12.4 둘씩은 독립이지만 한꺼번에는 독립이 아닌

학습지

문제

STEP III 2021, 12번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2021년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:

www.ocr.org.uk/Images/696588-2021-past-papers.zip

힌트

먼저 혼자 힘으로 문제를 풀어 보자. 막히면 힌트를 **하나씩** 읽고, 힌트를 하나 읽을 때마다 다시 문제로 돌아가자. **잠깐, 직접 해 보기**라고 표시된 힌트가 핵심 단계이다. 각 힌트 뒤에는 풀이를 적을 공간이 있다.

? **생각해 보기** $\mathrm{P}(X_{12}=1)$은 얼마인가? $\mathrm{P}(X_{12}=1$ 이고$X_{23}=1)$은 얼마인가?

1단계 쌍별 독립(pairwise independence)

다루는 내용: 소문항 (i), 독립성(independence)

? **생각해 보기** $\mathrm{P}(X_{12}=1)$은 얼마인가? 또 $\mathrm{P}(X_{12}=1$ 그리고$X_{23}=1)$은 얼마인가? 이를 두 확률의 곱과 비교해 보라.

2단계 점수의 평균(mean)과 분산(variance)

다루는 내용: 소문항 (i), 평균과 분산

잠깐, 직접 해 보기 점수는 $\binom{n}{2}$개 확률변수의 합이다. 기댓값(expectation)은 항상 더해진다. 합의 분산이 분산의 합과 같아지는 것은 언제인가? 쌍별 독립만으로 충분한가?

3단계 분산 공식

다루는 내용: 소문항 (ii)

생각해 보기 Y_i의 평균이 0이면, 합의 분산은 그 제곱의 E이다. 제곱을 전개해 보라. 각 곱 Y_iY_j는 몇 번씩 나타나는가?

4단계 홀짝 게임

다루는 내용: 소문항 (iii), 독립이 아님

생각해 보기 Z_{12}의 분포(distribution)와 평균을 구하라. $Z_{12}Z_{23}$이 0이 아닌 것은 언제이며, 그때의 값은 얼마인가? $\mathrm{E}(Z_{12}Z_{23})$를 $\mathrm{E}(Z_{12})\mathrm{E}(Z_{23})$와 비교해 보라.

5단계 **분산**

다루는 내용: 소문항 (iii), 분산

> **잠깐, 직접 해 보기** 소문항 (ii)를 이용하라. 교차항(cross terms)에 기여하는 쌍들의 쌍은 어느 것인가? 공통인 참가자가 없는 것인가, 아니면 정확히 한 명을 공유하는 것인가? 각각은 몇 개씩 있는가?

숨은 수학

여러 사건의 독립

사건 $A_1, \ldots, A_n$이 **독립**이라는 것은, 이들 가운데 어떤 일부를 골라도 그 교집합의 확률이 각 확률의 곱과 같다는 뜻이다. 세 사건의 경우 이는 네 가지 조건, 곧 쌍에 대한 세 조건과 $\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$를 뜻한다. 세르게이 베른시테인(Sergei Bernstein)은 1920년대에 그림 12.4.1 (온라인 전체 절에 수록)의 예를 들어 쌍별 조건이 세 사건에 대한 조건을 함의하지 않음을 보였다. (거꾸로, 세 사건에 대한 조건만으로는 쌍별 조건도 함의되지 않는다.) 문제의 일치 사건 X_{12}, X_{23}, X_{13}은 베른시테인의 동전과 정확히 같은 구조이다.

상관없음은 독립이 아니다

공분산이 0이라는 것, 곧 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$는 독립보다 더욱 약한 조건이다. STEP III 2000 Q14(연습문제에 있다)는 두 값만 가지는 확률변수에서는 두 개념이 일치하지만, 세 값을 가지면 그렇지 않음을 보여 준다. 간단한 예를 들어 보자. X가 $\{-1, 0, 1\}$ 위의 균등분포를 따르고 $Y = X^2$이라 하자. 그러면 $\mathrm{E}(XY) = \mathrm{E}(X^3) = 0 = \mathrm{E}(X)\mathrm{E}(Y)$이지만, Y는 X에 의해 완전히 결정된다. 공분산은 선형적인 관계만을 측정한다.

값싼 무작위성

컴퓨터가 난수를 만드는 데에는 많은 비용이 든다. 1979년 Larry Carter와 Mark Wegman은 데이터를 유사 무작위 위치에 저장하는 **해싱**(*hashing*)을 비롯한 많은 알고리즘에서, 쌍별 독립인 무작위 선택이 완전히 독립인 선택만큼이나 좋다는 것을 보였다. 쌍별 독립인 값은 매우 적은 비용으로 만들 수 있다. 예를 들어 a와 b를 무작위로 고르고 $k = 0, 1, \ldots, p-1$에 대해 $a + bk \pmod{p}$를 계산하면, 어느 두 값도 서로 독립인 p개의 값을 얻는다. 이것이 잘 작동하는 이유가 바로 이 문제의 분산 계산이다. 체비쇼프 부등식(Chebyshev's inequality)에는 분산만 필요하고, 분산에는 쌍만 필요하다.

더 도전하기

1 **두 값.** STEP III 2000, 14번 문제 (step.maths.org/questions/00-s3-q14)에 도전해 보라. 이 문제는 각각 두 값만 가지는 확률변수에서는 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$이면 반드시 독립임을 보이고, 세 값을 가지는 경우의 반례를 요구한다.

2 **더 큰 팀.** 소문항 (i)에서 $n = 7$일 때 점수의 평균과 표준편차(standard deviation)를 구하고, 예시로 나온 점수 7과 비교하라.

3 **다시 생일 문제.** 지시변수를 이용하여 k명 가운데 생일이 같은 쌍의 개수의 평균과 분산을 구하라(Section 9.9). 이 지시변수들은 쌍별 독립인가?

4 **세 사건 조건은 성립하지만 쌍별 독립은 아닌 경우.** 확률이 각각 $\frac{1}{2}$이고 $\mathrm{P}(A \cap B \cap C) = \frac{1}{8}$이지만 쌍별 독립은 아닌 세 사건을 찾아라.

더 읽을거리

- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes*** (Oxford University Press). 독립성을 다루는 앞부분의 장들.
- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing*** (Cambridge University Press). 쌍별 독립과 그 알고리즘에서의 활용.
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability*** (CRC Press, 온라인 무료 공개). 지시변수와 공분산에 관한 내용.

안내에 따른 전체 풀이는 제 12.4절에 있다:
https://beyondstep.keithmcnulty.org/ko/ch12/pairwise-independence.html

매주 STEP 문제 하나를 끝까지 풀어 본다. The STEP Up은 이메일로 무료로 받아 볼 수 있다: keithmcnulty.substack.com