Example : The Hiring Problem

- 회사에서 직원을 고용하는데, 직업 소개소에서 매일 한 명씩 지원자를 보냄
- 회사는 지원자를 면접하고 그 지원자가 **첫번째 지원자거나, 현재 고용되어 있는 직원보다 유능**하면 반드시 고용한다. 이 때 현재 고용되어 있는 직원이 있으면 해고한다
- 면접비용은 1인당 C_i
- 고용비용은 C_h (해고비용 포함)
- C_h > C_i
- 고용 알고리즘

```
HIRE-ASSISTANT(n)
best = 0          // candidate 0 is a least-qualified dummy candidate
for i = 1 to n
    interview candidate i
    if candidate i is better than candidate best
        best = i
        hire candidate i
```

  - Worst Case : $O(c_h n)$
    - 매번 다음 지원자가 이전 지원자보다 유능

확률적 분석

- 가능한 모든 입력 순서에 대한 평균값을 계산
  - 수행시간이라면, Average-Case Running Time 계산
  - 지표확률변수를 사용해 HIRE-ASSISTANT(n)의 비용을 계산

지표 확률 변수 : Indicator Random Variable

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

- 확률로부터 확률 변수의 기댓값을 쉽게 계산 가능
- 보조정리 5.1

Lemma
For an event A, let $X_A = \mathrm{I}\{A\}$. Then $\mathrm{E}[X_A] = \Pr\{A\}$.

Proof Letting $\overline{A}$ be the complement of A, we have

$$\begin{aligned} \mathrm{E}[X_A] &= \mathrm{E}[\mathrm{I}\{A\}] \\ &= 1 \cdot \Pr\{A\} + 0 \cdot \Pr\{\overline{A}\} \quad \text{(definition of expected value)} \\ &= \Pr\{A\} . \end{aligned}$$

사용 예 1

- 동전을 한 번 던져서 앞면이 나오는 횟수의 기댓값

$$\text{Sample Space } S = \{H,T\} \text{ with } Pr\{H\} = Pr\{T\} = ½$$

$$\begin{aligned} X_H &= \mathrm{I}\{H\} \\ &= \begin{cases} 1 & \text{if } H \text{ occurs}, \\ 0 & \text{if } T \text{ occurs}. \end{cases} \end{aligned}$$

$$\begin{aligned} \mathrm{E}[X_H] &= \mathrm{E}[\mathrm{I}\{H\}] \\ &= 1 \cdot \Pr\{H\} + 0 \cdot \Pr\{T\} \\ &= 1 \cdot (1/2) + 0 \cdot (1/2) \\ &= 1/2 . \end{aligned}$$

사용 예 2

- 동전을 n 번 던져서 앞면이 나오는 횟수 X의 기댓값

$$\text{Let } X_i = I\{i\text{ 번째 던졌을 때 앞면이 나오는 경우}\}$$

$$X = \sum_{i=1}^{n} X_i$$

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

확률적 분석

- 지표 확률 변수를 사용하여
  HIRE-ASSISTANT(n)의 비용 $O(c_h m)$ 을 계산해보면(입력의 순서는 random 분포)

$$\text{Let } \begin{aligned} X_i &= \mathrm{I}\{\text{candidate } i \text{ is hired}\} \\ &= \begin{cases} 1 & \text{if candidate } i \text{ is hired}, \\ 0 & \text{if candidate } i \text{ is not hired} \end{cases} \end{aligned}$$

and

$$X = X_1 + X_2 + \cdots + X_n .$$

$$m = \begin{aligned} \mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\ &= \sum_{i=1}^{n} \mathrm{E}[X_i] \\ &= \sum_{i=1}^{n} 1/i \\ &= \ln n + O(1) \end{aligned}$$

by eq.(A.7)

$\mathrm{E}[X_i] = \Pr\{\text{candidate } i \text{ is hired}\} = 1/i$ ←랜덤 분포의 i명의 후보자 중 i번째 후보자가 가장 우수한 사람일 확률

  - HIRE-ASSISTANT(n)의 비용(Average Case) $O(c_h m) = O(c_h \ln n)$
  - 입력의 분포에 따라 계산하여 **Average-Case Cost**를 계산
    - 각각의 입력에 대한 Cost가 정해져 있음
      ▶ Average-Case Cost란 이들의 평균을 구하는 것

랜덤화된 알고리즘

```
RANDOMIZED-HIRE-ASSISTANT(n)
randomly permute the list of candidates
best = 0            // candidate 0 is a least-qualified dummy candidate
for i = 1 to n
    interview candidate i
    if candidate i is better than candidate best
        best = i
        hire candidate i
```

Lemma 5.3

The expected hiring cost of the procedure RANDOMIZED-HIRE-ASSISTANT is $O(c_h \ln n)$.

- 입력의 분포와 상관없이 **Expected Cost**를 계산
    - 위 케이스의 결과가 같은 것은 확률적 분석 과정에서 입력 분포를 랜덤으로 가정했기 때문
    - 입력을 알아도 Cost가 정해져있지 않음 ▶ Expected Cost만 구할 수 있음