9. Határeloszlás-tételek

A korábbi fejezetek során már többször előkerült az ún. határeloszlások témaköre: megnéztük, hogy a Poisson-eloszlás valószínűségeit közelítik a binomiális eloszlás valószínűségei, megfelelő paraméterezés esetén, illetve a normális eloszlásnál szó esett a de Moivre–Laplace-tételről.[38] De egyáltalán mit jelent az, hogy eloszlások „tartanak” egy másik eloszláshoz? És csak speciális alakú eloszlások esetén lehet határeloszlást bizonyítani? A mostani előadásban erről is szó lesz.

Először olyan egyenlőtlenségeket nézünk, amelyekkel egy valószínűségi változó eloszlásának „széleinek” valószínűségei becsülhetők. Ezek az egyenlőtlenségek hasznos segédeszközök határeloszlás-tételek bizonyításában is. Ezután a valószínűségszámítás két alaptételéről fogunk beszélni: a nagy számok törvényéről, és a centrális határeloszlás-tételről.

9.1. Csebisev-egyenlőtlenség

Egy valószínűségi változó eloszlása nem feltétlenül ismert egy probléma esetében, különösen igaz ez a gyakorlatra. Igény viszont lenne rá, hogy ilyenkor is meg tudjuk becsülni (például felülről), mekkora valószínűséggel lesz az X értéke szélsőséges.

9.1.1. Állítás (Markov-egyenlőtlenség)**.** *Legyen X nemnegatív értékű valószínűségi változó. Ekkor minden $a > 0$ esetén*
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Bizonyítás. Definiáljuk a következő valószínűségi változót: legyen $Y = a$ ha $X \geq a$, és 0 egyébként. Más szavakkal, $Y = a\mathbf{1}_{\{X \geq a\}}$. Mivel X nemnegatív, így $Y \leq X$ a teljes eseménytéren. Emiatt $\mathbb{E}(Y) \leq \mathbb{E}(X)$ is teljesül, a várható érték monotonitásából adódóan. Tehát
$$\mathbb{E}(X) \geq \mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + a \cdot \mathbb{P}(Y = a) = a \cdot \mathbb{P}(X \geq a).$$
Az egyenlőtlenség átrendezéséből következik az állítás. □

A Markov-egyenlőtlenség önmagában nem feltétlenül erős becslés. Például, ha $a < \mathbb{E}(X)$, akkor annyit állít, hogy egy valószínűség kisebb, mint egy 1-nél nagyobb szám, ami azért nem egy mély tény. Tipikus alkalmazása ehelyett, amikor a jóval nagyobb $\mathbb{E}(X)$-nél. Ekkor intuitívan azt állítja, hogy annak a valószínűsége, hogy az X változó a-nál szélsőségesebb értéket vesz fel, legalább olyan ütemben csökken, mint az $x \mapsto \frac{c}{x}$ függvény, ahol $c = \mathbb{E}(X)$.

A Markov-egyenlőtlenség helyett gyakrabban alkalmazzák az alábbi következményét:

9.1.2. Következmény (Csebisev-egyenlőtlenség[39])**.** *Legyen X valószínűségi változó, amire $\mathbb{D}^2(X)$ véges. Ekkor minden $a > 0$ esetén*
$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq a\big) \leq \frac{\mathbb{D}^2(X)}{a^2}.$$

Bizonyítás. A bal oldal átrendezése után alkalmazzuk a Markov-egyenlőtlenséget az X helyett az $\big(X - \mathbb{E}(X)\big)^2$ valószínűségi változóra (és a helyett a^2-re):
$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq a\big) = \mathbb{P}\Big(\big(X - \mathbb{E}(X)\big)^2 \geq a^2\Big) \leq \frac{\mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big)}{a^2} = \frac{\mathbb{D}^2(X)}{a^2}.$$
felhasználva, hogy $\big(X - \mathbb{E}(X)\big)^2$ mindig nemnegatív. Ez épp a bizonyítandó állítás. □

Vegyük észre, hogy míg a Markov-egyenlőtlenség csak nemnegatív valószínűségi változókra érvényes, addig a Csebisev-egyenlőtlenség tetszőleges valós esetben is igaz.

---

[38] Szemfüles olvasó felfedezheti, hogy a geometriai eloszlásról is megemlítettük, hogy a pontok sűrítése esetén az exponenciális eloszláshoz tart.

[39] Angol szakirodalomban: Chebyshev's inequality

9.1.3. Példa. Egy adott adatbázis átlagosan 50 lekérést fogad egy időegység alatt. A lekérések számának szórása a tapasztalatok szerint 5.[40] Adjunk alsó becslést annak a valószínűségére, hogy 40-nél több, de 60-nál kevesebb lesz a lekérések száma egy időegység alatt.

Jelölje X az egy időegység alatti lekérések számát. Ekkor a Csebisev-egyenlőtlenség szerint

$$\mathbb{P}(40 < X < 60) = \mathbb{P}(|X-50| < 10) = 1 - \mathbb{P}(|X-50| \geq 10) \geq 1 - \frac{\mathbb{D}^2(X)}{a^2} = 1 - \frac{5^2}{10^2} = \frac{3}{4}.$$

Vegyük észre, hogy ez a korlát nem igényelt semmilyen feltételt az eloszlásra, a várható értékének és szórásának ismeretén felül.

A Markov-egyenlőtlenség további, erősebb becslések alapjául is szolgálhat, amennyiben a négyzetre emelés helyett egyéb függvényt alkalmazunk, illetve X-re erősebb feltételeket teszünk.

9.1.4. Következmény (Paraméteres Csernov-egyenlőtlenség[41])**.** *Legyen X valószínűségi változó. Ekkor minden $a, t > 0$ esetén*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}}.$$

Bizonyítás. Tetszőleges $t > 0$ esetén az $x \mapsto e^{tx}$ függvény monoton növő. Így a Markov-egyenlőtlenség miatt

$$\mathbb{P}\big(X \geq a\big) = \mathbb{P}\big(e^{tX} \geq e^{ta}\big) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}},$$

felhasználva, hogy e^{tX} és e^{ta} mindig pozitív. □

9.1.5. Példa. Legyen $X \sim \mathrm{Pois}(5)$. Adjunk felső becslést $\mathbb{P}(X \geq 10)$-re. Ekkor a Csernov-egyenlőtlenség szerint

$$\mathbb{P}(X \geq 10) \leq \frac{\mathbb{E}(e^{tX})}{e^{10t}} = e^{-10t} \sum_{k=0}^{\infty} e^{tk} \frac{5^k}{k!} e^{-5} = e^{-10t} \sum_{k=0}^{\infty} \frac{(5e^t)^k}{k!} e^{-5} = e^{-10t+5e^t-5}.$$

Ez a felső becslés bármilyen pozitív t választása esetén igaz, így akkor is, ha a kitevő a lehető legkisebb. Deriválással megkapható, hogy a függvény minimuma $t = \ln(2)$ helyen van, értéke: $e^5/1024 \approx 0{,}1449$. Vegyük észre, hogy ezzel szemben a Markov-egyenlőtlenség csak $\mathbb{E}(X)/10 = \frac{1}{2}$ értékkel becsülne felülről.

A fejezet további részében a fenti ötleteket alkalmazva jutunk el a téma két legfontosabb tételéhez.

9.2. Nagy számok törvénye

A köznyelvben a nagy számok törvényét sokszor abban az értelemben használjuk, hogy ha valamit nagyon sokszor próbálgatunk, akkor előbb-utóbb sikerülni fog. Ez az állítás (illetve a precíz megfogalmazása) speciális esete annak, amit a valószínűségszámításban nagy számok törvényének hívunk. Valójában a „törvény” ennél általánosabb, ugyanis nem csak valószínűségekről beszél, hanem várható értékről is.

Fogalmazzuk meg először a fenti köznyelvi igazságot precízen. Legyenek $A_1, \dots, A_n$ együttesen független események, amelyek valószínűsége egységesen p, ahol $0 < p < 1$. Ezek az események felelnek meg annak, hogy n-szer próbálgatjuk ugyanazt a kísérletet. Az állítás olyasmi, hogy elég nagy n esetén biztosan sikerül valamelyik, még akkor is, ha p nagyon kicsi. Precízen megfogalmazva,

$$\lim_{n\to\infty} \mathbb{P}(A_1 \cup \dots \cup A_n) \to 1.$$

Vegyük észre, hogy valójában a valószínűség fogalmának bevezetésénél már ennél többet is elfogadtunk: azt az intuitív igazságot, hogy független kísérletek esetén a sikerek aránya épp a siker valószínűségéhez

[40] Hogy hogyan lehet a tapasztalat alapján, például ismételt kísérletekkel információt szerezni az eloszlás szórásáról, az a statisztika témakörébe eső nem nyilvánvaló kérdés, amire most nem térünk ki.

[41] Angol szakirodalomban: Chernoff bound

közelít (és emiatt elég sok kísérlet esetén nem is lesz nulla). Ez szintén a nagy számok törvényének speciális esete.

Hogy az utóbbit formulával is felírhassuk, vezessük be a következő jelölést: legyen $X_i = 1$ ha A_i teljesül, egyébként 0 (azaz $X_i = \mathbf{1}_{A_i}$). Jelölje $\overline{X_n}$ az $\frac{X_1+\ldots+X_n}{n}$ átlagot. Ekkor a „korábban elfogadott intuitív igazság" azt állítja, hogy ha $n \to \infty$, akkor

$$\overline{X_n} \to p.$$

A probléma az, hogy ez az állítás nem precíz: milyen értelemben tud valószínűségi változók egy sorozata konvergálni? Ahogy az a következő tételből kiderül, többféleképp.

9.2.1. Tétel. *Legyen $X_1, X_2, \ldots, X_n, \ldots$ független, azonos eloszlású valószínűségi változók egy sorozata. Tegyük fel, hogy $\mathbb{E}X_n = \mu$ és $\mathbb{D}(X_n) = \sigma$ minden n-re, ahol $\mu, \sigma \in \mathbb{R}$ rögzített. Jelölje $\overline{X_n}$ az $\frac{X_1+\ldots+X_n}{n}$ átlagot.*[42]

- ***Nagy számok gyenge törvénye:*** *Tetszőleges $\varepsilon > 0$ esetén*

$$\lim_{n\to\infty} \mathbb{P}\Big(\left|\overline{X_n} - \mu\right| \geq \varepsilon\Big) = 0.$$

- ***Nagy számok erős törvénye:***

$$\mathbb{P}\Big(\lim_{n\to\infty} \overline{X_n} = \mu\Big) = 1,$$

ahol a határérték úgy értendő, mint az $\omega \mapsto \overline{X_n}(\omega)$ függvények pontonkénti határértéke.

Ahogy az a tételből is kiderül, a nagy számok törvénye nem egyetlen tétel, hanem egy ún. tételkör, adott témájú állítások egy halmaza, különböző feltételekkel.

Az állítás annyiban általánosabb az előtte szereplő diszkussziónál, hogy X_i nem feltétlenül $\{0,1\}$-értékű, hanem bármilyen egyéb eloszlású is lehet. Ha X_i $\{0,1\}$-értékű, akkor a gyenge törvény neve: **Bernoulli-féle nagy számok gyenge törvénye**.

Nagy számok gyenge törvényének bizonyítása. A várható érték linearitása miatt az $\overline{X_n}$ átlag várható értéke is μ. Így alkalmazhatjuk a Csebisev-egyenlőtlenséget az állításban szereplő valószínűségre:

$$\mathbb{P}\big(\left|\overline{X_n} - \mu\right| \geq \varepsilon\big) = \mathbb{P}\Big(\left|\overline{X_n} - \mathbb{E}(\overline{X_n})\right| \geq \varepsilon\Big) \leq \frac{\mathbb{D}^2(\overline{X_n})}{\varepsilon^2}.$$

Mivel $X_1, \ldots, X_n$ függetlenek, így a $\mathbb{D}^2$ tulajdonságai szerint:

$$\mathbb{D}^2(\overline{X_n}) = \mathbb{D}^2\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\Big(\mathbb{D}^2(X_1) + \cdots + \mathbb{D}^2(X_n)\Big) = \frac{n}{n^2}\sigma^2,$$

ahol a jobb oldal 0-hoz tart $n \to \infty$ esetén. Ebből az állítás már következik. □

Mitől gyenge törvény az első, és erős a második? És egyáltalán, mi a különbség a kettő közt? Egyrészt, az erős törvényből következik a gyenge (ezt nem bizonyítjuk). Másrészt, a két állítás abban különbözik, hogy milyen értelemben nézzük az $\overline{X_n}$ konvergenciáját. A gyenge törvényben szereplő konvergencia neve *sztochasztikus konvergencia*, míg az erős törvényben *1-valószínűségű konvergenciá*ról beszélünk.

Rendben, de mit jelent az, hogy „más a konvergencia"? Van itt valami kézzelfogható eltérés, vagy ez csak valami elméleti-technikai kötözködés? És egyébként is, ha igaz az erős törvény, minek egyáltalán beszélni a gyengéről? Ezek jogos kérdések, haladjunk sorban.

A valószínűségszámítást néha azért nehéz összevetni a valósággal, mert míg a valószínűségek az összes kimenetelről beszélnek egyszerre, addig a valóságban mi (egyszerre) csak egyetlen kimenetelt látunk, egy ω esetén felvett értéket tapasztalunk. A gyenge törvény (és általában a sztochasztikus konvergencia), csak arról beszél, hogy azon ω-k, ahol az átlagtól való eltérés lényeges (azaz ε-nál nagyobb), egyre kevesebben vannak (sőt, részarányuk nullához tart). De ettől még megtörténhetne, hogy akármilyen ω-t is választunk, egyre-másra felbukkan egy olyan lépés, ahol $\overline{X_n}$ nagyon eltér μ-től.

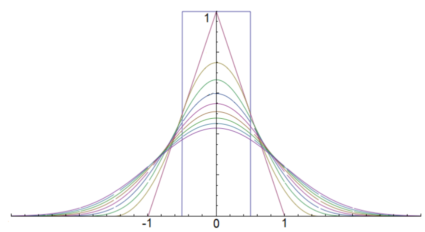
[42] A tételek a szórás végessége nélkül is teljesülnek, de a bizonyításuk komplikáltabb.

Az erős törvény éppen azt mondja, hogy ez lehetetlen. Vagyis 1 valószínűséggel olyanok az ω-k, hogy egy idő után már nem megyünk μ-től ε-nál távolabb.

Ennek fényében még jogosabb a kérdés: minek beszélünk egyáltalán az „elavult” gyenge törvényről? Azért, mert gyengébb feltételek esetén megeshet, hogy a gyenge tétel teljesül, de az erős már nem. Részletesebben: mind a függetlenség, mind az azonos eloszlásúság nagyon erős feltételek, amik a gyakorlatban sem mindig teljesülnek. (Például, ha elég ideig méregetjük a napi középhőmérsékleteket, azok nem lesznek függetlenek, továbbá előbb-utóbb nyár után tél lesz, legalábbis reméljük.) Szerencsére ezeket a feltételeket gyengítve is bizonyíthatók nagy számok törvényei. Viszont van olyan feltétel-gyengítés, ahol a gyenge törvény még igaz, de az erős már nem.

9.3. Centrális határeloszlás-tétel

A nagy számok törvénye csak arról beszél, hogy mihez konvergálunk (a várható értékhez, ugyebár), de arról nem, hogy ezt milyen gyorsan tesszük. Más szavakkal, arról nem mond semmit, hogy hogy a várható értéktől való eltérés milyen ütemben csökken, n növekedésének függvényében.



Ahhoz, hogy ezt megválaszoljuk, szükségünk van az eloszlásbeli konvergencia fogalmára, ugyanis az átlagtól való (egyre kisebb) eltérés már nem egy konkrét szám, hanem egy (egyre koncentráltabb) eloszlás a várható érték körül.

9.3.1. Definíció. Legyen $X_1, X_2, \dots, X_n, \dots$ valószínűségi változók egy sorozata, jelölje F_{X_n} az X_n eloszlásfüggvényét, és hasonlóan F_Z a Z eloszlásfüggvényét. Az $(X_n)_{n\in\mathbb{N}}$ sorozat **eloszlásban konvergál** egy Z valószínűségi változóhoz, ha

$$F_{X_n}(x) \to F_Z(x) \qquad (n \to \infty)$$

minden olyan $x \in \mathbb{R}$ esetén, ahol F_Z folytonos.[43] Jelölése: $X_n \xrightarrow{d} Z$.

9.3.2. Tétel (Centrális határeloszlás-tétel)**.** *Legyen $X_1, X_2, \dots, X_n, \dots$ független, azonos eloszlású valószínűségi változók egy sorozata. Tegyük fel, hogy $0 < \mathbb{D}^2 X_1 < \infty$. Legyen $Z \sim N(0;1)$. Ekkor*

$$\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} \xrightarrow{d} Z \qquad (n \to \infty).$$

Az eloszlásbeli konvergencia definícióját kibontva ez azt jelenti, hogy a bal oldal eloszlásfüggvénye tart Z eloszlásfüggvényéhez, azaz Φ-hez, a Φ minden folytonossági pontjában. Mivel Φ mindenhol folytonos, így a tétel következtetése a következő:

$$\mathbb{P}\left(\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} < a\right) \to \Phi(a)$$

minden $a \in \mathbb{R}$ esetén, ha $n \to \infty$.

Az utóbbi alakból már jobban látható, hogy a centrális határeloszlás-tétel a de Moivre–Laplace-tétel általánosítása. Valóban, hiszen itt olyan független, azonos eloszlású valószínűségi változókról beszélünk, amik véges szórásúak, míg a de Moivre–Laplace-tétel csak független indikátor valószínűségi változók összegéről (azaz egy binomiális eloszlású valószínűségi változóról) állít konvergenciát. De az állítás mindkét esetben az, hogy az eloszlásfüggvény Φ-vel közelíthető. Hogy ez a közelítés mennyire jó, arról külön tételek rendelkeznek.[44]

---

[43] Bebizonyítható, hogy az eloszlásbeli konvergencia fenti definíciójával ekvivalens a következő feltétel: $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ minden $f : \mathbb{R} \to \mathbb{R}$ folytonos, korlátos függvény esetén.

[44] Lásd Berry-Esseen tétel, [Terence Tao] Central limit theorem, Theorem 23

Megjegyzés. A tételben szereplő szumma más formában is írható:

$$\frac{X_1+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)}=\frac{\overline{X_n}-\mathbb{E}(\overline{X_n})}{\mathbb{D}(\overline{X_n})},$$

ami átrendezéssel és a szórás tulajdonságainak felhasználásával kapható. Más szavakkal, a tételben az $\overline{X_n}$ standardizáltjáról van szó.

9.3.3. Példa. Egy pincészetben munkanapokon átlagosan 100 liter bort mérnek ki, 20 szórással. Tegyük fel, hogy az egyes napok mérései függetlenek és azonos eloszlásúak. Az évből hátralévő 50 munkanap alatt 4750 liter bort kellene eladniuk ahhoz, hogy felülmúlják a tavalyi teljesítményt. Mi annak a valószínűsége, hogy ez sikerül?

Jelölje az egyes napok eredményeit $X_1, X_2, \ldots, X_{50}$. A feltételek szerint $\mathbb{E}(X_1) = 100$ és $\mathbb{D}(X_1) = 20$. A centrális határeloszlás-tétel szerint a megfelelően standardizált összeg közelítőleg normális, azaz

$$\mathbb{P}\Big(\sum_{i=1}^{50} X_i \geq 4750\Big) = 1-\mathbb{P}\Big(\frac{\sum_{i=1}^{50} X_i - 50\cdot 100}{\sqrt{50}\cdot 20} < \frac{4750-5000}{\sqrt{50}\cdot 20}\Big) \approx 1-\Phi\Big(\frac{4750-5000}{\sqrt{50}\cdot 20}\Big)$$
$$= 1-\Phi(-1{,}7678) = \Phi(1{,}7678) \approx 0{,}9615.$$

Tehát a siker valószínűsége nagyjából 96%.

A tétel bizonyításához az első alfejezetből meríthetünk ötletet, egész pontosan a Csernov-egyenlőtlenségben megjelenő mennyiségből: az $\mathbb{E}(e^{tX})$ várható értékből.

9.3.4. Definíció. Az X valószínűségi változó **momentumgeneráló függvény**e:[45]

$$M_X(t) \overset{\text{def}}{=} \mathbb{E}(e^{tX}),$$

azon t helyeken értelmezve, ahol $\mathbb{E}(e^{tX})$ véges.

A momentumgeneráló függvényt az alábbi hasznos tulajdonságai teszik alkalmassá, hogy felhasználjuk a centrális határeloszlás-tétel bebizonyításában.

9.3.5. Állítás. *Legyenek Y, Z illetve $X_1, X_2, \ldots$ valószínűségi változók. Tegyük fel, hogy $M_Y(t)$ és $M_Z(t)$ minden $t \in \mathbb{R}$ esetén értelmezett.*

(1) Ha $M_Y(t) = M_Z(t)$ minden $t \in \mathbb{R}$ esetén, akkor Y és Z azonos eloszlásúak.

(2) Ha $\lim_{n\to\infty} M_{X_n}(t) = M_Z(t)$ minden $t \in \mathbb{R}$ esetén, akkor $X_n \xrightarrow{d} Z$.

A bizonyításban feltesszük, hogy a momentumgeneráló függvények mindenhol értelmezettek. Az állítás enélkül is teljesülne, de az általános esetet itt nem bizonyítjuk. Továbbá a bizonyításban megjelenő lemmát szintén érvelés nélkül elfogadjuk.

Centrális határeloszlás-tétel bizonyításvázlata. A számolást egyszerűsítendő, csak azt az esetet vizsgáljuk, amikor $\mathbb{E}(X_1)$ =0 és $\mathbb{D}(X_1) = 1$. Ez valóban elegendő, ugyanis ha tudjuk a tételt a valószínűségi változók $\frac{X_i-\mathbb{E}(X_1)}{\mathbb{D}(X_1)}$ standardizáltjaira (amikre már teljesül, hogy a várható értékük 0, szórásuk 1), akkor ebből a tétel az alábbi átrendezéssel következik:

$$\frac{X_1+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n}\mathbb{D}(X_1)} = \sqrt{n}\frac{\sum_{i=1}^n X_i - \mathbb{E}(X_1)}{n\mathbb{D}(X_1)} = \sqrt{n}\frac{\sum_{i=1}^n \frac{X_i-\mathbb{E}(X_1)}{\mathbb{D}(X_1)}}{n}.$$

Tehát azt kell belátnunk, hogy $\sqrt{n}\cdot\overline{X_n} \xrightarrow{d} Z$. A momentumgeneráló függvényre vonatkozó állítás szerint ehhez elég belátni, hogy

$$\lim_{n\to\infty} M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = M_Z(t) \tag{5}$$

[45]A momentumgeneráló függvény a Laplace-transzformálttal rokon fogalom. A másik sűrűn előkerülő fogalom a témában a karakterisztikus függvény, ami pedig a Fourier-transzformáció valószínűségszámításbeli megfelelője. Fourier-transzformációról nagyszerű magyarázó videó: [youtube - 3blue1brown, Fourier Transform]

minden $t \in \mathbb{R}$ esetén.

Számoljuk ki először az (5) egyenlet jobb oldalát:

$$M_Z(t) = \mathbb{E}(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z-t)^2}{2} + \frac{t^2}{2}} \mathrm{d}z = e^{\frac{t^2}{2}}.$$

Más szavakkal, a standard normális eloszlás momentumgeneráló függvénye $t \mapsto e^{\frac{t^2}{2}}$.

Az (5) egyenlet bal oldalának vizsgálata előtt számoljuk ki az X_1 momentumgeneráló függvényét is:

$$M_{X_1}(t) = \mathbb{E}\Big(\sum_{k=0}^{\infty} \frac{1}{k!} t^k X_1^k\Big) = 1 + t\mathbb{E}X_1 + \frac{t^2}{2}\mathbb{E}(X_1^2) + \frac{t^2}{2} \cdot \mathbb{E}\Big(\sum_{k=3}^{\infty} \frac{2}{k!} t^{k-2} X_1^k\Big) = 1 + \frac{t^2}{2}\big(1 + r(t)\big), \tag{6}$$

ahol $r(t)$ jelöli az utolsó szumma várható értékét. Egyáltalán nem nyilvánvaló, de belátható a következő:

9.3.6. Lemma. $\lim_{t\to 0} r(t) = 0$.

A lemma miatt az (5) egyenlet bal oldaláról a következő mondható:

$$\begin{aligned} M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) &= \mathbb{E}\Big(e^{t\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}\Big) = \mathbb{E}\Big(\prod_{i=1}^n e^{\frac{t}{\sqrt{n}} X_i}\Big) = \\ &= \prod_{i=1}^n \mathbb{E}\Big(e^{\frac{t}{\sqrt{n}} X_i}\Big) = \prod_{i=1}^n M_{X_i}\Big(\frac{t}{\sqrt{n}}\Big) = M_{X_1}\Big(\frac{t}{\sqrt{n}}\Big)^n, \end{aligned} \tag{7}$$

ahol felhasználtuk a következő két tulajdonságot: Egyrészt az $e^{\frac{t}{\sqrt{n}} X_i}$ valószínűségi változók függetlenek, másrészt $M_{X_i}\big(\frac{t}{\sqrt{n}}\big)$ minden i esetén megegyezik $M_{X_1}\big(\frac{t}{\sqrt{n}}\big)$-nel, hiszen a momentumgeneráló függvény csak X_i eloszlásától függ, ami minden i esetén ugyanaz, mint X_1 eloszlása.

Helyettesítsük be az (7) egyenlet logaritmusába az $M_{X_1}\big(\frac{t}{\sqrt{n}}\big)$ (6) egyenlet szerinti értékét:

$$\ln M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = n \ln M_{X_1}\Big(\frac{t}{\sqrt{n}}\Big) = n \ln\left(1 + \frac{\big(\frac{t}{\sqrt{n}}\big)^2}{2}\Big(1 + r\Big(\frac{t}{\sqrt{n}}\Big)\Big)\right).$$

Jelölje $r\big(\frac{t}{\sqrt{n}}\big)$-et r_n. A lemma szerint $r_n \to 0$, ha $n \to \infty$. Ezért az egyenletet a következőképp folytathatjuk:

$$\ln M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = n \ln\Big(1 + \frac{t^2}{2n}\big(1 + r_n\big)\Big) = \frac{\ln\Big(1 + \frac{t^2}{2n}\big(1 + r_n\big)\Big)}{\frac{t^2}{2n}(1 + r_n)} \cdot \frac{t^2}{2} \cdot (1 + r_n).$$

A lemma miatt $\frac{t^2}{2n}(1+r_n) \to 0$, ha $n \to \infty$ (rögzített t esetén). Emiatt a hármas szorzat első tényezője 1-hez tart, felhasználva, hogy $\frac{\ln(1+y)}{y} \to 1$, ha $y \to 0$. Összességében:

$$\lim_{n\to\infty} \ln M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = 1 \cdot \frac{t^2}{2} \cdot (1 + 0) = \frac{t^2}{2}.$$

Erre alkalmazhatjuk az exponenciális függvényt (ami folytonos, így felcserélhető a határértékképzéssel). Tehát $\lim_{n\to\infty} M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t) = e^{\frac{t^2}{2}} = M_Z(t)$ minden $t \in \mathbb{R}$ esetén. □