풀이

STEP I 2017, 문제 12

확률과 통계 • 기댓값과 조건부

> **STEP I 2017, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/17-s1-q12

핵심 아이디어

주최자가 거두는 수입은 언제나 Nc로 같다. 유일한 불확실성은 상금을 지급해야 하는지 여부뿐이다. 따라서 기대 이익(expected profit)은 수입에서 J와 **누군가가** 당첨 번호를 골랐을 확률의 곱을 뺀 값이고, 결국 모든 것은 **아무도** 당첨 번호를 고르지 않는 여사건(complementary event)으로 귀결된다. 이 확률은 거듭제곱 $(1-p)^N$ 꼴이며, 근사식(approximation) $(*)$는 바로 이런 식을 다루기 위해 마련된 것이다.

풀이

소문항 (i)

당첨 번호를 W라 하고, 어떤 참가자도 W를 고르지 않는 사건을 E라 하자. W가 무엇이 되든, 각 참가자는 서로 독립적으로(independently) 확률 $1-\frac{1}{N}$로 그 번호를 피하므로

$$\mathrm{P}(E) = \left(1-\frac{1}{N}\right)^N.$$

주최자는 어떤 경우에든 Nc를 받는다. E가 일어나면 당첨자가 없으므로 이익은 Nc이다. 그렇지 않으면 상금이 당첨자들 사이에 나뉘지만, 당첨자가 몇 명이든 지급되는 총액은 J이므로 이익은 $Nc-J$이다. 따라서

$$\text{기대 이익} = Nc - J\,\mathrm{P}(E^c) = Nc - J\left[1-\left(1-\frac{1}{N}\right)^N\right].$$

$a=1$로 두고 $(*)$를 쓰면 $(1-\frac{1}{N})^N \approx \mathrm{e}^{-1}$이므로, 기대 이익은 근사적으로 $Nc-J(1-$

e^{-1})이다. $J = 2Nc$이면 이는

$$Nc - 2Nc\left(1 - e^{-1}\right) = Nc\left(\frac{2}{e} - 1\right).$$

가 된다. $e > 2$이므로 $2/e < 1$이고, 따라서 이 값은 음수이다. 즉 주최자는 손실을 볼 것으로 예상된다.

소문항 (ii)

인기 있는 번호는 γN개, 인기 없는 번호는 $(1-\gamma)N$개이다. 각 참가자는 정확히 하나의 번호를 고르므로, N개의 번호 전체에 대한 확률의 합은 1이다.

$$\gamma N \cdot \frac{a}{N} + (1-\gamma)N \cdot \frac{b}{N} = 1 \quad \Longrightarrow \quad \gamma a + (1-\gamma)b = 1.$$

주최자는 여전히 W를 균등하게(uniformly) 고르므로, W는 확률 γ로 인기 있는 번호이고 확률 $1-\gamma$로 인기 없는 번호이다.

> **생각해 보기** 참가자들은 더 이상 균등하게 고르지 않는다. 이것이 $\mathrm{P}(E)$를 구하는 방법을 바꾸는가?

바뀌는 것은 W를 피할 확률뿐이며, 이 확률은 이제 W가 어떤 종류의 번호인지에 따라 달라진다. 그러므로 이에 대해 조건을 건다(conditioning). W가 인기 있는 번호이면 각 참가자는 독립적으로 확률 $1 - \frac{a}{N}$로 그 번호를 피하고, 인기 없는 번호이면 확률 $1 - \frac{b}{N}$로 피한다. 전확률 법칙(law of total probability)과 $(*)$에 의해

$$\mathrm{P}(E) = \gamma\left(1 - \frac{a}{N}\right)^N + (1-\gamma)\left(1 - \frac{b}{N}\right)^N \approx \gamma\, e^{-a} + (1-\gamma)\, e^{-b}.$$

소문항 (i)에서와 같이 기대 이익은 $Nc - J(1 - \mathrm{P}(E))$이므로

$$\text{기대 이익} \approx \gamma J\, e^{-a} + (1-\gamma)J\, e^{-b} + (Nc - J),$$

이고, 이는

$$A = \gamma J, \qquad B = (1-\gamma)J, \qquad C = Nc - J.$$

로 둔 요구된 꼴이다.

$\gamma = \frac{1}{8}$, $a = 9b$인 경우. 관계식으로부터 $\frac{1}{8} \cdot 9b + \frac{7}{8}b = 1$, 즉 $2b = 1$이므로

$$b = \tfrac{1}{2}, \qquad a = \tfrac{9}{2}.$$

$J = 2Nc$이면 $A = \frac{1}{4}Nc$, $B = \frac{7}{4}Nc$, $C = -Nc$이므로 기대 이익은 근사적으로

$$Nc\left(\tfrac{1}{4}e^{-9/2} + \tfrac{7}{4}e^{-1/2} - 1\right).$$

이다. 괄호 안의 첫째 항은 양수이므로 $\frac{7}{4}e^{-1/2} > 1$, 즉 $\sqrt{e} < \frac{7}{4}$임을 보이면 충분하다. $e < 3$이므로

$$\sqrt{e} < \sqrt{3} < \frac{7}{4},$$

이고, 마지막 단계는 양변이 모두 양수이며 $3 < \frac{49}{16}$(즉 $48 < 49$)이기 때문에 성립한다. 따라서 괄호 안은 양수이고, 주최자는 이익을 볼 것으로 예상된다. 상금은 소문항 (i)과 같지만, 당첨 번호가 인기 없는 번호인 경우가 많기 때문에 상금이 주인을 찾지 못하는 일이 더 잦다.

검산

- $J = 2Nc$일 때 소문항 (i)의 기대 이익은 약 $(2/e - 1)Nc \approx -0.264\,Nc$이고, 마지막 부분의 기대 이익은 $\left(\frac{1}{4}e^{-4.5} + \frac{7}{4}e^{-0.5} - 1\right)Nc \approx 0.064\,Nc$이다. 주장한 대로 작은 이익이다.
- $N = 800$일 때 정확한 거듭제곱 $(1 - a/N)^N$을 쓰면 $0.0640\,Nc$가 나오므로, 여기서 근사식 $(*)$는 매우 정확하다.
- $\gamma = 1$(모든 번호가 "인기 있는" 번호)이면 관계식에 의해 $a = 1$이 되고, 소문항 (ii)의 식은 소문항 (i)의 $Je^{-1} + Nc - J$로 바뀐다.
- $N = 4$일 때(참가자들이 고를 수 있는 4^4가지 경우 전부) 컴퓨터로 정확히 열거한 결과는 근사하지 않은 두 식과 모두 일치한다.

흔한 실수

- 상금이 어떻게 나뉘는지를 걱정하는 것. 당첨자가 한 명 이상이면 당첨자 수와 관계없이 주최자가 지급하는 총액은 J이다.
- 당첨 번호가 인기 있는 번호일 확률이 a나 b에 따라 달라진다고 가정하는 것. 주최자는 균등하게 고르므로, 그 확률은 γ이다.
- 각 종류의 번호가 몇 개인지 세는 것을 잊고 관계식을 $a + b = 1$이나 $\gamma a + b = 1$로 쓰는 것.

STEP 너머

당첨 번호를 고르는 참가자의 수는 모수가 N과 p인 이항분포(binomial distribution)를 따르며, $(*)$는 N이 크고 Np가 고정되어 있을 때 이 분포가 근사적으로 푸아송 분포(Poisson distribution)임을 이해하는 첫걸음이다. $(1 - p)^N \to e^{-Np}$는 바로 푸아송 분포에서 0이 나올 확률이다. 이월(roll-over)이 있는 실제 복권도 정확히 이런 방식으로 분석되며, 인기 없는 번호를 고르는 사람들은 실제로 더 높은 기대 수익을 얻는다. Grimmett and Welsh,

Probability: An Introduction(확률론 입문)은 푸아송 극한과 여기서 사용한 것과 같은 조건부 논증을 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.