Verjetnostne metode v računalništvu - zapiski s predavanj prof. Marca

Tomaž Poljanšek

študijsko leto 2023/24

Kazalo

Poglavje 1

Introduction

1.1 Probability

(Ω, F, P_r):

- $\emptyset \in F$,
- $A \in F \implies A^c \in F$,
- $A_1, A_2 \dots \in F \implies \cup_{i=1}^{\infty} A_i \in F$.

$P_r(A) \geq 0$,
$P_r\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_r(A_i)$ if A_i disjoint,
$P_r\left(\cup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P_r(A_i)$,
$\Omega = \{\omega_1, \omega_2 \dots\}$ - countable case.
$\begin{pmatrix} \omega_1 & \omega_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$

Primer.

```
Alg():
  while True:
    B = sample as random from {0,1}  # 1 with probability p
    if B = 1:
```

```
        return
```

$\Omega = \{1, 01, 001, 0001 \dots\}$
$\begin{pmatrix} 1 & 01 & 001 & 0001 & \dots \\ p & (1-p)p & (1-p)^2p & (1-p)^3p & \dots \end{pmatrix}.$

1.2 Random variables

$X : \Omega \to \mathbb{Z}$.
$E[X] = \sum_{c \in \mathbb{Z}} c \cdot P_r(X = c)$ expected value of X.
Properties:

- $E[f(X)] = \sum_{c \in \mathbb{Z}} f(c) \cdot P_r(X = c)$,
- $E[aX + bY] = aE[X] + bE[Y]$,
- $E[X \cdot Y] = E[X] \cdot E[Y]$ if X, Y indepentent,
- $P_r(X \geq a) \leq \frac{E[X]}{a}$ $\forall a > 0$ $X \geq 0$ Markov inequality.

Primer. (Continuing from before).
$X =$ number of trials before return.
$X : \Omega \to \mathbb{Z}$.
$X : 1 \to 1, 01 \to 2, 003 \to 3 \dots$
$\begin{pmatrix} 1 & 2 & 3 & 4 & \dots \\ p & (1-p)p & (1-p)^2p & (1-p)^3p & \dots \end{pmatrix}$ - geometric distribution.

Trditev 1.2.1. $E[X] = \frac{1}{p}$.

Dokaz 1.2.2. $X = \sum_{i=1}^{\infty} X_i$.
$X_i = \begin{cases} 1 \text{ if trial } i \text{ is executed} \\ 0 \text{ else} \end{cases}$

$$E[X] = E[\sum_{i=1}^{\infty} X_i] = \sum_{i=1}^{\infty} E[X_i] =$$
$$= \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{i=0}{\infty}(1-p)^i = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

$E[X] = \frac{1}{p}$.
$P_r(X \geq 100 \cdot \frac{1}{p}) \leq \frac{E[X]}{\frac{1}{p}} = \frac{1}{100}$.

Definicija 1.2.3. $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{i=1}^{\infty} \frac{1}{i}$.

Izrek 1.2.4. $H_n \leq 1 + \ln(n)$.

Dokaz 1.2.5.

$$H_n = 1 + \sum_{i=2}^{n} \frac{1}{i} \overset{\text{integral}}{\leq} 1 + \int_1^n \frac{dx}{x} = 1 + \ln(x)|_1^n = 1 + \ln(n)$$

Poglavje 2

Quicksort, min-cut

2.1 Quicksort

```
Input: set (no equal element) (unordered list) S∈ ℝ
    (or whatever you can compare linearly)
Output: ordered list
Code:
  def Quicksort(S):
    if |S|= 0 or 1:
      return S
    else:
      a = uniformly at random from S
      S⁻ = {b ∈ S | b < a}
      S⁺ = {b ∈ S | a < b}
      return Quicksort(S⁻), a, Quicksort(S⁺)
```

$C(n)$ - random variable, the number of comparisons in evaluation of Quicksort with $|S| = n$

Izrek 2.1.1. $E[C(n)] = O(N \log(n))$

Dokaz 2.1.2. $C(0) = C(1) = 0$

$$
\begin{aligned}
E[C(n)] = n-1+\sum_{i=1}^{n}\left(E[C(i-1)]+E[C(n-i)]\right)\cdot P_r(a \text{ is } i\text{-it element}) \leq \\
\leq n+\frac{2}{n}\sum_{i=1}^{n-1}E[C(i)]
\end{aligned}
$$

Induction:

$n = 1 : \checkmark$

$n - 1 \to n$:

$$
\begin{aligned}
E[C(n)] &\leq n+\frac{2}{n}\sum_{i=1}^{n}E[C(i)] \leq \\
&\leq n+\frac{2}{n}\sum_{i=1}^{n}5i\log i \leq \\
&\leq n+\frac{2}{n}\sum_{i=1}^{\lfloor\frac{n}{2}\rfloor}5i\log i+\frac{2}{n}\sum_{i=1+\lfloor\frac{n}{2}\rfloor}^{n-1}5i\log i \leq \\
&\leq n+\frac{2}{n}\sum_{i=1}^{\lfloor\frac{n}{2}\rfloor}5i\log\frac{n}{2}+\frac{2}{n}\sum_{i=1+\lfloor\frac{n}{2}\rfloor}^{n-1}5i\log n \leq \\
(\log\frac{n}{2} &= \log n - 1) \\
&\leq n+\frac{2}{n}\left(\sum_{i=1}^{n}5i\log n-\sum_{i=1}^{\frac{n}{2}}5i\right) = \\
&= n+\frac{10}{n}\left(\frac{n(n-1)}{2}\log n-\frac{\frac{n}{2}(\frac{n}{2}+1)}{2}\right) \leq \\
&\leq n+5(n-1)\log n - n < \\
&< 5n\log n
\end{aligned}
$$

$P\left(C(n)\geq b\cdot 5n\log n\right) \overset{\text{Markov}}{\leq} \frac{1}{b}$

Dokaz 2.1.3.

2:

Let $S_1, S_2 \dots S_n$ sorted eleents of S

Define random variable $X_{ij} = \begin{cases} 1 \text{ if } S_i \text{ and } S_j \text{ are compared} \\ 0 \text{ else} \end{cases}$

$C(n) = \sum_{1\leq i<j\leq n} E[X_{ij}]$
$E[X_{ij}] = P(S_i$ and X_j compared)
S_{ij} - the last set including S_i and S_j
$E[X_{ij}] = \frac{2}{|S_{ij}|} \leq \frac{2}{j-i+1}$
$|S_{ij}| \geq j-i+1$
S_{ij} has everything in between

$$
\begin{aligned}
\implies E[C(n)] &\leq \sum_{1\leq i<j\leq n} \frac{2}{j-i+1} = \\
&\overset{k=j-i+1}{=} \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \leq \\
&\leq 2 \cdot n \cdot H_n \leq \\
&\leq 2n(1+\log n)
\end{aligned}
$$

Literatura