풀이

STEP I 2001, 문제 14

확률과 통계 • 무작위 순서

> **STEP I 2001, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/01-s1-q14

이 문제는 *Beyond STEP*의 12.12 절 **언제 그만 찾아야 하는가**에서 단계별 안내와 함께 자세히 다룬다: https://beyondstep.keithmcnulty.org/ko/ch12/secretary.html

핵심 아이디어

면접 순서 전체를 생각하지 말고, **특정 지원자들이 어느 자리에 있는지**만 생각한다. 처음 n명 중 최고가 전체 최고가 되는 것은 정확히 최고의 지원자가 처음 n명의 면접 대상 안에 있을 때이다. 또한 대칭성(symmetry)에 의해 최고의 지원자는 N개의 자리 중 어느 자리에나 같은 확률로(equally likely) 있다. 소문항 (ii)는 여기에 경우 하나를 더 고려하거나, 깔끔하게 여사건(complement)을 이용하면 된다.

풀이

지원자들에게 순위대로 번호를 붙인다. 즉 지원자 1은 전체 최고, 지원자 2는 두 번째로 좋은 지원자이고, 나머지도 이런 식이다. $N!$가지 순서가 모두 같은 확률이므로, 모든 지원자는 주어진 어느 자리에나 같은 확률로 있다. 또한 처음 n번의 면접에서 본 지원자들의 집합은 크기가 n인 $\binom{N}{n}$개의 부분집합 중 어느 것이 될 확률도 같다. 풀이 전체에서 $1 \leq n \leq N$이다.

소문항 (i)

지원자 1이 처음 n명의 면접 대상 안에 있는 사건(event)을 A라 하자. A가 일어나면 지원자 1은 확실히 처음 n명 중 최고이다. 일어나지 않으면 처음 n명 중 최고는 지원자 1보다 못한 누군가이다. 따라서 구하는 사건은 정확히 A이다.

지원자 1은 N개의 자리 각각에 같은 확률로 있고, 그중 n개가 처음 n개의 자리이므로

$$\mathrm{P}(\text{처음}n\text{명 중 최고가 전체 최고}) = \frac{n}{N}.$$

소문항 (ii)

이제 두 번째로 좋은 지원자가 존재하도록 $N \geq 2$로 둔다. 처음 n명 중 최고가 전체 1위 또는 2위일 필요충분조건(if and only if)은 다음과 같다.

- 지원자 1이 처음 n명 안에 있다(사건 A). 또는
- 지원자 1은 처음 n명 안에 **없지만**, 지원자 2는 있다.

> **생각해 보기** 두 번째 경우에 처음 n명 중 최고가 지원자 2이고, 그보다 더 나은 누군가가 아니라는 것을 어떻게 알 수 있는가?

지원자 2보다 나은 지원자는 지원자 1뿐인데, 이 경우 지원자 1은 앞쪽에서 면접을 보지 않았다. 거꾸로 1과 2가 모두 처음 n명 안에 없으면, 처음 n명 중 최고는 순위가 3 이하이다. 따라서 위의 두 경우를 합친 것이 정확히 구하는 사건이며, 두 경우는 서로 배반(mutually exclusive)이다.

첫 번째 경우의 확률은 소문항 (i)에서 $\frac{n}{N}$이다. 두 번째 경우에는 $\mathrm{P}(A') = \frac{N-n}{N}$이고, 지원자 1이 마지막 $N-n$개의 자리 중 하나에 있다는 조건 아래에서 지원자 2는 남은 $N-1$개의 자리 중 어느 곳에나 같은 확률로 있으며, 그중 n개가 처음 n개의 자리이다. 따라서

$$\begin{aligned}\mathrm{P}(\text{처음}n\text{명 중 최고가 전체 1위 또는 2위}) &= \frac{n}{N} + \frac{N-n}{N} \cdot \frac{n}{N-1} \\ &= \frac{n\left[(N-1)+(N-n)\right]}{N(N-1)} \\ &= \frac{n(2N-n-1)}{N(N-1)}.\end{aligned}$$

다른 방법으로, 여사건을 이용할 수도 있다. 이 사건이 일어나지 않는 것은 정확히 앞쪽의 n명이 모두 순위 3부터 N까지인 $N-2$명 중에서 나올 때이며, 그 확률은

$$\frac{\binom{N-2}{n}}{\binom{N}{n}} = \frac{(N-n)(N-n-1)}{N(N-1)},$$

이다. 그리고 $N(N-1) - (N-n)(N-n-1) = 2Nn - n^2 - n = n(2N-n-1)$이므로 같은 답을 얻는다.

$N = 4, n = 2$인 경우

공식에 따르면 (i)의 답은 $\frac{2}{4} = \frac{1}{2}$이고, (ii)의 답은 $\frac{2 \cdot 5}{4 \cdot 3} = \frac{5}{6}$이다.

가능한 경우를 나열하기 위해, 어느 두 지원자가 먼저 면접을 보는지만 중요하다는 점에 주목한다. $\binom{4}{2} = 6$개의 쌍 각각은 24가지 순서 중 정확히 $2! \times 2! = 4$가지에서 나타나므로(처음 두 명의 순서를 정하고, 이어서 마지막 두 명의 순서를 정한다) 여섯 쌍은 모두 같은 확률이다.

먼저 면접 본 두 명	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$
이들 중 최고	1	1	1	2	2	3

처음 두 명 중 최고가 전체 최고인 경우는 6가지 중 3가지로 확률은 $\frac{1}{2}$이고, 전체 1위 또는 2위인 경우는 6가지 중 5가지로 확률은 $\frac{5}{6}$이다. 둘 다 공식과 일치한다.

검산

- $n = N$: 모든 지원자를 보므로 두 확률 모두 1이어야 한다. 실제로 $\frac{N}{N} = 1$이고 $\frac{N(N-1)}{N(N-1)} = 1$이다.
- $n = 1$: 첫 번째 지원자는 균등하게 무작위로 뽑힌 한 명이므로 (ii)의 답은 $\frac{2}{N}$이어야 한다. 공식은 $\frac{2N-2}{N(N-1)} = \frac{2}{N}$을 준다.
- $2N - n - 1 \geq N - 1$이므로 (ii)의 답은 (i)의 답 이상이다.
- $N \leq 7$과 모든 n에 대해 $N!$가지 순서를 모두 나열해 보면 두 공식과 정확히 일치한다(예를 들어 $N = 5, n = 2$이면 $\frac{2}{5}$와 $\frac{7}{10}$을 얻는다).

흔한 실수

- (ii)에서 "전체 2위"는 전체 최고가 처음 n명 안에 **없을** 때에만 해당한다는 점을 잊는 것. 단순히 $\frac{n}{N} + \frac{n}{N}$을 더하면 중복해서 세게 된다.
- 지원자 1의 자리에 대한 조건을 준(conditioning) 뒤에도 지원자 2에 대해 $\frac{n}{N}$을 쓰는 것. 지원자 1의 자리가 정해지면 남은 자리는 $N - 1$개뿐이다.

STEP 너머

이것은 유명한 **비서 문제**(*secretary problem*)의 상황이다. 각 지원자를 그 자리에서 채용하거나 거절해야 한다면, 최선의 전략은 처음 n명을 지켜본 다음 그들 모두보다 나은 첫 지원자를 채용하는 것이다. 소문항 (i)의 논증을 확장하면 최적의 n은 대략 N/e이고, 이때 성공 확률은 약 $1/e$임을 보일 수 있다. 여기서 사용한 대칭성 논증("각 지원자는 어느 자리에나 같은 확률로 있다")은 무작위 순열(random permutation) 연구의 출발점이다. Blitzstein과

Hwang의 *Introduction to Probability*(확률론 입문)는 이와 같은 대칭성 논증을 훌륭하게 다룬다.

Beyond STEP 풀이: https://beyondstep.keithmcnulty.org/ko/. 풀이 글은 CC0 1.0이다. 문제 자체의 저작권은 © Cambridge University Press & Assessment에 있으며 여기에 싣지 않는다. 영어 원문을 기계 번역하였으며 아직 검토되지 않았다.