 április 3.

Először néhány feladat megoldását tárgyaljuk.

Idézzük fel, hogy ha ξ standard normális eloszlású valószínűségi változó, akkor $\sigma\xi + m$ m várható értékű és σ^2 szórásnégyztű normális eloszlású valószínűségi változó. Egy ξ valószínűségi változó akkor normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, ha sűrűségfüggvénye a $g(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}$ függvény.

Feladat:

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 $\eta_1 + \eta_2$ $m_1 + m_2$ várható értékú 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 $y = \sqrt{2}\dfrac{x-A}{\sqrt{B}}$ helyettesítéssel, és abból a tényből, hogy a $\varphi(y) = \dfrac{1}{\sqrt{2\pi}} e^{-y^2/2}$ függvény sűrűségfüggvény. Ez az észrevétel azért hasznos, mert ez lehetővé teszi, hogy amennyiben olyan integrált kell kiszámolni, melyben 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

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

Jegyezzük meg, hogy a fenti számolásban megjelenő egyszerűsítések megjelenésének mélyebb oka van. Ha tudjuk, hogy a vizsgált integrál értéke olyan alakú, melyben konstansszor egy kvadratikus alak szerepel, akkor mivel a konvolució eredménye egy sűrűségfüggvény, ezért a végeredmény szükségszerűen normális sűrűségfüggvény. Továbbá, tudjuk, hogy a sűrűségfüggvény egy olyan valószínűségi változó sűrűségfüggvénye, amelyik várható értéke $m_1 + m_2$ szórásnégyzete pedig $\sigma_1^2 + \sigma_2^2$.

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:

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

Legyenek ξ és η független, standard normális eloszlású valószínűségi változók. Lássuk be, hogy $\xi^2 + \eta^2$ exponenciális eloszlású valószínűségi változó.
Megoldás: $P(\xi^2 < x) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1$, ha $x \geq 0$. Írjuk fel ξ^2 sűrűségfüggvényét és konvolúció 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 $x \geq 0$, és $g(x) = 0$, ha $x < 0$, és $\xi^2 + \eta^2$ sűrűségfüggvénye

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

és $f(x) = 0$, ha $x \leq 0$.

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. Hogyan? Vagyük észre, hogy

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

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

Megoldás.

Tegyünk először egy általános megjegyzést. Ha adva van két valószínűségi változó ξ és η, melyek (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, mely 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.

A feladatban vizsgálandó hányados eloszlásfüggvénye

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

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

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

Legyen ξ egyenletes eloszlású valószínűségi változó a $[0,1]$ intervallumban, azaz legyen $f(\cdot)$ sűrűségfüggvénye $f(x) = 1$, ha $0 \le x \le 1$, $f(x) = 0$ egyébként. Számítsuk ki a ξ^2 valószínűségi változó várható értékét és szórásnégyzetét.

Első megoldás: $\operatorname{Var}\xi^2 = E\left(\left(\xi^2\right)^2\right) - \left(E\xi^2\right)^2 = E\xi^4 - \left(E\xi^2\right)^2 = \int x^4 f(x)\,dx - \left(\int x^2 f(x)\,dx\right)^2$, ahol $f(x) = 1$, ha $0 \le x \le 1$, $f(x) = 0$ egyébként, azaz $f(x)$ a ξ

valószínűségi változó sűrűségfüggvénye. Innen $E\xi^2 = \int_0^1 x^2\,dx = \dfrac{1}{3}$, és $\operatorname{Var}\xi^2 = \int_0^1 x^4\,dx - \left(\int_0^1 x^2\,dx\right)^2 = \dfrac{1}{5} - \left(\dfrac{1}{3}\right)^2 = \dfrac{1}{5} - \dfrac{1}{9} = \dfrac{4}{45}$.

Második megoldás: Számítsuk ki először a ξ^2 valószínűségi változó eloszlás és sűrűségfüggvényét. A ξ^2 valószínűségi változó $F(\cdot)$ eloszlásfüggvénye

$$F(x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = P(0 \le \xi \le \sqrt{x}) = \sqrt{x}, \quad \text{ha } 0 \le x \le 1,$$

$F(x) = 0$, ha $\xi < 0$, $F(x) = 1$, ha $x > 1$. Innen a ξ^2 valószínűségi változó $f(\cdot)$ sűrűségfüggvénye $f(x) = \dfrac{1}{2\sqrt{x}}$, ha $0 \le x \le 1$, és $f(x) = 0$ egyébként. Ezért

$$E\xi^2 = \int_0^1 x\frac{1}{2\sqrt{x}}\,dx = \frac{1}{2}\cdot\frac{2}{3} = \frac{1}{3},$$

és

$$\operatorname{Var}\xi^2 = \int_0^1 x^2\frac{1}{2\sqrt{x}}\,dx - \left(\int_0^1 x\frac{1}{2\sqrt{x}}\,dx\right)^2 = \frac{1}{2}\cdot\frac{2}{5} - \left(\frac{1}{2}\cdot\frac{2}{3}\right)^2 = \frac{4}{45}.$$

Tegyük fel, hogy a következő játékot játszhatjuk: Feldobnak egy szabályos pénzdarabot egymástól függetlenül egymás után. Amennyiben fej a dobás eredménye, akkor a feltett tét dupláját kapjuk, ha írás, akkor a tét $\frac{2}{3}$ részét elveszítjük, és csak $\frac{1}{3}$ részét őrizhetjük meg. Mivel ez a játék előnyös, ezért feltesszük minden játékban minden pénzünket. Lássuk be, hogy amennyiben A volt a vagyonunk a játék kezdete előtt, és Z_n jelöli vagyonunkat az n-ik játék után, akkor

a) $EZ_n = A\left(\dfrac{7}{6}\right)^n$, azaz vagyonunk várható értéke exponenciálisan nő.

b) Z_n sztochasztikusan tart nullához, azaz ha sokáig játszunk akkor közel egy valószínűséggel majdnem minden pénzünket elveszítjük. (Valójában Z_n egy valószínűséggel is tart nullához, csak ennek indoklásához szükséges az előadáson eddig nem tárgyalt nagy számok erős törvénye is.)

c) Értsük meg, hogy ez a két állítás nem mond egymásnak ellent.

Megoldás: Vezessük be a következő ξ_j valószínűségi változókat: $\xi_j = 2$, ha a j-ik dobás eredménye fej, $\xi_j = \dfrac{1}{3}$, ha a j-ik dobás eredménye írás. Ekkor $\xi_1, \xi_2, \dots$, független valószínűségi változók, $P(\xi_j = 2) = P\left(\xi_j = \dfrac{1}{3}\right) = \dfrac{1}{2}$, és ezenkívül $Z_n = A\xi_1\xi_2\cdots\xi_n$. Ezért $E\xi_j = \dfrac{1}{2}\left(2 + \dfrac{1}{3}\right) = \dfrac{7}{6}$, és $EZ_n = EA\xi_1\xi_2\cdots\xi_n = AE\xi_1 E\xi_2\cdots E\xi_n = A\left(\dfrac{7}{6}\right)^n$. Ez a feladat a) állítása.

A $Z_n = A\xi_1\xi_2\cdots\xi_n$ relációból következik, hogy $\dfrac{1}{n}\log Z_n = \dfrac{\log A}{n} + \dfrac{1}{n}\sum\limits_{j=1}^{n}\log\xi_j$. Továbbá, $E\log\xi_j = \dfrac{1}{2}\left(\log 2 + \log\dfrac{1}{3}\right) = -\dfrac{1}{2}\log\dfrac{3}{2}$. Ezért a nagy számok (gyenge) törvénye szerint $\dfrac{1}{n}\log Z_n$ sztochasztikusan konvergál a negatív $-\dfrac{1}{2}\log\dfrac{3}{2}$ számhoz. Innen tetszőleges $\varepsilon > 0$ számra $P(Z_n > \varepsilon) = P\left(\dfrac{1}{n}\log Z_n > \dfrac{\log\varepsilon}{n}\right)$, és ez a valószínűség nullához tart, ha $n\to\infty$. Valóban, a nagy számok gyenge törvénye, és a $-\dfrac{1}{10} > -\dfrac{1}{2}\log\dfrac{3}{2}$ egyenlőtlenség miatt $\lim\limits_{n\to\infty} P\left(\dfrac{1}{n}\log Z_n < -\dfrac{1}{10}\right) = 1$. Mivel $\lim\limits_{n\to\infty}\dfrac{\log\varepsilon}{n} = 0$ innen következik a feladat b) állítása.

Az a) rész bizonyítása azon alapult, hogy $E\xi_j > 1$, a b) részé pedig azon, hogy $E\log\xi_j < 0$. Ez a két egyenlőtlenség teljesülhet egyszerre, mert a várható érték és a logaritmus egymással nem felcserélhető. Igaz az $Ee^{\eta} \ge e^{E\eta}$ egyenlőtlenség (ez a konvex függvényekre vonatkozó úgynevezett Jensen egyenlőtlenség speciális esete), ahonnan $\xi = \log\eta$ választással $E\xi \ge e^{\log E\xi}$, de egyenlőség nem írható a fenti egyenlőtlenség helyett. Jegyezzük meg, hogy hasonló, de egyszerűbben érthető példát mutat a feladat a) és b) állításának egyszerre való teljesülésére a következő modell. Olyan játékot játszunk, melyben $\frac{1}{2}$ valószínűséggel elveszítjük, $\frac{1}{2}$ valószínűséggel pedig megháromszorozzuk a pénzünket. Az egyes játékok egymástól függetlenek, és minden időpontban minden pénzünket feltesszük. Ekkor annak a valószínűsége, hogy az n-ik játék után minden pénzünket elveszítjük, $1 - \left(\frac{1}{2}\right)^n$, ami rendkívül gyorsan tart egyhez, és pénzünk várható értéke $3^n\left(\frac{1}{2}\right)^n$, ami exponenciálisan gyorsan nő az n függvényében. Hasonló, csak kissé rejtettebb eset történik az általunk tárgyalt feladatban is. Tekintsük a feladatban tekintett játék nyereményét sok játék után. Azt állíthatjuk, hogy nagy n indexre az n-ik játék után nagy valószínűséggel alig marad pénzünk. Viszont kis valószínűséggel nagyon sok pénzt nyerünk, és ezért nyereményünk várható értéke nagy. Ez az oka annak, hogy nemcsak a b), hanem az a) állítás is teljesül.

Tekintsük az előző feladatban tekintett játékot azzal a különbséggel, hogy a játék minden egyes fordulójában vagyonunk u-ad részét, $0 \le u \le 1$, tesszük fel tétként. Jelölje $Z_n(u)$ vagyonunkat a játék n-ik lépése után. Ekkor az $\dfrac{1}{n}\log Z_n(u)$ valószínűségi változók sztochasztikusan konvergálnak egy $B(u)$ számhoz. Határozzuk meg a legjobb $\bar u$ számot, amelyikre $B(\bar u) = \sup\limits_{0\le u\le 1} B(u)$. Lássuk be, hogy $B(\bar u) > 0$.

Megoldás: Vezessük be a $\xi_j = \xi_j(u)$, $j = 1, 2, \dots$, valószínűségi változókat, melyekre $\xi_j = 1 + u$, ha a j-ik dobás eredménye fej, és $\xi_j = \xi_j(u) = 1 - \dfrac{2u}{3}$, ha a j-ik dobás eredménye írás. Ekkor az n-ik lépésben a vagyonunk $Z_n = A\xi_1\cdots\xi_n$, a ξ_j, $j = 1, 2, \dots$, valószínűségi változók függetlenek és egyforma eloszlásúak, ezért $\dfrac{1}{n}\log Z_n = \dfrac{\log A}{n} + \dfrac{1}{n}\sum\limits_{j=1}^{n}\log\xi_j$, és a nagy számok gyenge törvénye szerint az

$\frac{1}{n}\log Z_n$ valószínűségi változók sztochasztikusan konvergálnak a $B(u) = E\log\xi_1 = \frac{1}{2}\left(\log(1+u) + \log\left(1-\frac{2u}{3}\right)\right) = \frac{1}{2}\log(1+u)\left(1-\frac{2u}{3}\right)$ számhoz. A $B(u)$ függvény a maximumát az $\bar u = \frac{1}{4}$ helyen veszi fel, és $B(\bar u) = \frac{1}{2}\log\frac{25}{24} > 1$.

A centrális határeloszlástétel.

Ez a valószínűségszámítás legfontosabb eredménye. Azt mondja ki, hogy ha sok független valószínűségi változót összegezünk, az összeget alkalmasan normalizáljuk, akkor nagyon általános feltételek teljesülése esetén az összeg közelítőleg normális eloszlású. A tétel megfogalmazása előtt vezessük be az irodalomban gyakran használt fogalmat.

Egy valószínűségi változó normalizáltjának a fogalma. *Legyen ξ olyan valószínűségi változó, melyre $E\xi^2 < \infty$. A ξ valószínűségi változó normalizáltja a $\bar\xi = \frac{\xi - E\xi}{\sqrt{\operatorname{Var}\xi}}$ valószínűségi változó, azaz a ξ valószínűségi változónak az a lineáris transzformáltja, melynek várható értéke nulla, szórásnégyzete pedig 1.*

Jegyezzük meg, hogy ha $\xi_1,\dots,\xi_n$ független valószínűségi változók, melyekre $E\xi_j^2 < \infty$ minden $1 \le j \le n$ indexre, akkor az $S = \sum\limits_{j=1}^n \xi_j$ összeg normalizáltja az

$$\bar S = \frac{\sum\limits_{j=1}^n \xi_j - \sum\limits_{j=1}^n E\xi_j}{\sqrt{\sum\limits_{j=1}^n \operatorname{Var}\xi_j}}$$

kifejezés.

Centrális határeloszlástétel. *Legyenek $\xi_1,\xi_2,\dots,$ független, egyforma eloszlású valószínűségi változók, jelölje $S_n = \sum\limits_{j=1}^n \xi_j$, $n = 1,2,\dots,$ e valószínűségi változók részletösszegeit. Ekkor az S_n valószínűségi változók $\frac{S_n - ES_n}{\sqrt{\operatorname{Var}S_n}} = \frac{\sum\limits_{j=1}^n \xi_j - \sum\limits_{j=1}^n E\xi_j}{\sqrt{\sum\limits_{j=1}^n \operatorname{Var}\xi_j}}$ normalizáltjai teljesítik a*

$$\lim_{n\to\infty} P\left(\frac{S_n - ES_n}{\sqrt{\operatorname{Var}S_n}} < x\right) = \Phi(x), \quad \textit{minden } -\infty < x < \infty \textit{ számra}$$

relációt, ahol $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du$, a standard normális eloszásfüggvény.

Megjegyzés 1: Az előbb kimondott tétel valójában speciális esete az általános centrális határeloszlástételnek, mely hasonló eredményt mond ki független, de nem feltétlenül azonos eloszlású valószínűségi változók normalizált részletösszegeire nagyon általános feltételek mellett.

Megjegyzés 2: Láttuk, hogy független normális eloszlású valószínűségi változók részletösszegeinek normalizáltja standard normális eloszlású valószínűségi változó. A centrális határeloszlástétel azt állítja, hogy sok független, de nem feltétlenül normális eloszlású valószínűségi változó összegének a normalizáltja nemcsak nulla várható értékű és egy szórásnégyzetű valószínűségi változó, hanem ezenkívül közelítőleg normális eloszlású is. Ez azt jelenti, hogy sok összeadandó esetén az összeg eloszlása ,,elfelejti" az összeadandók eloszlását, az ő eloszlása közelítőleg normális eloszlású. Bár ezt az eredményt be lehet bizonyítani, az mégis rendkívül meglepő. A centrális határeloszlást tekinthetjük a valószínűségszámítás legnagyobb misztériumának.

Megjegyzés 3: A műszaki tudományokban beszélnek egy olyan megfigyelésről, melyet hibatörvénynek neveznek. Gyakran elófordul, hogy egy műszaki feladatban egy kisérletet többször egymástól függetlenül elvégeznek, azok eredményeitt többször megmérik, de az egyes mérések pontatlanok. A tapasztalat azt mutatja, hogy szinte mindig a mért eredmények diagramja egy az igazi érték körüli tipikus úgynevezett harang-görbét rajzol ki. A meglepő tény az, hogy a legkülönbözőbb feladatokban mindig ugyanaz a görbe rajzolódik ki. Ennek a ténynek a hátterében valójában a centrális határeloszlástétel van. Az eredmény oka az, hogy a hibák sok kis apró egymástól független hiba összegeként keletkezik. Ezért a centrális határeloszlástétel szerint a hiba normális eloszlású, és a ,,hibatörvényben" megjelenő haranggörbe valójáan a normális sűrűségfüggvény.

Megjegyzés 4: Ha a $\xi_1, \xi_2, \dots$ független valószínűségi változók normalizált részletösszegei teljesítik a centrális határeloszlálstételt, akkor minden $-\infty < x < y < \infty$ számra teljesül a

$$\lim_{n\to\infty} P\left(x < \frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} < y\right) = \Phi(y) - \Phi(x)$$

azonosság, mert

$$P\left(x < \frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} < y\right) = P\left(\frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} < y\right) - P\left(\frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} \le x\right).$$

Jegyezzük meg továbbá, hogy $\Phi(-y) = 1 - \Phi(y)$ minden $-\infty < y < \infty$ számra, mivel a $\Phi(\cdot)$ eloszlásfüggvény $\varphi(\cdot)$ sűrűségfüggvénye páros függvény. Valóban, $\Phi(-y) = \int_{-\infty}^{-y} \varphi(u)\,du = 1 - \int_{-y}^{\infty} \varphi(u)\,du = 1 - \int_{-\infty}^{y} \varphi(-u)\,du = 1 - \int_{-\infty}^{y} \varphi(u)\,du = 1 - \Phi(y)$. Ezen eredmény miatt elegendő megadni a normális eloszlásfüggvény értékeit pozitív argumentumokra.

Megjegyzés 5: Természetesen felvetődik a kérdés, hogy létezik-e a centrális határeloszlástételnek több-dimenziós változata, melyben független több-dimenziós véletlen vektorok összegeinek alkalmas normalizáltjának eloszlásai teljesítenek valamilyen határeloszlástételt nagyon általános feltételek mellett. Ezenkívül érdekel minket a lehetséges határeloszlások konkrét alakja is. Ezt a kérdést, melynek fontos jelentősége van mind a valószínűségszámításban mind a matematikai statisztikában megválaszolták. Ezek a vizsgálatok vezettek a több-dimenziós normális eloszlások bevezetéséhez. A centrális határeloszlástétel több-dimenziós általánosításával később még foglalkozunk.

Tekintsük a következő két példát, melyek némi információt nyújtanak a centrális határeloszlástétel gyakorlati következményeiről. A következő előadáson a centrális határeloszlástétel további alkalmazásaira mutatunk példát.

Egy államban, (nevezzük az egyszerűség kedvéért Floridának,) 5 000 000 választó választ két párt (hívjuk ezeket mondjuk republikánus és demokrata pártnak) jelöltje között. Tegyük fel, hogy a választók egymástól függetlenül $\frac{1}{2}$ valószínűséggel választják valamelyik párt jelöltjét. Mi annak a valószínűsége, hogy a két jelölt által összegyüjtött szavazatok különbsége nem haladja meg a háromszázat.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 5\ 000\ 000$, valószínűségi változókat: $\xi_j = 1$, ha a j-ik választó a demokrata, $\xi_j = 0$, ha a j-ik választó a republikánus jelöltre szavaz. Ekkor $S = \sum\limits_{j=1}^{5\ 000\ 000} \xi_j$ a demokrata, és $5\ 000\ 000 - S$ a republikánus jelöltre leadott szavazatok száma, és minket a $P(|2S - 5\ 000\ 000| \le 300)$ valószínűség nagysága érdekel. Vegyük észre, hogy a ξ_j valószínűségi változók függetlenek, $E\xi_j = \dfrac{1}{2}$, $\operatorname{Var}\xi_j = \dfrac{1}{4}$, ezért $ES_n = 2\ 500\ 000$, $\operatorname{Var} S_n = \dfrac{1}{4} \cdot 5\ 000\ 000$, és a $P\left(\left|\dfrac{S_n - 2\ 500\ 000}{\sqrt{\frac{1}{4} \cdot 5\ 000\ 000}}\right| < x\right)$ valószínűségek kiszámítására alkalmazhatjuk a centrális határeloszlástételt. Ennek alapján

$$
\begin{aligned}
P(|2S - 5\ 000\ 000| \le 300) &= P\left(\left|\frac{S - ES}{\sqrt{\operatorname{Var} S}}\right| \le \frac{150}{\sqrt{\frac{1}{4} \cdot 5\ 000\ 000}}\right) \\
&\sim \Phi\left(\frac{150}{\sqrt{\frac{1}{4} \cdot 5000000}}\right) - \Phi\left(-\frac{150}{\sqrt{\frac{1}{4} \cdot 5000000}}\right) \\
&= 2\Phi\left(\frac{6\sqrt{5}}{100}\right) - 1 \sim 2\Phi(0.124) - 1 \sim 0.1.
\end{aligned}
$$

Megjegyzés: Valójában a feladatban tárgyalt modell némileg irreális. Általában különböző körzetek vannak, ahol a jelöltek népszerűsége eltérő. Egy jobb, a valóságot jobban követő modellben például azt tételezhetjük fel, hogy különböző körzetek vannak, az egyes körzetekben az egyes vélemények függetlenek, de az, hogy milyen valószínűséggel választja egy választó valamelyik jelöltet attól is függ, hogy mely körzetben lakik. Ez a modell is vizsgálható a centrális határeloszlástétel segítségével, de itt már a centrális határeloszlástétel általánosabb, független, de nem feltétlenül egyforma eloszlású valószínűségi változók összegének határeloszlását leíró alakjára van szükség.

Az előző feladat azt mutatja, hogy annak valószínűsége, hogy független valószínűségi változók viszonylag közel vannak a várható értékükhöz elég nagy. Nem irreális például feltételezni, hogy 5 000 000 szavazó által egy jelöltre leadott szavazatok száma mindössze 150-nel különbözik annak várható értékétől. A következő feladatban hasonló problémát tekintünk. Egy szabályos pénzdarabot 10 000 alkalommal feldobunk, és azt akarjuk tudni, mi annak a valószínűsége, hogy a fej-dobások száma mindössze 100-zal vagy 200-zal tér el annak várható értékétől. Ezt a valószínűséget kiszámoljuk pontosan a centrális határeloszlátétel segítségével. Másrészt megvizsgáljuk milyen becslést ad a Csebisev egyenlőtlenség. Ilyen módon információt kapunk arról is, hogy bizonyos esetekben milyen jó becslést ad a Csebisev egyenlőtlenség.

Tekintsünk egy szabályos pénzdarab 10 000 egymás utáni (független) feldobásából származó fej-írás sorozatot. Adjunk becslést a Csebisev 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, $\leq j \leq 10\,000$ 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, $1 \leq j \leq 10\,000$ független valószínűségi változók, $E\xi_j = \frac{1}{2}$, $\operatorname{Var} \xi_j = \frac{1}{4}$, 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 becslést adnunk. A Csebisev egyenlőtlenség az első valószínűségre a

$$P\left(\left|\sum_{j=1}^{10000}(\xi_j - E\xi_j)\right| > 100\right) \leq \frac{10000 \cdot \operatorname{Var} \xi_1}{100^2} = \frac{1}{4},$$

a második valószínűségre pedig a

$$P\left(\left|\sum_{j=1}^{10000}(\xi_j - E\xi_j)\right| > 200\right) \leq \frac{10000 \cdot \operatorname{Var} \xi_1}{200^2} = \frac{1}{16}$$

első becslést adja.

A centrális határeloszlástétel szerint $P\left(\frac{\sum\limits_{j=1}^{10000}(\xi_j - E\xi_j)}{\sqrt{10000 \cdot \frac{1}{4}}} > u\right) \sim 1 - \Phi(u)$. 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). (A Csebisev egyenlőtlenség 0.25 illetve 0.0625 felső becslést adta ezekre a valószínűségekre.)