# Valószínűségszámítás előadás informatika BSC/A szakosoknak

2019/2020 2. félév

Zempléni András

zempleni@caesar.elte.hu

http://zempleni.elte.hu

# Generátorfüggvény

- Legyen X nemnegatív egész értékű valószínűségi változó. A generátorfüggvénye $g_X(z) := E(z^X) = P(X=0) + zP(X=1) + z^2P(X=2) + \ldots$
- Tulajdonságai:
  - Véges, ha $|z| \leq 1$
  - Meghatározza az X eloszlását:
    - $P(X=0) = g_X(0)$
    - $P(X=1) = g_X'(0)$
    - $P(X=2) = g_X''(0)/2$ stb.
  - Ha X és Y függetlenek, nemnegatív egész értékűek: $g_{X+Y}(z) = g_X(z)g_Y(z)$, mert $E(z^{X+Y}) = E(z^X z^Y) = E(z^X)\,E(z^Y)$ a függetlenség miatt.

# Példák

- Elfajult eloszlásra ($P(X=k)=1$): $g_X(z)= z^k$.
- Indikátorváltozóra $g_X(z)= pz+1-p$
- Binomiális eloszlásra $g_X(z)=(pz+1-p)^n$
- Poisson eloszlásra

$$E(z^X) = \sum_{k=0}^{\infty} z^k \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}$$

# Karakterisztikus függvény

- Komplex értékű valószínűségi változók: $Z=X+iY$, ahol X és Y is valószínűségi változók.
- $E(Z):=E(X)+iE(Y)$.
- X (valós) valószínűségi változó karakterisztikus függvénye: $\varphi_X(t):=E(e^{itX})=E(\cos tX)+iE(\sin tX)$
- Tulajdonságai:
  - $\varphi_X(t)$ $\boldsymbol{R}\to\boldsymbol{C}$ függvény, mely minden X-re létezik.
  - $\varphi_X(0)=1$ minden X-re
  - $|\varphi_X(t)|\leq 1$ (mert $|E(e^{itX})| \leq E(|e^{itX}|)=1$
  - Ha X és Y függetlenek, $\varphi_{X+Y}(t)= \varphi_X(t)\varphi_Y(t)$, mert $E(e^{it(X+Y)})= E(e^{itX}e^{itY})= E(e^{itX})\, E(e^{itY})$ a függetlenség miatt.

# További tulajdonságok

- Ha Y=aX+b, akkor $\varphi_Y(t) = e^{ibt}\varphi_X(at)$
- Bizonyítás: $\varphi_Y(t) = E(e^{i(aX+b)t}) = e^{ibt}\, E(e^{iaXt}) = e^{ibt}\varphi_X(at)$.
- Kiszámítása az abszolút folytonos esetre:

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx = \int_{-\infty}^{\infty} \cos(tx) f(x)dx + i\int_{-\infty}^{\infty} \sin(tx) f(x)dx$$

- Példa. Ha X egyenletes eloszlású a [-1,1] intervallumon, akkor

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx = \int_{-1}^{1} \frac{\cos(tx)}{2} dx = \left[\frac{\sin tx}{2t}\right]_{-1}^{1} = \frac{\sin t}{t}$$

- Általában is: $\varphi_X$ valós, ha X eloszlása szimmetrikus a 0-ra.

# További tulajdonságok/2

- A standard normális eloszlás karakterisztikus függvénye:

$$\varphi(t) = e^{-\frac{t^2}{2}}$$

- A karakterisztikus függvény meghatározza az eloszlást (azaz különböző eloszlásokhoz különböző karakterisztikus függvény tartozik).
- Taylor sorfejtés: tegyük fel, hogy $E(X^n)$ véges valamilyen $n \geq 1$ egész számra. Ekkor $t \to 0$ mellett

$$\varphi_X(t) = 1 + \frac{it}{1!}E(X) + \frac{(it)^2}{2!}E(X^2) + \ldots + \frac{(it)^n}{n!}E(X^n) + o(t^n)$$

ahol $o(t^n)$ jelentése, hogy $t^n$–nel osztva is 0-hoz tart, ha $t \to 0$.

# További tételek

Azaz $\varphi_X^{(k)}(0)= i^k E(X^k)$ és így a szokásos Taylor-sorfejtésből adódik a tétel.

- Ha $\varphi$ karakterisztikus függvény, akkor egyenletesen folytonos.
- *Folytonossági tétel.* Legyen $\varphi_n$ karakterisztikus függvények egy sorozata (jelölje $Q_n$ a hozzá tartozó eloszlást). Ha $\varphi_n$ pontonként konvergál egy $\varphi$-hez, mely a 0-ban folytonos, akkor $\varphi$ is karakterisztikus függvény, és a hozzá tartozó eloszlás éppen a eloszlások Q gyenge határértéke.

# Centrális határeloszlás tétel

- Legyenek $X_1$ , $X_2$ ,…, $X_n$ ,... független, azonos eloszlású valószínűségi változók. Tegyük fel, hogy $\sigma^2=D^2(X)$ véges ($m:=E(X_i)$). Tekintsük a standardizált összegüket:

$$Z_n := \frac{X_1 + \ldots + X_n - nm}{\sqrt{n}\sigma}$$

Ekkor $Z_n$ gyengén konvergál a standard normális eloszláshoz, azaz

$$P\left( \frac{X_1 + \ldots + X_n - nm}{\sqrt{n}\sigma} < z \right) \to \Phi(z)$$

ahol $\Phi$ a standard normális eloszlás eloszlásfüggvénye.

# Bizonyítás vázlata

- A folytonossági tétel miatt elegendő a $Z_n$ karakterisztikus függvényére belátni, hogy $\varphi_n(t) \to \exp\{-t^2/2\}$.
- Ha $\psi(t)$ jelöli az $X_n$-$m$ karakterisztikus függvényét, akkor $X_1+X_2+\ldots+X_n$-$nm$ karakterisztikus függvénye $\psi^n(t)$. Ebből

$$\varphi_n(t) = \left(\psi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

- A maradéktagos Taylor formula miatt

$$\psi(t) = 1 + it\frac{E(X-m)}{1!} + i^2t^2\frac{E(X-m)^2}{2!} + o(t^2) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2)$$

# Bizonyítás befejezése

$$\varphi_n(t) = \left(\psi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)^n = \left(1 - \frac{t^2}{2n} + \frac{1}{n}o(1)\right)^n \to e^{-\frac{t^2}{2}}$$

Megjegyzés. A tételt tulajdonképpen már megfigyeltük a szimulációknál.

# A nem azonos eloszlású eset

- Ekkor – a nagy számok törvényénél már látott okok miatt – erősebb feltételek kellenek.
- A legegyszerűbb eset: ha $X_1$, $X_2$ ,…, $X_n$ ,… független, egyenletesen korlátos valószínűségi változók (ekkor $\sigma_i^2=D^2(X_i)$ véges, $m_i:=E(X_i)$), akkor a standardizált összegük:

$$Z_n := \frac{X_1+\ldots+X_n-(m_1+\ldots+m_n)}{\sqrt{\sigma_1^{\;2}+\ldots+\sigma_n^{\;2}}}$$

Ha $\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2\to\infty$ akkor $Z_n$ gyengén konvergál a standard normális eloszláshoz, azaz

$$P\left(Z_n < z\right)\to\Phi(z)$$

ahol $\Phi$ a standard normális eloszlás eloszlásfüggvénye.

# Általánosítások

- Ha nem korlátosak a tagok, további feltételekre (pl. magasabb momentumok létezése, hasonló nagyságrendű összeadandók) van szükség.
- Gyenge, a Bernstein tételben látott összefüggőség esetére is általánosítható a tétel.

# Konvergenciasebesség

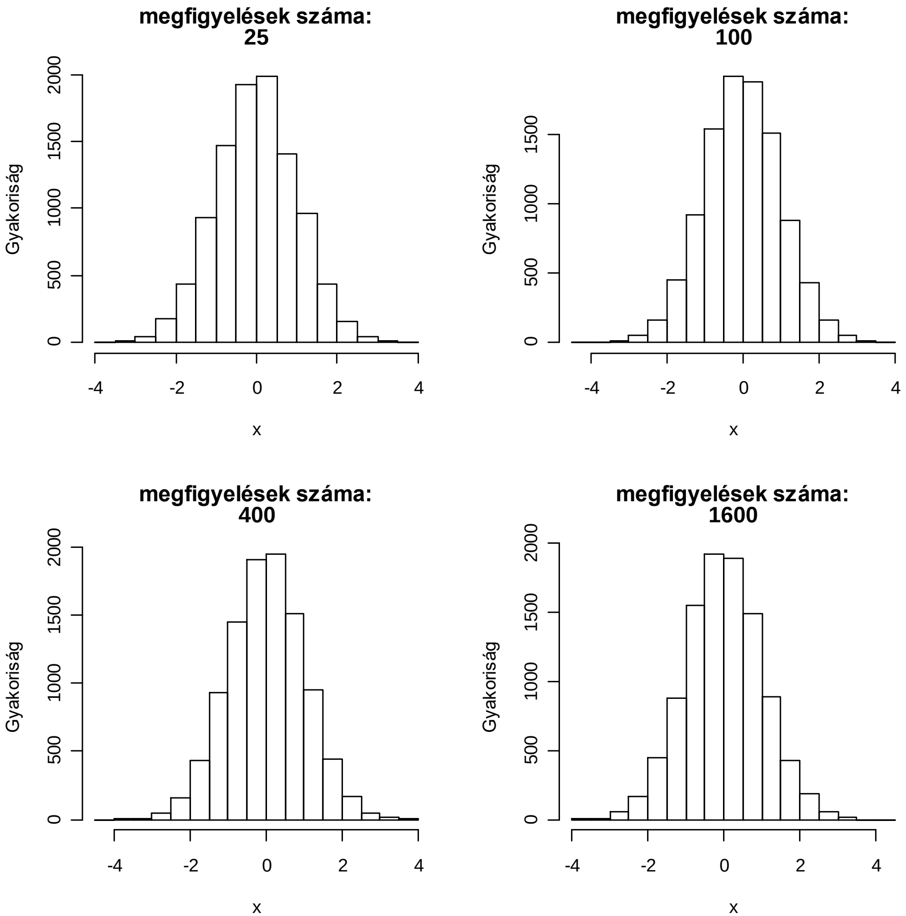
- Ha $X_1$, $X_2$,..., $X_n$,... független, azonos eloszlású valószínűségi változók, t.f.h. m=0, $\sigma$=1, akkor

$$\sup_z \left| P\left( \frac{X_1 + \ldots + X_n}{\sqrt{n}} < z \right) - \Phi(z) \right| \leq c \frac{E\,|\,X_1\,|^3}{\sqrt{n}}$$

  (Berry-Esséen tétel). c=0,47 a legjobb ismert érték.
- Gyakorlatban nagyon függ az eloszlás alakjától. Például az egyenletes eloszlásra n=12 elég jó közelítést ad, de az exponenciális eloszlásnál n=50 szükséges.

# Centrális határeloszlás-tétel: független, azonos, egyenletes eloszlások standardizált összege (n=25,100,400,1600)

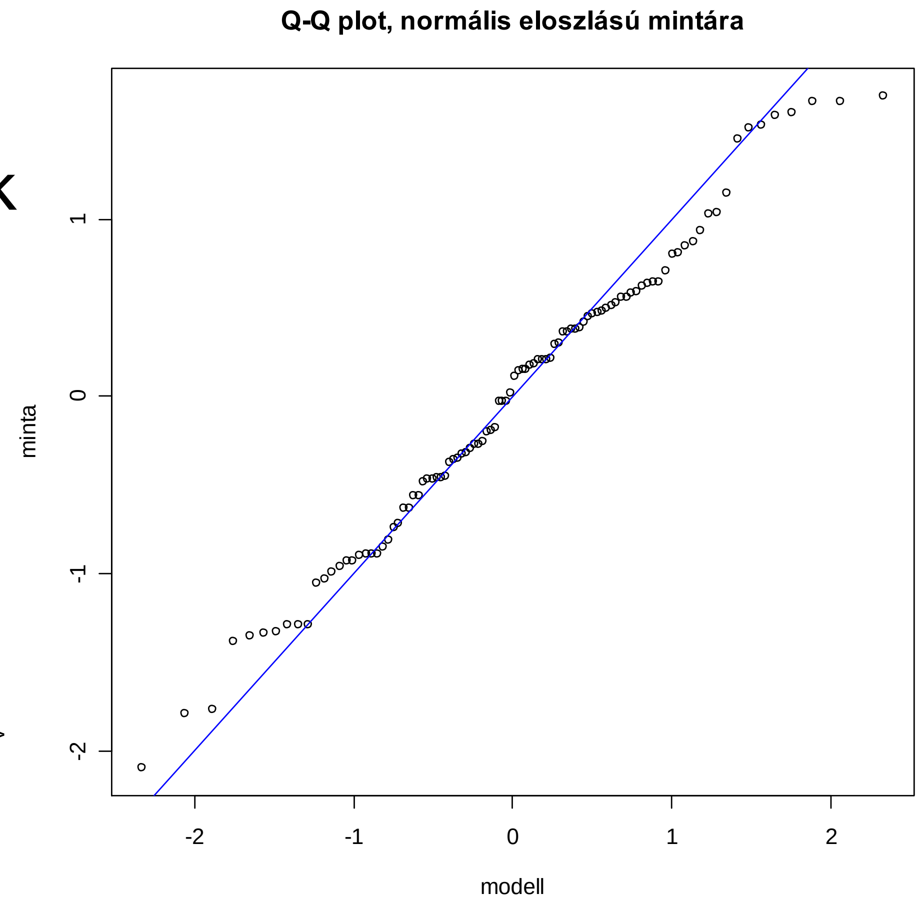
Itt és a következő oldalakon az adott n számú véletlen számot generáltuk, standardizáltuk az összegüket, és ezt 10000-szor megismételtük minden esetben. Ezek után azt vizsgáltuk, hogy a kapott 10000 véletlen szám mennyire van közel a standard normális eloszláshoz.



# Eloszlás illeszkedésének vizsgálata: Q-Q plot

A megfigyelt és az illesz-tett (G) eloszlás kvantiliseinek kétdimenziós ábrázolása.
Eloszlásfüggvény q-kvantilise: az az érték, amelynél q valószínűség-gel kapunk kisebbet: $G^{-1}(q)$

$$\left\{\left(G^{-1}\left(\frac{k}{n+1}\right), X_k^{(n)}\right) : k = 1,2,\ldots,n\right\}$$



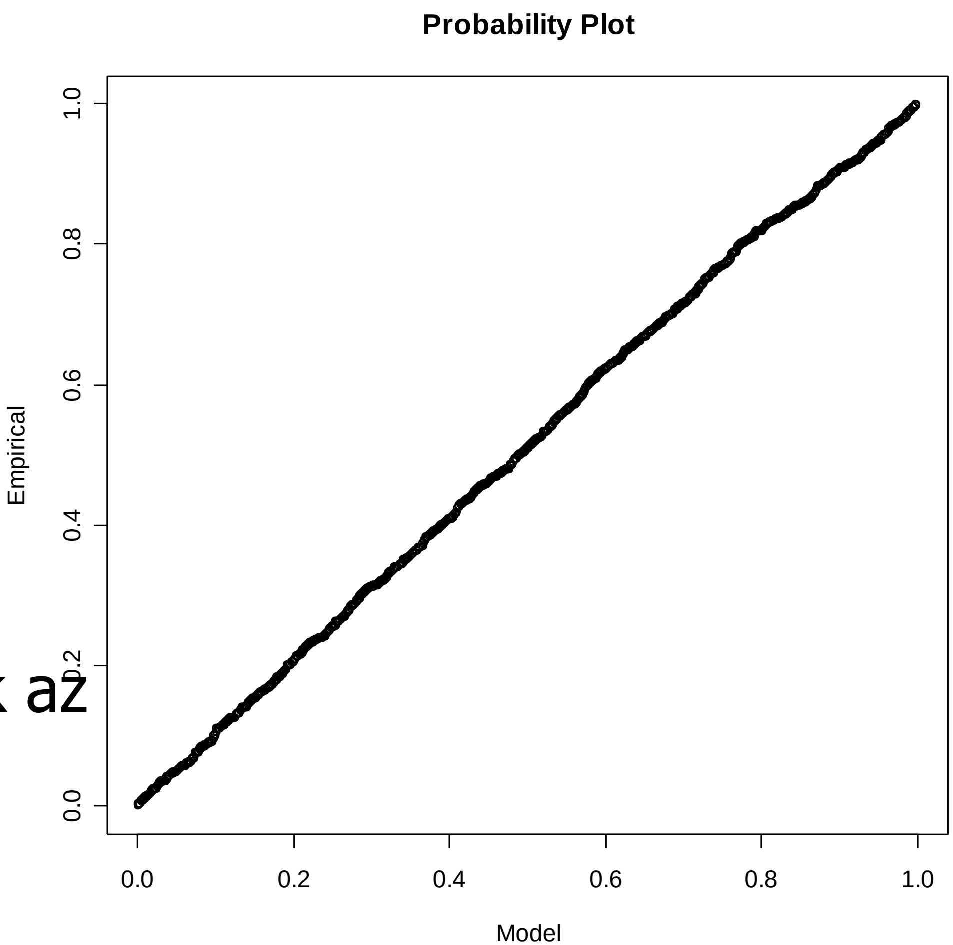

Ha jó az illeszkedés, a pontok az átló közelében vannak

# Eloszlás illeszkedésének vizsgálata: P-P plot

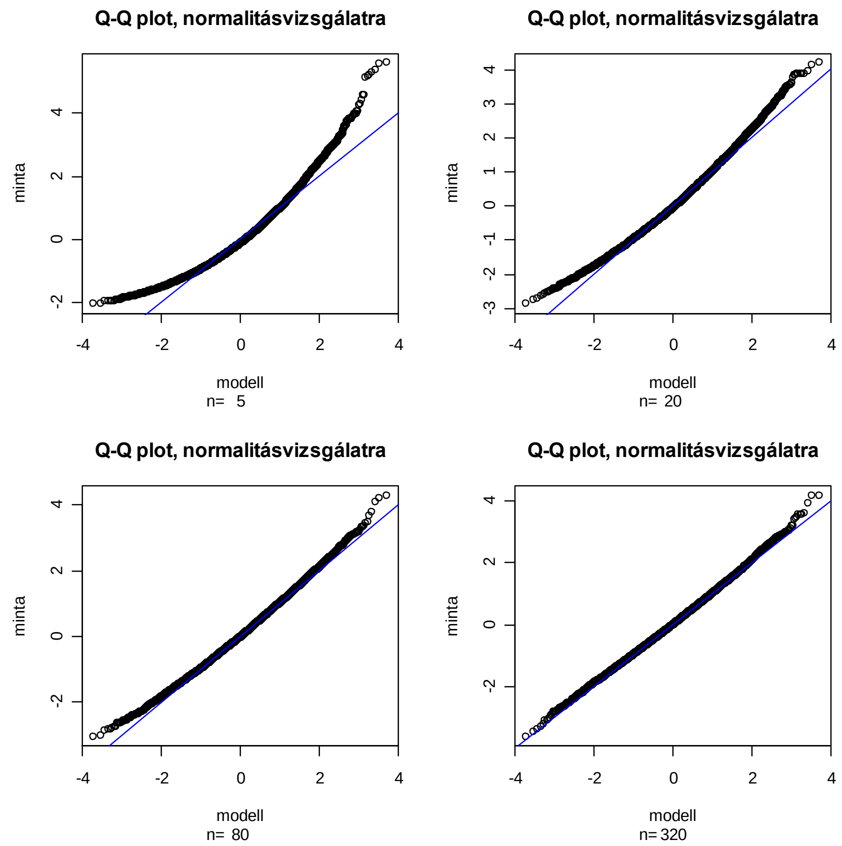
A megfigyelt és az illesz-tett eloszlás kétdimenziós ábrázolása

$$\left\{\left(\frac{k}{n+1}, G(X_k^{(n)})\right) : k = 1,2,\ldots,n\right\}$$

Ha jó az illeszkedés, a pontok az átló közelében vannak



# Centrális határeloszlás-tétel: független, azonos, exponenciális eloszlások standardizált összege (n=5,20,80,200)



# Centrális határeloszlás-tétel: független, azonos, Pareto (k=3) eloszlások standardizált összege (n=80,1280,20000)

Itt nagyon lassú a konvergencia