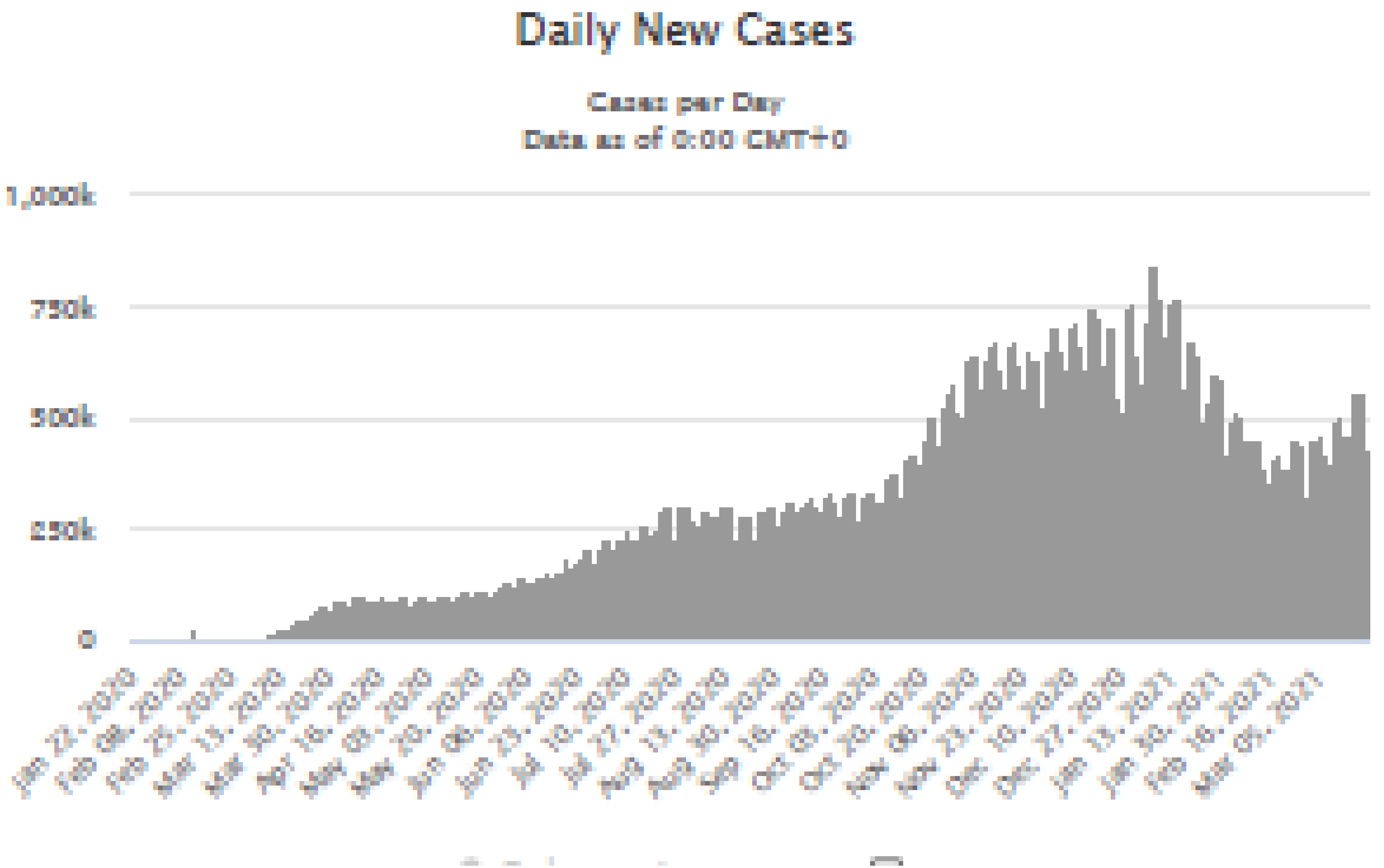
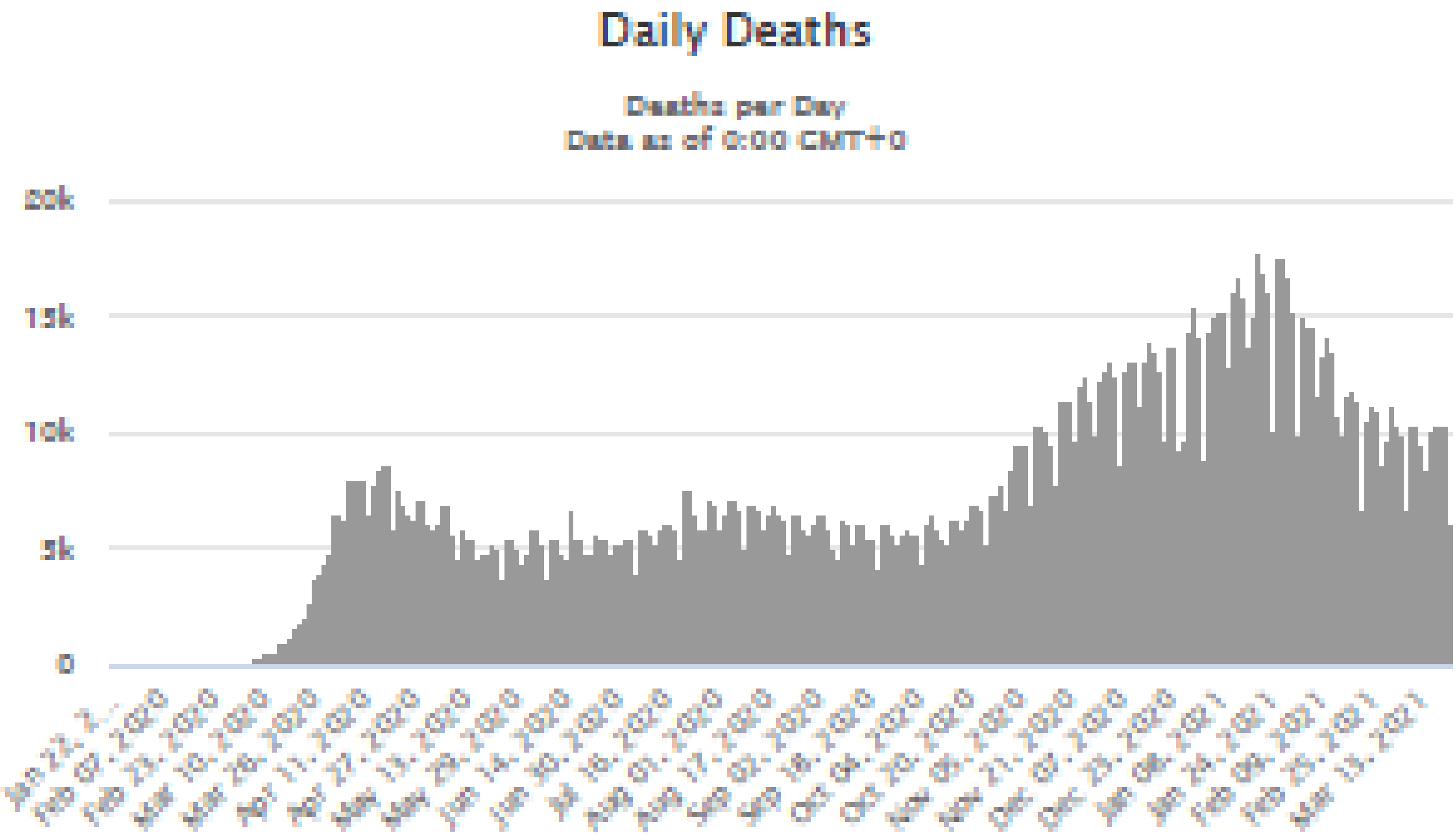
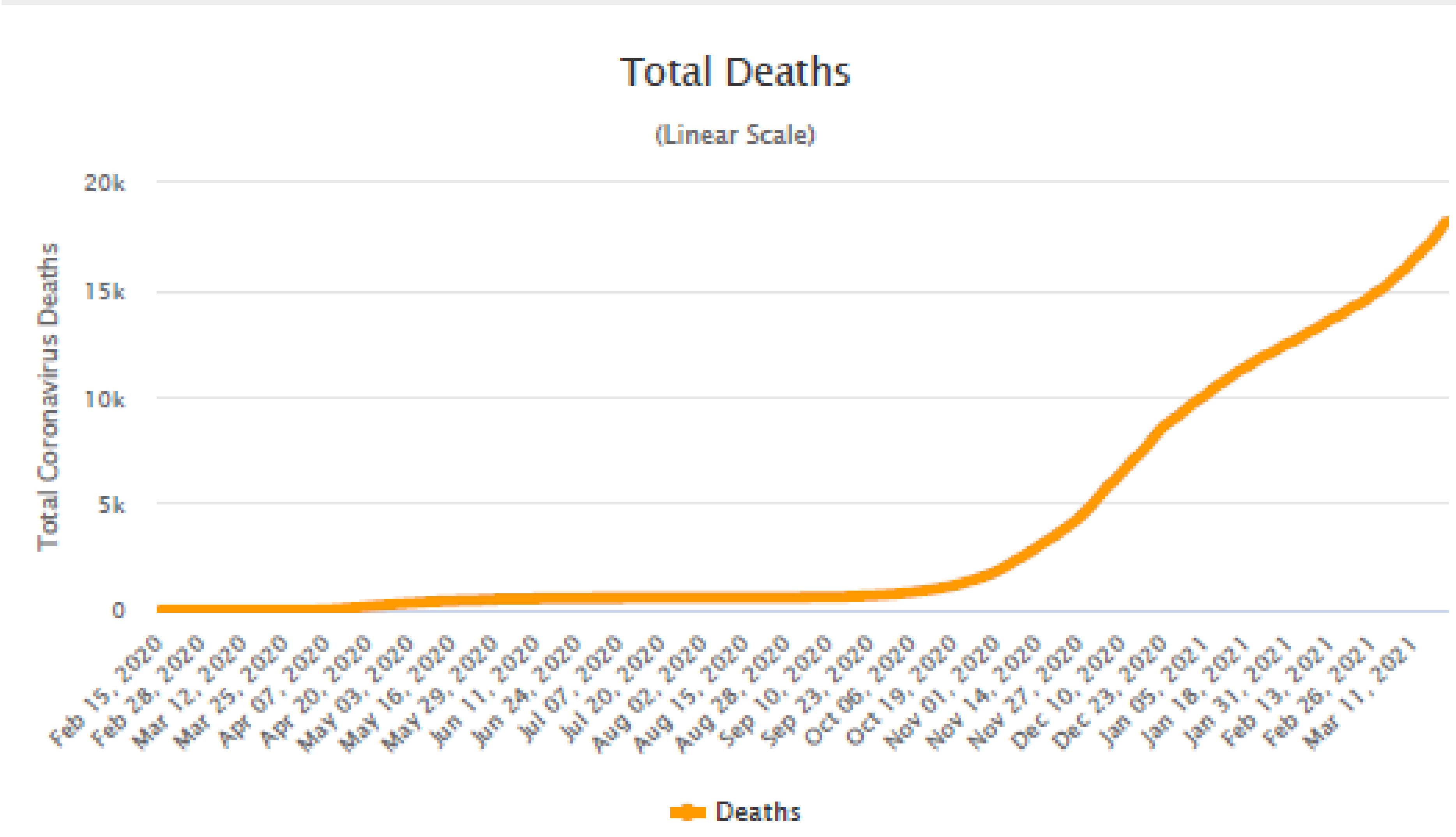
Valószínűségszámítás és Statisztika

8. előadás

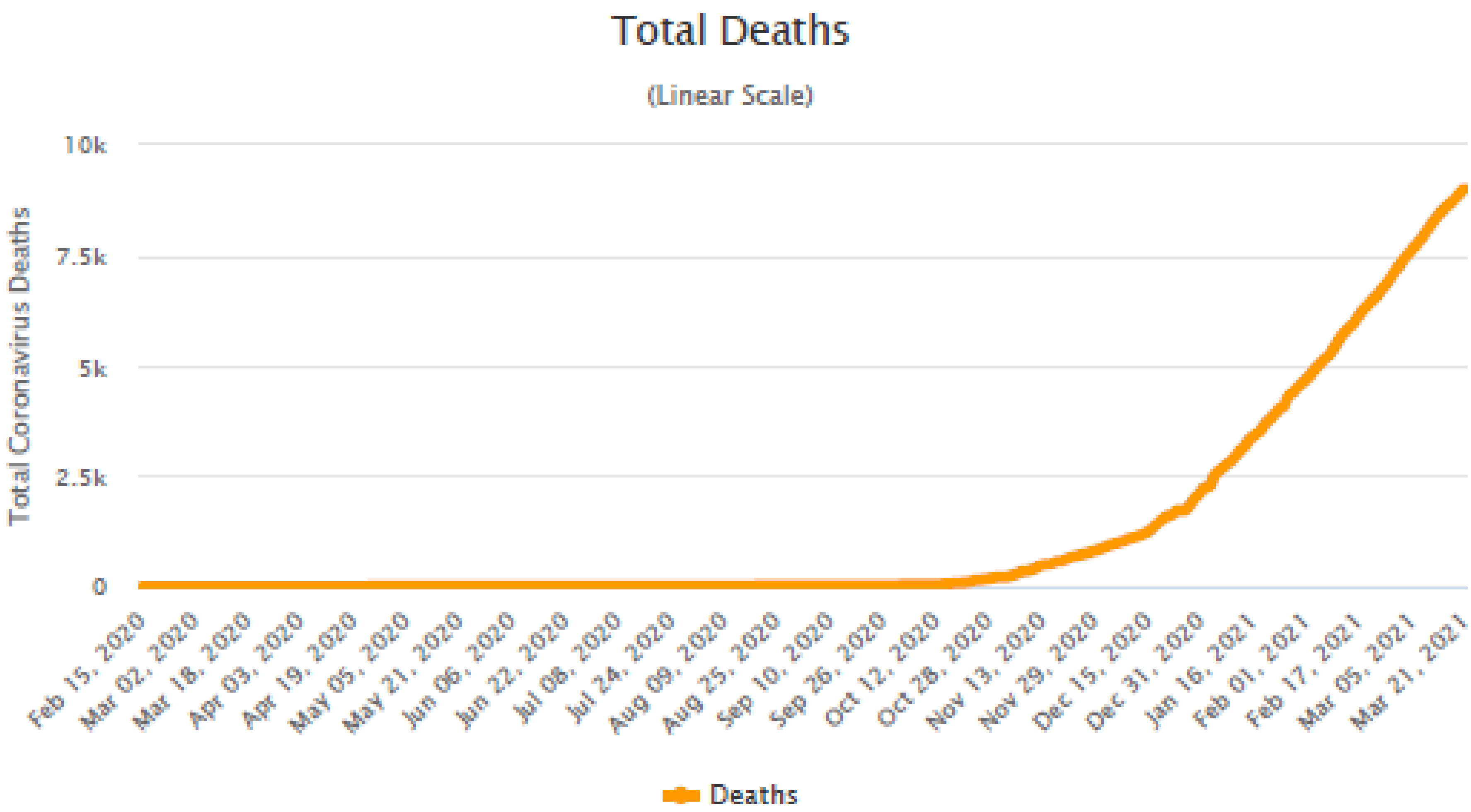
2023. április 20.


Daily New Cases
Cases per Day
Data as of 0:00 GMT+0
1,000k
750k
500k
250k
0
Jan 22, 2020
Feb 08, 2020
Feb 25, 2020
Mar 13, 2020
Mar 30, 2020
Apr 16, 2020
May 03, 2020
May 20, 2020
Jun 06, 2020
Jun 23, 2020
Jul 10, 2020
Jul 27, 2020
Aug 13, 2020
Aug 30, 2020
Sep 16, 2020
Oct 03, 2020
Oct 20, 2020
Nov 06, 2020
Nov 23, 2020
Dec 10, 2020
Dec 27, 2020
Jan 13, 2021
Jan 30, 2021
Feb 16, 2021
Mar 05, 2021


Daily Deaths
Deaths per Day
Data as of 0:00 GMT+0
20k
15k
10k
5k
0
Jan 22, 2...
Feb 07, 2020
Feb 23, 2020
Mar 10, 2020
Mar 26, 2020
Apr 11, 2020
Apr 27, 2020
May 13, 2020
May 29, 2020
Jun 14, 2020
Jun 30, 2020
Jul 16, 2020
Aug 01, 2020
Aug 17, 2020
Sep 02, 2020
Sep 18, 2020
Oct 04, 2020
Oct 20, 2020
Nov 05, 2020
Nov 21, 2020
Dec 07, 2020
Dec 23, 2020
Jan 08, 2021
Jan 24, 2021
Feb 09, 2021
Feb 25, 2021
Mar 13, 2021



Total Deaths
(Linear Scale)
Total Coronavirus Deaths
20k
15k
10k
5k
0
Feb 15, 2020
Feb 28, 2020
Mar 12, 2020
Mar 25, 2020
Apr 07, 2020
Apr 20, 2020
May 03, 2020
May 16, 2020
May 29, 2020
Jun 11, 2020
Jun 24, 2020
Jul 07, 2020
Jul 20, 2020
Aug 02, 2020
Aug 15, 2020
Aug 28, 2020
Sep 10, 2020
Sep 23, 2020
Oct 06, 2020
Oct 19, 2020
Nov 01, 2020
Nov 14, 2020
Nov 27, 2020
Dec 10, 2020
Dec 23, 2020
Jan 05, 2021
Jan 18, 2021
Jan 31, 2021
Feb 13, 2021
Feb 26, 2021
Mar 11, 2021
Deaths



Total Deaths
(Linear Scale)
Total Coronavirus Deaths
10k
7.5k
5k
2.5k
0
Feb 15, 2020
Mar 02, 2020
Mar 18, 2020
Apr 03, 2020
Apr 19, 2020
May 05, 2020
May 21, 2020
Jun 06, 2020
Jun 22, 2020
Jul 08, 2020
Jul 24, 2020
Aug 09, 2020
Aug 25, 2020
Sep 10, 2020
Sep 26, 2020
Oct 12, 2020
Oct 28, 2020
Nov 13, 2020
Nov 29, 2020
Dec 15, 2020
Dec 31, 2020
Jan 16, 2021
Feb 01, 2021
Feb 17, 2021
Mar 05, 2021
Mar 21, 2021
Deaths

Példa

- Milyen valószínűséggel születik fiúgyermek?
- Svájcban 1871 és 1900 között a 2.644.757 megszületett gyermekből 1.359.671 fiú és 1.285.086 lány volt.
- Fiúk relatív gyakorisága így 0,5141.
- Igaz-e, hogy a valószínűség 0,5? És 0,1?

$$X_i = \begin{cases} 1, & i.\,\text{fiú} \\ 0, & i.\,\text{lány} \end{cases} \Rightarrow$$

$$P(X_i = 1) = p, n = 2.644.757, \xi = \frac{\sum_{i=1}^{n} X_i}{n} \Rightarrow$$

$$EX_i = p, D^2 X_i$$

$$= p(1-p), P\left(\frac{\sum_{i=1}^{n} X_i - nEX_1}{DX_1\sqrt{n}} < x\right) \sim \Phi(x) \Rightarrow$$

$$P\left(-u < \sqrt{\frac{n}{p(1-p)}}(\xi - p) < u\right) \sim 2\Phi(u) - 1$$

$$p = 0.5 \Rightarrow \sqrt{\frac{n}{p(1-p)}}(\xi - p) = 37$$

$$u = 4 \Rightarrow 2\Phi(u) - 1 = 0{,}999936$$

$$p(1-p) \le \frac{1}{4} \Rightarrow$$

$$2\Phi(u)-1 \sim P\left(-u < \sqrt{\frac{n}{p(1-p)}}\left(\xi - p\right) < u\right) \le$$

$$\le P\left(-u < 2\sqrt{n}\left(\xi - p\right) < u\right) =$$

$$= P\left(\frac{-u}{2\sqrt{n}} < \left(\xi - p\right) < \frac{u}{2\sqrt{n}}\right) = P\left(\xi - \frac{u}{2\sqrt{n}} < p < \xi + \frac{u}{2\sqrt{n}}\right)$$

Esetünkben 0,9973 valószínűséggel $0{,}5132 \le p \le 0{,}5150$

Statisztikai mező

$$\left(\Omega, \mathcal{A}, P_{\vartheta}\right),\ \vartheta \in \Theta$$

statisztikai mező, ha Θ paraméterhalmaz

és $\left(\Omega, \mathcal{A}, P_{\vartheta}\right)$ minden paraméter

esetén valószínűségi mező.

Egy érmedobás modellje

- Nem ismerjük a fejdobás valószínűségét:

$$\Omega = \{\mathrm{F}, \mathrm{I}\}, A \ = \left\{\emptyset; \{\mathrm{F}\}; \{\ \mathrm{I}\}; \{\mathrm{F}, \mathrm{I}\}\right\},$$

$$P_p\left(\{\mathrm{F}\}\right) = p, P_p\left(\{\mathrm{I}\}\right) = 1 - p, p \in [0,1].$$

Minta

Def.: A $\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} : \Omega \to \mathcal{X}$

$\subseteq \mathbb{R}^n$ valószínűségi vektorváltozót

mintának nevezzük.

n: mintanagyság

ξ_i: i. mintaelem

Def.: minta realizációja:

$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ a konkrét megfigyelt számsorozat.

Mintatér

- Def: $\mathcal{X}$ mintatér: a minta lehetséges értékeinek halmaza. Elemei a mintaértékek.
- *n*-elemű valós minta esetén: $\mathcal{X} = \mathbb{R}^n$
- *n*-elemű pozitív egész értékű minta esetén: $\mathcal{X} = \mathbb{N}^n$
- Példa: egy biztosítónál 10 napon keresztül figyelték a bejelentett károk számát, ekkor $\mathcal{X} = \mathbb{Z}_0^{10}$

Az elmúlt 5 napban elhunyt koronavírusos betegek száma

- Megfigyelések: 0, 1, 2, 2, 1
- Minta és realizációja:

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} \text{ és } \begin{pmatrix} 0 \\ 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}$$

Mintanagyság: 5

A minták típusai

- Független minta: a mintaelemek függetlenek.
- Független azonos eloszlású minta: a mintaelemek függetlenek és azonos eloszlásúak.
- Diszkrét minta: a mintaelemek diszkrétek.
- Abszolút folytonos eloszlású minta: a mintaelemek abszolút folytonosak.

Eloszláscsaládok

$$F_\vartheta(\mathbf{s}) = P_\vartheta\left(\xi_1 < s_1, \ldots, \xi_n < s_n\right)$$

Független minta esetén:

$$F_\vartheta(\mathbf{s}) = \prod_{i=1}^{n} P_\vartheta\left(\xi_i < s_i\right)$$

Független azonos eloszlású minta esetén:

$$F_\vartheta(\mathbf{s}) = \prod_{i=1}^{n} P_\vartheta\left(\xi_i < s_i\right) = \prod_{i=1}^{n} F_\vartheta\left(s_i\right)$$

Jelölések:

E_ϑ: várható érték P_ϑ esetén,

D_ϑ: szórás P_ϑ esetén,

f_ϑ: sűrűségfüggvény P_ϑ esetén (absz. folyt. minta)

$p_\vartheta(s) = P_\vartheta\left(\xi_i = s\right)$ (diszkrét minta)

Példák

- Egy érmedobás. Fej esetén 1-et írunk, írás esetén 0-át.

$$p_p(k) = P_p(\xi_1 = k) = \begin{cases} p & k = 1 \\ 1-p & k = 0 \end{cases} = p^k (1-p)^{1-k}$$

Koronavírusos példa. Azt feltételezzük, hogy megfigyeléseink független, azonos eloszlású Poissonok.

$$p_\lambda(k) = P_\lambda(\xi_i = k) = \lambda^k e^{-\lambda} / k!,\ k = 0, 1, 2, \ldots$$

Statisztikák

Def.: Statisztika: a minta függvénye.

$$T : \mathcal{X} \to \mathbb{R}^k$$

Def'.: Statisztika:

$$T(\boldsymbol{\xi}), \text{ ha } T : \mathcal{X} \to \mathbb{R}^k \text{függvény.}$$

Tapasztalati momentumok

$\mathcal{X} = \mathbb{R}^n$

mintaközép:

$$T(\mathbf{x}) = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad T(\boldsymbol{\xi}) = \bar{\xi} = \frac{\sum_{i=1}^{n} \xi_i}{n},$$

tapasztalati k. momentum:

$$T(\mathbf{x}) = \frac{\sum_{i=1}^{n} x_i^k}{n}, \qquad T(\boldsymbol{\xi}) = \frac{\sum_{i=1}^{n} \xi_i^k}{n}.$$

Tapasztalati szórásnégyzet

$$\mathcal{X} = \mathbb{R}^n,$$

$$T(\mathbf{x}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n},$$

$$T(\boldsymbol{\xi}) = s^2 = \frac{\sum_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)^2}{n}$$

”Kis zh”

1. Mi a tapasztalati eloszlásfüggvény? (2p)
2. Határozza meg egy exponenciális eloszlású valószínűségi változó szórásnégyzetét! (5p)
3. Írja fel a Csebisev-egyenlőtlenséget! (2p)
4. Mi a sztochasztikus konvergencia? (2p)
5. Írja fel a Centrális határeloszlás tételt! (4p)