풀이

STEP II 1995, 문제 13

확률과 통계 • 푸아송 분포와 최적화

STEP II 1995, 13번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/95-s2-q13

핵심 아이디어

데이터베이스 분류와는 달리, 이 문제는 사업 이야기로 포장된 확률 문제이다. 탑승하지 않는 승객의 수는 시행 횟수가 매우 크고 성공 확률이 매우 작은 이항분포(binomial distribution)를 따르는 개수이며, 바로 이런 경우에 푸아송 분포(Poisson distribution)가 등장한다. 그다음 정수 k에 대해 최적화하는 요령은 미적분이 아니라 **차분**(first difference) $u_{k+1} - u_k$이다. 표를 한 장 더 팔 때 기대 이익이 늘어나는 한 계속 판다. 이 차분이 감소함을 보이고 나면, 가장 좋은 k는 차분이 처음으로 양수가 아니게 되는 곳이다.

풀이

푸아송 모형

$N + k$명의 표 소지자 각각이 서로 독립적으로 확률 ϵ으로 탑승하지 않는다고 하자. 그러면 탑승하지 않는 사람의 수 R는 이항분포 $\mathrm{B}(N+k, \epsilon)$를 따르고, 평균은 $(N+k)\epsilon$이다. k는 N에 비해 매우 작으므로 이 평균은 근사적으로 $N\epsilon = \lambda$이다. 시행 횟수가 크고 성공 확률이 매우 작은 이항분포는 평균이 같은 푸아송 분포로 잘 근사된다. 고정된 r에 대하여 $n\epsilon = \lambda$를 고정한 채 $n \to \infty$이면

$$\binom{n}{r}\epsilon^r(1-\epsilon)^{n-r} = \frac{n(n-1)\cdots(n-r+1)}{n^r}\cdot\frac{\lambda^r}{r!}\left(1-\frac{\lambda}{n}\right)^{n-r} \;\to\; \frac{\lambda^r}{r!}\,\mathrm{e}^{-\lambda},$$

이다. 첫 번째 인수는 1로 수렴(converge)하고 $(1-\lambda/n)^n \to \mathrm{e}^{-\lambda}$이기 때문이다. 따라서 근사적으로 $R \sim \mathrm{Po}(\lambda)$이다. 이제부터 이를 정확한 것으로 보고

$$p_r = \mathrm{P}(R = r) = \frac{\lambda^r}{r!}\,\mathrm{e}^{-\lambda}.$$

이라 쓴다.

이익

항공사는 $N+k$장의 표를 장당 £A의 이익으로 팔므로, 무슨 일이 일어나든 $A(N+k)$를 얻는다. r명의 승객이 탑승하지 않으면 N개의 좌석에 $N+k-r$명이 도착한다.

- $r \geq k$이면 도착하는 사람은 많아야 N명이므로 모두 탑승한다. 보상받는 사람은 없고, $\max(0, k-r) = 0$이다.
- $r < k$이면 $k-r$명이 탑승하지 못하고, 각각 £B의 비용이 들며, $\max(0, k-r) = k-r$이다.

두 경우 모두 파운드 단위의 이익은 $A(N+k) - B\max(0, k-r)$이다.

기대 이익 $u_0, \ldots, u_3$

기댓값(expected value)을 취하면 벌금에 기여하는 것은 $r = 0, 1, \ldots, k-1$뿐이므로

$$u_k = A(N+k) - B\sum_{r=0}^{k-1}(k-r)\,p_r.$$

이다. $p_0 = \mathrm{e}^{-\lambda}$, $p_1 = \lambda\mathrm{e}^{-\lambda}$, $p_2 = \frac{1}{2}\lambda^2\mathrm{e}^{-\lambda}$를 이용하면

$$\begin{aligned}
u_0 &= AN,\\
u_1 &= A(N+1) - B\mathrm{e}^{-\lambda},\\
u_2 &= A(N+2) - B\mathrm{e}^{-\lambda}(2+\lambda),\\
u_3 &= A(N+3) - B\mathrm{e}^{-\lambda}\left(3+2\lambda+\tfrac{1}{2}\lambda^2\right).
\end{aligned}$$

차분 v_k

일반식끼리 빼면

$$\begin{aligned}
v_k = u_{k+1} - u_k &= A - B\left[\sum_{r=0}^{k}(k+1-r)\,p_r - \sum_{r=0}^{k-1}(k-r)\,p_r\right]\\
&= A - B\left[p_k + \sum_{r=0}^{k-1}\big((k+1-r) - (k-r)\big)p_r\right]\\
&= A - B\sum_{r=0}^{k} p_r,
\end{aligned}$$

이다. 여기서 첫 번째 합의 $r = k$ 항인 $(k+1-k)p_k = p_k$는 두 번째 합에 짝이 없다. 말로 하면 $v_k = A - B\,\mathrm{P}(R \leq k)$이다. 이는 자연스럽다. $(k+1)$번째 추가 표는 확실히 £A를 벌어들이고, 그 승객이 결국 탑승을 거부당할 때, 즉 $R \leq k$일 때 정확히 £B의 추가 비용이 든다.

? **생각해 보기** 왜 괄호 안이 확률들의 단순한 합으로 줄어드는가?

각 $r \le k-1$에 대하여 k가 하나 늘면 p_r의 계수는 정확히 하나 늘어나고(그 결과에서 탑승을 거부당하는 승객이 한 명 더 생긴다), 결과 $r = k$는 새로 한 번의 보상 비용을 발생시키기 시작한다. 따라서 0부터 k까지의 모든 r이 p_r을 한 번씩 기여한다.

그러면

$$v_k - v_{k+1} = B\,p_{k+1} = B\,\frac{\lambda^{k+1}}{(k+1)!}\,\mathrm{e}^{-\lambda} > 0,$$

이다. $B > 0$이고 $\lambda > 0$이기 때문이다. 따라서 모든 k에 대하여 $v_k > v_{k+1}$이다. 즉 차분은 순감소한다.

마지막으로 $\sum_{r=0}^{k} p_r$는 $\mathrm{e}^{-\lambda}\sum_{r\ge 0}\lambda^r/r! = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = 1$의 부분합(partial sum)이므로 $k \to \infty$일 때 1로 수렴하고,

$$v_k \to A - B \qquad \text{단, } k \to \infty,$$

이다. 이것이 보이려던 것이다. $B > A$이므로 이 극한은 음수이다.

Fly By Night에 대한 조언

$F(k) = \mathrm{P}(R \le k) = \sum_{r=0}^{k} p_r$라 쓰면 $v_k = A - BF(k)$이고

$$u_{k+1} > u_k \iff v_k > 0 \iff F(k) < \frac{A}{B}.$$

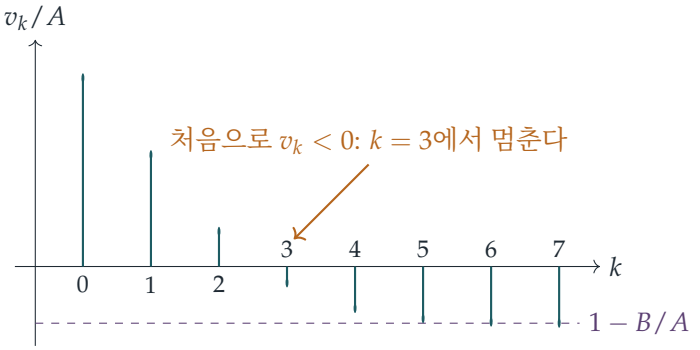
이다. v_k는 순감소하고 극한이 음수이므로, 수열(sequence) $v_0, v_1, v_2, \ldots$는 어느 지점까지는 양수이고 그 이후로는 음수이다(한 번 0이 될 수도 있다). 따라서 u_k는 증가하다가 감소하며, 최댓값은 다음 표가 더 이상 이익이 되지 않는 첫 번째 k에서 나온다.

> 다음을 만족하는 가장 작은k를 택한다: $F(k) = \displaystyle\sum_{r=0}^{k} \frac{\lambda^r}{r!}\mathrm{e}^{-\lambda} \;\ge\; \frac{A}{B}$.

특히 $\mathrm{e}^{-\lambda} \ge A/B$이면(승객이 매우 믿을 만하거나 벌금이 매우 크면) $v_0 \le 0$이므로 항공사는 초과 예약(overbooking)을 전혀 하지 말아야 한다. 즉 $k = 0$이다. 정확히 $F(k) = A/B$이면 $v_k = 0$이고, k와 $k+1$은 같은 기대 이익을 준다. $A/B < 1$이고 $F(k) \to 1$이므로 그러한 k는 항상 존재한다.

따라서 조언은 다음과 같다. 많아야 k명이 탑승하지 않을 확률이 표 한 장의 이익과 보상금의 비 A/B보다 작은 동안에는 추가 표를 계속 팔고, 그 확률이 처음으로 A/B에 도달하는 k에서 멈춘다.

예를 들어 $\lambda = 2$이고 $B = 1.25A$이면(따라서 $A/B = 0.8$) $F(0) \approx 0.135$, $F(1) \approx 0.406$, $F(2) \approx 0.677$, $F(3) \approx 0.857$이므로 가장 좋은 선택은 $k = 3$이다. 아래 그림은 이 경우의 v_k/A를 보여 준다.



검산

- 일반식으로부터 $u_1 - u_0 = A - Be^{-\lambda} = A - Bp_0$이고 $u_2 - u_1 = A - Be^{-\lambda}(1+\lambda) = A - B(p_0 + p_1)$이므로 v_0, v_1과 일치한다.
- 극한의 경우: $B \to A$이면 $A/B \to 1$이고 k는 한없이 커진다(탑승 거부의 비용이 거의 없다). $B \to \infty$이면 $A/B \to 0$이고 $k = 0$이다.
- 수치적으로, N, A, B, λ를 무작위로 수백 번 골라 이익 공식에서 직접 u_k를 계산하여, $u_0, \ldots, u_3$와 v_k의 공식, $v_k - v_{k+1} = Bp_{k+1}$, 그리고 u_k를 최대로 하는 k가 $F(k) \geq A/B$를 만족하는 가장 작은 k라는 것을 확인하였다. $N = 400, k = 3, \epsilon = 0.01$일 때 탑승하지 않는 사람이 $r = 3$명일 이항분포 확률과 푸아송 확률은 각각 0.194와 0.195이다.

흔한 실수

- 벌금 합에 $r \geq k$인 항을 포함하는 것. 이때는 도착한 사람 모두가 좌석을 얻으므로 이 항들은 아무것도 기여하지 않는다.
- u_k를 k에 대하여 미분하는 것. 변수는 정수이므로 차분 v_k가 올바른 도구이다.
- 답을 "$v_k = 0$"이라고 쓰는 것. 일반적으로 v_k를 정확히 0으로 만드는 k는 없다. 답은 $v_k \leq 0$이 되는 첫 번째 k이다.

STEP 너머

"차분이 감소하므로 부호가 바뀌는 곳에서 멈춘다"는 논증은 오목함수(concave function)가 도함수가 0이 되는 곳에서 최댓값을 갖는다는 사실의 이산 버전이다. "$\mathrm{P}(R \leq k)$가 A/B에 도달하면 멈춘다"는 규칙은 경영과학(operations research)의 고전적인 **신문 판매원**(newsvendor) 규칙(임계 분위수(critical fractile) 규칙)이며, 항공사의 초과 예약은 그

대표적인 응용이다. 위에서 엄밀하게 보인 이항분포의 푸아송 극한은 대학 확률론 강의에서 가장 먼저 배우는 극한 정리(limit theorem) 중 하나이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 푸아송 근사와 $\max(0, k - R)$ 같은 함수의 기댓값을 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.