풀이

STEP III 2020, 문제 12

확률과 통계 • 기하분포와 독립성

> **STEP III 2020, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2020년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:
>
> www.ocr.org.uk/Images/696589-2020-past-papers.zip

핵심 아이디어

모든 것은 서로 독립인 기하분포(geometric distribution) 확률변수의 쌍 (X, Y)의 함수이며, $\mathrm{P}(X = x,\, Y = y) = p^2 q^{x+y-2}$이다. 합과 차는 이 결합확률(joint probability)을 (x, y) 격자에서 한 직선을 따라 더해서 다루는 것이 가장 좋고, 최댓값과 최솟값은 $\mathrm{P}(T \le t)$와 같은 누적확률(cumulative probability)을 이용하는 것이 가장 쉽다. 독립성(independence)을 판단할 때는 한 변수의 어떤 값이 다른 변수를 완전히 결정해 버리는 경우를 찾아본다.

풀이

A는 처음 앞면이 나올 때까지 동전을 던지므로, 처음 $k-1$번이 모두 뒷면이고 k번째가 앞면일 때 X는 값 $k \ge 1$을 갖는다:

$$\mathrm{P}(X = k) = pq^{k-1}, \qquad k = 1, 2, 3, \ldots,$$

Y에 대해서도 마찬가지이다. 두 사람은 따로 동전을 던지므로 X와 Y는 독립이고,

$$\mathrm{P}(X = x,\, Y = y) = p^2 q^{x+y-2} \qquad (x, y \ge 1).$$

반복해서 사용할 두 가지 사실이 있다. $\mathrm{P}(X > k) = q^k$이고(처음 k번이 모두 뒷면), 따라서 $\mathrm{P}(X \le k) = 1 - q^k$이다. 또 $1 - q^2 = (1-q)(1+q) = p(1+q)$이다.

소문항 (i)

합. S는 $s \ge 2$인 값을 갖는다. $S = s$이려면 $1 \le x \le s-1$에 대하여 $X = x$이고 $Y = s - x$이어야 한다. 이 $s-1$개의 결과는 각각 확률이 $p^2 q^{s-2}$이므로

$$\mathrm{P}(S = s) = (s-1)p^2 q^{s-2}, \qquad s = 2, 3, 4, \ldots$$

최댓값. $T \le t$인 것은 $X \le t$와 $Y \le t$가 모두 성립할 때와 정확히 같으므로, 독립성에 의해 $\mathrm{P}(T \le t) = (1-q^t)^2$이다. 따라서 $t \ge 1$에 대하여

$$\begin{aligned}\mathrm{P}(T = t) &= \mathrm{P}(T \le t) - \mathrm{P}(T \le t-1) = (1-q^t)^2 - (1-q^{t-1})^2 \\ &= [(1-q^t) - (1-q^{t-1})][(1-q^t) + (1-q^{t-1})] \\ &= (q^{t-1} - q^t)(2 - q^{t-1} - q^t) \\ &= pq^{t-1}(2 - q^{t-1} - q^t),\end{aligned}$$

이며, 여기서 $q^{t-1} - q^t = q^{t-1}(1-q) = pq^{t-1}$을 이용하였다. 이것이 구하는 결과이다. ($t = 1$일 때 이 식은 $p(1-q) = p^2$을 주며, 올바르다.)

소문항 (ii)

차. $U = 0$이려면 어떤 $k \ge 1$에 대하여 $X = Y = k$이어야 한다:

$$\mathrm{P}(U = 0) = \sum_{k=1}^{\infty} p^2 q^{2k-2} = \frac{p^2}{1-q^2} = \frac{p^2}{p(1+q)} = \frac{p}{1+q}.$$

$u \ge 1$이면 $Y = X + u$ 또는 $X = Y + u$이다. 이 둘은 서로 배반이고 대칭성에 의해 확률이 같으므로

$$\mathrm{P}(U = u) = 2\sum_{k=1}^{\infty} \mathrm{P}(X = k,\, Y = k+u) = 2\sum_{k=1}^{\infty} p^2 q^{2k+u-2} = \frac{2p^2 q^u}{1-q^2} = \frac{2pq^u}{1+q}.$$

최솟값. $W > w$인 것은 $X > w$와 $Y > w$가 모두 성립할 때와 정확히 같으므로 $\mathrm{P}(W > w) = q^{2w}$이고, $w \ge 1$에 대하여

$$\mathrm{P}(W = w) = \mathrm{P}(W > w-1) - \mathrm{P}(W > w) = q^{2w-2} - q^{2w} = (1-q^2)\, q^{2(w-1)}.$$

따라서 W 자체도 "성공" 확률이 $1-q^2$인 기하분포를 따른다. A가 한 번, B가 한 번 던지는 각 회차에서 적어도 한 번 앞면이 나올 확률이 $1-q^2$이기 때문이다.

소문항 (iii)

$S = 2$이면 $X, Y \ge 1$이므로 반드시 $X = Y = 1$이고, 이때 $T = 1$이다. 따라서 $S = 2$와 $T = 3$은 동시에 일어날 수 없다:

$$\mathrm{P}(S = 2 \text{ 이고} T = 3) = 0.$$

한편 $0 < q < 1$이면 $q^2 + q^3 < 2$이므로 $\mathrm{P}(S = 2) = p^2 > 0$이고 $\mathrm{P}(T = 3) = pq^2(2 - q^2 - q^3) > 0$이다. 따라서 $\mathrm{P}(S = 2) \times \mathrm{P}(T = 3) > 0$이고, 양변은 서로 다르다.

소문항 (iv)

U와 W는 독립이다. 모든 $u \geq 0$과 $w \geq 1$에 대하여 $\mathrm{P}(U = u,\, W = w) = \mathrm{P}(U = u)\,\mathrm{P}(W = w)$임을 보여야 한다. U와 W를 모두 알면 작은 값이 w이고 큰 값이 $w + u$임을 알 수 있다.

> ? **생각해 보기** 어떤 쌍 (X, Y)가 $U = u$와 $W = w$를 주는가?

$u = 0$이면 가능한 쌍은 (w, w)뿐이다. $u \geq 1$이면 정확히 두 쌍 $(w,\, w + u)$와 $(w + u,\, w)$가 있으며, 각각의 확률은 $p^2 q^{2w+u-2}$이다.

$u = 0$일 때 $\mathrm{P}(U = 0,\, W = w) = p^2 q^{2w-2}$이고,

$$\mathrm{P}(U = 0)\,\mathrm{P}(W = w) = \frac{p}{1+q} \cdot p(1+q)\, q^{2w-2} = p^2 q^{2w-2}.$$

$u \geq 1$일 때 $\mathrm{P}(U = u,\, W = w) = 2p^2 q^{2w+u-2}$이고,

$$\mathrm{P}(U = u)\,\mathrm{P}(W = w) = \frac{2pq^u}{1+q} \cdot p(1+q)\, q^{2w-2} = 2p^2 q^{2w+u-2}.$$

모든 경우에 두 값이 일치하므로 U와 W는 독립이다.

다른 쌍은 독립이 아니다. 나머지 쌍은 (S, T), (S, U), (S, W), (T, U), (T, W)이다. 두 변수가 독립이 아님을 보이려면 $\mathrm{P}(A = a,\, B = b) \neq \mathrm{P}(A = a)\,\mathrm{P}(B = b)$인 값의 쌍 하나만 찾으면 충분하다.

> ? **생각해 보기** S 또는 T의 어떤 값 하나가 나머지를 모두 결정하는가?

사건 $S = 2$와 사건 $T = 1$은 같은 사건 $\{X = Y = 1\}$이며, 그 확률은 $p^2 > 0$이다. 이 사건에서 $T = 1, U = 0, W = 1$이다. $0 < p, q < 1$이므로 소문항 (i)과 (ii)의 식에 의해 아래 목록에 나오는 T, U, W의 모든 값은 양의 확률을 갖는다. 따라서

- (S, T): 소문항 (iii)에 의해 $\mathrm{P}(S = 2,\, T = 3) = 0 \neq \mathrm{P}(S = 2)\,\mathrm{P}(T = 3)$;
- (S, U): $\mathrm{P}(S = 2,\, U = 1) = 0$이지만 $\mathrm{P}(S = 2)\,\mathrm{P}(U = 1) > 0$;
- (S, W): $\mathrm{P}(S = 2,\, W = 2) = 0$이지만 $\mathrm{P}(S = 2)\,\mathrm{P}(W = 2) > 0$;
- (T, U): $\mathrm{P}(T = 1,\, U = 1) = 0$이지만 $\mathrm{P}(T = 1)\,\mathrm{P}(U = 1) > 0$;
- (T, W): $\mathrm{P}(T = 1,\, W = 2) = 0$이지만 $\mathrm{P}(T = 1)\,\mathrm{P}(W = 2) > 0$.

그러므로 S, T, U, W 가운데 독립인 쌍은 U와 W뿐이다.

검산

- 각 분포의 합은 1이다. 예를 들어 $\sum_{t \geq 1} \mathrm{P}(T = t)$는 망원급수(telescoping series)가 되어 $\lim(1 - q^t)^2 = 1$이고, $\frac{p}{1+q} + \sum_{u \geq 1} \frac{2pq^u}{1+q} = \frac{p}{1+q}\left(1 + \frac{2q}{p}\right) = \frac{p+2q}{1+q} = 1$이다.

- $p = q = \frac{1}{2}$일 때 $\mathrm{P}(T = 2) = \frac{1}{4}\left(2 - \frac{1}{2} - \frac{1}{4}\right) = \frac{5}{16}$이며, 이는 세 결과 $(1,2), (2,1), (2,2)$를 나열하여 얻은 확률 $\frac{1}{8} + \frac{1}{8} + \frac{1}{16}$과 일치한다.
- $p = q = \frac{1}{2}$일 때 $\mathrm{P}(U = 1,\, W = 1) = \mathrm{P}\{(1,2), (2,1)\} = \frac{1}{4}$이고, $\mathrm{P}(U = 1)\,\mathrm{P}(W = 1) = \frac{1}{3} \times \frac{3}{4} = \frac{1}{4}$이므로, 독립성이 요구하는 대로이다.
- 모든 식과 U, W의 독립성은 $p = 0.3, 0.55, 0.8$에 대하여 결합분포를 합산하는 컴퓨터 계산으로 확인하였다.

흔한 실수

- $u \geq 1$일 때 $\mathrm{P}(U = u)$에서 인수 2를 빠뜨리거나, $X = Y$가 한 번만 세어지는 $u = 0$일 때 잘못 포함하는 실수.
- 경우를 직접 더해서 $\mathrm{P}(T = t)$를 구하려다 $X = Y = t$를 두 번 세는 실수. 누적확률을 이용하면 이를 피할 수 있다.
- 값의 쌍 하나만 확인하면 독립성이 증명된다고 생각하는 실수. 독립성을 보이려면 **모든** u와 w를 확인해야 하지만, 종속성은 한 번 실패하는 것만으로 충분하다.

STEP 너머

U와 W의 독립성은 무기억성(memoryless property)의 이산 형태이다. 시각 W에 처음 앞면이 나오고 나면, 다른 사람이 남은 시간 동안 기다리는 시간은 그때까지 얼마나 걸렸는지와 상관없이 새로운 기하분포 확률변수가 된다. 서로 독립인 지수분포(exponential distribution) 확률변수에 대한 연속 버전은 최솟값과, 최솟값에서 최댓값까지의 간격이 서로 독립이라는 것으로, 이는 푸아송 과정(Poisson process)과 대기행렬 이론(queueing theory)의 바탕이 된다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)에서 이러한 개념들을 전개한다.

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.