A Valószínűségszámítás I. előadássorozat tizedik előadása.
2007. április 17.

Ezen az előadáson megfogalmazom a valószínűségszámítás talán legfontosabb eredményét, a centrális határeloszlástételt. (Az bizonyításával később fogok foglalkozni.) Most megmutatom az eredmény néhány alkalmazását gyakorlati feladatok megoldásában.

A centrális határeloszlástétel megfogalmazása előtt felidézek néhány korábbi fogalmat és eredményt. Egy ξ valószínűségi változó standard normális eloszlású, ha a sűrűségfüggvénye $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $-\infty < x < \infty$, alakú. Ha ξ standard normális eloszlású valószínűségi változó, akkor $E\xi = 0$, $\operatorname{Var}\xi = E\xi^2 = 1$. Ha ξ standard normális eloszlású valószínűségi változó, akkor $\sigma\xi + m$ egy m várható értékű és σ^2 valószínűségi változó. Az ilyen módon előállítható valószínűségi változókat nevezzük m várható értékű és σ^2 szórásnégyztű normális eloszlású valószínűségi változóknak. Be lehet látni, hogy egy ξ valószínűségi változó akkor és csak akkor normális eloszlású m várható értékkel és σ^2 szórásnégyzettel, ha sűrűségfüggvénye a $g(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m)^2/2\sigma^2}$ függvény. Ezt a tényt fogalmazom meg a következő észrevételben.

Észrevétel: *Egy η valószínűségi változó akkor és csak akkor m várható értékű és σ^2 szórásnégyzetű normális eloszlású valószínűségi változó, ha sűrűségfüggvénye a $g(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m)^2/2\sigma^2}$, $-\infty < x < \infty$ függvény.*

Indoklás: Ha $\eta = \sigma\xi + m$, ahol ξ standard normális valószínűségi változó, akkor η sűrűségfüggvénye $g(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m)^2/2\sigma^2}$, ahogy azt állítottuk. Megfordítva, ha η sűrűségfüggvénye az adott alakú, akkor a $\xi = \frac{\eta-m}{\sigma}$ valószínűségi változó sűrűségfüggvénye $f(x) = \sigma g(\sigma x + m) = \varphi(x)$, azaz ez standard normális eloszlású valószínűségi változó, és $\eta = \sigma\xi + m$. (Ezt az eredményt valójában már megtárgyaltuk a 8. előadás második feladatában.)

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

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

Megjegyzés: A fenti számolás meglehetősen kényelmetlen, de bizonyos általános elvek segíthetnek megérteni, hogy miért ilyen eredményt kaptunk. A felírt konvolució alakjából következik, hogy a végeredmény egy olyan kifejezés, amelynek exponensében egy kvadratikus alak szerepel, és ez meg van szorozva valamilyen konstanssal. Továbbá, a végeredmény egy sűrűségfüggvény, mivel két sűrűségfüggvény konvoluciója ismét sűrűségfüggvény. Ezért a végeredmény szükségszerűen normális sűrűségfüggvény. Ezenkívül, mivel független valószínűségi változók összegének a várható értéke és szórásngyzete egyenlő az összeadandók várható értkének illetve szórásnégyzetének az összegével, ezért a végeredmény egy olyan (normális eloszású) valószínűségi változó sűrűségfüggvénye, amelynek várható értéke $m_1 + m_2$ szórásnégyzete pedig $\sigma_1^2 + \sigma_2^2$.

Később megmutatom, hogy hogyan lehet a most tárgyalt azonosságot bizonyos eddig még nem tárgyalt eredmények segítségével egyszerűbben megkapni.

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 vezessünk be egy az irodalomban gyakran használt fogalmat.

Egy valószínűségi változó normalizáltjának a fogalma. *Legyen ξ olyan valószínűségi változó, amelyre $E\xi^2 < \infty$. 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, amelynek várható értéke nulla, szórásnégyzete pedig 1.*

Jegyezzük meg, hogy ha $\xi_1, \dots, \xi_n$ független valószínűségi változók, amelyekre $E\xi_j^2 < \infty$ 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.

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

2. megjegyzés. Láttuk, hogy független normális eloszlású valószínűségi változók részletösszegeinek normalizáltja standard normális eloszlású valószínűségi változó. Ez következik az előadás elején tárgyalt feladat eredményéből, amelyik szerint normális független normális eloszlású valószínűségi változók összege is normális eloszlású. 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.

3. megjegyzés. A műszaki tudományokban beszélnek egy olyan megfigyelésről, amelyet hibatörvénynek neveznek. Gyakran elófordul, hogy egy műszaki feladatban egy kisérletet többször egymástól függetlenül elvégeznek, azok eredményeit 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 hiba 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.

4. megjegyzés. 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.

5. megjegyzés. Természetesen felvetődik a kérdés, hogy létezik-e a centrális határeloszlástételnek több-dimenziós változata, amelyben 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, amelynek 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, amelyek némi információt nyújtanak a centrális határeloszlástétel gyakorlati következményeiről. A következő előadáson a centrális határeloszlástétel további alkalmazásaira mutatok példát.

Első példa:

A 2000. évben az Egyesült Államok Florida államában rendkívül szoros eredmény született. 5 000 000 választó választó választott két párt, a republikánus és demokrata párt jelöltjei között. A két jelölt által szerzett szavazatok száma (egy adott időpontbeli felmérés szerint) mindössze 300 volt. Tegyük fel, hogy a választók egymástól függetlenül $\frac{1}{2}$ valószínűséggel választották valamelyik párt jelöltjét. E feltevés teljesülése estén 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, $E\xi_j = \frac{1}{2}$, $\operatorname{Var} \xi_j = \frac{1}{4}$, ezért $ES_n = 2\ 500\ 000$, $\operatorname{Var} S_n = \frac{1}{4} \cdot 5\ 000\ 000$, és a $P\left(\left| \frac{S_n - 2\ 500\ 000}{\sqrt{\frac{1}{4} \cdot 5\ 000\ 000}} \right| < x \right)$ 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 kövelítő modellben például azt tételezhatjük 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 7. előadás végén (a 14. oldalon) tárgyalt Csebisev egyenlőtlenség. Ilyen módon információt kapunk arról is, hogy bizonyos esetekben milyen jó becslést ad a Csebisev egyenlőtlenség.

Második példa:

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, $\le j \le 10\,000$ valószínűségi változókat, amelyekre $\xi_j = 1$, 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 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

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

a második valószínűségre pedig a

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

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.)

További olyan feladatokat fogok tárgyalni, amelyeket a centrális határeloszlástétel segítségével lehet jól megoldani.

1.) Egy szabályos dobókockát feldobunk 1200 alkalommal egymástól függetlenül, és összeadjuk a páros értékű dobások eredményét. Adjunk jó közelítő becslést a centrális határeloszlástétel és egy normális eloszlástáblázat segítségével arra, hogy ez az összeg 2280 és 2500 közé esik.

Megoldás: Vezessük be a következő ξ_j, $1\le j\le 1200$, 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, $\xi_j=0$, ha a j-ik dobás eredménye 1, 3 vagy 5. Ekkor a $P\left(2280\le\sum\limits_{j=1}^{1200}\xi_j\le 2500\right)$ valószínűséget kell jól megbecsülnünk. Vegyük észre, hogy $E\xi_j=\frac16(2+4+6)=2$, $\operatorname{Var}\xi_j=\frac16(4+16+36)-4=\frac{16}{3}$. Innen a centrális határeloszlástétel alapján

$$P\left(2280\le\sum_{j=1}^{1200}\xi_j\le 2500\right)=P\left(\frac{-120}{\sqrt{1200\frac{16}{3}}}\le\frac{\sum\limits_{j=1}^{1200}\xi_j-\sum\limits_{j=1}^{1200}E\xi_j}{\sqrt{\sum\limits_{j=1}^{1200}\operatorname{Var}\xi_j}}\le\frac{100}{\sqrt{1200\frac{16}{3}}}\right)$$
$$\sim\Phi(1.25)-\Phi(-1.5)=0.8944+0.9322-1=0.8266.$$

2.) Vegyünk egy olyan pénzdarabot, amely $\frac23$ valószínűséggel esik a fej és $\frac13$ valószínűséggel az írás oldalra. Ezt a pénzdarabot annyiszor dobjuk fel, ameddig megjelenik 1200 fej dobás. Mi annak a valószínűsége, hogy az elvégzett dobások száma 1680 1830 közé esik? Adjunk erre a valószínűségre jó közelítő becslést.

Megoldás: Az elvégzett dobások száma egy η negatív binomiális eloszlású valószínűségi változó $n=1200$ és $p=\frac23$ paraméterekkel, azaz $P(\eta=k+n)=\binom{n+k-1}{n-1}(1-p)^kp^n$, $p=\frac23$, és $n=1200$ paraméterrel. Egy ilyen valószínűségi változónak ki lehet számolni a pontos eloszlását, azaz azt, hogy milyen értéket milyen valószínűséggel vesz fel. Elvileg, ez lehetőséget ad a kívánt valószínűség kiszámítására egy bonyolult összeg kiszámításának a segítségével. Ennél hasznosabb becslést tudunk kapni a következő érvelés segítségével, amely a kívánt valószínűséget jó pontossággal kiszámítja a centrális határeloszlástétel segítségével.

Jelölje ξ_j, $2\le j\le 1200$, a $j-1$-ik és j-ik fejdobás közötti dobások számát (a j-ik fejdobást beleszámítjuk a $j-1$-iket viszont nem számítjuk bele e dobások közé), és legyen ξ_1 az első fejdobásig (ezt is beleszámítva) elvégzett dobások száma. Ekkor a ξ_j valószínűségi változók függetlenek, negatív binomiális eloszlásúak $n=1$, $p=\frac23$ paraméterekkel, és minket a $P(1680<\xi_1+\cdots+\xi_{1200}<1830)$ valószínűség érdekel.

A 6. előadás 5. feladatában kiszámoltam ezt a várható értéket és szórásnégyzetet. Innen következik, hogy $E\xi_j = \frac{1}{p} = \frac{3}{2}$, $\operatorname{Var}\xi_j = \frac{1-p}{p^2} = \frac{3}{4}$. Ezért a centrális határeloszlástétel alapján $\eta = \xi_1 + \cdots + \xi_{1200}$ jelöléssel minket a

$$P\left(-4 < \frac{\eta - 1200E\xi_1}{\sqrt{1200\operatorname{Var}\xi_1}} < 1\right)$$

valószínűség érdekel. A centrális határeloszlástétel alapján

$$P\left(-4 < \frac{\eta - 1200E\xi_1}{\sqrt{1200\operatorname{Var}\xi_1}} < 1\right) \sim \Phi(1) + \Phi(4) - 1 \sim \Phi(1).$$

3.) Egy szabályos dobókockát és egy szabályos érmét feldobunk 3300 alkalommal egymástól függetlenül. (Az érme és kockadobások eredményei is függetlenek egymástól.) Ha a kockadobás eredménye páros és az érme a fej oldalra esett, akkor annyi forintot nyerünk, amennyi a kockadobás eredménye. Ha az érme az írás oldalra esett vagy a kockadobás eredménye páratlan szám, akkor nem nyerünk, és nem is vesztünk semmit. Mi a valószínűsége annak, hogy az össznyereményünk 3190 és 3520 forint közé esik? Adjunk erre jó közelítő becslést a centrális határeloszlástétel és egy normális eloszlástáblázat segítségével.

Megoldás: Vezessük be a következő ξ_j, és η_j $1 \le j \le 3300$, valószínűségi változókat: $\xi_j = 2$, ha a j-ik kockadobás eredménye 2, $\xi_j = 4$, ha a j-ik kockadobás eredménye 4, $\xi_j = 6$, ha a j-ik kockadobás eredménye 6, $\xi_j = 0$, ha a j-ik kockadobás eredménye 1, 3 vagy 5. Legyen $\eta_j = 1$, ha a j-ik érmedobás eredménye fej, és $\eta_j = 0$, ha a j-ik érmedobás írás. Legyen $\zeta_j = \xi_j\eta_j$. Ekkor a j-ik dobásnál a nyereményünk ζ_j lesz, $1 \le j \le 3300$, és a $P\left(3190 \le \sum\limits_{j=1}^{3300} \zeta_j \le 3520\right)$ valószínűséget kell jól megbecsülnünk. Ennek érdekében számoljuk ki a független ζ_j valószínűségi változók várható értékét és szórásnégyzetét. $E\zeta_j = E\xi_j\eta_j = E\xi_j E\eta_j = \frac{1}{6}(2+4+6)\cdot\frac{1}{2} = 1$, $E\zeta_j^2 = E\xi_j^2 E\eta_j^2 = \frac{1}{6}(4+16+36)\cdot\frac{1}{2} = \frac{14}{3}$, $\operatorname{Var}\zeta_j^2 = E\zeta_j^2 - (E\zeta_j)^2 = \frac{11}{3}$. Innen a centrális határeloszlástétel alapján

$$P\left(3190 \le \sum_{j=1}^{3300} \zeta_j \le 3520\right) = P\left(\frac{-110}{\sqrt{3300\frac{11}{3}}} \le \frac{\sum\limits_{j=1}^{3300} \zeta_j - \sum\limits_{j=1}^{3300} E\zeta_j}{\sqrt{\sum\limits_{j=1}^{3300} \operatorname{Var}\xi_j}} \le \frac{220}{\sqrt{3300\frac{11}{3}}}\right)$$

$$\sim \Phi(2) - \Phi(-1) = 0.9772 + 0.8413 - 1 = 0.9285.$$

4.) Legyen birtokunkban 100 lámpa, amelyek mindegyike egymástól független időtartamig működik, élettartamuk pedig exponenciális eloszlású $\lambda = \frac{1}{10}$ paraméterrel. (A lámpák élettartamának exponenciális eloszlása természetes feltételezés.) Egy termet bevilágítunk ezen lámpák valamelyikével, majd amikor az kiégett új lámpát használunk fel. Adjunk jó becslést arra, hogy a lámpák összélettartama legalább 1150 óra.

Megoldás: Jelölje ξ_j a j-ik lámpa éléttartamát, $1 \le j \le 100$. Ekkor a $P(\xi_1 + \cdots + \xi_{100} > 1150)$ valószínűségre kell jó becslést adnunk, ahol az összegben független exponenciális eloszlású valószínűségi változók szerepelnek $\lambda = \frac{1}{10}$ paraméterrel. Vezessük be az $\eta = \xi_1 + \cdots + \xi_{100}$ jelölést.

Kiszámoltuk, (lásd 8. előadás 7. feladatát), hogy jelen esetben $E\eta = mE\xi_1 = \frac{m}{\lambda} = 1000$, $\operatorname{Var}\eta = \frac{m}{\lambda^2} = 10000$ ($m = 100$ és $\lambda = \frac{1}{10}$ választással). Ezért a centrális határeloszlástétel szerint $\frac{\eta - E\eta}{\sqrt{\operatorname{Var}\eta}} = \frac{\eta - 1000}{100}$ jó közelítéssel standard normális eloszlású valószínűségi változó, és $P(\xi_1 + \cdots + \xi_{100} > 1150) = P\left(\frac{\eta - E\eta}{\sqrt{\operatorname{Var}\eta}} > 1.5\right) \sim 1 - \Phi(1.5)$.

A centrális határeloszlástétel nagy segítséget ad a matematikai statisztika két fontos problémakörében, a becsléselméletben és a hipotézisvizsgálatban. A becsléselméletben a következő problémákhoz hasonló kérdésekkel foglalkozunk. Mennyi egy lámpa várható élettartama, mennyi annak a valószínűsége annak, hogy egy kocka a hatos oldalra esik, mi a valószínűsége, annak, hogy egy (életbiztosítást kötni akaró ember) megéli a hetvenedik évet? Általában a kérdés a következő: Egy esemény eloszlása függ valamilyen ismeretlen paraméter(ek)től, és ez(eke)t a paraméter(eke)t akarjuk megbecsülni. Ennek érdekében kisérleteket végzünk, és valamilyen jó módszer segítségével próbáljuk megbecsülni az ismeretlen paramétert e kisérletek eredményének a függvényében.

A hipotézisvizsgálatban a feladat az, hogy bizonyos hipotézisünk van arról, hogy mennyi az élettartama egy lámpának, mennyire népszerű egy vélemény a választók között, hogy egy gyógyszer hatásosabb, mint egy másik. Általában, valamilyen hasonló problémáról van egy ellenőrizendő feltételezésünk. Annak érdekében, hogy eldöntsük, helyes-e a hipotézisünk kisérleteket végzünk. Ha kisérleteink eredménye nagyon valószínűtlen hipotézisünk teljesülése esetén, akkor hipotézisünket elutasítjuk, ha pedig valószínű, akkor elfogadjuk azt. De mikor tekintjük a bekövetkezett eredményt valószínűnek és mikor valószínűtlennek? E kérdés eldöntésében sokszor a centrális határeloszlástétel segít. Erre mutatok néhány példát. E példák tárgyalása előtt bevezetem a matematikai statisztika néhány egyszerű és természetes fogalmát. Jelen előadássorozatban csak a hipotézisvizsgálattal foglalkozom, és az ehhez kapcsolódó fogalmakat vezetem be.

Azt a feltételezést, amelyet ellenőrízni akarunk nevezik null-hipotézisnek, annak ellenkezőjét pedig ellenhipotézisnek. Vannak olyan feladatok, amelyekben a null-hipotézis egyetlen elemből áll, például, ha azt tételezzük fel, hogy egy dobókocka szabályos. Az ilyen hipotéziseket nevezik egyszerű hipotézisnek. Vannak olyan feladatok, amelyekben a null-hipotézis több elemből áll, például az a feltételezés, hogy egy javaslatot az emberek legalább fele támogat. Az ilyen hipotéziseket hívják összetett hipotézisnek. Két fajta hibát követhetünk el. Az egyik hiba az, hogy a hipotézis teljesül, mi mégis elutasítjuk azt. Ezt nevezik elsőfajú hibának. Azt, hogy a hipotézis nem teljesül,és mi mégis úgy döntünk, hogy az teljesül, másodfajú hibának nevezik. Az elsőfajú hiba csökkentése érdekében minél többször kell elfogadni, a másodfajú hipotézis csökkentése érdekében pedig minél többször kell elutasítani a hipotézist. A gyakorlatban általában úgy szokták a feladatokat megfogalmazni, hogy az elsőfajú hibára előírunk egy felső korlátot, és eme feltétel mellett próbáljuk a másodfajú hibát minél kisebbé tenni. Az elsőfajú hibára adott felső korlát összetett null-hipotézis esetén azt jelenti, hogy az elsőfajú hiba minden a null-hipotézis teljesülése esetén előfordulható eloszlás esetén legyen kisebb mint az előírt felső korlát.

5.) 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, $1 \le j \le 30\,000$, $S = S_{30000} = \sum\limits_{j=1}^{30\,000} \xi_j$. Ha a fejdobás eredményének valószínűsége pontosan $\frac{3}{4}$, akkor $E\xi_j = \frac{3}{4}$, $E\xi_j^2 = \frac{3}{4}$, $\operatorname{Var}\xi_j = E\xi_j^2 - (E\xi_j)^2 = \frac{3}{16}$, $ES = 30\,000 E\xi_j = 22\,500$, $\operatorname{Var} S = 30\,000 \operatorname{Var}\xi_j = 5625 = 75^2$. Innen és a centrális határeloszlástételből,

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

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 $k = 22\,500 - 75 \cdot 1.28$ és $p = \frac{3}{4}$ esetén annak valószínűsége, hogy a fejdobások száma nagyobb mint $k = 22\,500 - 75 \cdot 1.28 = 22\,212$ 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.

6.) Ledobunk egymástól függetlenül 24 000 pontot a $[0, 2]$ intervalumra egyenletes eloszlással, (azaz annak a valószínűsége, hogy egy ledobott pont értéke x-nél kisebb $\frac{x}{2}$-vel egyenlő, ha $0 \le x \le 2$, eggyel egyenlő, ha $x \ge 2$, és nulla, ha $x \le 0$.) Őrizzük meg azokat a ledobott pontokat, melyek értéke 1-nél kisebb, és hagyjuk el azokat, melyek értéke, nagyobb mint egy. Mi annak a valószínűsége, hogy a megőrzött pontok értékeinek az összege 5900 és 6075 közé esik? Adjunk erre a valószínűségre jó közelítő becslést a mellékelt normális eloszlástáblázat segítségével.

Megoldás: Vezessük be a következő ξ_j, $1 \le j \le 24\ 000$, valószínűségi változókat: $\xi_j = x$, ha a j-ik ledobott pont értéke x, és $0 \le x \le 1$, és $\xi_j = 0$, ha a j-ik ledobott pont értéke az $(1, 2]$ intervallumba esik. Ekkor a megőrzött pontok összege $S = \sum\limits_{j=1}^{24\ 000} \xi_j$, továbbá a ξ_j valószínűségi változók függetlenek és egyforma eloszlásúak. Ezért a centrális határeloszástétel segítségével jó becslést tudunk adni a minket érdeklő $P(5900 < S < 6075)$ valószínűségre. Ennek érdekében ki kell számolnunk a ξ_1 valószínűségi változó várható értékét és szórásnégyzetét. Vegyük észre, hogy ugyan a ξ_1 valószínúségi változónak nincs sűrűségfüggvénye, és nem diszkrét eloszlású, viszont anak F eloszlásfüggvénye felírható $F(x) = F_1(x) + F_2(x)$ alakban, ahol $F_1(x)$ nek van sűrűségfüggvénye, ami az $f(x) = \frac{1}{2}$ függvény a $[0,1]$ intervallumon, és $f(x) = 0$, ha $x < 0$ vagy $x > 1$, és $F_2(x)$ olyan mértéket határoz meg, amelyik a nullába van koncentrálva, és a nulla mértéke $\frac{1}{2}$. Pontosabban, tetszőleges A halmaz valószínűsége $P(A) = \int_A F(\,dx) = \int_A F_1(\,dx) + \int_A F_2(\,dx) = \int_A \frac{1}{2}\,dx + \int_A F_2(\,dx)$, ahol $\int_A F_2(\,dx) = \frac{1}{2}$, ha $0 \in A$, és $\int_A F_2(\,dx) = 0$, ha $0 \notin A$. Ekkor tetszőleges $h(x)$ függvényre $Eh(\xi) = \int h(x)F(\,dx) = \int h(x)F_1(\,dx) + \int h(x)F_2(\,dx) = \int_0^1 \frac{1}{2}h(x)\,dx + \frac{1}{2}h(0)$. Ezért

$$E\xi_1 = \frac{1}{2}\int_0^1 x\,dx = \frac{1}{4}, \quad E\xi_1^2 = \frac{1}{2}\int_0^1 x^2\,dx = \frac{1}{6}, \quad \operatorname{Var}\xi_1 = \frac{1}{6} - \frac{1}{16} = \frac{5}{48},$$

és $ES = 6000$, $\operatorname{Var} S = 2500$. Innen

$$\begin{aligned} P(5900 < S < 6075) &= P\left(-2 < \frac{S - ES}{\sqrt{\operatorname{Var} S}} < 1.5\right) \\ &\sim \Phi(1.5) - \Phi(-2) = \Phi(1.5) + \Phi(2) - 1. \end{aligned}$$

A ξ_j valószínűségi változók várható érékét és szórásnégyzetét hagyományos módon (Stieltjes integrálok használata nélkül) is kiszámolhattuk volna a következőképpen. Legyen η_j, a j-ik ledobott pont értéke, $f(x)$ az a függvény a $[0,2]$ intervallumon, amelyre $f(x) = x$, ha $0 \le x \le 1$, és $f(x) = 0$, ha $1 < x \le 2$. Ekkor $\xi_j = f(\eta_j)$, és innen az η_j sűrűségfüggvényének az ismeretében fel tudjuk írni, hogy $E\xi_j = Ef(\eta_j) = \int_0^2 f(x)\frac{1}{2}\,dx = \int_0^1 \frac{1}{2}x\,dx = \frac{1}{4}$, és $E\xi_j^2 = Ef^2(\eta_j) = \int_0^2 f^2(x)\frac{1}{2}\,dx = \int_0^1 \frac{1}{2}x^2\,dx = \frac{1}{6}$.

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.

A centrális határeloszlástétel egy fontos speciális esete a binomiális eloszlásfüggvény közelítéséről szól a normális eloszlásfüggvény segítségével. Történetileg is ez az eredmény született meg először. Ismertetem ezt az irodalomban Moivre–Laplace formula néven tárgyalt eredményt. Valójában ez az eredmény kissé más jellegű, mint a klasszikus centrális határeloszlástétel. Ez úgynevezett lokális centrális határeloszlástétel, amely annak valószínűségére ad jó becslést, hogy egy n és p paraméterű $B(n,p)$ binomiális eloszlású valószínűségi változó egy előírt 'tipikus' értéket vesz fel.

Emlékeztetek arra, hogy ha elvégzünk n független kisérletet, amelyek mindegyike egymástól függetlenül p valószínűséggel sikeres, akkor a sikeres kisérletek száma egy $B(n,p)$, n és p paraméterű binomiális eloszlású valószínűségi változó. Ezt a következőképpen is megfogalmazhatjuk. Vezessük be a ξ_j, $1 \le j \le n$, valószínűségi változókat, amelyekre $\xi_j = 1$, ha a j-ik kisérlet sikeres és $\xi_j = 0$, ha a j-ik kisérlet sikertelen volt. Ekkor $P(\xi_j = 1) = 1 - P(\xi_j = 0) = p$ minden $1 \le j \le n$ indexre, a ξ_j valószínűségi változók függetlenek, és az $S_n = \sum\limits_{j=1}^{n} \xi_j$ valószínűségi változó n és p paraméterű binomiális eloszlású valószínűségi változó. A Moivre–Laplace formula a $P(S_n = k)$ valószínűségekre ad jó közelítést. Vegyük észre, hogy ezeket a valószínűségeket explicit módon is fel tudjuk írni. Nevezetesen, $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$. A Moivre–Laplace formula valójában egy analízisbeli eredmény, amely erre a kifejezésre ad jó aszimptotikát. Bizonyításában kulcsszerepet játszik az alábbi Stirling formulának nevezett híres eredmény, amely az $n!$ mennyiség egy jó közelítését írja le. A következő előadáson ismertetni fogom a Stirling formulának egy olyan bizonyítását, amely részben valószínűségszámítási gondolatokon alapul.

Stirling formula.

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

azaz az első n egész szám $n! = 1 \cdot 2 \cdot 3 \cdots n$ szorzata teljesíti a

$$\lim_{n\to\infty} \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{n!} = 1$$

relációt.

A Stirling formula és alkalmas a Taylor sorfejtés segítségével kapható közelítések segítségével belátható az alábbi

Moivre–Laplace formula. *Legyen S_n binomiális eloszlású valószínűségi változó valamely n és p paraméterekkel, $n = 1, 2, \dots$, $0 < p < 1$. Rögzítsünk egy tetszőleges $A > 0$ számot. Ekkor*

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left\{-\frac{(k-np)^2}{2np(1-p)}\right\} + \varepsilon_{k,n},$$

ha $|k - np| < A\sqrt{np(1-p)}$ egy olyan $\varepsilon_{k,n}$ hibataggal, amely teljesíti a $\lim\limits_{n\to\infty} \frac{\varepsilon_{k,n}}{\sqrt{n}} = 0$, becslést, és a limesz egyenletes a k változóban $|k - np| < A\sqrt{npq}$ esetben.

Értsük meg a fenti közelítés jelentését. Az S_n valószínűségi változó várható értéke $ES_n = np$, szórásnégyzete $\operatorname{Var} S_n = np(1-p)$. Ezért a a centrális határeloszlástétel a $P\left(\frac{S_n - np}{\sqrt{np(1-p)}} < x\right) = P(S_n < np + x\sqrt{np(1-p)})$ valószínűségekre ad jó becslést.

Vezessük be a $x = x_k = \frac{k-np}{\sqrt{np(1-p)}}$ mennyiségeket. Az x_k mennyiségek, $k = 0$, ± 1, $\pm 2, \dots$, egy $\frac{1}{\sqrt{np(1-p)}}$ sűrűségű rácson vannak. Ezért nagy n indexre a centrális eloszlástétel differenciálásához hasonló formális számolás (differenciahányados képzés differenciálhányados helyett) azt sugallja, hogy

$$P(S_n = k) = P\left(\frac{S_n - np}{\sqrt{np(1-p)}} = x_k\right) = P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \in [x_k, x_{k+1})\right)$$
$$\sim \Phi(x_{k+1}) - \Phi(x_k) \sim \varphi(x_k)(x_{k+1} - x_k),$$

ahol $\Phi(x)$ a standard normális eloszlás eloszlásfüggvényét, $\varphi(x)$ pedig a sűrűségfüggvényét jelöli. A Moivre–Laplace formula azt mondja ki, hogy ezen heurisztikus módon levezetett közelítő azonosság két oldalán levő kifejezés hányadosa közel egy nagy n számokra és 'tipikus' x_k értékekre. A 'tipikus' jelző itt arra utal, hogy elég nagy A számokra, és nagy n indexekre az $|S_n - np| < A\sqrt{np(1-p)}$ esemény valószínűsége majdnem 1. (Ez például a Csebisev egyenlőtlenségből egyszerűen levezethető.)

Valójában a Moivre–Laplace formula az itt kimondottnál élesebb alakban is érvényes. Egyrészt az a k számokra előírt tartomány, ahol ez a formula érvényes jóval nagyobbnak is választható, másrészt az $\varepsilon_{k,n}$ hibatagokra sokkal jobb becslés is adható. Számunkra azonban nem a Moivre–Laplace formula minél élesebb alakja, hanem az eredmény tartalmának a megértése érdekes elsősorban. Ez a $P(S_n < x)$ eloszlásfüggvény helyett a $P(S_n = x)$ valószínűségekre ad jó becslést, ezért nevezik ezt lokális centrális határeloszlástételnek. A lokális centrális határeloszlástétel (amely csak bizonyos extra kikötések esetén érvényes) valójában erősebb állítás, mint a centrális határeloszlástétel, mert függvénysorozatok differenciálhányadosainak vagy 'majdnem' differenciálhánydosainak a konvergenciája erősebb tulajdonság, mint a függvénysorozatok konvergenciája. A Moivre–Laplace formula 'kiintegrálásának' a segítségével le lehet vezetni a centrális határeloszlástételt binomiális eloszlásokra. Az ehhez szükséges számolások részleteinek kidolgozását azonban elhagyom.

A Moivre–Laplace formula bizonyítása. Felírhatuk a

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

azonosságot, és az ebben az azonosságban szereplő $n!$, $k!$ és $(n-k)!$ kifejezésekre jó közelítést kaphatunk a Stirling formula segítségével. Azt kapjuk, hogy

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

ahol $\sim$ azt jelenti, hogy a két kifejezés hányadosa 1-hez tart $n \to \infty$ esetben. Sőt ez az 1-hez tartó konvergencia egyenletes, ha $|k - np| \le A\sqrt{np(1-p)}$.

A $P(S_n = k)$ valószínűségre adott utolsó kifejezés első tagjára a $\frac{1}{\sqrt{2\pi n \frac{k}{n}(1-\frac{k}{n})}} = \frac{1}{\sqrt{2\pi np(1-p)}} + O(n^{-1})$ közelítés írható fel a $|k - np| \le A\sqrt{np(1-p)}$ esetben, a második tag pedig

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

alakban írható. Alkalmazva a $\log(1+x) = x - \frac{x^2}{2} + O(x^3)$, Taylor sorfejtésből adódó közelítést, felhasználva, hogy mind $\frac{np-k}{k}$, mind $\frac{n-k}{k-np} n - k$ abszolut értéke kisebb, mint const. $n^{-1/2}$, azt kapjuk az utolsó formulából, hogy

$$\left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k} = \exp\left\{-\frac{(k-np)^2}{2k} - \frac{(k-np)^2}{2(n-k)} + O(n^{-1/2})\right\}$$
$$= \exp\left\{-\frac{(k-np)^2}{2n\frac{k}{n}\left(1-\frac{k}{n}\right)} + O(n^{-1/2})\right\}.$$

(Itt kihasználtuk, hogy a Taylor sorfejtés első tagja kiesik. Továbbá, mivel $\frac{k}{n}\left(1 - \frac{k}{n}\right) = p(1-p) + O(n^{-1/2})$ és $(k-np)^2 = O(n)$ a fenti relációból adódik, hogy

$$\left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k} = \exp\left\{-\frac{(k-np)^2}{2np(1-p)} + O(n^{-1/2})\right\}.$$

A kapott relációk alapján

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

A Moivre–Laplace formulát bebizonyítottuk.