5. 확률적 분석과 랜덤화된 알고리즘

통계에 관한 기초 용어 설명

확률함수

어떤 사건 X가 일어날 확률을 Pr(X)라고 표현한다. 주사위를 던져서 1이 나오는 사건을 A, 짝수가 나오는 사건을 B라 하면

$$Pr(A) = \frac{1}{6},\ \ Pr(B) = \frac{1}{2}$$

라고 표현할 수 있다.

기댓값

각 사건이 벌어졌을 때의 이득과 그 사건이 벌어질 확률을 곱한 것을 전체 사건에 대해 합한 값이다. 예를 들어 500원 동전을 던져서 앞면이 나오면 500원을 얻는다고 해보자. 앞면이 나올 확률 X는 2분의 1, 즉 50%이다. 따라서

$$E[X] = 500 * \frac{1}{2} = 250$$

라고 표현할 수 있다.

5.2 지표 확률 변수

표본 공간 S와 사건 A가 주어졌을 때 사건 A에 관한 **지표 확률 변수 I{A}**는 다음과 같이 정의된다.

I{A} =   1 사건 A가 일어날 경우

0 사건 A가 일어나지 않은 경우

예를 들어, 앞뒤가 동일한 동전을 던졌을 때 앞면이 나오는 기대 횟수에 대해 생각해보자. 앞, 뒤 둘 중 하나일 것이다. 표본 공간 S = { H, T } 이고, 이에 대한 확률 변수 Y는 H와 T를 각각 1/2의 확률로 갖도록 정의할 수 있다.이런 경우 Y = H로 표현되는, 즉 동전의 앞면이 나올 경우에 대한 지표 확률 변수 X_H로 를 정의할 수 있다.이 변수는 앞면이 나오면 1, 그렇지 않으면 0이다.

X_H = 1{H} = 1 H가 일어날 경우

0 T가 일어날 경우(H가 안일어나면 뒷면이니까)

동전을 한번 던졌을 때 앞면이 나올 경우의 기댓값

은 지표 변수 X_h의 기댓값과 일치한다.

$$\begin{aligned} E[X_H] &= E[IH] \\ &= 1 \cdot PrH + 0 \cdot PrT \\ &= 1 \cdot (1/2) + 0 \cdot (1/2) \\ &= 1/2 \end{aligned}$$

따라서 동전을 한 번 던질 경우 앞면이 나올 확률은 1/2이다. 사건 A에 관계된 지표 확률 변수의 기댓값은 사건 A가 발생할 확률과 일치한다.

보조정리 5.1

표본 공간 S와 표본 공간 안의 사건 A가 주어졌을 때

$$X_A = IA\text{라하자. 그러면}$$
$$E[X_A] = PrA\text{가 성립한다.}$$

지표 확률 변수를 이용한 고용 문제 분석

HIRE-ASSISTANT

```
1 best = 0 // 0번은 가장 낮은 점수를 갖는 가상의 지원자다.
2 for i 1 to n
3     지원자 i를 면접한다
4     if 지원자 i가 지원자 best 보다 나은가?
5         best = i
6         지원자 i를 고용한다.
```

전 단원에서 공부한 새로운 직원을 뽑게 되는 횟수에 대한 기댓값을 계산해보자. 지원자들은 무작위 순서로 온다고 가정한다. X를 새로운 직원을 고용한 횟수에 대한 확률 변수라 하자.

$$E[X] = \sum_{x=1}^{n} x \cdot PrX = x$$

계산을 단순화하기 위해 지표 확률 변수를 대신 사용하자.

새 직원의 고용 횟수에 대한 변수 하나를 통해 E[X]를 계산하는 대신에 지표 확률 변수를 이용해 계산하기 위해 각 지원자의 고용 여부를 나타내는 n개의 변수를 정의한다. 다시 말해, X_i를 i번째 지원자가 고용되었는지에 대한 지표 확률 변수로 정의한다.

X_i = I{지원자 i가 고용됨}

= 1 지원자i가 고용될 경우

= 0 지원자i가 고용되지 않을 경우

그리고 다음 식이 성립한다.

X = X_1 + X_2 + X_3 + ... + X_n

보조정리 5.1에 의해 다음 식이 성립한다.

E[X_i] = Pr{지원자i가 고용됨}

HIRE-ASSISTANT의 5-6행을 수행할 때 확률을 계산해야 한다. 6행에서 지원자 i는 1부터 i-1까지 모든 지원자보다 우수할 때이다. 각 지원자는 무작위 순서로 오므로 앞의 i명의 지원자 각각이 가장 뛰어난 직원일 가능성은 같다. 따라서 i번째 지원자가 1번부터 i-1번까지의 모든 지원자들보다 뛰어날 확률은 1/i 이고 고용될 확률 역시 1/i이다. E[X]를 계산해보자.

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

n명의 지원자를 면접하지만 실제로 고용하는 인원은 평균적으로 ln(n)명 뿐이다. 정리하면 다음과 같다.

보조정리 5.2

지원자가 무작위 순서로 온다고 할 때, HIRE-ASSISTANT 알고리즘은 총

$$O(c_h ln(n))$$

의 평균 고용비용이 든다.

고용 비용의 기댓값은 최악의 고용 비용 O(c_h n)보다 상당히 향상된다.