Valószínűségszámítás és statisztika – Valószínűségszámítási egyenlőtlenségek, nagy számok törvényei (ami az egydim. eloszlásokból eddig kimaradt)

Tartalom

- Várható érték általános esetben
- Markov- és Csebisev-egyenlőtlenség, paraméteres Csernov-korlát
- Nevezetes folytonos eloszlások: egyenletes, exponenciális, normális
- Konvergenciafajták, nagy számok törvényei

Előismeretek: Differenciál- és integrálszámítás egy dimenzióban, diszkrét és folytonos valószínűségi változók

Várható érték általános esetben

Eddig: 1. Ha X diszkrét, akkor $\mathbb{E}(X) = \sum_{k \in \mathrm{Ran}(X)} k\mathbb{P}(X = k)$,
2. ha X folytonos, akkor $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$ (mindkét esetben feltéve, hogy $\mathbb{E}(|X|)$ létezik),
és 3. ha X nemnegatív, akkor $\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x)) \mathrm{d}x \in [0, \infty]$.
Kérdés: Mi köze ezeknek a várhatóérték-fogalmaknak egymáshoz? Van-e a várható értéknek olyan definíciója, amely mindhárom esetet magába foglalja?

Várható érték általános esetben

Eddig: 1. Ha X diszkrét, akkor $\mathbb{E}(X) = \sum_{k \in \mathrm{Ran}(X)} k\mathbb{P}(X = k)$,
2. ha X folytonos, akkor $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$ (mindkét esetben feltéve, hogy $\mathbb{E}(|X|)$ létezik),
és 3. ha X nemnegatív, akkor $\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x)) \mathrm{d}x \in [0, \infty]$.
Kérdés: Mi köze ezeknek a várhatóérték-fogalmaknak egymáshoz? Van-e a várható értéknek olyan definíciója, amely mindhárom esetet magába foglalja?

1. válasz: $\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x)) \mathrm{d}x \in [0, \infty]$ bármilyen nemnegatív val. változóra igaz.

(Ezt már kimondtuk, bár csak folytonos esetben láttuk be.)

Ez alapján: akármilyen X val. változóra $X^+ = \max\{X, 0\}$ és $X^- = \max\{-X, 0\}$ nemnegatív val. változók, ezért $\mathbb{E}(X^+), \mathbb{E}(X^-)$ jóldefiniált.
Ha legalább az egyik véges, akkor $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \in [-\infty, \infty]$ is jóldefiniált. Ha mindkettő végtelen, akkor $\mathbb{E}(X)$-et nem értelmezzük.
Az általános esetben is igaz: mindkettő véges $\Leftrightarrow \mathbb{E}(|X|) < \infty$.

Várható érték általános esetben

Eddig: 1. Ha X diszkrét, akkor $\mathbb{E}(X) = \sum_{k \in \mathrm{Ran}(X)} k\mathbb{P}(X = k)$,
2. ha X folytonos, akkor $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$ (mindkét esetben feltéve, hogy $\mathbb{E}(|X|)$ létezik),
és 3. ha X nemnegatív, akkor $\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x)) \mathrm{d}x \in [0, \infty]$.
Kérdés: Mi köze ezeknek a várhatóérték-fogalmaknak egymáshoz? Van-e a várható értéknek olyan definíciója, amely mindhárom esetet magába foglalja?

2. válasz: szintén X^+-szal és X^--szal, de ezek várható értékét máshogy definiáljuk (amiről belátható, hogy az előző definícióval ekvivalens).

(3.4.1. Definíció.) Legyen X egy nemnegatív valószínűségi változó. Ekkor definíció szerint

$$\mathbb{E}(X) = \sup_{Z\colon Z \text{ egyszerű val. változó},\ Z \leq X} \mathbb{E}(Z) \in [0, \infty].$$

Vagyis X várható értéke a nála (ω-nként) nem nagyobb egyszerű Z val. változók várható értékeinek „lehető legnagyobb értéke”, precízebben legkisebb felső korlátja, azaz szuprémuma.

Várható érték általános esetben

$$X \geq 0 \qquad \Rightarrow \qquad \mathbb{E}(X) = \sup_{Z\colon Z \text{ egyszerű val. változó},\ Z \leq X} \mathbb{E}(Z) \in [0, \infty].$$

Egyszerű következmények (részben már találkoztunk velük):

- A várható érték monotonitása: Ha $\mathbb{P}(0 \leq X \leq Y) = 1$ (ahol az egyenlőtlenség ω-nként értendő), akkor $\mathbb{E}(X) \leq \mathbb{E}(Y)$. (Itt $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ is lehet végtelen. Konvenció: $\infty > c$ minden $c \in \mathbb{R}$-re és $\infty \geq \infty$).
- Ugyanez igaz annak feltevése nélkül is, hogy X és Y nemnegatívak, amennyiben a várható értékük definiált (az nem baj, ha $+\infty$ vagy $-\infty$).
- Speciális eset: pl. ha $c \in \mathbb{R}$ és $\mathbb{P}(X \geq c) = 1$, akkor $c \leq \mathbb{E}(X) \leq \infty$. Ha $\mathbb{P}(X \leq c) = 1$, akkor $-\infty \leq \mathbb{E}(X) \leq c$.
- Ha $a < b$ és $\mathbb{P}(a \leq X \leq b) = 1$, akkor $\mathbb{E}(X) \in [a, b]$.
- *Ezeket a vizsgafeladatok megoldásánál érdemes ellenőrizni!*

Markov-egyenlőtlenség

Előfordulhat, hogy egy val. változó pontos eloszlását nem ismerjük, de a momentumait (várható értékét, szórását stb.) becsülni tudjuk (pl. mérési eredmények/mintavétel alapján $\rightarrow$ lásd hamarosan a statisztika résznél).

A következőkben olyan egyenlőtlenségekkel ismerkedünk meg, amelyek egy val. változó momentumai alapján adnak becslést arra, hogy a val. változó milyen valószínűséggel vesz fel "extrém" (nagy/kicsi) értékeket.

(9.1.1. Állítás (Markov-egyenlőtlenség).) Legyen X nemnegatív értékű valószínűségi változó. Ekkor minden $a > 0$ esetén

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

(Bizonyítás)

Igaz diszkrét és folytonos (és általános) esetben is.
A Markov-egyenlőtlenséget néha $a > 0$ helyett csak $a > \mathbb{E}(X)$-re mondják ki (mert ha $a \leq X$, akkor triviális egyenlőtlenséget kapunk).

A Markov-egyenlőtlenség általánosításai

(9.1.1. Állítás (Markov-egyenlőtlenség).) Legyen $X \geq 0$, ekkor

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}, \qquad \forall a > 0.$$

Mit tehetünk, ha X negatív értékeket is felvesz, vagy ha a Markov-egyenlőtlenség nem ad túl jó becslést?

Transzformáljuk X-et valamilyen alkalmas nemnegatív val. változóvá. Az így kapott egyenlőtlenség, amit magyarul szokás Tóth Bálint nyomán *turbó Markov-egyenlőtlenség*nek is nevezni:

(Következmény.) Legyen X valószínűségi változó és $g\colon \underbrace{\mathrm{Ran}(X)}_{\subseteq \mathbb{R}} \to [0, \infty)$ folytonos, monoton növő függvény.

Ekkor minden $a > \mathbb{E}(g(X))$ esetén

$$\mathbb{P}(X \geq a) = \mathbb{P}(g(X) \geq g(a)) \leq \frac{\mathbb{E}(g(X))}{g(a)}.$$

A monoton növekvés fontos – különben az első egyenlőség nem feltétlenül teljesül!

A Markov-egyenlőtlenség általánosításai

(9.1.1. Állítás (Markov-egyenlőtlenség).) Legyen $X \geq 0$, ekkor

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}, \qquad \forall a > 0.$$

Mit tehetünk, ha X negatív értékeket is felvesz, vagy ha a Markov-egyenlőtlenség nem ad túl jó becslést?
Transzformáljuk X-et valamilyen alkalmas nemnegatív val. változóvá. Az így kapott egyenlőtlenség, amit magyarul szokás Tóth Bálint nyomán *turbó Markov-egyenlőtlenség*nek is nevezni:

(Következmény.) Legyen X valószínűségi változó és $g\colon \underbrace{\mathrm{Ran}(X)}_{\subseteq\mathbb{R}} \to [0,\infty)$ folytonos, monoton növő függvény.

Ekkor minden $a > \mathbb{E}(g(X))$ esetén

$$\mathbb{P}(X \geq a) = \mathbb{P}(g(X) \geq g(a)) \leq \frac{\mathbb{E}(g(X))}{g(a)}.$$

A monoton növekvés fontos – különben az első egyenlőség nem feltétlenül teljesül!

Csebisev-egyenlőtlenség

(„Turbó Markov-egyenlőtlenség.") Legyen X valószínűségi változó és $g\colon \mathrm{Ran}(X) \to [0,\infty)$ folytonos, monoton növő függvény. Ekkor minden $a > \mathbb{E}(g(X))$ esetén

$$\mathbb{P}(X \geq a) = \mathbb{P}(g(X) \geq g(a)) \leq \frac{\mathbb{E}(g(X))}{g(a)}.$$

(9.1.2. Következmény (Csebisev-egyenlőtlenség).) Legyen Y valószínűségi változó, ekkor minden $a > 0$ esetén

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq a) \leq \frac{\mathbb{D}^2(Y)}{a^2}.$$

Biz.: Mivel $\mathbb{D}^2(Y) = \mathbb{E}((|Y - \mathbb{E}(Y)|)^2)$, az egyenlőtlenség nem más, mint a turbó Markov, az alábbi választással:

- $X = |Y - \mathbb{E}(Y)|$,
- $g(x) = x^2$ (ami $\mathrm{Ran}(X)$-en, azaz $[0,\infty)$-en szig. mon. nő).

A turbó Markov és a szórásnégyzet definíciója miatt

$$\mathbb{P}(|Y-\mathbb{E}(Y)| \geq a) = \mathbb{P}((Y-\mathbb{E}(Y))^2 \geq a^2) \leq \frac{\mathbb{E}((Y - \mathbb{E}(Y))^2)}{a^2} = \frac{\mathbb{D}^2(Y)}{a^2}.$$

Csebisev-egyenlőtlenség

(„Turbó Markov-egyenlőtlenség.") Legyen X valószínűségi változó és $g\colon \mathrm{Ran}(X) \to [0,\infty)$ folytonos, monoton növő függvény. Ekkor minden $a > \mathbb{E}(g(X))$ esetén

$$\mathbb{P}(X \geq a) = \mathbb{P}(g(X) \geq g(a)) \leq \frac{\mathbb{E}(g(X))}{g(a)}.$$

(9.1.2. Következmény (Csebisev-egyenlőtlenség).) Legyen Y valószínűségi változó, ekkor minden $a > 0$ esetén

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq a) \leq \frac{\mathbb{D}^2(Y)}{a^2}.$$

Biz.: Mivel $\mathbb{D}^2(Y) = \mathbb{E}((|Y - \mathbb{E}(Y)|)^2)$, az egyenlőtlenség nem más, mint a turbó Markov, az alábbi választással:

- $X = |Y - \mathbb{E}(Y)|$,
- $g(x) = x^2$ (ami $\mathrm{Ran}(X)$-en, azaz $[0,\infty)$-en szig. mon. nő).

A Csebisev-egyenlőtlenség bármilyen val. változóra jó (bár $\mathbb{E}(Y^2) = \infty$ esetén triviális becslést ad.)

Csebisev-egyenlőtlenség: példák

Egy adott adatbázis szerver átlagosan 50 lekérést fogad egy időegység alatt. A lekérések számának szórása a tapasztalatok szerint 5. Adjunk alsó becslést annak a valószínűségére, hogy 40-nél több, de 60-nál kevesebb lesz a lekérések száma egy időegység alatt. (11.6. feladat, lásd gyakorlat)

A feladat tipikus példája annak, amikor egy val. változó eloszlását nem ismerjük, csak a momentumairól (itt a várható értékéről és a szórásáról) tudunk valamit.

A Csebisev-egyenlőtlenség nem mindig ad nagyon jó becslést, sőt $\mathbb{E}(Y^2) < \infty$ esetén is adhat triviális becslést (lásd gyakorlat, 11.7. feladat, itt a CHT egyébként kiváló becslést ad).

Paraméteres Csernov-egyenlőtlenség

(„Turbó Markov-egyenlőtlenség.") Legyen X valószínűségi változó és $g\colon \mathrm{Ran}(X) \to [0,\infty)$ folytonos, monoton növő függvény. Ekkor minden $a > \mathbb{E}(g(X))$ esetén

$$\mathbb{P}(X \geq a) = \mathbb{P}(g(X) \geq g(a)) \leq \frac{\mathbb{E}(g(X))}{g(a)}.$$

A $g(x) = x^2$ választáshoz képest még jobb becslést kaphatunk, ha a gyorsabban növő g-t választunk:

(9.1.4. Következmény (Paraméteres Csernov-egyenlőtlenség).) Legyen X valószínűségi változó. Ekkor minden $a, t > 0$ esetén

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}\left(\mathrm{e}^{tX}\right)}{\mathrm{e}^{ta}}.$$

Biz.: ez valóban a turbó Markov, ahol $X = X$, $g(x) = \mathrm{e}^{tx}$.

(9.1.5. Példa: Legyen $X \sim \mathrm{Pois}(5)$. Adjunk felső becslést $\mathbb{P}(X \geq 10)$-re. $\to$ A paraméteres Csernov-egyenlőtlenséggel, t-t optimalizálva jóval pontosabb felső korlátot kapunk, mint a Csebisev-egyenlőtlenséggel.) *(Mese a nagy eltérésekről... $\to$ előadás)*

Még egy példa Csernovra

A paraméteres Csernov-egyenlőtlenség pl. akkor is alkalmazható, ha X nemnegatív és $\mathbb{E}(\mathrm{e}^{tX})$ csak kis t esetén véges.
Nevezetes eloszlások ezzel a tulajdonsággal: exponenciális és geometriai eloszlás.
Ha $X \sim \mathrm{Exp}(\lambda)$, akkor $\mathbb{E}[\mathrm{e}^{tX}] = \frac{\lambda}{\lambda - t}$, ha $t < \lambda$ és $\mathbb{E}[\mathrm{e}^{tX}] = \infty$, ha $t \geq \lambda$. (Számolás: lásd előadás.)

Példa: Veszünk 10 gumibelsőt egy kerékpárhoz. Egy gumibelső élettartama folytonos, örökifjú eloszlást követ fél év várható értékkel, és a még be nem helyezett gumibelsők nem öregszenek. Ha az egyik gumi kilyukad, azonnal betesszük a következőt, és a gumik élettartamai függetlenek. Mi a valószínűsége, hogy a 10 gumbibelső kitart addig, amíg a biciklit 15 év múlva selejtezzük?

Még egy példa Csernovra

A paraméteres Csernov-egyenlőtlenség pl. akkor is alkalmazható, ha X nemnegatív és $\mathbb{E}(\mathrm{e}^{tX})$ csak kis t esetén véges.
Nevezetes eloszlások ezzel a tulajdonsággal: exponenciális és geometriai eloszlás.
Ha $X \sim \mathrm{Exp}(\lambda)$, akkor $\mathbb{E}[\mathrm{e}^{tX}] = \frac{\lambda}{\lambda - t}$, ha $t < \lambda$ és $\mathbb{E}[\mathrm{e}^{tX}] = \infty$, ha $t \geq \lambda$. (Számolás: lásd előadás.)

Példa: Veszünk 10 gumibelsőt egy kerékpárhoz. Egy gumibelső élettartama folytonos, örökifjú eloszlást követ fél év várható értékkel, és a még be nem helyezett gumibelsők nem öregszenek. Ha az egyik gumi kilyukad, azonnal betesszük a következőt, és a gumik élettartamai függetlenek. Mi a valószínűsége, hogy a 10 gumbibelső kitart addig, amíg a biciklit 15 év múlva selejtezzük?

($\leq 0{,}000127$, lásd előadás. Ha X_i az i-edik gumibelső élettartama, akkor $\mathbb{E}(\mathrm{e}^{X_1+\ldots+X_{10}}) = \mathbb{E}(\mathrm{e}^{X_1})^{10}$, mert függetlenek és azonos eloszlásúak. $\mathbb{E}(\mathrm{e}^{tX})$ fenti képletét alkalmazzuk $X = X_1$-re és $t < \lambda = 2$-re, majd az így kapott Csernov-korlátot t szerint deriváljuk és megnézzük, hogy hol 0 a derivált.)

Még egy példa Csernovra

A paraméteres Csernov-egyenlőtlenség pl. akkor is alkalmazható, ha X nemnegatív és $\mathbb{E}(\mathrm{e}^{tX})$ csak kis t esetén véges.
Nevezetes eloszlások ezzel a tulajdonsággal: exponenciális és geometriai eloszlás.
Ha $X \sim \mathrm{Exp}(\lambda)$, akkor $\mathbb{E}[\mathrm{e}^{tX}] = \frac{\lambda}{\lambda - t}$, ha $t < \lambda$ és $\mathbb{E}[\mathrm{e}^{tX}] = \infty$, ha $t \geq \lambda$. (Számolás: lásd előadás.)

Példa: Veszünk 10 gumibelsőt egy kerékpárhoz. Egy gumibelső élettartama folytonos, örökifjú eloszlást követ fél év várható értékkel, és a még be nem helyezett gumibelsők nem öregszenek. Ha az egyik gumi kilyukad, azonnal betesszük a következőt, és a gumik élettartamai függetlenek. Mi a valószínűsége, hogy a 10 gumbibelső kitart addig, amíg a biciklit 15 év múlva selejtezzük?

Miért nem CHT-vel csináltuk?
Próbáljuk ki CHT-vel is. A közelítés $1 - \Phi(6{,}32)$ (lásd előadás), de a $6{,}32$ hely túl nagy, nincs a normális eloszlás táblázatban...
A CHT csak *normális* eltérések valószínűségének megbecsülésére jó, *nagy* eltérésekére nem.

Nagy számok törvénye, konvergenciafajták

A relatív gyakoriság tart a valószínűséghez.
"Független, azonos eloszlású kísérletek átlaga a várható értékhez tart."
Mit értünk az alatt, hogy "tart"?
Egyfajta konvergenciatípust már láttunk: eloszlásban való konvergencia. $F_{X_n}(x) \to F_Z(x)$.

Kétféle kerül még elő az idén: $Z, X_1, X_2, \ldots$ val. változókra

- Sztochasztikus (avagy valószínűségben való) konvergencia: X_n valószínűségben konvergál Z-hez, ha $\forall \varepsilon > 0$-ra teljesül, hogy
$$\lim_{n\to\infty} \mathbb{P}(|X_n - Z| > \varepsilon) = 0.$$
- 1 valószínűséggel való (avagy majdnem biztos) konvergencia: X_n 1 valószínűséggel/majdnem biztosan konvergál Z-hez, ha
$$\mathbb{P}\Big(\lim_{n\to\infty} X_n = Z\Big) = \mathbb{P}\Big(\Big\{\omega \in \Omega \Big| \lim_{n\to\infty} X_n(\omega) = Z(\omega)\Big\}\Big) = 1.$$
Ez a legerősebb a háromból, míg az eloszlásban való konvergencia a leggyengébb.

Kapcsolat a konvergenciafajták között (kiegészítő anyag)

Mit értünk a konvergenciafajták erősségén/gyengeségén?

- Ha X_n majdnem biztosan/1 valószínűséggel konvergál Z-hez, akkor sztochasztikusan/valószínűségben is konvergál.
- A sztochasztikus konvergenciából viszont nem következik az 1 valószínűségű konvergencia.
  Igaz viszont: ha X_n sztochasztikusan konvergál Z-hez, akkor létezik egy olyan $X_{n_1}, X_{n_2}, \ldots$ végtelen részsorozat $(1 \leq n_1 < n_2 < \ldots)$, amely 1 valószínűséggel Z-hez tart.

Ezért szokás a sztochasztikus konvergenciát ,,gyenge", a majdnem biztos konvergenciát pedig ,,erős" konvergenciának is nevezni. Erre utal a következőkben a *nagy számok erős* és *gyenge* törvénye.

A fenti állításokat itt nem tudjuk bizonyítani (lásd mértékelmélet kurzus.)

Általában: ,,majdnem biztosan" = ,,1 valószínűséggel". Pl.: egy $\mathrm{U}(0;1)$ eloszlású X val. változó értéke majdnem biztosan nem $1/2$. $\{X = 1/2\}$ nem lehetetlen, de 0 valószínűségű esemény.

Nagy számok törvénye(i)

(9.2.1. Tétel.) Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók egy sorozata.
Tegyük fel, hogy $\mathbb{E}(X_n) = \mu \in \mathbb{R}$ és $\mathbb{D}(X_n) = \sigma > 0$ minden n-re.
Jelölje $\overline{X_n}$ az $\frac{X_1+\ldots+X_n}{n}$ átlagot.

- Nagy számok gyenge törvénye: Tetszőleges $\varepsilon > 0$ esetén

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X_n} - \mu| \geq \varepsilon) = 0.$$

- Nagy számok erős törvénye:

$$\mathbb{P}\Big(\lim_{n\to\infty} \overline{X_n} = \mu\Big) = \mathbb{P}\Big(\Big\{\omega \in \Omega \Big| \lim_{n\to\infty} \overline{X_n}(\omega) = \mu\Big\}\Big) = 1.$$

- (Gyenge törvény biz. Csebisevvel: lásd előadás/jegyzet)
- A gyenge törvény (indikátor X_i-k esetén) Bernoulli, az erős törvény Kolmogorov nevéhez fűződik.
- A törvények igazak akkor is, ha a szórások nem végesek, csak $\mathbb{E}(|X_1|) < \infty$ teljesül. Erős törvény bizonyítása az általános esetben: BME-TTK, Valszám 2, kb. két előadás.

Nagy számok törvénye, speciális eset: relatív gyakoriság

Legyenek $A_1, \ldots, A_n$ együttesen független, p valószínűségű események.

$$\mathbb{P}(A_1 \cup \ldots A_n) = \mathbb{P}\Big(\sum_{i=1}^{n} \mathbb{1}_{A_i} > 0\Big).$$

A tétel miatt $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i}$-re:

- Annak a valószínűsége, hogy p-től $\overline{X_n}$ egy akármilyen kicsi rögzített ε-nal is eltér, nullához tart. (A gyenge törvény miatt.)
- 1 valószínűséggel olyan kimenetelen vagyunk, ahol az $n \mapsto \overline{X_n}(\omega)$ számsorozat p -hez tart. (Az erős törvény miatt.)
- Következmény: ,,whatever can happen, will happen".
  1 valószínűséggel olyan kimenetelen vagyunk, ahol az $A_1, A_2, \ldots$ események közül végtelen sok bekövetkezik.

Nagy számok törvénye, speciális eset: relatív gyakoriság

Legyenek $A_1, \ldots, A_n$ együttesen független, p valószínűségű események.

$$\mathbb{P}(A_1 \cup \ldots A_n) = \mathbb{P}\Big(\sum_{i=1}^{n} \mathbb{1}_{A_i} > 0\Big).$$

A tétel miatt $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i}$-re:

- Annak a valószínűsége, hogy p-től $\overline{X_n}$ egy akármilyen kicsi rögzített ε-nal is eltér, nullához tart. (A gyenge törvény miatt.)
- 1 valószínűséggel olyan kimenetelen vagyunk, ahol az $n \mapsto \overline{X_n}(\omega)$ számsorozat p -hez tart. (Az erős törvény miatt.)
- Következmény: ,,whatever can happen, will happen''. 1 valószínűséggel olyan kimenetelen vagyunk, ahol az $A_1, A_2, \ldots$ események közül végtelen sok bekövetkezik. (Lásd még: majom az írógépnél, kép: `https://hu.wikipedia.org/wiki/V%C3%A9gtelen_sok_majom_%C3%A9s_%C3%ADr%C3%B3g%C3%A9p_t%C3%A9tele`)

Miért 1 valószínűséggel?

Jogos kérdés: a nagy számok erős törvényében miért kell az „1 valószínűséggel" feltétel? Nem lehetne esetleg kihagyni?

Válasz: Nem, még diszkrét valószínűségi változók esetén sem.
Intuitíció: Ha pl. $X_1, X_2, \ldots$ független indikátorváltozók, ahol $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = 1/2 \in (0, 1)$, akkor egy megfelelő Ω eseménytér elemei végtelen 0–1 sorozatok, ahol az i-edik bit az X_i értékét adja meg.
Ez az eseménytér tartalmaz olyan elemeket is, ahol a mintaátlag nem tart $1/2$-hez, pl. a $00000\ldots$ sorozatnál 0-hoz, az $101101110\ldots$ sorozatnál 1-hez, a $0010100101\ldots$ sorozatnál $2/5$-höz tart.
Így nem lehetetlen esemény, hogy a mintaátlag nem tart $1/2$-hez.
Az erős törvény szerint azonban 0 valószínűségű esemény. Az összes ilyen kimenetelből álló esemény valószínűsége 0.
Mértékelméleti hiányosságok miatt itt nem beszélünk arról, hogy a törvényben szereplő végtelen sok valószínűségi változót $(X_1, X_2, \ldots)$ milyen $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn lehet együttesen definiálni.

Kitérő: konvergenciák

Konvergencia-típusok, amiket említettünk:

- egy valószínűségű konvergencia $\mathbb{P}(\lim_{n\to\infty} \ldots) = 1$
- sztochasztikus konvergencia $\mathbb{P}(|\ldots| \geq \varepsilon) \overset{n\to\infty}{\longrightarrow} 0$
- eloszlásbeli konvergencia $F_{X_n}(x) \to F_Z(x)$.

Tegyük fel, hogy $\mathbb{E}(X_i) = 0$ és $\mathbb{D}(X_i) = 1$. Ekkor

- NSZT: $\overline{X_n} \to 0$ (1. és 2. értelemben)
- CHT: $\sqrt{n} \cdot \overline{X_n} \to Z \sim \mathrm{N}(0; 1)$ (3. értelemben)

NSZT további alkalmazásai: lásd a statisztika résznél (empirikus eloszlásfüggvény konvergenciája).