Modellezés magasabb rendű Markov láncokkal

Egyed Tünde
(Támavezető: Csiszár Villő)

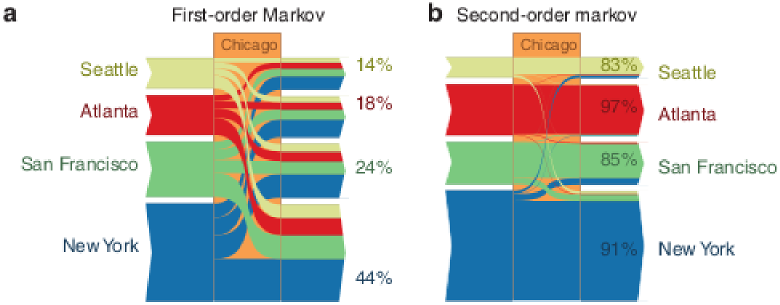
Eötvös Loránd Tudományegyetem
Természettudományi Kar

Önálló projekt 1
2021. december, Budapest

Bevezetés

- Elsőrendű Markov modell: a folyamat jövőbeli értéke csak a jelenbeli értéktől függ.
- Magasabb, k-ad rendű Markov modell: a folyamat k memóriájú, vagyis a következő lépés az előző k lépéstől függ.

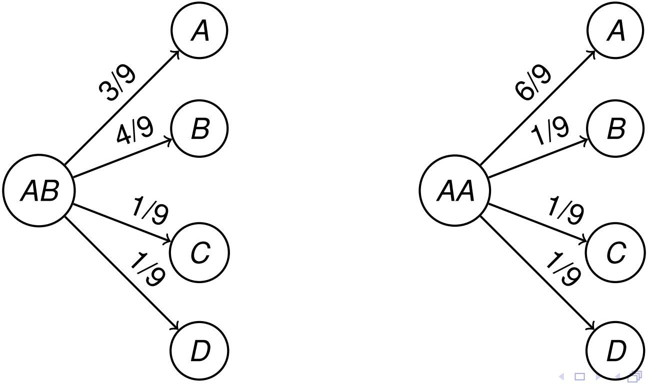
A másodrendű modellezés megmutat olyan mintákat, amiket a memória nélküli nem.



Cél: az optimális rend megtalálása

Szimuláció

- A modellezést egy szimulált mintán végeztem
- Generáltam 20 darab 100 elemű mintát
- Négy különböző lehetséges érték
- Az n. elem generálásánál az $n-1$. elemet négyszeres, az $n-2$. elemet háromszoros, a többi elemet egyszeres súllyal vettem figyelembe



Maximum likelihood

- Felírjuk a likelihood függvényt, ami az átmenetek valószínűségeinek a szorzata:

$$L(\theta_k, x) = p(x_1) \prod_i \prod_j p_{ij}^{n_{ij}}$$

- Belátható, hogy az átmenetvalószínűségek maximum likelihood becslései a relatív gyakoriságok
- Elsőrendű loglikelihood: -2 286
  Másodrendű loglikelihood: -2 092
- Szignifikancia tesztelése valószínűséghányados próbával: null modell $k = 1$, altenatív modell $k = 2$
- Eredményként azt kapjuk, hogy $p = 0$, vagyis a másodrendű modell valóban jobb.

Kétlépéses módszer

- Kétlépéses visszatérési valószínűségek becslése
- Stacionárius eloszlás kiszámítása
- A kétlépéses visszatérési valószínűségek összege a stacionárius eloszlással súlyozva

$$P(X_n = X_{n+2}) = \sum_{i,j} \pi_{ij} P(X_{n+2} = i | X_n = i, X_{n+1} = j)$$

- Eredmények összevetése a kétlépéses visszatérések relatív gyakoriságával
  Elsőrendű: 32,41%
  Másodrendű: 53,5%
  Relatív gyakoriság a mintában: 53,31%

Köszönöm a figyelmet!