풀이

STEP III 2017, 문제 12

확률과 통계 • 이산 결합 분포

> **STEP III 2017, 12번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/17-s3-q12

핵심 아이디어

모든 것은 두 가지 기본 합 공식 $\sum_{r=1}^{n} r = \frac{1}{2}n(n+1)$과 $\sum_{r=1}^{n} r^2 = \frac{1}{6}n(n+1)(2n+1)$에서 나온다. 먼저 확률의 합이 1이 되도록 하여 k를 정한다. 그다음 y에 대해 합하여 주변분포(marginal distribution)를 구한다. 공분산(covariance)을 구하려면 $\mathrm{E}(X)$와 $\mathrm{E}(XY)$가 필요하다. $x+y$가 x와 y에 대해 대칭(symmetric)이라는 사실 덕분에 계산량이 절반으로 준다.

풀이

소문항 (i)

n^2개의 모든 순서쌍 (x,y)에 대한 확률의 합은 1이어야 한다. 대칭성에 의해 $\sum_{x,y} x = \sum_{x,y} y$이다. 각 x는 y의 n개의 값마다 한 번씩 나타나므로, 이 두 합은 각각 $n\sum_{x=1}^{n} x = \frac{1}{2}n^2(n+1)$과 같다. 따라서

$$1 = k\sum_{x=1}^{n}\sum_{y=1}^{n}(x+y) = k\cdot 2\cdot \tfrac{1}{2}n^2(n+1) = k\,n^2(n+1), \qquad \text{따라서} \qquad k = \frac{1}{n^2(n+1)}.$$

X의 주변분포를 구하기 위해 x를 고정하고 y에 대해 합한다.

$$\begin{aligned}\mathrm{P}(X=x) &= k\sum_{y=1}^{n}(x+y) = k\Big(nx + \tfrac{1}{2}n(n+1)\Big) = \frac{kn\,(2x+n+1)}{2}\\ &= \frac{n+1+2x}{2n(n+1)},\end{aligned}$$

이것이 구하려던 결과이다. $x+y$는 x와 y에 대해 대칭이므로 Y도 같은 분포를 가진다. 즉 $\mathrm{P}(Y=y) = \dfrac{n+1+2y}{2n(n+1)}$이다.

생각해 보기 X와 Y가 독립이 아님을 보이려면 모든 순서쌍 (x,y)를 확인해야 할까?

그렇지 않다. 독립(independence)이려면 **모든** 순서쌍에 대해 $\mathrm{P}(X=x, Y=y) = \mathrm{P}(X=x)\,\mathrm{P}(Y=y)$가 성립해야 하므로, 이것이 성립하지 않는 순서쌍 하나만 찾으면 충분하다. $x=y=1$로 두면

$$\mathrm{P}(X=1, Y=1) = 2k = \frac{2}{n^2(n+1)}, \qquad \mathrm{P}(X=1)\,\mathrm{P}(Y=1) = \frac{(n+3)^2}{4n^2(n+1)^2}.$$

두 값이 같을 필요충분조건은 $8(n+1) = (n+3)^2$, 즉 $n^2 - 2n + 1 = 0$, 곧 $(n-1)^2 = 0$이다. 따라서 두 값은 $n=1$일 때만 같다. $n \geq 2$이므로 두 확률은 다르고, X와 Y는 **독립이 아니다**.

소문항 (ii)

$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$를 이용한다. 또 두 확률변수의 분포가 같으므로 $\mathrm{E}(Y) = \mathrm{E}(X)$임에 주목한다.

평균. 소문항 (i)에서 구한 주변분포를 이용하면

$$\begin{aligned}\mathrm{E}(X) &= \frac{1}{2n(n+1)} \sum_{x=1}^{n} \left((n+1)x + 2x^2\right) = \frac{1}{2n(n+1)} \left(\tfrac{1}{2}n(n+1)^2 + \tfrac{1}{3}n(n+1)(2n+1)\right) \\ &= \frac{1}{2}\left(\frac{n+1}{2} + \frac{2n+1}{3}\right) = \frac{7n+5}{12}.\end{aligned}$$

곱의 평균. 여기서 $xy(x+y) = x^2y + xy^2$이고, 대칭성에 의해 두 부분의 이중합은 같다. 또한 x^2y의 이중합은 두 단일합의 곱으로 인수분해된다.

$$\begin{aligned}\mathrm{E}(XY) &= k \sum_{x=1}^{n}\sum_{y=1}^{n} xy(x+y) = 2k\left(\sum_{x=1}^{n} x^2\right)\left(\sum_{y=1}^{n} y\right) = 2k \cdot \frac{n(n+1)(2n+1)}{6} \cdot \frac{n(n+1)}{2} \\ &= \frac{n^2(n+1)^2(2n+1)}{6\,n^2(n+1)} = \frac{(n+1)(2n+1)}{6}.\end{aligned}$$

공분산. 이들을 공통분모 144로 통분하여 합치면

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \frac{(n+1)(2n+1)}{6} - \frac{(7n+5)^2}{144} = \frac{24(2n^2+3n+1) - (49n^2+70n+25)}{144} \\ &= \frac{-n^2+2n-1}{144} = -\frac{(n-1)^2}{144}.\end{aligned}$$

$n \geq 2$이므로 $(n-1)^2 > 0$이고, 따라서 $\mathrm{Cov}(X,Y) = -\dfrac{(n-1)^2}{144} < 0$이다. 요구한 대로 공분산은 음수이다.

공분산이 음수인 것은 납득할 만하다. 결합확률은 $x+y$에만 의존하며, 합이 클수록 확률이 커진다. X가 이미 크다면 합을 크게 하기 위해 Y까지 클 필요는 없으므로, Y가 클 이유가 조금 줄어든다.

검산

- $n=2$로 두자. 그러면 $k=\frac{1}{12}$이고, 네 확률은 $\frac{2}{12}, \frac{3}{12}, \frac{3}{12}, \frac{4}{12}$로 합이 1이다. 주변확률은 $\mathrm{P}(X=1)=\frac{5}{12}$, $\mathrm{P}(X=2)=\frac{7}{12}$로 공식과 일치한다. 따라서 $\mathrm{E}(X)=\frac{19}{12}$이고 $\mathrm{E}(XY)=\frac{2+6+6+16}{12}=\frac{5}{2}$이다. 이로부터 $\mathrm{Cov}(X,Y)=\frac{5}{2}-\frac{361}{144}=-\frac{1}{144}$를 얻으며, 이는 $-(n-1)^2/144$와 일치한다.
- 주변확률의 합은 1이다. $\sum_x (n+1+2x)=n(n+1)+n(n+1)=2n(n+1)$이기 때문이다.
- 독립성 판정을 실패하게 만드는 $(n-1)^2$은 공분산을 영으로 만드는 것과 같은 값이다. 두 경우 모두 $n=1$은 확실하게 $X=Y=1$이 되는 퇴화된 경우이다.
- $n=2,\ldots,29$에 대해 정확한 분포를 컴퓨터로 확인하여 k, 주변분포, $\mathrm{E}(X)$, $\mathrm{E}(XY)$와 공분산 공식이 맞음을 확인했다.

흔한 실수

- $\sum_{x,y}$ x에 y에 대한 합에서 오는 인수 n이 있다는 것을 잊는 실수. 그러면 $k=1/(n(n+1))$이라는 틀린 값을 얻는다.
- 구체적인 순서쌍 하나를 제시하지 않고 비독립성을 "일반적으로" 보이려 하는 실수. 반례(counterexample) 하나면 충분하며, 그것이 가장 명확한 논증이다.
- 종속임을 보이려면 공분산이 영이 아니어야 한다고 생각하는 실수. 여기서는 공분산이 실제로 종속임을 보여 주지만, 공분산이 영이라고 해서 독립임이 증명되는 것은 **아니다**.

STEP 너머

$f(x,y)=g(x+y)$ 꼴의 결합분포(joint distribution)는 합을 통해 서로 연결된 확률변수의 간단한 모형이다. 이때 공분산의 부호는 g의 모양에 따라 달라진다. 여기서 g는 일차함수이고, 공분산은 작은 음수이다. 대학에서는 결합분포, 주변분포, 조건부 분포(conditional distribution), 상관(correlation)을 함께 공부하며, 연속형의 경우에는 이중합 대신 이중적분(double integral)을 사용한다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 STEP과 가까운 방식으로 이 내용을 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.