Cramér nagy eltérés tétele (H. Cramér, 1938)

Def: Egy $X \in \mathbb{R}$ valószínűségi változó **momentum-generáló függvénye** $M: \mathbb{R} \to [0, \infty]$, amire

$$M(\lambda) \stackrel{\text{def}}{=} \mathbb{E}(e^{\lambda X}) \qquad \forall \lambda \in \mathbb{R}.$$

Tulajdonságok:

0.) $e^{\lambda X} \geq 0$, ezért $M(\lambda) = \mathbb{E}(e^{\lambda X})$ mindig létezik, de előfordulhat, hogy $= \infty$. Ilyenkor kevésbé hasznos.

1.) $M(0) = \mathbb{E}(e^{0 \cdot X}) = 1$ mindig.
Az összes többi λ-ra viszont sajna lehet $M(\lambda) = \infty$

2.) Ha viszont $M(\lambda) < \infty$ a $\lambda = 0$ egy környezetében $\Rightarrow$ (mondjuk $\lambda \in (-\varepsilon, \varepsilon)$-ra), akkor $M(\lambda)$ ~~nagyon szép~~ fv.

a.) akárhányszor differenciálható

b.) $M'(\lambda) = \left[\mathbb{E}(e^{\lambda X})\right]' = \mathbb{E}(X e^{\lambda X}) \Rightarrow M'(0) = \mathbb{E}X$

$M''(\lambda) = \mathbb{E}(X^2 e^{\lambda X}) \qquad M''(0) = \mathbb{E}(X^2)$

$\vdots \qquad\qquad \vdots$

$M^{(k)}(\lambda) = \mathbb{E}(X^k e^{\lambda X}) \qquad M^{(k)}(0) = \mathbb{E}(X^k)$

ezért hívják momentum-generáló függvénynek

c.) Ha X és Y momentum-generáló függvénye azonos, akkor X és Y azonos eloszlású. Avagy: $M(\lambda)$ karakterizálja az eloszlást (teljes információt hordoz róla).

3.) Ha ~~$X \in \mathbb{R}$~~ $X \in \mathbb{N}$ nemnegatív egész értékű, akkor

$z := e^{\lambda}$ $\lambda := \ln z$ helyettesítéssel

$$g_X(z) = \mathbb{E}(z^X) = \mathbb{E}(e^{\lambda X}) = M(\lambda),$$

vagyis az $M(\lambda)$ ugyanaz, mint a generátorfüggvény, csak átskálázva. Persze ugyanarra jók, csak kicsit más nyelven.

⋮

és még sok minden egyéb, ami most nekünk nem kell.

Tétel (Cramér nagy eltérés tétel) ELSŐRE IJESZTŐ.

Legyen $X_1, X_2, X_3 \ldots$ független, azonos eloszlású val. változó,

$S_n = X_1 + \ldots + X_n$.

Tfh az X_i-k közös momentum-generáló függvénye, $M(\lambda)$ véges a $\lambda = 0$ egy környezetében.

Ekkor $\exists$ egy $I : \mathbb{R} \to [0, \infty]$ rátafüggvény, hogy

$\forall\ a < b \in [-\infty, \infty]$-re

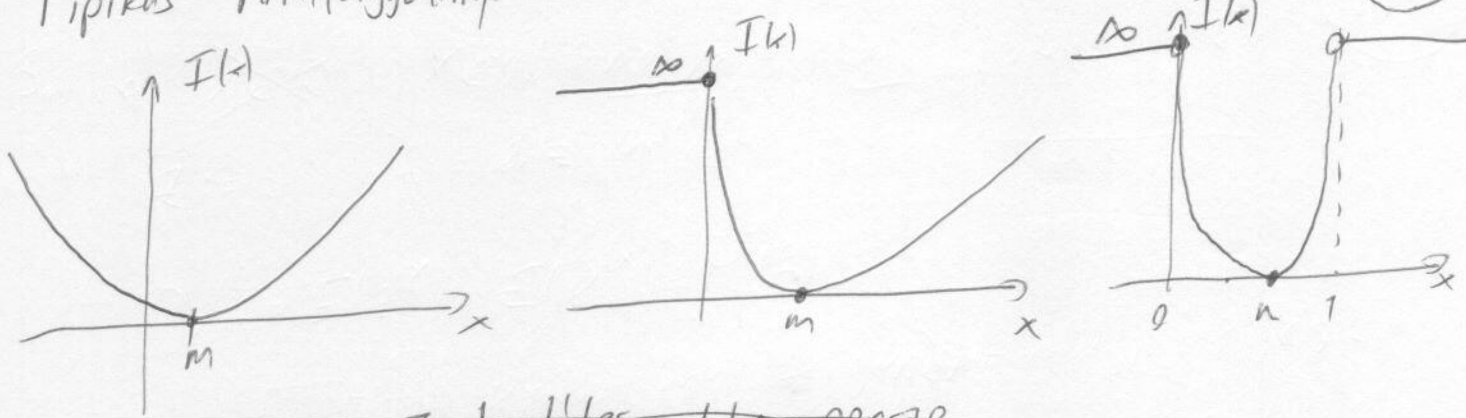
$$\mathbb{P}\left(\frac{S_n}{n} \in [a,b] \right) \lessapprox e^{-n \inf\limits_{x \in [a,b]} I(x)}$$

Továbbá $I(x)$ olyan, hogy

- ~~konv~~ konvex
- egyetlen minimuma $x = m := \mathbb{E}X$-ben $I(m) = 0$, minden más $x \neq m$-re $I(x) > 0$

Továbbá $I(x)$ számolására van recept – lásd később.

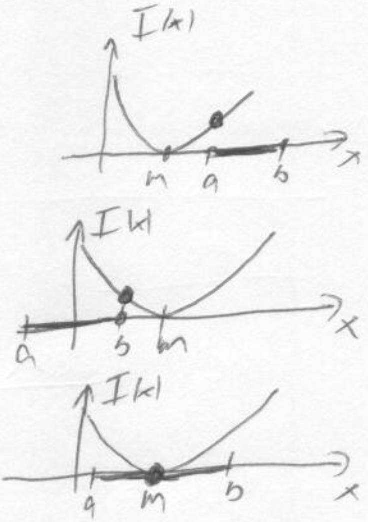
Tipikus rátafüggvény:



~~Pl. ha X ∈ [0,1] korlátos, akkor persze~~

<u>Köv</u>: A tételben lévő ~~m~~ rémisztő $\inf_{x \in [a,b]} I(x)$ képlettől nem kell megijedni: az $I(x)$ <u>legkisebb értékét</u> jelenti az $[a,b]$ intervallumon. Mivel $I(x)$ m-től távolodva csak nőni tud, ez nagyon könnyű:

$$\inf_{x \in [a,b]} I(x) = \begin{cases} I(a), & \text{ha } a > m \\ I(b), & \text{ha } b < m \\ 0, & \text{ha } a \le m \le b \end{cases}$$



<u>Avagy</u> (Cramér tétel, recept-változat):

$$\mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) \lessapprox \begin{cases} e^{-nI(a)}, & \text{ha } a > m \\ e^{-nI(b)}, & \text{ha } b < m \\ 1, & \text{ha } a \le m \le b \end{cases}$$

A rátafüggvény szemléletes jelentése:

1.) $I(x)$ az utálat mértéke, amit az $\frac{S_n}{n}$ átlag ~~bze~~ érez az $x \in \mathbb{R}$ szám iránt:

a.) $m = EX$ az egyetlen érték, amit az átlag szeret felvenni (vagyis nem utálja): $I(m) = 0$.

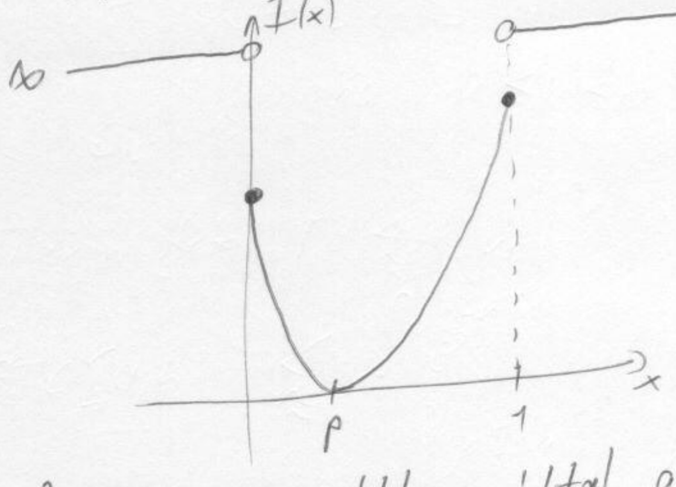
Minden más $x \neq m$-et utál: $I(x) > 0$.

b.) Minél jobban távolodik x az m-től, ~~am~~ annál jobban utálja az átlag felvenni: m-től távolodva $I(x)$ nő.

c.) ~~M~~ Ha pl. X felülről korlátos, $X \leq 1$, akkor az átlag ~~egy~~ egyáltalán nem lehet 1-nél nagyobb:

$x > 1$-re $I(x) = \infty$ (végtelenül utálja)

Pl ha $X \sim B(p)$ Bernoulli, akkor $0 \leq X \leq 1$, és



2.) Így $\inf_{x \in [a,b]} I(x)$ az átlag által az $[a,b]$ intervallum iránt érzett utálat mértéke: mennyire utálja az $[a,b]$ intervallum „legelviselhetőbb" pontját.

Megj.:

1.) Az is igaz, hogy ha már az $\frac{S_n}{n}$ átlag környezetében $[a,b]$-be esni (amit utálna, ha $m \notin [a,b]$) akkor azon belül a legkevésbé utált értéket fogja nagy val.séggel felvenni: ha $[a,b]$-n belül a minimum x-ben van, akkor

$$\mathbb{P}\left(\frac{S_n}{n} \approx x \,\middle|\, \frac{S_n}{n} \in [a,b]\right) \text{ nagy.}$$

2.) Ha $a < m < b$, akkor a Cramér tétel állítása, miszerint

$$\mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) \approx 1$$ éppen a nagy számok törvénye.

3.) Az $a > m$ és a $b < m$ eset pont a nagy eltérések esete: ilyenkor $\mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right)$ kicsi, ha n nagy.

4.) A tétel szigorú felső becslést $(\leq)$ és együttal jó közelítést $(\approx)$ is ad.

5.) A $\approx$ pontos jelentése: Ha $r \gneq \inf_{x \in [a,b]} I(x) =: r_0$ és n elég nagy, akkor ~~$\mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) > e^{-nr}$~~

$$e^{-nr} < \mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) \leq e^{-nr_0}$$

6.) A tételhez feltétel volt, hogy $M(\lambda) < \infty$ a $\lambda = 0$ egy környezetében. Ebből következik pl. ~~az~~ hogy $\mathbb{E}(X^k) < \infty$ minden k-ra (minden momentum véges), és nem is ~~nőnek~~ nőnek túl gyorsan.

6.) felyt. Még így is: ez egy viszonylag enyhe feltétel: minden nevezetes eloszlásra teljesül, ami az ismétlésben szerepelt.

Recept a rátafüggvény kiszámolásához

1.) $M(\lambda) := E(e^{\lambda X})$ a momentum-generáló függvény: számold ki!

[ez egy szumma, vagy egy integrál – többnyire könnyű]

2.) $\hat{I}(\lambda) := \ln M(\lambda)$ a logaritmikus momentum-generáló függvény

[az előző logaritmusa, nagyon könnyű]

3.) Számold ki az $\hat{I}'(\lambda)$ deriváltat [általában könnyű]

4.) Minden x-re oldd meg az $\hat{I}'(\lambda) = x$ egyenletet λ-ra

[Egyenlet-megoldás: általában NEHÉZ!!]

A megoldást jelöljük ~~$\lambda^*(x)$-szel~~! $\lambda^* = \lambda^*(x)$-szel!

5.) Ez után $\boxed{I(x) = x\,\lambda^*(x) - \hat{I}(\lambda^*(x))}$

[képlet-behelyettesítés, nagyon könnyű]

Rossz hír: ez az eljárás macerás, és általában nehéz (az egyenlet-megoldás miatt).

Jó hír: ZH-ban nem kell rátafüggvényt számolni: ha kell, meg lesz adva.

Példa rátafüggvény számolására

Legyen $X \sim B(p)$, vagyis

k	0	1
$\mathbb{P}(X=k)$	q	p

$q := 1-p$

1.) $M(\lambda) = \mathbb{E}(e^{\lambda X}) = q e^{\lambda \cdot 0} + p e^{\lambda \cdot 1} = q + p e^{\lambda}$

2.) $\hat{I}(\lambda) = \ln M(\lambda) = \ln(q + p e^{\lambda})$

3.) $\hat{I}'(\lambda) = \dfrac{p e^{\lambda}}{q + p e^{\lambda}}$

4.) Oldjuk meg az $\hat{I}'(\lambda) = x$ egyenletet:

$$\frac{p e^{\lambda}}{q + p e^{\lambda}} = x \qquad / \cdot (q + p e^{\lambda})$$

$$p e^{\lambda} = x q + x p e^{\lambda} \qquad / - x p e^{\lambda}$$

$$p e^{\lambda}(1-x) = x q \qquad / \div p(1-x)$$

$$e^{\lambda} = \frac{x q}{(1-x) p} \qquad / \ln [\ ]$$

$\lambda = \ln \dfrac{x q}{(1-x) p}$, nevezzük el $\lambda_*^*(x)$-nek.

$$\lambda_*^*(x) = \ln \frac{x q}{(1-x) p}$$

5.) $$\boxed{I(x)} = x \lambda_*^*(x) - \hat{I}(\lambda_*^*(x)) = x \ln \frac{x q}{(1-x) p} - \ln\left(q + p e^{\overbrace{\ln \frac{x q}{(1-x) p}}^{\frac{x q}{(1-x) p}}}\right)$$

$$= x \underbrace{\ln \frac{x q}{(1-x) p}}_{\ln \frac{x}{p} - \ln \frac{1-x}{q}} \underbrace{- \ln \Bigl(\underbrace{q + \frac{x q}{1-x}}_{\frac{q}{1-x}} \Bigr)}_{-\ln \frac{1-x}{q}} = \ldots = \boxed{x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{q}}$$

Példa a Cramér tétel alkalmazására

Egy véletlen algoritmus egy eldöntendő kérdésre („Igen/nem")
90% valószínűséggel ad helyes választ. Mivel ezt keveselljük,
lefuttatjuk az algoritmust 1000-szer, EGYMÁSTÓL FÜGGETLENÜL, megnézzük, hogy melyik
választ jön ki többször, és azt fogadjuk el.

Adjunk nagy eltérés
becslést annak valószínűségére, hogy végül ~~hibás~~ baj lesz:

baj van, ha → a végén hibás választ fogadunk el, vagy
→ az eredmény döntetlen (500−500), így
nem tudunk dönteni.

Megoldás: Legyen $n = 1000$, és $i = 1, 2, \ldots, n$-re legyen

$$X_i = \begin{cases} 1, & \text{ha az } i\text{-edik futás eredménye helyes} \\ 0, & \text{ha nem.} \end{cases}$$

[Az eredményt megnézve persze nem tudjuk, hogy helyes-e vagy
sem, de sebaj: attól még X_i értelmes.]

Így $S_n := X_1 + \ldots + X_n$ a helyes válaszok száma
[nem keverendő az „igen" válaszok számával].

$$\mathbb{P}(\text{baj van}) = \mathbb{P}(S_n \le 500) = \mathbb{P}\left(S_n \le \frac{n}{2}\right) = ?$$

~~A~~ A Cramér tétel nyelvére átfogalmazva:

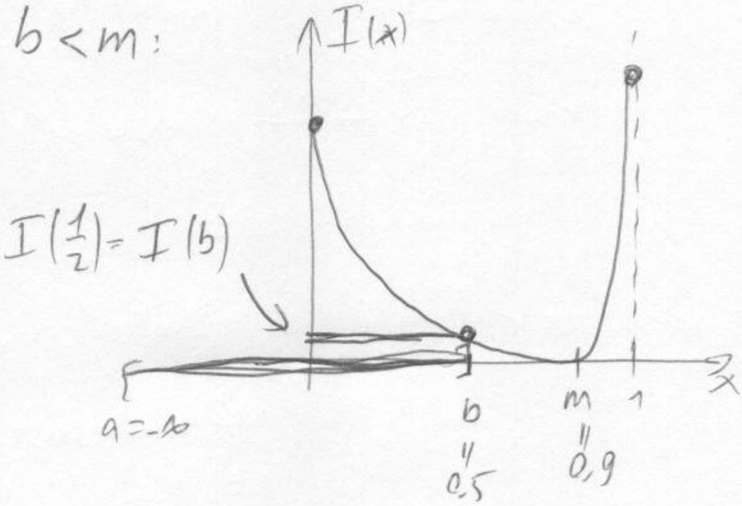
$$\mathbb{P}(\text{baj van}) = \mathbb{P}(S_n \le 500) = \mathbb{P}\left(\frac{S_n}{n} \le \frac{500}{n}\right) = \mathbb{P}\left(\frac{S_n}{n} \le \frac{1}{2}\right)$$

$$= \mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) = ? \quad \text{ahol } a = -\infty,\ b = \frac{1}{2},$$

és $X_1, \dots, X_n \sim B(p)$ függetlenek, $p = 0{,}9$

Mivel $m = p = 0{,}9$, esetünkben $b < m$:

a várható értéktől <u>lefelé</u> való nagy eltérés val. ségét keressük.



A Cramér tétel szerint

$$\mathbb{P}(\text{baj van}) = \mathbb{P}\left(\frac{S_n}{n} \in [a,b]\right) \lesssim e^{-n I(b)} = e^{-1000\, I(\frac{1}{2})},$$

ahol $I(x)$ a ~~...~~ a $B(p)$ eloszlás Cramér féle rátafüggvénye: ($p = 0{,}9$)

$$I(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{q} \qquad p = 0{,}9 \quad q = 0{,}1$$

$$\Rightarrow I\left(\frac{1}{2}\right) = \frac{1}{2} \ln \frac{1}{2p} + \frac{1}{2} \ln \frac{1}{2q} = \frac{1}{2} \ln \frac{1}{4pq} = \underline{\underline{-\ln \sqrt{4pq}}}$$

Esetünkben $I\left(\frac{1}{2}\right) = -\ln \sqrt{4 \cdot 0{,}1 \cdot 0{,}9} \approx 1.02165124$ (NE KEREKÍTS!!)

$$\Rightarrow \mathbb{P}(\text{baj van}) \lesssim e^{-1000 \cdot 1.02165124} = e^{-1021.65124} \approx 2 \cdot 10^{-444} \quad \text{(kicsi).}$$

<u>Megj.:</u> Ugyanerre a feladatra alkalmazható a Hoeffding-egyenlőtlenség is: $a_i = 0 \le X_i \le b_i = 1$, $t = 400$

választással $\sum_{i=1}^{n} (b_i - a_i)^2 = 1000 \cdot (1-0)^2 = 1000$

$$\Rightarrow \mathbb{P}(\text{baj van}) = \mathbb{P}(S_n \le 500) \underset{t=400}{\overset{\mathbb{E}S_n = 900}{=\!=}} \mathbb{P}(S_n \le ES_n - t) \le \exp\left(-\frac{2t^2}{1000}\right)$$

vagyis

$$\mathbb{P}(\text{baj van}) \le e^{-\frac{2 \cdot 400^2}{1000}} = e^{-320} \approx 1.06 \cdot 10^{-139}$$

Ez is kicsi, de azért a Cramér becslés sokkal jobb.

~~A Cramér tétel és a Hoef~~

A Cramér tétel és a Hoeffding-egyenlőtlenség összehasonlítása

<u>lítása</u> : mindkettőhöz $X_1, \dots X_n$ <u>független</u> kell.

<u>Hoeffding egyenlőtlenség</u>

- <u>előnyök</u>: → X_i-k-nek nem kell azonos eloszlásúnak lenni
  → X_i-kről $\boxed{\text{keveset}}$ kell tudni:
  csak a korlátokat és az összeg várható értékét
  → nagyon könnyű számolás
- <u>hátrányok</u>: → CSAK KORLÁTOS val. változókra jó
  → csak felső becslést ad, ami általában durva

<u>Cramér tétel</u>

- <u>előnyök</u>: → NEM CSAK KORLÁTOS val. változókra jó, hanem annál sokkal általánosabb esetben is
  → felső becslést <u>és</u> egyúttal <u>jó közelítést</u> ad
- <u>Hátrányok</u>: → csak (független és) <u>azonos eloszlású</u> val. változókra használható
  → az eloszlást $\boxed{\text{pontosan kell ismerni}}$ (mindent tudni kell róla)
  → általában macerás számolás