

# A sztochasztika alapjai

## MBNXK262

13. előadás: Függő változók

Kevei Péter

2023/24 tavasz

## Feltételes függetlenség

Tünetekből szeretnénk betegségekre következtetni. Tegyük fel, hogy allergia esetén az emberek 30%-a lázasodik be, míg Covid-19 esetén 80%-a. Tegyük fel továbbá, hogy az emberek 20%-a allergiás, és 10%-a covidos, és nem lehet mindkettő.

| betegség | láz | gyakoriság |
|----------|-----|------------|
| Allergia | 0,3 | 0,2        |
| Covid    | 0,8 | 0,1        |

| betegség | láz | gyakoriság |
|----------|-----|------------|
| Allergia | 0,3 | 0,2        |
| Covid    | 0,8 | 0,1        |

$$P(\text{allergia}|\text{láz}) =$$

$$P(\text{covid}|\text{láz}) =$$

## Több tünet

|          | Láz | Orrfolyás | Köhögés | gyakoriság |
|----------|-----|-----------|---------|------------|
| Allergia | 0,3 | 0,7       | 0,3     | 0,2        |
| Covid    | 0,8 | 0,2       | 0,7     | 0,1        |

Tegyük fel, hogy a tünetek egymástól függetlenek adott megbetegedés esetén, azaz

$$\mathbf{P}(\text{láz és orrfolyás}|\text{allergia}) = \mathbf{P}(\text{láz}|\text{allergia}) \cdot \mathbf{P}(\text{orrfolyás}|\text{allergia}).$$

|          | Láz | Orrfolyás | Köhögés | gyakoriság |
|----------|-----|-----------|---------|------------|
| Allergia | 0,3 | 0,7       | 0,3     | 0,2        |
| Covid    | 0,8 | 0,2       | 0,7     | 0,1        |

Egy beteg lázas és köhög, de nem folyik az orra. Mennyi a valószínűsége, hogy covidos?

$$\begin{aligned}
 & \mathbf{P}(\text{covid}|\text{láz, köhögés, nincs orrfolyás}) \\
 &= \frac{\mathbf{P}(\text{covid}) \cdot \mathbf{P}(\text{láz, köhögés, nincs orrfolyás}|\text{covid})}{\mathbf{P}(\text{láz, köhögés, nincs orrfolyás})}.
 \end{aligned}$$

A feltételes függetlenség szerint

$$\begin{aligned}
 & \mathbf{P}(\text{láz, köhögés, nincs orrfolyás}|\text{covid}) \\
 &= \mathbf{P}(\text{láz}|\text{covid}) \cdot \mathbf{P}(\text{köhögés}|\text{covid}) \cdot \mathbf{P}(\text{nincs orrfolyás}|\text{covid}) \\
 &= 0,8 \cdot 0,7 \cdot 0,8 = 0,448.
 \end{aligned}$$

A teljes valószínűség tétele szerint

$$\begin{aligned} & \mathbf{P}(\text{láz, köhögés, nincs orrfolyás}) \\ &= \mathbf{P}(\text{covid}) \cdot \mathbf{P}(\text{láz, köhögés, nincs orrfolyás}|\text{covid}) \\ &\quad + \mathbf{P}(\text{allergia}) \cdot \mathbf{P}(\text{láz, köhögés, nincs orrfolyás}|\text{allergia}) \\ &= 0,1 \cdot 0,448 + 0,2 \cdot 0,3 \cdot 0,3 \cdot 0,3 = 0,0502. \end{aligned}$$

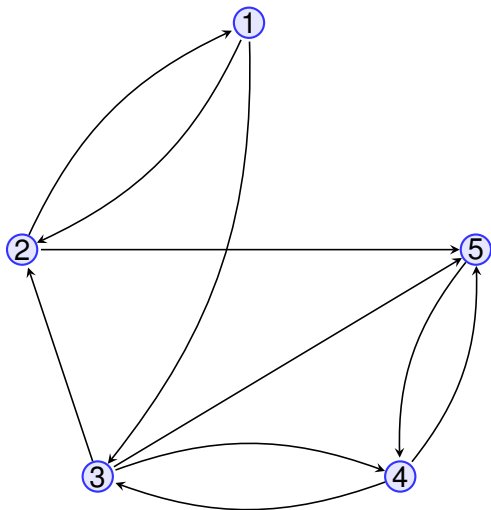
A kapott eredményt visszaírva

$$\mathbf{P}(\text{covid}|\text{láz, köhögés, nincs orrfolyás}) = \frac{0,0448}{0,0502} = 0,89.$$

Tehát ekkor már 89% a covid esélye.

Bayesi-hálókról -> Mesterséges intelligencia kurzus

# PageRank



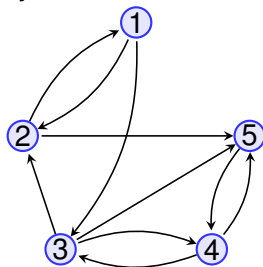
# PageRank - véletlen szörfös

Legyen az 1-es a kezdőoldalunk, és jelölje  $\xi_i$ ,  $i = 1, 2, \dots$ , azt az oldalt, ahova az  $i$ . lépés (kattintás) után jutunk.

$$\mathbf{P}(\xi_1 = 2) = \quad , \mathbf{P}(\xi_1 = 3) = \quad ,$$

$$\mathbf{P}(\xi_2 = 1) =$$

$$\mathbf{P}(\xi_2 = 2) =$$





# Markov-lánc

$\xi_1, \xi_2, \xi_3, \dots$  véletlen változók sorozata Markov-lánc, ha

$$\mathbf{P}(\xi_k = x_k | \xi_{k-1} = x_{k-1}, \dots, \xi_1 = x_1) = \mathbf{P}(\xi_k = x_k | \xi_{k-1} = x_{k-1}).$$

Átmenetvalószínűségek:

$$p_{j,i} = \mathbf{P}(\xi_k = j | \xi_{k-1} = i).$$

Átmenetmátrix:

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 1 \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} & 0 \end{pmatrix}$$

# Markov-lánc

$$P^2 = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{6} & 0 & 0 \\ \frac{1}{6} & \frac{1}{4} & 0 & \frac{1}{6} & 0 \\ 0 & \frac{1}{4} & \frac{1}{6} & 0 & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{5}{12} & 0 & \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{pmatrix}$$

Megmutatható, hogy  $\lim_{n \rightarrow \infty} P^n$  létezik, és a határérték mátrix minden oszlopa megegyezik  $(\pi_1, \dots, \pi_5)^\top$  sajátvektorral.

# PageRank

Az  $i$ . oldal fontossága  $\pi_i \geq 0$ , és  $\sum_{i=1}^5 \pi_i = 1$ .

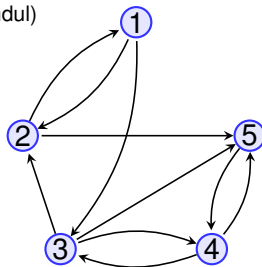
$$\pi_1 = \frac{1}{2}\pi_2 \quad (1\text{-be csak a 2-ből fut él, és onnan 2 él indul})$$

$$\pi_2 = \frac{1}{2}\pi_1 + \frac{1}{3}\pi_3 \quad (2\text{-be az 1-ből és a 3-ból fut él})$$

$$\pi_3 = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_4$$

$$\pi_4 = \frac{1}{3}\pi_3 + 1 \cdot \pi_5$$

$$\pi_5 = \frac{1}{2}\pi_3 + \frac{1}{3}\pi_3 + \frac{1}{2}\pi_4.$$



# PageRank

Mátrix alakban:

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 1 \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{pmatrix}.$$

# PageRank

Kurt Bryan, Tanya Leise: The \$25,000,000,000 eigenvector.  
The linear algebra behind Google.



By Joi Ito, via Wikimedia Commons



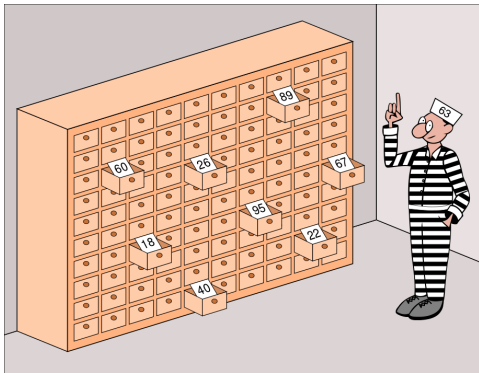
By Marcin Mycielski, from Wikimedia Commons

## 100 rab problémája

100 halálraítélt, 1-től 100-ig megszámozott rabnak a börtönigazgató ad egy utolsó esélyt. Az igazgató egy 100 fiókos (ugyan-csak 1-től 100-ig számozott) szekrény fiókjaiba beteszi (véletlen sorrendben) a számokat. Minden fiókba pontosan egy szám kerül. Ezután a rabok egyesével bemennek a szobába, ahol ki-nyithatnak 50 fiókot. A fiókokat visszacsukják, a szekrényben nem rendezhetik át a számokat, mindent úgy hagynak, ahogy volt. Ha *minden* rab megtalálja a saját számát, akkor kiszaba-dulnak, különben mindegyiküket kivégzik. (Gál Anna, Peter Bro Miltersen (2003))

# 100 rab problémája

- ▶ egyesével
- ▶ minden rab pontosan ugyanúgy hagyja a szobát
- ▶ 50 fiókot nyithat ki
- ▶ előtte lehet beszélni
- ▶ utána, közben nem
- ▶ mindenkinek meg kell találnia a számát



Forrás: Wikimedia

# 1. stratégia

Minden rab véletlenszerűen kinyit 50 fiókot.  
Mennyi a szabadulás esélye?



## 2. stratégia

Minden rab először kinyitja a sorszámanak megfelelő fiókot.  
Aztán kinyitja azt a fiókot, amelyiknek a sorszámat megtalálta az előbb, és ezt folytatja.

## Példa (6 fiók, 3 próba)

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 2 & 1 & 6 \end{pmatrix}$$

- ▶ 1. rab:  $1 \rightarrow 3 \rightarrow 5$
- ▶ 2. rab:  $2 \rightarrow 4$
- ▶ 3. rab:  $3 \rightarrow 5 \rightarrow 1$
- ▶ 4. rab:  $4 \rightarrow 2$
- ▶ 5. rab:  $5 \rightarrow 1 \rightarrow 3$
- ▶ 6. rab: 6 Hurrá!



## 2. Példa (6 fiók, 3 próba)

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 1 & 5 & 6 & 2 & 4 \end{pmatrix}$$

► 1. rab:  $1 \rightarrow 3 \rightarrow 5$  vége



## Szabadulás valószínűsége

A ciklusstratégia pontosan akkor nyer, ha a leghosszabb ciklus 50-nél nem hosszabb.

Azon permutációk száma, melyben a leghosszabb ciklus 51:

## Szabadulás valószínűsége

Hasonlóan, azon permutációk száma, melyben a leghosszabb ciklus  $k$ ,  $k \geq 51$ :

$$\binom{100}{k} (k-1)! (100-k)! = \frac{100!}{k! (100-k)!} \cdot (k-1)! \cdot (100-k)! = \frac{100!}{k}.$$

Összegezve, azon permutációk száma, melyben a leghosszabb ciklus legalább 51:

$$100! \left[ \frac{1}{51} + \frac{1}{52} + \dots + \frac{1}{100} \right].$$

## Szabadulás valószínűsége

Hasonlóan, azon permutációk száma, melyben a leghosszabb ciklus  $k$ ,  $k \geq 51$ :

$$\binom{100}{k} (k-1)! (100-k)! = \frac{100!}{k!(100-k)!} \cdot (k-1)! \cdot (100-k)! = \frac{100!}{k}.$$

Összegezve, azon permutációk száma, melyben a leghosszabb ciklus legalább 51:

$$100! \left[ \frac{1}{51} + \frac{1}{52} + \dots + \frac{1}{100} \right].$$

Tehát, annak a valószínűsége, hogy a leghosszabb ciklus legalább 51:

$$\frac{1}{51} + \frac{1}{52} + \dots + \frac{1}{100}$$

# Szabadulás valószínűsége

Annak a valószínűsége, hogy a leghosszabb ciklus legalább 51:

$$\frac{1}{51} + \frac{1}{52} + \dots + \frac{1}{100}$$

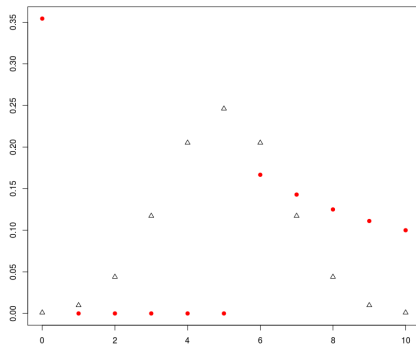
Tehát annak a valószínűsége, hogy a leghosszabb ciklus nem hosszabb, mint 50,

$$1 - \left[ \frac{1}{51} + \frac{1}{52} + \dots + \frac{1}{100} \right] \approx 1 - \ln 2 = 0,31,$$

ami éppen a szabadulás valószínűsége.  
Ez optimális.

# Mi történt?

Nemtalált rabok számának eloszlása (10 rab, 5 próba):



Várhatóan mindkét esetben 5 rab nem talál.