풀이

STEP I 1989, 문제 15

확률과 통계 • 분포의 비교

> **STEP I 1989, 15번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/89-s1-q15

핵심 아이디어

시속 $12\,\mathrm{km\,h^{-1}}$로 달리면 1킬로미터마다 정확히 5분이 걸리므로, 각 이동 시간은 고정된 분수에 하나의 확률변수(random variable)의 간단한 배수를 더한 것이 된다. 이 확률변수는 균등분포(uniform), 푸아송 분포(Poisson), 이항분포(binomial)를 따른다. 따라서 시간에 관한 모든 질문은 그 하나의 확률변수에 관한 질문이 된다. 가운데 부분은 세 경로가 똑같은 확률로 선택되는 상황에서의 베이즈 정리(Bayes' theorem)이다.

풀이

이동 시간

T_A, T_B, T_D를 세 경로 각각의 이동 시간(분 단위)이라 하자. $12\,\mathrm{km\,h^{-1}}$는 5분마다 1 km를 가는 속력이므로 다음이 성립한다.

- Angle Avenue: $T_A = 25 + A$이고, $A \sim \mathrm{U}(0, 10)$이다.
- Bend Boulevard: $T_B = 20 + 3B$이고, $B \sim \mathrm{Po}(4)$이다.
- Detour Drive: X를 우회 횟수라 하자. 다섯 곳의 도로 공사가 각각 확률 $\frac{4}{5}$로 서로 독립적으로(independently) 작용한다고 가정하면 $X \sim \mathrm{B}(5, \frac{4}{5})$이고, 거리는 $2 + X$ km이며 $T_D = 10 + 5X$이다.

27분 이상

Angle Avenue. $T_A \geq 27 \iff A \geq 2$이므로

$$\mathrm{P}(T_A \geq 27) = \frac{10-2}{10} = \frac{4}{5}.$$

Bend Boulevard. B는 정수이므로 $T_B \geq 27 \iff 3B \geq 7 \iff B \geq 3$이다. 따라서

$$\mathrm{P}(T_B \geq 27) = 1 - \mathrm{e}^{-4}\left(1 + 4 + \frac{4^2}{2!}\right) = 1 - 13\mathrm{e}^{-4} \approx 0.7619.$$

Detour Drive. $T_D \geq 27 \iff 5X \geq 17 \iff X \geq 4$이다. 따라서

$$\mathrm{P}(T_D \geq 27) = 5\left(\frac{4}{5}\right)^4 \frac{1}{5} + \left(\frac{4}{5}\right)^5 = \left(\frac{4}{5}\right)^4 \left(1 + \frac{4}{5}\right) = \frac{256}{625} \cdot \frac{9}{5} = \frac{2304}{3125} = 0.73728.$$

이동이 오래 걸렸다면 어느 경로였을까?

R을 선택한 경로라 하고 $\mathrm{P}(R = A) = \mathrm{P}(R = B) = \mathrm{P}(R = D) = \frac{1}{3}$이라 하자. 또 L을 이동 시간이 27분 이상인 사건이라 하자. 베이즈 정리에 의해 $\frac{1}{3}$인 인수들은 약분된다.

$$\begin{aligned}\mathrm{P}(R = B \mid L) &= \frac{\mathrm{P}(L \mid B)}{\mathrm{P}(L \mid A) + \mathrm{P}(L \mid B) + \mathrm{P}(L \mid D)} = \frac{1 - 13\mathrm{e}^{-4}}{\frac{4}{5} + 1 - 13\mathrm{e}^{-4} + \frac{2304}{3125}} \\ &= \frac{1 - 13\mathrm{e}^{-4}}{\frac{7929}{3125} - 13\mathrm{e}^{-4}} \approx \frac{0.7619}{2.2992} \approx 0.331.\end{aligned}$$

이 값은 $\frac{1}{3}$보다 아주 조금 클 뿐인데, 이는 납득할 만하다. 세 경로가 27분 이상의 이동 시간을 낼 확률이 거의 같기 때문이다.

소문항 (i)

$\mathrm{E}(A) = 5, \mathrm{E}(B) = 4, \mathrm{E}(X) = 5 \times \frac{4}{5} = 4$와 기댓값의 선형성(linearity of expectation)을 이용하면

$$\mathrm{E}(T_A) = 25 + 5 = 30, \qquad \mathrm{E}(T_B) = 20 + 3 \times 4 = 32, \qquad \mathrm{E}(T_D) = 10 + 5 \times 4 = 30.$$

생각해 보기 두 경로가 동점이다. 혹시 계산 실수일까?

아니다. Angle Avenue와 Detour Drive는 둘 다 평균이 정확히 30분이고, Bend Boulevard는 평균이 32분이다. 따라서 평균 시간을 가장 작게 하려면 Angle Avenue와 Detour Drive 중 어느 쪽이든 이용하면 된다. 두 경로는 똑같이 좋고, 둘 다 Bend Boulevard보다 낫다.

소문항 (ii)

이제 각 경로에 대하여 $\mathrm{P}(T < 32)$를 비교한다.

- $T_A < 32 \iff A < 7$이므로 $\mathrm{P}(T_A < 32) = \frac{7}{10} = 0.7$이다.
- $T_B < 32 \iff 3B < 12 \iff B \le 3$이므로

$$\mathrm{P}(T_B < 32) = \mathrm{e}^{-4}\left(1 + 4 + 8 + \frac{32}{3}\right) = \frac{71}{3}\mathrm{e}^{-4} \approx 0.4335.$$

- $T_D < 32 \iff 5X < 22 \iff X \le 4$이므로

$$\mathrm{P}(T_D < 32) = 1 - \left(\frac{4}{5}\right)^5 = 1 - \frac{1024}{3125} = \frac{2101}{3125} \approx 0.6723.$$

$0.7 > 0.6723 > 0.4335$이므로 항상 **Angle Avenue**를 이용해야 하며, 그러면 전체 날 중 70%에서 32분 안에 도착한다.

검산

- 400 000번의 이동을 시뮬레이션(simulation)한 결과 $\mathrm{P}(T \ge 27) \approx 0.801,\ 0.761,\ 0.740$, 평균 시간 $\approx 30.0,\ 32.0,\ 30.0$분, $\mathrm{P}(T < 32) \approx 0.700,\ 0.435,\ 0.670$, 그리고 $\mathrm{P}(R = B \mid L) \approx 0.331$을 얻었다. 모두 정확한 값과 일치한다.
- 경계값 점검: Detour Drive에서 우회가 3번이면 시간은 25분이고 4번이면 30분이므로, "27분 이상"은 정말로 "우회 4번 이상"을 뜻한다.

흔한 실수

- 정수 조건을 대충 다루는 것: $3B \ge 7$은 $B \ge 2$가 아니라 $B \ge 3$을 뜻하고, $3B < 12$는 $B \le 4$가 아니라 $B \le 3$을 뜻한다.
- Detour Drive의 기본 거리 2 km를 빠뜨리거나, 다른 경로에서 1km당 5분의 주행 시간을 빠뜨리는 것.
- (i)에서 하나의 경로만 최선이라고 주장하는 것. 두 평균은 정확히 같다.

STEP 너머

소문항 (i)과 (ii)는 "최선"이 무엇을 최적화하느냐에 따라 달라진다는 것을 보여 준다. 평균은 분포를 요약하는 하나의 값일 뿐이며, 평균이 같은 두 경로라도 퍼진 정도와 꼬리 확률(tail probability)은 매우 다를 수 있다. 결정 이론(decision theory)에서는 이것

이 **손실함수**(loss function)를 선택하는 문제이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 여기에 나온 분포들과 베이즈 정리를 다룬다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)에는 이런 종류의 문제가 많이 실려 있다.

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.