풀이

STEP I 1996, 문제 13

확률과 통계 • 이산확률분포와 기댓값

> **STEP I 1996, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/96-s1-q13

이 문제는 *Beyond STEP*의 12.11 절 **우표, 컵, 그리고 e**에서 단계별 안내와 함께 자세히 다룬다:
https://beyondstep.keithmcnulty.org/ko/ch12/stamps-and-e.html

핵심 아이디어

게임은 처음으로 틀리게 추측하는 순간 끝나므로, $N = k$는 "컵이 $2, 3, \ldots, k-1$개일 때는 맞히고, 컵이 k개일 때 틀린다"는 뜻이다. 이 확률들을 곱하면 분모에 계승(factorial)이 나타나고, 합 속에 계승이 보이면 지수급수(exponential series) $e^x = \sum x^n/n!$가 곧 등장하리라는 강한 신호이다. 기댓값(expectation)을 구하는 요령은 k를 $k!$와 약분하여 각 합이 $\sum 1/n!$의 첨자를 평행이동한 꼴이 되도록 만드는 것이다.

풀이

N의 확률분포

컵 c개가 뒤섞여 있으면 짐은 사실상 무작위로 추측하는 셈이므로, 확률 $1/c$로 맞히고 확률 $1 - 1/c$로 틀린다. 매 회마다 새로 섞으므로 각 회는 서로 독립(independent)이다.

컵은 항상 적어도 두 개 있으므로 $N \geq 2$이다. $k \geq 2$일 때, 사건 $N = k$는 짐이 컵이 $2, 3, \ldots, k-1$개일 때 올바르게 추측하고 컵이 k개일 때 실패하는 경우에 정확히 일어난다. 독립성에 의하여

$$\begin{aligned} \mathrm{P}(N = k) &= \frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{k-1} \cdot \left(1 - \frac{1}{k}\right) = \frac{1}{(k-1)!} \cdot \frac{k-1}{k} \\ &= \frac{k-1}{k!}, \qquad k = 2, 3, 4, \ldots \end{aligned}$$

이다. ($k = 2$일 때는 올바른 추측들의 곱이 빈 곱이 되어, 마땅히 그래야 하듯 $\frac{1}{2}$을 얻는다.)

이와 동치인 깔끔한 꼴도 눈여겨볼 만하다. $N > k$는 짐이 컵이 $2, \ldots, k$개일 때 모두 맞혔다는 뜻이므로

$$\mathrm{P}(N > k) = \frac{1}{k!} \qquad (k \geq 1).$$

이다. 실제로 $\frac{1}{(k-1)!} - \frac{1}{k!} = \frac{k-1}{k!}$이므로 위의 식과 일치한다. $k \to \infty$일 때 이 값은 0으로 수렴하므로, 게임은 확률 1로 끝나고 확률들의 합은 1이다.

기댓값

k를 $k!$와 약분하면

$$\begin{aligned}\mathrm{E}(N) &= \sum_{k=2}^{\infty} k \cdot \frac{k-1}{k!} = \sum_{k=2}^{\infty} \frac{k(k-1)}{k!} = \sum_{k=2}^{\infty} \frac{1}{(k-2)!} \\ &= \sum_{j=0}^{\infty} \frac{1}{j!} = \mathrm{e},\end{aligned}$$

이다. 여기서 $j = k - 2$로 놓고 $x = 1$에서 $\mathrm{e}^x = \sum_{j \geq 0} x^j / j!$를 이용하였다. 이것이 구하려던 결과이다.

분산

> **생각해 보기** 여기서는 어느 이차 적률(second moment)을 합하기가 가장 쉬운가? $\mathrm{E}(N^2)$인가, $\mathrm{E}\big(N(N-1)\big)$인가?

어느 쪽이든 된다. $k^2 \,\mathrm{P}(N = k) = k \cdot \dfrac{k(k-1)}{k!}$에서 인수 $k(k-1)$이 이미 깔끔하게 약분되기 때문이다. 계산하면

$$\begin{aligned}\mathrm{E}(N^2) &= \sum_{k=2}^{\infty} \frac{k^2(k-1)}{k!} = \sum_{k=2}^{\infty} \frac{k}{(k-2)!} = \sum_{j=0}^{\infty} \frac{j+2}{j!} \\ &= \sum_{j=1}^{\infty} \frac{1}{(j-1)!} + 2\sum_{j=0}^{\infty} \frac{1}{j!} = \mathrm{e} + 2\mathrm{e} = 3\mathrm{e},\end{aligned}$$

이다. 가운데 단계에서 $\sum j/j!$의 $j = 0$인 항은 0이고, $j \geq 1$일 때 $j/j! = 1/(j-1)!$이다. 따라서 분산(variance)은

$$\mathrm{Var}(N) = \mathrm{E}(N^2) - \big(\mathrm{E}(N)\big)^2 = 3\mathrm{e} - \mathrm{e}^2 = \mathrm{e}(3 - \mathrm{e}) \approx 0.766.$$

이다.

검산

- 꼬리 꼴 $\mathrm{P}(N > k) = 1/k!$가 보여 주듯이, 확률 $\frac{1}{2}, \frac{1}{3}, \frac{1}{8}, \frac{1}{30}, \ldots$의 합은 1이다.
- 꼬리 공식을 이용하면 $\mathrm{E}(N) = \sum_{k\geq 0} \mathrm{P}(N > k) = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots = \mathrm{e}$이므로, 독립적인 다른 경로로 같은 답을 얻는다.
- $\mathrm{e} < 3$이므로 분산은 마땅히 그래야 하듯 양수이다. 정확한 합으로 계산하면 $\mathrm{E}(N) = 2.71828\ldots$, $\mathrm{Var}(N) = 0.76579\ldots$이고, 게임 400 000회를 모의실험한 결과 평균은 2.720, 분산은 0.767이었다.

흔한 실수

- 확률분포를 $k = 1$부터 시작하는 것: 처음부터 컵이 두 개 있으므로 $N \geq 2$이다.
- 실패한 추측에 해당하는 마지막 인수 $1 - \frac{1}{k}$를 빠뜨려 $1/(k-1)!$을 얻는 것. 이 "확률"들의 합은 1이 아니라 $\mathrm{e} - 1$이다.
- $\big(\mathrm{E}(N)\big)^2$을 빼는 대신 $\mathrm{Var}(N) = \mathrm{E}(N^2) - \mathrm{E}(N)$이라고 쓰는 것.

STEP 너머

꼬리합 공식(tail-sum formula) $\mathrm{E}(N) = \sum_{k\geq 0} \mathrm{P}(N > k)$는 $\{0, 1, 2, \ldots\}$에서 값을 갖는 모든 확률변수(random variable)에 대하여 성립하며, 대학 1학년에서 쓰는 표준적인 도구이다. 이 문제에서처럼 복잡한 기댓값 계산을 한 줄짜리 계산으로 바꾸어 주는 경우가 많다. 모든 적률을 더 깔끔하게 한데 묶는 방법은 확률생성함수(probability generating function) $G(s) = \mathrm{E}(s^N)$이다. 여기서는 $G(s) = 1 + (s-1)\mathrm{e}^s$이 되고, $G'(1) = \mathrm{e}$와 $G''(1) = 2\mathrm{e}$로부터 평균과 분산을 곧바로 얻는다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 바로 이러한 관점에서 생성함수를 전개한다.

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.