



Programovanie v jazyku Python

výpočtové modely a simulácia
prednáška 10

Katedra kybernetiky a umelej inteligencie
Technická univerzita v Košiciach
Ing. Ján Magyar

Biased random walk

- v prípade random walku každý prvok populácie je vybraný s rovnakou pravdepodobnosťou
- biased random walk je stochastická metóda, kde výber nie je uniformný, vieme definovať rozdelenie
- niekedy výber je časovo závislý alebo závisí od predošlého výberu

Veľkosť vzorky

- pri simuláciách je kľúčový problém určenie potrebného počtu iterácií a potrebnej veľkosti vzoriek
- s reprezentatívnymi vzorkami nikdy nedosiahneme 100% presnosť
- šancu na nevyrovnanú vzorku vieme eliminovať len ak uvažujeme celú populáciu
- distribúcia hodnôt vo vzorke by mala byť približne rovnaká, ako distribúcia hodnôt v celej populácii

Distribúcia hodnôt - hod jednou kockou

Aká bude distribúcia hodnôt pri hode jednou kockou 1000 krát?

Aká bude pravdepodobnosť toho, že padne 1?

Aká bude pravdepodobnosť toho, že padne 2?

Uniformné rozdelenie

- každá hodnota (alebo skupina hodnôt) je reprezentovaná v rovnakom pomere v populácii
- každá hodnota bude výsledkom pokusu s rovnakou pravdepodobnosťou
- rozdelenie vieme popísať pomocou intervalu (najmenšia a najväčšia hodnota)
- nie je časté v skutočnosti, iba v modeloch



Distribúcia hodnôt - hod dvoma kockami

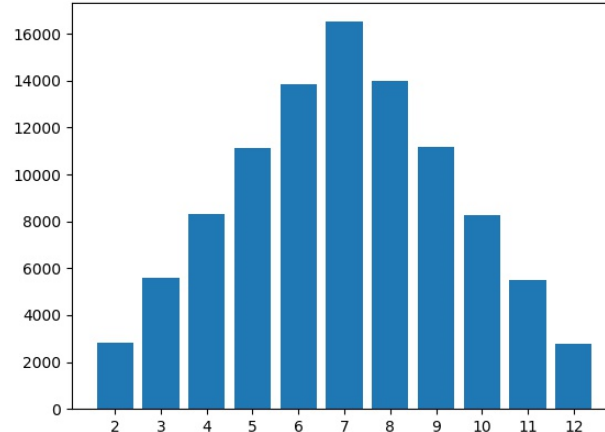
Aká bude distribúcia súčtu hodnôt pri hode dvoma kockami 1000 krát?

Aká bude pravdepodobnosť toho, že súčet čísel bude 2?

Aká bude pravdepodobnosť toho, že súčet čísel bude 7?

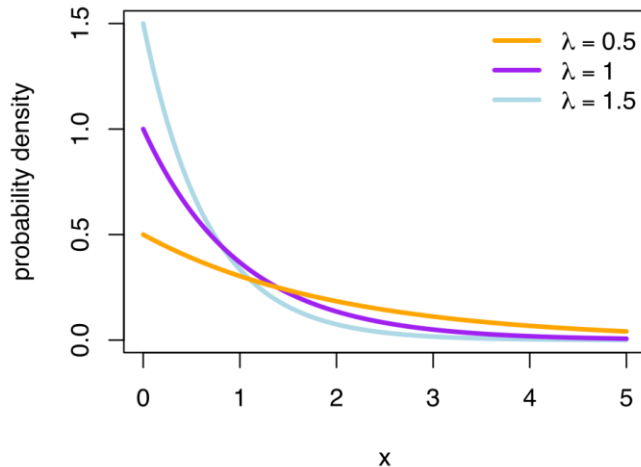
Normálne rozdelenie

- Gaussovo rozdelenie
- stabilné rozdelenie - dostaneme približne rovnaký výsledok aj pri viacerých pokusoch
- najpravdepodobnejšie je priemerná hodnota (súčasne aj medián)
- rozdelenie vieme popísať pomocou priemeru a smerodajnej odchýlky
- asi najčastejšie rozdelenie v prírode



Exponenciálne rozdelenie

- nesimetrické rozdelenie
- najpravdepodobnejšia je najmenšia hodnota
- vieme ho popísať konštantou λ
- vyjadruje rozdelenie uplynulého času medzi dvoma inštanciami toho istého javu, ktorý nastane s konštantou priemernou frekvenciou (doprava, návštevnosť web stránky)



Náhodnosť v Pythone

- metódy implementované v module `random`
- defaultne založené na systémovom čase (v ms)
- môže použiť OS-špecifické zdroje náhodnosti
- `random.seed(a=None, version=2)`
 - inicializácia generátora náhodných hodnôt
 - `a` - parameter generovania
 - `version` - rozsah seed hodnôt
 - podpora reprodukovanie výsledkov

Generovanie celých čísel

- `random.randint(a, b)`
 - náhodné číslo N , kde $a \leq N \leq b$
- `random.randrange([start,] stop[, step])`
 - náhodný prvok z `range(start, stop, step)`
 - reálne sa nevygeneruje objekt typu `range`
 - viac uniformné rozdelenie, ako `int(random.random() * n)`

Náhodnosť so sekvenciami

- `random.choice(seq)`
 - vyberie náhodný prvok zo sekvencie
- `random.choices(population, weights=None, cum_weights=None, k=1)`
 - vyberie k prvkov z populácie a vráti ich v jednom zozname
 - `weights/cum_weights` - definuje pravdepodobnosť výberu jednotlivých prvkov
- `random.shuffle(x[, random])`
 - pomieša prvky sekvencie v náhodnom poradí
 - ak sekvencia je nemenná, je možné zavolať metódu `sample`
- `random.sample(population, k)`
 - vyberie k jedinečných prvkov z populácie
 - vracia nový zoznam s hodnotami

Generovanie desatinných čísel

- `random.random()`
 - náhodné desatinné číslo z rozsahu $[0.0, 1.0)$
- `random.uniform(a, b)`
 - náhodné desatinné číslo z rozsahu $[a, b]$ alebo $[b, a]$
- `random.gauss(mu, sigma)`
 - náhodné číslo z normálneho rozdelenia
 - `mu` - najpravdepodobnejšia hodnota
 - `sigma` - smerodajná odchýlka
- `random.expovariate(lambd)`
 - náhodné číslo z exponenciálneho rozdelenia
 - čísla sú z rozsahu $[0, \infty)$ ak `lambd > 0`, $(-\infty, 0]$ ak `lambd < 0`

Craps / Kocky

Craps je populárna kocková hra v kasínach, kde hráč hodí dvoma kockami naraz a vyhrá podľa nasledovných pravidiel:

- ak súčet čísel je 7 alebo 11, vyhrá
- ak súčet čísel je 2, 3 alebo 12, prehrá
- v ostatných prípadoch súčet čísel sa stane cieľovým počtom bodov, a hráč hodí kockami dovedy, kým nehodí
 - 7, v tomto prípade prehrá
 - cieľový počet bodov, v tomto prípade vyhrá

Akú šancu má hráč na výhru? Má kasíno väčšiu šancu?

Craps / Kocky

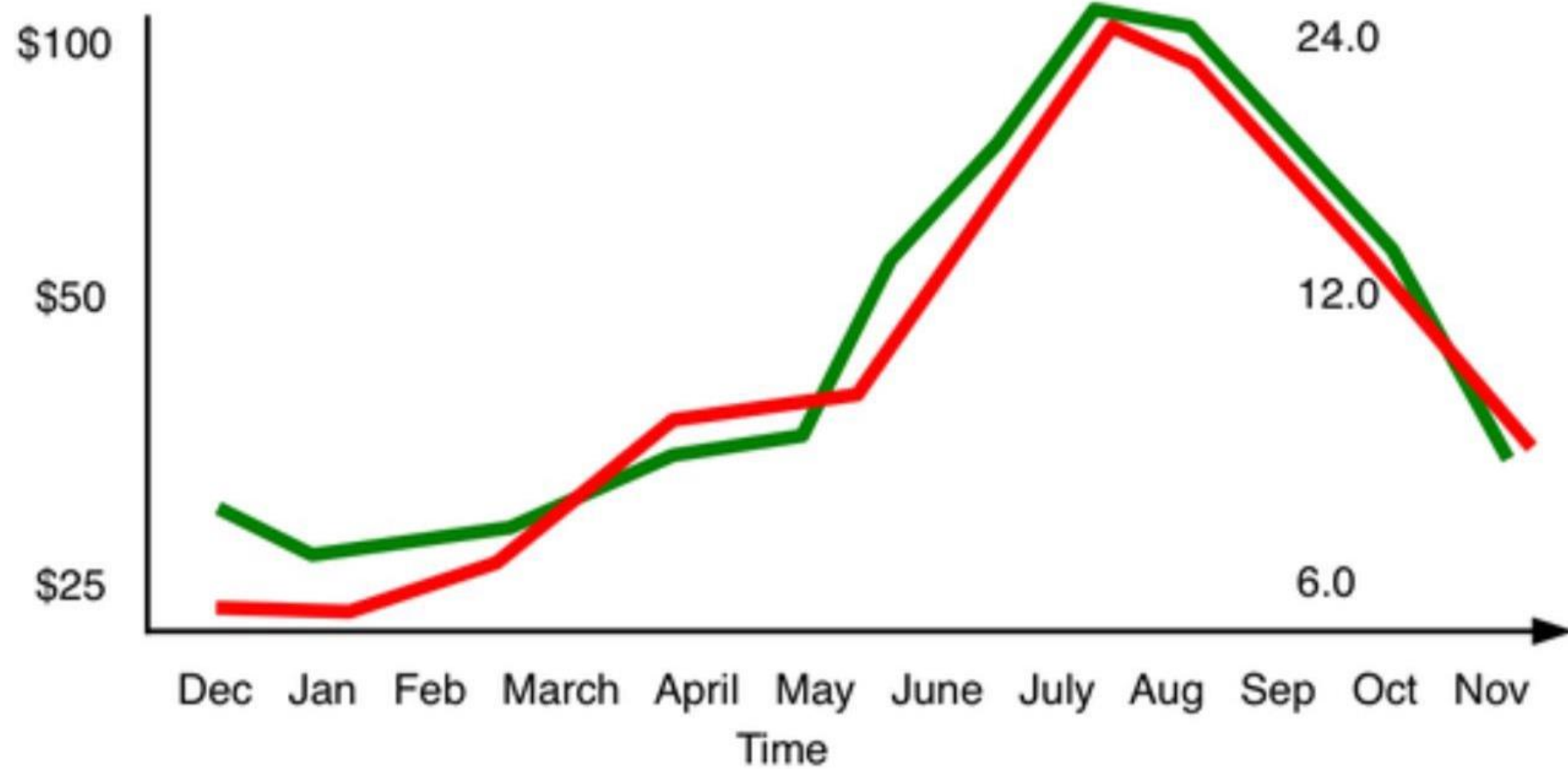
- Ako zvýšime šancu výhry hráča?
- Ako zvýšime šancu výhry kasína?

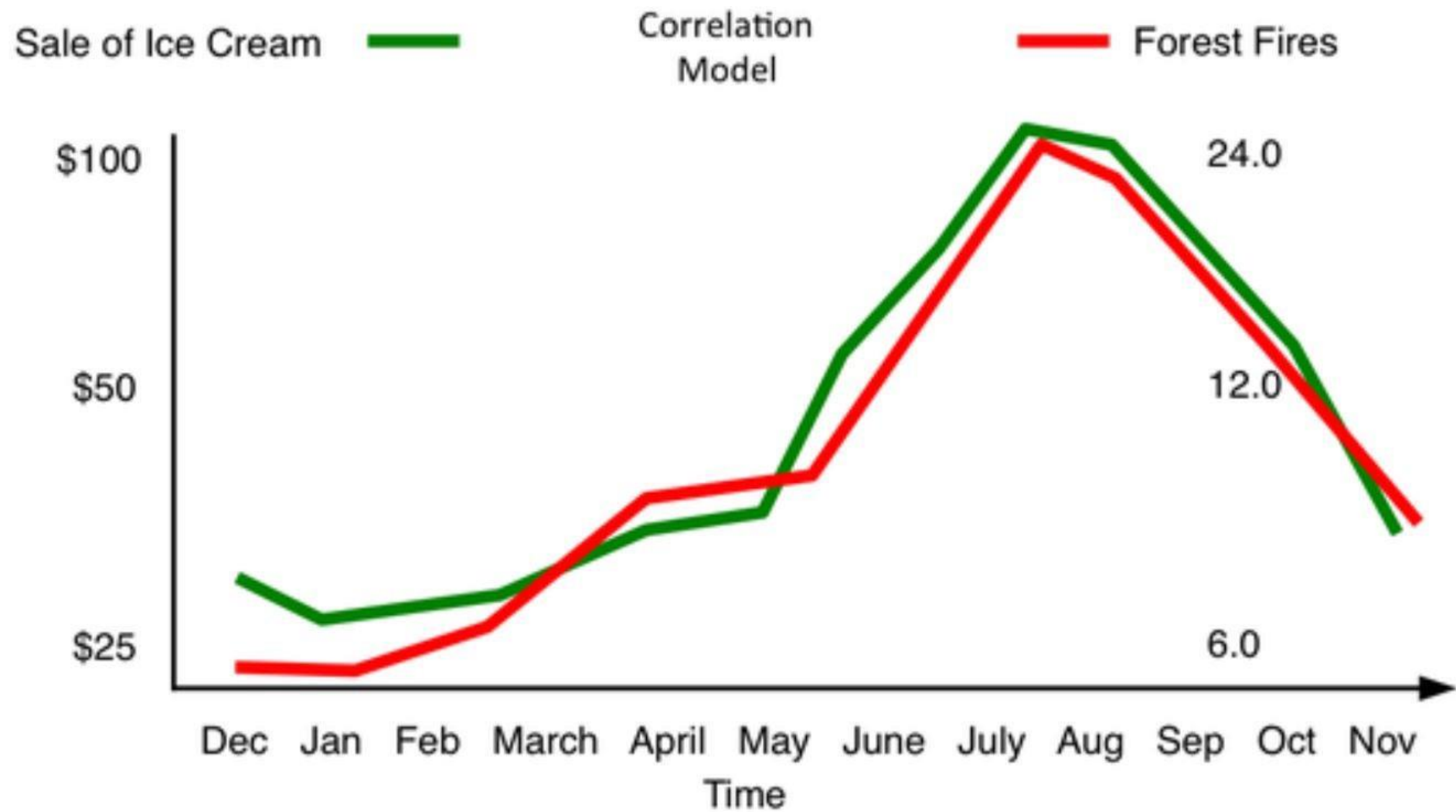
Vyhodnotenie výsledkov

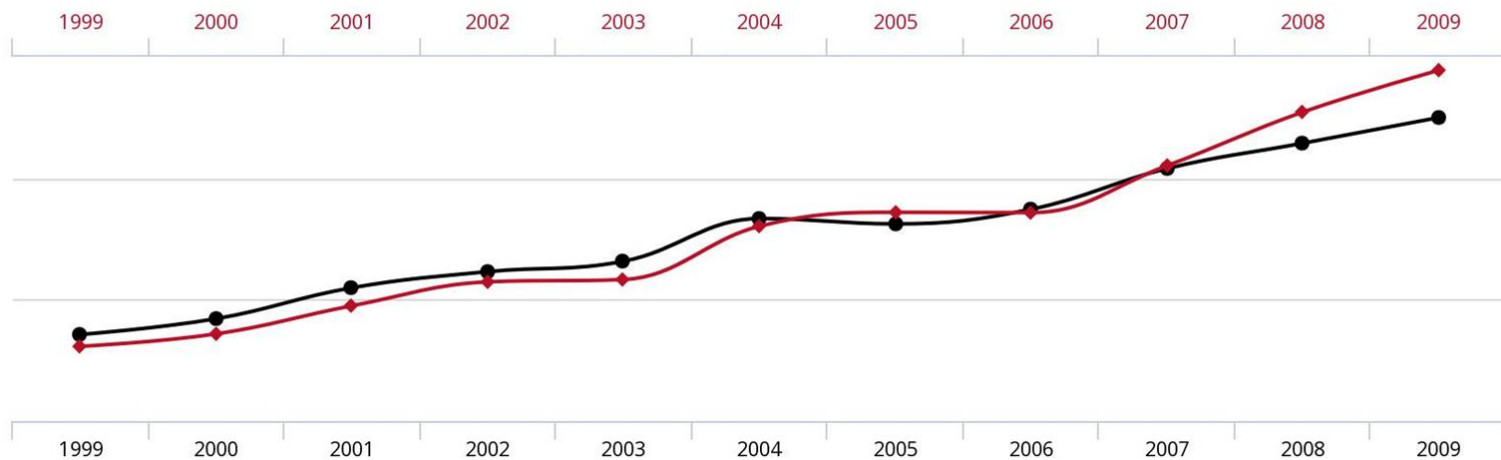
- Odkiaľ vieme, že výsledok simulácie zodpovedá realite?
- je potrebné otestovať implikácie výsledku na reálnom príklade (napr. aproximácia π)
- výstupom simulácie sú
 - údaje
 - modely
 - implikácie alebo následky

Vyhodnotenie výsledkov

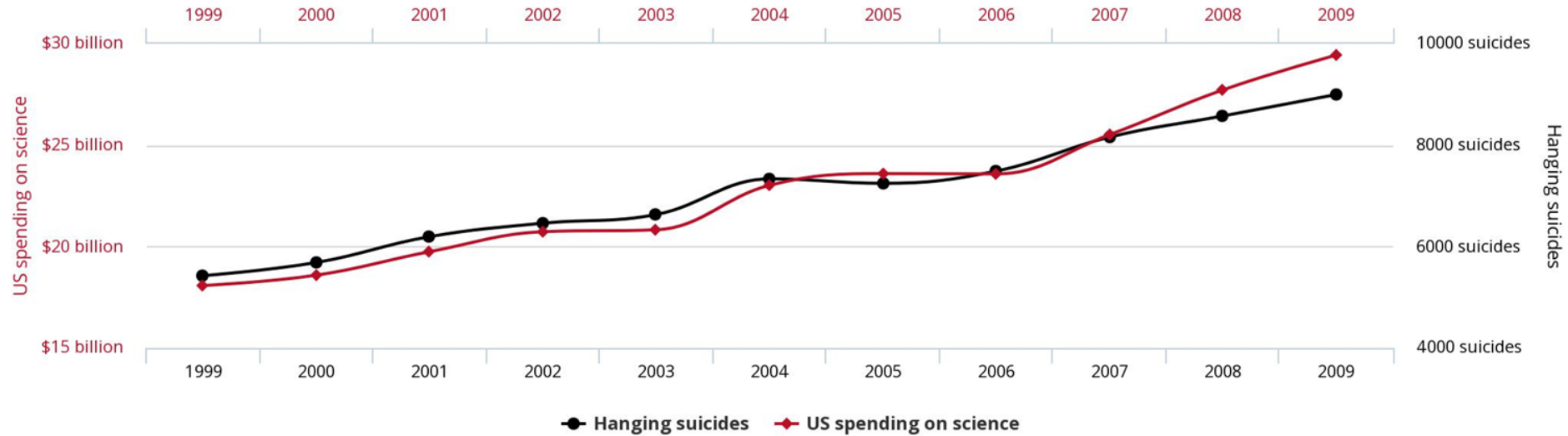
1. je potrebné analyzovať údaje, nielen ich vlastnosti!
napr.: priemer získaných bodov na teste
2. máme reprezentatívnu vzorku? AKA data enhancement
cieľom je otestovať, a nie potvrdiť hypotézu
3. paradox texaského strelca
neignorujeme časť údajov
4. uponáhľané závery
napr.: vyššia nehodovosť v okolí domu
5. korelácia \neq kauzalita







US spending on science, space, and technology correlates with Suicides by hanging, strangulation and suffocation





MAY, 2011

1/5/11: RAMSEY SCORES
2/5/11: OSAMA BIN LADEN DIES



October, 2011

2/10/11: RAMSEY SCORES
5/10/11: STEVE JOBS DIES
19/10/11: RAMSEY SCORES
20/10/11: MUAMMAR GADDAFI DIES



Febuary, 2012

11/2/12: RAMSEY SCORES
11/2/12: WHITNEY HOUSTON DIES



March, 2013

21/3/13: RAMSEY SCORES
22/3/13: (BASKETBALLER) RAY
WILLIAMS DIES
23/3/13: BORIS BEREZOVSKY DIES



November, 2013

30/11/13: RAMSEY SCORES
30/11/13: PAUL WALKER DIES



August, 2014

10/8/14: RAMSEY SCORES
11/8/14: ROBIN WILLIAMS DIES



January, 2016

9/1/16: RAMSEY SCORES
10/1/16: DAVID BOWIE DIES
13/1/16: RAMSEY SCORES
14/1/16: ALAN RICKMAN DIES



Simulácia vírusových ochorení

<https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-00-introduction-to-computer-science-and-programming-fall-2008/assignments/pset12.pdf>

Zhrnutie

- biased random walk
- štatistické rozdelenia
- vybrané metódy náhodnosti v Pythone
- vyhodnotenie výsledkov
- časté chyby pri interpretácií výsledkov