8. cvičení

Datové struktury I, 22. 11. 2022 `https://iuuk.mff.cuni.cz/~chmel/2223/ds1/`

Úloha 1 (Lehké opakování pravděpodobnosti)
Matt střílí, a snaží se trefit se do terče. Protože je začátečník, pravděpodobnost, že se trefí, je $p \in (0, 1]$, a je nezávislá na jakýchkoliv jeho předchozích pokusech. Nechť X je náhodná veličina, která označuje počet pokusů do Mattovy první trefy (včetně). Ukažte, že $\mathbb{E}[X] = \frac{1}{p}$.

Úloha 2 (Špatná verze kukačky)
Proč je následující implementace insertu pro kukačkové hashování problematická?

```
for i=1 to n
  if T[h1(x)] je prázdné
    T[h1(x)] = x
    return
  swap(T[h1(x)], x)
  if T[h2(x)] je prázdné
    T[h2(x)] = x
    return
  swap(T[h2(x)], x)
```

Úloha 3 (Pravděpodobnost kolize)
Ukažte, že v hashovací tabulce velikosti $m = n^2$ s n prvky dojde ke kolizi s pravděpodobností maximálně $\frac{1}{2}$, předpokládáme-li zcela náhodnou hashovací funkci. (Zkuste to dokázat bez použití přímo padnoucího pozorování z přednášky.)

Úloha 4 (Rehashujeme)
Jednoduchá implementace rehashe u kukačkového hashování je, že si všechny hodnoty vložíme do pomocného pole, a potom je po jednom insertujeme. Vymyslete implementaci rehashe, která pomocné pole nepotřebuje. (Pozor na to, že během rehashe můžeme znovu začít s rehashem.)

Bonusové úlohy

Úloha 5 (Pevné body permutací)
Mějme uniformně náhodnou permutaci na n prvcích. Určete střední hodnotu počtu pevných bodů této permutace.

Opáčko pravděpodobnosti

Definice (Pravděpodobnostní prostor)**.** Pravděpodobnostní prostor je trojice (Ω, Σ, P), taková, že Ω je tzv. nosná množina (všech jevů), Σ je množina přípustných jevů (podmnožin Ω), P je funkce $P : \Sigma \to [0, 1]$ a jsou splněny nějaké axiomy (omezující Σ na *σ-algebru* a P na *pravděpodobnost*).
Pravděpodobnostní prostor je *diskrétní*, pokud Ω je konečná nebo spočetná, $\Sigma = 2^\Omega$ a máme $p : \Omega \to [0, 1]$, že $P[A] = \sum_{\omega \in A} p(\omega)$ a $P[\Omega] = 1$.

Definice (Náhodná veličina, její střední hodnota)**.** Nechť $(\Omega, 2^\Omega, P)$ je diskrétní pravděpodobnostní prostor. Potom náhodná veličina je libovolná funkce $X : \Omega \to \mathbb{R}$ (každému jevu přiřazujeme nějaké číslo).
Střední hodnota této veličiny je definována jako $\mathbb{E}[X] = \sum_{\omega \in \Omega} P[\omega] \cdot X(\omega)$

Tvrzení (Union bound)**.** Pro jevy A_1, A_2 platí, že $P[A_1 \cup A_2] \leq P[A_1] + P[A_2]$.

Tvrzení (Linearita střední hodnoty)**.** Pro náhodné veličiny X, Y a $\alpha \in \mathbb{R}$ platí:

- $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Definice (Indikátor, nezávislost náhodných veličin)**.** Nechť A je jev v diskrétním pravděpodobnostním prostoru. Potom indikátor A je náhodná veličina I_A definovaná: $I_A(\omega) = 0 \Leftrightarrow \omega \notin A$, jinak $I_A(\omega) = 1$.
Náhodné veličiny X, Y na diskrétním pravděpodobnostním prostoru $(\Omega, 2^\Omega, P)$ jsou nezávislé, pokud $\forall \alpha, \beta \in \mathbb{R}$ jsou jevy $\{\omega \in \Omega : X(\omega) \leq \alpha\}, \{\omega \in \Omega : Y(\omega) \leq \beta\}$ nezávislé.

Věta (Markovova nerovnost)**.** Buď X nezáporná náhodná veličina. Pak $\forall \varepsilon > 0$ platí $P[X \geq \varepsilon] \leq \frac{\mathbb{E}[X]}{\varepsilon}$.
Ekvivalentně pro jakékoliv $d > 1$, $P[X \geq d \cdot \mathbb{E}[X]] \leq \frac{1}{d}$.

Věta (Chernoffova mez)**.** Mějme $X = \sum_{i=1}^{k} X_i$, kde X_i jsou nezávislé Bernoulliovské náhodné veličiny, $\mu = \mathbb{E}[X], c > 1$. Pak $P[X \geq c\mu] \leq \left(\frac{e^{c-1}}{c^c}\right)^\mu$.