풀이

STEP III 2006, 문제 12

확률과 통계 • 정규근사와 기댓값

STEP III 2006, 12번 문제

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

step.maths.org/questions/06-s3-q12

핵심 아이디어

추가 버스의 가치는 평균적으로 벌어들이는 금액만큼일 뿐이므로, 회사가 고려할 만한 면허 수수료는 연간 추가 이익의 **기댓값**(expectation)을 넘지 않는다. 기존 버스들만으로 충분히 감당되는 날에는 추가 버스가 아무것도 벌지 못하고, 또 좌석 32개보다 많은 승객을 태울 수도 없다. 따라서 "잘린(clipped)" 확률변수의 기댓값을 구해야 한다. 수치는 정규근사(normal approximation)를 했을 때 기존 수용 인원과 새 수용 인원이 분포에서 아주 깔끔한 점에 오도록 정해져 있다.

풀이

모형 세우기

관광객이 한 번 도착할 때, 1024명의 관광객 중 버스 정류장으로 안내되는 사람의 수를 X라 하자. 각 관광객은 확률 $\frac{1}{2}$로 그곳에 가며, 관광객들은 서로 독립(independent)적으로 안내된다고 가정한다. 그러면

$$X \sim \mathrm{B}(1024, \tfrac{1}{2}), \qquad \mathrm{E}(X) = 512, \qquad \mathrm{Var}(X) = 1024 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 256.$$

n이 크고 $p = \frac{1}{2}$이므로 정규근사 $X \approx \mathrm{N}(512, 16^2)$를 사용하고, $Z \sim \mathrm{N}(0,1)$로 두어 $X = 512 + 16Z$로 쓴다. 문제의 지시대로 연속성 수정(continuity correction)은 하지 않는다.

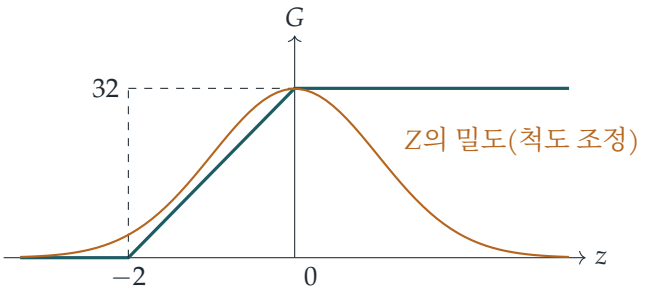
버스 15대는 $15 \times 32 = 480$명의 승객을 태우므로, 한 번 도착할 때의 현재 이익은 $£\min(X, 480)$이다. 16번째 버스가 있으면 수용 인원은 512이고 이익은 $£\min(X, 512)$이다.

따라서 한 번 도착할 때의 추가 이익은

$$G = \min(X, 512) - \min(X, 480) = \begin{cases} 0, & X \le 480, \\ X - 480, & 480 \le X \le 512, \\ 32, & X \ge 512. \end{cases}$$

Z로 나타내면 두 꺾이는 점은 $X = 480 \iff Z = -2$와 $X = 512 \iff Z = 0$이고, 가운데 구간에서는 $X - 480 = 32 + 16Z$이다. 그러므로

$$G = \begin{cases} 0, & Z \le -2, \\ 32 + 16Z, & -2 \le Z \le 0, \\ 32, & Z \ge 0. \end{cases}$$



추가 이익의 기댓값

$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}$를 표준정규분포의 확률밀도함수(probability density function)라 하자. 기댓값을 세 구간으로 나누면

$$\mathrm{E}(G) = 32\,\mathrm{P}(Z \ge 0) + \int_{-2}^{0} (32 + 16z)\,\phi(z)\,\mathrm{d}z.$$

첫째 항은 $32 \times \frac{1}{2} = 16$이다. 적분에서 상수 부분은

$$32 \int_{-2}^{0} \phi(z)\,\mathrm{d}z = 32(\Phi(0) - \Phi(-2)) = 32(\tfrac{1}{2} - (1 - \Phi(2))) = 32\Phi(2) - 16,$$

이 되는데, 여기서 대칭성 $\Phi(-2) = 1 - \Phi(2)$를 사용하였다.

생각해 보기 표 없이 $z\,\phi(z)$를 어떻게 적분할까?

밀도함수를 미분하면 $\phi'(z) = -z\,\phi(z)$이다. 따라서 $z\,\phi(z)$는 정확한 원시함수(antiderivative) $-\phi(z)$를 가지며,

$$16 \int_{-2}^{0} z\,\phi(z)\,\mathrm{d}z = 16[-\phi(z)]_{-2}^{0} = -16(\phi(0) - \phi(2)) = -\frac{16}{\sqrt{2\pi}}(1 - \mathrm{e}^{-2}).$$

(구간 전체에서 $z < 0$이므로, 이 값이 음수인 것은 당연하다.)

세 부분을 더하면 16들이 상쇄된다.

$$\mathrm{E}(G) = 16 + (32\Phi(2) - 16) - \frac{16}{\sqrt{2\pi}}(1 - \mathrm{e}^{-2}) = 32\Phi(2) - \frac{16}{\sqrt{2\pi}}(1 - \mathrm{e}^{-2}).$$

면허 수수료

1년에 50번 도착하므로, 기댓값의 선형성(linearity of expectation)에 의해 16번째 버스로 얻는 연간 추가 이익의 기댓값은

$$50\,\mathrm{E}(G) = 1600\,\Phi(2) - \frac{800}{\sqrt{2\pi}}(1 - \mathrm{e}^{-2}) \text{ 파운드.}$$

이보다 큰 면허 수수료를 내면 회사는 평균적으로 손해를 보고, 이보다 작은 수수료라면 기대 이익이 남는다. 따라서 이 값이 (정규근사 때문에 근사적으로) 고려할 만한 가장 큰 연간 수수료이며, 이것이 요구된 결과이다. 수치로는 약 £1288이다.

검산

- 범위 확인: 추가 버스는 한 번 운행할 때 0에서 32 사이를 벌므로, 연간 금액은 0과 1600 사이에 있어야 한다. 우리가 구한 값 $1600 \times 0.97725 - 319.15 \times 0.86466 \approx 1287.6$은 이 범위에 있다.
- 정확한 계산: 정확한 $\mathrm{B}(1024, \frac{1}{2})$ 확률로 $\min(k, 512) - \min(k, 480)$을 합하고 50을 곱하면 1287.69가 되어, 정규근사 값 1287.64와 6펜스 이내로 일치한다.

흔한 실수

- 새 버스가 필요할 때마다 항상 가득 찬다고 여겨 추가 이익을 $32\,\mathrm{P}(X > 480)$로 잡는 것. 버스 승객이 480명에서 512명 사이인 날에는 새 버스가 일부만 찬다.
- $\int_{-2}^{0} z\phi(z)\,\mathrm{d}z$에서의 부호 실수: 피적분함수가 음수이므로 답도 음수여야 한다.

STEP 너머

$\mathrm{E}\big(\min(X, c)\big)$라는 양은 재고가 c인 신문 판매상의 기대 판매량이며, c를 고르는 "신문팔이 문제(newsvendor problem)"는 경영과학(operations research)의 고전적인 문제이다. 같은 형태의 잘린 수익 $\min(\max(X - a, 0), b - a)$는 금융수학에서 "콜 스프레드(call spread)"라 불린다. $\phi'(z) = -z\phi(z)$라는 기법은 절단 정규분포의 적률(truncated normal moments) 이론의 출발점이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론

입문)은 이항분포의 정규근사와 확률변수의 함수의 기댓값을 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.