풀이

STEP II 2025, 문제 12

확률과 통계 • 푸아송 분포

STEP II 2025, 12번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2025년 STEP 시험지 묶음(zip 파일) 가운데 STEP II 시험지에 있다:

www.ocr.org.uk/Images/750486-2025-past-papers.zip

핵심 아이디어

푸아송 분포(Poisson distribution)에서는 이웃한 두 확률의 비가 $p_r/p_{r-1} = \lambda/r$로 간단하다. 따라서 모든 차분(difference) d_r를 이미 아는 확률과 r에 대한 간단한 유리식의 곱으로 나타낼 수 있다. d_r의 부호를 보면 p_r가 어디에서 증가하고 감소하는지 알 수 있다. 한 단계 더 나아가 이계 차분(second difference) $d_{r+1} - d_r$의 부호를 보면 d_r가 어디에서 증가하고 감소하는지 알 수 있는데, 이 부호는 r에 대한 이차식(quadratic)이 결정한다. 복잡해 보이는 수 $\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}$는 그 이차식의 한 근(에 1을 더한 값)일 뿐이다.

풀이

풀이 전체에서 $p_r = \mathrm{e}^{-\lambda}\lambda^r/r!$이므로 $r \geq 1$에 대하여

$$p_r = \frac{\lambda}{r}\, p_{r-1}, \qquad p_{r-1} = \frac{r}{\lambda}\, p_r,$$

이고, 모든 p_r는 양수이다.

소문항 (i)

$r \geq 1$에 대하여

$$d_r = p_r - p_{r-1} = p_{r-1}\left(\frac{\lambda}{r} - 1\right) = \frac{p_{r-1}}{r}\,(\lambda - r).$$

$p_{r-1}/r > 0$이므로 d_r의 부호는 $\lambda - r$의 부호와 같고, λ가 정수가 아니므로 d_r는 결코 0이 아니다. λ보다 작은 가장 큰 정수를 $m = \lfloor \lambda \rfloor$라 하면 $\lambda - 1 < m < \lambda$이다($\lambda \notin \mathbb{Z}$이므로 두 부등식 모두 등호가 성립하지 않는다). 그러면

- $1 \leq r \leq m$이면 $r < \lambda$이므로 $d_r > 0$이고 $p_0 < p_1 < \cdots < p_m$이다.

▶ $r \geq m+1$이면 $r > \lambda$이므로 $d_r < 0$이고 $p_m > p_{m+1} > p_{m+2} > \cdots$이다.

($\lambda < 1$이면 $m = 0$이고 첫 번째 경우에 해당하는 r는 없다.) 따라서 $r \neq m$인 모든 r에 대하여 $p_r < p_m$이다. 즉 m은 가장 일어날 가능성이 큰 유일한 값이고, $\lambda - 1 < m < \lambda$이다.

소문항 (ii)

이제 d_r 자체가 어떻게 변하는지 살펴보자. $r \geq 1$에 대하여 $e_r = d_{r+1} - d_r = p_{r+1} - 2p_r + p_{r-1}$이라 하자. 모든 항을 p_r로 나타내면

$$e_r = p_r\left(\frac{\lambda}{r+1} - 2 + \frac{r}{\lambda}\right) = \frac{p_r}{\lambda(r+1)}\Big(\lambda^2 - 2\lambda(r+1) + r(r+1)\Big).$$

앞의 인수가 양수이므로 e_r의 부호는

$$Q(r) = r^2 + (1-2\lambda)r + \lambda^2 - 2\lambda.$$

의 부호와 같다. 이 이차식의 판별식(discriminant)은 $(1-2\lambda)^2 - 4(\lambda^2 - 2\lambda) = 1 + 4\lambda$이므로 두 근은

$$\alpha = \lambda - \tfrac{1}{2} - \sqrt{\lambda + \tfrac{1}{4}}, \qquad \beta = \lambda - \tfrac{1}{2} + \sqrt{\lambda + \tfrac{1}{4}},$$

이고, $Q(r) < 0$인 것은 정확히 $\alpha < r < \beta$일 때이다. 따라서

$$d_{r+1} < d_r \iff \alpha < r < \beta, \qquad d_{r+1} > d_r \iff r < \alpha \text{ 또는} r > \beta.$$

$\beta + 1 = \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}$는 정수가 아니므로 β도 정수가 아니다. 그러므로 $k = \lfloor \beta + 1 \rfloor = \lfloor \beta \rfloor + 1$은 바로

$$k < \lambda + \tfrac{1}{2} + \sqrt{\lambda + \tfrac{1}{4}} < k + 1.$$

을 만족하는 정수이다.

생각해 보기 $r < \alpha$일 때에도 d_r가 증가한다는 점을 생각하면, d_r의 최솟값은 도대체 어디에서 찾아야 할까?

소문항 (i)에서 $r < \lambda$이면 $d_r > 0$이고 $r > \lambda$이면 $d_r < 0$이므로, 최솟값은 음수인 항들, 즉 $r \geq m+1$인 항들 가운데 있어야 한다. 또 $\beta > \lambda$이고($\sqrt{\lambda + \frac{1}{4}} > \frac{1}{2}$이므로) $\alpha < \lambda$이므로, $r \geq m+1$인 모든 r는 $r > \alpha$를 만족하며, $k > \beta > \lambda$에서 $k \geq m+1$이다. 따라서 $r \geq m+1$인 정수에 대하여 다음이 성립한다.

▶ $m+1 \leq r \leq k-1$이면 $r < \beta$이므로 $e_r < 0$이고, $d_{m+1} > d_{m+2} > \cdots > d_k$이다.

▶ $r \geq k$이면 $r > \beta$이므로 $e_r > 0$이고, $d_k < d_{k+1} < d_{k+2} < \cdots$이다.

($k = m+1$이면 첫 번째 경우에 해당하는 r는 없다.) 따라서 d_k는 다른 모든 음수 d_r보다 작고, 물론 모든 양수 d_r보다도 작다. 그러므로 d_r의 최솟값은 $r = k$에서 나타나며, 이것이 증명하려던 것이다. 말로 하면, p_r는 $r = k-1$과 $r = k$ 사이에서 가장 가파르게 감소한다.

소문항 (iii)

먼저 $\lambda < 1$이면 소문항 (i)에 의해 모든 d_r가 음수이다. 그런데 $r \to \infty$일 때 $d_r \to 0$이므로($p_r \to 0$이기 때문에), 어떤 r에 대해서도 그 뒤에 d_r보다 큰 항이 있다. 즉 최댓값이 아예 존재하지 않으며, 당연히 $r = 1$에서 최댓값을 갖지도 않는다. 따라서 $\lambda > 1$이어야 한다(λ는 정수가 아니므로 $\lambda = 1$은 제외된다).

이제 $\lambda > 1$이라 하면 $d_1 > 0$이다. 양수인 항은 $d_1, \dots, d_m$이고, 최댓값은 이들 가운데 있다.

- 최댓값이 $r = 1$에서 나타나면 $d_2 < d_1$, 즉 $e_1 < 0$이다.
- 역으로 $e_1 < 0$이면 $\alpha < 1 < \beta$이므로 $1 \le r < \beta$인 모든 정수 r에 대하여 $e_r < 0$이다. $m < \lambda < \beta$이므로 $d_1 > d_2 > \cdots > d_m$이고, 그 뒤의 항은 모두 음수이다. 따라서 최댓값은 $r = 1$에서 나타난다.

그러므로 $\lambda > 1$일 때 최댓값이 $r = 1$에서 나타날 필요충분조건은 $e_1 < 0$, 즉 $Q(1) < 0$이다.

$$Q(1) = 1 + 1 - 2\lambda + \lambda^2 - 2\lambda = \lambda^2 - 4\lambda + 2 < 0 \iff 2 - \sqrt{2} < \lambda < 2 + \sqrt{2}.$$

이를 $\lambda > 1$과 합치면($2 - \sqrt{2} < 1$이므로) 구하는 조건은

$$1 < \lambda < 2 + \sqrt{2}.$$

소문항 (iv)

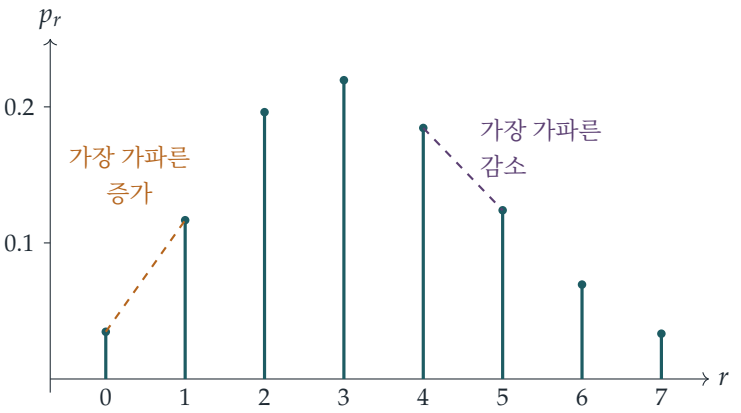
$\lambda = 3.36$이면 $\lambda + \frac{1}{4} = 3.61 = 1.9^2$이므로

- 소문항 (i)에 의해 최빈값(mode)은 $m = 3$이다.
- $\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} = 3.86 + 1.9 = 5.76$이므로 $k = 5$이다. 즉 가장 가파른 감소는 p_4에서 p_5로 갈 때이다.
- $1 < 3.36 < 2 + \sqrt{2} \approx 3.41$이므로 가장 가파른 증가는 p_0에서 p_1로 갈 때이다(아슬아슬하게 그렇다. $\alpha = 0.96$이 1에 가까우므로 d_1과 d_2는 거의 같다).

$p_r = \frac{3.36}{r} p_{r-1}$과 $p_0 = \mathrm{e}^{-3.36} \approx 0.035$를 이용하면 다음 근삿값을 얻는다.

$$p_0, \dots, p_7 \approx 0.035,\ 0.117,\ 0.196,\ 0.220,\ 0.185,\ 0.124,\ 0.069,\ 0.033.$$

그래프의 개형은 처음부터 가파르게 증가하다가 완만해지며 $r = 3$에서 최고점에 이르고, 이어서 점점 가파르게 감소하여 $r = 4$와 $r = 5$ 사이에서 가장 가파르게 감소한 뒤 다시 완만해지는 모양이어야 한다.



검산

- $\lambda \in (0, 40)$인 임의의 값 20,000개에 대하여 p_r와 d_r를 계산해 보면, 최빈값은 $\lfloor \lambda \rfloor$이고, d_r의 최솟값은 $\lfloor \lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} \rfloor$에서 나타나며, ($\lambda > 1$일 때) d_r의 최댓값이 $r = 1$에서 나타나는 것은 정확히 $\lambda < 2 + \sqrt{2} \approx 3.414$일 때임이 확인된다.
- $\lambda = 3.36$일 때: $d_1 \approx 0.0820$, $d_2 \approx 0.0794$이므로 d_1이 (근소한 차이로) 가장 크다. $d_4, d_5, d_6 \approx -0.035, -0.061, -0.055$이므로 예측대로 d_5가 가장 작다.
- λ가 작을 때: $\lambda \to 0$이면 $\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}} \to 1$이므로 $k = 1$이다. 실제로 이때 확률은 거의 모두 0에 몰려 있고, 가장 큰 감소는 첫 번째 감소인 $p_1 - p_0$이다.

흔한 실수

- $r < \alpha$일 때에도 d_r가 **증가**한다는 점을 잊고, 그 구간에 최솟값이 있을 수 없는 이유(그 구간에서는 $d_r > 0$이다)를 설명하지 않는 실수.
- 소문항 (iii)에서 $d_2 < d_1$만 확인하고, 모든 d_r가 음수여서 수열이 최댓값을 갖지 않는 $\lambda < 1$인 경우를 빠뜨리는 실수.
- $\lambda - 1 \leq m \leq \lambda$라고 쓰는 실수. λ가 정수가 아니기 때문에 바로 부등식에서 등호가 성립하지 않는다.

STEP 너머

차분 d_r와 e_r는 일계 도함수와 이계 도함수(second derivative)의 이산적인 유사물이며, k는 푸아송 분포의 이산적인 "변곡점(point of inflection)"을 나타낸다. λ가 크면 푸아송 분포는 평균(mean)이 λ이고 표준편차(standard deviation)가 $\sqrt{\lambda}$인 정규분포(normal

distribution)에 가까운데, 이 정규분포의 변곡점은 $\lambda \pm \sqrt{\lambda}$이다. 여기서 얻은 $\alpha, \beta \approx \lambda \pm \sqrt{\lambda}$와 비교해 보라. 비 p_r / p_{r-1}을 이용하는 기법은 이산 분포의 최빈값을 구하는 표준적인 방법이며, Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)이나 Blitzstein과 Hwang의 *Introduction to Probability*(확률 입문)에서 더 자세히 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.