풀이

STEP I 1997, 문제 14

확률과 통계 • 기대 비용의 최소화

> **STEP I 1997, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/97-s1-q14

핵심 아이디어

비용 T는 확률변수(random variable) X의 함수이므로, 그 기댓값(expectation)은 이 함수에 확률밀도함수(density)를 곱해 적분한 것이다. 적분을 $x = y$에서 나누면 $\mathrm{E}(T)$는 y에 대한 평범한 함수가 되고, 이를 최소화하는 것은 대학 1학년 수준의 미적분 연습이다. 미분하여 0으로 놓고, 그 정류점(stationary point)이 실제로 최솟값(minimum)인지 확인하면 된다.

풀이

대비의 양이 음수인 것은 의미가 없으므로 $y \geq 0$이라 하자. $T = g(X)$로 쓰고, 여기서

$$g(x) = \begin{cases} y & x \leq y, \\ y + r + s(x - y) & x > y. \end{cases}$$

이다. 그러면

$$\begin{aligned} \mathrm{E}(T) &= \int_0^\infty g(x)\,\mathrm{e}^{-x}\,\mathrm{d}x \\ &= \int_0^\infty y\,\mathrm{e}^{-x}\,\mathrm{d}x + \int_y^\infty (r + s(x - y))\mathrm{e}^{-x}\,\mathrm{d}x, \end{aligned}$$

이다. 여기서 어떤 경우에도 지불하는 비용 y를 따로 떼어 냈다. 확률밀도함수를 적분하면 1이므로 첫 번째 적분은 y이다.

두 번째 적분에서는 $u = x - y$로 치환한다. 그러면 $\mathrm{e}^{-x} = \mathrm{e}^{-y}\mathrm{e}^{-u}$이고 u는 0부터 ∞까지 변

한다.

$$\begin{aligned}\int_y^\infty (r+s(x-y))\mathrm{e}^{-x}\,\mathrm{d}x &= \mathrm{e}^{-y}\int_0^\infty (r+su)\,\mathrm{e}^{-u}\,\mathrm{d}u \\ &= \mathrm{e}^{-y}\Big(r\int_0^\infty \mathrm{e}^{-u}\,\mathrm{d}u + s\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u\Big) \\ &= \mathrm{e}^{-y}(r+s),\end{aligned}$$

이때 $\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u = [-u\mathrm{e}^{-u}]_0^\infty + \int_0^\infty \mathrm{e}^{-u}\,\mathrm{d}u = 1$ (부분적분(integration by parts))을 이용하였다. 따라서

$$C(y) = \mathrm{E}(T) = y + (r+s)\,\mathrm{e}^{-y}.$$

생각해 보기 적분하지 않고 인수 $\mathrm{e}^{-y}(r+s)$를 알아볼 방법이 있을까?

있다. 홍수 피해는 확률 $\mathrm{P}(X > y) = \mathrm{e}^{-y}$로 일어난다. 피해가 일어났다는 조건 아래에서 초과량 $X - y$는 다시 평균이 1인 지수분포를 따르므로(지수분포(exponential distribution)는 **무기억성**(*memoryless*)을 가진다), 기대 피해액은 $r + s\cdot 1$이다. 둘을 곱하면 $(r+s)\mathrm{e}^{-y}$를 얻는다.

이제 최소화하자. 미분하면

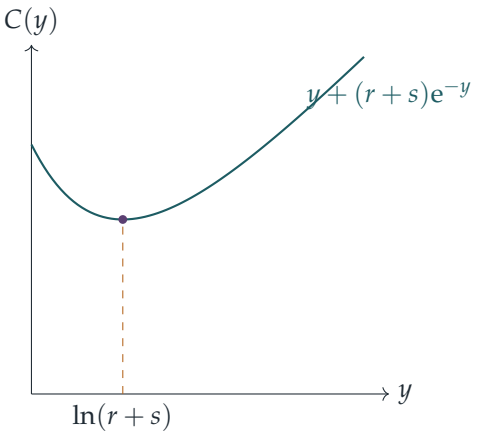
$$C'(y) = 1 - (r+s)\mathrm{e}^{-y}, \qquad C''(y) = (r+s)\mathrm{e}^{-y} > 0.$$

이다. 따라서 $C'(y) = 0$인 것은 정확히 $\mathrm{e}^y = r+s$, 즉 $y = \ln(r+s)$일 때이다. 모든 y에 대해 $C'' > 0$이므로 C는 볼록(convex)하고 C'은 증가한다. 따라서 C는 $y < \ln(r+s)$에서 감소하고 $y > \ln(r+s)$에서 증가한다. 그러므로 이 정류점은 전역 최솟값(global minimum)이고, 기대 총비용은

$$y = \ln(r+s),$$

로 택할 때 최소가 되어 원하는 결과를 얻는다. 최소 기대 비용은 $C(\ln(r+s)) = \ln(r+s) + 1$이다.

이를 위해서는 $\ln(r+s) \geq 0$, 즉 $r+s \geq 1$이어야 하며, 이는 암묵적으로 가정되어 있다. 만약 $r+s < 1$이면 모든 $y \geq 0$에 대해 $C'(y) > 0$이므로 최선의 선택은 $y = 0$이다. 피해가 너무 싸서 대비할 가치가 전혀 없는 것이다.



($r+s=3$으로 그린 개형이다.)

검산

- $y=0$이면 항상 피해를 입으며, $\mathrm{E}(T)=r+s\,\mathrm{E}(X)=r+s$이므로 $C(0)$과 일치한다.
- $r=2$, $s=3$이면 공식은 $y=\ln 5\approx 1.609$와 최소 비용 $1+\ln 5\approx 2.609$를 준다. $C(y)$를 촘촘한 격자 탐색(grid search)으로 조사해도 같은 최소점을 얻으며, $y=1$에서 4×10^5년을 모의실험(simulation)한 결과 평균 비용은 2.844로, 공식의 값 $1+5\mathrm{e}^{-1}\approx 2.839$와 비슷하였다.
- $r+s$가 커지면 y도 커진다. 상식대로, 홍수 피해가 클수록 더 많이 대비할 만하다.

흔한 실수

- 대비 비용 y를 $X\le y$일 때만이 아니라 **매년** 지불한다는 사실을 잊는 것.
- $sx\,\mathrm{e}^{-x}$는 적분하면서 $-sy\,\mathrm{e}^{-x}$ 항을 빠뜨려, $\mathrm{e}^{-y}(r+s)$ 대신 $\mathrm{e}^{-y}(r+s+sy)$를 얻는 것.
- 정류점을 찾고 나서 그것이 최솟값임을 보이지 않는 것.

STEP 너머

이것은 불확실성 아래에서의 의사결정(decision-making under uncertainty)의 작은 예로, 기대 손실(expected loss)을 최소화하는 행동을 고르는 문제이다. 확실한 비용과 기대 벌칙 사이의 균형을 잡는 같은 구조가 운용과학(operations research)의 "신문팔이 문제(newsvendor problem)"와 통계학의 베이즈 결정(Bayes decisions)이라는 개념의 바

탕에 있다. 위에서 사용한 무기억성은 연속분포 가운데 지수분포를 특징짓는 성질이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이 성질과 함께 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.