풀이

STEP I 1988, 문제 14

확률과 통계 • 중도절단 정규분포의 기댓값

> **STEP I 1988, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/88-s1-q14

핵심 아이디어

확률변수 X_M은 그냥 $\min(X, M)$이다. 즉 정규분포를 따르는 확률변수에서 M보다 큰 값을 모두 M으로 눌러 놓은 것이다. 그 기댓값(expectation)은 $x < M$ 범위에서의 적분과, 상수에 꼬리 확률(tail probability)을 곱한 항으로 나뉜다. 이때 $x\phi(x)$는 부호를 제외하면 $\phi(x)$의 도함수이므로 적분은 쉽다. 관광객 이야기는 같은 아이디어가 모습만 바꾼 것이다. 버스 차량단이 태우는 승객 수는 min(수요, 수용 인원)이고, 정규근사(normal approximation)를 하면 이것은 정확히 X_M 꼴이 된다.

풀이

X_M의 기댓값

$\{X < M\}$에서는 $X_M = X$이고 $\{X \geq M\}$에서는 $X_M = M$이므로

$$\mathrm{E}(X_M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M\,\mathrm{P}(X \geq M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M(1-\Phi(M)).$$

> **생각해 보기** 도함수가 $x\phi(x)$인 함수가 있을까?

있다. $\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$이므로 연쇄법칙(chain rule)에 의해 $\phi'(x) = -x\phi(x)$이다. 따라서

$$\int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x = \Big[-\phi(x)\Big]_{-\infty}^{M} = -\phi(M) + 0,$$

이다. $x \to -\infty$일 때 $\phi(x) \to 0$이기 때문이다. 그러므로

$$\mathrm{E}(X_M) = -\phi(M) + M(1-\Phi(M)),$$

이고, 이것이 증명하려던 식이다.

간단히 확인해 보자. $M \to \infty$이면 $M(1-\Phi(M)) \to 0$이므로 $\mathrm{E}(X_M) \to 0 = \mathrm{E}(X)$이다. $M \to -\infty$이면 $\mathrm{E}(X_M) \approx M$이 되는데, 당연히 그래야 한다.

버스

한 번의 방문에서 버스 정류장으로 보내지는 관광객 수를 N이라 하자. 1024명의 관광객이 각각 독립적으로, 양쪽 방향에 각각 확률 $\frac{1}{2}$로 안내된다고 가정한다. 그러면

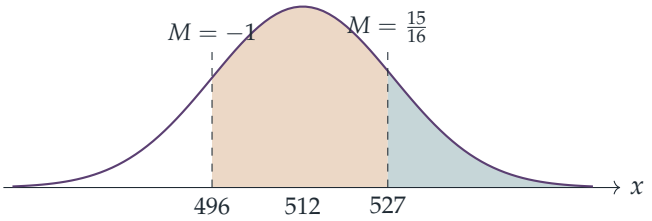
$$N \sim \mathrm{B}(1024, \tfrac{1}{2}), \qquad \mathrm{E}(N) = 512, \qquad \mathrm{Var}(N) = 1024 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 256 = 16^2.$$

이다. $n = 1024$처럼 충분히 크면 N은 X가 표준정규분포를 따를 때 $512 + 16X$로 매우 잘 근사된다.

버스가 16대이면 수용 인원은 $16 \times 31 = 496$이고, 17대이면 527이다. 수용 인원이 C인 차량 단은 승객 $\min(N, C)$명을 태우므로, 한 번의 방문에서 얻는 이익은 블로스카 단위로

$$\begin{aligned}\min(N, C) &\approx \min(512 + 16X,\ C) \\ &= 512 + 16\min\Big(X, \frac{C-512}{16}\Big) = 512 + 16\,X_M, \qquad \text{여기서} M = \frac{C-512}{16}.\end{aligned}$$

이다. 따라서 $C = 496$이면 $M = -1$이고, $C = 527$이면 $M = \frac{15}{16}$이다.



수요가 호박색 구간에 떨어지면 17번째 버스는 496을 넘는 초과분을 태운다. 수요가 청록색 꼬리에 떨어지면 17번째 버스는 꽉 찬 31명을 태운다.

$\phi(1) \approx 0.24197$, $\Phi(1) \approx 0.84134$를 쓰고 $1 - \Phi(-1) = \Phi(1)$임에 주목하면

$$\mathrm{E}(X_{-1}) = -\phi(1) - \big(1 - \Phi(-1)\big) = -\phi(1) - \Phi(1) \approx -0.24197 - 0.84134 = -1.08332.$$

이다. $\phi(0.9375) \approx 0.25707$, $1 - \Phi(0.9375) \approx 0.17425$를 쓰면

$$\mathrm{E}(X_{15/16}) \approx -0.25707 + 0.9375 \times 0.17425 \approx -0.25707 + 0.16336 = -0.09371.$$

이다. 따라서 버스를 한 대 더 두면 방문당 기대 이익이

$$16\big(\mathrm{E}(X_{15/16}) - \mathrm{E}(X_{-1})\big) \approx 16 \times 0.98961 \approx 15.83 \text{ 블로스카}.$$

만큼 늘어난다. 방문은 한 해에 50번이므로 추가 버스의 가치는 약

$$50 \times 15.83 \approx 792 \text{ 블로스카(연간)}.$$

이다. 이것이 고려할 만한 연간 뇌물의 최댓값이다. 기대 추가 이익보다 많이 지불하면 협동조합은 평균적으로 손해를 보게 된다.

> **생각해 보기** N이 이산적이니 연속성 보정(continuity correction)을 해야 하지 않을까?

여기서는 필요 없다. 연속성 보정은 $\mathrm{P}(N \le c)$처럼 평균을 내는 함수가 정수에서 불연속적으로 뛰는 확률을 근사할 때 도움이 된다. 함수 $k \mapsto \min(k, C)$에는 그런 도약이 없고, 보정하지 않은 정규근사가 이 함수에 대해 매우 정확하다. 이항분포(binomial distribution)로 정확히 계산하면 791.68이 나오는데, 이는 위의 값과 소수점 아래 두 자리까지 일치한다. 수용 인원을 $\frac{1}{2}$만큼 옮기면 오히려 결과가 나빠져서 775가 나온다.

검산

- $\int_{-\infty}^{M} x\phi(x)\,\mathrm{d}x + M(1-\Phi(M))$를 중점 구적법(midpoint quadrature)으로 계산하면 여러 M값에 대해 공식과 소수점 아래 여덟 자리까지 일치한다. 표본 4×10^5개로 $\min(X, M)$을 몬테카를로 시뮬레이션(Monte Carlo simulation)한 결과도 일치한다.
- $\mathrm{B}(1024, \frac{1}{2})$에 대해 정확히 합을 구하면 방문당 기대 추가 이익은 15.8337, 즉 한 해에 791.68이 되어 정규근사와 들어맞는다.
- 대략적인 타당성 확인: 추가 버스는 $N > 496$일 때마다 무언가를 벌어들이는데, 이는 확률 약 0.84로 일어나며 버스가 꽉 차는 경우도 많다. 따라서 방문당 31의 절반을 조금 넘는 값은 그럴듯하다.

흔한 실수

- 척도 변환을 잊는 것: $\min(N, C)$는 $512 + 16X_M$이므로 $\mathrm{E}(X_M)$의 차에 표준편차(standard deviation) 16을 곱하고, 다시 50을 곱해야 한다.
- $M = -1$에서의 부호 실수: 항 $M(1-\Phi(M))$은 $-(1-\Phi(1))$이 아니라 $-(1-\Phi(-1)) = -\Phi(1)$이다.

STEP 너머

$\mathrm{E}\min(X, M)$은 **중도절단**(*censored*) 정규 확률변수의 기댓값이다. $\mathrm{E}(X-M)^+ = \phi(M) - M(1-\Phi(M))$과 같은 밀접한 식들은 재고 이론(inventory theory, "신문팔이 문제(newsvendor problem)")과 옵션(option)의 블랙-숄즈(Black–Scholes) 가격의 근간이다. 항등식 $\mathrm{E}\min(N, C+1) - \mathrm{E}\min(N, C) = \mathrm{P}(N > C)$는 수용 인원을 한 단위 늘릴 때마다 얻는 가치를 깔끔하게 보여 준다. 이러한 아이디어를 더 알고 싶다면 Grimmett and Welsh, *Probability: An Introduction*(확률론 입문서)을 참고하라.

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.