Algorytmika
Ćwiczenia 6

May 28, 2025

Adrian Herda
Politechnika Wrocławska

1. Zadanie 27

Niech X będzie licznikiem morrisa. Licznik rozpoczyna z wartością 0 i dla każdej operacji inkremetowany jest z prawdopodobieństwem 2^{-X}. Estymator morrisa jest zdefiniowany jako:

$$\hat{n} = 2^{X_n} - 1$$
$$\hat{n} + 1 = 2^{X_n}$$

gdzie X_n jest licznikiem morrisa po n operacjach. Wartość oczekiwana tego estymatora jest równa:

$$E[\hat{n}] = n$$

Dowód przez indukcję:
1. Dla $n = 0$ mamy $E[\hat{n}] = 2^{X_0} - 1 = 2^0 - 1 = 0$.
2. Załóżmy, że dla dowolnego n mamy $E[\hat{n}] = n$.
3. Dla $n + 1$ mamy:

$$\begin{aligned}
E\left[\widehat{n+1}\right] &= E\left[2^{X_{n+1}} - 1\right] \\
&= \sum_{i=1}^{\infty} P(X_n = i) \cdot E\left[2^{X_{n+1}} \mid X_n = i\right] - 1 \\
&= \sum_{i=1}^{\infty} P(X_n = i) \cdot \left(\underbrace{\frac{1}{2^i} \cdot 2^{i+1}}_{\text{increment}} + \underbrace{\left(1 - \frac{1}{2^i}\right) \cdot 2^i}_{\text{no increment}} \right) - 1 \\
&= \sum_{i=1}^{\infty} P(X_n = i) \cdot (2^i + 1) - 1 \\
&= \sum_{i=1}^{\infty} P(X_n = i) 2^i + \sum_{i=1}^{\infty} P(X_n = i) - 1 \\
&= E\left[2^{X_n}\right] + 1 - 1 \\
&= E[\hat{n} + 1] = n + 1
\end{aligned}$$

■

1.1. Wariancja

$$\begin{aligned}
E\big[2^{2X_n}\big] &= \sum_{i=1}^{\infty} P(X_n = i) \cdot 2^{2i} \\
&= \sum_{i=1}^{\infty} \left(P(X_{n-1} = i-1) \cdot \frac{1}{2^{i-1}} + P(X_{n-1} = i) \cdot \left(1 - \frac{1}{2^{i-1}}\right)\right) \cdot 2^{2i} \\
&= \sum_{i=1}^{\infty} 2^{i+1} P(X_{n-1} = i-1) + \sum_{i=1}^{\infty} 2^{2i} P(X_{n-1} = i) - \sum_{i=1}^{\infty} 2^{i} P(X_{n-1} = i) \\
&= 4 \sum_{i=1}^{\infty} 2^{i-1} P(X_{n-1} = i-1) + E\big[2^{2X_{n-1}}\big] + E\big[2^{X_{n-1}}\big] \\
&= 4E\big[2^{X_{n-1}}\big] + E\big[2^{2X_{n-1}}\big] + E\big[2^{X_{n-1}}\big] \\
&= 3E\big[2^{X_{n-1}}\big] + E\big[2^{2X_{n-1}}\big] \\
&= 3n + E\big[2^{2X_{n-1}}\big]
\end{aligned}$$

Ta formuła jest rekurencyjna, więc możemy j rozwin i otrzymamy ciąg arytmetyczny:

$$E\big[2^{2X_n}\big] = 1 + \sum_{i=1}^{n} 3i = 1 + \frac{3}{2} n(n+1)$$

A więc:

$$\begin{aligned}
\operatorname{Var}(\hat{n}) &= E\Big[\big(2^{X_n} - 1\big)^2\Big] - E\big[2^{X_n} - 1\big]^2 \\
&= E\big[2^{2X_n} - 2 \cdot 2^{X_n} + 1\big] - n^2 \\
&= E\big[2^{2X_n}\big] + 1 - 2E\big[2^{X_n}\big] - n^2 \\
&= 1 + \frac{3}{2} n(n+1) + 1 - \big(2E\big[2^{X_n}\big] - 2\big) - n^2 - 2 \\
&= \frac{3}{2} n^2 + \frac{3}{2} n - \big(2E\big[2^{X_n} - 1\big]\big) - n^2 \\
&= \frac{3}{2} n^2 + \frac{3}{2} n - 2n - n^2 \\
&= \frac{1}{2} n^2 - \frac{1}{2} n \\
&= \frac{1}{2} n(n-1)
\end{aligned}$$

1.2. Błąd standardowy

Błąd standardowy estymatora morrisa jest równy pierwiastkowi z wariancji podzielonej przez liczbę operacji:

$$\operatorname{SE}[\hat{n}] = \sqrt{\operatorname{Var}\left(\frac{\hat{n}}{n}\right)} = \sqrt{\frac{1}{2} \frac{n(n-1)}{n^2}} = \sqrt{\frac{n-1}{2n}}$$

2. Zadanie 28

Niech $h(x)$ będzie funkcją hashującą, która dla każdego x zwrca wartość z tablicy haszującej o długości m.

$$P(\text{kolizja}) = P\Big(\exists_{i\neq j}\big(h(x_i) = h(x_j)\big)\Big)$$

2.1. Dowód wzoru

Załóżmy że wstawione zostało już n elementów. To znaczy że pozstało $m-n$ miejsc w tablicy haszującej. Prawdopodobieństwo kolizji przy wstawianiu następnego elementu $x_{\{n+1\}}$ jest równe:

$$P(\text{kolizja}) = \frac{n}{m}$$

Przy wstawainiu kolejnch n elementów prawdopodobieństwo kolizji liczone jesst w następujący sposób:

$$\left.\begin{array}{l} P(\text{brak kolizji}) = \prod_{i=0}^{n-1} \frac{m-i}{m} \approx \exp\big(-\frac{n(n-1)}{2m}\big) \\ \frac{1}{2} \leq P(\text{kolizja}) \Longrightarrow \frac{1}{2} \geq P(\text{brak kolizji}) \end{array}\right\} \Longrightarrow \frac{1}{2} \approx \exp\left(-\frac{n(n-1)}{2m}\right)$$

$$\begin{aligned} \ln(2) &\approx \frac{n(n-1)}{2m} \\ n(n-1) &\approx 2\ln(2)m \\ n^2 - n &\approx 2\ln(2)m \end{aligned}$$

Dla dostatecznie dużych n możemy przyjąć, że

$$n^2 - n \approx n^2$$

więc:

$$\begin{aligned} n^2 &\approx 2\ln(2)m \\ n &\approx \sqrt{2\ln(2)m} \end{aligned}$$

■

2.2. Wartość n dla $m = 2^{16}$

$$\begin{aligned} n &\approx \sqrt{2\ln(2)2^{16}} \\ &\approx \sqrt{2 * 0.693147 * 65536} \\ &\approx \sqrt{90852.18725} \\ &\approx 301.417 \end{aligned}$$

3. Zadanie 29

Niech $h_b(x)$ będzie funkcją skrótu zwracającą b-bitowe ciągi. Rozmiar przestrzeni obrazów funkcji $h_b(x)$ dla dowolnego b wynosi:

$$|\{y : h_b(x) = y\}| = 2^b = m$$

Z poprzedniego zadania mamy wzór przybliżający wartość n dla której prawdopodobieństwo kolizji wynosi conajmniej $\frac{1}{2}$:

$$n \approx \sqrt{2m\ln(2)} = \sqrt{2 \cdot 2^b \cdot \ln(2)} = 2^{\frac{b}{2}} \cdot \sqrt{2\ln(2)}$$

3.1. Dla $b = 64$

$$n \approx 2^{\frac{64}{2}} \cdot \sqrt{2\ln(2)} = 2^{32}\sqrt{2\ln(2)} \approx 5.05694 \cdot 10^9$$

3.2. Dla $b = 128$

$$n \approx 2^{\frac{128}{2}} \cdot \sqrt{2\ln(2)} = 2^{64}\sqrt{2\ln(2)} \approx 2.17194 \cdot 10^{19}$$

3.3. Dla $b = 256$

$$n \approx 2^{\frac{256}{2}} \cdot \sqrt{2\ln(2)} = 2^{128}\sqrt{2\ln(2)} \approx 4.00652 \cdot 10^{38}$$