Probabilistic Analysis and Randomized Algorithms

Departemen Informatika

Universitas Diponegoro

Probabilistic Analysis

- Selain membatasi analisis pada kasus terbaik atau kasus terburuk, juga dapat dianalisis berdasarkan distribusi probabilitas setiap kasus

- Secara implisit, kita telah menggunakan analisis probabilistic saat mengatakan bahwa “dengan mengasumsikan berdistribusi uniform, membutuhkan perbandingan n/2 untuk menemukan elemen di dalam sebuah list yang berjumlah n elemen”

- Catatan :
  - Distribusi uniform adalah distribusi yang mempunyai peluang yang sama untuk setiap kejadian

Hiring Problem (Masalah Perekrutan)

- Misalkan kita menggunakan Agen tenaga kerja untuk merekrut asisten kantor
  - Agen tenaga kerja akan mengirimkan satu kandidat setiap hari : interview dan putuskan
  - Cost untuk interview adalah c_i per kandidat (Biaya ke agensi)
  - Cost untuk mempekerjakan adalah c_h per kandidat (termasuk memecat asisten sebelumnya dan biaya ke agensinya)
  - $c_h > c_i$
  - Kita akan selalu mempekerjakan kandidat terbaik yang terlihat hingga saat itu.

Algoritma

```
Hire-Assistant(n)
best = 0                // fictional least qualified candidate
for i = 1 to n
   interview candidate i // paying cost c_i
   if candidate i is better than candidate best
      best = i
      hire candidate i    // paying cost c_h
```

What is the cost of this strategy?

- Jika kita interview *n* kandidat dan mempekerjakan m dari kandidat tersebut maka cost nya $\Theta(c_i n + c_h m)$
- Misalkan kita interview semua n dan c_i kecil, maka kita akan focus pada $c_h m$.
- $c_h m$ bervariasi di setiap prosesnya dan tergantung dari urutan interview
- Ini seperti paradigma secara umum:
  - Temukan maksimum atau minimum dalam sebuah barisan dengan memeriksa setiap elemen
  - Dan merubah pemenangnya sebanyak m kali

Best case dan Worst case

- **Best Case**
- Jika setiap kandidat lebih jelek dibandingkan semua yang datang sebelumnya, sehingga kita hanya mempekerjakan satu kandidat:
  $\Theta(c_i n + c_h) = \Theta(c_i n)$
- **Worst Case**
- Jika setiap kandidat lebih baik dari semua yang datang sebelumnya, sehingga kita mempekerjakan semua n (m=n) :
  $\Theta(c_i n + c_h n) = \Theta(c_h n)$ since $c_h > c_i$

- But this is pessimistic. What happens in the average case?

Probabilistic Analysis

- Kita harus mengetahui atau membuat asumsi tentang distribusi input
- Analisis dapat menentukan waktu komputasi untuk average case

- Untuk masalah Hiring:
  - Anggap kandidat datang dalam urutan acak
  - Analisis dapat dihitung menggunakan permutasi
  - Urutan kandidat mempunyai permutasi dari n kandidat
  - Bagaimana dengan permutasi yang menunjukkan banyaknya yang kita pekerjakan? Apa sekali? dua kali? …atau n-1 kali ? n kali?
  - Tergantung dari seberapa banyak kandidat yang datang sebelum kandidat terbaik.
  - Analisis ini lebih mudah dengan menggunakan indicator variabel

Randomized Algorithms

- Sebuah algoritma adalah **randomized** jika perilakunya ditentukan oleh sebagian nilai yang digenerate secara random.

- Contoh dalam masalah Hiring, dilakukan perubahan sbb:
  - Agen tenaga kerja akan mengirimkan ke kita list dari n kandidat dan kita akan memilih urutan interview
  - Pemilihan dilakukan secara random
  - Jadi kita mengambil kendali atas pertanyaan apakah masukan diurutkan secara acak:
    - Ya, urutan akan diacak, sehingga kasus rata-rata menjadi nilai yang diharapkan.

Probabilistic Analysis with Indicator Random Variables

- Definisi :
- **Random variable** (mis., *X*) adalah variable yang mengambil nilai sesuai dengan distribusi probabilitasnya

- **Expected value** dari random variable (mis, E[*X*]) is nilai rata-rata yang kita amati jika kita mengambil random variable secara berulang-ulang.

Indicator random variable

- Diberikan ruang sample S dan kejadian A dalam S maka didefinisikan indicator random variable adalah :

$$I\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs ,} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

Lemma 1

For an event A, let X_A = I{A}. Then the expected value **E[X_A] = Pr{A}** (the probability of event A).

Proof:
Let ¬A be the complement of A. Then

E[X_A]	= E[I{A}]	(by definition)
	= 1*Pr{A} + 0*Pr{¬A}	(definition of expected value)
	= Pr{A}.	

Contoh

- Misalkan sebuah koin memiliki dua sisi, “Gambar” dan “Angka”. Berapakah nilai harapan dari munculnya “Gambar” saat koin dilempar satu kali?

- Sample space S is {G, A}
- Pr{G} = Pr{A} = 1/2
- Definisikan indicator random variable X_G= I{G}, menghitung jumlah munculnya “Gambar” saat dilempar satu kali.
- Karena Pr{G} = 1/2, berdasarkan Lemma 1 maka E[X_G] = 1/2.

Contoh

- Berapa nilai harapan dari “Gambar” ketika koin dilempar sebanyak n kali?
- Misalkan X adalah random variable untuk jumlah “Gambar” dari n lemparan.
- Dapat dihitung $E[X] = \sum_{i=0,n} i \Pr\{X=i\}$

Nilai Harapan

- Untuk $i = 1, 2, \ldots n$ definisikan X_i = I {Hasil lemparan ke i dalam kejadian G}.
- Sehingga $X = \sum_{i=1,n} X_i$.
- Lemma 1 mengatakan bahwa $E[X_i] = Pr\{G\} = 1/2$ for $i = 1, 2, \ldots n$.
- Nilai Harapan dari "Gambar" adalah $E[X] = E[\sum_{i=1,n} X_i]$
- *Problem:* We don't have $\sum_{i=1,n} X_i$; we only have $E[X_1]$, $E[X_2]$, ... $E[X_n]$.

Lanjutan Nilai Harapan

- *Solusi:* **Linearity of expectation**: ***expectation of sum equals sum of expectations.***
- Oleh karena itu:

$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \sum_{i=1}^{n} 1/2 \\
&= n/2\,.
\end{aligned}
$$

Hiring Problem Revisited

- Asumsikan bahwa kandidat datang dalam urutan random (kita dapat lakukan dengan randomization jika diperlukan).
- Misalkan X adalah random variable untuk berapa kali kita memperkerjakan asisten kantor baru.
- Definisikan indicator random variables $X_1, X_2, \ldots X_n$ dimana X_i = I {candidate *i* is hired}.
- Dipunyai:
- $X = X_1 + X_2 + \ldots + X_n$ *(Jumlah total yang dipekerjakan adalah jumlah apakah masing-masing individu :hire (1) or not (0).)*
- Lemma 1 didapatkan bahwa E[X_i] = Pr{candidate *i* is hired}.

Lanjutan

- We need to compute Pr{candidate *i* is hired}:
- Candidate *i* is hired iff candidate *i* is better than candidates 1, 2, ..., *i*−1
- Assumption of random order of arrival means any of the first *i* candidates are equally likely to be the best one so far.
- Thus, Pr{candidate *i* is the best so far} = 1/i.
  (Intuitively, as you add more candidates each candidate is less and less likely to be better than all the ones prior. More precisely, candidate 2 has proability 1/2 of being the best of the first 2 candidates; candidate 3 has probability 1/3 of being the best of the first 3 candidates; etc.)

Lanjutan

- By Lemma 1, $E[X_i]$ = *1/i*, a fact that lets us compute E[X] (*notice that this follows the proof pattern outlined above!)*:

$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \sum_{i=1}^{n} 1/i
\end{aligned}
$$

Lanjutan

- The sum is a harmonic series. The n^{th} **harmonic number** is:

$$
\begin{aligned}
H_n &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} \\
&= \sum_{k=1}^{n} \frac{1}{k} \\
&= \ln n + O(1) .
\end{aligned}
$$

- Thus, the expected hiring cost is $\Theta(c_h \ln n)$, much better than worst case $\Theta(c_h n)$! (catatan : $\ln n = \Theta(\lg n)$)

Referensi

- http://www2.hawaii.edu/~suthers/courses/ics311f20/Notes/Topic-05.html

Ada materi yang belum jelas ?

Silakan dicoba untuk soal Latihan UTS tahun kemarin