

Nagy számok törvényei

Statisztikai mintavétel

Várható érték becslése

Dr. Berta Miklós

Fizika és Kémia Tanszék

Széchenyi István Egyetem



SZÉCHENYI
ISTVÁN
EGYETEM

Fizika és Kémia
Tanszék



A mérés mint statisztikai mintavétel

- A méréssel az eloszlásfüggvénnyel leírt valószínűségi változó végtelen lehetséges értékei közül *véges mintát veszünk*.
- A matematikai statisztika olyan kérdésekre keresi a választ, hogy a *véges mintavétel alapján milyen valószínűséggel következtethetünk az eredeti valószínűségi változó eloszlásfüggvényére?*

Csebishev-egyenlőtlenség

- Ha az X tetszőleges valószínűségi változónak van várható értéke (m) és szórása (D), valamint $k > 1$ tetszőleges valós szám, akkor annak a valószínűsége, hogy X a várható értékétől abszolút értékben a szórása k -szorosánál nagyobb mértékben térjen el legfeljebb $\frac{1}{k^2}$.

$$P(|X - m| \geq k.D) \leq \frac{1}{k^2}$$

Példa: Egy gyártósoron csapágygolyókat gyártunk, és megállapítottuk, hogy a golyók átmérőjének várható értéke 5 mm 0,1 mm-es szórással.

Legfeljebb mekkora a valószínűsége annak, hogy ha a gyártott csapágygolyók közül kiválasztunk egyet, akkor annak átmérője nagyobb lesz, mint 5,3 mm? Mekkora ez a valószínűség, ha tudjuk, hogy a gyártott golyók átmérője Gauss – eloszlású valószínűségi változó?

Megoldás: Az 5,3 mm-nél nagyobb átmérő a szórás több mint 3-szorosával tér el a várható értéktől. A Csebisev-egyenlőtlenség alapján a keresett valószínűség:

$$P((d - 5) \geq 0,3) \leq \frac{1}{3^2} = 0,11.$$

Ha feltesszük a Gauss – eloszlást, akkor:

$$P((d - 5) \geq 0,3) = 1 - F_G(3) = 0,0013.$$

Megjegyzés: Az $F_G(3)$ a standard Gauss – eloszlás értéke 3-nál. A MATLAB ezt a `normcdf(3)` paranccsal számolja.

Nagy számok törvénye

- Legyenek $X_1, X_2, \dots, X_N$ valószínűségi változók függetlenek, azonos m várható értékkel, és azonos D szórással, valamint ε legyen tetszőleges pozitív valós szám. Az

$$\bar{X}_N = \frac{X_1 + X_2 + \dots + X_N}{N}$$

számtani középére igaz, hogy:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_N - m| < \varepsilon) = 1.$$

Következmény

- Ha valamely mennyiség értékét n -szer megmérjük, akkor a mért értékek **számtani közepe** egyre közelebb lesz a mennyiség „**valódi értéke**hez”, ha a mérések N számát növeljük.
- A Csebisev – egyenlőtlenségből mindig meghatározható, hogy adott N számú mérés számtani közepe legfeljebb mekkora valószínűséggel tér el a „valódi” értéktől a mért értékek szórásának legalább k -szorosával.

Statisztikai becslések

- Láttuk, hogy a számtani közép, statisztikus értelemben, a várható értékhez tart, ha növeljük a mérések számát.
- Azaz a számtani közép a várható érték egy *statisztikus becslése*.
- Egy statisztikus becslést *torzítatlannak* nevezünk, ha n növekedésével a becsült értékhez tart.

A szórás becslése

- Az

$$s_0^{*2} = \frac{\sum_{i=1}^N (X_i - m)^2}{N - 1}$$

mennyiséget *korrigált tapasztalati szórás-négyzet*nek nevezzük, és bebizonyítható, hogy *a valószínűségi eloszlás szórásnégyzetének torzítatlan becslését adja.*

A számtani közép szórásának becslése

- A mérések során fellépő véletlen hibákról feltesszük, hogy normál eloszlásúak.
- Ha egy mennyiséget többször mérünk, akkor a mért értékek számtani közepe a mérendő mennyiség torzítatlan becslése.
- *Mivel a számtani közép is mérési eredményekből számítható, így maga is valószínűségi változó!*

A számtani közép szórásának becslése

- A centrális határeloszlás tétele szerint:
N független, azonos valószínűségi eloszlású változó (μ_0, σ_0) összege $N \rightarrow \infty$ határesetben normális eloszlású valószínűségi változó, melynek *várható értéke* $\mu = N \cdot \mu_0$, *szórásnégyzete* pedig $\sigma^2 = N \cdot \sigma_0^2$.
- A számtani közép definíciója alapján:

$$\sum_{i=1}^N x_i = N \bar{X}_N$$

A számtani közép szórásának becslése

- Tudjuk, hogy az előbbi összeg várható értéke:

$$\left\langle \sum_{i=1}^N x_i \right\rangle = N \cdot \mu_0 = \langle N \bar{X}_N \rangle = N \langle \bar{X}_N \rangle \Rightarrow \langle \bar{X}_N \rangle = \mu_0$$

- Tudjuk, hogy az előbbi összeg szórásnégyze-
te:

$$\sigma^2\left(\sum_{i=1}^N x_i\right) = N\sigma_0^2 = \sigma^2(N\bar{X}_N) = N^2\sigma^2(\bar{X}_N) \Rightarrow \sigma^2(\bar{X}_N) = \frac{\sigma_0^2}{N}$$

Megjegyzés:

$$\begin{aligned}\sigma^2(N\bar{X}_N) &= \int_{\forall x} (N\bar{X}_N - \mu(N\bar{X}_N))^2 \rho_G(x) dx = \\ &= N^2 \int_{\forall x} (\bar{X}_N - \mu(\bar{X}_N))^2 \rho_G(x) dx = N^2 \sigma^2(\bar{X}_N)\end{aligned}$$

Méréselméleti alkalmazás

- Ha egy X mérendő mennyiség mért értékeit (x_i) sorozat jelöli, akkor a mérendő mennyiség várható értékének torzítatlan becslése a mért értékek $\bar{X}_N$ számtani közepe:

$$\bar{X}_N = \frac{\sum_{i=1}^N x_i}{N},$$

a következő tapasztalati szórással:

$$s_{\bar{X}}^{*2} = \frac{\sum_{i=1}^N (x_i - \bar{X}_N)^2}{N(N-1)} = \frac{s_0^{*2}}{N}.$$

Példa: Egy rúd hosszát többször megmértük. A mért értékek a következők:

i	1	2	3	4	5	6	7	8	9
L_i [cm]	102	103	105	101	102	104	105	102	104

Becsülje meg a rúd hosszát és a mérés „hibáját”!

Megoldás: (MATLAB segítségével)

$$L = (103.1111 \pm 0.4843) \text{ cm}$$

Példa: Egy fonálinga lengésidejét mértük. Az alábbi 15 értéket kaptuk. Határozza meg ezek alapján a fonálinga lengésidejét és a mérés bizonytalanságát!

Sorszám	$t[s]$
1	29,729
2	26,711
3	27,787
4	26,457
5	27,158
6	27,200
7	28,801
8	34,318
9	28,157
10	26,264
11	27,404
12	28,274
13	28,964
14	29,513
15	28,193
Átlag	27,9
Korrigált szórás	0,3

Megoldás: (MATLAB segítségével)