**A Valószínűségszámítás II. előadássorozat hatodik előadása.**

*2002. október 15.*

**A centrális határeloszlástétel.**

Ebben az előadásban a független valószínűségi változók összegére vonatkozó centrális határeloszlástétellel foglalkozunk. Ezt az eredményt a lehető legáltalánosabb formában fogalmazzuk meg, és tárgyaljuk annak legfontosabb speciális eseteit, melyek feltételeit viszonylag könnyen ellenőrízhetjük. Annak érdekében, hogy ezt a programot végrehajtsuk, be kell vezetnünk néhány fontos fogalmat.

Először a következő fogalmakat vezetjük be:

**Szériasorozatok definiciója.** *A*

$$\begin{array}{l} \xi_{1,1},\dots,\xi_{1,n_1} \\ \vdots \qquad\quad \vdots \\ \xi_{k,1},\dots,\xi_{k,n_k} \\ \vdots \qquad\quad \vdots \end{array}$$

*valószínűségi változók rendszere egy $(\Omega,\mathcal{A},P)$ valószínűségi mezőn, $k\to\infty$, szériasorozat, ha az egy sorban levő $\xi_{k,1}$, …, $\xi_{k,n_k}$ valószínűségi változók függetlenek. (A különböző sorokban levő valószínűségi változók kapcsolatáról nem tételezünk fel semmit.)*

Tekintsük egy $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, szériasorozat egy sorban lévő valószínűségi változóinak $S_k=\sum\limits_{j=1}^{n_k}\xi_{k,j}$ összegeit. Azt fogjuk vizsgálni, hogy az $S_k$ összegek $k\to\infty$ esetben mikor konvergálnak eloszlásban a normális eloszláshoz. A kérdés hátterében az a kép van, hogy a bevezető valószínűségszámítás előadásban tanult centrális határeloszlástétel azt sugallja, hogy nagyon általlanos esetben, ha sok kis független valószínűségi változót összeadunk, akkor az összeg eloszlása ,,majdnem elfelejti" az egyes összeadandók eloszlását, és közel normális eloszlású lesz. Ilyen eredményt azonban csak akkor várhatunk, ha az egyes összeadandók az összeghez képest nagyon kicsik. Ezt a tulajdonságot viszont meg kell fogalmaznunk pontosabban. Ez a célja az alábbi egyenletes kicsiség definiciójának.

**(Normalizált) szériasorozat egyenletes kicsiségének a definiciója.** *Legyen*

$$\begin{array}{l} \xi_{1,1},\dots,\xi_{1,n_1} \\ \vdots \qquad\quad \vdots \\ \xi_{k,1},\dots,\xi_{k,n_k} \\ \vdots \qquad\quad \vdots \end{array}$$

*szériasorozat, melynek tagja teljesítik a* $E\xi_{k,j}=0$*,* $E\xi_{k,j}^2=\sigma_{k,j}^2<\infty$*,* $k=1,2,\dots$*,* $1\le j\le n_k$*, feltételeket, továbbá az* $S_k^2$ *sorösszegek* $ES_k^2=\sum\limits_{j=1}^{n_k}\sigma_{k,j}^2$ *szórásnégyzetei teljesítik a* $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}\sigma_{k,j}^2=1$ *relációt. Azt mondjuk, hogy ez a szériasorozat teljesíti az egyenletes kicsiség feltételét, ha teljesül a*

$$\lim_{k\to\infty}\sup_{1\le j\le n_k}\sigma_{k,j}^2=0$$

*reláció.*

Az egyenletes kicsiség szemléletes tartalma világos. Nulla várható értékű valószínűségi változókat tekintünk, és ezek nagyságát természetes a szórásnégyzettel mérni. Nagy $k$-ra a $k$-ik sorösszeg szórásnégyzete közel 1, és azt kívánjuk, hogy az összegben szereplő változók szórásnégyzetének a maximuma is sokkal kisebb ennél.

Az érdekel minket, hogy az egyenletes kicsiség feltéelét teljesítő (normalizált) szériasorozat $S_k$ sorösszegei mikor teljesítik a centrális határeloszlástételt, azaz mikor konvergálnak eloszlásban a standard normális eloszláshoz. Kimondunk és bebizonyítunk egy tételt, mely szerint egy nagyon általános feltétel, az úgynevezett Lindeberg feltétel teljesülése esetén a centrális határeloszlástétel érvényes. Megfogalmazok egy olyan eredményt is, mely szerint a Lindeberg feltétel nemcsak elégséges, de szükséges feltétele is a centrális határeloszlástételnek. Ennek az eredménynek a bizonyítását azonban csak a (nem kötelező tananyaghoz tartozó) kiegészítésben ismertetem. Tárgyalni fogjuk még azt, hogy a szériasorozatokra kimondott tétel mit mond ki független valószínűségi változók normalizált részletösszegeire. Továbbá megfogalmazom a bebizonyított eredmény néhány következményét, melyek a centrális határeloszlás teljesülését fogalmazzák meg bizonyos a Lindeberg feltételnél szigorúbb, de gyakran könnyebben ellenőrizhető feltétel teljesülése esetén.

Először megfogalmazzuk a Lindeberg feltételt.

**Lindeberg feltétel definiciója szériasorozatokra:** *Legyen* $\xi_{k,j}$*,* $k=1,2,\dots$*,* $1\le j\le n_k$*, olyan szériasorozat, melyre* $E\xi_{k,j}=0$*,* $E\xi_{k,j}^2<\infty$*,* $k=1,2,\dots$*,* $1\le j\le n_k$*, és* $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\xi_{k,j}^2=1$*. Ez a szériasorozat akkor és csak akkor teljesíti a Lindeberg feltételt, ha tetszőleges* $\varepsilon>0$ *számra*

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}E\xi_{k,j}^2 I\left(\{|\xi_{k,j}|>\varepsilon\}\right)=0,$$

*ahol* $I(A)$ *egy* $A$ *halmaz indikátor függvénye.*

**Centrális határeloszlástétel szériasorozatokra a Lindeberg feltétel teljesülése esetén.** *Legyen* $\xi_{k,j}$*,* $k=1,2,\dots$*,* $1\le j\le n_k$*, olyan szériasorozat, melyre* $E\xi_{k,j}=0$*,*

*$E\xi_{k,j}^2 < \infty$, $k = 1, 2, \dots$, $1 \le j \le n_k$, $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} E\xi_{k,j}^2 = 1$, és teljesítse e szériasorozat a Lindeberg feltételt. Ekkor*

*a.) a szériasorozat tagjai teljesítik a $\lim\limits_{k\to\infty} \left( \sup\limits_{1\le j\le n_k} E\xi_{k,j}^2 \right) = 0$ kicsiségi feltételt.*

*b.) Az $S_k = \sum\limits_{j=1}^{n_k} \xi_{k,j}$, $1 \le k < \infty$, véletlen összegek eloszlásban konvergálnak a standard, azaz nulla várható értékű és 1 szórásnégyzetű normális eloszláshoz, ha $k \to \infty$.*

Érvényes az előbbi eredmény következő megfordítása:

**Tétel arról, hogy a Lindeberg feltétel a centrális határeloszlástételnek szükséges feltétele.** *Legyen $\xi_{k,j}$, $k = 1, 2, \dots$, $1 \le j \le n_k$, olyan szériasorozat, melyre $E\xi_{k,j} = 0$, $E\xi_{k,j}^2 < \infty$, $k = 1, 2, \dots$, $1 \le j \le n_k$, $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} E\xi_{k,j}^2 = 1$, és teljesítse ez a széroasorozat a kicsiségi feltételt, azaz legyen $\lim\limits_{k\to\infty} \left( \sup\limits_{1\le j\le n_k} E\xi_{k,j}^2 \right) = 0$. Tegyük fel ezen kívül azt, hogy az $S_k = \sum\limits_{j=1}^{n_k} \xi_{k,j}$, $1 \le k < \infty$, véletlen összegek eloszlásban konvergálnak egy 1 szórásnégyzetű és tetszőleges várható értékű normális normális eloszláshoz, ha $k \to \infty$. Ekkor a $\xi_{k,j}$, $1 \le k < \infty$, $1 \le j \le n_k$, szériasorozat teljesíti a Lindeberg feltételt.*

*Megjegyzés:* Érdemes külön hangsúlyozni, hogy a centrális határeloszlástétel fent megfogalmazott megfordításában nem csak azt követeltük meg, hogy a szériasorozat $S_k = \sum\limits_{j=1}^{n_k} \xi_{k,j}$, $1 \le k < \infty$, részletösszegei a normális eloszláshoz konvergáljanak, hanem azt is, hogy a határeloszlásnak, „a várt szórásnégyzete” legyen, azaz, hogy egy olyan valószínűségi változó szórásnégyzete melynek eloszása megegyezi a tekintett eloszlások határeloszlásával legyen egyenlő a tekintett valószínűségi változók szórásnégyzetének a határértékével. Később fogunk látni példát arra, hogy az előző, a centrális határeloszlástétel megfordításáról szóló tételből ez a feltétel nem hagyható el.

Bár a Lindeberg féle centrális határeloszlástétel fő állítását csak később bizonyítjuk, már most megmutatjuk, hogy a Lindeberg feltételből következik az egyenletes kicsiség feltétele. Ennek érdekében jelöljük $F_{k,j}(\cdot)$-vel a $\xi_{k,j}$ valószínűségi változó eloszlásfüggvényét, és írjuk fel a Lindeberg feltételt az $F_{k,j}$ eloszlásfüggvények segítségével. Vegyük észre, hogy

$$\sum_{j=1}^{n_k} E\xi_{k,j}^2 I\left(\{|\xi_{k,j}| > \varepsilon\}\right) = \sum_{j=1}^{n_k} \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du),$$

és $E\xi_{k,j}^2 = \int_{-\infty}^{\infty} u^2 F_{k,j}(\,du)$. Innen

$$E\xi_{k,j}^2 = \int_{-\varepsilon}^{\varepsilon} u^2 F_{k,j}(\,du) + \int_{|u|\ge-\varepsilon} u^2 F_{k,j}(\,du) \le \varepsilon^2 + \sum_{j=1}^{n_k} \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du)$$

tetszőleges $\varepsilon > 0$ számra. Viszont a Lindeberg feltétel azt jelenti, hogy az utolsó egyenlőtlenség jobboldalán szereplő kifejezés második tagja $k \to \infty$ esetén nullához tart, ezért $\limsup\limits_{k\to\infty} \left( \sup\limits_{1\le j\le n_k} E\xi_{k,j}^2 \right) \le \varepsilon$. Mivel ez az állítás minden $\varepsilon > 0$ számra érvényes, innen következik, hogy a $\xi_{k,j}$, $k = 1, 2, \dots$, $1 \le j \le n_k$ szériasorozat teljesíti az egyenletes kicsiség feltételét.

Most pedig tárgyaljuk meg, mit jelent a Lindeberg féle centrális határeloszlástétel független valószínűségi változók normalizált részletösszegeire. Legyen adva független valószínűségi változók $\xi_1, \xi_2, \dots$, sorozata egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, és tegyük fel, hogy mindegyik $\xi_k$ valószínűségi változónak létezik az $E\xi_k^2 < \infty$ második momentuma. Vezessük be az $E\xi_k = m_k$, $\operatorname{Var} \xi_k = E\xi_k^2 - (E\xi_k)^2 = \sigma_k^2$, $k = 1, 2, \dots$, jelöléseket, valamint legyen $S_k = \sum\limits_{j=1}^{n} \xi_j$, $j = 1, 2, \dots$. Természetes a következő szériasorozatot definiálni, $\xi_{k,j} = \dfrac{\xi_k - E\xi_k}{\sqrt{\operatorname{Var} S_k}}$, $k = 1, 2, \dots$, $1 \le j \le k$. Nyílván, $\operatorname{Var} S_k = \sum\limits_{j=1}^{k} \sigma_k^2$, $\dfrac{S_k - ES_k}{\sqrt{\operatorname{Var} S_k}} = \sum\limits_{j=1}^{k} \xi_{k,j}$. Továbbá az így definiált $\xi_{k,j}$ valószínűségi változók $k = 1, 2, \dots$, $1 \le j \le k$, szériasorozatot alkotnak (rögzített $k$ indexre a $\xi_{k,j}$ valószínűségi változók, $1 \le j \le k$, függetlenek). Tehát, ha a $\xi_k$ valószínűségi változók olyanok, hogy a belőlük készített $\xi_{k,j}$ sorozat teljesíti a Lindeberg féle centrális határeloszlástétel feltételeit, akkor érvényes a $\xi_k$ valószínűségi változók normalizált részletösszegeire a centrális határeloszlástétel. Megfogalmazzuk először a szükséges fogalmakat a $\xi_k$ valószínűségi változók segítségével, majd a centrális határeloszlástétel Lindeberg-féle változatát független valószínűségi változók normalizált részletösszegeire. Végül megfogalmazom annak néhány érdekes következményét, azt, hogy hogyan lehet bebizonyítani a segítségével a centrális határeloszlástételt bizonyos könnyen ellenőrizhető feltétel segítségével.

**Lindeberg feltétel független valószínűségi változók sorozataira:** *Legyen $\xi_n$, $n = 1, 2, \dots$, független valószínűségi változók sorozata, melyre $E\xi_n = 0$, $\sigma_n^2 = E\xi_n^2 < \infty$, $n = 1, 2, \dots$, és az $s_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$, $n = 1, 2, \dots$, sorozat teljesíti a $\lim\limits_{n\to\infty} s_n^2 = \infty$ feltételt. Azt mondjuk, hogy a $\xi_n$, $n = 1, 2, \dots$, sorozat teljesíti a Lindeberg feltételt, ha minden $\varepsilon > 0$ számra*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} E\xi_k^2 I(\{|\xi_k| > \varepsilon s_n\}) = 0.$$

Megfogalmazzuk a centrális határeloszlástételt független valószínűségi változók normalizált összegére, ha ez a sorozat teljesíti a Lindeberg feltételt. Ez az eredmény egyszerű következménye a szériasorozatokra kimondott centrális határeloszlástételnek a Lindeberg feltétel teljesülése esetén valamint annak az előbb leírt egyszerű konstrukciónak, melyben független valószínűségi változók sorozatának segítségével definiáltunk egy szériasorozatot.

**A centrális határeloszlástétel független valószínűségi változók normalizált részletösszegeire a Lindeberg feltétel teljesülése esetén.** *Legyen $\xi_j$, $j = 1, 2, \dots$, független valószínűségi változók sorozata, melyre $E\xi_j^2 < \infty$, $j = 1, 2, \dots$, és a $\xi_j$ valószínűségi változók* $\operatorname{Var}\xi_j = E\xi_j^2 - (E\xi_j)^2$ *szórásnégyzeteire érvényes a* $\lim\limits_{n\to\infty} \sum\limits_{j=1}^{n} \sigma_j^2 = \infty$ *feltétel, és a $\bar\xi_j = \xi_j - E\xi_j$, $j = 1, 2, \dots$, valószínűségi változók teljesítik a Lindeberg feltételt. Ekkor*

*a.) a $\xi_j$, $j = 1, 2, \dots$, valószínűségi változók teljesítik az alábbi kicsiségi feltételt:*

$$\lim_{k\to\infty} \left( \sup_{1\le j\le k} \frac{\operatorname{Var}\xi_j}{\sum\limits_{j=1}^{k} \operatorname{Var}\xi_j} \right) = 0.$$

*b.) Az $S_n = \sum\limits_{k=1}^{n} \xi_k$ véletlen összegek $\bar S_n = \dfrac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}}$, $1 \le n < \infty$, normalizáltjai eloszlásban konvergálnak a standard, azaz nulla várható értékű és 1 szórásnégyzetű normális eloszláshoz, ha $n \to \infty$.*

Az alább megfogalmazott eredmény arról szól, hogy néhány egyszerű és fontos esetben teljesül a Lindeberg feltétel független valószínűségi változók sorozatára, ezért ezek normalizált részletösszegei teljesíti a centrális határeloszlástételt.

**1. Tétel.** *Legyen $\xi_n$, $n = 1, 2, \dots$, független valószínűségi változók sorozata, melyre $E\xi_n = 0$, $\sigma_n^2 = E\xi_n^2 < \infty$, $n = 1, 2, \dots$, $\lim\limits_{n\to\infty} s_n^2 = \infty$, ahol $s_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$. Ez a sorozat teljesíti a Lindeberg feltételt, ha a következő tulajdonságok egyike teljesül.*

*a.) A független $\xi_n$, $n = 1, 2, \dots$, valószínűségi változók egyforma eloszlásúak.*

*b.) $E|\xi_k|^{2+\alpha} < \infty$, minden $k = 1, 2, \dots$ számra valamilyen $\alpha > 0$ konstanssal, és* $\lim\limits_{n\to\infty} \dfrac{\left( \sum\limits_{k=1}^{n} E|\xi_k|^{2+\alpha} \right)^{2/(2+\alpha)}}{s_n^2} = 0$. *Speciálisan, ez a feltétel teljesül akkor, ha $E\xi_k^2 \ge K$ valamilyen $K > 0$ számmal minden $k = 1, 2, \dots$ indexre, és ezenkívül érvényes a $\lim\limits_{k\to\infty} k^{-\alpha/2} E|\xi_k|^{2+\alpha} = 0$ reláció.*

*Az 1. Tétel bizonyítása:* Először az a) részt bizonyítjuk be. Ha $\xi_1, \xi_2, \dots$, független, egyforma eloszlású valószínűségi változók sorozata, $E\xi_1 = 0$, $0 < E\xi_1^2 < \infty$, akkor $\dfrac{1}{s_n^2} \sum\limits_{k=1}^{n} E\xi_k^2 I(|\xi_k| > \varepsilon s_n) = \dfrac{1}{E\xi_1^2} E\xi_1^2 I\left(|\xi_1| > \varepsilon\sqrt{nE\xi_1^2}\right) \to 0$, ha $n \to \infty$, mert az $A_n = \{\omega\colon |\xi_1(\omega)| > \varepsilon\sqrt{nE\xi_1^2}\}$, halmazok egymásba skatulyázottak, és metszetük az üres halmaz, azaz $A_1 \supset A_2 \supset A_3 \supset \cdots$, és $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$.

A b) állítás bizonyítása érdekében vegyük észre, hogy a Hölder egyenlőtlenség alapján

$$\begin{aligned}\sum_{k=1}^n E\xi_k^2 I(|\xi_k|>\varepsilon s_n) &\le \sum_{k=1}^n \left(E|\xi_k|^{(2+\alpha)}\right)^{(2/\alpha+2)} P(|\xi_k|>\varepsilon s_n)^{\alpha/(2+\alpha)}\\ &\le \left(\sum_{k=1}^n E|\xi_k|^{(2+\alpha)}\right)^{(2/\alpha+2)} \left(\sum_{k=1}^n P(|\xi_k|>\varepsilon s_n)\right)^{\alpha/(2+\alpha)}.\end{aligned}$$

(A Hölder egyenlőtlenség azt állítja, hogy véve egy $(X,\mathcal{X})$ téren definiált $f$ és $g$ függvény szorzatának egy az ezen a téren definiált $\mu$ mérték szerinti (Lebesgue integrálját) e kifejezés teljesíti az $\left|\int f(\omega)g(\omega)\,d\mu(\omega)\right| \le \left(\int |f(\omega)|^p\,d\mu(\omega)\right)^{1/p}\left(\int |g(\omega)|^q\,d\mu(\omega)\right)^{1/q}$ minden olyan $(p,q)$ számpárra, melyre $p>0$, $1>0$ és $\frac{1}{p}+\frac{1}{q}=1$. Jelen esetben ezt a felírt egyenlőtlenségsor első egyenlőtlenségénak a bizonyításában az $(\Omega,\mathcal{A})$ téren a $P$ mérték szerint integrálunk, és az $f(\omega)=\xi_k(\omega)^2$ és $g(\omega)=I(|\xi_k(\omega)|>\varepsilon s_n)$ függvényekre alkalmazzuk a Hölder egyenlőtlenséget minden $k=1,\dots,n$ indexre $p=\frac{2+\alpha}{2}$ és $q=\frac{2+\alpha}{\alpha}$ választással. A második egyenlőtlenség bizonyításában a Hölder egyenlőtlenségnek a következő speciális esetét alkalmazzuk: $\left|\sum\limits_{k=1}^n A_kB_k\right| \le \left(\sum\limits_{k=1}^n |A_k|^p\right)^{1/p}\left(\sum\limits_{k=1}^n |B_k|^q\right)^{1/q}$, ha $p>0$ és $q>0$, $\frac{1}{p}+\frac{1}{q}=1$. (Ezt az egyenlőtlenséget állítja a Hölder egyenlőtlenség abban a speciális esetben, ha az $X=\{1,\dots,n\}$ véges halmaz részhalmazain bevezetjük azt a mértéket, amelyik minden halmazhoz annak elemszámát rendeli mértékként.) Most az $A_k=\left(E|\xi_k|^{(2+\alpha)}\right)^{(2/\alpha+2)}$, $B_k=P(|\xi_k|>\varepsilon s_n)^{\alpha/(2+\alpha)}$, $p=\frac{2+\alpha}{2}$ és $q=\frac{2+\alpha}{\alpha}$ választással alkalmazzuk a Hölder egyenlőtlenséget.)

Másrészt, a Csebisev egyenlőtlenség alapján $\sum\limits_{k=1}^n P(|\xi_k|>\varepsilon s_n) \le \sum\limits_{k=1}^n \frac{E\xi_k^2}{\varepsilon^2 s_n^2} = \frac{1}{\varepsilon^2}$. Ezért ha teljesülnek a 1. Tétel b) részének a feltételei, akkor

$$\limsup_{n\to\infty} 1\frac{1}{s_n^2}\sum_{k=1}^n E\xi_k^2 I(|\xi_k|>\varepsilon s_n) \le \limsup_{n\to\infty} \frac{\left(\sum\limits_{k=1}^n E|\xi_k|^{(2+\alpha)}\right)^{(2/\alpha+2)}}{s_n^2}\left(\frac{1}{\varepsilon^2}\right)^{\alpha/(2+\alpha)} = 0$$

minden $\varepsilon>0$ számra, azaz teljesül a Lindeberg feltétel.

Ha teljesülnek a b.) rész végén említett feltételek, akkor $s_n^2 \ge \text{const.}\, n$, és

$$\frac{\sum\limits_{k=1}^n E|\xi_k|^{2+\alpha}}{n^{(\alpha+2)/2}} \to 0, \qquad \text{ha } n\to\infty.$$

(Itt azt használtuk ki, hogy $\sum\limits_{k=1}^n k^{-\alpha/2} \ge \text{const.}\, n^{1+\alpha/2}$.) Ezért ekkor teljesülnek a b.) részben megfogalmazott feltételek.

**A centrális határeloszlástétel bizonyítása szériasorzatokra a Lindeberg feltétel teljesülése esetén.**

A bizonyítás gondolatmenete a következő. Legyen adva valamely $\xi_{k,j}$, $E\xi_{k,j}=0$, $k=1,2,\dots$, $1\le j\le n_k$, szériasorozat, amelyik teljesíti a Lindeberg feltételt, és $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\xi_{k,j}^2=1$. Ahhoz, hogy belássuk a centrális határeloszlástételt, elég megmutatni azt, hogy az $S_k=\sum\limits_{j=1}^{n_k}\xi_{k,j}$ sorösszegek karakterisztikus függvényei teljesítik a $\lim\limits_{k\to\infty}Ee^{itS_k}=e^{-t^2/2}$ relációt minden $t$ valós számra. Viszont tudjuk, hogy $Ee^{itS_k}=\prod\limits_{j=1}^{n_k}Ee^{it\xi_{k,j}}$. Ezért elég megmutatni, hogy a $\xi_{k,j}$ valószínűségi változók karakterisztikus függvényére megengedett az $Ee^{it\xi_{k,j}}\sim e^{-t^2E\xi_{k,j}^2/2}$ helyettesítés. Ez ugyanis azt jelenti, hogy

$$Ee^{itS_k}=\prod_{j=1}^{n_k}Ee^{it\xi_{k,j}}\sim\prod_{j=1}^{n_k}e^{-t^2E\xi_{k,j}^2/2}=\exp\left\{-\frac{t^2}{2}\sum_{j=1}^{n_k}E\xi_{k,j}^2\right\}\sim e^{-t^2/2}.$$

A $\xi_{k,j}$ valószínűségi változó karakterisztikus függvényére alkalmazott közelítés azért látszik elfogadhatónak, mert tudjuk, hogy ez a valószínűségi változó kicsi, így bízhatunk abban, hogy a $\varphi_{k,j}(t)=Ee^{it\xi_{j,k}}$ függvényt jól közelíti e függvény nulla körüli Taylor sorának első két tagból álló szelete. Viszont egy valószínűségi változó karakterisztikus függvényének deriváltjait a nulla helyen meghatározzák a valószínűségi változó momentumai. Ezt a meghatározást pontosan megadtuk a harmadik előadásban. Ezt felhasználva azt kapjuk, hogy $Ee^{it\xi_{k,j}}\sim 1+itE\xi_{k,j}-\frac{t^2}{2}E\xi_{k,j}^2=1-\frac{t^2}{2}E\xi_{k,j}^2$. Ez azt sugallja, hogy $Ee^{it\xi_{k,j}}\sim e^{E\xi_{k,j}^2t^2/2}$. Egy precíz bizonyításban a fenti heurisztikus meggondolásokat pontosítani kell, és azokat be kell bizonyítani. Ennek érdekében először egy olyan egyszerű, de hasznos lemmát bizonyítunk, amelyik becslést ad arra, hogy az $e^{it}$ függvényt milyen jól közelítik Taylor sorának szeletei.

**Lemma.** *Tetszőleges nem negatív egész $k$ és valós $t$ számra*

$$\left|e^{it}-\left(1+\frac{it}{1!}+\cdots+\frac{(it)^k}{k!}\right)\right|\le\frac{|t|^{k+1}}{(k+1)!}.$$

*Bizonyítás:* Vezessük be az $F(t)=e^{it}-\left(1+\frac{it}{1!}+\cdots+\frac{(it)^k}{k!}\right)$ függvényt, és tekintsük ennek $F^{(j)}(t)$ deriváltjait. $F^{(j)}(0)=0$, ha $0\le j\le k$, és $|F^{(k+1)}(t)|=|e^{ikt}|=1$ minden $t\in R^1$ számra. Teljes indukcióval kapjuk, hogy $|F^{(j)}(t)|\le\int_0^t|F^{(j+1)}(s)|\,ds\le\int_0^t\frac{|s|^{k-j}\,ds}{(k-j)!}=\frac{|t|^{k+1-j}}{(k+1-j)!}$ minden $j=k+1,k,\dots,0$ számra. Tehát $|F(t)|\le\frac{|t|^{k+1}}{(k+1)!}$, és ez a Lemma állítása.

Most rátérünk a Tétel bizonyítására. Jelölje a $\xi_{k,j}$ valószínűségi változó karakterisztikus függvényét $\varphi_{k,j}(t) = Ee^{it\xi_{k,j}}$, $-\infty < t < \infty$, eloszláfüggvényét, $F_{k,j}(x) = P(\xi_{k,j} < x)$ szórásnégyzetét $\sigma^2_{k,j} = E\xi^2_{k,j}$, $k = 1, 2, \dots$, $1 \le j \le n_k$. Emlékeztetek arra, hogy fetettük, hogy $E\xi_{k,j} = 0$. Jeleztük, hogy érdemes a $\varphi_{k,j}(t)$ karakterisztikus függvényt az $1 - \dfrac{\sigma^2_{k,j}t^2}{2}$ függvénnyel közelíteni. Annak érdekében, hogy a Tételt bebizonyítsuk érdemes jó becslést adni ennek a közelítés jóságára. Ebben segít az előbb tárgyalt lemma. Ugyanis felírhatjuk, hogy

$$\begin{aligned} \varphi_{k,j}(t) - \left(1 - \frac{\sigma^2_{k,j}t^2}{2}\right) &= \varphi_{k,j}(t) - \left(1 + iE\xi_{k,j}t - \frac{E\xi^2_{k,j}}{2}t^2\right) \\ &= \int_{-\infty}^{\infty} \left(e^{itu} - \left(1 + iut - \frac{(ut)^2}{2}\right)\right) F_{k,j}(\,du) \end{aligned} \tag{1}$$

Az (1) formula jobboldalán szereplő kifejezést jól tudjuk becsülni az előző Lemma segítségével. Nevezetesen felírhatjuk, hogy $\left|e^{itu} - \left(1 + iut - \dfrac{(ut)^2}{2}\right)\right| \le \dfrac{(tu)^3}{6}$. De ugyancsak felírhatjuk a következő durvábbnak látszó becslést is:

$$\left|e^{itu} - \left(1 + iut - \frac{(ut)^2}{2}\right)\right| \le \left|e^{itu} - (1 + iut)\right| + \frac{(ut)^2}{2} \le (ut)^2.$$

Melyiket érdemes alkalmazni a két becslés közül? Az első becslés akkor valójában csak akkor hasznos, ha $|ut|$ kicsi, nagy $ut$ esetén érdemesebb a második becslést alkalmazni. (Ez a körülmény azzal függ össze, hogy egy függvény Taylor-sorának egy véges szelete csak azon pont környezetében ad jó becslést, amely körül a függvényt sorba fejtettük.) Mi a következő eljárást fogjuk követni: Rögzítünk egy $\varepsilon > 0$ számot, és $|u| \le \varepsilon$ esetben az első, $|u| > \varepsilon$ esetben a második becslést alkalmazzuk. Jegyezzük meg, hogy mint láttuk a Lindeberg feltétel teljesüléséből következik az egyenletes kicsiség feltétele, azaz minden $\varepsilon > 0$ számhoz létezik olyan $k = k(\varepsilon)$ küszöbindex, melyre $E\xi^2_{k,j} < \varepsilon$, ha $k \ge k(\varepsilon)$, ahonnan a Csebisev egyenlőtlenség szerint $P(\xi_{k,j} > \varepsilon) \le \varepsilon$, azaz $F_{k,j}(-\varepsilon) + (1 - F_{k,j}(\varepsilon) < \varepsilon$, ami azt jelenti, hogy az $F_{k,j}$ mérték szerint nagy halmazon az első becslést fogjuk alkalmazni. Az elmondottak alapján, az (1) formulából azt kapjuk, hogy

$$\begin{aligned} \left|\varphi_{k,j}(t) - \left(1 - \frac{\sigma^2_{k,j}t^2}{2}\right)\right| &\le \int_{-\infty}^{\infty} \left|e^{itu} - \left(1 + iut - \frac{(ut)^2}{2}\right)\right| F_{k,j}(\,du) \\ &\le \int_{-\varepsilon}^{\varepsilon} \frac{|ut|^3}{6} F_{k,j}(\,du) + \int_{|u|>\varepsilon} (ut)^2 F_{k,j}(\,du) \\ &\le \frac{\varepsilon|t|^3}{6} \int_{-\varepsilon}^{\varepsilon} u^2 F_{k,j}(\,du) + t^2 \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du) \le \varepsilon \frac{|t|^3}{6} \sigma^2_{k,j} + t^2 \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du) \end{aligned} \tag{2}$$

Jegyezzük meg, hogy az utolsó becslés jobboldalán levő két tag közül az első a benne szereplő $\varepsilon$ tényező miatt lesz kicsi (az $\varepsilon$ paramétert kicsinek fogjuk választani), a második pedig a Lindeberg feltétel teljesülése miatt. A most kapott becslés segítségével jó becslést akarunk kapni a $\log \varphi_{k,j}(t) + \dfrac{\sigma_{k,j}^2 t^2}{2}$ kifejezés abszolut értékére elég nagy $k$ számra minden $1 \le j \le n_k$ esetében. Megmutatjuk, hogy minden $\varepsilon > 0$ és $t$ számra

$$\left| \log \varphi_{k,j}(t) + \frac{\sigma_{k,j}^2 t^2}{2} \right| \le \varepsilon |t|^3 \sigma_{k,j}^2 + 2t^2 \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du), \qquad \text{ha } k \ge k_0(\varepsilon, t), \text{ és } 1 \le j \le n_k \tag{3}$$

Ezt a becslést a (2) képlet és a $|\log(1+z) - z| \le \dfrac{z^2}{10}$, egyenlőtlenség segítségével mutatjuk meg, $z = \log \varphi_{k,j}(t) - 1$ választással. (Ez az egyenlőtlenség a $\log(1+z)$ függvényt kifejező Taylor sor alakjából egyszerűen következik. Nem törekedtünk pontos konstansokat írni a becslésben. Vegyük észre, hogy ez az egyenlőtlenség nem csak valós, hanem komplex számokra is érvényes.) Válasszuk az $\varepsilon > 0$ számot olyan kicsinek, hogy $\varepsilon |t|^3 \le \frac{1}{1000}$, és legyen $k_0$ olyan nagy, hogy $k \ge k_0$ és tetszőleges $1 \le j \le n_k$ esetén $\sigma_{k,j}^2 \le \frac{1}{1000}$. (Az egyenletes kicsiség tulajdonsága miatt ezt megtehetjük.) Ekkor alkalmazhatjuk az $|\log(1+z) - z|$ kifejezésre felírt becslést, és azt kapjuk, hogy

$$|\log \varphi_{k,j}(t) - (\varphi_{k,j}(t) - 1)| \le \frac{1}{10} (\varphi_{k,j}(t) - 1)^2$$

Ezért a (2) formula alapján

$$\left| \log \varphi_{k,j}(t) + \frac{\sigma_{k,j} t^2}{2} \right| \le \varepsilon \frac{|t|^3}{6} \sigma_{k,j}^2 + t^2 \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du) + \frac{1}{10} (\varphi_{k,j}(t) - 1)^2.$$

A (3) formula bizonyítása az utolsó becsléből viszonylag egyszerűen kiolvasható, azt kell észrevenni, hogy a becslés jobboldalán szereplő $\frac{1}{10}(\varphi_{k,j}(t) - 1)^2$ tag „másodrendben kicsi”. Pontosabban, a (2) formula alapján

$$\begin{aligned} (\varphi_{k,j}(t) - 1)^2 &\le \left( \frac{\sigma_{k,j}^2 t^2}{2} + \varepsilon \frac{|t|^3}{6} \sigma_{k,j}^2 + t^2 \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du) \right)^2 \\ &\le 3 \frac{\sigma_{k,j}^4 t^4}{4} + 3\varepsilon^2 \left( 2 \frac{|t|^3}{6} \sigma_{k,j}^2 \right)^2 + 3t^4 \left( \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du) \right)^2 \le \varepsilon |t|^3 \sigma_{k,j}^2, \end{aligned}$$

ha $k \ge k_0$ elég nagy $k_0 = k_0(\varepsilon, t)$ küszöbindex-szel, mivel ekkor $\int_{|u|>\varepsilon} u^2 F_{k,j}(\,du) \le \sigma_{k,j}^2$ $\sigma_{k,j}^2$ és elég kicsi. A fenti becslésekből adódik a (3) formula.

Vezessük be a

$$\delta(\varepsilon, k, j, t) = \varepsilon |t|^3 \sigma_{k,j}^2 + 2t^2 \int_{|u|>\varepsilon} u^2 F_{k,j}(\,du)$$

kifejezést. Ekkor a (3) formula alapján

$$e^{-\delta(\varepsilon,k,j,t)} \le \varphi_{k,j}(t) e^{\sigma_{k,j}^2 t^2/2} \le e^{\delta(\varepsilon,k,j,t)} \quad \text{ha } k \ge k_0(\varepsilon,t),\ 1 \le j \le n_k.$$

Innen

$$\exp\left\{-\sum_{j=1}^{n_k} \delta(\varepsilon,k,j,t)\right\} \le \prod_{j=1}^{n_k} \varphi_{k,j}(t) \exp\left\{\frac{t^2}{2}\sum_{j=1}^{n_k}\sigma_{k,j}^2\right\} \le \exp\left\{\sum_{j=1}^{n_k}\delta(\varepsilon,k,j,t)\right\},$$

ha $k \ge k_0(\varepsilon,t)$.

A Lindeberg feltétel teljesüléséből, és a $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} \sigma_{k,j}^2 = 1$ relációból következik, hogy $\limsup\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} \delta(\varepsilon,k,j,t) \le \varepsilon|t|^3$, ezért az utolsó becslésből kapjuk, hogy

$$\begin{aligned} e^{-\varepsilon|t|^3} &\le \liminf_{k\to\infty}\left(\prod_{j=1}^{n_k}\varphi_{k,j}(t)\exp\left\{\frac{t^2}{2}\sum_{j=1}^{n_k}\sigma_{k,j}^2\right\}\right) \\ &\le \limsup_{k\to\infty}\left(\prod_{j=1}^{n_k}\varphi_{k,j}(t)\exp\left\{\frac{t^2}{2}\sum_{j=1}^{n_k}\sigma_{k,j}^2\right\}\right) \le e^{\varepsilon|t|^3}. \end{aligned}$$

Mivel ez az egyenlőtlenség minden $\varepsilon > 0$ számra teljesül, ezért

$$\lim_{k\to\infty}\left(\prod_{j=1}^{n_k}\varphi_{k,j}(t)\exp\left\{\frac{t^2}{2}\sum_{j=1}^{n_k}\sigma_{k,j}^2\right\}\right) = 1,$$

ami viszont a $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}\sigma_{k,j}^2 = 1$ reláció miatt azt jelenti, hogy $\lim\limits_{k\to\infty}\prod\limits_{j=1}^{n_k}\varphi_{k,j}(t)e^{t^2/2} = 1$, azaz $\lim\limits_{k\to\infty} Ee^{itS_k} = \lim\limits_{k\to\infty}\prod\limits_{j=1}^{n_k}\varphi_{k,j}(t) = e^{-t^2/2}$, és ezt kellett bizonyítani.

Kiegészítés: *Annak a tételnek a bizonyítása, mely szerint a Lindeberg feltétel a centrális határeloszlástételnek szükséges feltétele.*

Először felidézem a bizonyítandó eredményt.

**Tétel.** *Legyen $\xi_{k,j}$, $k = 1,2,\dots$, $1 \le j \le n_k$, olyan szériasorozat, melyre $E\xi_{k,j} = 0$, $E\xi_{k,j}^2 < \infty$, $k = 1,2,\dots$, $1 \le j \le n_k$, $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k} E\xi_{k,j}^2 = 1$, és teljesítse ez a széroasorozat a kicsiségi feltételt, azaz legyen $\lim\limits_{k\to\infty}\left(\sup\limits_{1\le j\le n_k} E\xi_{k,j}^2\right) = 0$. Tegyük fel ezen kívül azt,* *hogy az* $S_k = \sum_{j=1}^{n_k} \xi_{k,j}$*,* $1 \le k < \infty$*, véletlen összegek eloszlásban konvergálnak egy 1 szórásnégyzetű és tetszőleges várható értékű normális normális eloszláshoz, ha* $k \to \infty$*. Ekkor a* $\xi_{k,j}$*,* $1 \le k < \infty$*,* $1 \le j \le n_k$*, szériasorozat teljesíti a Lindeberg feltételt.*

*Bizonyítás:* Jelölje $\varphi_{j,k}(t) = Ee^{it\xi_{k,j}}$ a $\xi_{k,j}$ valószínűségi változó karakterisztikus függvényét, és $F_{k,j}(x)$ az eloszlásfüggvényét. Az a tény, hogy az $S_k$ valószínűségi változók eloszlásban konvergálnak egy egy szórású és valamilyen $m$ várható értékű normális eloszlású valószínűségi változóhoz a karakterisztikus függvények nyelvén úgy fejezhető ki, hogy

$$\lim_{k\to\infty} Ee^{itS_k} = \lim_{k\to\infty} \prod_{j=1}^{n_k} \varphi_{k,j}(t) = e^{-t^2/2+imt} \tag{A1}$$

minden $-\infty < t < \infty$ számra. A konvergencia egyenletes minden véges intervallumban.

Vegyünk egy véges $[a,b]$ intervallumot, és definiáljuk a $\varphi_{k,j}(t)$ függvény logaritmusát. Bár a logaritmus függvény a komplex számsíkon többértékű, (egy $z = re^{i\varphi}$ komplex szám logaritmusa a $\log z = \log r + i\varphi + i2k\pi$ számok valamelyike, ahol $k$ tetszőleges egész szám lehet) nagy $k$ számra az egyenletes kicsiség feltétele alapján a $\log \varphi_{k,j}(t)$ szám természetes módon definiálható az $[a,b]$ intervallumban, és megmutatjuk, hogy teljesül egy számunkra hasznos, a későbbiekben (A2)-vel jelölt reláció.

E program végrehajtása érdekében vegyük észre, hogy az egyenletes kicsiség miatt minden $T > 0$ számra alkalmas $k_0 = k_0(T)$ küszöbindex esetén $|\varphi_{k,j}(t) - 1| \le \frac{1}{4}$. (Válasszuk a $T = \max(|a|,|b|)$ számot.) Valóban,

$$\begin{aligned} |\varphi_{k,j}(t) - 1| = |\varphi_{k,j}(t) - 1 - Eit\xi_{k,j}| &= \left| \int (e^{itu} - 1 - itu) F_{k,j}(\,du) \right| \\ &\le \int \frac{t^2u^2}{2} F_{k,j}(\,du) = \frac{t^2}{2} E\xi_{k,j}^2 \le \frac{1}{4}, \quad \text{ha } k \ge k_0, \text{ és } |t| \le T. \end{aligned}$$

Ezután a logaritmust úgy definiálhatjuk, a $\{z\colon |z-1| \le \frac{1}{4}\}$ tartományban, mint a logaritmusfüggvény azon analitikus ágát, melyre $\log 1 = 0$.

Ezzel a választással a $|1 - \varphi_{k,j}(t)| \le \frac{1}{4}$ és $\log(1+z) - z| \le |z|^2$, ha $|z| \le \frac{1}{4}$ egyenlőtlenségek miatt

$$\begin{aligned} |\log \varphi_{k,j}(t) + (1 - \varphi_{k,j}(t))| &= |\log(1 - (1 - \varphi_{k,j}(t))) + (1 - \varphi_{k,j}(t))| \\ &\le |1 - \varphi_{k,j}(t)|^2 \le t^4 \left(E\xi^2\right)^2. \end{aligned}$$

Innen

$$\left| \sum_{j=1}^{n_k} \log \varphi_{k,j}(t) - \sum_{j=1}^{n_k} (\varphi_{k,j}(t) - 1) \right| \le t^4 \sum_{j=1}^{n_k} \left(E\xi_{k,j}^2\right)^2 \le \text{const.}\, t^4 \max_{1\le j\le n_k} E\xi_{k,j}^2,$$

mert $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k} E\xi_{j,k}^2=1$. Innen, mivel az az egyenletes kicsiség feltétele miatt teljesül a $\lim\limits_{k\to\infty}\max\limits_{1\le j\le n_k} E\xi_{k,j}^2=0$ reláció, érvényes a

$$\lim_{k\to\infty}\left|\sum_{j=1}^{n_k}\log\varphi_{k,j}(t)-\sum_{j=1}^{n_k}(\varphi_{k,j}(t)-1)\right|=0 \tag{A2}$$

becslés.

Tekintsük az (A1) azonosság két oldalán szereplő kifejezések abszolut értékét, és vegyük ezek logaritmusát. Felhasználva, hogy egy $z$ komplex szám logaritmusára érvényes a $\log|z|=\operatorname{Re}\log z$ az (A1) azonosságból azt kapjuk, hogy

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}\operatorname{Re}(\log\varphi_{k,j}(t))=-\frac{t^2}{2} \tag{A3}$$

Az (A2) reláció alapján

$$\begin{aligned}0&=\lim_{k\to\infty}\left|\operatorname{Re}\left(\sum_{j=1}^{n_k}\log\varphi_{k,j}(t)-\sum_{j=1}^{n_k}(\varphi_{k,j}(t)-1)\right)\right|\\&=\lim_{k\to\infty}\left|\sum_{j=1}^{n_k}\operatorname{Re}(\log\varphi_{k,j}(t))-\sum_{j=1}^{n_k}\operatorname{Re}(\varphi_{k,j}(t)-1)\right|.\end{aligned}$$

Innen és az (A3) formulából kapjuk, hogy

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}\operatorname{Re}(\varphi_{k,j}(t)-1)=-\frac{t^2}{2}.$$

Továbbá, mivel nagy $k$ indexre a szériasorozat sorainak az összege közel egy szórásnégyzetű, ezért innen következik, hogy

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}E\left(\cos(t\xi_{k,j})-1+\frac{t^2\xi_{k,j}^2}{2}\right)=0 \quad \text{minden } t\in R^1 \text{ számra.}$$

Vegyük észre, hogy $\cos u-1+\dfrac{u^2}{2}\ge 0$ minden $u\in R^1$ számra, ugyanis az $F(u)=\cos u-1+\dfrac{u^2}{2}$ függvényre $F(0)=F'(0)=0$, és $F''(u)=1-\cos u\ge 0$ minden $u$ számra. Továbbá $\cos u-1+\dfrac{u^2}{2}\ge\dfrac{u^2}{4}$, ha $|u|>3$. Ezért a fenti egyenlőtlenségből következik, hogy

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}\frac{t^2}{4}E\xi_{k,j}^2 I\left(\left\{|\xi_{k,j}|\ge\frac{3}{t}\right\}\right)=0.$$

$t=\dfrac{3}{\varepsilon}$ választással megkapjuk a tétel állítását.