풀이

STEP III 2000, 문제 14

확률과 통계 • 공분산과 독립

STEP III 2000, 14번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/00-s3-q14

핵심 아이디어

빠진 네 번째 확률을 먼저 채운 다음, $\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$를 항별로 전개한다. 모든 항이 하나의 인수 $a - pq$와 결코 0이 될 수 없는 어떤 식의 곱으로 모이며, $a = pq$는 바로 왼쪽 위 칸이 곱으로 분해된다는 진술이다. 2×2 표에서 한 칸이 곱으로 분해되면, 주변합이 나머지 세 칸도 곱으로 분해되도록 강제한다. 값이 세 개이면 여유가 더 생기고, 대칭인 표를 쓰면 이 연결이 끊어진다.

풀이

항등식

네 확률의 합은 1이므로

$$\mathrm{P}(X = x_2,\, Y = y_2) = 1 - a - (q - a) - (p - a) = 1 - p - q + a.$$

행과 열을 따라 더하면 주변분포(marginal distribution)를 얻는다.

$$\mathrm{P}(X = x_1) = a + (q - a) = q, \qquad \mathrm{P}(Y = y_1) = a + (p - a) = p,$$

따라서 $\mathrm{P}(X = x_2) = 1 - q$이고 $\mathrm{P}(Y = y_2) = 1 - p$이다. 이것이 p와 q의 진짜 의미이다.

	y_1	y_2	
x_1	a	$q - a$	q
x_2	$p - a$	$1 - p - q + a$	$1 - q$
	p	$1 - p$	

이제

$$\begin{aligned}\mathrm{E}(XY) &= a\,x_1y_1 + (q-a)\,x_1y_2 + (p-a)\,x_2y_1 + (1-p-q+a)\,x_2y_2,\\ \mathrm{E}(X)\mathrm{E}(Y) &= \big(qx_1 + (1-q)x_2\big)\big(py_1 + (1-p)y_2\big)\\ &= pq\,x_1y_1 + q(1-p)\,x_1y_2 + p(1-q)\,x_2y_1 + (1-p)(1-q)\,x_2y_2.\end{aligned}$$

계수끼리 빼면 다음과 같다.

$$\begin{aligned}x_1y_1:&\quad a-pq,\\ x_1y_2:&\quad (q-a)-(q-pq) = -(a-pq),\\ x_2y_1:&\quad (p-a)-(p-pq) = -(a-pq),\\ x_2y_2:&\quad (1-p-q+a)-(1-p-q+pq) = a-pq.\end{aligned}$$

그러므로

$$\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = (a-pq)(x_1y_1 - x_1y_2 - x_2y_1 + x_2y_2) = (a-pq)(x_1-x_2)(y_1-y_2).$$

따라서 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$이면 좌변이 0이 되어 $(a-pq)(x_1-x_2)(y_1-y_2) = 0$이고, 이것이 보여야 할 것이었다.

독립성

서로 다른 두 값을 각각 취하는 확률변수(random variable)의 쌍은 적당한 a, p, q에 대하여 (각 칸에 이름만 붙이면) 위와 같은 꼴의 결합분포표(joint table)를 가진다. $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$이면 위의 곱은 0이다. $x_1 \neq x_2$이고 $y_1 \neq y_2$이므로 뒤의 두 인수는 0이 아니고, 따라서 $a = pq$이다.

> **생각해 보기** 이제 한 칸이 곱으로 분해된다는 것을 안다. 왜 이것으로 나머지 세 칸이 해결되는가?

나머지 각 칸은 주변합에서 이미 아는 칸을 뺀 것이기 때문이다.

$$\begin{aligned}&\mathrm{P}(X=x_1, Y=y_2) = q-pq = q(1-p) = \mathrm{P}(X=x_1)\,\mathrm{P}(Y=y_2),\\ &\mathrm{P}(X=x_2, Y=y_1) = p-pq = p(1-q) = \mathrm{P}(X=x_2)\,\mathrm{P}(Y=y_1),\\ &\mathrm{P}(X=x_2, Y=y_2) = 1-p-q+pq = (1-p)(1-q) = \mathrm{P}(X=x_2)\,\mathrm{P}(Y=y_2).\end{aligned}$$

모든 칸이 주변확률의 곱이므로 X와 Y는 독립(independent)이다. (역, 즉 독립이면 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$라는 것은 잘 알려진 사실이므로, 두 값을 취하는 확률변수에 대해서는 두 조건이 동치이다.)

세 값을 취하는 예

A와 B는 각각 $\{-1, 0, 1\}$ 위의 균등분포(uniform distribution)를 따르므로 $\mathrm{E}(A) = \mathrm{E}(B) = 0$이다. 따라서 모든 행과 열의 합이 $\frac{1}{3}$이고, $\mathrm{E}(AB) = 0$이지만 모든 칸이 $\frac{1}{9}$인 것은 아닌 표가 필요하다. 대칭인 선택이 통한다.

	$B=-1$	$B=0$	$B=1$
$A=-1$	$\frac{1}{6}$	0	$\frac{1}{6}$
$A=0$	0	$\frac{1}{3}$	0
$A=1$	$\frac{1}{6}$	0	$\frac{1}{6}$

각 행과 각 열의 합이 $\frac{1}{3}$이므로 A와 B는 각각 $-1, 0, 1$을 확률 $\frac{1}{3}$로 취한다. 0이 아닌 곱 ab는 네 귀퉁이에서만 나온다.

$$\mathrm{E}(AB) = \tfrac{1}{6}(1) + \tfrac{1}{6}(-1) + \tfrac{1}{6}(-1) + \tfrac{1}{6}(1) = 0 = \mathrm{E}(A)\mathrm{E}(B).$$

그러나 $\mathrm{P}(A=0, B=0) = \frac{1}{3} \neq \frac{1}{9} = \mathrm{P}(A=0)\,\mathrm{P}(B=0)$이므로 A와 B는 독립이 아니다. 말로 하면, $B=0$인 것은 정확히 $A=0$일 때이고, 그렇지 않을 때에는 A가 무엇이든 B는 같은 확률로 ± 1이다.

검산

- $x_1 = y_1 = 1$, $x_2 = y_2 = 0$으로 두고 $X = Y$인 경우를 시험해 보자. 그러면 $a = p = q$이고, 항등식은 $\mathrm{Var}(X) = p - p^2$를 말해 주는데, 이는 베르누이(Bernoulli) 확률변수에 대해 옳은 값이다.
- 이 항등식은 다항식 항등식으로서 성립한다. x_i, y_i, a, p, q를 무작위로 2000번 골라 유리수 연산으로 정확히 확인하였다. 세 값을 취하는 표도 같은 방법으로 확인하였다. 주변확률은 모두 $\frac{1}{3}$이고, $\mathrm{E}(AB) = 0$이며, $\mathrm{P}(A=0, B=0) = \frac{1}{3}$이다.

흔한 실수

- $a = pq$를 보이고 멈추는 것. 독립이려면 **모든** 칸이 곱으로 분해되어야 하므로, 나머지 세 칸도 (간단히) 확인해야 한다.
- $(x_1 - x_2)(y_1 - y_2)$로 나누어도 되는 이유를 말하지 않는 것. 값들이 서로 다르므로 이 식은 0이 아니다.
- 마지막 소문항에서 $B = A^2$나 $B = |A|$를 제시하는 것. 이들은 공분산이 0이면서 종속이지만, B가 결코 -1이 되지 않으므로 요구된 주변분포를 갖지 않는다.

STEP 너머

양 $\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$는 공분산(covariance)이며, 이 문제는 "상관이 없다(uncorrelated)"가 "독립이다"보다 약한 조건임을 보여 준다. 단, 두 값을 취하는 (지시변수와 비슷한) 확률변수는 예외인데, 이 경우 항등식이 모든 것을 두 지시변수(indicator)의 공분산으로 환원하기 때문이다. 앞으로 만나게 될 관련된 사실이 하나 있다. 결합정규분포(jointly normal)를 따르는 확률변수에 대해서는 공분산이 0이면 실제로 독립이다. 더 알고 싶다면 Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문서), 또는 Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문서)를 보라.

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.