A sztochasztika alapjai
MBNXK262

3. előadás: Feltételes valószínűség, függetlenség

Kevei Péter

2022/23 tavasz

Teljes eseményrendszer

Definíció

A $B_1, B_2, \ldots$ események teljes eseményrendszert alkotnak, ha

- minden $i \neq j$ párra $B_i \cap B_j = \emptyset$;
- $\cup_{n=1}^{\infty} B_n = \Omega$.

Tétel (Teljes valószínűség tétele)

Legyen $B_1, B_2, \ldots$ teljes eseményrendszer, melyre $\mathbf{P}(B_n) > 0$ minden n-re. Ekkor tetszőleges $A \in \mathcal{A}$ esemény esetén

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(A|B_n)\mathbf{P}(B_n).$$

Bayes-formula

Tétel

Legyenek A és B olyan események, hogy $\mathbf{P}(A) > 0$, $\mathbf{P}(B) > 0$. *Ekkor*

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(A)}.$$

Bayes-tétel

Tétel

Legyen $B_1, B_2, \ldots$ *teljes eseményrendszer, melyre* $\mathbf{P}(B_n) > 0$ *minden* n*-re. Ekkor tetszőleges pozitív valószínűségű* $A \in \mathcal{A}$ *esemény esetén, tetszőleges* k*-ra*

$$\mathbf{P}(B_k|A) = \frac{\mathbf{P}(A|B_k)\mathbf{P}(B_k)}{\sum_{n=1}^{\infty} \mathbf{P}(A|B_n)\mathbf{P}(B_n)}.$$

Monty Hall probléma

Egy játékos 3 csukott ajtó közül 1-et választhat. Kettő mögött kecske van, egy mögött autó, de nem tudja, hogy melyik hol van. A játékos autót szeretne nyerni. Miután egy ajtót kiválaszt, a játékvezető kinyittat egy ajtót, amit nem választott a játékos, és ami mögött kecske van. Ezután felajánlja a játékosnak, hogy cserélhet ajtót. Megéri-e váltani?

Monty Hall probléma

Tegyük fel, hogy a játékos az 1-es ajtót választotta, a játékvezető a 2-est nyittatta ki. Jelölje A_i azt az eseményt, hogy az autó az i-edik ajtó mögött van, B pedig azt, hogy a játékvezető a 2-es ajtót nyittatta ki.

Doppingteszt

Kifejlesztenek egy új doppingtesztet, mely a doppingolók 99%-ánál pozitív eredményt ad, azonban a nem doppingoló sportolók 1%-nál is tévesen pozitív eredményt ad. Tegyük föl, hogy a sportolók 1%-a doppingol. Mennyi annak a valószínűsége, hogy egy véletlenül kiválasztott sportoló

(a) doppingtesztje pozitív?

(b) doppingolt, ha tudjuk, hogy a doppingtesztje pozitív?

T: a teszt eredménye pozitív; D: a sportoló doppingolt

Doppingteszt

Szindbád

Szindbádnak jogában áll N háremhölgy közül egyet kiválasztania oly módon, hogy az előtte egyenként elvonuló hölgyek valamelyikére rámutat. Tegyük fel, hogy egyértelmű szigorúan monoton szépségi sorrendet tud felállítani, és a háremhölgyek bármely elvonulási sorrendje egyformán valószínű. Szindbád k hölgyet elenged, majd kiválasztja az elsőt, aki szebb az összes előtte elvonultnál. Mennyi a valószínűsége, hogy a legszebb hölgyet választja ki? Milyen k esetén lesz ez a valószínűség a legnagyobb, ha N elég nagy?

Szindbád

Jelölje A_i azt az eseményt, hogy a i-edik lány a legszebb, $i = 1, 2, \ldots, N$, és B azt, hogy Szindbád a legszebb lányt választja.

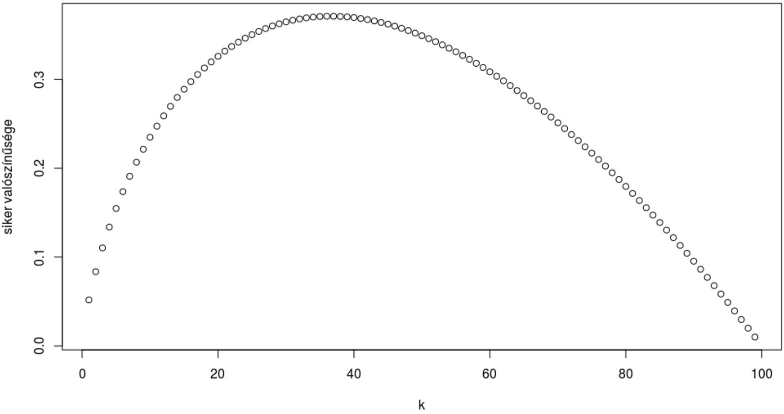
Szindbád

Ha $i > k$, akkor Szindbád pontosan akkor választja a legszebbet, ha az első $i-1$ lány közül a legszebb az első k-ban volt.
Tehát
$$\mathbf{P}(B|A_i) = \frac{k}{i-1}.$$

Összegezve

$$\mathbf{P}(B) = \sum_{i=1}^{N} \mathbf{P}(A_i)\,\mathbf{P}(B|A_i) = \sum_{i=k+1}^{N} \frac{1}{N}\frac{k}{i-1}.$$

Szindbád



ábra: A siker valószínűsége k függvényében, $N = 100$

Szindbád

És ez mennyi?

$$\mathbf{P}(B) = \frac{k}{N}\sum_{i=k}^{N-1}\frac{1}{i}.$$

Integrálközelítő összeg becslése integrállal:

$$\log\frac{N}{k} = \int_k^N \frac{1}{x}\mathrm{d}x \le \sum_{i=k}^{N-1}\frac{1}{i} \le \int_{k-1}^{N-1}\frac{1}{x}\mathrm{d}x = \log\frac{N-1}{k-1}.$$

Tehát

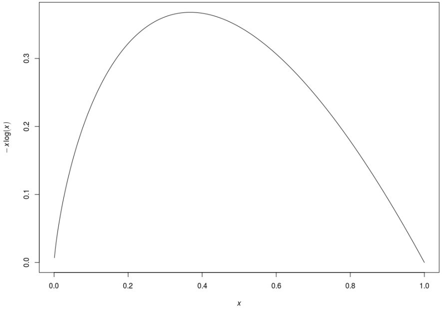
$$\mathbf{P}(B) \sim \frac{k}{N}\log\frac{N}{k}.$$

Szindbád

Hogy válassza Szindbád k-t?

$$\mathbf{P}(B) \sim \frac{k}{N} \log \frac{N}{k}.$$

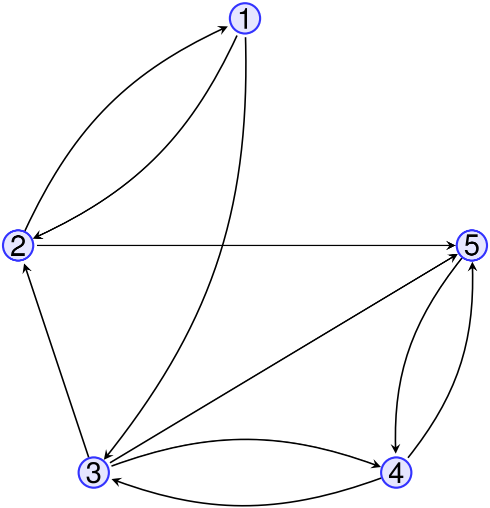
Az $f(x) = x \log \frac{1}{x}$ függvény maximuma kell $[0, 1]$-en. HF
Maximumhely $x = 1/e$,
maximum értéke: $1/e$,
tehát $k = [N/e]$.



Szindbád

- Szindbád megtudja N-et.
- Elenged $k = [N/e]$ lányt.
- Majd kiválasztja az elsőt, aki szebb az összes eddiginél.
- Ekkor nagyjából $1/e \approx 0,368$ valószínűséggel a legszebbbbet választja.

PageRank



PageRank - véletlen szörfös

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

$$\mathbf{P}(\xi_1 = 2) = \mathbf{P}(\xi_1 = 3) = \frac{1}{2},$$

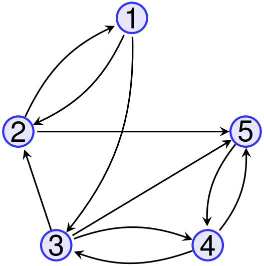
$$\mathbf{P}(\xi_2 = 1) = \mathbf{P}(\xi_2 = 1 | \xi_1 = 2) \cdot \mathbf{P}(\xi_1 = 2)$$

$$= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\mathbf{P}(\xi_2 = 5) = \mathbf{P}(\xi_2 = 5 | \xi_1 = 2) \cdot \mathbf{P}(\xi_1 = 2)$$

$$+ \mathbf{P}(\xi_2 = 5 | \xi_1 = 3) \cdot \mathbf{P}(\xi_1 = 3)$$

$$= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{12}.$$



PageRank

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

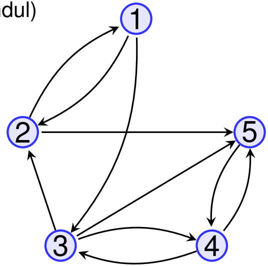
$$\pi_1 = \frac{1}{2}\pi_2 \quad \text{(1-be csak a 2-ből fut él, és onnan 2 él indul)}$$

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

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

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

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



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

Függetlenség

A B esemény bekövetkezése nem befolyásolja az A bekövetkezését, ha

$$\mathbf{P}(A|B) = \mathbf{P}(A),$$

ahonnan $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Definíció

Az A és B események függetlenek, ha $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Definíció

Az A és B események függetlenek, ha $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

- a függetlenség szimmetrikus
- a biztos és a lehetetlen eseménytől minden esemény független HF

Példa

Francia kártyapakliból húzunk egy lapot. D: dámát húzunk, K kőrt húzunk.
Ekkor $D \cap K$ az az esemény, hogy a kőr dámát húztuk ki, így $\mathbf{P}(D \cap K) = 1/52$. Ugyanakkor $\mathbf{P}(D) = 4/52 = 1/13$ és $\mathbf{P}(K) = 13/52 = 1/4$, azaz a két esemény független.

Három esemény függetlensége

Definíció

Az A, B, C események **függetlenek**, ha

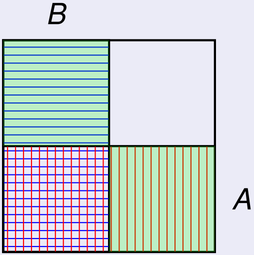
- $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$,
- $\mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C)$,
- $\mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C)$,
- és $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$.

Továbbá, az A, B, C események **páronként függetlenek**, ha bármely kettő független.

Példa

Válasszunk egyenletes eloszlás szerint egy pontot a $[0,1]^2$ egységnégyzetben.

- A: a pont a $[0,1] \times [0,1/2]$ téglalapba esik (piros);
- B: a pont az $[0,1/2] \times [0,1]$ téglalapba esik (kék);
- C: a pont a $[0,1/2] \times [1/2,1] \cup [1/2,1] \times [0,1/2]$ halmazba esik (zöld);



Könnyen ellenőrizhető, hogy A, B, C páronként függetlenek, de *nem* függetlenek. HF

Több esemény függetlensége

Definíció

Az $A_1, A_2, \ldots, A_n$ események **függetlenek**, ha bármely $k \in \{2, 3, \ldots, n\}$ és $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ esetén

$$\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \ldots \mathbf{P}(A_{i_k}).$$

Végtelen sok esemény akkor független, ha közülük bármely véges sok független.

Állítás

Ha az $A_1, \ldots, A_n$ események függetlenek, akkor tetszőleges $k \in \{1, 2, \ldots, n\}$ esetén az $\{A_1, \ldots, A_k\}$ eseményekből ill. az $\{A_{k+1}, \ldots, A_n\}$ eseményekből alkotott események függetlenek.

Ezt nem bizonyítjuk. Az állítás szerint például ha A, B, C, D független események, akkor $A \cup B$ és $C \cap D$ is függetlenek. HF

Állítás

Független események közül ha néhányat kicserélünk a komplementerére, akkor is független eseményeket kapunk.

Üzenetküldés zajos csatornán

$0-1$ sorozatot akarunk küldeni zajos csatornán keresztül. Tegyük fel, hogy a csatorna minden bitet egymástól függetlenül p valószínűséggel elront, azaz a 0-ból 1-est, az 1-ből 0-át csinál. Ő egy *bináris szimmetrikus csatorna*.

Üzenetküldés zajos csatornán

$$\mathbf{P}(\text{ 5 bit hiba nélkül átmegy }) =$$

Ez nem túl jó, ha $p = 0,1$, akkor ez a valószínűség $0,59$, míg $p = 0,2$ esetén csak $0,33$.

Üzenetküldés zajos csatornán

$$\mathbf{P}(\text{ 5 bit hiba nélkül átmegy }) =$$

Ez nem túl jó, ha $p = 0,1$, akkor ez a valószínűség $0,59$, míg $p = 0,2$ esetén csak $0,33$. Minden bitet háromszor küldünk el, tehát a 0 helyett 000, az 1 helyett 111.

$$\begin{aligned} p' &= \mathbf{P}(\text{ hibásan dekódol }) \\ &= \mathbf{P}(\text{ pontosan 2 bit fordul át }) + \mathbf{P}(\text{ mindhárom átfordul }) \\ &= \end{aligned}$$

Ha $p =$0,1, akkor $p' =$0,028.

Üzenetküldés zajos csatornán

$$\mathbf{P}(\text{ 5 bit hiba nélkül átmegy }) =$$

Ez nem túl jó, ha $p = 0,1$, akkor ez a valószínűség $0,59$, míg $p = 0,2$ esetén csak $0,33$. Minden bitet háromszor küldünk el, tehát a 0 helyett 000, az 1 helyett 111.

$$\begin{aligned} p' &= \mathbf{P}(\text{ hibásan dekódol }) \\ &= \mathbf{P}(\text{ pontosan 2 bit fordul át }) + \mathbf{P}(\text{ mindhárom átfordul }) \\ &= \end{aligned}$$

Ha $p =$0,1, akkor $p' =$0,028.

$$\mathbf{P}(\text{ 5 bitnyi üzenet háromszorozva hiba nélkül átmegy })$$
$$= (1 - p')^5 = \left(1 - 3p^2 + 2p^3\right)^5.$$

0,86, ha $p =$0,1.