EXTRÉM STAT

2. előadás (09.27.)

A téma motivációja: tudunk-e extrapolálni olyan események valószínűségére, amelyre nem történtek meg a tapasztalat alapján?

Pl.:
- Hány hurrikán lesz 100 év alatt?
- Mi lesz a legnagyobb szélerősség?
- mennyit bír a legnagyobb terhelésen? stb...

Konkrét fizikai példák:
- Asztrofizikai klaszterekben lévő galaxisok száma a legfényesebb és második legfényesebb fv-ben.
- random-modell of spin glasses..
  Az energiaszintek eloszlását követik, mi lesz az alapállapot?

<u>Szélsőérték problémája</u>

Y változót mérjük N-szer: $\{y_1, \dots, y_N\}$

$z_N = \max\{y_1, \dots, y_N\}$

Kérdés: Mi lesz z_N eloszlása?

- $P(z)$ eloszlás nyilván függ $P_0(y)$ szülőeloszlástól, a valószínűség csak a $P_0(y)$ farkától függ. (Attól, hogy milyen jellegű a lecsengés)
- Ha $P_0(y)$ a végtelenbe megy le, akkor N növelésével $P(z)$ is tart a végtelenbe. De akkor mit tudunk mondani? Egyrészt, hogy z_N hogy tolódik ki, másrészt, hogy $P(z)$ lecsengése milyen.

Látni fogjuk, hogy ha y-k függetlenek, akkor $P(z)$ egyetemlegesen meghatározott, és akár több $P_0(y)$ is vezethet ugyanarra a $P(z)$-re.

Becslések:
- Mit tudunk mondani z_N-ről? TFH $P_0(y) = e^{-y}$
  z_n maximum, ha a $P_0(z_N)\Delta$ valószínűség $\frac{1}{N}$ valószínűségű, vagyis ha 1-szer fordul csak elő: $P_0(z_N)\Delta \cdot N \approx 1 \Rightarrow e^{-z_N} N \approx 1 \Rightarrow z_N \approx \ln N$
- Mi az eloszlás szélessége? z_N-től δ-nyira a valószínűség e-edére csökken.
  $P(z+\delta)\Delta \cdot N \approx 1/e$

Mit tudunk analitikusan?

Exponenciális eloszlás esetén: $P_0(y) = e^{-y}$

A Gumbel-eloszlás:

1. kérdés: Tudunk-e mit mondani $P_N(z_N)$-ről?

2. kérdés: Van-e olyan $z_N = a_N x + b_N$ trafó, amire $P_N(x_N)$-ről többet tudunk mondani?

Praktikus lenne az, hogy $\bar{x}_N = 0$ legyen.

Ezzel megszüntetjük a végtelenbe menés problémáját.

z_N eloszlására kumulatív eloszlásfv-e:

$$M_N(z) = \int_{-\infty}^{z} P_N(z_N)\,dz = \left[\int_0^z P_0(y)\,dy\right]^N = (1-e^{-z})^N$$

úgy kéne eltolni, hogy megszüntessük a végtelenbe tolódást.

Tudjuk, hogy $\bar{z}_N \to \ln N \Rightarrow$ legyen $x := z - \ln N$

$$M_N(x) = \left(1 - e^{-x-\ln N}\right)^N = \left(1 - \frac{e^{-x}}{N}\right)^N = e^{-e^{-x}}$$

Ebből $\underline{P(x) = e^{-x-e^{-x}}}$ Fisher–Tippett–Gumbel-eloszlás

Általános alak: $P(x) = \frac{1}{b} e^{-\frac{x-a}{b} - e^{-\frac{x-a}{b}}}$

a: hely paraméter

b: skála paraméter

Kérdés, hogy ugyan mivel kell eltolni $\bar{z}_N$-t, hogy $\bar{x}_N = 0$ legyen?

$$M_N(z) = (1-e^{-z})^N \Rightarrow \bar{z}_N = \int_0^\infty z \frac{dM_N}{dz}\,dz = N\int_0^\infty z\, e^{-z}\frac{(1-e^{-z})^N}{1-e^{-z}}\,dz = \qquad x = z - \ln N$$

$$= N\int_{-\ln N}^{\infty} (x+\ln N)\frac{e^{-x}}{N}\frac{\left(1-\frac{e^{-x}}{N}\right)^N}{1-\underbrace{\frac{e^{-x}}{N}}_{\to 0}}\,dx \to$$

$$\to \int_{-\infty}^{\infty} (x+\ln N)\, e^{-x-e^{-x}}\,dx = \ln N + \int_{-\infty}^{\infty} x\, e^{-x-e^{-x}}\,dx =$$

$$= \ln N + \gamma_E \leftarrow \text{Euler konstans} = 0{,}57\ldots$$

Szóval $\bar{z}_N$ $\ln N$-nel tart a ∞-be, de ahhoz, hogy $\bar{x}_N$ 0 legyen, még γ_E-vel el kell tolni.

Egyébként a második momentum $\langle x^2 \rangle = \frac{\pi^2}{6} + \gamma_E^2$

Egyszóval, az FTG-eloszlás széles határon számolt alakja:

$$P_{FTG}(x) = \frac{\pi}{\sqrt{6}} \cdot e^{-\frac{\pi}{\sqrt{6}} \cdot x - \gamma_E - e^{-\frac{\pi}{\sqrt{6}} x - \gamma_E}}$$

← (ez lehet, hogy nem jó és majd kijavítja)

Ugyanezt elvégezhetjük Gaussi szülőeloszlásra

$$P_0(y) = 2y\, e^{-y^2} \quad \Rightarrow \quad M_N(z) = \left[\int_0^z P_0(y)\, dy\right]^N = (1 - e^{-z^2})^N$$

$$z = \frac{x}{\sqrt{\ln N}} + \sqrt{\ln N}, \quad x = \sqrt{\ln N} \cdot z - \ln N$$

$$N_N(x) = \left(1 - \frac{\exp\left(-x - \frac{x^2}{\ln N}\right)}{N}\right)^N = e^{-e^{-x}}$$

⇒ ugyanúgy FTG-t kapunk.

EXTRÉM STAT

3. előadás (10.04.)

Az extrém értékekről szóló ismereteinket $p_0(y)$ alapján determináljuk, de valójában a farka az, ami igazán számít.

Nagy x: $x \to \infty$ -re

- $p_0(y) \sim e^{-y^\alpha} \Rightarrow M_{FTG}(x) = e^{-e^{-x}}$
- $p_0(y) \sim y^{-\frac{1}{\delta}-1} \Rightarrow M_{FTF}(x) = \begin{cases} e^{-x^{1/\delta}} & x > 0 \\ 0 & x < 0 \end{cases}$
- $p_0(y) \sim (y_0 - y)^{\frac{1}{\delta}-1} \Rightarrow M_W(x) = \begin{cases} e^{-(-x)^{1/\delta}} & x < 0 \\ 1 & x > 0 \end{cases}$

Vezessük le a tételeit is:

FTF

$$p_0(y) = \frac{\alpha}{y^{\alpha+1}}$$

$$M_N(z) = \left[\int_1^z p_0(y)dy\right]^N = \left[\int_1^z \frac{\alpha}{y^{\alpha+1}} dy\right]^N = \left(1 - \frac{1}{z^\alpha}\right)^N =$$

$$= \left(1 - \frac{1}{N(N^{-\frac{1}{\alpha}} z)^\alpha}\right)^N = \qquad x := N^{-\frac{1}{\alpha}} z$$

$$= \left(1 - \frac{1}{N x^\alpha}\right)^N = e^{-x^{-\alpha}}$$

általános alak: $P_{FTF}(x) = a\left(\frac{x-a}{b}\right)^{-\alpha-1} e^{-\left[\frac{x-a}{b}\right]^{-\alpha}}$

$x \to \infty$ -re $p_{FTF}(x) \sim x^{-\alpha-1}$

Weibull

$$p_0(y) = \frac{\beta+1}{a^{\beta+1}} (a-y)^\beta \qquad z_{max} \xrightarrow[N \to \infty]{} a$$

emiatt csak becsülhetjük a.a. logikát

$p_0(\bar{z}_N)\Delta \cdot N \approx 1$ ahol $\Delta \simeq a - \bar{z}_N$

$$\Rightarrow a - \bar{z}_N \simeq N^{-\frac{1}{\beta+1}}$$

valóságban: $M(z) = \left(\int_0^z p_0(y)dy\right)^N = \left(1 - \left(1 - \frac{z}{a}\right)^{\beta+1}\right)^N \qquad x := \frac{z-a}{a} N^{\frac{1}{\beta+1}}$

$$= \left(1 - (-x)^{\beta+1}/N\right)^N = e^{-(-x)^{\beta+1}}$$

általános alak: $p_W(x) = \frac{\beta+1}{b}\left(\frac{a-x}{b}\right)^\beta e^{-\left(\frac{a-x}{b}\right)^{\beta+1}}$

Általános alakja a renormalizációs csoport egyenletnek.

Ising-modell: $H = -\sum_i J_i s_i s_{i+a}$

$$\beta J = K \xrightarrow{T\to\infty} 0, \qquad K \xrightarrow{T\to 0} \infty$$

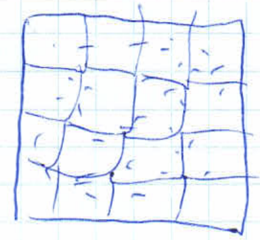
$\Rightarrow$ nagy T-n kis K, kis T-n nagy.

mert $e^{\beta J s_i s_{i+a}}$ $T=0$-nál ∞, $T\to\infty$-nél 1

Nagyon alacsony T-n mind egyforma mutat, nagy magas T-n random.

A kettő között van T_c ahol fázis-átalakulás van. Mi történik itt?

Osszuk fel a rendszert részekre és tekintsük egy dobozt egy egységnek, átlagoljuk a dobozok spinjét.



Ahol a távol levő spinek kevésbé korreláltak, mint a szomszédosak, ezért a dobozok kevésbé korreláltak, mint a részecskék.

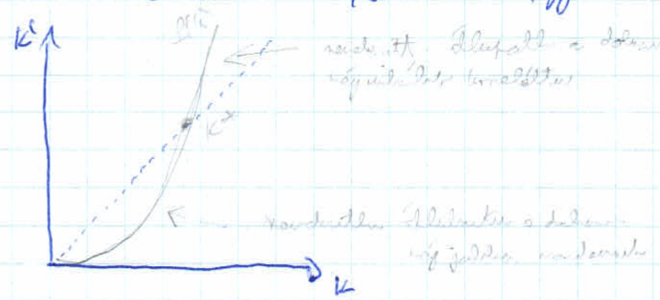
DE! Mi van a kritikus pontban? Nincs se rend, se rendezetlenség $\Rightarrow$ skálainvariancia, ezért ugyanazt látjuk kisebb dobozokkal.

Ha a eredeti csatolás $K s_i s_{i+a} \Rightarrow K' s_i s_{i+a}$ lesz, akkor K' függ az eredetitől: $K' = f(K)$

Lesz egy pont, ahol $K' = K$

$\Rightarrow K^* = f(K^*)$

itt a skála faktor már nem változtat a rendszeren.



Mi köze van az extrém statisztikához?

$$M(x) = \lim_{N\to\infty} \left[\mu(z)\right]^N \Big|_{z=a_N x+b_N} \qquad \text{ahol } \mu(z)=\int_{-\infty}^{z} \rho_c(y)\,dy$$

TFH $M(x)$ a határeloszlás! Ekkor az a fenti egyenletet kielégíti

$$\left[M(a_p x+b_p)\right]^p = M(x)$$

[Hogy oldunk meg ilyen egyenleteket? Pl.: $f(x+a) = f(x)+a$.

Deriválva: $f'(x+a) = f'(x)$. A bal oldal függ a-tól, a jobb oldal nem, tehát

$f'(x) = C \Rightarrow f(x) = Cx+B$ $\quad C$ és B [illegible]]

Legyen $M(x)=e^{-e^{-f(x)}}$! Ekkor az egyenlet:

$$e^{-e^{-f(x)}} = e^{-e^{-f(a_p x+b_p)+\ln p}} \Rightarrow f(x)=f(a_p x+b_p)-\ln p$$

Legyen M normálisan olyan, hogy $M(0)=M'(0)=\frac{1}{e}$ $\Rightarrow$ $f(0)=0;\ f'(0)=1$

deriválva az egyenletet: $f'(x)=a_p f'(a_p x+b_p)$

A bal oldal nem függ p-től $\Rightarrow$ $f'(x)=\frac{1}{1+\gamma x}$ $\Rightarrow$ $f(x)=\frac{1}{\gamma}\ln(1+\gamma x)+C$ ~

$$\left.\begin{aligned} M(x) &= e^{-(1+\gamma x)^{-1/\gamma}} \\ p(x) &= (1+\gamma x)^{-\frac{1+\gamma}{\gamma}}\, e^{-(1+\gamma x)^{-1/\gamma}} \end{aligned}\right\}$$

Általánosított szélsőérték statisztika (GEV-stat)

Ha $\gamma\to 0$, akkor az épp a FTG-eloszlás

EXTRÉM STAT

4. előadás (10.11.)

A GEV stat kumulatív eloszlásfüggvénye: $M(x) = e^{-(1+\gamma x)^{-1/\gamma}}$

ha $\gamma > 0$ és $-\frac{1}{\gamma} < x$ $\quad P(x) = (1+\gamma x)^{-\frac{(1+\gamma)}{\gamma}} \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$ $\quad$ FTF

⋮

ha $\gamma < 0$, $x < \frac{1}{|\gamma|}$ $\quad P(x) = (1-|\gamma| x)^{\frac{1}{|\gamma|}-1} \exp\left[-(1-|\gamma| x)^{\frac{1}{|\gamma|}}\right]$ $\quad$ W

ha $\gamma \to 0$ $\quad P(x) = e^{-x-e^{-x}}$ $\quad$ FTG

SK-spinmodell

$$H = -\sum_{i,j} J_{ij} \sigma_i \sigma_j$$ ahol J-k eloszlása: $P(J_{ij}) \sim e^{-\frac{J_{ij}^2 N}{J^2}}$

miért fontos? TFH Fe atomot szennyezünk Cu-val. Hogyan változnak a mágneses tulajdonságok?

Az egyedi [illegible] ilyen:

mivel a szennyező atomok helye random, ezért nem tudjuk a pontos d-ket, tehát J is jó közelítéssel random.

(Aztán kiderült, hogy ez a probléma máshol is fontos.)

A konkrét kísérletben nem, de TFH az energiaszintek eloszlása is hasonló:

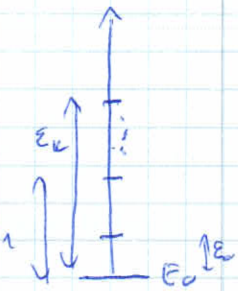
$$P_0(E_i) \propto e^{-\frac{E_i^2}{J^2 N}}$$ (a szimulációkkal egész jól egyezik)

[Az N azért kerül be, mert a H összegben N^2-tel arányos tag van.]

ez nem jó ok, amit kidumálta, de kell kamu.

A hagyományos T-n csak azok az energiák érdekesek, amik az E_{min} fölött $k_B T$-vel vannak $\Rightarrow$ A legkisebbek érdekelnek.

$$Z = \sum_{\{E_n\}} e^{-\frac{E_n}{k_BT}} = e^{-\frac{E_0}{k_BT}} + e^{-\frac{E_0+E_1}{k_BT}} + \dots + e^{-\frac{E_0+E_k}{k_BT}} =$$



$$= e^{-\frac{E_0}{k_BT}}\left(1 + e^{-\frac{E_0}{k_BT}} + e^{-\frac{E_1}{k_BT}} + \dots\right)$$

$\lesssim$
1 ha $E_0 \ll k_BT$

addig érdekes, amíg $E_k \approx k_BT$, mert utána ≈ 0

Kérdés: mennyi az alsó völgy E-nek értéke az SK-modellben?

A elérhető E-minták száma $N_{EVS} = 2^N$

Legyen $y^2 := \frac{E^2}{J^2 N}$ $\Rightarrow$ $P_0(y) = e^{-y^2}$

$u := \frac{\sqrt{\ln 2}}{J} E_{min} - N \ln 2$ $\Rightarrow$ $y_{min} = \frac{u}{\sqrt{\ln N_{EVS}}} + \sqrt{\ln N_{EVS}}$

Az E_{min} eloszlása FTG: $P(u) = e^{u - e^u} \sim e^u$ $\Rightarrow$ exponenciális eloszlás

Mekkora a gap az alsó kettő között?

$z_1 = \max\{y_1, \dots, y_N\}$

$z_2 = 2^{nd}\max\{y_1, \dots, y_N\}$

$\langle z_1\rangle - \langle z_2\rangle = ?$

Láttuk, hogy megfelelő $z = a_N x + b_N$ trafóval $\langle x_1\rangle$ $O(1)$ lesz.

$\langle z_1\rangle = a_N\langle x_1\rangle + b_N$

Azt várjuk, hogy $\langle z_2\rangle$ nem ettől túl messze, és a nagyságrendben se tér el.

$\langle z_2\rangle = a_N\langle x_2\rangle + b_N$ és $\langle x_2\rangle$ is $O(1)$-es lesz.

Legyen $x_{1,2}$ eloszlása $P_{1,2}(x)$! Ekkor a keresett különbség

$\langle z_1\rangle - \langle z_2\rangle = a_N(\langle x_1\rangle - \langle x_2\rangle) = O(a_N)$

a_N-t már ismertük: $a_N = \begin{cases} 1 & \text{ha } P_0 \text{ exponenciális} \\ 1/\sqrt{\ln N} & \text{ha } P_0 \text{ gaussi} \\ N^{\delta} & \text{ha } P_0 \sim \frac{1}{y^{\delta+1}} \\ N^{-\delta} & \text{ha } P_0 \text{ véges a levágás} \end{cases}$

Tehát a spinüveg modell esetén a gap fix lesz N-től függetlenül.

~~[illegible]~~

Számoljuk ki pontosan az eloszlást! Azt a valószínűséget keressük, hogy z körül dz-be 1 db essen, és felette csak 1 db legyen!

Láttuk, hogy $N(z) = \int_{-\infty}^{z} P_0(y)\,dy$ és $M_1(z) = (N(z))^N$

Az kell, hogy z-dz alatt N-2, z+dz felett 1 db legyen, közte pedig 1.

$$P_2(z)dz = N(N-1)(N(z))^{N-2}\,(1-N(z))\,P_0(z)\,dz$$

mert ennyi féle képpen tudjuk kiválasztani az első 2-t rendezett sorrendben.

EXTRÉM STAT

5. előadás (10.18)

Mennyi a különbség a legnagyobb és 2. legnagyobb érték között

$$P_2(z)\,dz = N(N-1)\,\mu^{N-2}(z)\;P_1(z)\,dz\,\left(1-\mu(z)\right)$$

$$\Rightarrow\quad P_2(z) = N(N-1)\,(\mu(z))^{N-2}\,\frac{d\mu}{dz}\,(1-\mu(z)) =$$

$$= N(N-1)\,\mu^{N-2}\,\frac{d\mu}{dz} - N(N-1)\,\mu^{N-1}\,\frac{d\mu}{dz}$$

$$M_2(z) = \int P_2(z)\,dz = N(\mu(z))^{N-1} - (N-1)\,(\mu(z))^N =$$

$$= N(\mu(z))^N\left[\frac{1}{\mu(z)} + \frac{1}{N} - 1\right]$$

változás: $z = a_N x + b_N$. Mivel $M_1(x) = (\mu(x))^N = e^{-(1+\gamma x)^{-1/\gamma}}$

$$\Rightarrow \mu(x) = e^{-\frac{(1+\gamma x)^{-1/\gamma}}{N}} \approx 1 - \frac{(1+\gamma x)^{-1/\gamma}}{N}$$

$$M_2(x) = N M_1(x)\left[1 + \frac{(1+\gamma x)^{-1/\gamma}}{N} + \frac{1}{N} - 1\right] = M_1(x)\left[1 + (1+\gamma x)^{-1/\gamma}\right]$$

$$P_2(x) = \frac{dM_2}{dx} = P_1(x)\left[1+(1+\gamma x)^{-1/\gamma}\right] - \underbrace{M_1(x)\,(1+\gamma x)^{-\frac{1}{\gamma}-1}}_{P_1} =$$

$$= P_1(x)\,(1+\gamma x)^{-1/\gamma} = (1+\gamma x)^{-\frac{2}{\gamma}-1}\,e^{-(1+\gamma x)^{-1/\gamma}}$$

(Belátható, hogy a k-ik legnagyobbnál is csak $(1+\gamma x)$ kitevője változik.)

Ha alkalmazásokat nézünk, akkor az a_N is beleszámít pl.: $D_{12} = a_N(\langle x_1\rangle - \langle x_2\rangle)$

De pl.: galaxisoknál elosztva a legfényesebb fluktuációjával, ami szintén a_N-nel arányos, akkor a_N kiesik, és marad egy eloszlást jellemző cucc, amiből aztán következtetni pl. N-re, vagy másra.

$$M_2(x) = M_1(x) + (1+\gamma x)^{-1/\gamma} M_1(x) = M_1(x) + (1+\gamma x)\,P_1(x)$$

$$P_2(x) = P_1(x) + \gamma P_1(x) + (1+\gamma x)\,P_1'(x)$$

$$\langle x_2\rangle = (1+\gamma)\langle x_1\rangle + \int x(1+\gamma x)P_1'(x)\,dx = \text{parciális} = (1+\gamma)\langle x_1\rangle - 1 - 2\gamma\langle x_1\rangle =$$

$$= (1-\gamma)\langle x_1\rangle - 1 \qquad \Rightarrow \qquad \underline{\langle x_1\rangle - \langle x_2\rangle = \gamma\langle x_1\rangle + 1}$$

Vajon hogy viselkedik az abszolút különbség?

$$\Delta_{12} = a_N (\langle x_1\rangle - \langle x_2\rangle) = a_N (\gamma \langle x_1 \rangle + 1)$$

1) FTF. $\gamma > 0$, $a_N \sim N^{\gamma}$ $\Rightarrow$ Δ_{12} növekszik

2) W. $\gamma < 0$ $a_N \sim N^{-|\gamma|}$ $\Rightarrow$ Δ_{12} csökken

3) FTG. e^{-x} $\gamma = 0$ $a_N = \text{const} \Rightarrow \Delta_{12}$ konstans

e^{-x^2} $\gamma = 0$ $a_N = \frac{1}{\sqrt{2 \ln N}}$ $\Rightarrow$ Δ_{12} csökken.

⋮

$e^{-x^{\delta}}$ $\Delta \sim \begin{cases} \to 0 & \delta > 1 \\ \text{konst} & \delta = 1 \\ \to \infty & \delta < 1 \end{cases}$

vagy a formula: $a_N \sim [\ln N]^{-1+\frac{1}{\delta}}$

Sorrend statisztikák

A k-ik érték fölött van k-1 másik, alatta pedig N-k.

$$z_k = k^{th} \max \{y_1, y_2, \cdots y_N\}$$

$$P_k(z) = \frac{N!}{(N-k)!(k-1)!} (\mu(z))^{N-k} \frac{d\mu}{dz} [1-\mu(z)]^{k-1} =$$

$$= \frac{N!}{(N-k)!(k-1)!} (\mu(z))^{N-1} \frac{d\mu}{dz} \left(\frac{1}{\mu(z)} - 1\right)^{k-1} \approx \quad k \ll N$$

$$\approx \frac{N^k}{(k-1)!} (\mu(z))^{N-1} \frac{d\mu}{dz} \left(\frac{1}{\mu(z)} - 1\right)^{k-1}$$

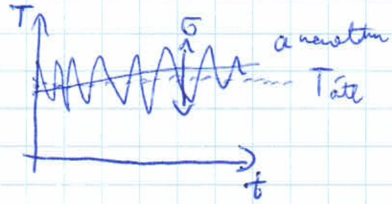
végig az órán, de a eredmény

$$P_k(z) = \frac{1}{(k-1)!} (1+\gamma x)^{-\frac{k}{\gamma}-1} e^{-(1+\gamma x)^{-1/\gamma}}$$

EXTRÉM STAT

6. előadás (11.08.)

Kimaradt a témából: Hogyan állapítsuk meg, hogy egy adatsorban van-e trend?



Generáljunk $T_{átl}$ körül σ szórású eloszlást, és illesszünk erre is a-kat

Ha $|a_{fit}| > \sqrt{\overline{a^2}}$, akkor az egy hihető trend.

Kérdés, hogy másképp állapítsuk meg, melyik a legnagyobb.

$$T_1 = \frac{\sqrt{\langle (z_1 - \langle z_1 \rangle)^2 \rangle}}{\langle z_1 - z_2 \rangle} = \frac{\sqrt{\langle x_1^2 \rangle - \langle x_1 \rangle^2}}{\langle x_1 - x_2 \rangle}$$

Gumbelre kiszámoltuk, hogy $\langle x_1 \rangle = \gamma_E$

$$\langle x_1^2 \rangle = \frac{\pi^2}{6} + \gamma_E^2$$

$$\Rightarrow T_1^{FTG} = \frac{\pi}{\sqrt{6}} \simeq 1{,}3$$

a plussz lap 5. ábráján látjuk, hogy nem a közelében van.

$\Rightarrow$ valószínűleg nem FTG eloszlásból van az adat.

NEM, mert a többi még nagyobb, mint 1,3, tehát a gond az, hogy a galaxisok fénye nem független eloszlásból származik.

$$T_2 = \frac{\sqrt{\langle (x_1 - x_2)^2 \rangle - (\langle x_1 - x_2 \rangle)^2}}{\Delta_{12}} = \frac{\sqrt{\langle \Delta_{12}^2 \rangle - \langle \Delta_{12} \rangle^2}}{\langle \Delta_{12} \rangle} = ?$$

$$P(z_1, z_2)\, dz_1\, dz_2 = N(N-1)\, F^{N-2}(z_2)\, P_0(z_2)\, dz_2\, P(z_1)\, dz_1\, \Theta(z_1 - z_2) =$$

$$= N(N-1)\, F^{N-2}(z_2) \left.\frac{dF}{dz}\right|_{z_2} dz_2 \left.\frac{dF}{dz}\right|_{z_1} dz_1\, \Theta(z_1 - z_2)$$

$$P(x_1, x_2)\, dx_1\, dx_2 = N(N-1) \frac{M_1(x_2)}{F^2(x_2)} \left.\frac{dF}{dx}\right|_{x_2} dx_2 \left.\frac{dF}{dx}\right|_{x_1} dx_1\, \Theta(x_1 - x_2)$$

Tudjuk, hogy $M_1(x) = e^{-(1+\gamma x)^{-1/\gamma}}$

$$\Rightarrow \quad \mu(x) = e^{-\frac{(1+\gamma x)^{-1/\gamma}}{N}} = 1 - \frac{(1+\gamma x)^{-1/\gamma}}{N} \to 1$$

$$\frac{d\mu}{dx} = \frac{(1+\gamma x)^{-\frac{1}{\gamma}-1}}{N}$$

$$P(x_1, x_2) \to M_1(x_2) \, (1+\gamma x_2)^{-\frac{1}{\gamma}-1} \, (1+\gamma x_1)^{-\frac{1}{\gamma}-1} \, \Theta(x_1 - x_2) =$$

$$= P_1(x_2) (1+\gamma x_1)^{-\frac{1}{\gamma}-1} \, \Theta(x_1 - x_2)$$

$$P^{\gamma=0}(x_1, x_2) = e^{-x_2 - e^{-x_2}} e^{-x_1} \, \Theta(x_1 - x_2)$$

$$\tilde{P}(\Delta) = \int dx_1 dx_2 \, P(x_1, x_2) \, \delta(x_1 - x_2 - \Delta) = \int dx_1 dx_2 \, P_1(x_2) (1+\gamma x_1)^{-\frac{1}{\gamma}-1} \Theta(x_1 - x_2) \, \delta(x_1 - x_2 - \Delta)$$

= a Θ miatt az csak $\Delta > 0$-ra van értelmes, de végig ezt tesszük.

$$= \int dx_1 \, P_1(x_1 - \Delta) \, (1+\gamma x_1)^{-\frac{1}{\gamma}-1}$$

$$\tilde{P}^{\gamma=0}(\Delta) = \int dy \, P_1(y) \underbrace{\left[1+\gamma(y+\Delta)\right]^{-\frac{1}{\gamma}-1}}_{e^{-y-\Delta}} = e^{-\Delta} \underbrace{\int dy \, P_1(y) \, e^{-y}}_{\tilde{P} \text{ normáltsága miatt ez tuti } 1.^{*}} = e^{-\Delta}$$

*: nem csak azért is, mert ez a 2. legnagyobbak értékek eloszlása.

$$T_2^{\gamma=0} = \frac{\sqrt{2-1}}{1} = 1$$

EXTRÉM STAT

7. előadás (11.15.)

sorrendstatisztika: $P(y)$-ból húzok N-et, mekkora lesz a legnagyobb, a 2. legnagyobb, ... a k-ik legnagyobb?

TFH iid-változók!

$P_0(z)\,dz$ a sűrűség, $\mu(z)=\int_{-\infty}^{z} P_0(y)\,dy$ az eloszlás. $z_k = a_N x_k + b_N$

$$P_k(z_k)\,dz_k = \frac{N!}{(N-k)!\,(k-1)!}\left(\mu(z_k)\right)^{N-k} P_0(z_k)\,dz_k\left[1-\mu(z_k)\right]^{k-1} =$$

$$= \frac{N!}{(N-k)!\,(k-1)!}\left(\mu(z_k)\right)^{N-k} \left.\frac{d\mu}{dz}\right|_{z_k} dz_k\left[1-\mu(z_k)\right]^{k-1} =$$

Tudjuk, hogy $M_1(z)=(\mu(z))^N$ ahol $M_1(z)\xrightarrow[N\to\infty]{}$ GEV

$$P_k(x_k)\,dx_k = \frac{N!}{(N-k)!\,(k-1)!}\underbrace{\frac{(\mu(z_k))^N}{(\mu(z_k))^k}}\left.\frac{d\mu}{dx}\right|_{x_k} dx_k\left[1-\mu(z_k)\right]^{k-1}$$

$$\rightarrow \frac{M_1(x_k)}{\mu^k(x_k)} = M_1(x_k) \quad \text{mert } \mu(x)=e^{-\frac{(1+\gamma x)^{-1/\gamma}}{N}} \approx 1-\frac{(1+\gamma x)^{-1/\gamma}}{N}\to 1$$

és feltesszük, hogy $k \ll N$.

$$\frac{d\mu}{dx}\approx\frac{1}{N}(1+\gamma x)^{-\frac{1}{\gamma}-1}$$

$$(1-\mu(z_k))^{k-1} \approx \frac{(1+\gamma x)^{-\frac{k-1}{\gamma}}}{N^{k-1}}$$

$$P_k(x_k)\,dx_k = \frac{N!}{(N-k)!\,(k-1)!} M_1(x_k)\frac{1}{N^k}(1+\gamma x)^{-\frac{k}{\gamma}-1}\,dx_k =$$

$$= \frac{1}{(k-1)!}\underbrace{\frac{N(N-1)\cdots(N-k+1)}{N^k}}_{\approx 1} M_1(x_k)(1+\gamma x)^{-\frac{k}{\gamma}-1}\,dx_k =$$

$$= \frac{1}{(k-1)!}(1+\gamma x)^{-\frac{k}{\gamma}-1} e^{-(1+\gamma x)^{-1/\gamma}}\,dx$$

$\gamma\to 0$ esetén: $P_k(x_k)\,dx_k = \frac{1}{(k-1)!}e^{-kx_k-e^{-x_k}}\,dx_k$

$\langle x_k\rangle \approx k^{(0)}$

Példa:

Legyen QM-ben $V(x) = |x|^{\alpha}$ potenciál.

Ekkor a E-spektrum: $\varepsilon_k = (hk)^{\frac{2\alpha}{\alpha+2}}$

$\Rightarrow$ Weibull-eloszlás lesz $|\gamma| = \frac{2\alpha}{\alpha+2}$ gammával.

Végesméret korrekciók

N pontot mintavételezve, mikor tér el a valószínűségi eloszlás a GEV-től?

$$P(z,N) = P(a_N x + b_N, N) \xrightarrow{N\to\infty} P_1(x)$$

$$= P_1(x) + \underbrace{q(N)}_{\to 0}\,\Phi_1(x) \quad \leftarrow \text{TFH így felírható}$$

közelít, elmélet

$$\frac{P_1(x,N) - P_1(x)}{q(N)} \to \Phi_1(x)$$

$q(N)$-t nem ismerjük, de egy adatbázist vizsgálva megállapították, hogy milyen kell legyen, hogy $\Phi_1(x)$ azt kapjuk, mert ekkor $P_1(x,N) - P_1(x)$ ugyanazt az x-függvényt mutatják, csak a normálás más különböző N-ekre.

Tapasztalat: exp esetén $q(N) = \frac{1}{N}$

hogy kapjuk ezt meg elméletileg?

$$M_1(z,N) = (1-e^{z})^N = (1-e^{a_N x + b_N})^N$$

tehát $z = a_N x + b_N$

ahol $\langle x \rangle = 0 \Rightarrow b_N = \langle z \rangle_N$

$\langle x^2 \rangle = 1 \Rightarrow \sigma^2 = a_N^2$

EXTRÉM STAT

8. előadás (11. 22.)

A relaxáció határeloszlásban $z \to a_N x + b_N$

$$P(x,N) = P(x) + \underbrace{q(N)}_{\xrightarrow[N\to\infty]{} 0}\,\underbrace{\Phi(x)}_{\text{alak korrekció}}$$

$\Rightarrow$ valamilyen N-függő $q(N)$-re: $\dfrac{P(x,N)-P(x)}{q(N)} = \Phi(x)$

hogyan tudjuk meg $q(N)$-t?

TFH a változókat úgy választottuk, hogy $\langle x\rangle = 0$ és $\langle x^2\rangle = 1$

$$\Rightarrow \quad b_N = \langle z\rangle \quad \text{és} \quad a_N = \sigma(z)$$

Gumbel-osztályra számoljuk végig, hogyan közelít ezt!

- $P_0(y) = e^{-y} \qquad \mu(z) = 1 - e^{-z} \qquad M(z,N) = \left(1-e^{-z}\right)^N$

$$\Rightarrow P(z,N) = N e^{-z}\left(1-e^{-z}\right)^{N-1}$$

$$\langle z\rangle = \int_0^\infty z N e^{-z}\left(1-e^{-z}\right)^{N-1} dz \qquad u := z - \ln N$$

$$= \int_{-\ln N}^{\infty} (u+\ln N)\, e^{-u}\left(1-\frac{e^{-u}}{N}\right)^{N-1} du =$$

$$= \ln N + \int_{-\ln N}^{\infty} u e^{-u}\, e^{\overbrace{(N-1)\ln\left(1-\frac{e^{-u}}{N}\right)}} du \qquad \ln(1-x) = x - \frac{x^2}{2} + \mathcal{O}(x^3)$$

$$(N-1)\left(-\frac{e^{-u}}{N} - \frac{e^{-2u}}{2N^2}\right) = -e^{-u} + \frac{e^{-u}}{N} - \frac{e^{-2u}}{2N}$$

$$= \ln N + \int_{-\ln N}^{\infty} u e^{-u-e^{-u}} du + \frac{1}{N}\int_{-\ln N}^{\infty} u e^{-u-e^{-u}}\left(e^{-u} - \frac{e^{-2u}}{2}\right)$$

ha $u < -\ln N$, az integrandus kicsi, ezért integrálhatunk $-\infty$-ig

$$= \ln N + \gamma_E + \frac{1}{N}\Big[\underbrace{(\gamma_E - 1)}_{\langle u_2\rangle} - \underbrace{\left(\gamma_E - \frac{3}{2}\right)}_{\langle u_3\rangle}\Big] = \underbrace{\ln N + \gamma_E}_{\text{határérték tagja}} + \frac{1}{2N}$$

$$\Downarrow$$

$$q(N) \sim \frac{1}{N}$$

$$\langle z^2 \rangle = \int_0^\infty z^2 N e^{-z} (1-e^{-z})^{N-1} \, dz \quad = \text{u.a. lépések}$$

$$= \frac{\pi^2}{6} \cdots$$

$$\langle z^2 \rangle - \langle z \rangle^2 = \frac{\pi^2}{6} - \frac{1}{N} \qquad \Rightarrow \quad a_N \simeq \frac{\pi}{\sqrt{6}} - \frac{\sqrt{6}}{\pi} \frac{1}{2N}$$

A műveletsor-átrendezés után:

$$P(x,N) = N e^{-\left(\frac{\pi}{\sqrt{6}} - \frac{\sqrt{6}}{\pi}\frac{1}{2N}\right) x - \ln N - \gamma_E + \frac{1}{2N}} \left(1 - e^{-a_N x - b_N}\right)^{N-1}$$

$$= \underbrace{e^{-\frac{\pi}{\sqrt{6}} x - \gamma_E - e^{-\frac{\pi}{\sqrt{6}} x - \gamma_E}}}_{P(x)} + \frac{1}{N} \phi_{exp}(x)$$

ahol
$$\phi_{exp}(x) = \frac{1}{2} P(x) \left[-\frac{6}{\pi^2} - 1 + \frac{\sqrt{6}}{\pi} x + \left(3 - \frac{\sqrt{6}}{\pi}\right) e^{-\frac{\pi}{\sqrt{6}} x - \gamma_E} - e^{-\frac{2\pi}{\sqrt{6}} - 2\gamma_E} \right]$$

- $P_0(y) = e^{-y^2}$ ez még bonyolultabb

  $P(x,N) = P(x) + \frac{1}{\ln N} \phi_G(x)$ $\qquad \phi_G$ a független.

- $P_0(y) = \frac{e^{-x^b}}{x^a}$

  $P(x,N) = P(x) + (\qquad) \phi_G(x)$

  esetek: $a = b = 1$ $\qquad q(N) = \frac{1}{(\ln N)^2}$

  a többit nem mondta ☺

EXTRÉM STAT

9. előadás (11. 29.)

Korrelált változók:

Legyen $y_1, y_2, \dots y_N$ N db változó!

Kiírtuk a korrelációt: $C(i,j) = \langle y_i y_j \rangle - \langle y_i \rangle \langle y_j \rangle$

A fluktuáció: $\langle (\delta Y)^2 \rangle = \sum_{i=1}^{N} (\langle y_i^2 \rangle - \langle y_i \rangle^2) = N \langle (\delta y)^2 \rangle$

$$Y = \sum_{i=1}^{N} y_i$$

Ekkor $Y \sim N$, $\frac{\sqrt{\langle (\delta Y)^2 \rangle}}{\langle Y \rangle} \sim \frac{1}{\sqrt{N}}$ $\Rightarrow$ „extenzív változó"

TFH: $C(i \neq j, j) = Q$

$\Rightarrow \langle (\delta Y)^2 \rangle = N \langle (\delta y_1)^2 \rangle + N(N-1) Q$

$\Rightarrow N \gg 1$ -re $\langle (\delta Y)^2 \rangle \sim N^2$ $\Rightarrow$ $\frac{\sqrt{\langle (\delta Y)^2 \rangle}}{\langle Y \rangle} \sim O(1)$

TFH: $C(i,j) = \phi(|i-j|)$

$\Rightarrow \langle (\delta Y)^2 \rangle = N \langle (\delta y_1)^2 \rangle + 2N \sum_{n=1}^{N/2} \phi(n)$

ha a $\phi(n)$-nek integrálja véges, $\langle (\delta Y)^2 \rangle \sim N$

tehát olyan, mint a korrelálatlan rendszer „gyenge korreláció"

ennek feltétele, hogy $\phi(n)$ elég gyorsan csengjen le.

ha $\phi(n) \sim 1/n^{-1+\eta}$ $\Rightarrow$ $\langle (\delta Y)^2 \rangle \sim N^{1+\eta}$ $\Rightarrow$ $\frac{\sqrt{\langle (\delta Y)^2 \rangle}}{\langle Y \rangle} \sim N^{-\frac{1}{2}+\frac{\eta}{2}}$

ha $\eta > 1$ $\Rightarrow$ a fluktuáció egyre nagyobb

védekezés kérdés, hogy mi van erős korrelációnál, de vizsgáljuk előbb a random walk-ot.

Random walk: véletlen lépések fel vagy le.

$\Rightarrow$ A lépés korrelálatlan, de a pozíció nem. $\Rightarrow$ erős korreláció

$\Delta x_i = a e_i$ $e_i = \pm 1$ $P(e_i = 1) = P(e_i = -1) = \frac{1}{2}$

$C(i,j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$

$x_i = \sum_{k=1}^{i} a e_k$ $\langle x_i \rangle = a \sum_{k=1}^{i} \langle e_k \rangle = 0$

$$\langle x_i^2 \rangle = a^2 \sum_{k=1}^{i} \sum_{l=1}^{i} \langle e_k e_l \rangle = \text{kezdetben} = a^2 i$$

$$\langle x_i x_j \rangle = a^2 \sum_{k=1}^{i} \sum_{l=1}^{j} \langle e_k e_l \rangle = a^2 \min(i,j)$$

$$\langle X \rangle = \sum_{k=1}^{N} \langle x_k \rangle = 0$$

$$\langle X^2 \rangle = \sum_{k=1}^{N} \sum_{l=1}^{N} \langle x_k x_l \rangle = a^2 \sum_{k=1}^{N} \sum_{l=1}^{N} \min(k,l)$$

- diagonális elemek: $a^2 \sum_{k=1}^{N} k = a^2 \frac{N(N+1)}{2}$
- nem diag. elemek: $(k<l)$ $\quad a^2 \sum_{k=1}^{N} \underbrace{k}_{\min(k,l)} \underbrace{(N-k)}_{\# l>k} = a^2 \frac{N(N+1)(N-1)}{6} \approx a^2 \frac{N^3}{6}$

$$\Rightarrow \langle X^2 \rangle = a^2 \frac{N^3}{3} \Rightarrow \text{erős korreláció}$$

Kérdés: Mi lesz $\max_x (x-x_0)$ eloszlása?

$$x_{max}(t) = \max_t (x(t)-x_0)$$

$M(x<z|x_0,t)$: eloszlás fv. az a valség, hogy t-ig a válto nem lépi át z-t.

$$M(x<z|x_0,t) = \int_{-\infty}^{z} P_z(x,t|x_0,0)\,dx$$

P-t a diffúziós egyenlet adja meg:

$$\partial_t P_z(x,t|x_0,0) = D\,\partial_x^2 P_z(x,t|x_0,0)$$

kezdőfeltétel nélkül: $P_z(x,t|x_0) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{(x-x_0)^2}{4Dt}}$

A határfeltétel: $P_z(z,t|x_0,0) = 0$, hiszen nem lépheti át z-t.

ez kielégíthető, ha hozzáadunk egy tükörmegoldást:

$$-e^{-\frac{(x+x_0-2z_0)^2}{4Dt}}$$

$$P_z(x,t|x_0,0) = \frac{1}{\sqrt{4\pi D t}} \left[e^{-\frac{(x-x_0)^2}{4Dt}} - e^{-\frac{(x+x_0-2z_0)^2}{4Dt}} \right]$$

$$P(z|x_0,t) = \frac{\partial M(x<z|x,t)}{\partial z} = \underbrace{P_z(x,t|x_0,0)\Big|_{x=z}}_{0} + \int_{-\infty}^{z} \frac{\partial P_z(x,t|x_0,0)}{\partial z}\,dx =$$

$$= \frac{1}{\sqrt{\pi D t}} e^{-\frac{(x_0-z)^2}{4Dt}} \qquad z > x_0$$

EXTRÉM STAT

10. előadás (12.06.)

Rekord statisztikák

$y_0, \dots y_{n-1}$ i.i.d változók. y_n egy rekord, ha $y_n > \max\{y_0, \dots y_{n-1}\}$

Mi a valség annak, hogy y_n egy rekord? Hány db rekord van egy n-méretű adathalmazban?

$$q_i = \begin{cases} 1 & y_i \text{ rekord} \\ 0 & y_i \text{ nem rekord} \end{cases}$$

$$R_n = \sum_{i=0}^{n} q_i$$

$$P_i = \Pr\{q_i = 1\}$$

Ha csak annyit tudunk, hogy i.i.d-k a változók, akkor $P_n = \frac{1}{n+1}$

$$\langle R_n \rangle = \sum_{i=0}^{n} P_n = \sum_{i=0}^{n} \frac{1}{i+1} \to \ln n + \gamma_E$$

- Mi van, ha rövid korreláció van az adatban?

  Állítás: egy n-ben az $\langle R_n \rangle$ u.a. az, mert ha $\ell > \xi$ távolságra vizsgálódunk, ahol ξ a korrelációs hossz, akkor y_n és $y_{n-\ell}$ független lesz, tehát az ℓ-lel ezelőtti rekord megdöntése u.a.

  DE a röviddel ezelőtti rekord megdöntése változik

  $\Rightarrow \langle R_n \rangle = A \ln n + B$.

- Mi van, ha változók függetlenek, de az eloszlás változik?

  pl.: • normál eloszlás, de $\sigma_i \sim i$ (szélesedő tartományból) de nincs trend.

  $\langle R_n \rangle \sim (\ln n)^2$

  • van trend, mert $\langle x_i \rangle \sim i$

  $\langle R_n \rangle \sim n$

  • van erős korreláció

  random walknál $\langle R_n \rangle \sim \sqrt{n}$