Galton-Watson elágazó folyamat

[Francis Galton, Henry William Watson 1874 – angol családnevek kihalásáról]

Egy populáció létszámát / lélekszámát / elemszámát modellezzük

diszkrét időben: $n = 0, 1, 2, 3, \dots$

A létszámok legyenek $Z_0, Z_1, Z_2, Z_3 \dots$

n gyakran nem esik egybe a tényleges idővel, így Z_n jelentése ~~a populáció elemszáma n-kor~~ az n-edik generáció elemszáma.

Modell:
- Kezdetben van egyetlen egyed, ő egyedül alkotja a nulladik generációt: $Z_n \equiv 1$ ← Az egyszerűség kedvéért
- $X \in \mathbb{N}$ valószínűségi változó az egylépéses utódszám: a populáció minden egyede véletlen számú utódot ~~...~~hoz létre egymástól teljesen függetlenül és X-szel (és egymással is) azonos eloszlással.
- Az n-edik generáció tagjainak utódai alkotják az $(n+1)$-edik generációt.

értsd:

- csak a közvetlen utódok (unokák nem)
- ~~a sz~~ az n-edik generáció tagjai nem számítanak bele az $(n+1)$-edikbe

[Pl. ha n tényleges idő, akkor úgy kell venni, hogy mindenki létrehoz valahány utódot, ő maga pedig meghal.]

Pl:

Mat. modell:

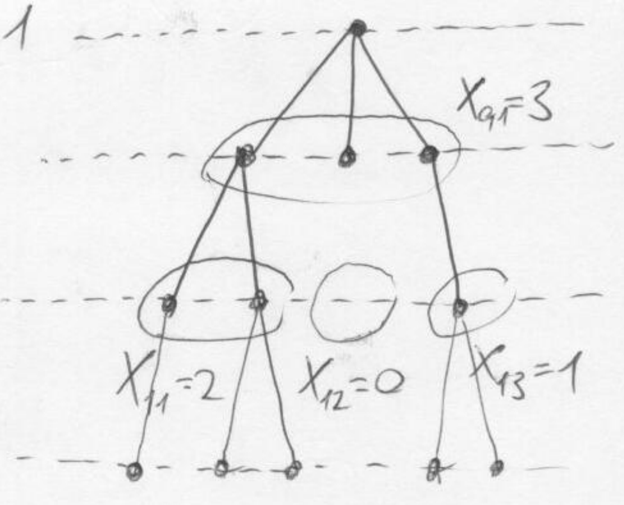
$n=0,1,2,\ldots$ generációk

$k=1,2,3\ldots$: egyedek számozása a generációkon belül

$n=0 : Z_n=1$

$n=1 : Z_1=3$

$n=2 : Z_2=3$



$X_{n,k}:=$ az n-edik generáció k-adik tagja által létrehozott utódok száma $\quad n=0,1,2,\ldots$
$\quad k=1,2,3,\ldots$

Feltevés: Az $X_{n,k}\in\mathbb{N}$, $n=0,1,2\ldots$; $k=1,2,3\ldots$ család teljesen független és ~~azo~~ X-szel (és egymással) azonos eloszlású.

[Persze $X_{n,k}$ csak akkor érdekes, ha az n-edik generáció k-adik tagja ténylegesen megszületik – pl. a fenti példában lehet, hogy $X_{1,4}=100$, de ez mindegy.]

Ezek után

Def (Galton-Watson folyamat)

Legyen $X\in\mathbb{N}$ valószínűségi változó, és legyenek $n=0,1,2,\ldots$ ill. $k=1,2,3\ldots$-re $X_{n,k}$ teljesen független, X-szel azonos eloszlású val. változók. Legyen $Z_0=1$, és legyen $n=0,1,2,\ldots$-re

$$Z_{n+1}=\sum_{k=1}^{Z_n} X_{n,k}.$$

Ekkor $Z_0, Z_1\ldots$ Galton-Watson elágazó folyamat, X egylépéses utódszám-eloszlással.

Megjegyzések:

1.) Pistike ~~...~~ gyerekeinek száma független

- attól, hogy ő hanyadik generáció tagja
- attól, hogy a testvéreinek/unokatestvéreinek hány gyereke van
- attól, hogy hány testvére van
- attól, hogy ő hanyadik gyerek a családban
- és úgy átalában <u>minden előzménytől</u>,

feltéve persze, hogy Pistike egyáltalán megszületik!

2.) Ennek megfelelően $Z_{n+1} = \sum_{k=1}^{\boxed{Z_n}} X_{n,k}$ <u>véletlen tagszámú összeg</u>.

3.) Persze Pistike gyerekeinek számától <u>nem</u> független

- az unokáinak száma
- a teljes családfa össz-elemszáma

4.) A nulladik generáció 1-etlen elemének neve: „A gyökér".

5.) Ha $Z_n = 0$, akkor garantáltan $Z_{n+1} = Z_{n+2} = \ldots = 0$:

ha a folyamat kihalt, akkor kihalva is marad (nincs bevándorlás).

$\{\text{kihalás}\} = \{\exists n : Z_n = 0\}$

Jelölések:

- Az $X_{n,k}$-k közös várható értéke $m := EX$

  generátorfüggvénye $g(z) := g_X(z)$

- Z_n várható értéke $m_n := \mathbb{E}Z_n$

  generátorfv-e $g_n(z) := g_{Z_n}(z)$

- $r_n := \mathbb{P}(Z_n = 0)$ a kihalás val.sége az n-edik generációig

- $r_\infty = \mathbb{P}(\text{kihalás}) = \mathbb{P}(\exists n : Z_n = 0)$ a kihalás valószínűsége (bármeddig)

- $N := \sum_{n=0}^{\infty} Z_n$ a populáció össz-elemszáma, a gyökeret is beleértve.

  Ez persze véletlen, N véges $\iff$ a folyamat kihal,

- N generátorfv-e $g_N(z)$

Kérdések:

1.) $m_n = ?$ Ez hogy viselkedik $n \to \infty$-re?

2.) Mi X_n eloszlása?

3.) $r_n = ?$

4.) $r_\infty = \mathbb{P}(\text{kihalás}) = ?$

5.) $\mathbb{E}N = ?$

6.) Mi N eloszlása?

Válaszok:

- $Z_0 \equiv 1$, így $m_0 = 1$, $g_0(z) = z$, $r_0 = \mathbb{P}(Z_0 = 0) = 0$
- $Z_1 = X_{0,1} \sim X$, így $m_1 = m$, $g_1(z) = g(z)$, $r_1 = \mathbb{P}(X = 0) = g(0)$
- Általában $n = 0, 1, 2, \dots$-re $Z_{n+1} = \sum_{k=1}^{Z_n} X_{n,k}$ véletlen tagszámú

$$\text{összeg} \xrightarrow[\text{tételek}]{\text{korábbi}} \left. \begin{array}{l} m_{n+1} = m_n \cdot m \\ g_{n+1}(z) = g_n(g(z)), \text{ vagyis } \underbrace{g_{n+1} = g_n \circ g}_{\text{kompozíció}} \end{array} \right\} n = 0, 1, \dots$$

Köv.:

Tétel 1) $m_n = m^n$ $\quad n = 0, 1, 2, \dots$

2) $g_n = g \circ g \circ \dots \circ g$ $\quad$ n-szeres fv-kompozíció, $n = 0, 1, 2, \dots$

[Megj: Ez még $n = 0$-ra is igaz:
- a „nulla tagú összeg” $:= 0$ = az összeadás neutrális eleme = semmit nem hozzáadás
- a „nulla tényezős szorzat” $:= 1$ = a szorzás neutrális eleme = semmivel sem szorzás
- a nulla tagú fv-kompozíció := identitás függvény $g(z) \equiv z$ = a fv-kompozíció neutrális eleme = semmit nem csinálás.]

Köv.: $g_{n+1} = g \circ g_n$ $\quad n = 0, 1, 2, \dots$

[Megj: Ez nem nyilvánvaló. A véletlen tagszámú összegből csak az jött ki, hogy $g_{n+1} = g_n \circ g$, és a kompozíció általában nem kommutatív. Az, hogy g_n és g most kivételesen felcserélhető, azért van, mert csak a $Z_0 \equiv 1$ esetet nézzük.]

Ebből jön ki a kihalás valószínűségére vonatkozó

<u>Tétel</u> Az $r_0, r_1, r_2 \dots$ kihalási valószínűségek sorozatát az

$$r_0 := 0$$

$$r_{n+1} := g(r_n) \quad n = 0, 1, 2 \dots$$

rekurziós szabály adja.

<u>Biz.</u>: A generátorfüggvény egy nagyon hasznos tulajdonsága, hogy $r_n = \mathbb{P}(Z_n = 0) = g_n(0)$. Ebből

$$r_{n+1} = g_{n+1}(0) = g(g_n(0)) = g(r_n) \quad \square$$

[<u>Megj</u>: Igaz az $r_n = \underbrace{g(g(g(\dots(g(0))\dots)))}_{n\text{-szeres kompozíció}}$ zárt formula is, de nem túl hasznos: nagyon nem érdemes egy n-szeres fv-kompozíciót kiszámolni (ami többnyire nagyon csúnya, ha $n > 2$), ha elég n db függvény-behelyettesítés.]

Kritikusság, kihalási valószínűség

Emlékeztető: $\mathbb{E}Z_n = m_n = m^n$, ahol ~~m~~ $m = \mathbb{E}X$ az 1-lépéses utódszám várható értéke

Lényeg: $m_n = m^n \begin{cases} \to 0, & \text{ha } m<1 \\ \equiv 1, & \text{ha } m=1 \\ \to \infty, & \text{ha } m>1 \end{cases}$

Ez motiválja a következőt:

Def: A Galton-Watson folyamat

- szuperkritikus, ha $m>1$
- kritikus, ha $m=1$
- szubkritikus, ha $m<1$

Sejthető: ha $m<1$, akkor $\mathbb{P}(\text{kihalás}) = 1$

ha $m>1$, akkor $\mathbb{P}(\text{kihalás}) < 1$

Nehezebb: Mi van, ha $m=1$?

Tétel ~~A Galton-Watson~~ Szubkritikus Galton-Watson folyamatra

$$\mathbb{P}(\text{kihalás}) = 1.$$

Tétel Kritikus Galton-Watson folyamatra $\mathbb{P}(\text{kihalás}) = 1$, hacsak nem $X \equiv 1$ (vagyis mindenkinek pontosan 1 utódja van), mert akkor persze $\mathbb{P}(\text{kihalás}) = 0$

Tétel Szuperkritikus Galton-Watson folyamatra $\mathbb{P}(\text{kihalás}) = r_\infty < 1$, a $g = g_X$ generátorfüggvény egyetlen $[0,1)$-beli fixpontja: $r_\infty = z$, ha $z = g(z)$ [és r_∞ az egyetlen ilyen pont $[0,1)$-ben].

A 3 tételt egyszerre bizonyítjuk.

Biz: NAGYON TANULSÁGOS! $r_n = \mathbb{P}(Z_n = 0) = \mathbb{P}(n \text{ lépésben kihal})$.

Tudjuk, hogy $r_{n+1} = g(r_n)$ rekurzívan megadott sorozat,

és persze $r_{n+1} \geq r_n$, mert $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$

Ebből következik, hogy $\cancel{r_\infty = \lim_{n\to\infty} r_n}$

$$\mathbb{P}(\text{kihalás}) = r_\infty = \lim_{n\to\infty} r_n .$$

Vajon miről lehet megismerni egy rekurzívan definiált sorozat határértékét?

Hát onnan, hogy a rekurziót újra és újra alkalmazva, egy idő után már nem változik (észrevehetően) a sorozat.

Ezt mondja ki a következő <u>nagyon általános</u> tétel:

<u>Tétel</u>: Legyen H halmaz, $g: H \to H$ folytonos, és legyen az $a_0, a_1, a_2 \ldots$ sorozat rekurzívan definiált: $a_{n+1} = g(a_n)$ $n = 0, 1, 2 \ldots$

Tegyük fel, hogy $A = \lim_{n\to\infty} a_n$.

Ekkor $A = g(A)$.

<u>Magyarul</u>: <u>Folytonos</u> rekurziós függvénnyel adott rekurzív sorozat határértéke (ha van egyáltalán), csak a rekurziós függvény fixpontja lehet.

<u>Biz</u>: $a_{n+1} = g(a_n)$, vegyük mindkét oldalon $n \to \infty$-t:

$$\begin{array}{ccc} a_{n+1} & = & g(a_n) \\ \downarrow & & \downarrow \\ A & & A \\ & & \downarrow \\ & = & g(A) \end{array}$$

mert g folytonos.

$\square$

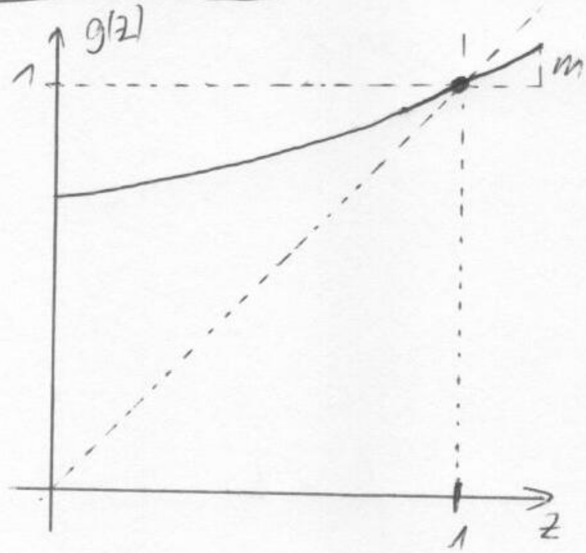
[Megj: A tétel tényleg nagyon általános: H bármilyen halmaz lehet, ahol a folytonosság és a határérték fogalma értelmes]

Ott tartunk, hogy $r_\infty = g(r_\infty)$, és persze $r_\infty = \mathbb{P}(\text{kihalás}) \in [0,1]$.

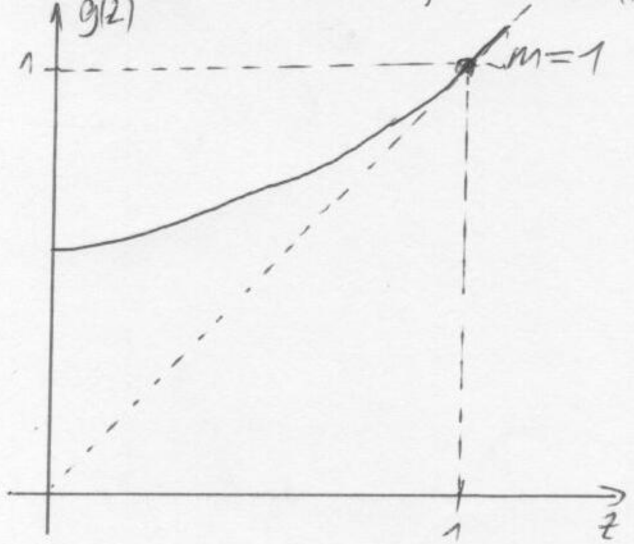
g-ről tudjuk, hogy

- $g: [0,1] \to [0,1]$ folytonos, sőt konvex
- $g(1) = 1$, $g'(1) = m$
- g szigorúan konvex, hacsak nem $X \sim$ Bernoulli, mert akkor lineáris.

Szubkritikus eset:

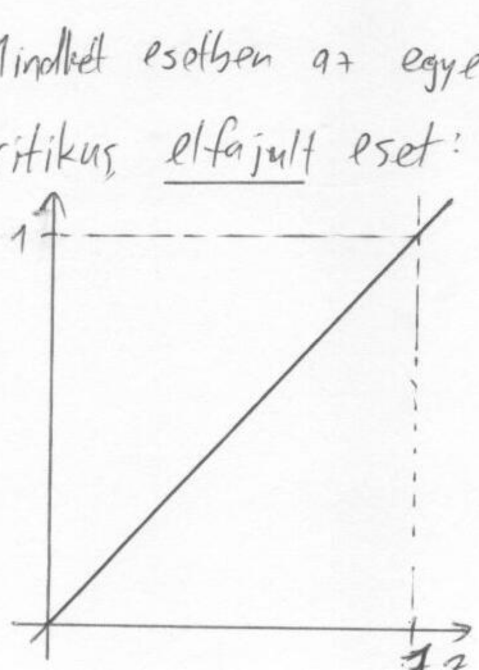


Kritikus, nem elfajult eset ($X \not\equiv 1$)



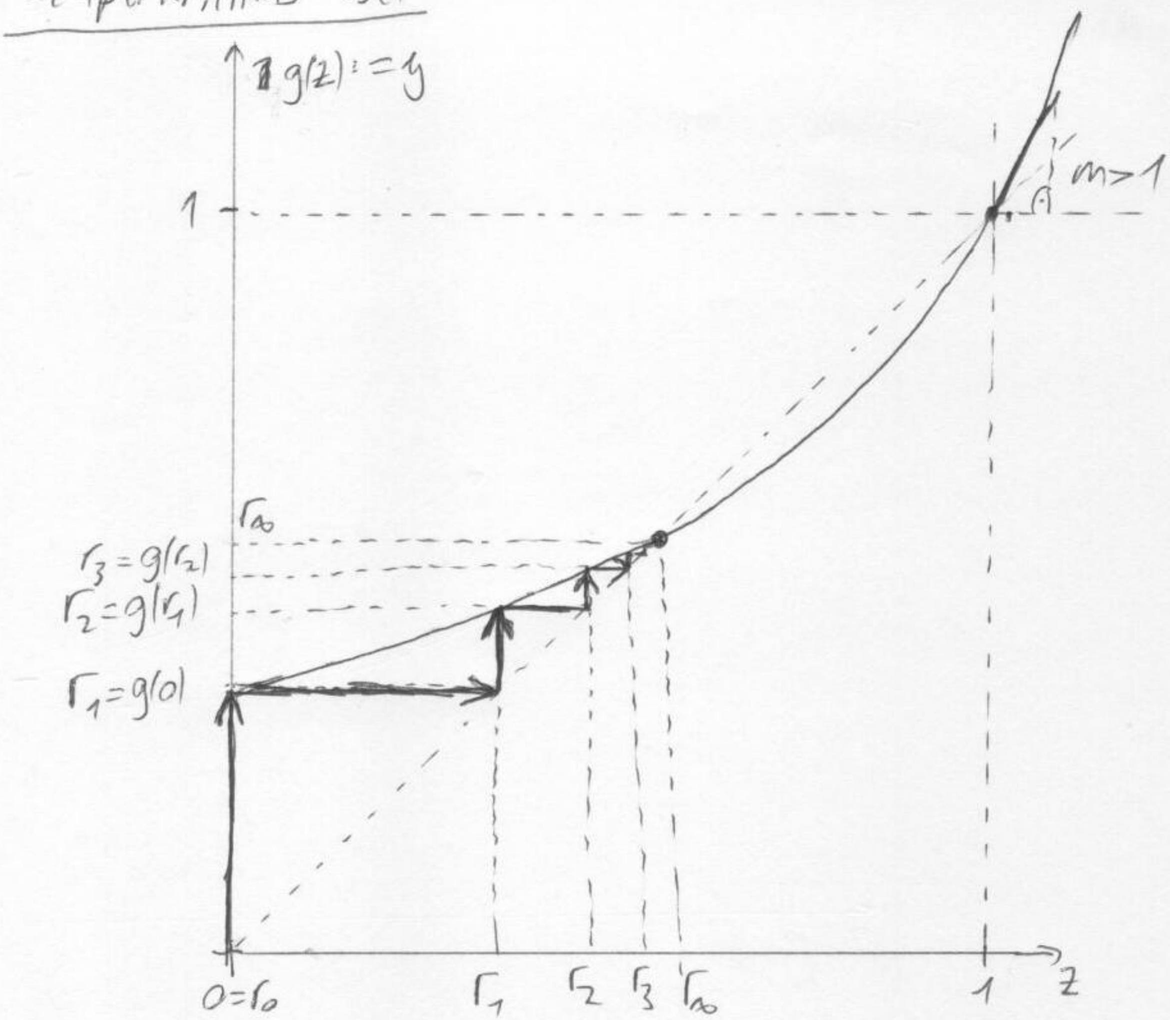
Mindkét esetben az egyetlen fixpont a $z=1$, tehát $r_\infty = 1$.

Kritikus elfajult eset: $X \equiv 1$, így $g(z) = z$



Minden $z \in [0,1]$ fixpont, de az iteráció az $r_0 = 0$ fixpontból indul, és ott is marad örökre $\Rightarrow$ $r_\infty = 0$

Szuperkritikus eset



- $z=1$ persze most is fixpont, de $m = g'(1) > 1$, így ha z 1-nél kicsivel kisebb, akkor $g(z) \lneqq 1$. Mivel g folytonos, kell lenni még 1 fixpontnak, hiszen $g(0) \geqq 0$.
- A konvexitás miatt pedig <u>pontosan 1</u> fixpont ~~lege~~ lesz $[0,1)$-ben.
- <u>Melyiket</u> találja meg az iteráció?

Ezt mutatja az „1-dimenziós iteráció grafikus ábrázolása": a mindenkori $r_{n+1} = g(r_n)$ függvényértékeket visszavetítjük a függőleges tengelyről a vízszintesre az átló segítségével – <u>lásd az ábrát</u>: lényegében az $y = g(z)$ ~~és~~ fv-grafikon és az $y = z$ átlós egyenes között kell cikcakkozni a nyilak mentén. Így a kisebbik, $r_\infty \lneqq 1$ fixpontba futunk bele. □

Megj.: Általában igaz (minden m-re), hogy r_∞ a $z \mapsto g(z)$ függvény legkisebb $[0,\infty)$-beli fixpontja, de ez csak a szuperkritikus esetben hasznos: $m \le 1$-re kár a fixpontokat keresgélni.

A teljes családfa mérete

Emlékeztető: $N = \sum_{n=0}^{\infty} Z_n = Z_0 + Z_1 + Z_2 + \dots$

és $\mathbb{E} Z_n = m^n$.

Ebből $\mathbb{E}N = m^0 + m^1 + m^2 + \dots$ könnyű: mértani sor összege.

Tétel: $\mathbb{E}N = \begin{cases} \infty, & \text{ha } m \ge 1 \\ \frac{1}{1-m}, & \text{ha } m < 1. \end{cases}$

Érdekes: A KRITIKUS $(m=1)$ folyamat 1 val.széggel kihal (ha nem elfajult), de $\mathbb{E}(\text{össz-lélekszám}) = \mathbb{E}N = \infty$:

nem elfajult Galton-Watson	szubkritikus $m<1$	kritikus $m=1$	szuperkritikus $m>1$
$\mathbb{P}(\text{kihalás})$	$=1$	$=1$	<1
$\mathbb{E}N$	$<\infty$	$=\infty$	$=\infty$

A teljes családfa-méret eloszlása

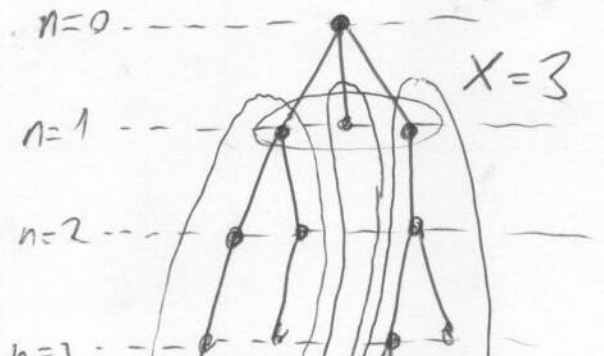
~~Kérdés~~ Feladat: Keressük meg $N = Z_0 + Z_1 + Z_2 + \dots$ eloszlását!

Megoldás: N generátorfüggvényét fogjuk megkeresni: ez teljes információt hordoz az eloszlásról.

Trükk: Legyen

- $Z_1 = X = X_{0,1}$ az 1. generáció elemszáma = a gyökér gyerekeinek száma
- N_1 az 1. generáció 1. tagjának teljes családfaméretе
- N_2 az 1. — // — 2. — // —
- N_3 az 1. — // — 3. — // —

  ⋮ — // —
- N_k az 1. — // — k.



$N_1 = 7 \quad N_2 = 1 \quad N_3 = 9$

Észrevétel: • az 1. generáció k. tagjának születésével egy ugyanolyan Galton-Watson folyamat indul, mint az eredeti, csak ennek máshol van a gyökere

- Ezért N_k ugyanolyan eloszlású, mint N, minden k-ra.
- Ráadásul $N_1, N_2, N_3 \dots$ függetlenek (egymástól, teljesen), és még X-től is.

- Persze N_k nem független N-től (pl. mindig $N > N_k$), de sebaj.
- Ezekkel $N = \underset{\substack{\uparrow \\ \text{gyökér}}}{1} + \sum_{k=1}^{X} N_k$, ahol $S_X := \sum_{k=1}^{X} N_k$ véletlen tagszámú összeg

Így, ha $N, N_1, N_2, N_3 \ldots$ közös generátorfüggvénye g_N,

akkor $S_X = \sum_{k=1}^{X} N_k$ generátorfv-e $g \circ g_N$;

az $U :\equiv 1$ val. változó generátorfv-e $g_U(z) = z$, és persze

$U \equiv 1$ független S_X-től,

így $1 + \sum_{k=1}^{X} N_k = U + S_N$ generátorfv-e $z \cdot g(g_N(z))$

De, mint láttuk $1 + \sum_{k=1}^{X} N_k = N$, vagyis beláttuk, hogy

<u>Tétel</u> $\boxed{g_N(z) = z\, g(g_N(z))}$

Avagy: A $g_N(z)$ generátorfüggvény $y := g_N(z)$ függvényértéke minden $z \in [0,1]$-re eleget tesz az $\cancel{y = z\, g(z)}$ $y = z\, g(y)$ egyenletnek.

<u>Megj:</u> 1.) $g_N(z)$ megtalálásához ezt az $y = z\, g(z)$ egyenletet kell megoldani ~~minden~~ y-ra (minden z esetén).

2.) Az egyenletnek több megoldása is lehet: a tétel csak annyit állít, hogy a keresett $g_N(z)$ a megoldások <u>egyike</u>.

3.) Az egyenlet csak <u>szubkritikus</u> vagy <u>kritikus</u> esetben ad értelmes generátor függvényt, amikor $\mathbb{P}(N<\infty) = 1$.

Olyan val. változóknak, amik ∞-ek is lehetnek, nem is beszéltünk a generátorfüggvényéről. [Ha nagyon akarnánk, lehetne.]

14/14