풀이

STEP II 2001, 문제 12

확률과 통계 • 푸아송 근사

> **STEP II 2001, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/01-s2-q12

핵심 아이디어

공동 구매단(syndicate)은 당첨 복권을 정확히 한 장 갖는 것이 보장되므로, 모든 것은 다른 참가자들 가운데 몇 명이 당첨 번호를 골랐는가에 달려 있다. 이 개수는 시행 횟수가 매우 크고 성공 확률이 매우 작은 이항분포(binomial distribution)를 따르는데, 이것이 바로 푸아송 분포(Poisson distribution)가 대신 쓰이는 상황이다. 그러면 기대 배당액은 W에 $1/(1+X)$의 기댓값(expectation)을 곱한 것이 되고, 이 합은 $k+1$을 계승(factorial)에 흡수시켜 닫힌 형태로 계산할 수 있다.

풀이

설정하기

공동 구매단은 복권을 N장 샀으므로, 다른 참가자들은 복권을 $3.8N - N = 2.8N$장 샀다. 공동 구매단이 사지 않은 이 복권들 가운데 당첨 번호가 적힌 복권의 수를 X라 하자.

공동 구매단에 속하지 않은 각 참가자는 다른 참가자들과 추첨에 독립적으로 번호를 균등하게 무작위로 고르므로, 당첨 번호가 무엇이든 그들의 복권 한 장 한 장은 독립적으로 확률 $1/N$로 당첨 번호와 일치한다. 따라서

$$X \sim \mathrm{B}\left(2.8N,\ \tfrac{1}{N}\right).$$

N이 매우 크므로 $2.8N$은 크고 $1/N$은 작은 반면 평균 $2.8N \cdot \frac{1}{N} = 2.8$은 고정되어 있다. 따라서 X는 근사적으로 평균이 $\lambda = 2.8$인 푸아송 분포를 따른다.

$$\mathrm{P}(X = k) \approx \frac{\mathrm{e}^{-2.8}\,(2.8)^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

공동 구매단은 각 번호의 복권을 정확히 한 장씩 갖고 있으므로 언제나 당첨 복권을 정확히

한 장 가지며, 당첨 복권의 총수는 $T = 1 + X$이다.

가장 가능성 높은 당첨 복권의 수

$p_k = \mathrm{P}(X = k)$라 하자. 이웃한 확률을 비교하면

$$\frac{p_k}{p_{k-1}} = \frac{\lambda}{k} = \frac{2.8}{k},$$

이고, 이 값은 $k \leq 2$일 때 1보다 크고 $k \geq 3$일 때 1보다 작다. 그러므로 $p_0 < p_1 < p_2$이고 $p_2 > p_3 > p_4 > \cdots$이며, X의 가장 가능성 높은 값은 2이다. (정확한 이항분포에서도 마찬가지이다. 이 경우 $p_k / p_{k-1} = \frac{(2.8N-k+1)}{k(N-1)}$이고, 이 값은 정확히 $k < 2.8 + \frac{1}{N}$일 때 1보다 크다.)

따라서 가장 가능성 높은 당첨 복권의 수는

$$T = 1 + 2 = 3.$$

기대 순손실

당첨금은 $T = 1 + X$장의 당첨 복권에 똑같이 나뉘고 공동 구매단은 그중 한 장을 가지므로, 공동 구매단은 $\dfrac{W}{1+X}$를 받는다. 공동 구매단은 £N을 지불했으므로 순손실은

$$L = N - \frac{W}{1+X}, \qquad \text{그리고} \qquad \mathrm{E}(L) = N - W\,\mathrm{E}\left(\frac{1}{1+X}\right).$$

? **생각해 보기** $\sum \frac{1}{k+1} \cdot \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}$를 어떻게 더할까?

요령은 $(k+1) \cdot k! = (k+1)!$이므로 다루기 까다로운 인수 $\frac{1}{k+1}$이 계승에 합쳐진다는 것이다. λ를 곱하고 나누면 이 합은 한 항이 빠진 푸아송 급수가 된다.

$$\begin{aligned}\mathrm{E}\left(\frac{1}{1+X}\right) &\approx \sum_{k=0}^{\infty} \frac{1}{k+1} \cdot \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} = \frac{\mathrm{e}^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} \\ &= \frac{\mathrm{e}^{-\lambda}}{\lambda} \sum_{j=1}^{\infty} \frac{\lambda^j}{j!} = \frac{\mathrm{e}^{-\lambda}}{\lambda}\left(\mathrm{e}^{\lambda} - 1\right) = \frac{1 - \mathrm{e}^{-\lambda}}{\lambda},\end{aligned}$$

여기서 $j = k + 1$로 놓고 $\sum_{j \geq 0} \lambda^j / j! = \mathrm{e}^{\lambda}$를 이용하여, 값이 1인 $j = 0$ 항을 뺐다.

$\lambda = 2.8 = \frac{14}{5}$이면 이 값은 $\frac{5}{14}\left(1 - \mathrm{e}^{-2.8}\right)$이므로

$$\mathrm{E}(L) \approx N - \frac{5\left(1 - \mathrm{e}^{-2.8}\right)}{14} W,$$

이고, 이것이 구하려던 결과이다.

검산

- 같은 요령이 정확한 이항분포 $X \sim \mathrm{B}(n, p)$에도 통한다. $\mathrm{E}\left(\frac{1}{1+X}\right) = \dfrac{1-(1-p)^{n+1}}{(n+1)p}$이고, 이 값은 $np = \lambda$를 유지하며 $n \to \infty$일 때 $(1-\mathrm{e}^{-\lambda})/\lambda$로 수렴한다. $N = 1000$일 때 정확한 값은 0.33536이고, 푸아송 근삿값은 $\frac{5}{14}(1-\mathrm{e}^{-2.8}) \approx 0.33542$이다.
- $N = 200$으로 추첨을 10^5번 모의실험하였더니 평균 배당액은 $0.335W$였고, 당첨 복권이 3장인 경우가 가장 자주 나타났다.
- 극한의 경우: 아무도 참가하지 않았다면$(\lambda \to 0)$ $(1-\mathrm{e}^{-\lambda})/\lambda \to 1$이고, 마땅히 그래야 하듯 공동 구매단이 당첨금 전부를 가져간다.

흔한 실수

- 다른 참가자들의 복권 수로 $2.8N$ 대신 $3.8N$을 쓰는 것. 공동 구매단의 복권은 무작위가 아니며, 정확히 한 장의 당첨 복권을 보탠다.
- $\mathrm{E}\left(\frac{W}{1+X}\right) = \frac{W}{1+\mathrm{E}(X)} = \frac{W}{3.8}$로 쓰는 것. 기댓값은 비선형 함수를 그대로 통과하지 않는다. 실제로 $\frac{1}{3.8} \approx 0.263$은 0.335보다 눈에 띄게 작다. $(1/(1+x)$가 볼록(convex)하므로, 옌센 부등식(Jensen's inequality)에 따르면 이 값은 더 클 수 없다.)

STEP 너머

$\mathrm{B}(n, \lambda/n)$을 $\mathrm{Po}(\lambda)$로 근사하는 것을 "희귀 사건의 법칙(law of rare events)"이라 하며, 대학 1학년 과정에서는 확률들 또는 확률생성함수(probability generating function)가 수렴함을 보여 이를 증명한다. $\mathrm{E}\big(1/(1+X)\big)$의 계산은 사실 생성함수의 적분 $\int_0^1 G_X(s)\,\mathrm{d}s$이며, 이는 평균과 분산 계산을 넘어서는 생성함수의 멋진 활용이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)에서 두 아이디어를 모두 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.