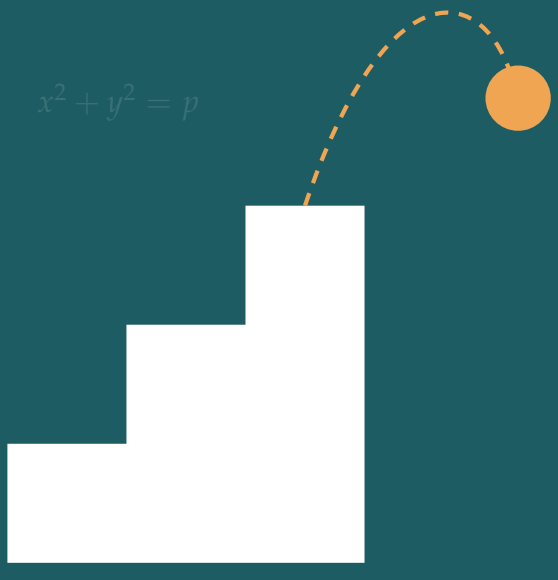
STEP 대비 · 대학 수학

Beyond STEP

제12장

확률



KEITH McNULTY

Beyond STEP

STEP 기출문제로 여는
대학 수학의 세계

제12장

확률

Keith McNulty

Beyond STEP: STEP 기출문제로 여는 대학 수학의 세계

제12장: 확률

이 PDF는 *Beyond STEP*의 전체 13개 장 가운데 제12장이다. 책 전체, 각 장의 개별 PDF, 모든 절의 인쇄용 연습지, 모든 STEP 문제의 자세한 풀이를 `beyondstep.keithmcnulty.org`에서 무료로 읽고 내려받을 수 있다.

2026년 11월 온라인 초판 발행.

이 PDF에서 지은이가 직접 만든 모든 것, 곧 본문, 풀이, 그림, 연습문제와 해답은 크리에이티브 커먼즈 CC0 1.0 Universal 라이선스에 따라 퍼블릭 도메인(public domain)으로 공개한다. 누구든 허락을 구하지 않고 어떤 목적으로든 복사하고, 고쳐 쓰고, 공유할 수 있다. 자세한 내용은 `creativecommons.org/publicdomain/zero/1.0`을 보라.

여기에서 다루는 STEP 문제의 저작권은 Cambridge University Press & Assessment에 있다. 문제는 이 책에 싣지 않았으며, 각 문제마다 공식 출처인 STEP Support Programme의 문제 데이터베이스나 OCR의 기출 시험지로 가는 링크를 달았다.

오류 제보는 `github.com/keithmcnulty/step_book_mcnulty`에 이슈로 남겨 주기 바란다.

장

12

확률

불확실성도 정확하게 추론할 수 있다. 그러나 직관은 그렇지 못하다는 것이 드러난다.

이 장의 내용

12.1	조건을 정확하게 걸기	STEP I 2007
12.2	기저율 오류	STEP II 1998
12.3	"무작위"란 무슨 뜻인가?	STEP III 2022
12.4	둘씩은 독립이지만 한꺼번에는 독립이 아닌	STEP III 2021
12.5	마지막 승객의 좌석	STEP II 2021
12.6	동전 하나 더	STEP I 2016
12.7	토너먼트에서 누가 누구와 만나는가	STEP II 견본 문제지
12.8	페니의 게임	STEP III 2001
12.9	치우친 동전으로 공정하게 정하기	STEP I 2019
12.10	느슨한 끈 끝 묶기	STEP III 2008
12.11	우표, 컵, 그리고 e	STEP I 1996
12.12	언제 그만 찾아야 하는가	STEP I 2001
12.13	삼자 결투	STEP II 1995
12.14	혼합 전략	STEP I 1989
12.15	멸종인가 폭발인가	STEP II 2003
12.16	앞으로 보고 거꾸로 보는 마르코프 연쇄	STEP III 1996
12.17	바나흐의 성냥갑	STEP II 1992
12.18	무기억성	STEP III 2021
12.19	평면 위의 무작위 점	STEP II 2012
12.20	뷔퐁의 바늘	STEP I 1993
12.21	막대 부러뜨리기	STEP III 1990
12.22	평균이 없는 분포	STEP III 1995
12.23	확률로 얻는 스털링 공식	STEP II 2016
12.24	중심극한정리	STEP III 2008
12.25	여럿 가운데 가장 큰 값	STEP III 2025
A	연습문제 해답	157쪽

우연에 대해 추론하기

> 확률론은 계산으로 환원된 상식에 지나지 않는다.
>
> — 피에르시몽 라플라스(Pierre-Simon Laplace), *Théorie analytique des probabilités* (1812)

라플라스의 유명한 말은 절반만 옳다. 확률(probability)은 상식을 정밀하게 만든 것이다. 그러나 이 장에서 거듭 보게 되듯이, 상식을 그대로 내버려 두면 틀리고 만다. 사람들은 희귀병 검사에서 양성이 나오면 아마 그 병에 걸렸을 것이라고 생각한다. 비행기에 처음 탑승한 승객이 엉뚱한 자리에 앉으면 마지막 승객이 자기 자리에 앉을 가능성은 거의 없다고 생각한다. 동전 던지기에서 어떤 앞뒷면 배열이 같은 길이의 다른 배열을 이길 수는 없다고 생각한다. 많은 측정값의 평균은 언제나 어떤 값으로 안정된다고 생각한다. 이것들은 모두 거짓이며, 각각의 경우 짧은 계산 하나가 진실을, 더 나아가 그 이유까지 드러낸다.

확률론은 17세기에 블레즈 파스칼(Blaise Pascal)과 피에르 드 페르마(Pierre de Fermat)가 주고받은 편지 속에서 도박에 관한 질문들로부터 시작되었다. 확률론이 엄밀한 수학의 한 분야가 된 것은 1933년에 이르러서였는데, 이때 안드레이 콜모고로프(Andrey Kolmogorov)가 그 공리(axiom)를 제시하였다. 그 사이에, 그리고 그 이후로, 확률론은 수학에서 가장 널리 쓰이는 분야 가운데 하나로 성장하였다.

확률은 왜 중요한가

불확실성을 다루는 모든 분야는 확률을 사용한다. 의학(진단 검사와 임상 시험), 법학(증거의 무게), 금융(위험과 보험), 공학(신뢰성과 홍수), 컴퓨터 과학(무작위 알고리즘, 암호학, 기계 학습), 물리학(양자역학과 통계역학), 생물학(유전학과 전염병)이 모두 그렇다. 현대 인공지능은 상당 부분 응용 확률론이다. 언어 모델은 다음 단어가 나올 확률을 예측한다. 그리고 위험, 우연, 증거에 대해 올바르게 추론하는 능력은 교육받은 사람이라면 누구나 지닐 수 있는 가장 값진 능력 가운데 하나이다.

대학에서 만나게 될 내용

대학 1학년 확률론에서는 공리, 조건부 확률(conditional probability)과 베이즈 정리(Bayes' theorem), 독립(independence), 이산 확률변수와 연속 확률변수(random variable), 기댓값(expectation)과 분산(variance), 생성함수(generating function), 그리고 두 가지 위대한 극한 정리인 큰 수의 법칙(law of large numbers)과 중심극한정리(central limit theorem)를 다룬다. 이후의 과목에서는 확률과정(random process)(마르코프 연쇄(Markov chain), 푸

아송 과정(Poisson process), 무작위 행보(random walk), 분기 과정(branching process)), 측도론적 확률론(measure-theoretic probability)(극한과 연속분포를 튼튼한 기반 위에 올려놓는다), 마팅게일(martingale)을 공부하며, 확률론을 통계학, 해석학, 조합론, 물리학과 연결한다.

이 장을 관통하는 질문들

세 가지 큰 질문

1. **질문이 정확히 무엇인가?** 확률 실험은 무엇이고, 무엇을 조건으로 삼고 있으며, "무작위로"라는 말은 무슨 뜻인가?
2. **영리한 접근법은 무엇인가?** 언제 대칭(symmetry), 지시변수(indicator), 조건부화, 잘 고른 상태가 긴 계산을 대신할 수 있는가?
3. **장기적으로 무슨 일이 일어나는가?** 확률과정은 시간이 흐름에 따라 어떻게 행동하며, 많은 확률적인 양의 합, 평균, 최댓값은 어떤 모습인가?

이 장은 다섯 부분으로 이루어져 있다. **명료하게 생각하기**(Section 12.1에서 Section 12.4까지)에서는 함정들, 즉 조건부화, 기저율(base rate), "무작위"의 의미, 독립의 의미를 다룬다. **대칭과 영리한 논증**(Section 12.5에서 Section 12.9까지)에는 아름답고 짧은 풀이를 가진 문제들을 모았다. **기댓값, 게임, 멈추기**(Section 12.10에서 Section 12.14까지)에서는 평균과 의사 결정, 즉 언제 멈출지, 누구를 쏠지, 어떻게 무작위화할지로 눈을 돌린다. **확률과정**(Section 12.15에서 Section 12.19까지)에서는 시간과 공간 속에서 무작위성을 따라간다. 그리고 **극한과 연속 확률**(Section 12.20에서 Section 12.25까지)은 기하학적 확률(geometric probability)과 위대한 극한 정리들로 마무리한다.

이 장의 아이디어

§	주제	아이디어
12.1	조건부화	"특정한 하나" 대 "적어도 하나"
12.2	베이즈 정리	기저율과 자연 빈도(natural frequency)
12.3	베르트랑(Bertrand)의 역설	"무작위"에는 정의가 필요하다
12.4	독립	쌍별 독립(pairwise independence)은 상호 독립(mutual independence)이 아니다
12.5	비행기 좌석	두 좌석 사이의 대칭
12.6	동전 하나 더	서로 여사건인 대칭적 사건
12.7	토너먼트	경기와 쌍 세기
12.8	Penney의 게임	첫 걸음 분석(first-step analysis); 비추이성(non-transitivity)
12.9	공정한 동전	치우침에서 공정함 끌어내기
12.10	무작위 고리	무기억적인 지시변수의 합
12.11	컵과 e	계승(factorial)과 지수급수
12.12	최적 정지(optimal stopping)	$\frac{1}{e}$ 규칙
12.13	삼자 결투(truel)	전략적 약함
12.14	게임 이론(game theory)	무작위화 전략; 최소최대(minimax)
12.15	분기 과정	고정점(fixed point)으로서의 소멸
12.16	마르코프 연쇄	정상분포(stationary distribution); 시간 역전(reversal)
12.17	바나흐(Banach)의 성냥갑	음이항분포(negative binomial distribution)
12.18	지수분포(exponential distribution)	무기억성(memorylessness)과 독립
12.19	공간적 무작위성	최근접 이웃(nearest neighbour); 레일리(Rayleigh) 분포
12.20	뷔퐁(Buffon)의 바늘	기하학적 확률과 π
12.21	무작위 삼각형	넓이로서의 확률
12.22	코시(Cauchy) 분포	평균이 없는 분포
12.23	스털링(Stirling) 공식	정규곡선에서 얻는 계승
12.24	중심극한정리	적률생성함수(moment generating function)
12.25	극값(extreme value)	검벨 분포(Gumbel distribution)

Section 12.24 한 절은 적률생성함수를 사용하는데, 이 내용은 2019년에 STEP 출제 범위에서 빠졌다. 그 절에서는 필요한 내용을 직접 설명한다.

이 장을 읽는 방법

확률 문제가 헷갈릴 때에는 작은 경우에 대해 표본공간(sample space)을 적어 보고 세어 보라. 답이 수상할 만큼 간단하게 나오면 그것을 설명해 주는 대칭을 찾아보라. 그리고 직관과 계산이 어긋나면 계산을 믿고, 그다음에 직관이 왜 틀렸는지 알아내라. 바로 거기에서 배움이 일어난다.

12.1 조건을 정확하게 걸기

문제

STEP I 2007, 13번 문제

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

`step.maths.org/questions/07-s1-q13`

도입

확률(probability)에서 가장 위험한 말은 " 라는 조건에서"이다. 이 문제의 마지막 네 소문항은 거의 똑같이 들린다. 3번 원판이 처음에 뽑힌다, 3번 원판이 뽑힌다, 번호가 적힌 원판이 처음에 뽑힌다, 번호가 적힌 원판이 뽑힌다. 이 가운데 셋은 답이 $\frac{1}{2}$이다. 나머지 하나는 그렇지 않다.

이 차이는 유명한 "아들 딸" 퍼즐(어느 가족에 아이가 둘 있고, 적어도 한 명은 아들이다. 둘 다 아들일 확률은 얼마인가?)에 숨어 있는 바로 그 차이이다. 이 퍼즐은 여러 세대의 학생들은 물론이고 적지 않은 전문 수학자들까지 헷갈리게 했다. 여기서 얻을 교훈은, 조건(condition)이란 어떤 결과들이 여전히 가능한지에 대한 정확한 진술이며, 정보를 어떻게 얻었는지가 정보 그 자체만큼이나 중요하다는 것이다.

시작하기 전에

생각해 보기 원판 열한 개 가운데 네 개를 뽑는다. "네 개 모두 번호가 적혀 있다"는 사건에 뽑는 순서가 중요한가? 똑같이 가능한 네 원판의 집합은 몇 개인가?

네 원판으로 이루어진 모든 집합은 뽑힐 가능성이 똑같으므로 순서는 무시해도 된다. 똑같이 가능한 집합은 $\binom{11}{4} = 330$개이다.

차근차근 풀기

1단계 조건이 없을 때

번호가 적힌 원판 네 개로 이루어진 집합은 $\binom{5}{4} = 5$개이므로

$$\mathrm{P}(\text{네 개 모두 번호가 적혀 있음}) = \frac{5}{330} = \frac{1}{66}.$$

2단계 3이 처음에 나올 때

생각해 보기 3번 원판이 처음에 뽑혔다면, 나머지 세 번의 뽑기는 어떻게 되는가?

나머지 세 원판은 남은 열 개(번호가 적힌 것 네 개, 빈 것 여섯 개)에서 무작위로 고른 세 개의 집합이고, $\binom{10}{3} = 120$개의 집합이 모두 똑같이 가능하다. 따라서

$$\text{(ii)}\quad \frac{\binom{4}{3}}{120} = \frac{4}{120} = \frac{1}{30}, \qquad \text{(iii)}\quad \frac{\binom{4}{1}\binom{6}{2}}{120} = \frac{4 \times 15}{120} = \frac{1}{2},$$

이다. (iii)에서는 나머지 세 개 가운데 정확히 하나에 번호가 적혀 있어야 하기 때문이다.

3단계 3이 (어느 시점에) 뽑힐 때

생각해 보기 3번 원판이 언제 뽑혔는지가 중요한가? 네 개 가운데 3이 들어 있다는 조건에서, 나머지 세 개에 대해 무엇을 말할 수 있는가?

대칭성(symmetry)에 의해, 네 원판의 집합이 3을 포함한다는 조건에서 나머지 세 원판은 2단계와 똑같이 남은 열 개에서 무작위로 고른 세 개의 집합이다. 따라서 (iv)도 $\frac{1}{2}$이다.

4단계 번호가 적힌 원판이 처음에 나올 때

첫 번째 원판에 번호가 적혀 있다는 조건에서(그 번호가 무엇이든), 나머지 세 개는 역시 남은 열 개의 원판에서 무작위로 고른 세 개의 집합이며, 그 열 개 가운데 네 개는 번호가 적혀 있고 여섯 개는 비어 있다. 따라서 (v)도 $\frac{1}{2}$이다.

5단계 번호가 적힌 원판이 뽑힐 때

생각해 보기 이제 조건은 "네 개 가운데 적어도 하나에 번호가 적혀 있다"이다. 이것은 앞의 조건들 가운데 어느 하나와 같은가? 330개의 집합 가운데 몇 개가 이 조건을 만족하는가?

이 조건은 다르다. 특정한 원판 하나를 지목하지 않기 때문이다. 번호가 적힌 원판이 하나도 없는 집합은 $\binom{6}{4} = 15$개뿐이므로, 315개가 조건을 만족한다. 이 가운데 번호가 적힌 원판을

정확히 두 개 포함하는 집합은 $\binom{5}{2}\binom{6}{2} = 10 \times 15 = 150$개이다. 따라서

$$\text{(vi)} \quad \frac{150}{315} = \frac{10}{21}. \qquad \blacksquare$$

> **잠깐, 직접 해 보기** (vi)은 왜 (v)와 다른가? (v)에서는 **어느** 원판에 번호가 적혀 있는지(첫 번째 원판) 알고, 나머지 세 개에는 아무런 제약이 없다. (vi)에서는 **어떤** 원판에 번호가 적혀 있다는 것만 안다. 번호가 적힌 원판이 세 개나 네 개인 집합은 "어떤 원판에 번호가 적혀 있다"를 더 여러 방식으로 만족하므로, (v)에서처럼 적게 반영되지 않는다.

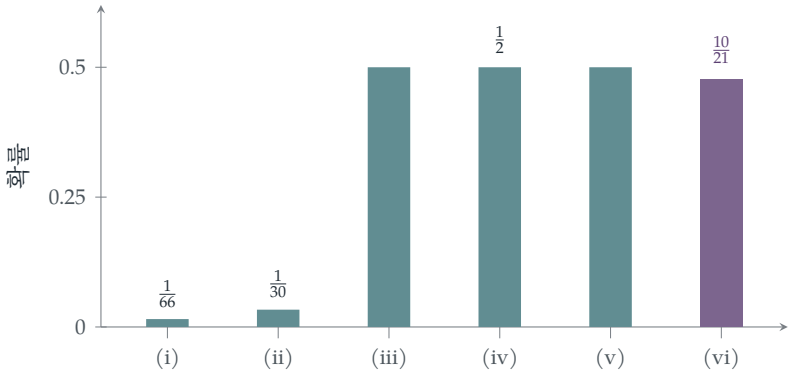


그림 12.1.1 여섯 개의 답. 소문항 (iii), (iv), (v)는 모두 **특정한** 원판에 번호가 적혀 있다는 조건을 걸며, 답은 $\frac{1}{2}$이다. 소문항 (vi)은 **어떤** 원판에 번호가 적혀 있다는 조건만 걸며, 답은 $\frac{10}{21}$이다.

돌아보기

- **조건은 표본공간(sample space)을 제한한다.** $\mathrm{P}(A \mid B)$는 B에 속하는 결과들 가운데 A에도 속하는 결과의 비율이다.
- **대칭성은 수고를 덜어 준다.** 3이 처음에 뽑혔는지 어느 시점에 뽑혔는지는 아무런 차이를 만들지 않았다. 모든 위치가 대등하기 때문이다.
- **"특정한 하나"는 "적어도 하나"가 아니다.** 비슷하게 들리는 조건도 상당히 다른 결과의 집합을 고를 수 있다.

숨은 수학

아들 딸 퍼즐

마틴 가드너(Martin Gardner)는 1959년에 다음 문제를 냈다. 스미스 씨에게는 아이가 둘 있고, 그 가운데 적어도 한 명은 아들이다. 둘 다 아들일 확률은 얼마인가? 똑같이 가능한 가족 구성 BB, BG, GB, GG를 나열하고 GG를 지우면 세 가지가 남고, 그 가운데 하나가 BB이다. 따라서 답은 $\frac{1}{3}$이다. 그러나 **첫째** 아이가 아들이라는 것을 안다면 답은 $\frac{1}{2}$이다. 이것은 우리 문제의 소문항 (vi)과 (v)의 축소판이다. 가드너는 나중에, 처음 낸 그대로의 문제가 모호했다고 인정했다. 답은 아들이 있다는 사실을 우리가 어떻게 알게 되었는지에 달려 있다. 아이들 가운데 한 명을 무작위로 만났는데 그 아이가 아들이었다면 답은 $\frac{1}{2}$이고, "아들이 적어도 한 명 있습니까?"라고 물어서 그렇다는 대답을 들었다면 답은 $\frac{1}{3}$이다. 확률은 정보 자체뿐 아니라 그 정보를 만들어 낸 과정에 따라 달라진다.

몬티 홀

확률에서 가장 많은 논쟁을 불러일으킨 퍼즐에도 같은 교훈이 깔려 있다. 어느 게임 쇼에서 세 개의 문 가운데 하나 뒤에 자동차가 있다. 당신은 1번 문을 고른다. 자동차가 어디 있는지 아는 진행자는 항상 나머지 두 문 가운데 하나를 열어 염소를 보여 주고, 항상 선택을 바꿀 기회를 준다. 자동차가 당신이 고른 문 뒤에 있어서 진행자가 고를 수 있는 경우에는, 나머지 두 문 가운데 하나를 무작위로 고른다. 진행자가 3번 문을 열고, 2번 문으로 바꾸겠느냐고 묻는다. 바꿔야 할까? 그렇다. 바꾸면 확률 $\frac{2}{3}$으로 이긴다. 진행자가 문을 고르는 것은 무작위가 아니다. 그것은 자동차가 어디 있는지에 따라 달라지며, 따라서 정보를 담고 있다. 정확한 규칙이 중요하다. 진행자가 가능할 때마다 항상 3번 문을 연다면, 3번 문이 열린 것을 보았을 때 1번 문과 2번 문은 똑같이 가능해지고, 바꾸어서 이길 확률은 $\frac{1}{2}$에 불과하다. 그런데도 미리 항상 바꾸기로 정한 참가자는 여전히 $\frac{2}{3}$의 비율로 이긴다. 바꾸면 처음 선택이 틀렸을 때 정확히 이기기 때문이다. 1990년에 메릴린 보스 사반트(Marilyn vos Savant)가 잡지 칼럼에서 이 답을 내놓았을 때, 그녀는 틀렸다는 편지를 수천 통 받았고, 그 가운데에는 수학자들이 보낸 것도 있었다. 컴퓨터 시뮬레이션을 해 보거나 똑같이 가능한 경우를 꼼꼼히 나열해 보면 논란은 끝난다.

베이즈와 라플라스

조건부 확률(conditional probability)을 체계적인 토대 위에 올려놓은 사람은, 사후인 1763년에 논문이 출판된 토머스 베이즈(Thomas Bayes)와, 1774년에 오늘날 베이즈 정리(Bayes' theorem)라고 부르는 법칙을 일반적인 형태로 서술한 피에르시몽 라플라스(Pierre-Simon Laplace)이다. 오늘날 모든 대학 강의의 출발점인 안드레이 콜모고로프(Andrey Kolmogorov)의 1933년 공리(axioms)는 $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$로 정의하며, 확률론의 나머지 부분은 이 정의 위에 세워진다. 다음 절 Section 12.2에서는 베이즈 정리가 실제로 어떤 일을 하는지 보여 준다.

사고 기술

조건을 정확히 하라

조건부 확률을 계산하기 전에, 어떤 결과들이 조건을 만족하는지 정확히 적어 보라. 그 정보가 어떻게 얻어졌는지 물어보라.

대학에서 조건부 확률과 조건부 기댓값(conditional expectation)은 확률론과 통계학에서 가장 중요한 도구에 속하며, 동시에 가장 흔한 오류의 원천이기도 하다. 적어도 작은 경우에 대해서라도 표본공간을 직접 적어 보는 것이 가장 좋은 방어책이다.

더 도전하기

1 **일반화하기.** 빈 원판이 b개, 번호가 적힌 원판이 n개 있고 k개를 뽑을 때, (iv)와 (v)의 답이 항상 같음을 보이고, (vi)이 그것들과 같아지는 때를 구하여라.

2 **화요일에 태어난 아들.** 어느 가족에 아이가 둘 있고, 적어도 한 명은 화요일에 태어난 아들이다. 둘 다 아들일 확률이 $\frac{13}{27}$임을 보여라. (똑같이 가능한 14×14가지 경우를 나열하거나, 영리하게 세어 보라.)

3 **몬티 홀.** 결과를 나열하여 $\frac{2}{3}$을 확인하여라. 진행자가 자동차가 어디 있는지 모른 채 나머지 문 가운데 하나를 무작위로 열었는데, 우연히 염소가 나왔다면 어떻게 되는가?

4 **에이스 두 장.** 카드 13장으로 된 브리지 패에 에이스가 적어도 한 장 들어 있다. 에이스가 적어도 두 장 들어 있을 확률은 얼마인가? 스페이드 에이스가 들어 있다는 조건에서의 확률과 비교하여라.

더 읽을거리

- **Jason Rosenhouse, *The Monty Hall Problem*** (Oxford University Press). 이 퍼즐과 그 역사, 그리고 수많은 변형을 다룬다.
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University Press), 조건부 확률을 다룬 장들.
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability*** (CRC Press; 온라인 무료 공개). 조건부 확률을 훌륭하게 다루며, 역설도 많이 실려 있다.

12.2 기저율 오류

문제

STEP II 1998, 12번 문제

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

`step.maths.org/questions/98-s2-q12`

도입

어떤 질병 검사가 그 병에 걸린 사람에 대해서는 90% 정확하고, 잘못된 경보는 거의 내지 않는다고 하자. 프랭키가 양성 판정을 받았다. 프랭키는 얼마나 걱정해야 할까? 대부분의 사람들은, 그리고 여러 연구에 따르면 대부분의 의사들도, "약 90%"라고 답한다. 이 문제에서 참된 답은 $\frac{1}{11}$, 즉 약 9%이다. 열 배나 작다.

이런 실수는 **기저율**(*base rate*)을 무시하는 데서 나온다. 이 병은 드물기 때문에, 수많은 건강한 사람들 사이에서는 위양성(false positive) 비율이 작더라도 실제 환자 수보다 훨씬 많은 잘못된 경보가 나온다. 이러한 **기저율 오류**(*base-rate fallacy*)는 의학, 법률(법률에서는 검사의 오류(prosecutor's fallacy)라고 부른다), 보안 검색, 기계 학습에서 중요한 문제이다. 베이즈 정리(Bayes' theorem)가 그 처방이다.

시작하기 전에

? **생각해 보기** 1000명의 사람을 떠올려 보자. 그중 몇 명이 병에 걸렸는가? 그중 몇 명이 양성 판정을 받는가? 건강한 사람 중에서는 몇 명이 양성 판정을 받는가?

1000명 가운데 10명이 병에 걸렸고, 그중 9명이 양성 판정을 받는다. 건강한 990명 가운데 $\frac{1}{11}$이 양성 판정을 받으므로 위양성은 90명이다. 따라서 양성 판정을 받은 사람은 99명이고, 그중 실제로 아픈 사람은 9명뿐이다.

차근차근 풀기

1단계 베이즈 정리

생각해 보기 "프랭키가 병에 걸렸다"를 D, "AL 검사 결과가 양성이다"를 $+$로 나타내자. 우리는 $\mathrm{P}(+ \mid D)$를 알고 있고 $\mathrm{P}(D \mid +)$를 구하려 한다. 둘은 어떤 관계인가?

조건부확률(conditional probability)의 정의에 따라 $\mathrm{P}(D \mid +) = \frac{\mathrm{P}(D \cap +)}{\mathrm{P}(+)}$이고, $\mathrm{P}(+)$는 프랭키가 병에 걸렸는지 아닌지에 따라 나누어진다.

$$\begin{aligned}\mathrm{P}(D \mid +) &= \frac{\mathrm{P}(+ \mid D)\mathrm{P}(D)}{\mathrm{P}(+ \mid D)\mathrm{P}(D) + \mathrm{P}(+ \mid \bar{D})\mathrm{P}(\bar{D})} \\ &= \frac{0.9 \times 0.01}{0.9 \times 0.01 + \frac{1}{11} \times 0.99} = \frac{0.009}{0.009 + 0.09} = \frac{1}{11}.\end{aligned}$$

■

이것이 바로 천 명의 사람에서 얻은 $\frac{9}{99}$이다.

2단계 두 번째 검사

생각해 보기 두 번째 검사는 AL에서 양성 판정을 받은 사람들에게만 실시한다. 그 집단에서 환자의 비율은 얼마인가? 그것을 새로운 "기저율"로 사용하라.

AL에서 양성인 사람들 가운데 $\frac{1}{11}$은 환자이고 $\frac{10}{11}$은 환자가 아니다. 이것을 사전확률(prior probability)로 삼아 베이즈 정리를 다시 적용하면

$$\mathrm{P}(D \mid \text{두 검사 모두 양성}) = \frac{0.9 \times \frac{1}{11}}{0.9 \times \frac{1}{11} + p \times \frac{10}{11}} = \frac{0.9}{0.9 + 10p}.$$

이 값을 $\frac{1}{2}$로 놓으면 $10p = 0.9$이므로 $p = 0.09$이다. ■

(문제에서 암묵적으로 가정하듯이, 프랭키가 아픈지 아닌지가 주어졌을 때 두 검사의 오류는 서로 독립(independent)이라고 가정한다.)

잠깐, 직접 해 보기 $p = 0.09$이면 STEP 검사의 위양성 비율은 AL과 거의 같은데도, 두 번째 양성 결과는 확률을 $\frac{1}{11}$에서 $\frac{1}{2}$로 끌어올린다. 똑같은 수준의 검사가 두 번째에는 왜 훨씬 큰 효과를 낼까?



그림 12.2.1 전형적인 백 명의 사람. 한 명은 병에 걸렸고(자주색), 열 번 중 아홉 번은 양성 판정을 받는다. 건강한 99명 가운데 약 9명은 잘못하여 양성 판정을 받는다(호박색). 양성 판정을 받은 사람들 가운데 건강한 사람이 아픈 사람보다 대략 열 배 많다.

돌아보기

- **자연 빈도(natural frequencies).** 1000명의 사람을 떠올리자 답이 분명해졌고 확인하기도 쉬웠다.
- **사전확률 곱하기 가능도.** 사후확률(posterior probability)은 검사의 정확도와 기저율 둘 다에 달려 있다.
- **단계별 갱신.** 어제의 사후확률이 오늘의 사전확률이다. 두 번째 검사는 $\frac{1}{100}$이 아니라 $\frac{1}{11}$에서 출발했다.

숨은 수학

오즈와 증거

베이즈 정리는 **오즈**(*odds*)로 나타낼 때 가장 명료하다. D의 오즈는 $\mathrm{P}(D) : \mathrm{P}(\bar{D})$이다. 하나의 증거 E는 오즈에 **가능도비**(*likelihood ratio*) $\frac{\mathrm{P}(E|D)}{\mathrm{P}(E|\bar{D})}$를 곱한다.

$$\text{사후 오즈} = \text{사전 오즈} \times \text{가능도비}.$$

여기서 사전 오즈는 $1 : 99$이고 AL의 가능도비는 $0.9 \div \frac{1}{11} = 9.9$이므로, 사후 오즈는 $9.9 : 99 = 1 : 10$, 즉 확률로는 $\frac{1}{11}$이다. 가능도비가 10인 두 번째 독립적인 검사($p = 0.09$인 STEP 검사가 그렇다)는 오즈를 $1 : 1$로 만든다. 로그(logarithm)를 취하면 증거는 **더해진**다. 즉 각각의 독립적인 검사는 자신의 가능도비의 로그만큼 기여한다. 앨런 튜링(Alan Turing)은 바로 이 방법을 사용해 증거를 "밴(ban)"과 "데시밴(deciban)" 단위로 측정하며 블레츨리 파크(Bletchley Park)에서 에니그마 암호문을 해독했다.

검사의 오류

1999년, 사무 변호사(solicitor) Sally Clark는 두 아기 아들을 살해한 혐의로 유죄 판결을 받았는데, 그 근거의 일부는 한 가정에서 영아 돌연사(cot death)가 두 번 일어날 확률이

"7300만분의 1"이라는 증거였다. 그 수치 자체의 오류(두 죽음이 독립이라고 가정하여 유전적·환경적 요인을 무시했다)는 제쳐 두더라도, 그 추론은 P(증거 | 무죄)와 P(무죄 | 증거)를 혼동했다. 어머니가 두 아이를 살해하는 일 역시 극히 드물며, 따져야 할 것은 두 설명 사이의 비교이다. 영국 왕립통계학회(Royal Statistical Society)는 공개 서한을 보내 이 오류를 지적했고, 그녀에 대한 유죄 판결은 2003년에 뒤집혔다. 그 수학은 바로 이 문제의 수학이다.

기거렌처와 자연 빈도

1990년대에 심리학자 게르트 기거렌처(Gerd Gigerenzer)는 검사 통계를 조건부확률의 형태로 제시받은 의사들이 이 문제와 같은 질문에서 대부분 크게 틀렸지만, 같은 정보를 자연 빈도의 형태("여성 1000명 가운데 10명이 유방암에 걸렸고, 그중 8명이 양성 판정을 받는다...")로 제시받으면 대부분 맞힌다는 것을 보였다. 그의 연구는 여러 나라에서 의학적 위험을 환자에게 전달하는 방식을 바꾸어 놓았다.

사고 기술

베이즈 정리로 조건을 뒤집기

$P(A \mid B)$와 $P(B \mid A)$는 서로 다른 수이다. 한쪽에서 다른 쪽으로 가려면 기저율 $P(A)$가 필요하다.

베이즈 추론, 즉 증거에 비추어 믿음을 갱신하는 것은 스팸 필터에서 의학적 진단에 이르기까지 현대 통계학과 기계 학습의 토대이다. 대학에서는 하나의 확률 대신 사전분포와 사후분포를 다루는 베이즈 추론(Bayesian inference)으로 이것을 만나게 된다.

더 도전하기

1 **냉장고 속 시체.** STEP II 1987, 16번 문제 (`step.maths.org/questions/87-s2-q16`)에 도전해 보자. 신뢰도가 서로 다른 두 목격자가 등장하는 베이즈 정리 문제로, STEP 문제 가운데서도 손꼽히게 기억에 남는 설정이다.

2 **얼마나 드물어야 할까?** 민감도(sensitivity)가 0.9이고 위양성 비율이 $\frac{1}{11}$인 검사에서, 양성 결과가 병에 걸렸을 가능성이 절반보다 크다는 뜻이 되려면 그 병은 얼마나 흔해야 하는가?

3 **검사를 반복하기.** AL 검사 자체를 반복하고, 그 사람의 실제 상태가 주어졌을 때 결과들이 서로 독립이라면, AL에서 두 번 양성 결과가 나온 뒤 병에 걸렸을 확률은 얼마인가? 세 번이라면?

4 **선별 검사.** 어떤 선별 검사 프로그램이 유병률(prevalence)이 10 000명 중 1명인 질환에 대해 백만 명을 검사한다. 사용하는 검사의 민감도는 99%이고 위양성 비율은 1%이다. 그 질환이 있을 수 있다는 통보를 받는 사람은 몇 명이며, 그중 실제로 질환이 있는 사람은 몇 명인가?

더 읽을거리

- **Gerd Gigerenzer, *Reckoning with Risk*** (Penguin). 자연 빈도, 의학 검사, 그리고 법정.
- **David Spiegelhalter, *The Art of Statistics*** (Pelican). 의학적 선별 검사와 Sally Clark 사건을 다룬다.
- **Sharon Bertsch McGrayne, *The Theory That Would Not Die*** (Yale University Press). 튜링의 활용을 포함한 베이즈 정리의 역사.

12.3 "무작위"란 무슨 뜻인가?

문제

STEP III 2022, 12번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2022년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:

`www.ocr.org.uk/Images/696587-2022-past-papers.zip`

도입

"원의 현(chord)을 무작위로 하나 고른다. 그 길이의 기댓값(expected value)은 얼마인가?" 이 질문에는 답이 있을 것처럼 들린다. 그러나 "무작위로"가 무슨 뜻인지 정하기 전까지는 답이 없다. 원 위의 두 점을 무작위로 골라 이으면 길이의 기댓값은 $\frac{4a}{\pi} \approx 1.27a$이다. 원 내부의 한 점과 그 점을 지나는 방향을 무작위로 고르면 답은 $\frac{16a}{3\pi} \approx 1.70a$이다. 두 방법 모두 "무작위 현"을 고르는 지극히 타당한 방법이다.

이것이 **베르트랑의 역설**(*Bertrand's paradox*)로, 조제프 베르트랑(Joseph Bertrand)이 1889년에 일종의 경고로서 발표한 것이다. 확률은 무작위 시행이 구체적으로 정해진 뒤에야 비로소 잘 정의된다. 이 역설은 수학자들이 20세기에 확률론의 엄밀한 기초를 세우는 데 한몫했는데, 이 기초에서는 모든 확률이 정확하게 정의된 결과들의 공간 위에서 다루어진다.

시작하기 전에

생각해 보기 각 AOB가 θ일 때 현 AB의 길이는 얼마인가? (O에서 수선을 내려 보자.)

O에서 내린 수선은 현과 각을 모두 이등분하므로 $AB = 2a\sin\frac{\theta}{2}$이다.

차근차근 풀기

1단계 원 위의 두 점

$$\mathrm{E}(AB) = \frac{1}{2\pi}\int_0^{2\pi} 2a\sin\frac{\theta}{2}\,\mathrm{d}\theta = \frac{a}{\pi}\Big[-2\cos\frac{\theta}{2}\Big]_0^{2\pi} = \frac{a}{\pi}\cdot 4 = \frac{4a}{\pi}. \qquad \blacksquare$$

2단계 내부의 무작위 점

> **생각해 보기** $0 \le r \le a$일 때 $\mathrm{P}(R \le r)$는 얼마인가? 이를 미분하여 확률밀도를 구하라.

$\mathrm{P}(R \le r)$는 반지름이 r인 원판의 넓이를 전체 원판의 넓이로 나눈 값, 즉 $\frac{r^2}{a^2}$이다. 따라서 R는 $[0,a]$에서 확률밀도(probability density) $\frac{2r}{a^2}$를 가진다. (중심에서 먼 곳일수록 넓이가 더 크므로, 점은 중심에서 멀리 있을 가능성이 더 크다.)

> **생각해 보기** C를 지나는 현이 OC와 각 t를 이룬다. 이 현은 O에서 얼마나 떨어져 있는가?

O에서 이 직선까지의 거리는 $R\sin t$이므로 현의 길이는 $2\sqrt{a^2 - R^2\sin^2 t}$이다.

3단계 C에 대한 평균

$s = \sin t$라 하면

$$L(t) = \int_0^a 2\sqrt{a^2 - r^2 s^2}\cdot\frac{2r}{a^2}\,\mathrm{d}r = \frac{4}{a^2}\Big[-\frac{(a^2 - r^2s^2)^{3/2}}{3s^2}\Big]_0^a$$
$$= \frac{4}{a^2}\cdot\frac{a^3 - a^3(1-s^2)^{3/2}}{3s^2} = \frac{4a(1-\cos^3 t)}{3\sin^2 t}. \qquad \blacksquare$$

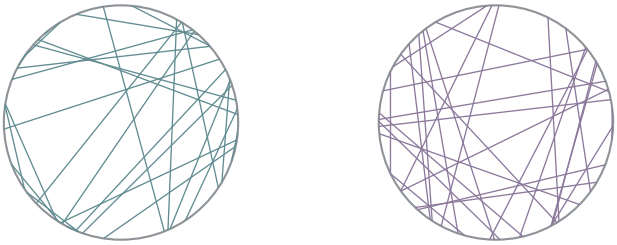
> **생각해 보기** $1-\cos^3 t$와 $\sin^2 t = 1-\cos^2 t$를 인수분해하라. 무엇이 약분되는가?

$c = \cos t$라 하면 $\frac{1-c^3}{1-c^2} = \frac{(1-c)(1+c+c^2)}{(1-c)(1+c)} = \frac{1+c+c^2}{1+c} = c + \frac{1}{1+c}$이고, $1+\cos t = 2\cos^2\frac{t}{2}$이다. 따라서

$$L(t) = \frac{4a}{3}\Big(\cos t + \tfrac{1}{2}\sec^2\tfrac{t}{2}\Big). \qquad \blacksquare$$

4단계 방향도 무작위로

$$\mathrm{E}\big(L(T)\big) = \frac{2}{\pi}\int_0^{\pi/2} \frac{4a}{3}\Big(\cos t + \tfrac{1}{2}\sec^2\tfrac{t}{2}\Big)\,\mathrm{d}t$$
$$= \frac{8a}{3\pi}\Big[\sin t + \tan\tfrac{t}{2}\Big]_0^{\pi/2} = \frac{8a}{3\pi}(1+1) = \frac{16a}{3\pi}. \qquad \blacksquare$$



무작위 양 끝점
평균 길이 $\frac{4a}{\pi} \approx 1.27a$

원판 내부의 무작위 중점
평균 길이 $\frac{4a}{3} \approx 1.33a$

그림 12.3.1 서로 다른 두 방법으로 고른 28개의 "무작위 현". 왼쪽: 원 위에서 독립적으로 균등하게 고른 두 점. 오른쪽: 원판에서 균등하게 무작위로 고른 점을 현의 중점으로 삼은 경우. 오른쪽 방법에서는 가장자리 근처의 짧은 현이 더 적게 나온다. 방법이 다르면 분포가 다르고, 답도 다르다.

돌아보기

- **시행을 구체적으로 정하라.** 각 소문항에서 "무작위 현"은 정확한 절차가 되었고, 절차마다 다른 답이 나왔다.
- **넓이에서 확률밀도로.** $\mathrm{P}(R \le r)$는 기하에서 곧바로 나왔고, 이를 미분하여 확률밀도를 얻었다.
- **단계별로 평균하라.** 먼저 (t를 고정하고) C의 위치에 대해 평균하고, 그다음 t에 대해 평균했다.

숨은 수학

베르트랑의 세 가지 답

베르트랑이 원래 던진 질문은 다음과 같았다. 무작위 현이 원에 내접하는 정삼각형(inscribed equilateral triangle)의 한 변 $\sqrt{3}a$보다 길 확률은 얼마인가? 그는 세 가지 방법과 세 가지 답을 제시했다.

- 무작위 양 끝점: 한 끝점을 고정하면, 다른 끝점이 원의 가운데 3분의 1 부분에 있을 때 현이 길다. 확률은 $\frac{1}{3}$이다.

- 무작위 반지름과 그 위의 무작위 점을 중점으로: 현이 O로부터 $\frac{a}{2}$보다 가까이 있으면 현이 길다. 확률은 $\frac{1}{2}$이다.
- 원판 전체에서 고른 무작위 중점: 중점이 O로부터 $\frac{a}{2}$ 이내에 있으면 현이 길다. 확률은 $\frac{1}{4}$이다.

각 방법은 현들의 집합 위에 주어진 서로 다른 확률분포이며, 여기에 모순은 없다. 세 방법은 서로 다른 질문에 답하고 있을 뿐이다. 교훈은 "균등하게 무작위로"라는 말이 어떤 매개화(parametrisation)를 택했는지에 상대적으로만 의미가 있으며, 연속적인 집합에서는 대개 유일하게 자연스러운 매개화가 존재하지 않는다는 것이다.

정답이 있을까?

1973년 물리학자 Edwin Jaynes는 원을 조금 옮기거나 크기를 조금 바꾸어도 답이 변하지 않아야 한다는 근거로 두 번째 방법이 "올바른" 방법이라고 주장했다. 무작위 현이란 바닥에 그린 원 위로 빨대를 던지는 것과 같아야 한다는 것이다. 중심으로부터의 거리가 균등한 분포만이 이러한 불변성(invariance)을 가지며, 빨대를 던지는 실험도 $\frac{1}{2}$를 확인해 준다. 올바른 분포란 상황의 대칭성을 존중하는 분포라는 같은 생각이 무작위 직선의 수학인 **적분기하학**(*integral geometry*)의 바탕에 깔려 있으며, 이것은 Section 12.20의 뷔퐁(Buffon)의 바늘 문제도 설명해 준다.

역설에서 공리로

베르트랑은 *Calcul des probabilités*(1889)에서 현의 역설을 발표했는데, 부분적으로는 그가 보기에 부주의한 "기하학적 확률"의 사용을 비꼬기 위해서였다. 푸앵카레(Henri Poincaré)도 이 역설을 논의했다. 1900년 힐베르트(David Hilbert)는 새 세기를 위한 유명한 문제 목록에 확률론의 공리화(axiomatisation)를 포함시켰고, 콜모고로프(Andrey Kolmogorov)가 1933년에 이를 해결했다. 확률은 결과들의 공간 위의 측도이며, "무작위" 대상에 관한 모든 질문은 먼저 어떤 공간과 어떤 측도를 쓰는지 밝혀야 한다. 베르트랑의 역설은 왜 이것이 필요한지를 보여 주는 대표적인 예이다.

사고 기술

무작위 시행을 정의하라

확률을 계산하기 전에 무작위 대상이 정확히 어떻게 생성되는지 기술하라. 기술 방식이 다르면 답도 달라질 수 있다.

대학에서는 모든 확률 계산이 표본공간(sample space)과 확률측도(probability measure)에서 출발한다. 통계학과 물리학에서는 흔히 대칭성을 이용해 올바른 무작위성 모형을 고르는 일이 작업의 큰 부분을 차지한다.

더 도전하기

1 **베르트랑의 확률.** 위의 세 확률 $\frac{1}{3}$, $\frac{1}{2}$, $\frac{1}{4}$를 확인하라.

2 **또 하나의 평균.** 현의 중점을 원판에서 균등하게 고르면(오른쪽 그림) 길이의 기댓값이 $\frac{4a}{3}$임을 보여라. (**힌트:** 이것은 $L(\frac{\pi}{2})$이다. 왜 그런가?)

3 **무작위 거리.** O에서 현까지의 거리가 $[0, a]$에서 균등분포를 따르면 길이의 기댓값이 $\frac{\pi a}{2}$임을 보여라.

4 **세 방법, 세 분포.** 세 방법 각각에 대해 $[0, 2a]$에서 현의 길이 ℓ의 확률밀도를 구하라.

더 읽을거리

- **Edwin Jaynes, "The well-posed problem"**, *Foundations of Physics* 3 (1973). 불변성 논증을 담은 글로, 읽기 쉽고 도발적이다.
- **Gábor Székely, *Paradoxes in Probability Theory and Mathematical Statistics*** (Reidel). 베르트랑의 역설을 비롯한 많은 역설을 다룬다.
- **Daniel Klain & Gian-Carlo Rota, *Introduction to Geometric Probability*** (Cambridge University Press). 적분기하학을 다룬 책으로, 나중에 읽어 볼 만하다.

12.4 둘씩은 독립이지만 한꺼번에는 독립이 아닌

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`

도입

1번 참가자와 2번 참가자가 같은 수를 굴렸다. 이 사실이 2번과 3번 참가자가 같은 수를 굴리는지에 대해 무언가를 알려 줄까? 놀랍게도 그렇지 않다. 두 사건은 서로 독립(independent)이다. 1번과 3번 참가자에 대해서도 마찬가지이다. 그런데 1번이 2번과 같고 2번이 3번과 같다면, 1번은 반드시 3번과 같다. 세 사건 가운데 어느 두 사건도 서로 독립이지만, 세 사건 전체는 독립이 아닌 것이다.

쌍별 독립(*pairwise independence*)은 완전한 독립보다 약한 조건이며, 이 차이는 중요하다. 분산(variance)을 다룰 때에는 쌍별 독립만으로 충분하고, 이 문제는 바로 이 점을 이용한다. 그러나 다른 많은 목적에는 쌍별 독립으로는 부족하다. 컴퓨터 과학자들은 일부러 쌍별 독립인 난수를 사용하기도 하는데, 완전히 독립인 난수보다 훨씬 적은 비용으로 만들 수 있으면서도 많은 알고리즘에는 그것으로 충분하기 때문이다.

시작하기 전에

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

1번 참가자가 무엇을 굴리든, 2번 참가자가 그와 같은 수를 굴릴 확률(probability)은 $\frac{1}{6}$이다. $X_{12}=1$과 $X_{23}=1$이 동시에 성립한다는 것은 세 사람 모두 같은 수를 굴렸다는 뜻이므로, 그 확률은 $\frac{6}{216}=\frac{1}{36}$이다.

차근차근 풀기

1단계 쌍별 독립

$\mathrm{P}(X_{12}=1, X_{23}=1) = \frac{1}{36} = \frac{1}{6}\cdot\frac{1}{6} = \mathrm{P}(X_{12}=1)\mathrm{P}(X_{23}=1)$이고 각 확률변수는 두 값만 가지므로, X_{12}와 X_{23}은 독립이다. (두 값만 가지는 확률변수에서는 결합 분포표(joint table)의 한 칸이 나머지 칸을 모두 결정한다. 연습문제 1에서 이를 엄밀하게 다룬다.) X_{12}와 X_{34}처럼 공통인 참가자가 없는 쌍은 당연히 독립이다. 따라서 모든 X_{ij}는 쌍별 독립이다.

2단계 점수의 평균과 분산

점수는 $S = \sum_{i<j} X_{ij}$로, $\binom{n}{2}$개의 확률변수의 합이며, 각 확률변수의 평균(mean)은 $\frac{1}{6}$, 분산은 $\frac{1}{6}\cdot\frac{5}{6} = \frac{5}{36}$이다. 기댓값(expectation)은 언제나 더해지므로 $\mathrm{E}(S) = \frac{1}{6}\binom{n}{2}$이다.

> **생각해 보기** 합의 분산은 언제 분산의 합과 같은가?

합의 분산은 분산의 합에 공분산(covariance)의 합의 두 배를 더한 것이고, 독립인 쌍의 공분산은 0이다. 따라서 쌍별 독립이면 충분하다.

$$\mathrm{E}(S) = \frac{n(n-1)}{12}, \qquad \mathrm{Var}(S) = \frac{5}{36}\binom{n}{2} = \frac{5n(n-1)}{72}.$$

3단계 분산 공식

Y_i의 평균이 모두 0이면 $\mathrm{Var}(\sum Y_i) = \mathrm{E}((\sum Y_i)^2)$이고, 제곱을 전개하면 각 Y_i^2이 한 번씩, 각 곱 Y_iY_j $(i<j)$가 두 번씩 나온다. 여기에 기댓값을 취하면 공식을 얻는다. ■

4단계 홀짝 게임

> **생각해 보기** $\mathrm{E}(Z_{12})$는 얼마인가? $Z_{12}Z_{23}$은 어떤 값을 가질 수 있으며, 언제 0이 아닌가?

$Z_{12}=1$일 확률은 $\frac{3}{36} = \frac{1}{12}$이고(둘 다 2, 둘 다 4, 또는 둘 다 6을 굴리는 경우), -1일 확률도 $\frac{1}{12}$이다. 따라서 $\mathrm{E}(Z_{12}) = 0$이고 $\mathrm{E}(Z_{12}^2) = \frac{1}{6}$이다.

$Z_{12}Z_{23} \neq 0$인 것은 세 참가자가 모두 같은 수를 굴릴 때뿐이며, 이때 $Z_{12} = Z_{23}$(둘 다 $+1$이거나 둘 다 -1)이므로 곱은 1이다. 따라서 $\mathrm{E}(Z_{12}Z_{23}) = \frac{1}{36}$이고, 이는 $\mathrm{E}(Z_{12})\mathrm{E}(Z_{23}) = 0$과 다르다. 즉 두 확률변수는 독립이 아니다. ■

(더 직접적으로는, 각 사건의 확률은 $\frac{1}{12}$인데 $\mathrm{P}(Z_{12}=1, Z_{23}=-1) = 0$이다.)

5단계 분산

총점 $T = \sum Z_{ij}$의 평균은 0이다. 소문항 (ii)에 의해 $\mathrm{Var}(T)$는 $\binom{n}{2} \cdot \frac{1}{6}$에 쌍들의 쌍에 대한 $\mathrm{E}(Z_{ij}Z_{kl})$의 합의 두 배를 더한 것이다. 공통인 참가자가 없는 쌍들의 쌍은 0을 기여한다(서로 독립이고 각각의 평균이 0이기 때문이다). 정확히 한 명의 참가자를 공유하는 쌍들의 쌍은 각각 $\frac{1}{36}$을 기여하며, 그러한 쌍들의 쌍은 $n\binom{n-1}{2}$개 있다(공유하는 참가자를 고른 뒤 나머지 두 명을 고른다). 따라서

$$\begin{aligned}\mathrm{Var}(T) &= \frac{n(n-1)}{12} + 2 \cdot \frac{1}{36} \cdot \frac{n(n-1)(n-2)}{2} \\ &= \frac{n(n-1)}{36}\bigl(3 + (n-2)\bigr) = \frac{n(n^2-1)}{36}. \qquad \blacksquare\end{aligned}$$

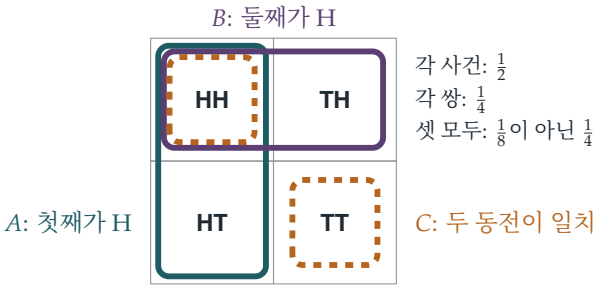


그림 12.4.1 베른시테인의 예: 공정한 동전 두 개. 사건 "첫째 동전이 앞면"(청록색), "둘째 동전이 앞면"(자주색), "두 동전이 일치"(호박색, 점선)의 확률은 각각 $\frac{1}{2}$이고, 어느 두 사건도 확률 $\frac{1}{4}$인 한 칸에서만 겹치므로 쌍별 독립이다. 그러나 세 사건은 모두 HH에서 겹치고, 그 확률은 $\frac{1}{8}$이 아니라 $\frac{1}{4}$이다. 어느 두 사건이 나머지 한 사건을 결정하기 때문이다.

돌아보기

- **지시변수.** 점수를 지시변수(indicator variable)의 합으로 나타내자 모든 문제가 참가자 쌍에 대한 문제로 바뀌었다.
- **분산에는 쌍만 있으면 된다.** 합의 분산은 쌍별 공분산에만 의존하므로, 분산을 더하는 데에는 쌍별 독립으로 충분하다.
- **독립성은 가정하지 말고 확인하라.** 겉보기에 비슷한 두 게임이 서로 다른 종속 구조를 가지고 있었다.

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 12.4.1의 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.

사고 기술

지시변수로 분해하기

어떤 개수의 평균과 분산을 구하려면, 셀 수 있는 대상마다 하나씩 지시변수를 두어 그 개수를 지시변수의 합으로 나타내라. 평균에는 지시변수 하나하나가 필요하고, 분산에는 쌍이 필요하다.

이것은 확률론에서 가장 유용한 기법 가운데 하나이며, 이미 Section 9.8에서도 등장했다. 대학에서는 무작위 그래프, 알고리즘, 통계적 검정의 분석에서 핵심 도구로 쓰인다.

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.

12.5 마지막 승객의 좌석

문제

STEP II 2021, 11번 문제

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

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

도입

승객 백 명이 만석인 비행기에 탑승한다. 첫 번째 승객은 탑승권을 잃어버려서 아무 자리에나 앉는다. 그 뒤의 승객은 각자 자기 자리가 비어 있으면 그 자리에 앉고, 그렇지 않으면 비어 있는 자리 중 하나를 무작위로 골라 앉는다. 마지막 승객이 자기 자리에 앉을 확률(probability)은 얼마일까? 혼란이 비행기 전체로 연쇄적으로 번질 수 있으니 확률이 작으리라 예상할 수도 있다. 아니면 대부분의 승객은 영향을 받지 않으니 1에 가까워야 한다고 생각할 수도 있다.

답은 승객 수와 상관없이 정확히 $\frac{1}{2}$이다. 이 문제는 점화식(recurrence)과 수학적 귀납법(induction)으로 이를 증명한다. 그런데 대칭(symmetry)을 이용한 아름다운 한 줄짜리 논증도 있다. 이 논증은 답이 왜 $\frac{1}{2}$인지 설명해 줄 뿐 아니라, 어느 승객에 대해서든 답을 곧바로 적을 수 있게 해 준다.

시작하기 전에

잠깐, 직접 해 보기 일어날 수 있는 경우를 모두 나열하여 P_2와 P_3을 구하여라.

좌석이 두 개이면 T_1은 S_1에 앉거나(그러면 T_2는 S_2에 앉는다) S_2에 앉으며, 각각의 확률은 $\frac{1}{2}$이다. 따라서 $P_2 = \frac{1}{2}$이다. 좌석이 세 개일 때, T_1이 S_1에 앉으면(확률 $\frac{1}{3}$) 모두 문제가 없다. T_1이 S_3에 앉으면 T_3은 자기 자리를 잃는다. T_1이 S_2에 앉으면 T_2는 S_1과 S_3 중 하나를 무작위로 고르고, T_3은 T_2가 S_1을 고를 때만 자기 자리에 앉는다. 따라서 $P_3 = \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}$이다.

차근차근 풀기

1단계 점화식

생각해 보기 T_1이 S_k에 앉는다고 하자. 단, $2 \le k \le n-1$이다. 승객 $T_2, \dots, T_{k-1}$에게는 어떤 일이 일어나는가? T_k가 도착했을 때 어느 좌석이 비어 있으며, 이 상황은 더 작은 열차의 경우와 어떻게 비교되는가?

승객 $T_2, \dots, T_{k-1}$은 자기 자리가 비어 있으므로 그 자리에 앉는다. T_k가 도착하면 빈 좌석은 S_1과 $S_{k+1}, \dots, S_n$이며, 모두 $n-k+1$개이다. T_k는 이 중 하나를 무작위로 고르고, 그 뒤로 각 승객은 자기 자리가 비어 있으면 그 자리에, 아니면 무작위로 고른 자리에 앉는다. 이는 좌석이 $n-k+1$개인 원래 문제와 정확히 같다. 여기서 T_k가 첫 번째 승객의 역할을 하고, S_1이 T_k의 "자기" 좌석(누군가 앉으면 모든 것이 정리되는 좌석)의 역할을 한다. 따라서 T_n이 S_n에 앉을 조건부 확률(conditional probability)은 P_{n-k+1}이다.

이제 각각 확률 $\frac{1}{n}$인 T_1의 좌석에 대하여 조건을 걸자. S_1이면 반드시 성공하고, S_n이면 반드시 실패하며, S_k이면 성공 확률이 P_{n-k+1}이다. 따라서

$$P_n = \frac{1}{n}\left(1 + 0 + \sum_{k=2}^{n-1} P_{n-k+1}\right) = \frac{1}{n}\left(1 + \sum_{r=2}^{n-1} P_r\right). \qquad \blacksquare$$

2단계 답

생각해 보기 $P_2 = P_3 = \frac{1}{2}$이다. P_n을 추측하고 강한 귀납법(strong induction)으로 증명하여라.

주장: 모든 $n \ge 2$에 대하여 $P_n = \frac{1}{2}$이다. $n = 2$일 때 성립한다. $2 \le r \le n-1$에 대하여 $P_r = \frac{1}{2}$이면

$$P_n = \frac{1}{n}\left(1 + \frac{n-2}{2}\right) = \frac{1}{n} \cdot \frac{n}{2} = \frac{1}{2}. \qquad \blacksquare$$

3단계 끝에서 두 번째 승객

생각해 보기 더 간결한 논증을 찾아보아라. 마지막에 T_n에게 남는 좌석은 어느 것인가? S_2일 수 있을까?

마지막으로 T_n에게 남는 좌석은 S_1 아니면 S_n이다. $2 \le k \le n-1$인 S_k일 수는 없는데, T_k가 그 자리에 앉았을 것이기 때문이다. 그리고 그 전에 이루어지는 모든 무작위 선택은 S_1과 S_n을 대칭적으로 다룬다. 자리를 빼앗긴 승객이 좌석을 고를 때마다 S_1과 S_n이 선택될 가능성은 같다. 두 좌석 중 먼저 선택되는 쪽이 결과를 결정하므로, 각 좌석이 남는 좌석이 될 가능성은 같다. 따라서 $P_n = \frac{1}{2}$이다.

같은 논증은 $k \geq 2$인 임의의 승객 T_k에 대해서도 통한다. T_k는 앞선 어떤 승객이 자기 자리에 앉은 경우에만 자리를 잃는다. 자리 빼앗기는 누군가가 S_1을 고를 때까지 계속된다. 좌석 $S_1, S_k, S_{k+1}, \ldots, S_n$은 T_k가 도착하기 전에 좌석을 고르는 승객들에게 모두 똑같이 취급되므로, 이 중 가장 먼저 선택되는 좌석은 $n-k+2$개의 좌석 중 어느 것일 가능성이나 같고, T_k는 그것이 S_k일 때만 자리를 잃는다. 따라서

$$\mathrm{P}(T_k\text{가 앉는 좌석이}S_k) = \frac{n-k+1}{n-k+2}.$$

$k = n-1$일 때($n \geq 3$) 이로부터 $Q_n = \frac{2}{3}$을 얻는다. 또 $Q_2 = \frac{1}{2}$인데, 이때는 $T_{n-1} = T_1$이고 이 승객은 확률 $\frac{1}{2}$로 S_1에 앉기 때문이다. ■

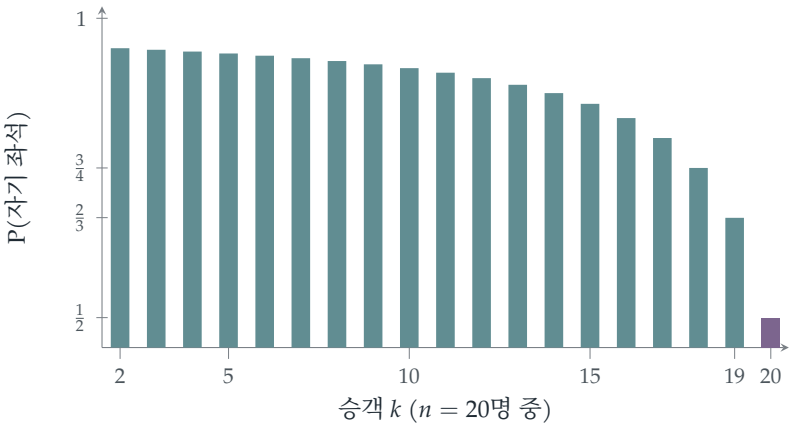


그림 12.5.1 $n = 20$인 열차에서 승객 k가 자기 자리에 앉을 확률. $2 \leq k \leq n$일 때 이 확률은 $\frac{n-k+1}{n-k+2}$이며, T_2의 $\frac{19}{20}$에서 T_{19}의 $\frac{2}{3}$까지 줄어들고, 마지막 승객은 $\frac{1}{2}$이다(자주색).

돌아보기

- **더 작은 복사본으로 줄이기.** T_1이 S_k에 앉은 뒤, 나머지 문제는 좌석 수가 더 적은 같은 문제였다.
- **강한 귀납법.** 점화식에 더 작은 경우가 모두 등장했으므로, 귀납 가정에서 그 모든 경우에 대하여 결과를 가정했다.
- **대칭 찾기.** 좌석 S_1과 S_n은 처음부터 끝까지 서로 바꿔도 차이가 없으며, 이 사실이 아무런 계산 없이 $\frac{1}{2}$를 설명해 준다.

숨은 수학

대칭 논증

두 번째 증명은 확률론의 강력한 아이디어의 한 예이다. 두 결과가 무작위 과정의 대칭으로 서로 연결되어 있으면 두 결과의 확률은 같다. 여기서의 과정은 S_1과 S_n 중 하나가 선택될 때까지 둘을 구별하지 않으므로, 둘이 끝까지 남을 가능성은 같아야 한다. 대칭 논증은 긴 계산을 단 하나의 관찰로 대신하는 경우가 많다. 다음 몇 절에도 이런 논증이 가득하다. Section 12.6의 동전 한 번 더 던지기, Section 12.7의 토너먼트 대진 추첨이 그렇다.

자리를 빼앗기는 승객은 몇 명인가?

자리를 빼앗기는 승객(자기 자리에 앉지 못하는 승객)의 수도 깔끔하게 다룰 수 있다. T_1은 S_1을 고르지 않는 한 자기 자리에 앉지 못하고, $T_k (k \geq 2)$는 확률 $\frac{1}{n-k+2}$로 자리를 빼앗긴다. 이를 더하면 자리를 빼앗기는 승객 수의 기댓값(expected value)은

$$\left(1 - \frac{1}{n}\right) + \left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2}\right) = H_n - \frac{1}{n},$$

으로, 대략 $\ln n$이다. 300명이 탄 비행기에서도 혼란은 평균적으로 겨우 여섯 명 정도에게만 영향을 미친다. 여기서도 조화수(harmonic number)가 등장한다(Section 9.11).

사고 기술

대칭을 찾아라

무작위 과정이 두 결과를 구별하지 못한다면, 두 결과의 가능성은 같다. 두 결과가 처음으로 구별되는 순간을 찾아라.

대학에서는 반사 원리(reflection principle, Section 9.10)에서부터 교환가능성(exchangeability)에 이르기까지 확률론 전반에 대칭 논증이 등장한다. 교환가능성은 확률변수(random variable)의 수열이 어떤 순서로 나열하든 같은 분포(distribution)를 갖는 성질로, 현대 통계학의 상당 부분의 토대가 된다.

더 도전하기

1 **건망증 있는 두 승객.** 처음 두 승객이 모두 무작위로 자리를 고른다면, 마지막 승객이 자기 자리에 앉을 확률은 얼마인가?

2 **자리를 빼앗기는 승객.** $n = 3$일 때 가능한 경우를 모두 나열하여 자리를 빼앗기는 승객 수의 기댓값을 확인하여라.

3 **우회로.** 점화식을 이용하여 $\sum_{r=2}^{n-1} P_r = \frac{n-2}{2}$가 대칭 논증과 모순되지 않음을 보여라.

4 **마지막 좌석.** 마지막 승객이 자기 자리에 앉지 못했다는 조건 아래에서, 마지막 승객이 S_1에 앉을 확률이 1임을 보여라. 왜 그런가?

더 읽을거리

- **Peter Winkler, *Mathematical Puzzles: A Connoisseur's Collection*** (A K Peters). 수학 퍼즐 모음집. 비행기 좌석 퍼즐을 비롯하여 대칭으로 풀리는 여러 퍼즐이 실려 있다.
- **Frederick Mosteller, *Fifty Challenging Problems in Probability*** (Dover). 확률 문제 50선. 고전적인 문제 모음집이다.
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability*** (CRC Press; 온라인 무료 공개). 확률론 입문서.

12.6 동전 하나 더

문제

STEP I 2016, 12번 문제

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

`step.maths.org/questions/16-s1-q12`

도입

밥은 앨리스보다 동전을 하나 더 던진다. 밥이 앞면을 더 많이 얻을 확률(probability)은 얼마일까? 밥이 유리하기는 하지만 앞면이 보장되는 것은 아니며, 비기는 경우는 밥에게 불리하게 처리된다. 처음 두 경우를 일일이 계산해 보면 두 번 모두 답이 $\frac{1}{2}$이다. 이것은 우연일까?

우연이 아니다. 밥은 앨리스보다 앞면을 더 많이 얻거나 뒷면을 더 많이 얻되 둘 다는 아니라는 사실을 알아차리기만 하면, 한 문장으로 끝나는 증명(proof)이 있다. 이 문제는 계산이 어떻게 패턴을 암시하고, 교묘한 대칭(symmetry)이 그 패턴을 어떻게 설명하는지를 보여 주는 모범적인 예이다.

시작하기 전에

생각해 보기 앨리스가 동전을 두 번 던질 때, 앨리스가 얻는 앞면의 개수의 분포(distribution)는 무엇인가? 밥이 세 번 던질 때는 어떠한가?

앨리스는 앞면을 $0, 1, 2$개 얻을 확률이 각각 $\frac{1}{4}, \frac{2}{4}, \frac{1}{4}$이고, 밥은 $0, 1, 2, 3$개 얻을 확률이 각각 $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$이다.

차근차근 풀기

1단계 직접 계산하기

소문항 (i)에서는 앨리스의 앞면 개수에 대하여 모두 더한다.

$$\begin{aligned}\mathrm{P}(B > A) &= \tfrac{1}{4}\mathrm{P}(B \geq 1) + \tfrac{2}{4}\mathrm{P}(B \geq 2) + \tfrac{1}{4}\mathrm{P}(B = 3) \\ &= \tfrac{1}{4}\cdot\tfrac{7}{8} + \tfrac{1}{2}\cdot\tfrac{4}{8} + \tfrac{1}{4}\cdot\tfrac{1}{8} = \frac{7+8+1}{32} = \frac{1}{2}.\end{aligned}$$

소문항 (ii)는 항이 더 많을 뿐 같은 계산이다. 같은 정도로 일어날 수 있는 $8 \times 16 = 128$가지 결과를 세어 보면 그중 64가지에서 밥이 앞서므로(그림 12.6.1), 이번에도 답은 $\frac{1}{2}$이다.

2단계 밥의 던지기를 나누기

생각해 보기 밥의 $n+1$번 던지기를 n번 던진 다음 한 번을 더 던지는 것으로 생각해 보자. 처음 n번을 던진 뒤 밥은 앨리스보다 앞서 있거나, 같거나, 뒤처져 있다. 각 경우에 어떤 일이 일어나는가?

둘 다 n번씩 던진 뒤, 밥이 앞서 있을 확률은 p_2, 같을 확률은 p_1, 뒤처져 있을 확률은 (앨리스와 밥 사이의 대칭에 의해) p_2이다. 밥이 앞서 있다면 계속 앞선다. 같다면 추가로 던진 동전으로 확률 $\frac{1}{2}$로 앞서게 된다. 뒤처져 있다면 앞면이 하나 더 나와도 기껏해야 같아질 뿐이다. 따라서

$$\mathrm{P}(\text{밥이 앞면을 더 많이 얻음}) = p_2 + \tfrac{1}{2}p_1.$$

세 경우가 모든 경우를 포괄하므로 $2p_2 + p_1 = 1$이고, 따라서 $p_2 + \frac{1}{2}p_1 = \frac{1}{2}$이다. 동전이 하나 더 있으면, 모든 n에 대하여 밥이 앞면을 더 많이 얻을 확률은 정확히 $\frac{1}{2}$이다. ■

잠깐, 직접 해 보기 한 문장 증명은 다음과 같다. 밥은 $n+1$번 던지므로 앞면과 뒷면의 개수의 합이 $n+1$로, 앨리스보다 하나 많다. 밥이 앨리스보다 앞면을 더 많이 얻는 것은 밥이 앨리스보다 뒷면을 더 많이 얻지 **않는** 것과 필요충분조건임을 보여라. 앞면과 뒷면은 대칭이므로 이 두 사건의 확률은 같고, 따라서 각각 $\frac{1}{2}$이다.

실제로 $H_A + T_A = n$, $H_B + T_B = n+1$이라 쓰자. $H_B > H_A$이면 $H_B \geq H_A + 1$이므로 $T_B = n+1-H_B \leq n - H_A = T_A$이다. 역으로 $T_B \leq T_A$이면 $H_B \geq H_A + 1$이다. 따라서 "앞면이 더 많음"과 "뒷면이 더 많음"은 서로 여사건(complementary event)이고, 앞면과 뒷면을 맞바꾸면(동전이 공정하므로 확률은 변하지 않는다) 두 사건이 서로 맞바뀐다.



그림 12.6.1 소문항 (ii): 같은 정도로 일어날 수 있는 128가지 결과 중 각 앞면 개수 조합에 해당하는 결과의 수. 밥이 앞서는 칸은 색칠되어 있으며, 정확히 전체 결과의 절반을 담고 있다.

돌아보기

- **작은 경우를 계산하라.** 두 번의 계산이 답 $\frac{1}{2}$을 암시했다.
- **자연스러운 분할에 대해 조건을 걸어라(conditioning).** 밥의 추가 던지기를 따로 떼어 내자 문제가 대등한 승부에 동전 하나를 더한 것으로 줄어들었다.
- **서로 여사건인 쌍을 찾아라.** "앞면이 더 많음"과 "뒷면이 더 많음"은 서로 여사건이며, 대칭에 의해 서로 맞바뀐다.

숨은 수학

대칭과 여사건

이 깔끔한 증명은 두 가지 사실을 결합한다. 대칭에 의해 서로 맞바뀌는 두 사건은 확률이 같고, 서로 여사건인 두 사건은 확률의 합이 1이다. 두 조건이 모두 성립하면 각 사건의 확률은 $\frac{1}{2}$이다. 같은 추론이 확률론 곳곳에 등장한다. 비행기 좌석 문제(Section 12.5)에서 S_1과 S_n은 대칭에 의해 맞바뀌었고, 마지막에 남는 것은 정확히 둘 중 하나였다. 헥스(Hex) 게임에서는 판의 대칭성과 반드시 정확히 한 명이 이긴다는 사실로부터 선수가 유리함을 보일 수 있지만, 큰 판에서의 필승 전략은 아무도 모른다.

비길 확률은 얼마나 될까?

던지는 횟수가 같을 때는 비길 확률 p_1이 중요하다. 앨리스와 밥은 앨리스의 앞면 수가 밥의 앞면 수와 같을 때 비기는데, 이는 (대칭에 의해 밥의 앞면 수를 뒷면 수로 바꾸면) 앨리스의

앞면 수와 밥의 앞면 수의 합이 n일 확률, 즉 $2n$번 던져서 앞면이 모두 n개 나올 확률과 같다. 따라서 스털링(Stirling) 근사(Section 12.23)를 이용하면

$$p_1 = \binom{2n}{n}\frac{1}{4^n} \approx \frac{1}{\sqrt{\pi n}},$$

이다. 비기는 일은 드물어지지만 그 속도는 느리다. 각자 100번씩 던져도 비길 확률은 여전히 약 5.6%이다.

사고 기술

> **계산하고, 추측하고, 그다음 설명하라**
>
> 계산 결과가 의심스러울 만큼 단순한 답을 내면, 그 이유를 찾아라. 단순한 답에는 대개 단순한 설명이 있다.
>
> 대학에서는 이 습관이 이해와 계산을 가른다. 여기서 첫 번째 증명은 답을 확인해 주고, 두 번째 증명은 답을 설명하며 곧바로 일반화된다(연습문제 2를 풀어 보라).

더 도전하기

1. **동전 두 개 더.** 밥이 동전을 $n+2$개, 앨리스가 n개 던질 때, H'을 밥의 처음 $n+1$번 던지기에서 나온 앞면의 개수라 하면 $\mathrm{P}(\text{밥이 앞면을 더 많이 얻음}) = \frac{1}{2} + \frac{1}{2}\mathrm{P}(H' = H_A)$임을 증명하여라.

2. **편향된 동전.** 두 사람 모두 앞면이 나올 확률이 p인 동전을 사용한다면, 한 번 더 던질 때 답은 여전히 $\frac{1}{2}$인가? 깔끔한 증명의 어느 단계가 성립하지 않는가?

3. **비기기.** 본문의 요령을 이용하여 $p_1 = \binom{2n}{n}/4^n$임을 보이고, $n = 1, 2, 3$일 때 그 값을 계산하여라.

4. **주사위.** 앨리스는 주사위 n개를, 밥은 $n+1$개를 던진다. 밥이 6의 눈을 더 많이 얻을 확률은 $\frac{1}{2}$인가?

더 읽을거리

- **Peter Winkler, *Mathematical Mind-Benders*** (A K Peters). 대칭을 이용한 풀이가 있는 확률 퍼즐이 많이 실려 있다.
- **Gabor Székely, *Paradoxes in Probability Theory and Mathematical Statistics*** (Reidel). 확

률론과 수리통계학의 역설.

- **Frederick Mosteller, *Fifty Challenging Problems in Probability*** (Dover). 확률의 도전적인 문제 50선.

12.7 토너먼트에서 누가 누구와 만나는가

문제

STEP II Specimen paper, 16번 문제

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

`step.maths.org/questions/spec-s2-q16`

도입

윔블던(Wimbledon) 대회에는 128명의 선수가 대진표에 오른다. 두 선수를 무작위로 고를 때, 이들이 언젠가 맞붙을 확률(probability)은 얼마일까? 각 라운드에서 만날 확률을 모두 더해 볼 수도 있다. 각각은 살아남을 확률과 서로 대진이 짜일 확률의 곱이다. 이 방법도 통하지만 훨씬 좋은 방법이 있다. 경기 수를 세는 것이다. 모든 경기는 두 선수 사이에서 치러지고, 경기는 정확히 127번 열린다. 대칭성(symmetry)에 의해 모든 선수 쌍은 각 경기의 대진이 될 가능성이 똑같다. 따라서 구하는 확률은 127을 선수 쌍의 개수로 나눈 값이다.

이 문제는 올바른 대상을 세어서 확률 문제를 푸는 완벽한 예이다. 답 $\frac{1}{2^{n-1}}$은 단 한 줄로 나온다.

시작하기 전에

생각해 보기 2^n명의 선수가 참가하는 녹아웃 토너먼트(knockout tournament)에서는 몇 경기가 열리는가? (각 경기에서 선수 한 명이 탈락한다.)

각 경기에서 정확히 한 명이 탈락하고, 한 명을 제외한 모든 선수가 탈락하므로 경기는 $2^n - 1$번 열린다.

차근차근 풀기

1단계 첫 라운드

고른 두 선수를 X와 Y라 하자. 첫 라운드에서 X의 상대는 나머지 2^n-1명 중 누구든 될 가능성이 같으므로

$$\mathrm{P}(\text{만나는 라운드가}1) = \frac{1}{2^n-1}.$$

2단계 핵심이 되는 대칭성

> **생각해 보기** 각 라운드의 대진 추첨은 무작위이고 모든 경기는 공정한 동전 던지기이다. 이 과정이 어떤 선수 쌍을 다른 쌍과 다르게 취급하는가?

그렇지 않다. 대진 추첨이 무작위이고 모든 경기가 50–50이므로, 선수들의 이름을 바꾸어 붙여도 어떤 확률도 달라지지 않는다. 따라서 특정한 경기(예를 들어 결승전)가 $\binom{2^n}{2}$개의 쌍 중 어느 쌍 사이에서 치러질 가능성도 모두 같다.

3단계 결승전

결승전은 특정한 한 경기이므로, 그 경기가 X와 Y 사이에서 치러질 확률은

$$\frac{1}{\binom{2^n}{2}} = \frac{2}{2^n(2^n-1)}.$$

(검산: 두 선수 모두 결승에 올라야 하므로, 서로 다른 절반에 배치되고 각자 $n-1$경기를 이겨야 한다.)

4단계 어느 단계에서든

> **생각해 보기** X와 Y는 많아야 한 번 만날 수 있다. 이 사실이 어떻게 경기별 확률을 더할 수 있게 해 주는가?

경기 m이 X와 Y 사이의 경기이면 1인 값을 I_m이라 하자. 2^n-1개의 경기 각각에 대해 $\mathrm{P}(I_m=1) = \binom{2^n}{2}^{-1}$이고, 이 사건들은 서로 배반(mutually exclusive)이므로(두 선수는 두 번 만날 수 없다)

$$\mathrm{P}(\text{만남}) = \frac{2^n-1}{\binom{2^n}{2}} = \frac{2(2^n-1)}{2^n(2^n-1)} = \frac{1}{2^{n-1}}. \qquad \blacksquare$$

$n=7$인 윔블던의 경우, 주어진 두 선수가 만날 확률은 $\frac{1}{64}$이다. 검산으로, 첫 라운드에는 2^n-1개의 경기 중 2^{n-1}개가 들어 있으므로 $\frac{2^{n-1}}{\binom{2^n}{2}} = \frac{1}{2^n-1}$이 되어 단계 1의 결과와 일치한다.

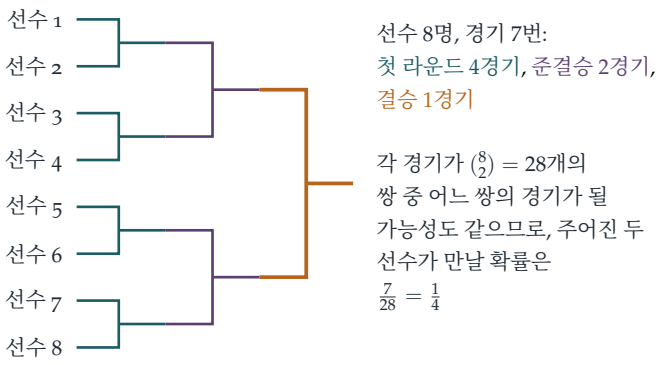


그림 12.7.1 $2^3 = 8$명의 선수가 참가하는 녹아웃 토너먼트(문제에서는 매 라운드 무작위로 대진을 다시 짜지만, 여기서는 고정된 대진표로 그렸다). 경기는 7번 열린다. 대칭성에 의해 각 경기가 28개의 쌍 중 어느 쌍의 경기가 될 가능성도 같으므로, 고른 두 선수가 만날 확률은 $\frac{7}{28} = \frac{1}{4} = \frac{1}{2^{n-1}}$이다.

돌아보기

- **경기 수를 센다.** 경기 수 $2^n - 1$이 핵심 사실이었다.
- **쌍 사이의 대칭성.** 모든 선수 쌍은 특정한 경기를 치를 가능성이 같다.
- **경기에 대한 선형성.** 두 선수는 많아야 한 번 만나므로, 만날 확률은 만나는 횟수의 기댓값이며 이는 경기에 대한 합이다.

숨은 수학

개수의 기댓값

단계 4의 논증은 사실 기댓값(expectation) 계산이다. X와 Y가 만나는 횟수는 경기마다 하나씩인 지시변수(indicator)들의 합이고, 그 기댓값(이 횟수는 0 또는 1이므로 만날 확률과 같다)은 지시변수들의 기댓값의 합이다. 이것이 기댓값의 선형성(linearity of expectation)이며, Section 9.11와 Section 12.4에서도 사용되었다. 이 방법의 큰 장점은 라운드 사이의 복잡한 종속 관계를 전혀 신경 쓰지 않아도 된다는 것이다.

시드가 배정된 토너먼트

실제 토너먼트에는 시드(*seed*)가 배정된다. 가장 강한 선수들은 후반 라운드가 되기 전에는 서로 만나지 않도록 배치된다. 또 실제 경기는 50–50이 아니다. 두 가지 모두 대칭성을 깨뜨리므로, 주어진 두 선수가 만날 확률을 구하는 일은 진짜 계산이 된다. 스포츠를 모형화하는 통계학자들은 선수들의 레이팅으로부터 각 경기의 확률을 추정하여 바로 이런 계산으로 토

너먼트 결과를 예측한다. 대칭적인 경우는 실제 효과를 측정하는 기준선이 된다.

사고 기술

올바른 대상을 세어라

올바른 대상을 고르기만 하면, 확률은 흔히 (유리한 대상의 개수) ÷ (가능성이 같은 대상의 개수)로 구할 수 있다. 여기서 그 대상은 경기와 쌍이었다.

무엇을 셀지 고르는 것이 조합적 확률(combinatorial probability)의 기술이다. 대학에서는 같은 아이디어가 **이중 세기**(*double counting*)(Section 7.9)로, 그리고 어떤 구조의 개수의 기댓값으로 그 구조의 존재를 증명(proof)하는 확률적 방법(probabilistic method)으로 나타난다.

더 도전하기

1 **라운드별로.** X와 Y가 라운드 k에서 만날 확률이 $\frac{2^{n-k}}{\binom{2^n}{2}}$임을 직접 보이고, 이를 모두 더하면 $\frac{1}{2^{n-1}}$이 됨을 확인하여라.

2 **고정된 대진표.** 대신 대진표를 (그림처럼) 미리 고정해 두고 모든 선수를 그 안에 무작위로 배치한다면 답이 달라지는가?

3 **강한 선수.** 선수 X가 모든 경기에서 이긴다고 하자. 이제 X가 Y를 만날 확률은 얼마인가?

4 **리그.** 모든 쌍이 한 번씩 맞붙는 $2n$개 팀의 리그(league)에서는 몇 경기가 열리는가? 이에 대응하는 "만날 확률"은 무엇인가?

더 읽을거리

- **Frederick Mosteller, *Fifty Challenging Problems in Probability*** (Dover). 토너먼트 문제가 실려 있다.
- **David Spiegelhalter, *The Art of Statistics*** (Pelican). 모형화와 예측에 관한 책이다.
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability*** (CRC Press; 온라인 무료 공개). 지시변수와 대칭성을 다룬다.

12.8 페니의 게임

문제

STEP III 2001, 13번 문제

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

`step.maths.org/questions/01-s3-q13`

도입

여러분과 친구가 각자 동전 세 번 던지기의 결과로 이루어진 수열을 하나씩 고르고, 둘 중 한 수열이 나타날 때까지 동전을 계속 던진다고 하자. 어느 순간에나 세 번 던지기의 모든 수열은 나타날 확률이 같으므로, 어떤 수열을 고르든 똑같이 유리하지 않을까? 전혀 그렇지 않다. 친구가 어떤 수열을 고르든, 여러분은 그것을 이기는 수열을 고를 수 있으며, 때로는 그 승산이 7 대 1에 이른다.

이것이 1969년 Walter Penney가 고안한 **페니 게임**(*Penney's game*)이다. 이 문제는 A가 B를 이기고, B가 C를 이기고, C가 D를 이기고, D가 A를 이긴다는 것을 보여 준다. 가위바위보처럼 순환이 생기는 것이다. "이긴다"는 관계는 추이적(transitive)이지 않다. 나중에 고르는 사람은 장기적으로 언제나 이길 수 있다.

시작하기 전에

생각해 보기 지금까지 던진 결과가 TT로 끝났고 아직 아무도 이기지 않았다고 하자. THH가 TTH보다 먼저 나타날 수 있을까?

TT에서 T가 더 나올 때마다 마지막 두 결과는 계속 TT이고, 처음 나오는 H가 TTH를 완성한다. THH가 나타나려면 TH가 필요한데, 이는 T 다음에 H가 나올 때만 생기며, 그 H는 이미 TTH를 완성한다. 따라서 일단 TT가 나오면 A가 이긴다.

차근차근 풀기

1단계 상태

생각해 보기 두 번 던져서는 어느 수열도 완성될 수 없다. 두 번 던진 뒤, 각 상태에서 다음 한 번의 던지기는 어떤 결과를 가져오는가?

게임은 마지막 두 결과에만 의존하므로, 상태(state)(마지막 두 결과)를 추적하면 된다.

- TT에서: A가 이기므로(위에서 보였다) $p_{\mathrm{TT}} = 1$이다.
- HT에서: 다음 던지기에서 TH 또는 TT가 되므로 $p_{\mathrm{HT}} = \frac{1}{2}p_{\mathrm{TH}} + \frac{1}{2}p_{\mathrm{TT}}$이다.
- TH에서: H가 나오면 THH가 완성되고(B가 이긴다), T가 나오면 HT가 된다. 따라서 $p_{\mathrm{TH}} = \frac{1}{2}\cdot 0 + \frac{1}{2}p_{\mathrm{HT}}$이다.
- HH에서: H가 나오면 다시 HH가 되고, T가 나오면 HT가 된다. 따라서 $p_{\mathrm{HH}} = \frac{1}{2}p_{\mathrm{HH}} + \frac{1}{2}p_{\mathrm{HT}}$이고, 이로부터 $p_{\mathrm{HH}} = p_{\mathrm{HT}}$를 얻는다.

2단계 풀기

$p_{\mathrm{TH}} = \frac{1}{2}p_{\mathrm{HT}}$를 HT에 대한 식에 대입하면 $p_{\mathrm{HT}} = \frac{1}{4}p_{\mathrm{HT}} + \frac{1}{2}$이므로 $p_{\mathrm{HT}} = \frac{2}{3}$, $p_{\mathrm{HH}} = \frac{2}{3}$, $p_{\mathrm{TH}} = \frac{1}{3}$이다. 네 가지 시작 상태는 각각 확률이 $\frac{1}{4}$이므로

$$\mathrm{P}(A\text{ 가 이김}) = \tfrac{1}{4}\left(\tfrac{2}{3} + \tfrac{2}{3} + \tfrac{1}{3} + 1\right) = \tfrac{2}{3}. \qquad \blacksquare$$

3단계 B 대 C

생각해 보기 C는 HHT를, B는 THH를 원한다. C는 언제 이기는가? (처음 나오는 HH 바로 앞에 무엇이 와야 하는지 살펴보라.)

처음 두 결과가 HH이면, 그 뒤 처음 나오는 T가 THH가 나타나기 전에 HHT를 완성하므로 C가 이긴다. 그렇지 않으면 HH가 처음 나타나기 전 어딘가에 T가 있으므로, 처음 나오는 HH 바로 앞에는 T가 있다. 그 순간 THH가 나타나며, 이는 HHT(HH 뒤에 T가 필요하다)보다 먼저이다. 따라서 C는 게임이 HH로 시작할 때만 이긴다.

$$\mathrm{P}(B\text{ 가 이김}) = \tfrac{3}{4}. \qquad \blacksquare$$

4단계 순환

같은 방법(또는 "숨은 수학" 절의 지름길)을 쓰면, C(HHT)가 D(HTT)를 확률 $\frac{2}{3}$으로 이기고, D(HTT)가 A(TTH)를 확률 $\frac{3}{4}$로 이긴다는 것을 알 수 있다. 마지막 것은 앞면과 뒷면을

바꾼 3단계의 거울상이다. 따라서

A 는 다음을 이긴다: B, B 는 다음을 이긴다: C, C 는 다음을 이긴다: D, D 는 다음을 이긴다:

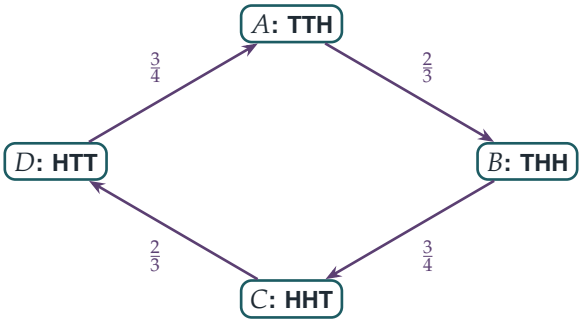


그림 12.8.1 페니의 순환. X에서 Y로 가는 화살표는 X의 수열이 Y의 수열을 표시된 확률로 이긴다는 뜻이다. 모든 수열은 다른 어떤 수열에게 지므로, 나중에 고르는 사람은 언제나 이기는 수열을 고를 수 있다.

돌아보기

- **알맞은 상태를 고른다.** 마지막 두 결과만 중요했으므로, 네 개의 상태와 네 개의 방정식으로 게임 전체를 나타낼 수 있었다.
- **첫걸음 분석.** 각 방정식은 다음 던지기를 조건으로 하였다. 한 상태에서의 확률은 다음 던지기가 이끄는 곳들에서의 확률의 평균이다.
- **결정적인 사건을 찾는다.** 3단계에서는 "처음 나오는 HH 앞에 T가 있다"는 사실이 방정식 없이 게임의 승부를 결정했다.

숨은 수학

어떤 수열이 더 유리한 이유

세 번 던지기의 모든 수열은 주어진 어느 위치에서나 나타날 확률이 같고, 각 수열이 나타날 때까지의 기대 대기 시간은 쉽게 계산할 수 있다. 그러나 대기 시간은 서로 다르다. TTH는 평균 8번, TTT는 평균 14번을 기다려야 하는데, 자기 자신과 겹치는 수열(TTT처럼, 아깝게 빗나간 TTH가 두 번 던지기만큼의 진행을 망쳐 버리는 경우)은 늦게 나타나기 때문이다. 그리고 두 수열의 경주에서 중요한 것은 각 수열이 **상대** 수열과 어떻게 겹치느냐이다. THH가 HHT를 이기는 까닭은, HHT가 나타날 때마다(맨 처음인 경우를 제외하고) 그 앞에 T가 있고, 이는 THH가 막 나타났다는 뜻이기 때문이다. 두 번째 선수의 비결은 언제나 같다. 상대 수열의 처음 두 기호를 가져와서, 그 앞에 상대 수열의 가운데 기호와 반대인 기

호를 붙이면 된다.

콘웨이의 공식

콘웨이(John Conway)는 승산에 대한 놀라운 공식을 발견했다. 길이가 n인 두 수열 X와 Y에 대하여 **상관**(*correlation*) XY를 n자리 이진수로 정의한다. 왼쪽부터 읽을 때, X의 i번째 기호에서 시작하는 부분이 같은 길이의 Y의 앞부분과 같으면 i번째 자리를 1로 둔다. 그러면 Y가 X보다 먼저 나타날 승산은

$$(XX - XY) : (YY - YX).$$

이다. $X = \text{HHT}$, $Y = \text{THH}$인 경우, $XX = 100_2 = 4$(HHT 전체만이 자기 자신과 일치한다), $XY = 001_2 = 1$(HHT의 마지막 T만이 THH의 앞부분과 일치한다), $YY = 100_2 = 4$, $YX = 011_2 = 3$(THH 끝의 HH와 H가 모두 HHT의 앞부분과 일치한다)이다. 따라서 THH가 먼저 나올 승산은 $(4-1) : (4-3) = 3 : 1$이고, 이것이 3단계의 $\frac{3}{4}$이다. 콘웨이의 공식은 길이와 알파벳에 관계없이 모든 수열에 대해 성립하며, 그 증명에는 **마팅게일**(*martingale*) 이론이 쓰인다. 도박꾼들이 수열이 나타나는 데 돈을 거는 카지노를 상상하고, 공정한 게임의 기대 이익은 0이라는 사실을 이용하는 것이다.

Penney와 가드너

Walter Penney는 1969년 *Journal of Recreational Mathematics*에 실린 짧은 글에서 이 게임을 소개했다. 마틴 가드너(Martin Gardner)는 1974년 *Scientific American*의 칼럼에서 이 게임을 널리 알렸는데, 여기에는 콘웨이가 발견했지만 발표하지 않았던 콘웨이의 공식도 포함되어 있었다. 이런 종류의 비추이적 관계(주사위 A가 B를, B가 C를, C가 D를, D가 A를 이기는 에프론(Efron)의 주사위도 그 예이다)는 거듭 놀라움을 주는 대상이며, 다수의 선호가 순환을 이룰 수 있는 투표 이론에서도 중요하다. 이것이 콩도르세(Condorcet)의 역설이다.

사고 기술

첫걸음 분석

상태 사이를 옮겨 다니는 과정에서 확률이나 기댓값을 구하려면 첫 단계를 조건으로 삼는다. 그러면 상태마다 일차방정식이 하나씩 생긴다.

이것은 마르코프 연쇄(Markov chain)의 기본 기법으로, Section 12.16에서, 그리고 대학 확률론 과정 전반에서 만나게 된다. 같은 방법으로 한 상태에 다른 상태보다 먼저 도달할 확률, 흡수될 때까지의 기대 시간, 도박 문제에서 파산할 확률을 구할 수

있다.

더 도전하기

1 **네 명이 함께.** STEP II 2015, 12번 문제 (`step.maths.org/questions/15-s2-q12`)에 도전해 보라. 네 명이 동시에 하는 페니 게임이다.

2 **대기 시간.** HH가 처음 나타날 때까지 던지는 횟수의 기댓값이 6이고, HT의 경우는 4임을 보여라. (첫걸음 분석을 이용하라.)

3 **최선의 응수.** 길이가 3인 여덟 개의 수열 각각에 대하여, 그것을 가장 큰 확률로 이기는 수열을 구하라. 어느 수열이 가장 쉽게 지는가?

4 **콘웨이의 공식.** 콘웨이의 공식을 이용하여 HHT가 HTT를 확률 $\frac{2}{3}$으로 이긴다는 것을 확인하라.

더 읽을거리

- **Martin Gardner, *Time Travel and Other Mathematical Bewilderments*** (W. H. Freeman). 페니 게임과 콘웨이의 공식이 실려 있다.
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics*** (Addison-Wesley), 8.4절. 생성함수를 이용한 페니 게임.
- **David Williams, *Probability with Martingales*** (Cambridge University Press). 마팅게일을 이용한 증명. 나중을 위한 책이다.

12.9 치우친 동전으로 공정하게 정하기

문제

STEP I 2019, 11번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2019년 STEP 시험지 묶음(zip 파일) 가운데 STEP I 시험지에 있다:

`www.ocr.org.uk/Images/696590-2019-past-papers.zip`

도입

공정한 결정을 내려야 하는데, 쓸 수 있는 동전은 휘어진 것 하나뿐이고 얼마나 치우쳐 있는지도 모른다. 그래도 공정하게 고를 수 있을까? 1951년 존 폰 노이만(John von Neumann)이 내놓은 답은 이렇다. 동전을 두 번 던진다. HT가 나오면 호텔을, TH가 나오면 찻집을 고르고, HH나 TT가 나오면 다시 두 번 던진다. HT와 TH는 둘 다 확률(probability)이 pq이므로, p가 무엇이든 이 결정은 완벽하게 공정하다.

이 문제의 절차는 그 반대이다. HH와 TT에서 결정하므로 빠르지만 공정하지 않다(호텔을 고를 확률은 $\frac{p^2}{p^2+q^2}$이다). 문제는 어떤 동전을 쓰더라도 이 절차가 결코 오래 걸리지 않음을 보여 준다. 두 절차를 비교하면 무작위성을 만들어 낼 때의 기본적인 맞교환, 곧 속도와 공정성 사이의 맞교환이 드러난다.

시작하기 전에

생각해 보기 한 라운드에서 결정이 내려질 확률은 얼마인가? 결정이 내려지지 않을 확률은? $2pq$가 가질 수 있는 가장 큰 값은 얼마인가?

동전들이 일치하면 결정이 내려지며, 그 확률은 $p^2 + q^2$이다. 그렇지 않으면 확률 $2pq$로 계속 진행한다. $pq = p(1-p) \leq \frac{1}{4}$이고(등호는 $p = \frac{1}{2}$일 때 성립한다), 따라서 $2pq \leq \frac{1}{2}$이다.

차근차근 풀기

1단계 n번째 라운드에서의 결정

각 라운드는 독립(independent)이다. n번째 라운드에서 결정이 내려지려면 $n-1$번의 라운드 동안 동전이 일치하지 않고, 그다음에 일치해야 한다.

$$\mathrm{P}(\text{결정이 일어나는 라운드} n) = (2pq)^{n-1}(p^2+q^2). \qquad \blacksquare$$

2단계 n번째 라운드까지의 결정

? **생각해 보기** 처음 n번의 라운드에서 결정이 **내려지지 않을** 확률은 얼마인가?

n번의 라운드 동안 결정이 내려지지 않을 확률은 $(2pq)^n \le \left(\frac{1}{2}\right)^n$이므로, p가 무엇이든 n번째 라운드까지 결정이 내려질 확률은 적어도 $1-\frac{1}{2^n}$이다. $\blacksquare$

3단계 세 사람

? **생각해 보기** 첫 라운드에서 앞면이 둘, 뒷면이 하나 나온 뒤, 세 번째 사람이 뒷면을 앞면으로 뒤집어 놓는다. 둘째 라운드에서 결정이 내려지려면 무엇이 일어나야 하는가? 뒷면이 둘, 앞면이 하나 나온 경우에는 어떠한가?

첫 라운드에서 결정이 내려질 확률은 p^3+q^3이다. 그렇지 않으면 두 동전은 일치하고 하나는 다르다. 다수가 앞면이면(확률 $3p^2q$) 다른 동전 하나를 앞면으로 뒤집어 놓으므로, 둘째 라운드에서 결정이 내려지려면 나머지 두 사람이 앞면을 던져야 한다. 그 확률은 p^2이다(이제 모두 뒷면이 되는 것은 불가능하다). 마찬가지로 다수가 뒷면이면(확률 $3pq^2$) 결정이 내려지려면 뒷면이 두 번 나와야 하며, 그 확률은 q^2이다. 따라서

$$\mathrm{P}(\text{결정, 늦어도 라운드2}) = p^3+q^3+3p^2q\cdot p^2+3pq^2\cdot q^2.$$

? **생각해 보기** 모든 것을 $s=pq$로 나타내어라. ($p+q=1$을 이용한다.)

$p^3+q^3=(p+q)^3-3pq(p+q)=1-3s$이고 $3p^4q+3pq^4=3s(p^3+q^3)=3s(1-3s)$이다. 따라서 $s\le\frac{1}{4}$이므로

$$\mathrm{P}=1-3s+3s-9s^2=1-9s^2\ge 1-\frac{9}{16}=\frac{7}{16},$$

이다. $\blacksquare$

돌아보기

- **여사건(complement).** "n번째 라운드까지 적어도 한 번"은 1− ("n번의 라운드 동안 한 번도 없음")으로 계산하는 것이 가장 쉬웠다.
- **최악의 경우를 한정하기.** 알 수 없는 p는 부등식(inequality) $pq \leq \frac{1}{4}$로 처리했다.
- **대칭함수(symmetric function).** p와 q에 대하여 대칭인 식은 $p+q=1$과 pq로 나타내면 간단해진다.

숨은 수학

폰 노이만의 기법

문제의 절차는 빠르지만 치우쳐 있다. 폰 노이만의 절차는 공정하다. 두 번 던져서 HT나 TH이면 결정하고, HH나 TT이면 반복한다. 각 시도는 확률 $2pq$로 성공하므로 평균 $\frac{1}{2pq}$번의 시도가 필요하며, 동전이 매우 치우쳐 있으면 느리다(그림 12.9.1). 문제의 절차는 라운드마다 확률 $p^2+q^2 \geq \frac{1}{2}$로 결정을 내리지만, $p > \frac{1}{2}$이면 그 답은 호텔 쪽으로 기운다.

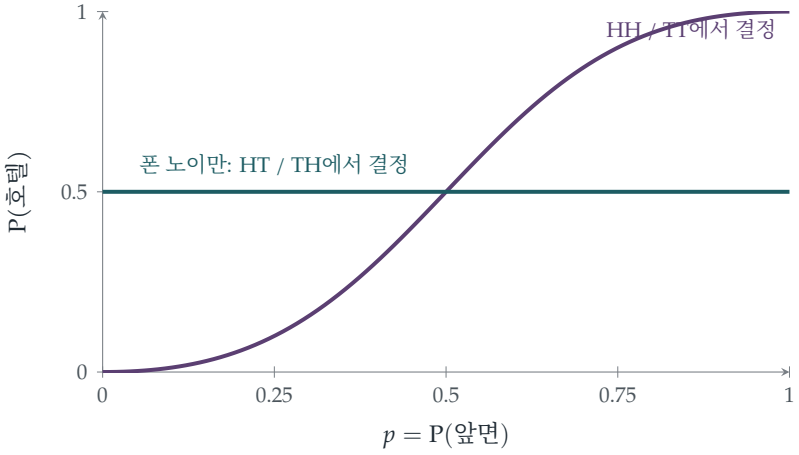


그림 12.9.1 동전의 치우침에 따른 호텔을 고를 확률. 동전들이 일치할 때 결정하는 방법(자주색)은 빠르지만 더 잘 나오는 면 쪽으로 기운다. 폰 노이만의 절차(청록색)는 모든 p에 대하여 정확히 공정하지만, 동전이 매우 치우쳐 있으면 시간이 더 오래 걸린다.

무작위성 추출

폰 노이만의 기법은 최초의 **무작위성 추출기**(*randomness extractor*), 곧 불완전한 무작위성을 완전한 무작위성으로 바꾸는 절차이다. 이 기법은 낭비가 많다(두 번 던질 때마다 공정

한 비트를 많아야 하나 얻고, 하나도 얻지 못할 때도 많다). 더 나은 방법들은 버려진 HH와 TT 쌍을 재활용하여 동전에 담긴 무작위성을 거의 전부 추출하는데, 그 양은 던질 때마다 엔트로피(*entropy*) $-p\log_2 p - q\log_2 q$비트로 측정된다. 추출기는 컴퓨터의 하드웨어 난수 생성기에 쓰인다. 이 생성기는 예측할 수 없지만 치우친 물리적 잡음을 모아서, 암호에 쓸 수 있도록 치우치지 않은 비트로 바꾸어야 한다.

폰 노이만과 난수

존 폰 노이만은 1951년 논문 "Various techniques used in connection with random digits"(난수와 관련하여 쓰이는 여러 기법)에서 이 기법을 설명했다. 이 논문은 그와 동료들이 핵무기 설계를 위해 최초의 컴퓨터 시뮬레이션 몇 가지("몬테카를로 방법(Monte Carlo method)")를 수행하던 시기에 쓰였다. 같은 논문에서 그는 산술로 난수를 만들어 내는 일에 대해 유명한 말을 남겼다. "산술적 방법으로 난수를 만들어 내려는 사람은 누구든, 당연히, 죄를 짓고 있는 상태이다."

사고 기술

대칭성을 이용하여 미지수를 소거하기

매개변수를 모를 때는 그 값과 상관없이 확률이 같은 사건을 찾아, 그 사건들을 바탕으로 결정하라.

HT와 TH는 확률이 pq로 같으므로, 둘을 비교하면 p가 완전히 사라진다. 대학에서는 같은 아이디어가 통계학의 **피벗량**(*pivotal quantity*)과, 무작위화(randomisation)와 대칭성으로 알 수 없는 효과를 상쇄하는 실험 설계(design of experiments)에서 나타난다.

더 도전하기

1 **라운드 수의 기댓값.** 소문항 (i)의 절차와 폰 노이만의 절차에 대하여 라운드 수의 기댓값(expected value)을 구하여라. 어떤 p에 대하여 폰 노이만의 절차가 더 느린가?

2 **공정한 셋 중 하나 고르기.** 치우친 동전을 사용하여 세 가지 선택지 가운데 하나를 각각 정확히 확률 $\frac{1}{3}$로 고르는 절차를 고안하여라.

3 **공정한 동전으로 만드는 치우친 동전.** 공정한 동전을 사용하여 앞면이 나올 확률이 정확히 $\frac{1}{3}$인 동전을 흉내 내는 절차를 고안하여라. 던지는 횟수의 기댓값은 얼마인가?

4 다시 세 사람. 소문항 (ii)에서 그들이 호텔에 갈 확률을 p의 함수로 나타내면 무엇인가? $p = \frac{1}{2}$일 때 이 절차는 공정한가?

더 읽을거리

- **John von Neumann, "Various techniques used in connection with random digits"**, *National Bureau of Standards Applied Mathematics Series* 12 (1951). 짧고 유명한 논문이다.
- **Donald Knuth, *The Art of Computer Programming*, Vol. 2** (Addison-Wesley), 난수를 다룬 3장.
- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing*** (Cambridge University Press), 무작위성 추출과 엔트로피를 다룬다.

12.10 느슨한 끈 끝 묶기

문제

STEP III 2008, 13번 문제

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

`step.maths.org/questions/08-s3-q13`

도입

끈 백 가닥을 주머니에 넣어 끝만 삐져나오게 한다. 끝 두 개를 무작위로 골라 묶고, 남은 끝이 없을 때까지 이것을 되풀이한다. 그러면 고리 여러 개가 뒤엉킨 덩어리가 남는다. 고리는 몇 개일까? 끈 백 가닥으로 원리상 작은 고리 백 개를 만들 수도 있고, 거대한 고리 하나를 만들 수도 있다. 놀랍게도 고리 개수의 기댓값(expected value)은 겨우 3.3 정도이고, 끈이 사만 가닥이어도 여전히 6 정도에 불과하다.

핵심 관찰 하나로 이 문제는 거의 자명해진다. 매듭을 하나 지을 때마다, 무슨 일이 일어났든 상자 안의 끈은 그저 한 가닥 줄어든다. 이 과정은 자신의 이력을 "잊어버리며", 고리의 개수는 확률이 바뀌어 가는 독립적인 동전 던지기들의 합이 된다.

시작하기 전에

생각해 보기 두 끝을 묶은 뒤에는 끈이 몇 가닥 남는가? (고리가 된 것은 더 이상 "가닥"으로 세지 않는다.) 이 개수는 고리가 만들어졌는지에 따라 달라지는가?

두 끝이 같은 끈에 속하면 그 끈은 고리가 되고 $n-1$가닥이 남는다. 서로 다른 끈에 속하면 두 끈은 더 긴 한 가닥이 되고, 역시 $n-1$가닥이 남는다. 어느 경우든 끝이 두 개씩인 끈 $n-1$가닥이 남는다.

차근차근 풀기

1단계 첫 번째 매듭

끝은 $2n$개이다. 하나를 고르면, 그 짝은 나머지 $2n-1$개의 끝 중 어느 것이든 같은 확률로 될 수 있고, 그중 정확히 하나가 같은 끈에 속한다. 따라서 첫 번째 매듭이 고리를 만들 확률은 $\frac{1}{2n-1}$이고, 첫 단계에서 만들어지는 고리 개수의 기댓값은 $\frac{1}{2n-1}$이다.

2단계 모든 매듭

생각해 보기 위의 관찰에 따르면, 두 번째 매듭을 짓기 전의 상황은 어떤가? 세 번째는? 끈이 k가닥 남았을 때 짓는 매듭에서 고리가 생기는지를 나타내는 지시확률변수(indicator)를 I_k라 하자. $\mathrm{E}(I_k)$는 각각 얼마인가?

각 매듭을 짓기 전에 상자에는 $k = n, n-1, \dots, 1$에 대해 k가닥의 끈이 있고, 그 매듭이 고리를 만들 확률은 $\frac{1}{2k-1}$이다. (한 가닥이 남았을 때 짓는 마지막 매듭은 항상 고리를 만든다.) 고리의 총 개수는 $R = I_n + I_{n-1} + \cdots + I_1$이므로

$$\mathrm{E}(R) = \sum_{k=1}^{n} \frac{1}{2k-1} = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1}. \qquad \blacksquare$$

3단계 분산

생각해 보기 지시확률변수 I_k들은 독립(independent)인가? (어떤 단계에서 고리가 생길 확률이 그 전에 일어난 일에 따라 달라지는가?)

각 단계에서 고리가 생길 확률은 그 전에 무슨 일이 있었든 $\frac{1}{2k-1}$이다. 상태가 언제나 "끈 k가닥"이기 때문이다. 따라서 I_k들은 독립이고, 분산(variance)은 더해진다.

$$\mathrm{Var}(R) = \sum_{k=1}^{n} \frac{1}{2k-1}\left(1 - \frac{1}{2k-1}\right) = \sum_{k=1}^{n} \frac{2k-2}{(2k-1)^2}. \qquad \blacksquare$$

4단계 끈 사만 가닥

생각해 보기 $2n-1$까지의 홀수의 역수의 합을 조화수(harmonic number)를 써서 나타내어라.

$2n$까지의 모든 수의 역수의 합에서 짝수의 역수의 합을 빼면

$$\sum_{k=1}^{n} \frac{1}{2k-1} = H_{2n} - \tfrac{1}{2}H_n \approx \ln(2n) - \tfrac{1}{2}\ln n = \tfrac{1}{2}\ln(4n).$$

$n = 40\,000$이면 $4n = 160\,000 = 20^4$이므로 $\mathrm{E}(R) \approx \frac{1}{2} \cdot 4 \ln 20 \approx 6$이다. ■(정확한 값은 6.28이다. 차이는 오일러(Euler) 상수(Euler's constant)에서 오는데, $H_n \approx \ln n + 0.577$이기 때문이다.)

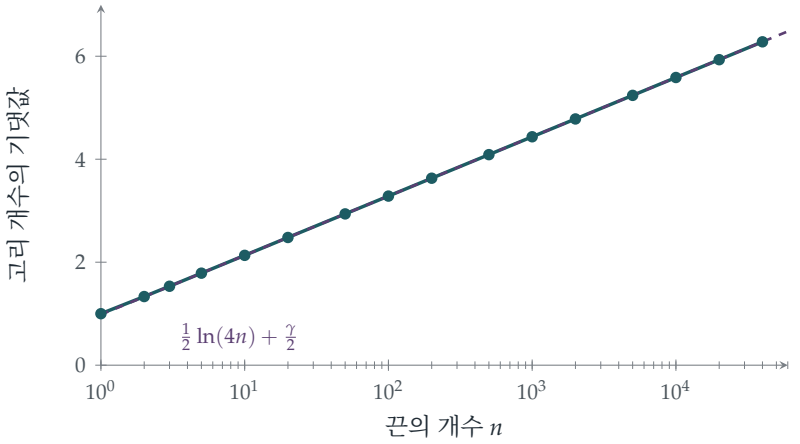


그림 12.10.1 끈의 개수에 대한 고리 개수의 기댓값을 로그 눈금으로 나타낸 그래프. 정확한 값(청록색)은 직선 $\frac{1}{2}\ln(4n) + \frac{\gamma}{2}$(자주색, 점선) 위에 놓인다. 여기서 $\gamma \approx 0.577$은 오일러 상수이다. 끈의 개수를 10배로 늘려도 고리는 약 1.15개만 늘어난다.

돌아보기

- **과정이 무엇을 잊어버리는지 찾는다.** 매듭을 지을 때마다 상자의 상태는 이력과 무관하게 "끈 k가닥"이었다.
- **다시 지시확률변수.** 고리의 개수는 매듭 하나에 하나씩 대응하는 지시확률변수들의 합이었다.
- **독립성이 분산을 준다.** 각 단계의 확률이 과거에 의존하지 않았으므로 지시확률변수들은 독립이었다.

숨은 수학

로그적 증가

고리 개수의 기댓값은 $\frac{1}{2}\ln n$처럼 증가하고, 분산은 평균에 가깝다(둘 다 작은 확률들의 합이다). 중심극한정리(central limit theorem, Section 12.24)에 따르면, n이 클 때 고리의 개수는 평균과 분산이 약 $\frac{1}{2}\ln(4n)$인 정규분포(normal distribution)를 근사적으로 따른다. 끈이 40 000가닥이면 고리는 대개 2개에서 11개 사이로 생긴다.

무작위 순열의 순환

같은 로그가 n개 대상의 무작위 순열(permutation)이 갖는 순환(cycle)의 개수에도 나타난다. 그 평균은 $H_n \approx \ln n$이다(끈 문제는 이와 가까운 친척으로, $\frac{1}{2}\ln n$이다). 무작위 순열은 순환이 적다. 대개 거대한 순환 하나가 원소 대부분을 포함하며, 가장 큰 순환에 속한 원소의 비율은 평균적으로 약 0.62로, 이것이 골롬–딕맨 상수(Golomb–Dickman constant)이다. 끈을 무작위로 엉키게 하는 것도 같은 방식으로 움직인다. 끈의 대부분은 긴 고리 한두 개에 들어간다. 이러한 사실들은 **무작위 조합 구조**(*random combinatorial structures*) 연구의 일부이며, 이 분야는 정수론(number theory)과도 연결되어 있다(무작위 정수의 소인수(prime factors)는 무작위 순열의 순환처럼 행동한다).

무작위 매듭과 고분자

끈 묶기 문제는 물리계의 장난감 모형이다. 용액 속의 긴 고분자(polymer) 분자들은 끝과 끝이 이어져 고리와 사슬을 이루며, 고리 형성의 통계는 비슷한 규칙을 따른다. 매듭 이론(knot theory)에서 수학자들은 무작위로 만든 닫힌 고리가 매듭지어져 있을 확률이 얼마인가라는 더 어려운 질문을 던진다. 긴 무작위 고리에 대해서는 그 답이 1로 수렴하며, 이 결과는 1988년 De Witt Sumners와 Stuart Whittington이 증명하였다.

사고 기술

무기억 구조를 찾아라

각 단계 뒤에 계의 상태가 지금까지 거친 단계의 수에만 의존한다면, 각 단계는 독립적으로 행동하고 기댓값과 분산은 그냥 더해진다.

이것이 쿠폰 수집가 문제(coupon collector, Section 9.11)의 바탕에 있는 아이디어로, 거기서는 각 단계가 새로운 기하분포(geometric distribution) 대기 시간이다. 이 아이디어는 현재 상태만이 중요한 마르코프 연쇄(Markov chain) 이론에서도 다시 나타난다.

더 도전하기

1 **작은 경우.** $n = 2$일 때 가능한 경우를 모두 나열하고, 고리 개수의 기댓값이 $\frac{4}{3}$임을 확인하여라.

2 **하나의 큰 고리.** 모든 끈을 포함하는 고리 하나로 과정이 끝날 확률이 $\prod_{k=2}^{n}\left(1-\frac{1}{2k-1}\right)$임을 보이고, n이 클 때 이 값이 근사적으로 $\sqrt{\frac{\pi}{4n}}$임을 보여라. (**힌트:** 이 곱을 $\binom{2n}{n}$과 관련지어라.)

3 **순열의 순환.** 순열을 원소 하나씩 만들어 가는 방법("중국 식당(Chinese restaurant)" 구성)으로, n개 대상의 무작위 순열이 갖는 순환 개수의 기댓값이 H_n임을 보여라.

4 **분산.** $\operatorname{Var}(R) = \mathrm{E}(R) - \sum_{k=1}^{n}\frac{1}{(2k-1)^2}$임을 보이고, 빼는 합이 $\frac{\pi^2}{8}$로 수렴함을 보여라.

더 읽을거리

- **Peter Winkler, *Mathematical Puzzles: A Connoisseur's Collection*** (A K Peters). 끈 묶기 퍼즐이 실려 있다.
- **Richard Arratia, Andrew Barbour & Simon Tavaré, *Logarithmic Combinatorial Structures*** (European Mathematical Society). 순환, 소수, 무작위 구조를 다루며, 나중을 위한 책이다.
- **Colin Adams, *The Knot Book*** (American Mathematical Society). 무작위 매듭을 포함한 매듭 이론 입문서이다.

12.11 우표, 컵, 그리고 e

문제

STEP I 1996, 13번 문제

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

`step.maths.org/questions/96-s1-q13`

도입

우표를 파는 기묘한 방법이 있다. 사는 사람은 어느 컵 아래에 우표가 숨어 있는지 계속 맞혀야 하고, 한 판이 지날 때마다 컵이 하나씩 늘어난다. 그리고 처음으로 틀렸을 때 탁자 위에 놓인 컵 하나당 1파운드씩 낸다. 평균적으로 그가 내는 돈은 £e, 즉 약 £2.72이다.

수 e는 확률에서 묘할 정도로 자주 나타난다. 아무도 자기 셔츠를 돌려받지 못할 확률(Section 9.8), 비서 문제(secretary problem, Section 12.12), 포아송 분포(Poisson distribution)가 그 예이다. 이 문제는 e가 등장하는 사례 가운데 아마 가장 단순한 것이며, 놀랍게도 또 다른 유명한 퍼즐과 정확히 같은 분포를 가진다. 0과 1 사이의 난수를 몇 개 더해야 합이 1을 넘을까? 평균적으로 e개이다.

시작하기 전에

생각해 보기 $N = 4$가 되려면 무슨 일이 일어나야 하는가? 그 확률은 얼마인가?

짐(Jim)은 컵이 2개일 때(확률 $\frac{1}{2}$)와 3개일 때($\frac{1}{3}$) 맞히고, 컵이 4개일 때 틀려야 한다($\frac{3}{4}$). 따라서 $\mathrm{P}(N = 4) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{4} = \frac{3}{4!} = \frac{1}{8}$이다.

차근차근 풀기

1단계 분포

$N = k$ ($k \geq 2$)가 되려면 짐은 컵이 $2, 3, \ldots, k-1$개일 때 맞히고 k개일 때 틀려야 한다.

$$\mathrm{P}(N = k) = \frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{k-1} \cdot \frac{k-1}{k} = \frac{k-1}{k!}.$$

(검산: $\frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \cdots = \sum_{k \geq 2} \left(\frac{1}{(k-1)!} - \frac{1}{k!} \right) = 1$이며, 이는 망원급수(telescoping sum)이다.)

2단계 기댓값

생각해 보기 $k \cdot \frac{k-1}{k!}$를 간단히 하여라. 어떤 급수가 나오는가?

$$\mathrm{E}(N) = \sum_{k \geq 2} \frac{k(k-1)}{k!} = \sum_{k \geq 2} \frac{1}{(k-2)!} = \frac{1}{0!} + \frac{1}{1!} + \cdots = e. \qquad \blacksquare$$

3단계 분산

생각해 보기 $k^2 = k(k-1) + k$로 쓰거나 $\frac{k^2(k-1)}{k!}$를 직접 간단히 하여 $\mathrm{E}(N^2)$를 구하여라.

$$\mathrm{E}(N^2) = \sum_{k \geq 2} \frac{k^2(k-1)}{k!} = \sum_{k \geq 2} \frac{k}{(k-2)!} = \sum_{j \geq 0} \frac{j+2}{j!} = \sum_{j \geq 1} \frac{1}{(j-1)!} + 2 \sum_{j \geq 0} \frac{1}{j!} = e + 2e = 3e.$$

따라서 분산(variance)은 $\mathrm{Var}(N) = 3e - e^2 = e(3-e) \approx 0.766$이다. $\blacksquare$

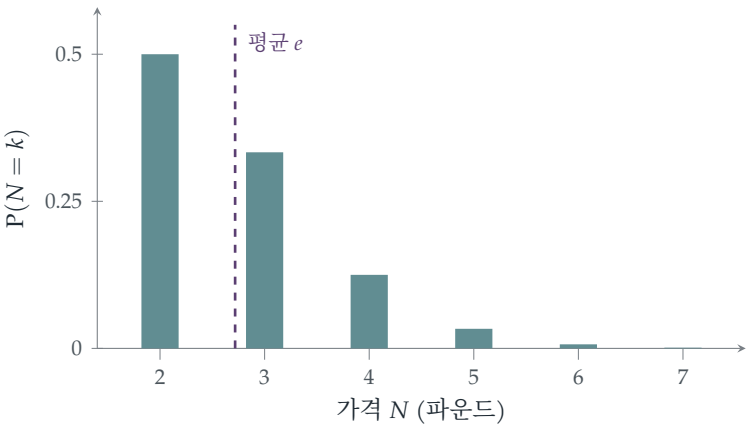


그림 12.11.1 가격의 분포 $\mathrm{P}(N = k) = \frac{k-1}{k!}$. 짐은 절반의 경우에 £2를 내지만, 길고 얇은 꼬리가 평균을 $£e \approx £2.72$까지 끌어올린다.

돌아보기

- **확률의 곱.** 독립(independent)인 성공이 이어지다가 실패로 끝나는 사건의 확률은 곱으로 주어지고, 이 곱은 $\frac{k-1}{k!}$로 간단히 정리되었다.
- **계승을 약분하기.** k나 k^2를 곱한 뒤 약분하자 합이 지수급수(exponential series)로 바뀌었다.
- **첨자 이동으로 구한** $\mathrm{E}(N^2)$**.** $j = k - 2$로 첨자를 바꾸자 합이 e 두 개로 나뉘었다.

숨은 수학

모습만 바꾼 같은 분포

난수 $U_1, U_2, \ldots$를 $[0,1]$에서 균등하게(uniformly) 뽑고, 그 합이 1을 넘는 데 필요한 개수를 M이라 하자. 처음 k개의 수를 더한 값이 1보다 작을 확률은 영역 $\{u_1 + \cdots + u_k < 1, u_i \geq 0\}$의 부피이며, 이 값은 $\frac{1}{k!}$이다(넓이가 $\frac{1}{2}$인 삼각형, 부피가 $\frac{1}{6}$인 사면체(tetrahedron) 등). 따라서

$$\mathrm{P}(M > k) = \frac{1}{k!}, \qquad \mathrm{P}(M = k) = \frac{1}{(k-1)!} - \frac{1}{k!} = \frac{k-1}{k!},$$

이고, 이는 N의 분포와 정확히 같다. 두 문제 모두에서 $\mathrm{P}(N > k) = \frac{1}{k!}$이다. 컵 문제에서 이것은 $k-1$번 연속으로 맞힐 확률이다. 공식 $\mathrm{E}(N) = \sum_{k \geq 0} \mathrm{P}(N > k)$(Section 9.9에서 나온 공식)를 쓰면 한 줄 만에 $\mathrm{E}(N) = \sum \frac{1}{k!} = e$를 얻는다.

e가 자꾸 나타나는 이유

k개의 대상이 특정한 순서로 놓이는 경우이든, 확률이 $\frac{1}{2}, \frac{1}{3}, \ldots$인 사건 k개가 연달아 일어나는 경우이든, 단체(simplex)의 부피이든, 확률에 $\frac{1}{k!}$이 들어 있으면 k에 대한 합은 지수급수가 된다. 교란순열(derangement, Section 9.8)에서는 $\sum \frac{(-1)^k}{k!} = e^{-1}$이 나왔고, 여기서는 $\sum \frac{1}{k!} = e$가 나왔다. 그리고 일어날 가능성이 낮은 독립 사건이 많이 있을 때 생기는 $\left(1 - \frac{1}{n}\right)^n \to e^{-1}$ 꼴의 극한은 비서 문제와 포아송 분포에 e가 나타나는 이유를 설명해 준다.

몬테카를로 방법으로 구하는 e

"합이 1을 넘는" 문제를 이용하면 실험으로 e를 추정할 수 있다. 난수를 만들어 1을 넘는 데 몇 개가 필요한지 세고, 이를 여러 번 반복하여 평균을 내면 된다. 이는 π를 추정하는 뷔퐁(Buffon)의 바늘(Section 12.20)과 사촌 격이다. 카지노의 이름을 딴 이러한 **몬테카를로**(Monte Carlo) 방법은 직접 계산하기 어려운 양을 무작위 표본추출로 추정한다. 오늘날 이 방법은 입자물리학에서 금융에 이르기까지 과학 전반에서 쓰이며, 주로 고차원 적분에 사용된다.

사고 기술

급수를 알아보기

기댓값이 계승을 포함한 합으로 주어지면, 표준적인 급수(지수급수, 기하급수, 이항급수)가 나타날 때까지 항을 간단히 하라.

대학에서는 생성함수(generating function)가 이 과정을 체계화한다. 확률생성함수(probability generating function) $\mathrm{E}(t^N)$는 분포 전체를 하나로 담고 있으며, $t = 1$에서의 도함수 값으로부터 평균과 분산을 얻는다.

더 도전하기

1 **생성함수.** $\mathrm{E}(t^N) = 1 + (t-1)e^t$임을 보이고, 이를 이용하여 $\mathrm{E}(N)$과 $\mathrm{E}(N(N-1))$을 구하여라.

2 **꼬리 공식.** 음이 아닌 정수 값을 갖는 확률변수(random variable)에 대하여 $\mathrm{E}(N) = \sum_{k \geq 0} \mathrm{P}(N > k)$임을 증명하고, 이를 이용하여 한 줄 만에 $\mathrm{E}(N)$을 구하여라.

3 **컵 하나로 시작하기.** 게임이 컵 하나(짐이 틀릴 수 없는 경우)로 시작한다면 분포는 어떻게 달라지는가? 가격의 기댓값은 얼마인가?

4 **단체.** $\{u_1 + u_2 + u_3 < 1, u_i \geq 0\}$의 부피가 $\frac{1}{6}$임을 보이고, 수학적 귀납법(induction)을 이용하여 k차원에서는 그 부피가 $\frac{1}{k!}$임을 보여라.

더 읽을거리

- **Eli Maor, *e: The Story of a Number*** (Princeton University Press). 확률에서의 등장을 포함한 e의 역사를 다룬다.
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University Press). 생성함수를 다룬다.
- **Paul Nahin, *Digital Dice*** (Princeton University Press). 몬테카를로 시뮬레이션으로 푸는 확률 퍼즐 모음이다.

12.12 언제 그만 찾아야 하는가

문제

STEP I 2001, 14번 문제

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

`step.maths.org/questions/01-s1-q14`

도입

지원자를 한 명씩 면접하면서 그 자리에서 결정을 내려야 한다고 하자. 이 지원자를 합격시키고 멈추거나, 아니면 영원히 불합격시키고 다음 지원자를 본다. 목표는 단 한 명의 최고 지원자를 뽑는 것이다. 너무 일찍 합격시키면 더 나은 사람을 놓치기 쉽고, 너무 오래 기다리면 최고의 지원자는 이미 지나가 버렸기 쉽다. 어떻게 해야 할까?

그 답은 확률(probability)론에서 가장 유명한 결과 가운데 하나이다. 처음 37%의 지원자는 모두 불합격시키고, 그 뒤로는 앞선 모든 지원자보다 나은 첫 지원자를 합격시킨다. 이 방법은 약 37%의 확률로 성공하며, 37%는 바로 $\frac{1}{e}$이다. 이 문제에서는 기본적인 확률을 계산하고, 짝을 이루는 STEP II 2006의 문제(연습문제에 있다)에서는 이 전략의 가장 단순한 형태, 즉 첫 번째 지원자만 불합격시키는 전략을 분석한다.

시작하기 전에

생각해 보기 면접 순서가 무작위일 때, 전체에서 최고인 지원자가 처음 면접한 n명 안에 있을 확률은 얼마인가?

최고의 지원자는 N개의 자리 중 어느 자리에나 같은 확률로 있으므로, 구하는 확률은 $\frac{n}{N}$이다.

차근차근 풀기

1단계 처음 n명 중 최고가 전체 최고인 경우

처음 n명 중 최고인 사람이 전체 최고인 것은 전체 최고가 처음 n명 안에 있을 때이고, 그때뿐이다. 따라서 확률은 $\frac{n}{N}$이다. ■

2단계 최고 또는 두 번째

? **생각해 보기** 처음 n명 중 최고인 사람이 전체에서 두 번째인 것은 두 번째 지원자가 처음 n명 안에 있고 최고의 지원자는 그 안에 없을 때이고, 그때뿐이다. 그 확률은 얼마인가?

두 번째 지원자가 처음 n명 안에 있을 확률은 $\frac{n}{N}$이다. 이 조건 아래에서 최고의 지원자는 나머지 $N-1$개의 자리 중 하나에 있고, 그중 $N-n$개가 처음 n명 뒤의 자리이다. 따라서

$$\mathrm{P}(\text{최고 또는 두 번째}) = \frac{n}{N} + \frac{n}{N}\cdot\frac{N-n}{N-1}. \qquad ■$$

$N=4, n=2$이면 $\frac{1}{2}+\frac{1}{2}\cdot\frac{2}{3}=\frac{5}{6}$이다. 상위 두 지원자가 네 자리 중 어디에 오는지를 나열해 보면(순서를 고려한 12가지 배치가 모두 같은 확률이다), 상위 두 지원자가 모두 3번째와 4번째 자리에 오는 2가지 배치를 제외하면 처음 두 명 중 최고인 사람은 언제나 상위 두 명 중 하나이다. 따라서 확률은 $\frac{10}{12}=\frac{5}{6}$이다.

3단계 최적 전략

? **생각해 보기** 전략(strategy): 처음 r명의 지원자를 불합격시키고, 그 뒤로는 앞선 r명 모두보다 나은 첫 지원자를 합격시킨다. 최고의 지원자가 $k>r$번째 자리에 있을 때, 이 전략이 그 사람을 뽑는 것은 언제인가?

이 전략이 k번째 자리에 있는 최고의 지원자를 뽑는 것은 $r+1,\dots,k-1$번째 자리의 어떤 지원자도 처음 r명 모두를 앞서지 못할 때, 즉 처음 $k-1$명 중 최고인 사람이 처음 r명 안에 있을 때이고, 그때뿐이다. 1단계에서 N을 $k-1$로 바꾸면 이 확률은 $\frac{r}{k-1}$이다. 각 자리 k의 확률은 $\frac{1}{N}$이므로

$$\mathrm{P}(\text{성공}) = \frac{1}{N}\sum_{k=r+1}^{N}\frac{r}{k-1} = \frac{r}{N}\left(\frac{1}{r}+\frac{1}{r+1}+\cdots+\frac{1}{N-1}\right).$$

? **생각해 보기** N이 클 때 $x=\frac{r}{N}$이라 하자. 이 합은 근사적으로 $\ln\frac{N}{r}$이다. $x\ln\frac{1}{x}$를 최대로 하는 x의 값은 무엇인가?

성공 확률은 근사적으로 $x \ln \frac{1}{x}$이다. 미분(differentiation)하면 $x = \frac{1}{e}$일 때 $\ln \frac{1}{x} - 1 = 0$이고, 최댓값은 $\frac{1}{e} \approx 0.368$이다(그림 12.12.1).

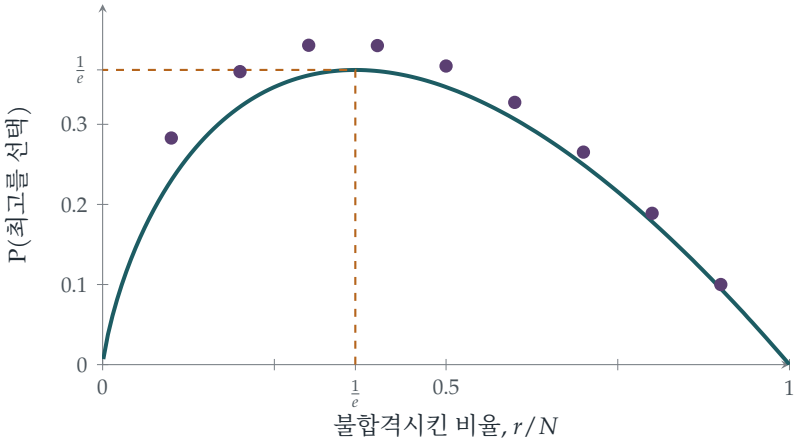


그림 12.12.1 처음 r/N 비율의 지원자를 불합격시킬 때 최고의 지원자를 뽑을 확률: 극한 곡선 $x \ln \frac{1}{x}$(청록색)와 $N = 10$일 때의 정확한 값(자주색). N이 아무리 커도, 최선의 전략은 지원자의 약 $\frac{1}{e} \approx 37\%$를 불합격시키고 약 $\frac{1}{e}$의 확률로 성공한다.

돌아보기

- **자리의 대칭성(symmetry).** 무작위 순서에서는 각 지원자가 어느 자리에나 같은 확률로 있으므로, $\frac{n}{N}$을 바로 얻었다.
- **최고의 지원자가 어디 있는지로 조건을 건다.** 전략의 성공 확률은 최고의 지원자의 자리에 대해 조건부(conditioning)로 나누어 계산했다.
- **합이 적분이 된다.** N이 클 때 조화합(harmonic sum)은 로그가 되었고, 최적화를 하여 $\frac{1}{e}$을 얻었다.

숨은 수학

최적 멈춤

비서 문제(secretary problem)는 **최적 멈춤**(*optimal stopping*)의 출발점이 된 예이다. 최적 멈춤이란 시간에 따라 들어오는 정보를 바탕으로, 기대 보상을 최대로 하도록 언제 행동할지를 결정하는 문제이다. ("r명을 불합격시키고, 그 뒤 지금까지의 최고가 나오면 뽑는다"는 전략만이 아니라) 모든 전략 가운데 이러한 문턱 규칙이 최적임을 마지막 지원자에서부터 거꾸로 따져 가는 논증으로 증명할 수 있다. 주식을 언제 팔지, 금융 옵션을 언제 행사할지, 주차 자리를 얼마나 오래 찾을지도 같은 수학이 지배한다. 놀라운 점은 극한에서 답이

N에 의존하지 않는다는 것이다. 지원자가 백 명이든 백만 명이든 $\frac{1}{e}$의 확률로 성공한다.

변형

상위 두 지원자 중 누구라도 만족한다면, 최선의 전략은 약 57%의 확률로 성공한다. 최고의 지원자를 뽑을 확률 대신 뽑힌 지원자의 기대 **순위**(expected rank)를 최적으로 하려 한다면, 최적 전략은 더 복잡해지고 기대 순위는 약 3.87로 수렴한다(Chow, Moriguti, Robbins, Samuels, 1964). 불합격시킨 지원자를 어떤 확률로 다시 부를 수 있거나, 지원자 점수의 분포를 알고 있다면("완전 정보" 문제) 답은 또 달라진다. 연습문제의 아이스크림 문제는 가장 단순한 경우이다. 한 명만 불합격시키고, 그다음에 더 나은 것이 나오면 택한다.

누가 풀었을까?

이 문제는 1950년대에 수학자들 사이에서 돌았고, 1960년에 마틴 가드너(Martin Gardner)가 "구골(googol) 게임"이라는 이름으로 널리 알렸다(종이쪽지에 수를 적으면, 상대가 한 장씩 뒤집어 보다가 가장 큰 수에서 멈춰야 한다). Thomas Ferguson의 글 "Who solved the secretary problem?"(1989)은 이 문제의 얽히고설킨 역사를 추적한다. 한 전승에 따르면 천문학자 요하네스 케플러(Johannes Kepler)는 1611년에 두 번째 아내를 고르면서 열한 명의 후보를 만났고, 결국 다섯 번째 후보에게 청혼했다고 한다.

사고 기술

전략을 매개변수로 나타내고 최적화하기

최선의 전략을 모를 때는 매개변수를 가진 자연스러운 전략의 모임을 생각하고, 성공 확률을 그 매개변수의 함수(*function*)로 계산한 뒤 최적화한다.

대학에서는 이 방법을 결정 이론(decision theory), 게임 이론(game theory), 통계학(statistics)에서 쓴다. 그 모임 안에서 최선인 전략이 전체에서도 최선임을 보이는 것은 더 어렵고 더 흥미로운 단계로, 흔히 **역진 귀납법**(*backward induction*)으로 해낸다.

더 도전하기

1 **아이스크림.** STEP II 2006, 13번 문제 (`step.maths.org/questions/06-s2-q13`)에 도전해 보라. 간단한 규칙을 쓰는 작은 비서 문제이다. 첫 번째 것을 거절하고, 그다음 그것보다 나은 첫 번째 것을 택한다.

2 **작은 N.** $N = 4$일 때 최선의 r의 값과 성공 확률을 구하라. 모든 경우를 나열하여 확인하라.

3 **최고 또는 두 번째.** 소문항 (ii)를 이용하여, "처음 n명 중 최고가 상위 두 명 중 하나이다"가 "처음 n명 중 최고가 전체 최고이다"보다 훨씬 일어나기 쉬움을 논증하라. 이것은 상위 두 명 중 하나를 뽑기 위한 전략에 대해 무엇을 시사하는가?

4 **정확한 최댓값.** $N = 10$일 때 그림에서 $r = 3$이 최적임을 확인하라. $N = 100$일 때 $r = 37$이 최적임을 보여라($r = 36, 37, 38$일 때의 성공 확률은 소수점 아래 넷째 자리까지 0.3710, 0.3710, 0.3708이다).

더 읽을거리

- **Thomas Ferguson, "Who solved the secretary problem?"**, *Statistical Science* 4 (1989). 역사와 여러 변형을 다룬다. 온라인에서 무료로 볼 수 있다.
- **Brian Christian & Tom Griffiths, *Algorithms to Live By*** (William Collins). 최적 멈춤을 비롯한 여러 알고리즘을 일상의 결정에 적용한다.
- **Martin Gardner, *New Mathematical Diversions*** (Mathematical Association of America). 구골 게임을 다룬다.

12.13 삼자 결투

문제

STEP II 1995, 12번 문제

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

`step.maths.org/questions/95-s2-q12`

도입

세 사람이 빵을 던지며 결투(duel)를 벌인다. 커스버트는 절대 빗맞히지 않고, 버티는 다섯 번에 세 번 맞히며, 가장 서툰 앨저넌은 다섯 번에 두 번밖에 맞히지 못한다. 앨저넌이 먼저 던진다. 그는 누구를 겨냥해야 할까? 놀랍게도 답은 "아무도 겨냥하지 않는다"이다. 그는 일부러 빵을 엉뚱한 곳으로 던져야 한다.

이유는 전략에 있다. 세 사람이 모두 남아 있는 동안에는 위험한 두 사수인 버티와 커스버트가 서로를 겨냥한다. 앨저넌이 둘 중 하나를 쓰러뜨리면 살아남은 쪽이 그를 노리게 된다. 일부러 빗맞히면 두 사람이 먼저 승부를 가리게 두고, 살아남은 사람에게 첫 번째로 던질 기회를 얻는다. 실제로 최선의 계획을 따르면 가장 서툰 사수가 승리할 가능성이 **가장** 높아진다. 이 "삼자 결투(truel)"는 약점이 어떻게 강점이 될 수 있는지 보여 주는, 게임 이론(game theory)에서 즐겨 쓰는 예이다.

시작하기 전에

생각해 보기 두 사람의 결투에서 퍼시가 먼저 던진다. 그가 곧바로 맞히거나, 아니면 둘 다 빗맞혀서 상황이 처음과 똑같이 되풀이된다. 퍼시가 이길 확률(probability)에 대한 방정식을 세워 보아라.

퍼시가 이길 확률을 P라 하자. 그는 확률 p로 곧바로 맞힌다. 그렇지 않으면(확률 $1-p$) 로드니가 던지는데, 로드니가 확률 $1-r$로 빗맞히면 처음 상황으로 돌아간다. 따라서 $P = p + (1-p)(1-r)P$이다.

차근차근 풀기

1단계 두 사람의 결투

이를 풀면 $P(1-(1-p)(1-r))=p$이고, $1-(1-p)(1-r)=p+r-pr$이므로

$$P=\frac{p}{p+r-pr}. \qquad \blacksquare$$

앨저넌($p=\frac{2}{5}$)과 버티($r=\frac{3}{5}$)가 결투하고 앨저넌이 먼저 던지면 이 값은 $\frac{2/5}{2/5+3/5-6/25}=\frac{10}{19}$이다. 버티가 먼저 던지면 앨저넌이 이길 확률은 $\left(1-\frac{3}{5}\right)\cdot\frac{10}{19}=\frac{4}{19}$이다.

2단계 앨저넌의 세 가지 선택

생각해 보기 각 경우에 누가 남고, 다음에 누가 던지는가?

(i) **앨저넌이 커스버트를 맞힌다.** 다음에는 버티가 앨저넌을 향해 던진다. 앨저넌은 버티가 먼저 던지는 결투에서 이겨야 하므로 확률은 $\frac{4}{19}$이다.

(ii) **앨저넌이 버티를 맞힌다.** 다음에는 커스버트가 앨저넌을 향해 던지는데, 커스버트는 절대 빗맞히지 않는다. 앨저넌의 승리 확률은 0이다.

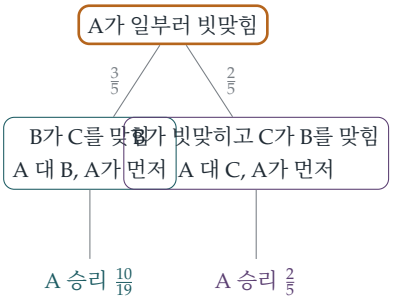
(iii) **앨저넌이 빗맞힌다.** 버티가 커스버트를 향해 던진다. 확률 $\frac{3}{5}$로 버티가 맞히면 앨저넌과 버티가 남고 앨저넌이 던질 차례이므로, 앨저넌은 확률 $\frac{10}{19}$로 이긴다. 확률 $\frac{2}{5}$로 버티가 빗맞히면 커스버트가 버티를 맞히고, 앨저넌은 단 한 번의 던지기로 커스버트와 맞선다. 그는 맞혀야만 이기므로 확률은 $\frac{2}{5}$이다. 이를 모두 합하면

$$\frac{3}{5}\cdot\frac{10}{19}+\frac{2}{5}\cdot\frac{2}{5}=\frac{6}{19}+\frac{4}{25}=\frac{150+76}{475}=\frac{226}{475}\approx 0.476.$$

3단계 조언

생각해 보기 앨저넌이 커스버트를 겨냥하면 맞힐 수도 있고(확률 $\frac{2}{5}$) 빗맞힐 수도 있다. 그의 전체 승리 확률은 얼마인가? 일부러 빗맞히는 경우와 비교해 보아라.

커스버트를 겨냥하면 승리 확률은 $\frac{2}{5}\cdot\frac{4}{19}+\frac{3}{5}\cdot\frac{226}{475}=\frac{200+678}{2375}=\frac{878}{2375}\approx 0.370$이다. 버티를 겨냥하는 것은 이보다도 나쁘다. 일부러 빗맞히면 $\frac{226}{475}\approx 0.476$으로 가장 좋다. 따라서 앨저넌은 엉뚱한 곳으로 던져야 하고, 그러면 그가 이길 확률은 $\frac{226}{475}$이다. $\blacksquare$



이 계획을 따르면:
A 승리 $\frac{226}{475} \approx 0.48$
B 승리 $\frac{135}{475} \approx 0.28$
C 승리 $\frac{114}{475} \approx 0.24$

그림 12.13.1 앨저넌의 최선의 계획. 일부러 빗맞힘으로써 그는 더 강한 두 상대가 먼저 싸우게 하고, 살아남은 사람에게 첫 번째로 던질 기회를 확보한다. 가장 서툰 사수가 결국 승리할 가능성이 가장 높고, 완벽한 사수가 가장 낮다.

돌아보기

- **되풀이되는 상황은 방정식을 낳는다.** 결투는 두 번 빗맞힌 뒤 "다시 시작"되므로, 승리 확률에 대한 일차방정식(linear equation)을 얻는다.
- **상대의 대응을 생각하라.** 앨저넌의 선택 하나하나가 버티와 커스버트가 다음에 할 행동을 바꾼다.
- **아무것도 하지 않는 것까지 포함해 모든 선택을 비교하라.** 가장 나빠 보이던 수가 최선의 수였다.

숨은 수학

전략과 게임 이론

삼자 결투는 게임 이론의 핵심 아이디어를 보여 준다. 최선의 행동은 다른 참가자들이 어떻게 대응할지에 달려 있다는 것이다. 가장 위험한 상대인 커스버트를 맞히는 것이 뻔한 수이지만, 그 수는 앨저넌에게 불리한 방향으로 게임을 바꾼다. 이 문제에서는 버티와 커스버트의 전략(strategy)이 고정되어 있다(두 사람은 서로를 겨냥한다). 완전한 게임 이론적 분석에서는 각 참가자가 다른 참가자들도 똑같이 한다는 것을 알면서 자신의 전략을 스스로 고른다. 1944년 폰 노이만(John von Neumann)과 Oskar Morgenstern의 연구, 1950년 내시(John Nash)의 연구에서 시작된 이러한 분석은 오늘날 경제학, 진화생물학, 정치학의 토대를 이루고 있다.

약점이 되는 강점

삼자 결투에서 누가 이길까? 앨저넌이 최선의 계획을 따르면 승리 확률은 앨저넌이 $\frac{226}{475}$, 버티가 $\frac{135}{475}$, 커스버트가 $\frac{114}{475}$이다. 가장 뛰어난 사수가 살아남을 가능성이 가장 낮은데, 선택권이 있는 사람은 모두 그를 겨냥하기 때문이다. 같은 양상은 후보가 여럿인 선거, 기업 간 경쟁, 그리고 가장 강한 경쟁자가 첫 번째 표적이 될 수 있는 자연에서도 나타난다. 1950년대에 삼자 결투를 게임 이론에 도입한 Martin Shubik은 이를 이용해 가장 적합한 자가 반드시 살아남는가 하는 질문을 던졌다.

사고 기술

한 수 앞을 내다보라

행동을 고르기 전에, 그 행동이 다른 모든 사람의 상황을 어떻게 바꾸는지, 그리고 그들이 어떻게 대응할지 따져 보아라.

대학에서는 이것이 게임 이론의 역진 귀납법(backward induction)이 된다. 먼저 게임의 끝을 분석한 다음, 거꾸로 거슬러 올라가며 최선의 첫 수를 정하는 것이다.

더 도전하기

1 **버티와 커스버트.** 앨저넌이 일부러 빗맞힐 때, 버티가 이길 확률은 $\frac{135}{475}$이고 커스버트가 이길 확률은 $\frac{114}{475}$임을 확인하여라.

2 **더 나은 사수.** 앨저넌이 $\frac{2}{5}$ 대신 확률 a로 맞힌다면, a가 어떤 값일 때 일부러 빗맞히는 것이 여전히 그의 최선의 첫 수인가?

3 **버티는 커스버트를 겨냥해야 할까?** 앨저넌이 일부러 빗맞힌다고 할 때, 버티는 앨저넌보다 커스버트를 겨냥하는 것이 유리한가? (커스버트의 행동은 바뀌지 않는다고 가정한다.)

4 **던지는 횟수가 정해진 결투.** 각 참가자가 두 번씩만 던질 수 있는 결투에서 퍼시가 이길 확률을 구하고, 횟수 제한이 없는 결투와 비교하여라.

더 읽을거리

► **D. Marc Kilgour & Steven Brams, "The truel"**, *Mathematics Magazine* 70 (1997). 삼자 결투를 쉽게 개관한 글이다.

- **Avinash Dixit & Barry Nalebuff, *Thinking Strategically*** (Norton). 일반 독자를 위한 게임 이론 책으로, 삼자 결투도 다룬다.
- **Ken Binmore, *Game Theory: A Very Short Introduction*** (Oxford University Press). 게임 이론 입문서이다.

12.14 혼합 전략

문제

STEP I 1989, 16번 문제

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

`step.maths.org/questions/89-s1-q16`

도입

가위바위보에서는 어떤 고정된 선택이든 상대에게 공략당할 수 있다. 매번 바위만 내면 상대는 곧 보를 내는 법을 익힌다. 유일하게 안전한 전략은 무작위화(randomisation)하는 것이다. 이 문제는 그보다 더 정교한 형태로, 보상이 서로 같지 않고 올바른 무작위화 방법도 분명하지 않다. 그런데 1, 2, 3을 각각 확률 $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$로 고르면, 상대가 무엇을 하든 장기적으로 지지 않음을 보장할 수 있다는 사실이 밝혀진다.

이것은 게임 이론(game theory)의 출발점이 된 결과인, 1928년 존 폰 노이만(John von Neumann)의 **최소최대 정리**(*minimax theorem*)를 처음 엿보는 것이다. 이 정리에 따르면, 한 사람의 이득이 곧 다른 사람의 손실인 모든 2인 게임에서, 완벽하게 플레이하는 상대에 맞서 가능한 최선의 결과를 보장하는 무작위 전략이 각 플레이어에게 존재한다.

시작하기 전에

생각해 보기 두 사람이 말하는 수의 각 쌍에 대하여 B의 상금을 나타내는 표를 만들어 보라(B의 수는 세로로, A의 수는 가로로 놓는다).

규칙을 차례로 적용하면(예를 들어 B가 1을, A가 3을 말하면 두 수의 차가 2이므로 더 작은 수를 말한 B가 £2를 낸다) 다음 표를 얻는다.

B \ A	1	2	3
1	0	+1	−2
2	−1	0	+1
3	+2	−1	0

이 표는 **반대칭**(*antisymmetric*)이다(두 플레이어를 바꾸면 부호가 바뀐다). 따라서 어느 플레이어도 처음부터 유리하지 않다는 의미에서 이 게임은 공정하다.

차근차근 풀기

1단계 고정된 선택에 맞서기

A가 항상 1을 말하면, B는 3을 말하여 매 판 £2를 따야 한다. A가 항상 2를 말하면, B는 1을 말하여 £1을 딴다. A가 항상 3을 말하면, B는 2를 말하여 £1을 딴다. 어떤 고정된 선택이든 공략당할 수 있다.

2단계 균등한 무작위 선택에 맞서기

생각해 보기 A가 균등하게 무작위로 수를 고른다면, B가 j를 말할 때 B의 기대 이득(expected gain)은 얼마인가? (표의 j번째 행의 평균을 구하라.)

행의 평균은 $j=1$일 때 $\frac{1}{3}(0+1-2)=-\frac{1}{3}$, $j=2$일 때 0, $j=3$일 때 $\frac{1}{3}(2-1+0)=\frac{1}{3}$이다. 따라서 한 판당 B의 기대 상금은

$$-\tfrac{1}{3}b_1+0\cdot b_2+\tfrac{1}{3}b_3,$$

이고, 이것은 $b_3=1$일 때 최대가 된다. 즉 항상 3을 말하면 한 판에 평균 $\frac{1}{3}$파운드를 딴다. 균등한 무작위화는 A를 지켜 주지 못한다.

3단계 일반적인 경우

생각해 보기 이제 A가 확률 a_j로 j를 말한다고 하자. B의 기대 상금을 A가 말하는 수에 따라 묶어 보라. a_1의 계수(가중치 b_j를 둔 1열의 평균)는 무엇인가?

표의 열을 위에서 아래로 읽으면

$$\text{B의 기대 상금} = (-b_2+2b_3)a_1+(b_1-b_3)a_2+(-2b_1+b_2)a_3,$$

이므로 $X=-b_2+2b_3$, $Y=b_1-b_3$, $Z=-2b_1+b_2$이다.

생각해 보기 세 값을 모두 0으로 만들 수 있는가? (방정식은 세 개지만 서로 독립이 아니다. 왜 그런가?)

이 방정식들은 독립이 아니다. 항등적으로 $X+2Y+Z=0$이기 때문이다(확인해 보라). 따라서 $Y=0$에서 $b_1=b_3$을, $X=0$에서 $b_2=2b_3$을 얻으면 $Z=0$도 저절로 성립한다.

$b_1 + b_2 + b_3 = 1$과 함께 풀면

$$b_1 = \tfrac{1}{4}, \qquad b_2 = \tfrac{1}{2}, \qquad b_3 = \tfrac{1}{4},$$

이고, 이때 $X = Y = Z = 0$이다. 그러면 a_j가 무엇이든 B의 기대 상금은 0이고, B는 장기적으로 아무것도 잃지 않는다. ■

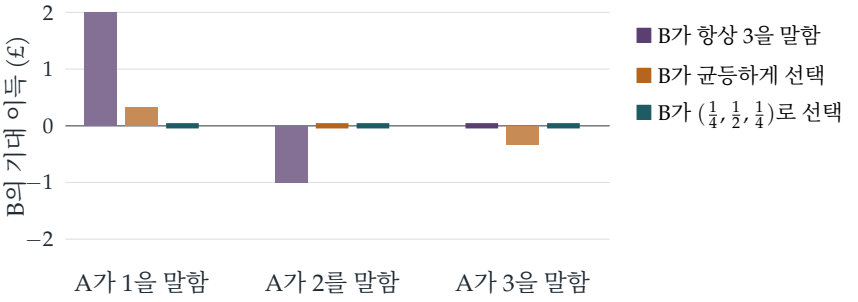


그림 12.14.1 B의 세 가지 전략에 대하여, A의 각 선택에 맞선 B의 기대 이득. 항상 3을 말하면 A의 1에는 이기지만 A의 2에는 진다. 균등한 선택은 A의 3에 진다. 전략 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$는 A가 무엇을 하든 정확히 0(가는 막대로 표시)을 주므로 결코 공략당하지 않는다.

돌아보기

- **보수표(payoff table)를 만들라.** 모든 추론은 3×3 행렬의 행과 열에 관한 것이었다.
- **순수 전략(pure strategy)은 공략당할 수 있다.** 어떤 고정된 선택에 대해서도, 심지어 균등한 무작위 선택에 대해서도 상대에게는 이익이 되는 대응이 있다.
- **균등하게 맞추라.** 안전한 전략은 상대가 어느 쪽을 골라도 차이가 없게 만든다. 즉 A의 모든 선택이 같은 기대 결과를 준다.

숨은 수학

최소최대 정리

보수 행렬 M으로 주어지는 **2인 영합**(*zero-sum*) 게임에서 B가 확률 $\mathbf{b}$를, A가 확률 $\mathbf{a}$를 고르면 B의 기대 상금은 $\mathbf{b}^{\mathsf{T}} M \mathbf{a}$이다. 폰 노이만의 최소최대 정리는

$$\max_{\mathbf{b}} \min_{\mathbf{a}} \mathbf{b}^{\mathsf{T}} M \mathbf{a} = \min_{\mathbf{a}} \max_{\mathbf{b}} \mathbf{b}^{\mathsf{T}} M \mathbf{a}.$$

임을 말한다. 말로 하면, B가 무작위 전략을 미리 공표하여 보장할 수 있는 금액은, A가 자신의 전략을 공표하여 B를 묶어 둘 수 있는 금액과 같다. 이 공통의 값을 게임의 **값**(*value*)이라

한다. 여기서는 반대칭성 때문에 값이 0이고, 전략 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$는 두 플레이어 모두에게 최적이다. 이 정리는 선형대수와 볼록성(convexity)을 이용하여 증명하며(선형 계획법(linear programming)의 쌍대 정리(duality theorem)와 동치이다), 이를 여러 플레이어로 일반화한 내시 균형(*Nash equilibrium*)으로 존 내시(John Nash)는 노벨 경제학상을 받았다.

왜 무작위화하는가?

여기서 무작위성은 운의 문제가 아니라 정보의 문제이다. B의 선택을 예측할 수 있다면 A는 그것을 공략할 수 있다. B는 올바른 확률로 무작위화함으로써 자신의 선택을 예측할 수 없게 만들 **뿐 아니라** 어떤 예측도 도움이 되지 않도록 한다. 실제 예로는 축구의 페널티킥이 있다. 프로 선수들을 연구한 결과, 키커와 골키퍼가 게임 이론의 최적값에 가까운 비율로 왼쪽과 오른쪽을 무작위로 선택한다는 사실이 밝혀졌다. 테니스의 서브도 또 다른 예이다.

폰 노이만의 1928년 논문

존 폰 노이만은 1928년 논문 "On the theory of games of strategy"(전략 게임 이론에 관하여)에서 최소최대 정리를 증명하였다. 이 연구의 동기는 포커였는데, 포커에서 블러핑은 무작위화의 한 형태이다. 1944년에 그는 경제학자 오스카어 모르겐슈테른(Oskar Morgenstern)과 함께 *Theory of Games and Economic Behavior*(게임 이론과 경제 행동)를 펴내어 게임 이론을 하나의 학문 분야로 세웠다. 폰 노이만은 훗날, 최소최대 정리를 핵심 결과로 여겼으며 이 정리가 없었다면 출판할 가치가 있는 게임 이론은 없었을 것이라고 회고하였다.

사고 기술

무작위화로 상대의 이점을 없애라

어떤 고정된 선택이든 공략당할 수 있을 때에는, 상대가 자신의 선택지들 사이에서 차이를 느끼지 못하게 만드는 확률로 무작위로 선택하라.

대학에서는 게임 이론, 선형 계획법, 알고리즘 이론이 모두 이 아이디어를 사용한다. 예를 들어 무작위 알고리즘(randomised algorithm)은 어떤 적대자도 예측할 수 없는 무작위 선택을 함으로써 최악의 입력으로부터 스스로를 보호한다.

더 도전하기

1 **A의 최선의 대응.** B가 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$를 사용하면 A의 모든 전략에 대해 기대 상금이 0임을 보여라. B가 대신 $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$을 사용하면 A의 최선의 대응을 구하라.

2 **가위바위보.** 가위바위보의 보수 행렬을 쓰고 최적 전략을 구하라. 여기서는 왜 균등 전략이 최적이고, 이 문제에서는 그렇지 않은가?

3 **유일성.** 지지 않음을 보장하는 B의 전략은 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ **하나뿐**임을 보여라. (**힌트**: 그 전략은 X, Y, Z를 모두 음이 아니게 만들어야 한다. $X + 2Y + Z = 0$을 이용하라.)

4 **변형된 게임.** 차가 2일 때의 벌금이 £2가 아니라 £3이라면 최적 전략을 구하라.

더 읽을거리

- **Ken Binmore, *Playing for Real*** (Oxford University Press). 철저하면서도 친절한 게임 이론 입문서.
- **Avinash Dixit & Barry Nalebuff, *The Art of Strategy*** (Norton). 스포츠와 비즈니스의 예를 곁들인 게임 이론.
- **William Poundstone, *Prisoner's Dilemma*** (Anchor). 폰 노이만, 게임 이론, 그리고 냉전.

12.15 멸종인가 폭발인가

문제

STEP II 2003, 12번 문제

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

`step.maths.org/questions/03-s2-q12`

도입

1873년 골턴(Francis Galton)은 *Educational Times*에 다음과 같은 질문을 실었다. 왜 귀족 가문의 성(姓)은 그렇게 자주 사라지는가? 각 남자는 무작위 수의 아들을 두고, 아들 각각이 성을 물려준다. 각 남자가 평균적으로 한 명보다 많은 아들을 둔다고 해도, 그 성은 영원히 살아남을까? 이 질문에서 **분지 과정**(branching process)의 이론이 시작되었고, 오늘날 이 이론은 핵 연쇄 반응, 전염병의 확산, 세포 집단의 성장, 돌연변이의 생존을 모형화하는 데 쓰인다.

이 문제는 가장 단순한 모형을 다룬다. 각 입자는 0개, 1개 또는 2개의 입자로 갈라진다. 개체 수의 기댓값(expectation)은 기하급수적으로 늘어나거나 줄어든다. 그러나 기댓값이 이 이야기의 전부는 아니다. 개체 수의 기댓값이 폭발적으로 늘어날 때에도 혈통이 완전히 끊어질 확률은 양수이며, 아름다운 고정점(fixed point) 논증으로 이 확률을 구할 수 있다.

시작하기 전에

생각해 보기 어떤 단계에서 입자가 두 개 있다면, 다음 단계의 입자 수는 어떤 분포(distribution)를 따르는가?

두 입자는 각각 독립적으로(independently) $\mathrm{B}(2, p)$를 따르는 수만큼 자손을 낳으므로, 그 합은 $\mathrm{B}(4, p)$를 따른다. 즉 각각 확률 p로 성공하는 독립적인 "기회"가 네 번 있는 것이다.

차근차근 풀기

1단계 둘, 그리고 다시 둘

$\mathrm{P}(X_1 = 2) = p^2$이고, 이어서 $\mathrm{P}(X_2 = 2 \mid X_1 = 2) = \binom{4}{2}p^2q^2 = 6p^2q^2$이다. 따라서

$$\mathrm{P}(X_1 = 2, X_2 = 2) = 6p^4q^2. \qquad \blacksquare$$

2단계 거꾸로 추론하기

> **생각해 보기** $X_2 = 2$가 일어나는 경우는 몇 가지인가? (X_1은 어떤 값일 수 있는가?)

$X_2 = 2$가 되려면 $X_1 = 1$(확률 $2pq$)이고 그 뒤 자손이 2개(p^2)이거나, $X_1 = 2$이고 그 뒤 $\mathrm{B}(4, p)$에서 2가 나와야 한다. ($X_1 = 0$이면 아무것도 남지 않는다.) 따라서 $\mathrm{P}(X_2 = 2) = 2p^3q + 6p^4q^2$이고,

$$\mathrm{P}(X_1 = 2 \mid X_2 = 2) = \frac{6p^4q^2}{2p^3q + 6p^4q^2} = \frac{3pq}{1 + 3pq}.$$

이 값을 $\frac{9}{25}$와 같다고 놓으면 $75pq = 9 + 27pq$이므로 $pq = \frac{3}{16}$, 즉 $p(1-p) = \frac{3}{16}$이고, $p = \frac{1}{4}$ 또는 $p = \frac{3}{4}$를 얻는다. $\blacksquare$

3단계 개체 수의 기댓값

> **생각해 보기** $n-1$단계에 입자가 k개 있다면, n단계 입자 수의 기댓값은 얼마인가? 그다음 k에 대해 평균을 내어 보자.

각 입자는 평균 $\mathrm{E}(X) = 2p$개의 자손을 낳으므로, 입자 k개는 평균 $2pk$개의 자손을 낳는다. 즉 $\mathrm{E}(X_n \mid X_{n-1} = k) = 2pk$이다. k에 대해 평균을 내면 $\mathrm{E}(X_n) = 2p\,\mathrm{E}(X_{n-1})$이다. $\mathrm{E}(X_0) = 1$이므로

$$\mathrm{E}(X_n) = (2p)^n.$$

$p = \frac{1}{4}$이면 $(2p)^n = 2^{-n} \to 0$이고, $p = \frac{3}{4}$이면 $(2p)^n = \left(\frac{3}{2}\right)^n \to \infty$이다. $\blacksquare$

돌아보기

- **독립인 이항분포(binomial distribution)의 합.** 각각 $\mathrm{B}(2, p)$를 따르는 입자 k개의 자손 수를 모두 합하면 $\mathrm{B}(2k, p)$를 따른다.
- **시간을 거슬러 적용한 베이즈 정리(Bayes' theorem).** $\mathrm{P}(X_1 = 2 \mid X_2 = 2)$는 나중의 관측으로부터 이전 상태를 거꾸로 추론한 것이다.

▶ **조건부 기댓값(conditional expectation).** 바로 앞 세대에 대해 조건을 걸었더니 기댓값이 기하수열(geometric sequence)이 되었다.

숨은 수학

멸종 확률

$p = \frac{3}{4}$일 때 개체 수의 기댓값은 한없이 커진다. 그렇다면 개체군은 살아남는가? 꼭 그렇지는 않다. 첫 입자가 자손을 하나도 낳지 않으면(확률 $q^2 = \frac{1}{16}$) 혈통은 곧바로 끊긴다. 결국 멸종할 확률(extinction probability)을 e라 하자. 첫 입자가 자손 k개를 낳으면, 그 자손들 각각의 혈통은 독립적으로 확률 e로 끊어지므로

$$e = q^2 + 2pq\,e + p^2e^2 = (q + pe)^2 = \mathrm{G}(e),$$

이다. 여기서 $\mathrm{G}(s) = (q + ps)^2$은 자손 수의 **확률생성함수**(probability generating function)이다. 멸종 확률은 G의 고정점이며, 사실은 음이 아닌 고정점 가운데 **가장 작은** 것이다. $p = \frac{3}{4}$일 때 $(1 + 3e)^2 = 16e$에서 $9e^2 - 10e + 1 = 0$이므로 $e = \frac{1}{9}$ 또는 1이고, 멸종 확률은 $\frac{1}{9}$이다. $p = \frac{1}{4}$일 때 $[0, 1]$에 있는 고정점은 $e = 1$뿐이므로, 멸종은 확실하다(그림 12.15.1).

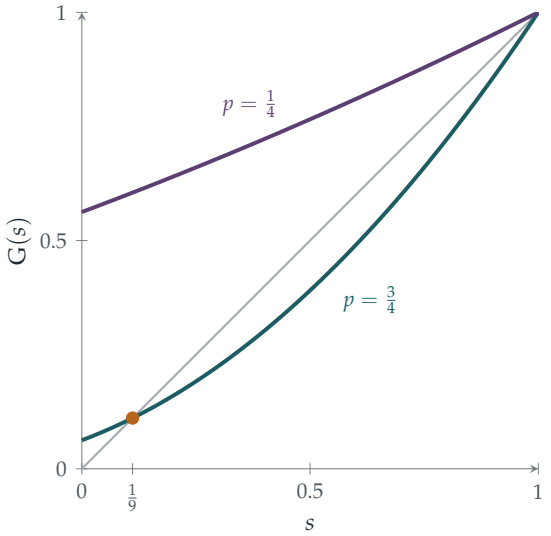


그림 12.15.1 자손 수의 생성함수 $\mathrm{G}(s) = (q + ps)^2$과 직선 $y = s$. 멸종 확률은 두 그래프가 만나는 가장 작은 점이다. $p = \frac{1}{4}$(자주색, 평균 $\frac{1}{2}$)이면 그 점은 $s = 1$이므로 멸종은 확실하다. $p = \frac{3}{4}$(청록색, 평균 $\frac{3}{2}$)이면 그 점은 $s = \frac{1}{9}$(호박색)이므로 개체군은 확률 $\frac{8}{9}$로 영원히 살아남는다.

임계 문턱

일반적으로 분지 과정은 자손 수의 평균이 1보다 작으면 확실히 멸종하고, 평균이 1과 같을 때에도 모든 개체가 정확히 한 명의 자손을 두는 자명한 경우(이때 혈통은 한 번에 한 명씩 그냥 영원히 이어진다)를 제외하면 확실히 멸종한다. 평균이 1보다 크면 양의 확률로 살아남는다. 자손 수의 평균은 역학(疫學)에서 말하는 **재생산수**(reproduction number) R이다. $R > 1$인 전염병은 크게 번질 수 있지만, 그런 경우에도 초기 발병이 우연히 사그라들 수 있으며, 그 확률은 바로 위의 고정점 방정식으로 주어진다. 핵물리학에서 연쇄 반응은 각 핵분열이 평균적으로 정확히 한 번의 핵분열을 더 일으킬 때 **임계**(critical) 상태라고 하며, 원자로를 운전하는 기술은 바로 그 상태를 유지하는 것이다.

골턴, 왓슨, 비에네메

골턴의 질문에는 1874년 Henry William Watson 목사가 답했지만, 그 답은 부분적으로 틀렸다. 그는 고정점 방정식을 찾아냈지만 두 번째 근을 놓치는 바람에 모든 성은 결국 사라진다고 결론지었다. 사실 올바른 분석은 이미 1845년에 프랑스의 통계학자 비에네메(Irénée-Jules Bienaymé)가 내놓았으나, 그의 연구는 1970년대까지 잊혀 있었다. 이 모형은 오늘날 **골턴–왓슨 과정**(Galton–Watson process)(때로는 비에네메–골턴–왓슨 과정)이라 불리며, 확률 과정을 다루는 모든 강의에서 핵심적인 예로 등장한다.

사고 기술

첫 세대에 대해 조건을 걸어라

스스로를 새롭게 반복하는 과정(각 자손이 전체 과정의 독립적인 복사본을 새로 시작하는 과정)은 첫 단계에 대해 조건을 걸어 얻은 방정식을 만족한다.

방정식 $e = \mathrm{G}(e)$는 고정점 문제이며, Section 2.2와 Section 8.7에서 만난 것과 같은 종류이다. 대학에서는 생성함수와 고정점이 분지 과정, 무작위 트리, 무작위 네트워크를 다루는 주요 도구이다.

더 도전하기

1 **일반적인 경우의 멸종.** 자손 수의 분포가 $\mathrm{B}(2, p)$일 때, 멸종 확률이 $p \le \frac{1}{2}$이면 1이고 $p > \frac{1}{2}$이면 $\left(\frac{q}{p}\right)^2$임을 보여라.

2 **분산.** $\mathrm{Var}(X_n)$이 $\mathrm{Var}(X_n) = (2p)^2\mathrm{Var}(X_{n-1}) + 2pq\,\mathrm{E}(X_{n-1})$을 만족함을 보이고, $p = \frac{1}{2}$일 때 분산(variance)을 구하여라.

3 **n세대까지의 멸종.** $e_n = \mathrm{P}(X_n = 0)$이라 하자. $e_n = \mathrm{G}(e_{n-1})$임을 보이고, $p = \frac{3}{4}$일 때 e_1, e_2, e_3을 계산하여라. 이 값들은 $\frac{1}{9}$에 가까워지는가?

4 **성씨.** 각 남자가 아들을 0명, 1명, 2명, 3명 둘 확률이 각각 $\frac{1}{4}$라면, 그의 성이 결국 사라질 확률은 얼마인가?

더 읽을거리

- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes*** (Oxford University Press). 분지 과정을 다루는 절.
- **Adam Kucharski, *The Rules of Contagion*** (Profile). 일반 독자를 위한 재생산수와 감염병 발생 이야기.
- **David Kendall, "Branching processes since 1873"**, *Journal of the London Mathematical Society* 41 (1966). 비에네메를 포함한 분지 과정의 역사.

12.16 앞으로 보고 거꾸로 보는 마르코프 연쇄

문제

STEP III 1996, 12번 문제

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

`step.maths.org/questions/96-s3-q12`

도입

엑스(Ecks) 교수의 정리에는 일정한 규칙이 있다. 다음에 무엇을 증명할지는 바로 직전에 무엇을 증명했는지에만 달려 있다. 미래가 현재에는 의존하지만 과거에는 의존하지 않는 과정을 **마르코프 연쇄**(Markov chain)라고 하며, 마르코프 연쇄는 어디에나 있다. 날씨 모형에도, 보드게임 뱀과 사다리(Snakes and Ladders)에도, DNA 모형에도, 구글의 초기 페이지랭크(PageRank) 알고리즘에도, 그리고 통계학자들이 복잡한 확률분포를 시뮬레이션할 때 쓰는 방법에도 있다.

이 문제는 장기적으로 정리의 유형이 어떤 비율로 섞이는지, 즉 **정상분포**(stationary distribution)를 구한다. 장기적으로 엑스 교수의 정리 여섯 개 중 하나는 새롭고 옳은 정리이고, 절반은 거짓이다. 이어서 문제는 연쇄를 시간을 거슬러 거꾸로 돌려 보고, 오늘의 정리가 주어졌을 때 어제의 정리는 무엇이었는지를 묻는다.

시작하기 전에

생각해 보기 "각 행의 합이 1이다"라는 조건은 무슨 뜻인가? 각 열의 합도 1인가?

i행은 현재 유형이 i일 때 다음 유형이 될 확률들을 나열한 것이다. 세 유형 중 하나는 반드시 일어나므로 각 행의 합은 1이다. 열은 그럴 필요가 없다. 둘째 열의 합은 정확히 1이지만, 첫째 열의 합은 0.6이고 셋째 열의 합은 1.4이다.

차근차근 풀기

1단계 한 단계

현재 정리의 유형 i에 대해 조건을 걸고 전확률 법칙(law of total probability)을 쓰면

$$b_j = \sum_i \mathrm{P}(\text{현재} i)\,\mathrm{P}(\text{다음} j \mid \text{현재} i) = a_1 p_{1j} + a_2 p_{2j} + a_3 p_{3j}. \qquad \blacksquare$$

행렬로 쓰면 행벡터(row vector)로 $\mathbf{b} = \mathbf{a}P$이다.

2단계 정상분포

> **생각해 보기** $a_1 + a_2 + a_3 = 1$이면서 $\mathbf{a} = \mathbf{a}P$인 것을 찾고자 한다. 방정식을 모두 써 보아라. 어느 것이 가장 간단한가?

방정식은 다음과 같다.

$$\begin{aligned} a_1 &= 0.3a_1 + 0.2a_2 + 0.1a_3, \\ a_2 &= 0.3a_1 + 0.4a_2 + 0.3a_3, \\ a_3 &= 0.4a_1 + 0.4a_2 + 0.6a_3. \end{aligned}$$

(이 방정식들은 독립이 아니다. 모두 더하면 양변이 모두 $a_1 + a_2 + a_3$이 된다.) 둘째 식에서 $0.6a_2 = 0.3(a_1 + a_3) = 0.3(1 - a_2)$이므로 $a_2 = \frac{1}{3}$이다. 첫째 식에서 $7a_1 = 2a_2 + a_3 = \frac{2}{3} + a_3$이고 $a_1 + a_3 = \frac{2}{3}$이므로 $8a_1 = \frac{4}{3}$, $a_1 = \frac{1}{6}$, $a_3 = \frac{1}{2}$이다.

$$(a_1, a_2, a_3) = \left(\tfrac{1}{6}, \tfrac{1}{3}, \tfrac{1}{2}\right). \qquad \blacksquare$$

3단계 연쇄를 거꾸로 돌리기

> **생각해 보기** $q_{ij} = \mathrm{P}(\text{직전} i \mid \text{현재} j)$이다. 직전 정리와 현재 정리 모두에 정상분포를 적용하여 베이즈 정리(Bayes' theorem)를 써라.

베이즈 정리에 의해

$$q_{ij} = \frac{\mathrm{P}(\text{직전} i)\,\mathrm{P}(\text{현재} j \mid \text{직전} i)}{\mathrm{P}(\text{현재} j)} = \frac{a_i p_{ij}}{a_j}.$$

예를 들어 $q_{11} = \frac{\frac{1}{6}\times 0.3}{\frac{1}{6}} = 0.3$이고 $q_{21} = \frac{\frac{1}{3}\times 0.2}{\frac{1}{6}} = 0.4$이다. 모두 모으면 ($i$는 행, j는 열을 나타낸다)

$$(q_{ij}) = \begin{pmatrix} 0.3 & 0.15 & \frac{2}{15} \\ 0.4 & 0.4 & \frac{4}{15} \\ 0.3 & 0.45 & 0.6 \end{pmatrix}. \qquad \blacksquare$$

이번에는 각 열의 합이 1이다. j열이 직전 정리가 될 수 있는 유형들을 나열하기 때문이다.

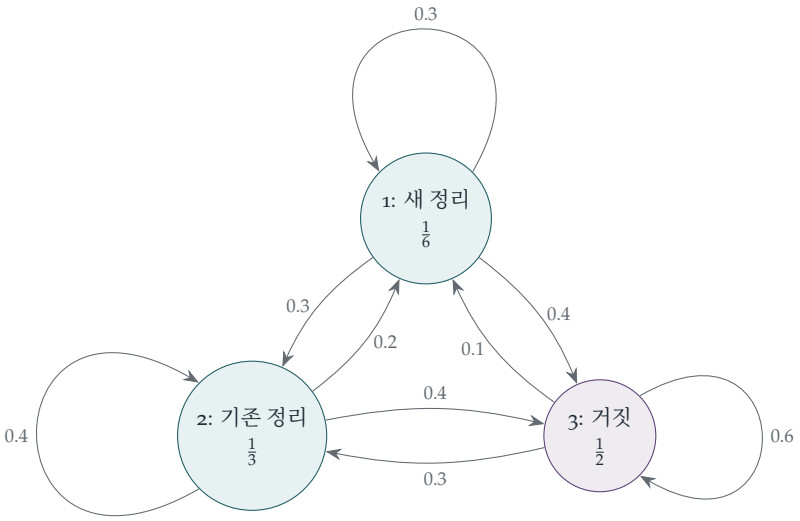


그림 12.16.1 마르코프 연쇄로 나타낸 엑스 교수의 정리. 화살표는 다음 정리가 각 유형일 확률을 나타낸다. 원 안의 분수는 정상분포이다. 장기적으로 정리 여섯 개 중 하나는 새롭고 옳으며, 세 개 중 하나는 이미 알려져 있고, 절반은 거짓이다.

돌아보기

- **행렬로 본 전확률 법칙.** 연쇄의 한 단계는 행벡터에 행렬 P를 곱하는 것이다.
- **정상성은 고유벡터이다.** $\mathbf{a} = \mathbf{a}P$는 $\mathbf{a}$가 고윳값(eigenvalue) 1에 대한 P의 좌고유벡터(left eigenvector)라는 뜻이다.
- **베이즈 정리로 거꾸로 돌리기.** 거꾸로 가는 확률들은 순방향 행렬과 정상분포로부터 만들어지는 또 하나의 확률행렬(stochastic matrix)을 이룬다.

숨은 수학

평형으로의 수렴

엑스 교수가 어떤 정리에서 출발하든, n개의 정리를 더 증명한 뒤 유형의 분포는 $\mathbf{a}_0 P^n$이다. 이 연쇄처럼 모든 유형 다음에 모든 유형이 올 수 있는 연쇄에서는, 출발점에 관계없이 P^n이 모든 행이 정상분포 $(\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$와 같은 행렬로 수렴(converge)한다. 수렴 속도는 P의 나머지

고윳값들(여기서는 0.2와 0.1)이 결정하며, 이들의 거듭제곱은 기하급수적으로 줄어든다. 단 몇 개의 정리만 지나도 엑스 교수의 과거는 잊혀진다. 마르코프 연쇄에 대한 이 **에르고딕 정리**(ergodic theorem)는 대학 2학년 확률론 과정의 핵심 결과이며, 마르코프 연쇄 몬테카를로(Markov chain Monte Carlo) 방법이 작동하는 근거이다. 복잡한 분포에서 표본을 뽑으려면, 그 분포를 정상분포로 갖는 마르코프 연쇄를 설계하여 오랫동안 돌리면 된다.

가역성

거꾸로 가는 연쇄 (q_{ij})는 순방향 연쇄 (p_{ij})와 다르다. 엑스 교수의 정리를 오랫동안 기록한 것을 보면 시간의 방향을 알아낼 수 있다는 뜻이다. 거꾸로 가는 연쇄와 순방향 연쇄가 일치하는 연쇄, 즉 모든 i, j에 대해 $a_i p_{ij} = a_j p_{ji}$인 연쇄를 **가역적**(reversible)이라고 한다. 네트워크 위의 무작위 행보(random walk)와 열평형 상태에 있는 물리계의 모형이 여기에 속한다. 가역성은 정상분포를 쉽게 찾게 해 주므로 강력한 성질이다.

마르코프와 푸시킨

안드레이 마르코프(Andrey Markov)는 1906년에 자신의 연쇄를 도입했는데, 큰 수의 법칙(law of large numbers)에 독립성(independence)이 꼭 필요하지는 않다는 것을 보이려는 것이 한 가지 이유였다. 이는 독립성이 필요하다고 주장한 수학자 Pavel Nekrasov와의 논쟁에서 나온 것이다. 1913년에 마르코프는 이 아이디어를 푸시킨(Pushkin)의 운문 소설 **예브게니 오네긴**(Eugene Onegin)의 처음 20 000개 글자에 적용했다. 그는 각 글자를 모음과 자음으로 분류하고, 각 글자의 유형이 바로 앞 글자의 유형에 의존하며 그 방식이 두 상태 연쇄로 잘 설명된다는 것을 보였다. 이것은 최초의 텍스트 통계 분석이었으며, 오늘날 모든 예측 입력 시스템의 조상이다.

사고 기술

확률은 벡터로, 전이는 행렬로

어떤 계가 고정된 확률로 상태 사이를 옮겨 다닐 때, 분포는 벡터로, 전이는 행렬로 써라. 그러면 장기적인 거동은 고유벡터에 관한 문제가 된다.

대학에서는 이를 통해 7장의 선형대수가 확률론으로 들어온다. 페론–프로베니우스 정리(Perron–Frobenius theorem)는 정상분포의 존재를 보장하고, 나머지 고윳값들은 그 분포에 얼마나 빨리 도달하는지를 잰다.

더 도전하기

1 **토끼 굴.** STEP III 2003, 13번 문제 (`step.maths.org/questions/03-s3-q13`)에 도전해 보아라. 각 꼭짓점에 출구가 있는 정사각형의 꼭짓점 위를 움직이는 무작위 행보를 첫 단계 방정식으로 푸는 문제이다.

2 **두 단계.** 엑스 교수의 마지막 정리가 새롭고 옳은 정리였다면, 그다음다음 정리가 세 유형 각각일 확률을 구하여라.

3 **고윳값.** P의 고윳값이 1, 0.2, 0.1임을 보여라.

4 **가역적인가?** 엑스 교수의 연쇄에 대해 $a_i p_{ij} = a_j p_{ji}$가 성립하는지 확인하여라. 그 답은 무엇을 말해 주는가?

더 읽을거리

- **James Norris, *Markov Chains*** (Cambridge University Press). 학부 과정의 표준 교재이다.
- **Olle Häggström, *Finite Markov Chains and Algorithmic Applications*** (Cambridge University Press). 짧으며, 마르코프 연쇄 몬테카를로를 잘 다룬다.
- **Brian Hayes, "First links in the Markov chain"**, *American Scientist* 101 (2013). 마르코프의 푸시킨 분석을 다시 들려준다.

12.17 바나흐의 성냥갑

문제

STEP II 1992, 16번 문제

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

`step.maths.org/questions/92-s2-q16`

도입

전해지는 이야기에 따르면 수학자 바나흐(Stefan Banach)는 담배를 몹시 많이 피웠는데, 양쪽 주머니에 성냥갑을 하나씩 넣고 다녔다. 성냥이 필요할 때마다 그는 주머니 하나를 무작위로 골랐다. 어느 날 그는 성냥갑 하나가 비어 있음을 알게 되었다. 다른 성냥갑에는 성냥이 몇 개 남아 있었을까? 이것이 바로 슈타인하우스(Hugo Steinhaus)가 바나흐를 기리며 고안한 **바나흐의 성냥갑 문제**이며, 이 문제는 성냥 대신 사탕이 등장하는 택시 운전사 버전이다.

미묘한 점은 "발견한다"라는 말에 있다. 운전사는 봉지에서 마지막 사탕을 꺼낼 때 봉지가 비었다는 것을 알아채지 못하고, 다음에 그 봉지에 손을 뻗을 때 비로소 알게 된다. 따라서 문제의 사건은 다음과 같다. 토피 봉지에 손을 뻗었는데, n개의 토피가 이미 모두 없어진 상태이다. 이런 일이 N번째 손을 뻗을 때 일어나는 경우의 수를 세는 것은 이항분포(binomial distribution)의 깔끔한 연습 문제이며, 바나흐 문제의 답에는 놀라운 특징이 있다. 다른 상자에 남은 성냥의 수로 가장 가능성이 높은 값은 0이다(1과 공동 1위이다).

시작하기 전에

생각해 보기 운전사가 N번째로 손을 뻗을 때 토피가 다 떨어졌음을 발견하려면, N번째에는 어느 봉지에 손을 뻗어야 하는가? 그 이전의 $N-1$번에는 무엇이 들어 있어야 하는가?

N번째에는 토피 봉지에 손을 뻗어야 하고, 그 이전의 $N-1$번에서는 정확히 n개의 토피(전부)와 $N-1-n$개의 박하사탕을 꺼냈어야 한다. 또한 그보다 먼저 박하사탕 봉지가 빈 것

을 발견하지 않았어야 한다.

차근차근 풀기

1단계 토피가 다 떨어졌음을 발견하기

> ? **생각해 보기** 처음 $N-1$번의 시행은 어떤 순서로 일어날 수 있는가? $N-1-n$에 대한 어떤 조건이 있어야 박하사탕 봉지가 빈 것을 한 번도 발견하지 않게 되는가?

토피를 꺼낸 n번은 처음 $N-1$번 가운데 어느 n번이어도 된다. 박하사탕을 최대 m개 꺼냈다면, 즉 $N-1-n \leq m$이면, 박하사탕이 떨어진 뒤에 박하사탕 봉지에 손을 뻗은 적은 없다. 따라서 $n+1 \leq N \leq n+m+1$일 때

$$\mathrm{P}(T_N) = \binom{N-1}{n} p^{n+1} q^{N-1-n},$$

이다. 여기서 T_N은 "N번째로 손을 뻗을 때 토피가 다 떨어졌음을 발견한다"는 사건이다(인수 p^{n+1}에는 N번째 시행 자체가 포함되어 있다).

2단계 박하사탕이 다 떨어졌음을 발견하기

역할을 바꾸어 같은 논증을 하면, $m+1 \leq N \leq n+m+1$일 때

$$\mathrm{P}(M_N) = \binom{N-1}{m} q^{m+1} p^{N-1-m}.$$

3단계 조건부 확률

> ? **생각해 보기** N번째 시행에서 발견이 일어났다는 사실이 주어졌다. 그것이 토피였을 확률은 얼마인가?

사건 T_N과 M_N은 배반 사건(exclusive events)이므로

$$\mathrm{P}(\text{토피} \mid \text{발견 시점} N) = \frac{\binom{N-1}{n} p^{n+1} q^{N-1-n}}{\binom{N-1}{n} p^{n+1} q^{N-1-n} + \binom{N-1}{m} q^{m+1} p^{N-1-m}},$$

이다. 단, 이항계수(binomial coefficient)는 위의 범위 밖에서는 0으로 본다. ($N \leq m$이면 토피만 떨어졌을 수 있으므로 확률은 1이다. 마찬가지로 $N \leq n$이지만 $N > m$이면 확률은 0이다.) ■

잠깐, 직접 해 보기 $n = m$이고 $p = q = \frac{1}{2}$일 때 이 공식을 확인해 보아라. 대칭성에 의해 가능한 모든 N에 대해 답은 $\frac{1}{2}$이어야 한다. 실제로 그러한가?

돌아보기

- **사건을 정확히 읽는다.** "발견한다"는 마지막 사탕을 꺼내는 것이 아니라 빈 봉지에 손을 뻗는 것을 뜻했다.
- **마지막 단계를 고정하고 나머지를 센다.** N번째 시행을 고정했고, 처음 $N-1$번은 어떤 순서여도 되므로 이항계수가 나왔다.
- **경계를 확인한다.** 제약 조건 $N-1-n \le m$과 $N-1-m \le n$이 불가능한 진행 과정을 걸러 냈다.

숨은 수학

바나흐의 성냥갑

바나흐의 버전에서는 각 성냥갑에 처음에 성냥이 n개씩 들어 있고, 두 주머니가 선택될 가능성은 같다. 그가 처음으로 빈 성냥갑을 발견했을 때 다른 성냥갑에 남은 성냥의 수를 K라 하자. $K = k$이려면, 빈 성냥갑으로 한 번 더 손을 뻗기 전에 빈 성냥갑에서 n개, 다른 성냥갑에서 $n-k$개를 어떤 순서로든 꺼냈어야 한다. 또한 어느 성냥갑이든 빈 쪽이 될 수 있다. 따라서

$$\mathrm{P}(K=k) = 2\binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k+1} = \binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k}.$$

이 확률들은 k가 커질수록 감소하는데, 다만 $\mathrm{P}(K=0) = \mathrm{P}(K=1)$이다(그림 12.17.1). 남은 성냥 수의 기댓값(expected value)은 약 $2\sqrt{n/\pi} - 1$이다. 각 성냥갑에 성냥이 50개씩 있으면 약 7이다. 여기서 $\sqrt{n}$은 $2n$번의 무작위 선택 뒤 두 성냥갑 사이에 생기는 불균형의 전형적인 크기로, 생일 문제(Section 9.9)와 무작위 행보(random walk)에서 나오는 것과 같은 제곱근 법칙이다.

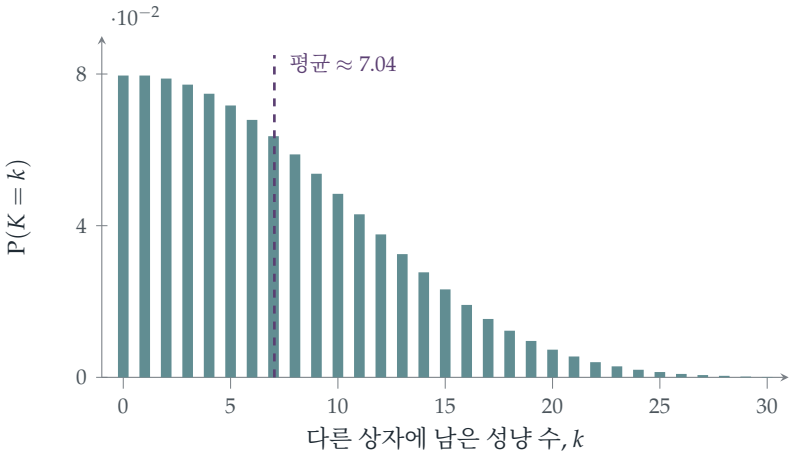


그림 12.17.1 각 상자에 성냥이 50개씩 있는 바나흐의 성냥갑 문제: 한 상자가 처음으로 빈 것이 발견되었을 때 다른 상자에 남은 성냥 수의 분포. 가장 가능성이 높은 값은 0과 1로 서로 같지만, 평균은 약 7이다.

음이항분포

각각 성공 확률이 p인 독립 시행에서 r번째 성공을 얻을 때까지 필요한 시행 횟수는 **음이항분포**(negative binomial distribution)를 따른다. 즉 $\mathrm{P}(N = k) = \binom{k-1}{r-1} p^r q^{k-r}$이다. 택시 운전사 문제는 두 음이항 과정(토피와 박하사탕) 사이의 경주이며, 각 과정은 사탕 수보다 한 번 더 많은 "성공"을 필요로 한다. 음이항분포는 보험(청구 건수), 생태학(무리 지어 분포하는 동물의 개체 수), 유전학 등에 나타나며, 흔히 푸아송 분포(Poisson distribution)보다 유연한 대안으로 쓰인다.

바나흐와 스코틀랜드 카페

함수해석학(functional analysis)의 창시자 가운데 한 사람인 바나흐는 르부프(Lwów, 지금의 리비우(Lviv))의 스코틀랜드 카페에서 많은 수학 연구를 했다. 그곳에서 그와 동료들은 웨이터가 보관하던 공책인 **스코틀랜드 책**(Scottish Book)에 문제를 적어 두었고, 상품은 맥주 한 잔부터 살아 있는 거위 한 마리까지 다양했다. 바나흐의 스승인 슈타인하우스는 바나흐의 흡연을 기리는 연설을 위해 성냥갑 문제를 고안했다. 이 문제를 처음으로 출판물에서 분석한 사람은 펠러(William Feller)로, 1950년에 나온 그의 유명한 확률론 교과서에서였다.

사고 기술

마지막 사건을 고정하고, 그곳에 이르는 경로를 센다

어떤 과정이 특정한 시점에 특정한 방식으로 끝날 확률을 구하려면, 마지막 단계에서 일어나는 일을 고정하고 그곳에 이르는 진행 과정의 수를 센다.

대학에서 이것은 음이항분포, 무작위 행보의 첫 도달 시간(first-passage time), 그리고 Section 9.10의 반사 원리(reflection principle)의 기초가 된다.

더 도전하기

1 **작은 경우.** $n = m = 1$이고 $p = \frac{1}{2}$일 때 가능한 순서를 모두 나열하고 N의 분포를 구하여라.

2 **바나흐의 공식.** $n = 1$과 $n = 2$일 때 $\sum_{k=0}^{n} \binom{2n-k}{n} 2^{-(2n-k)} = 1$임을 확인하여라.

3 **같은 꼭대기.** 바나흐 문제에서 모든 n에 대해 $\mathrm{P}(K = 0) = \mathrm{P}(K = 1)$임을 보여라.

4 **어느 쪽이 먼저 떨어지는가?** 택시 운전사 문제에서 $n = m = 2$이고 $p = \frac{2}{3}$일 때, (언제든) 처음으로 빈 것이 발견되는 봉지가 토피 봉지일 확률을 구하여라.

더 읽을거리

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1** (Wiley), 제VI장. 바나흐 문제를 세상에 처음 소개한 책이다(확률론 입문과 그 응용).
- **R. Daniel Mauldin (ed.), *The Scottish Book*** (Birkhäuser). 스코틀랜드 카페의 문제들을 해설과 함께 실었다.
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University Press). 음이항분포를 다룬다(확률론 입문).

12.18 무기억성

문제

STEP III 2021, 11번 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`

도입

어떤 방사성 원자가 백만 년 동안 붕괴하지 않은 채 실험실에 놓여 있었다고 하자. 이 원자는 이제 곧 붕괴할 "때가 된" 것일까? 그렇지 않다. 다음 1초 동안 이 원자가 붕괴할 확률(probability)은 방금 만들어진 원자의 경우와 정확히 같다. 방사성 원자는 나이를 먹지 않는다. 붕괴할 때까지의 대기 시간은 **지수분포**(*exponential distribution*)를 따르며, 이 분포를 규정하는 성질은 바로 기억이 없다는 것이다.

이 문제는 무기억성을 인상적인 형태로 보여 준다. 지수분포를 따르는 대기 시간을 정수 단위 부분과 소수 부분으로 나누어 보자. 놀랍게도 이 두 조각은 서로 **독립**(*independent*)이다. 즉 몇 시간을 꽉 채워 기다렸는지 안다고 해서 나머지 분에 대해 알 수 있는 것은 아무것도 없다. 또한 이 문제는 연습문제에서, 버스를 기다리는 시간이 왜 버스 사이의 간격보다 길게 느껴질 수 있는지도 설명해 준다.

시작하기 전에

생각해 보기 $x \geq 0$일 때 $\mathrm{P}(X > x)$는 무엇인가? $\mathrm{P}(X > s + t \mid X > s)$는 무엇인가?

$\mathrm{P}(X > x) = \int_x^\infty \lambda e^{-\lambda u}\,\mathrm{d}u = e^{-\lambda x}$이다. 따라서 $\mathrm{P}(X > s + t \mid X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathrm{P}(X > t)$이다. 즉 이미 s만큼 기다렸더라도, 남은 대기 시간은 처음부터 새로 기다리는 경우와 같은 분포를 따른다. 이것이 **무기억성**(*memoryless property*)이다.

차근차근 풀기

1단계 정수 부분

$Y = n$인 것은 정확히 $n \le X < n+1$일 때이다.

$$\mathrm{P}(Y = n) = e^{-\lambda n} - e^{-\lambda(n+1)} = (1 - e^{-\lambda})e^{-n\lambda}. \qquad \blacksquare$$

따라서 Y는 기하분포(geometric distribution)를 따른다. 정수 단위 하나하나를 확률 $e^{-\lambda}$로 "버텨 내는" 셈이다.

2단계 소수 부분

> **생각해 보기** $Z < z$는 X가 구간 $[n, n+z)$ 중 하나에 놓인다는 뜻이다. 이 구간들의 확률을 모두 더하라.

$$\mathrm{P}(Z < z) = \sum_{n \ge 0} \left(e^{-\lambda n} - e^{-\lambda(n+z)}\right) = (1 - e^{-\lambda z}) \sum_{n \ge 0} e^{-\lambda n} = \frac{1 - e^{-\lambda z}}{1 - e^{-\lambda}}. \qquad \blacksquare$$

3단계 Z의 평균

> **생각해 보기** $[0,1]$에서 값을 갖는 확률변수에 대해 $\mathrm{E}(Z) = \int_0^1 \mathrm{P}(Z > z)\,dz$이다. 이를 이용하라.

$c = 1 - e^{-\lambda}$로 두면 $\mathrm{P}(Z > z) = 1 - \frac{1 - e^{-\lambda z}}{c}$이므로

$$\mathrm{E}(Z) = 1 - \frac{1}{c}\left(1 - \frac{1 - e^{-\lambda}}{\lambda}\right) = 1 - \frac{1}{c} + \frac{1}{\lambda} = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1 - e^{-\lambda}} = \frac{1}{\lambda} - \frac{1}{e^{\lambda} - 1}. \qquad \blacksquare$$

(λ가 작을 때, 즉 대기 시간이 길 때 이 값은 $\frac{1}{2}$에 가까워진다. 소수 부분이 거의 균등분포를 따르기 때문이다. λ가 클 때는 이 값이 $\frac{1}{\lambda}$에 가까운데, X가 거의 항상 1보다 작기 때문이다.)

4단계 독립성

$$\mathrm{P}(Y = n, z_1 < Z < z_2) = \mathrm{P}(n + z_1 < X < n + z_2) = e^{-\lambda n}\left(e^{-\lambda z_1} - e^{-\lambda z_2}\right).$$

> **생각해 보기** 소문항 (i)과 (ii)를 이용하여 $\mathrm{P}(Y = n)\,\mathrm{P}(z_1 < Z < z_2)$와 비교하라.

(i)과 (ii)에서 $\mathrm{P}(Y = n)\mathrm{P}(z_1 < Z < z_2) = (1 - e^{-\lambda})e^{-n\lambda} \cdot \frac{e^{-\lambda z_1} - e^{-\lambda z_2}}{1 - e^{-\lambda}}$이고, 이는 위의 값과 같다. 따라서 Y와 Z는 독립이다. $\blacksquare$

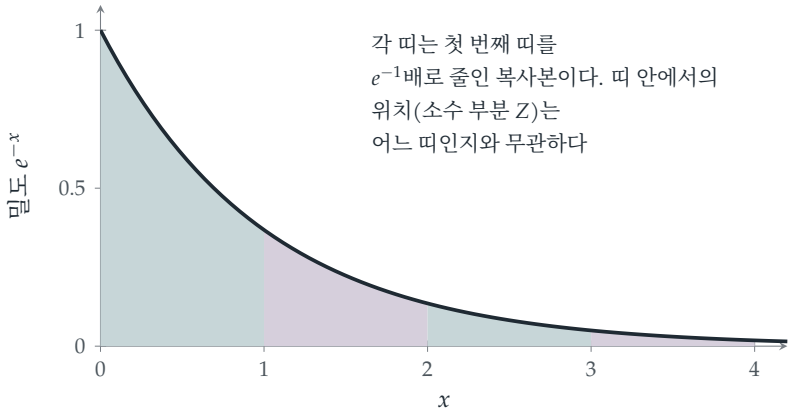


그림 12.18.1 $\lambda = 1$인 지수분포의 확률밀도함수를 폭이 1인 띠로 나눈 그림. $e^{-(n+z)} = e^{-n}e^{-z}$이므로 모든 띠는 같은 모양이다. X가 어느 띠에 떨어지는지(정수 부분 Y)와 그 띠 안의 어디에 떨어지는지(소수 부분 Z)는 서로 독립이다.

돌아보기

- **조각별로 더하기.** Z의 분포와 결합확률 모두 구간 $[n, n+1)$들에 대한 확률을 더해서 얻었다.
- **꼬리 공식.** $\mathrm{E}(Z) = \int \mathrm{P}(Z > z)\,\mathrm{d}z$를 이용하여 확률밀도함수를 구하지 않고 넘어갈 수 있었다.
- **인수분해는 곧 독립성.** 결합확률이 n의 함수와 z_1, z_2의 함수의 곱으로 분리되었다.

숨은 수학

지수분포가 특별한 이유

Y와 Z의 독립성은 $e^{-\lambda(n+z)} = e^{-\lambda n}e^{-\lambda z}$에서 곧바로 나오며, 이는 모습을 바꾼 무기억성이다. 사실 지수분포는 기억이 없는 **유일한** 연속분포이다. 모든 $s, t \geq 0$에 대해 $\mathrm{P}(X > s+t) = \mathrm{P}(X > s)\mathrm{P}(X > t)$이면, $g(t) = \ln \mathrm{P}(X > t)$는 $g(s+t) = g(s) + g(t)$를 만족하고, (g가 단조함수이면) 이로부터 반드시 $g(t) = -\lambda t$가 된다. 그래서 지수분포는 누적되는 것 없이 "무작위로" 일어나는 사건의 대기 시간을 모형화하는 데 쓰인다. 방사성 붕괴, 손님의 도착, 마모되지 않는 전자 부품의 고장이 그 예이다.

최솟값과 대기 시간의 역설

X_1과 X_2가 각각 비율 λ_1과 λ_2를 갖는 서로 독립인 지수분포 확률변수이면 $\mathrm{P}(\min > x) = e^{-\lambda_1 x} e^{-\lambda_2 x}$이므로, 최솟값은 비율 $\lambda_1 + \lambda_2$인 지수분포를 따른다. 서로 독립인 두 무작위 사건 중 먼저 일어나는 것을 기다리는 일은, 두 비율을 합친 비율로 일어나는 사건 하나를 기다리는 일과 같다(연습문제의 STEP III 2002 Q13). 무기억성은 **대기 시간의 역설**(*waiting-time paradox*)도 설명해 준다. 버스가 평균 간격 10분으로 무작위로 도착한다면, 무작위 시각에 도착한 승객은 평균 5분이 아니라 10분을 기다린다. 남은 대기 시간에는 기억이 없기 때문이다. 그리고 승객이 도착한 시점을 포함하는 간격의 길이는 평균 20분인데, 긴 간격일수록 그 안에 도착할 가능성이 크기 때문이다.

러더퍼드와 푸아송 과정

1910년에 어니스트 러더퍼드(Ernest Rutherford)와 한스 가이거(Hans Geiger)는 방사성 원천에서 방출되는 알파 입자의 수를 7.5초 간격으로 세었고, 그 개수가 푸아송(Poisson) 분포를 따른다는 것을 발견하였다. 수학자 Harry Bateman은 방출이 일정한 비율로 서로 독립적으로 일어난다면 바로 그렇게 되어야 함을 보였다. 이때 방출 사이의 시간 간격은 지수분포를 따른다. 지수분포를 따르는 간격으로 이어지는 사건들의 흐름인 **푸아송 과정**(*Poisson process*)은 오늘날 시간 속의 무작위성을 나타내는 기본 모형으로, 전화 통화, 보험금 청구, 신경세포의 발화 등에 쓰인다. 다음 절인 Section 12.19에서는 이를 공간으로 확장한다.

사고 기술

인수분해로 독립성 판정하기

두 확률변수의 결합확률이 한 변수의 함수와 다른 변수의 함수의 곱으로 인수분해되면 두 확률변수는 독립이다.

대학에서는 결합확률밀도함수의 인수분해가 독립성을 판정하는 표준적인 방법이며, 다변량 문제에서 독립인 성분을 알아보는 방법이기도 하다. 예를 들어 Section 12.22에서 정규분포를 따르는 점의 반지름과 각이 그러하다.

더 도전하기

1 **두 버스 노선.** STEP III 2002, 13번 문제 (`step.maths.org/questions/02-s3-q13`)에 도전해 보라. 두 지수분포 확률변수의 최솟값을 이용하여 다음 버스까지의 대기 시간을

구하는 문제이다.

2 **기억이 없는 이산분포.** 기하분포에 기억이 없음을, 즉 $\mathrm{P}(Y \geq m+n \mid Y \geq m) = \mathrm{P}(Y \geq n)$임을 보여라.

3 **대기 시간의 역설.** 버스가 비율 λ인 푸아송 과정으로 도착한다. (t가 클 때) 고정된 시각 t를 포함하는 간격의 기대 길이가 평균 간격의 두 배인 $\frac{2}{\lambda}$임을 보여라.

4 **Z의 확률밀도함수.** Z의 확률밀도함수를 구하고, $\lambda = 1$과 $\lambda = 5$일 때 그 그래프의 개형을 그려라.

더 읽을거리

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University Press). 지수분포와 푸아송 과정을 다룬다.
- **Sheldon Ross, *Introduction to Probability Models*** (Academic Press). 지수분포, 푸아송 과정, 대기행렬을 다룬다.
- **David Spiegelhalter, *The Art of Uncertainty*** (Pelican). 일상생활 속의 무작위성과 기다림을 다룬다.

12.19 평면 위의 무작위 점

문제

STEP II 2012, 13번 문제

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

`step.maths.org/questions/12-s2-q13`

도입

슈퍼마켓, 나무, 별, 휴대전화 기지국이 어떤 영역에 "무작위로" 흩어져 있다면, 가장 가까운 것까지는 얼마나 가야 할까? 답은 오직 밀도(density) k(단위 넓이당 평균 개수)에만 의존하며, **레일리 분포**(Rayleigh distribution)라는 특징적인 모양을 갖는다. 가장 가까운 것까지의 거리는 평균 $\frac{1}{2\sqrt{k}}$이다. 제곱킬로미터당 슈퍼마켓이 하나라면 500미터이다.

이 모형은 **공간 푸아송 과정**(spatial Poisson process)으로, 시간에 따른 무작위 도착(Section 12.18)의 2차원판이다. 계산은 짧고 아름답지만, 진짜 흥미로운 점은 이것이 기준점 역할을 한다는 데 있다. 생태학자들은 실제 최근접 이웃 거리(nearest-neighbour distance)를 이 공식과 비교하여 나무나 둥지가 뭉쳐 있는지, 무작위로 흩어져 있는지, 아니면 일부러 간격을 두고 있는지 판단한다.

시작하기 전에

생각해 보기 평균이 μ인 푸아송 확률변수(Poisson random variable)가 값 0을 가질 확률은 얼마인가?

$\mathrm{P}(0) = e^{-\mu}$이다. 반지름이 y인 원의 넓이는 πy^2이므로, 그 안에 있는 슈퍼마켓 수의 평균은 $k\pi y^2$이다.

차근차근 풀기

1단계 가까이에 슈퍼마켓이 없을 확률

거리 y 이내에 슈퍼마켓이 하나도 없을 확률은 $e^{-\pi k y^2}$이다.

2단계 가장 가까운 거리의 분포

> **생각해 보기** $Y \geq y$는 거리 y 이내에 슈퍼마켓이 없다는 뜻이다. 그렇다면 $\mathrm{P}(Y < y)$는 무엇인가?

$\mathrm{P}(Y < y) = 1 - e^{-\pi k y^2}$이고, 이를 미분하면 확률밀도함수

$$\mathrm{f}(y) = 2\pi k y e^{-\pi k y^2}, \qquad y \geq 0. \qquad \blacksquare$$

를 얻는다.

3단계 평균

> **생각해 보기** $y\,\mathrm{f}(y)$를 부분적분(integration by parts)하거나, $\mathrm{E}(Y) = \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y$를 이용한다. 그런 다음 치환하여 주어진 가우스 적분(Gaussian integral)을 쓴다.

$$\mathrm{E}(Y) = \int_0^\infty e^{-\pi k y^2}\,\mathrm{d}y.$$

$x = y\sqrt{2\pi k}$로 치환하면 $\pi k y^2 = \frac{1}{2}x^2$이 되므로 $\mathrm{E}(Y) = \frac{1}{\sqrt{2\pi k}}\sqrt{\frac{\pi}{2}} = \frac{1}{2\sqrt{k}}$이다. $\blacksquare$

4단계 분산

$$\mathrm{E}(Y^2) = \int_0^\infty y^2 \cdot 2\pi k y e^{-\pi k y^2}\,\mathrm{d}y.$$

$u = \pi k y^2$로 치환하면($\mathrm{d}u = 2\pi k y\,\mathrm{d}y$) 이 값은 $\int_0^\infty \frac{u}{\pi k} e^{-u}\,\mathrm{d}u = \frac{1}{\pi k}$이다. 따라서 분산(variance)은

$$\mathrm{Var}(Y) = \frac{1}{\pi k} - \frac{1}{4k} = \frac{4-\pi}{4\pi k}. \qquad \blacksquare$$

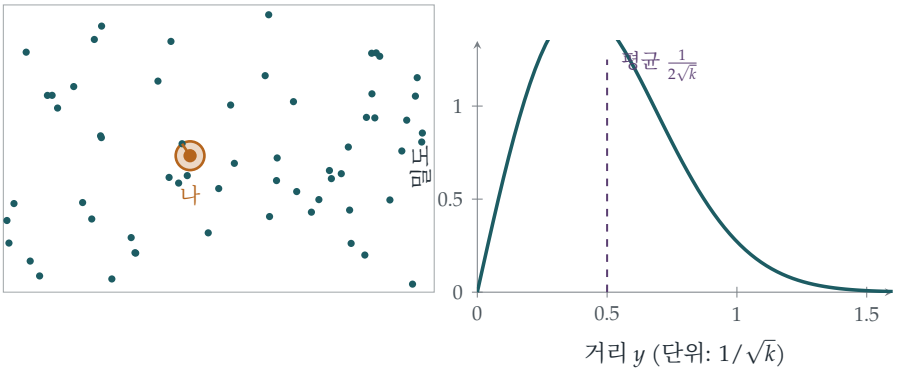


그림 12.19.1 왼쪽: 넓이가 24인 영역에 무작위로 흩어진 점 60개($k = 2.5$)와, 선택한 한 점에서 그중 가장 가까운 점까지의 거리(호박색). 오른쪽: 그 거리의 레일리 밀도 $2\pi k y e^{-\pi k y^2}$. 거리는 $1/\sqrt{k}$를 단위로 잰다. 평균은 $\frac{1}{2\sqrt{k}}$이다.

돌아보기

- **거리에 관한 질문을 개수에 관한 질문으로 바꾼다.** "가장 가까운 것이 y보다 멀다"는 "y 이내에 하나도 없다"와 같다.
- **누적분포함수를 먼저, 밀도는 나중에.** $\mathrm{P}(Y < y)$는 바로 구해졌고, 이를 미분하여 밀도를 얻었다.
- **알려진 적분으로 치환한다.** 각 적률(moment)은 가우스 적분이나 지수함수의 적분으로 귀결되었다.

숨은 수학

평면 위의 푸아송 과정

강도(intensity)가 k인 공간 푸아송 과정이란, 서로소인(disjoint) 영역들에 들어 있는 점의 개수가 서로 독립(independent)이고, 넓이가 A인 영역에 들어 있는 점의 개수가 평균 kA인 푸아송 분포를 따르는 무작위 점 배치이다. 이것은 매우 많은 점을 매우 넓은 영역에 독립적으로, 균일하게, 알맞은 평균 밀도로 떨어뜨렸을 때 얻어지는 것으로, 교란순열(derangement)(Section 9.8)을 설명하는 것과 같은 "희귀 사건의 법칙(law of rare events)"이다. 이는 완전한 공간적 무작위성을 나타내는 자연스러운 모형이며, 실제 패턴은 이것을 기준으로 평가된다. 두 생태학자가 1954년에 고안한 *Clark–Evans* **검정**(Clark–Evans test)은 관측된 평균 최근접 이웃 거리를 $\frac{1}{2\sqrt{k}}$와 비교한다. 이보다 훨씬 작으면 군집(씨앗이 어미 나무 근처에 떨어지는 경우)을, 훨씬 크면 규칙적인 간격(텃세를 부리는 새, 또는 식물들이 물을 두고 벌이는 경쟁)을 시사한다.

3차원에서

같은 계산을 3차원에서 하면, 가장 가까운 별, 분자, 은하까지의 거리에 대해 $\mathrm{P}(Y > y) = e^{-\frac{4}{3}\pi k y^3}$를 얻는다. 찬드라세카르(Subrahmanyan Chandrasekhar)는 1943년에 이를 이용하여 무작위로 배치된 이웃 별들이 한 별에 미치는 중력을 연구했고, 물리학자들은 이를 기체 분자의 평균 자유 행로(mean free path)에 사용한다.

레일리

밀도 $\frac{y}{\sigma^2} e^{-y^2/(2\sigma^2)}$는 레일리 경(Lord Rayleigh)의 이름을 딴 것이다. 그는 1880년에 소리의 이론에서 이 분포를 만났는데, 위상이 무작위인 많은 파동을 합한 것의 진폭을 나타낸다. 이 분포는 또한 두 좌표가 서로 독립인 정규확률변수(normal random variable)인 점과 원점 사이의 거리를 나타내기도 하는데, 이것이 이 분포가 여기에 나타나는 이유를 보는 또 다른 방법이다(연습문제 3). 오늘날에는 무작위 위상을 가진 여러 경로를 따라 수신기에 도달하는 휴대전화 신호의 페이딩(fading)을 모형화하는 데 쓰인다.

사고 기술

분포함수를 먼저 구하라

최솟값, 최댓값, 또는 가장 가까운 거리의 분포를 구할 때는 $\mathrm{P}(Y > y)$를 곧바로 적을 수 있는 경우가 많다. 이를 미분하여 밀도를 얻는다.

이 방법은 여러 무작위 수명 가운데 가장 긴 것을 다루는 Section 12.25에서 다시 등장한다. 대학에서는 모든 "순서통계량(order statistic)"을 다룰 때 표준적인 첫 단계이다.

더 도전하기

1 **중앙값.** 가장 가까운 슈퍼마켓까지 거리의 중앙값(median)을 구하고, 평균과 비교하여라.

2 **두 번째로 가까운 것.** 두 번째로 가까운 슈퍼마켓까지 거리의 분포를 구하여라. (힌트: $\mathrm{P}(Y_2 > y)$는 y 이내에 슈퍼마켓이 많아야 하나 있을 확률이다.)

3 **정규분포를 따르는 좌표.** U와 V가 서로 독립이고 $\mathrm{N}(0, \sigma^2)$를 따르면, $\sqrt{U^2 + V^2}$의 밀도가 $\frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}$임을 보여라. σ가 어떤 값일 때 슈퍼마켓 거리의 분포가 되는가?

4 Clark–Evans. 나무 400그루가 있는 1 km^2의 숲을 조사했더니, 나무에서 가장 가까운 이웃 나무까지의 평균 거리가 35 m였다. 이 결과는 완전한 공간적 무작위성과 부합하는가, 아니면 군집이나 규칙적인 간격을 시사하는가?

더 읽을거리

- **Peter Diggle, *Statistical Analysis of Spatial Point Patterns*** (Arnold). 공간 점 패턴의 통계적 분석에 관한 표준 참고서로, Clark–Evans 검정과 그 후속 방법들을 다룬다.
- **John Kingman, *Poisson Processes*** (Oxford University Press). 임의의 차원에서의 푸아송 과정을 다룬 아름다운 소책자이다.
- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes*** (Oxford University Press). 푸아송 과정에 관한 장을 보라.

12.20 뷔퐁의 바늘

문제

STEP I 1993, 16번 문제

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

`step.maths.org/questions/93-s1-q16`

도입

평행한 마룻널이 깔린 바닥에 바늘을 떨어뜨려 보자. 바늘이 널 사이의 틈에 걸칠 확률에는 π가 들어 있다. 어디에도 원은 보이지 않는데도 그렇다. 여러 번 던져서 교차한 횟수를 세면 실험으로 π를 추정할 수 있다. 이것이 **뷔퐁의 바늘**(*Buffon's needle*) 문제이다. 뷔퐁 백작 조르주루이 르클레르(Georges-Louis Leclerc, Comte de Buffon)가 1733년에 제기하고 1777년에 직접 풀었다. 이는 **기하학적 확률**(*geometric probability*)의 첫 번째 문제이며, 모든 몬테카를로 방법(Monte Carlo method)의 조상이다.

π는 각도에서 나온다. 바늘의 방향은 반 바퀴 위에서 균등하게 분포하며, 선들에 수직인 방향으로 드리운 바늘의 "그림자" 길이를 모든 방향에 대해 평균하면 $\frac{2}{\pi}$라는 인수가 나온다. 이 문제는 $\cos^{-1}$의 적분으로 이 계산을 수행한다.

시작하기 전에

생각해 보기 핀의 중심은 가장 가까운 선에서 거리 x만큼 떨어져 있고, 핀은 선들의 수선과 각 y를 이룬다. 핀은 선 쪽으로 얼마나 멀리 뻗어 있는가?

핀의 각 절반은 길이가 a이고, 선들에 수직인 성분은 $a\cos y$이다. 따라서 핀은 정확히 $a\cos y \geq x$일 때 가장 가까운 선과 교차한다. ($a < b$이므로 다른 선에는 닿을 수 없다.)

차근차근 풀기

1단계 적분

$t = \cos y$로 놓으면, t가 0에서 1까지 변할 때 $y = \cos^{-1} t$는 $\frac{\pi}{2}$에서 0까지 변하고, $\mathrm{d}t = -\sin y\,\mathrm{d}y$이다. 따라서

$$\int_0^1 \cos^{-1} t\,\mathrm{d}t = \int_0^{\pi/2} y \sin y\,\mathrm{d}y = \left[-y\cos y\right]_0^{\pi/2} + \int_0^{\pi/2} \cos y\,\mathrm{d}y = 0 + 1 = 1. \qquad \blacksquare$$

(또는 이렇게 볼 수도 있다. 이 적분은 $\cos^{-1}$의 그래프와 두 축 사이의 넓이이며, 이는 0에서 $\frac{\pi}{2}$까지 $y = \cos x$ 아래의 넓이를 $y = x$에 대해 대칭이동한 것과 같다.)

2단계 위치가 주어졌을 때

생각해 보기 x를 고정하면 어떤 각 y에서 교차가 일어나는가? 그 확률은 얼마인가?

$x \geq a$이면 핀은 선에 닿을 수 없다. $x < a$이면 교차하려면 $\cos y \geq \frac{x}{a}$, 즉 $y \leq \cos^{-1}\frac{x}{a}$이어야 한다. Y는 $[0, \frac{\pi}{2}]$에서 균등분포(uniform distribution)를 따르므로

$$\mathrm{P}(\text{교차} \mid X = x) = \frac{2}{\pi}\cos^{-1}\frac{x}{a} \qquad (0 \leq x < a). \qquad \blacksquare$$

3단계 위치에 대해 평균하기

X는 $[0, b]$에서 균등분포를 따르고 확률밀도(density)는 $\frac{1}{b}$이므로, $x = at$로 치환하면

$$\mathrm{P}(\text{교차}) = \frac{1}{b}\int_0^a \frac{2}{\pi}\cos^{-1}\frac{x}{a}\,\mathrm{d}x = \frac{2a}{\pi b}\int_0^1 \cos^{-1} t\,\mathrm{d}t = \frac{2a}{\pi b}, \qquad \blacksquare$$

이다.

잠깐, 직접 해 보기 바늘의 길이가 간격의 절반의 두 배, 즉 간격과 정확히 같다면($a = \frac{b}{2}$이므로 바늘의 길이 $2a$가 b와 같다) 교차 확률은 $\frac{1}{\pi}$이다. π를 소수점 아래 둘째 자리까지 추정하려면 몇 번이나 던져야 할까?

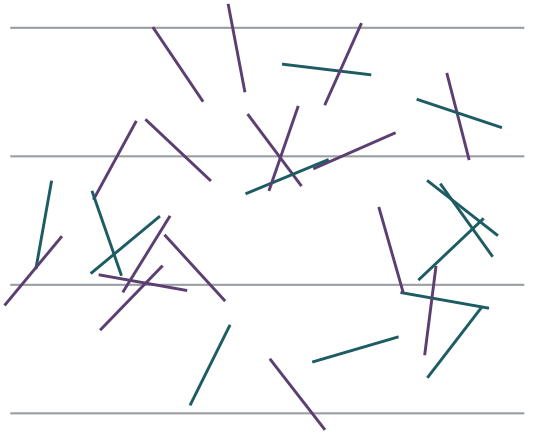


그림 12.20.1 간격이 $2b = 2$인 선들 위에 길이 $2a = 1.4$인 바늘 서른 개를 던졌다. 선과 교차하는 바늘은 자주색이다. 여기서는 기댓값 13.4개에 비해 17개가 교차했다. 표본이 작으면 추정이 거칠며, 뷔퐁의 바늘은 π를 계산하는 방법으로는 느리기로 유명하다.

돌아보기

- **두 개의 독립인 좌표.** 핀의 위치와 각도는 서로 독립(independent)이고 균등분포를 따르므로, 확률은 한 번에 한 변수씩 계산할 수 있는 이중적분(double integral)이었다.
- **조건을 걸고, 그다음 평균하기.** 먼저 x를 고정하고 각도에 대해 평균한 다음, x에 대해 평균한다.
- **치환이 상수를 드러낸다.** $x = at$로 척도를 바꾸자 답이 하나의 보편적인 적분으로 귀결되었다.

숨은 수학

뷔퐁의 국수

거의 계산 없이 답을 알아내는 방법이 있다. 바늘을 아무 모양으로나 구부려 "국수(noodle)"를 만들고, 그것이 선과 교차하는 횟수를 C라 하자. 기댓값의 선형성(linearity of expectation)에 의해 곡선의 교차 횟수의 기댓값은 그 작은 조각들의 기댓값의 합이므로 길이에 비례한다. 즉 어떤 상수 κ에 대해 $\mathrm{E}(C) = \kappa L$이다. κ를 구하기 위해 국수를 지름이 선의 간격 $2b$와 같은 원으로 잡자. 이 원은 항상 정확히 두 번 교차하고, 길이는 $2\pi b$이다. 따라서 $\kappa \cdot 2\pi b = 2$이고, $\kappa = \frac{1}{\pi b}$이다. 길이가 $2a < 2b$인 곧은 바늘은 많아야 한 번 교차하므로 $\mathrm{P}(\text{교차}) = \mathrm{E}(C) = \frac{2a}{\pi b}$이다. 이 논증은 Joseph-Émile Barbier(1860)가 제시한 것으로, 확률적 추론의 보석이라 할 만하다. π는 상수를 보정하기 위해 고른 원에서 나온다.

π 추정하기, 그리고 하지 말아야 할 방법

바늘 n개를 던져서 c개가 교차했다면 $\frac{c}{n} \approx \frac{2a}{\pi b}$이고 $\pi \approx \frac{2an}{bc}$이다. 오차는 $\frac{1}{\sqrt{n}}$ 정도로만 줄어들기 때문에, π를 소수점 아래 셋째 자리까지 구하려면 약 백만 번을 던져야 한다. 1901년 이탈리아의 수학자 Mario Lazzarini는 3408번 던져서 $\pi \approx \frac{355}{113} = 3.1415929\ldots$(소수점 아래 여섯째 자리까지 정확)를 얻었다고 주장했다. 통계학자들은 이 결과가 믿기에는 지나치게 좋다고 지적했다. 3408번 던지고 바늘을 신중하게 고르면 $\frac{355}{113}$은 정확히 1808번의 교차에서 나오며, 적절한 순간에 실험을 멈추면 이 값을 만들어 낼 수 있다. 이는 선택적 보고에 관한 고전적인 교훈담이다.

뷔퐁

뷔퐁은 박물학자로, 파리 왕립 식물원의 원장이었고 36권짜리 박물지를 썼다. 그는 바늘 문제(그리고 정사각형 타일을 쓰는 변형인 "franc-carreau" 게임)를 1733년 과학 아카데미(Académie des Sciences)에 발표했고, 1777년에 그 풀이를 출판했다. 그의 목적은 산술뿐 아니라 기하학도 확률에 쓰일 수 있음을 보이는 것이었다. 여기에서 무작위 직선과 그것이 곡선 및 곡면과 만나는 교점을 다루는 이론인 적분기하학(integral geometry)이 자라났고, 이는 오늘날 2차원 단면으로부터 3차원 구조(예를 들어 조직 속의 세포)의 크기와 모양을 추정하는 학문인 입체해석학(stereology)에 쓰인다.

사고 기술

영역 위의 적분으로서의 확률

결과가 서로 독립인 여러 연속확률변수에 의존할 때, 어떤 사건의 확률은 그 사건이 일어나는 영역 위에서 결합확률밀도(*joint density*)를 적분한 것이다.

대학에서는 이것이 연속확률의 기초가 된다. 결합확률밀도, 주변분포(marginal distribution), 변수변환(change of variable)이 그것이며, 이들은 모두 Section 12.21과 Section 12.22에서 다시 등장한다.

더 도전하기

1 **긴 바늘.** $a > b$이면 바늘이 두 개 이상의 선과 교차할 수 있다. 이때도 교차 횟수의 기댓값이 여전히 $\frac{2a}{\pi b}$임을 보여라.

2 **국수.** 길이가 $2a$인 바늘을 직각으로 구부린다(두 팔의 길이는 각각 a). 교차 횟수의 기댓값은 얼마인가?

3 **정사각형 타일.** 반지름이 r인 동전을 한 변의 길이가 $s > 2r$인 정사각형 타일이 깔린 바닥에 던진다. 동전이 한 타일 안에 완전히 놓일 확률을 구하여라.

4 **Lazzarini의 바늘.** $a = \frac{5}{6}b$일 때($\frac{2a}{\pi b} = \frac{5}{3\pi}$이다), 3408번 던져서 1808번 교차하면 추정값 $\pi \approx \frac{355}{113}$을 얻음을 보여라.

더 읽을거리

- **Daniel Klain & Gian-Carlo Rota, *Introduction to Geometric Probability*** (Cambridge University Press). 뷔퐁의 바늘에서 출발하는 기하학적 확률 입문서이다.
- **Lee Badger, "Lazzarini's lucky approximation of π"**, *Mathematics Magazine* 67 (1994). 의심스러운 실험의 통계학을 다룬다.
- **Paul Nahin, *Digital Dice*** (Princeton University Press). 확률 문제를 위한 몬테카를로 방법을 다룬다.

12.21 막대 부러뜨리기

문제

STEP III 1990, 16번 문제

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

`step.maths.org/questions/90-s3-q16`

도입

막대를 무작위로 고른 두 점에서 부러뜨린다. 세 조각으로 삼각형을 만들 수 있을까? 잘 알려진 답은 확률(probability) $\frac{1}{4}$이다. 그렇다면 삼각형이 만들어질 때, 그 삼각형은 예각삼각형(acute triangle)일까, 둔각삼각형(obtuse triangle)일까? 대개는 둔각삼각형이다. 만들어지는 삼각형 가운데 예각삼각형은 약 3분의 1에 불과하다. 무작위 도형에 관한 이런 질문들은 19세기부터 확률론 연구자들을 즐겁게 해 왔으며, 연속 확률(continuous probability)을 다루는 연습에 안성맞춤이다. 영역 안의 무작위 점, 부분 영역을 정하는 조건, 그리고 그 넓이를 구하는 적분이 모두 등장하기 때문이다.

이 문제는 막대를 부러뜨리는 서로 다른 두 가지 방법을 제시하고, 각 경우에 둔각이 생길 확률을 묻는다. 답인 $3-2\sqrt{2}$와 $3-4\ln 2$는 간단한 기하에서 나오는 뜻밖의 수이다.

시작하기 전에

생각해 보기 세 길이로 삼각형을 만들 수 있을 필요충분조건은 각 길이가 나머지 두 길이의 합보다 작은 것이다. 세 길이의 합이 L이라면, 이 조건은 가장 긴 길이에 대해 무엇을 말해 주는가? 변 c의 대각은 언제 둔각인가?

$a+b+c=L$일 때 $c<a+b$는 $c<\frac{L}{2}$를 뜻한다. 따라서 세 조각으로 삼각형을 만들 수 있는 것은 정확히 모든 조각이 전체 길이의 절반보다 짧을 때이다. 코사인 법칙(cosine rule)에 따라, c의 대각이 둔각인 것은 정확히 $c^2>a^2+b^2$일 때이다.

차근차근 풀기

1단계 한 번 자른 뒤 이등분하기

조각은 X, $\frac{1-X}{2}$, $\frac{1-X}{2}$이다. 길이가 같은 두 조각은 항상 $\frac{1}{2}$보다 작으므로, 삼각형이 만들어질 필요충분조건은 $X < \frac{1}{2}$이다. 이 삼각형은 이등변삼각형(isosceles triangle)이고 두 밑각은 예각이므로, 둔각삼각형이 되는 것은 X의 대각이 둔각일 때뿐이다.

$$X^2 > 2\Big(\frac{1-X}{2}\Big)^2 = \frac{(1-X)^2}{2}, \qquad \text{즉} \qquad \sqrt{2}X > 1-X, \quad X > \frac{1}{1+\sqrt{2}} = \sqrt{2}-1.$$

$X < \frac{1}{2}$라는 조건 아래에서 X는 $[0, \frac{1}{2}]$ 위의 균등분포(uniform distribution)를 따르므로

$$\mathrm{P}(\text{둔각} \mid \text{삼각형}) = \frac{\frac{1}{2} - (\sqrt{2}-1)}{\frac{1}{2}} = 3 - 2\sqrt{2} \approx 0.172. \qquad \blacksquare$$

2단계 두 번째 구성

> **생각해 보기** 꼭짓점이 $(1,0)$, $(1,1)$, $(0,1)$인 삼각형 안의 (x,y)에 대하여, 변의 길이가 x, y, $2-x-y$일 때 세 삼각부등식(triangle inequality)을 확인하라.

영역은 $\{x \le 1, y \le 1, x+y \ge 1\}$이다. 부등식은 $x+y \ge 2-x-y$, 즉 $x+y \ge 1$; $x+(2-x-y) \ge y$, 즉 $y \le 1$; 그리고 $y+(2-x-y) \ge x$, 즉 $x \le 1$이다. 이들은 모두 영역 안에서 성립한다(영역의 경계에서는 등호가 성립하며, 이때 삼각형은 퇴화(degenerate)한다). $\blacksquare$

> **생각해 보기** B에서의 각은 변 $CA = y$의 대각이다. 조건 $y^2 > x^2 + (2-x-y)^2$을 y에 대한 부등식으로 나타내라.

전개하면 $y^2 > x^2 + 4 + x^2 + y^2 - 4x - 4y + 2xy$이므로 $4y - 2xy > 2x^2 - 4x + 4$, 즉 $2y(2-x) > 2(x^2 - 2x + 2)$이다. $2 - x > 0$이므로

$$\angle ABC\text{가 둔각} \iff y > \frac{x^2 - 2x + 2}{2 - x}. \qquad \blacksquare$$

3단계 넓이

> **생각해 보기** $x^2 - 2x + 2$를 $2 - x$로 나누어라. 그다음 영역 가운데 곡선 위쪽 부분의 넓이를 구하고, 이를 영역의 넓이 $\frac{1}{2}$로 나누어라.

$x^2 - 2x + 2 = -x(2-x) + 2$이므로 곡선은 $y = -x + \frac{2}{2-x}$이다. 이 곡선은 $(0,1)$에서 $(1,1)$까지 이어지며, $0 < x < 1$에서 직선 $y = 1 - x$보다 위에, $y = 1$보다 아래에 있다(그

림 12.21.1). 둔각 영역은 곡선과 $y = 1$ 사이에 있고, 그 넓이는

$$\int_0^1 \left(1 + x - \frac{2}{2-x}\right) \mathrm{d}x = 1 + \frac{1}{2} + \left[2\ln(2-x)\right]_0^1 = \frac{3}{2} - 2\ln 2.$$

이를 $\frac{1}{2}$로 나누면 $\mathrm{P}(\angle ABC\text{가 둔각}) = 3 - 4\ln 2 \approx 0.227$이다. ■

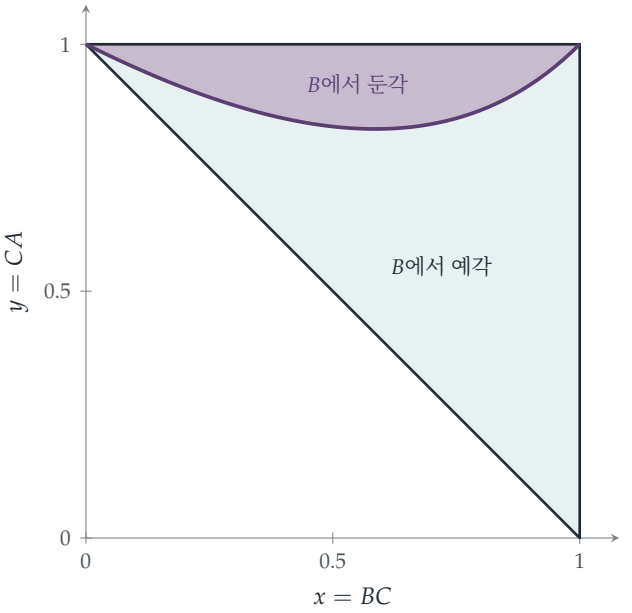


그림 12.21.1 소문항 (ii)에서 가능한 (x, y)가 이루는 삼각형. 곡선 $y = \frac{x^2-2x+2}{2-x}$ 위쪽(자주색 영역)에서 B에서의 각이 둔각이며, 이 영역은 삼각형의 $3 - 4\ln 2 \approx 0.23$만큼을 차지한다.

돌아보기

- **기하를 부등식으로 옮겨라.** "삼각형을 이룬다"와 "둔각이다"는 둘 다 길이 사이의 부등식이 되었다.
- **연속확률변수의 조건부 확률(conditional probability).** $X < \frac{1}{2}$라는 조건 아래에서 X는 더 작은 구간 위의 균등분포를 따랐다.
- **넓이의 비로서의 확률.** 영역 안에서 균등하게 고른 점에 대해, 확률은 넓이를 전체 넓이로 나눈 값이다.

숨은 수학

고전적인 부러진 막대 문제

막대를 서로 독립인 두 균등분포 점 U와 V에서 부러뜨린다. 순서쌍 (U, V)는 단위정사각형 위의 균등분포를 따르고, 세 조각이 삼각형을 이루는 것은 정확히 각 조각이 $\frac{1}{2}$보다 짧을 때이다. 이 조건은 넓이의 합이 $\frac{3}{4}$인 작은 삼각형 네 개를 잘라 내므로 확률은 $\frac{1}{4}$가 남는다. 삼각형이 만들어진다는 조건 아래에서 예각삼각형일 확률은 $\frac{1}{4}$ 가운데 $3\ln 2 - 2 \approx 0.079$, 즉 약 32%이다. 이 조각들은 **균등 무작위 분할**(*uniform random partition*)의 한 예이며, 그 결합분포는 **디리클레 분포**(*Dirichlet distribution*)이다. 디리클레 분포는 베이즈 통계학(Bayesian statistics)과 자원이 무작위로 나뉘는 방식을 다루는 모형에 등장한다.

막대를 부러뜨리는 방법이 중요하다

문제의 두 방법이 비슷하게 들리는 질문에 대해 서로 다른 답 $3 - 2\sqrt{2}$와 $3 - 4\ln 2$를 내며, 둘 다 고전적인 방법과도 다르다는 점에 주목하라. 베르트랑(Bertrand)의 현 문제(Section 12.3)에서처럼, 절차를 명시하기 전까지 "무작위"라는 말은 아무 의미가 없다.

무작위 삼각형

무작위 삼각형에 관한 문제는 빅토리아 시대에 인기 있던 퍼즐이었다. 루이스 캐럴(Lewis Carroll)의 *Pillow Problems*(1893)는 "무작위로" 고른 삼각형이 둔각삼각형일 확률을 묻는데, 그의 답도 이 절의 답들처럼 그가 택한 무작위성 모형에 전적으로 달려 있다. 무작위 삼각형에 관한 문제는 그 뒤로도 계속 놀라움을 낳았다. 예를 들어 정사각형 안에서 균등하게 고른 세 점이 둔각삼각형을 이룰 확률은 약 0.725인데, 이 값은 20세기에 들어서야 정확히 계산되었다.

사고 기술

표본공간(sample space)의 기하

무작위 결과를 영역 안의 점으로, 사건을 부분 영역으로 나타내고, 확률을 넓이의 비(또는 밀도(*density*)의 적분)로 계산하라.

대학에서 이 그림은 다변수 확률론(multivariate probability)의 토대이며, 좌표를 바꿀 때 밀도가 어떻게 변환되는지 알려 주는 변수변환 공식(change-of-variables formula)의 토대이기도 하다.

더 도전하기

1 **고전적인 문제.** 막대를 서로 독립인 두 균등분포 점에서 부러뜨릴 때, 조각들이 삼각형을 이룰 확률이 $\frac{1}{4}$임을 보여라.

2 **긴 조각을 부러뜨리기.** 막대를 균등분포 점에서 부러뜨린 다음, 더 긴 조각을 균등분포 점에서 부러뜨린다. 세 조각이 삼각형을 이룰 확률이 $2\ln 2 - 1 \approx 0.386$임을 보여라.

3 **예각인가 둔각인가.** 소문항 (ii)에서 삼각형 ABC가 C에서($AB = 2 - x - y$의 대각) 둔각을 가질 확률을 구하라. 이를 이용하여 삼각형이 어딘가에 둔각을 가질 확률을 구하라.

4 **가장 긴 조각의 기댓값.** 고전적인 부러진 막대 문제에서 가장 긴 조각의 길이의 기댓값(expected value)이 $\frac{11}{18}$임을 보여라.

더 읽을거리

- **Martin Gardner, *The Colossal Book of Mathematics*** (Norton). 부러진 막대 문제가 실려 있다.
- **Frederick Mosteller, *Fifty Challenging Problems in Probability*** (Dover). 확률에 관한 도전적인 문제 50선.
- **Lewis Carroll, *Pillow Problems*** (Dover, *A Tangled Tale*과 합본). 무작위 삼각형을 포함한 빅토리아 시대의 퍼즐들.

12.22 평균이 없는 분포

문제

STEP III 1995, 12번 문제

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

`step.maths.org/questions/95-s3-q12`

도입

측정을 여러 번 하고 평균을 내면 무작위 오차는 서로 상쇄되어, 평균은 참값으로 안정된다. 이 **큰 수의 법칙**(*law of large numbers*)은 모든 통계학의 토대이다. 그러나 여기에는 숨은 가정이 있다. 측정값의 평균(mean)이 유한해야 한다는 것이다. 이 문제는 서로 독립(independent)인 두 표준정규(standard normal) 확률변수의 비라는, 지극히 자연스러운 확률변수(random variable)를 만드는데, 이 확률변수는 평균이 아예 없다. 그 평균들은 결코 안정되지 않는다. 사실 이러한 확률변수 백만 개의 평균은 하나의 확률변수와 정확히 같은 분포(distribution)를 가진다.

이것이 **코시 분포**(*Cauchy distribution*)이다. 그 밀도(density)는 무해한 종 모양처럼 보이지만, 꼬리(tail)가 너무 두꺼워서 이따금 나타나는 엄청나게 큰 값이 어떤 평균이든 좌우해 버린다. 이 문제는 아름다운 기하학적 논증으로 이 분포를 찾아낸다. 좌표의 비는 직선의 기울기이고, 방사 대칭성 때문에 그 직선의 각은 균등분포를 따른다.

시작하기 전에

생각해 보기 (U, V)가 원점으로부터의 거리에만 의존하는 밀도를 가진 무작위 점이라면, 원점에서 그 점까지 이은 선분이 한 축과 이루는 각 Θ에 대해 무엇을 말할 수 있는가?

평면을 회전해도 밀도는 변하지 않으므로, 모든 방향이 똑같이 일어날 법하다. 즉 각 Θ는 $[0, 2\pi)$에서 균등분포를 따른다. 그리고 $\frac{U}{V}$는 거리가 아니라 방향에만 의존한다.

차근차근 풀기

1단계 결합 밀도

X와 Y가 독립이므로, 결합 밀도(joint density)는 각각의 밀도의 곱이다.

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}.$$

이 밀도는 원점으로부터의 거리의 제곱인 x^2+y^2를 통해서만 x와 y에 의존한다. 즉 원점을 중심으로 하는 원 위에서 밀도가 일정하다. 이것이 바로 "방사 대칭(radially symmetrical)"의 뜻이다.

2단계 대칭성 이용하기

> **생각해 보기** $\frac{U}{V} < k$는 V의 부호에 따라 두 가지 방식으로 일어난다. 180° 회전 $(u,v) \mapsto (-u,-v)$에 의해 이 둘은 어떻게 연결되는가?

$\frac{U}{V} < k$는 $V > 0$이고 $U < kV$이거나, $V < 0$이고 $U > kV$라는 뜻이다. 회전 $(u,v) \mapsto (-u,-v)$는 첫 번째 영역을 두 번째 영역으로 옮기고, 밀도를 보존한다. 따라서 두 영역의 확률은 같고,

$$P\left(\frac{U}{V} < k\right) = 2P(U < kV, V > 0). \qquad \blacksquare$$

3단계 각

> **생각해 보기** u를 가로축, v를 세로축으로 하는 평면에 영역 $\{V > 0, U < kV\}$를 그려 보자. 이것은 꼭짓점이 원점에 있는 쐐기 모양이다. 그 각은 얼마인가?

영역 $V > 0$은 위쪽 반평면이다. 원점을 지나는 직선 $u = kv$는 양의 v축과 $\arctan k$의 각을 이룬다($k > 0$일 때 양의 u 쪽으로 기운다). 위쪽 반평면 중 이 직선의 왼쪽 부분은 각이 $\frac{\pi}{2} + \arctan k$인 쐐기이다. 방사 대칭성에 의해 쐐기의 확률은 그 각을 2π로 나눈 값이므로,

$$P\left(\frac{U}{V} < k\right) = 2 \cdot \frac{\frac{\pi}{2} + \arctan k}{2\pi} = \frac{1}{2} + \frac{1}{\pi} \arctan k.$$

이를 k에 대해 미분하면 밀도

$$g(k) = \frac{1}{\pi(1+k^2)}. \qquad \blacksquare$$

를 얻는다.

돌아보기

- **독립인 정규분포는 방사 대칭이다.** 두 정규분포 밀도의 곱은 거리에만 의존한다.
- **비는 곧 방향이다.** $\frac{U}{V}$는 점의 각에만 의존하며, 그 각은 균등분포를 따른다.
- **쐐기의 확률.** 방사 대칭이 있으면 쐐기의 확률은 그 각에 비례한다.

숨은 수학

평균이 없다

코시 분포의 평균은 $\int_{-\infty}^{\infty} \frac{k}{\pi(1+k^2)}\,\mathrm{d}k$여야 하겠지만, $\int_0^K \frac{k}{1+k^2}\,\mathrm{d}k = \frac{1}{2}\ln(1+K^2) \to \infty$이다. 즉 이 적분은 수렴(converge)하지 않고, 평균은 존재하지 않는다. (대칭성 때문에 답이 0일 것 같지만, 양의 부분과 음의 부분이 모두 무한대이고 $\infty - \infty$는 값을 갖지 않는다.) 분산(variance)도 무한대이다.

더 나쁜 것은, 독립인 코시 확률변수 n개의 평균이 다시 정확히 같은 분포의 코시 분포를 따른다는 점이다. (이를 확인하는 한 가지 방법은 특성함수(characteristic function) $\mathrm{E}(e^{itX}) = e^{-|t|}$를 이용하는 것이다.) 따라서 평균을 내는 것은 전혀 도움이 되지 않는다. 백만 번 측정한 값의 평균이 한 번 측정한 값보다 조금도 더 믿을 만하지 않다. 그림 12.22.1는 무슨 일이 일어나는지 보여 준다. 정규분포 표본의 누적 평균(running average)은 안정되지만, 코시 분포 표본의 누적 평균은 이따금 나타나는 거대한 값에 흔들리며 끝없이 요동친다.

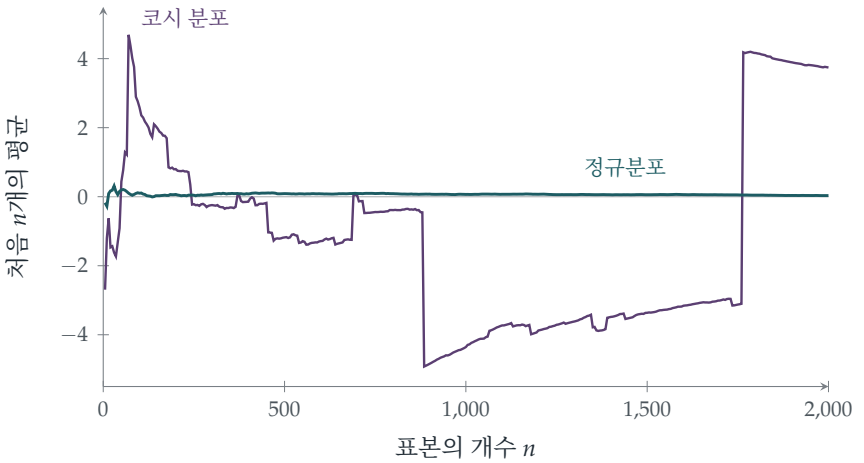


그림 12.22.1 독립인 표본 2000개의 누적 평균. 표준정규분포 표본(청록색)의 경우, 큰 수의 법칙이 예측하는 대로 평균이 0으로 안정된다. 코시 분포 표본(자주색)의 경우 평균은 결코 안정되지 않는다. 이따금 엄청나게 큰 값 하나가 평균을 새로운 곳으로 내던진다.

코시 확률변수는 어디에서 나오는가

균등하게 무작위인 방향을 향하도록 손전등을 벽에 비추면, 벽에 생기는 빛점의 위치는 코시 분포를 따른다. 물리학에서는 같은 밀도를 **로렌츠 분포**(*Lorentzian*)라고 부르며, 스펙트럼선과 공명 봉우리의 모양을 기술한다(Section 11.1와 비교해 보라). 금융에서는 정규분포 모형이 크게 과소평가하는 이따금의 폭락을 모형화하기 위해 코시 분포처럼 꼬리가 두꺼운(heavy-tailed) 분포를 사용한다. 통계학에 주는 교훈은 심각하다. 평균과 분산에 기반한 방법은 꼬리가 두꺼운 자료에서는 실패하며, 중앙값(median)을 사용하는 것과 같은 로버스트한(robust) 방법이 필요하다(코시 분포에서도 중앙값은 완벽하게 잘 정의되고 잘 작동한다).

푸아송, 코시, 그리고 논쟁

이 분포는 1824년 시메옹 드니 푸아송(Siméon Denis Poisson)이 연구했는데, 그는 이 분포가 큰 수의 법칙을 위반한다는 것을 알아차렸다. 오귀스탱루이 코시(Augustin-Louis Cauchy)는 1853년 최소제곱법(method of least squares)을 둘러싸고 Irénée-Jules Bienaymé와 공개 논쟁을 벌이던 중 이 분포를 만났다. 코시는 오차의 꼬리가 두꺼울 때 최소제곱법이 형편없는 방법일 수 있음을 보이는 데 이 분포를 사용했다. 이름은 코시에게 붙었는데, 이는 Stephen Stigler가 "스티글러의 명명 법칙(Stigler's law of eponymy)"이라고 부른 것의 한 예이다. 어떤 과학적 발견도 최초 발견자의 이름을 따서 불리지 않는다는 법칙이다.

사고 기술

기댓값이 존재하는지 확인하라

평균이나 분산을 사용하기 전에, 그것을 정의하는 적분이 수렴하는지 확인하라. 대칭성이 발산하는 적분을 0으로 만들어 주지는 않는다.

대학에서는 적분 가능성(integrability) 조건이 확률론의 모든 정리에 포함된다. 큰 수의 법칙에는 유한한 평균이, 중심극한정리(central limit theorem, Section 12.24)에는 유한한 분산이 필요하다. 코시 분포는 이러한 조건이 성립하지 않을 때의 표준적인 반례(counterexample)이다.

더 도전하기

1 **손전등.** 길고 곧은 벽에서 거리 1만큼 떨어진 손전등이 벽의 수선과 각 Θ를 이루는 방향을 가리킨다. 여기서 Θ는 $(-\frac{\pi}{2}, \frac{\pi}{2})$에서 균등분포를 따른다. 빛점의 위치 $\tan\Theta$가 코시 분포를 따름을 보여라.

2 **역수.** C가 코시 분포를 따르면 $\frac{1}{C}$도 코시 분포를 따름을 보여라. (**힌트**: $\frac{U}{V}$와 $\frac{V}{U}$.)

3 **중앙값은 잘 작동한다.** 코시 분포의 중앙값이 0이고 사분위수(quartile)가 ± 1임을 보여라.

4 **반지름과 각.** X, Y가 독립인 표준정규 확률변수일 때, 극좌표(polar coordinates)로 바꾸어 $R^2 = X^2 + Y^2$가 평균이 2인 지수분포(exponential distribution)를 따르고 각과 독립임을 보여라. (이것이 정규분포 난수를 생성하는 Box–Muller 방법의 기초이다.)

더 읽을거리

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University Press). 확률변수의 변환을 다룬다.
- **Nassim Nicholas Taleb, *The Black Swan*** (Penguin). 일반 독자를 위해 두꺼운 꼬리와 그것이 왜 중요한지를 설명한다.
- **Stephen Stigler, *The History of Statistics*** (Harvard University Press). 코시와 Bienaymé의 논쟁을 담고 있다.

12.23 확률로 얻는 스털링 공식

문제

STEP II 2016, 13번 문제

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

`step.maths.org/questions/16-s2-q13`

도입

100!은 얼마나 큰 수일까? 직접 곱해 보면 시간이 꽤 걸리는데, 이 수는 158자리이다. **스털링 공식**(*Stirling's formula*) $n! \approx \sqrt{2\pi n}\,(n/e)^n$을 쓰면 단 한 줄로 오차 0.1% 이내의 값을 얻는다. 이것은 수학에서 가장 유용한 근사(approximation) 가운데 하나로, 확률론, 통계물리학, 알고리즘 분석에서 늘 쓰인다.

이 공식의 증명은 여러 가지가 있는데, 적분을 이용하는 방법(Section 4.4)이나 오일러-매클로린 공식(Euler–Maclaurin formula)을 이용하는 방법이 있다. 이 문제는 그중 가장 놀라운 증명, 즉 확률을 이용한 증명을 보여 준다. 평균이 n인 푸아송 분포(Poisson distribution)는 근사적으로 정규분포(normal distribution)를 따르고, 봉우리에서 두 분포를 비교하면 $n!$의 값이 저절로 정해진다. 스털링 공식에 나오는 $\sqrt{2\pi}$는 말 그대로 정규분포의 확률밀도(density)에 나오는 바로 그 $\sqrt{2\pi}$이다.

시작하기 전에

생각해 보기 평균이 μ이고 분산이 σ^2인 이산 확률변수(discrete random variable) X가 근사적으로 정규분포를 따른다면, $\mathrm{P}(X = \mu)$는 대략 얼마일까? (μ를 중심으로 하는 폭 1인 구간의 확률을 생각해 보자.)

정규근사(normal approximation)에서 $\mathrm{P}(X = \mu) \approx \mathrm{P}(\mu - \frac{1}{2} < X < \mu + \frac{1}{2})$이고, 이는 대략 폭 1에 봉우리에서의 확률밀도를 곱한 값, 곧 $\frac{1}{\sigma\sqrt{2\pi}}$이다.

차근차근 풀기

1단계 동전 열여섯 개

$X \sim \mathrm{B}(16, \tfrac{1}{2})$의 평균(mean)은 8이고 분산(variance)은 $16 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 4$이므로 $\sigma = 2$이다. 정규근사에 의해

$$\mathrm{P}(X = 8) \approx \frac{1}{2\sqrt{2\pi}} \approx 0.1995.$$

정확한 값은 $\binom{16}{8}/2^{16} = \tfrac{12870}{65536} \approx 0.1964$이므로 오차는 2% 이내이다.

2단계 중심 이항계수

> **생각해 보기** $\mathrm{B}(2n, \tfrac{1}{2})$의 평균과 표준편차(standard deviation)는 무엇인가? $\mathrm{P}(X = n)$을 정확한 값과 근삿값으로 각각 써 보자.

평균은 n이고 분산은 $\tfrac{n}{2}$이므로 $\sigma = \sqrt{n/2}$이다. 정확한 값은 $\mathrm{P}(X = n) = \binom{2n}{n}\frac{1}{2^{2n}}$이고, 근삿값은 $\frac{1}{\sqrt{n/2}\sqrt{2\pi}} = \frac{1}{\sqrt{\pi n}}$이다. 두 값을 같다고 놓으면

$$\frac{(2n)!}{(n!)^2\,2^{2n}} \approx \frac{1}{\sqrt{\pi n}}, \qquad (2n)! \approx \frac{2^{2n}(n!)^2}{\sqrt{n\pi}}. \qquad \blacksquare$$

3단계 스털링 공식

> **생각해 보기** 모수가 n인 푸아송 분포는 평균과 분산이 모두 n이다. $\mathrm{P}(X = n)$을 정확한 값과 근삿값으로 각각 써 보자.

정확한 값은 $\mathrm{P}(X = n) = e^{-n}\frac{n^n}{n!}$이다. n이 클 때 푸아송 분포는 근사적으로 정규분포를 따르고(서로 독립(independent)인 n개의 푸아송(1) 확률변수의 합이기 때문이다), $\sigma = \sqrt{n}$이므로 $\mathrm{P}(X = n) \approx \frac{1}{\sqrt{2\pi n}}$이다. 두 값을 같다고 놓으면

$$n! \approx \sqrt{2\pi n}\,e^{-n}n^n. \qquad \blacksquare$$

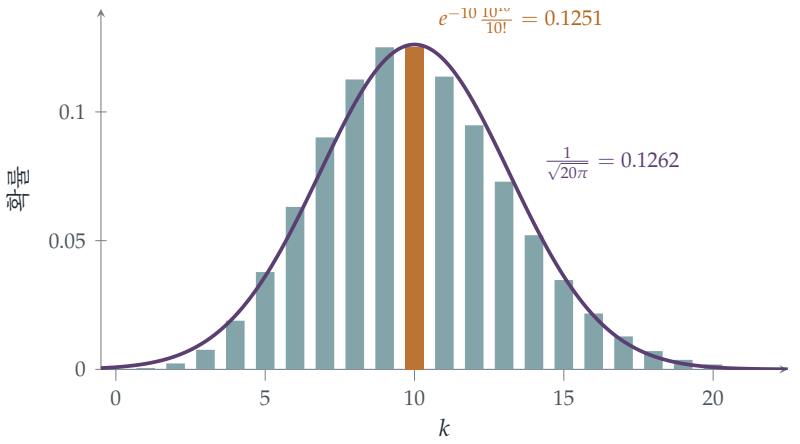


그림 12.23.1 평균이 10인 푸아송 분포(막대)와, 평균과 분산이 같은 정규분포의 확률밀도(자주색). 봉우리에서 푸아송 확률 $e^{-10}10^{10}/10!$(호박색)은 정규분포의 확률밀도 $\frac{1}{\sqrt{2\pi\cdot 10}}$에 가깝다. 이 근사적 등식을 정리한 것이 바로 스털링 공식이다.

돌아보기

- **한 점의 확률 ≈ 확률밀도 곱하기 폭.** 격자 분포(lattice distribution)에서는 한 점에서의 정규분포 확률밀도가 그 점의 확률의 근삿값이 된다.
- **정확한 공식과 근사의 만남.** 이항분포(binomial distribution)와 푸아송 분포의 확률은 정확한 값이다. 이것을 정규근사와 같다고 놓아 계승(factorial)의 추정값을 얻었다.
- **공식에 맞는 분포를 고른다.** $\binom{2n}{n}$은 $\mathrm{B}(2n, \frac{1}{2})$를, $\frac{n^n}{n!}$은 푸아송(n)을 떠올리게 했다.

숨은 수학

근사는 얼마나 정확한가?

이 논증은 휴리스틱(heuristic)이다. 정규근사를 **한 점에서** 사용하는데, 보통의 중심극한정리(central limit theorem)는 이를 정당화해 주지 않는다. 이를 엄밀하게 다룬 **국소 중심극한정리**(*local central limit theorem*)가 바로 그 일을 해 주며, 이로써 이 논증은 진정한 증명이 된다. 사실 스털링 공식은 놀라울 만큼 정확하다. 비 $\frac{n!}{\sqrt{2\pi n}(n/e)^n}$은 대략 $1+\frac{1}{12n}$이므로, $n=1$일 때조차 1 대신 0.92를 주고, $n=10$일 때는 0.8%만큼 작다. 전체 급수 $1+\frac{1}{12n}+\frac{1}{288n^2}-\cdots$는 **점근급수**(*asymptotic series*)의 한 예이다. n이 클수록 점점 더 좋은 근삿값을 주지만, n을 고정하면 결국 발산(diverge)한다.

계승과 정규분포가 연결되는 이유

스털링 공식의 $\sqrt{2\pi}$와 정규분포의 확률밀도에 나오는 $\sqrt{2\pi}$는 같은 상수이다. $\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$이기 때문이다. 드무아브르(de Moivre)는 1733년에 바로 이런 방식으로 이항분포의 정규근사를 발견했는데, 그 과정에서 큰 n에 대한 $\binom{n}{k}$의 근삿값이 필요했다. 그는 $n! \approx Cn^{n+1/2}e^{-n}$ 꼴은 찾았지만 상수는 찾지 못했다. 스털링(James Stirling)이 $C = \sqrt{2\pi}$임을 보였다. 정규분포는 계승의 점근적 성질에서 태어난 것이다.

드무아브르와 스털링

드무아브르(Abraham de Moivre)는 런던으로 망명한 프랑스 개신교도로, 개인 교습과 도박꾼들에게 조언을 해 주며 생계를 꾸렸다. 그는 1718년에 *The Doctrine of Chances*(우연의 이론)를 출간했고, 1733년의 소책자에서 이항분포의 정규근사를 발견했는데, 이것이 정규곡선이 처음 등장한 곳이다. 스코틀랜드의 수학자 스털링은 *Methodus Differentialis*(차분법, 1730)에서 이 상수의 값을 발표했고, 드무아브르는 너그럽게 그의 공로를 인정했다. 계승의 근사는 지금도 활발히 연구되는 주제이다. 아주 큰 n에 대해 $n!$을 계산하는 일은 조합론과 암호학에서 필요하다.

사고 기술

확률과 무관한 것을 확률로 증명하기

정확한 확률을 근삿값과 비교하면 그 공식에 나오는 수들에 대한 정보를 얻을 수 있다.

이 "확률론적 방법(probabilistic method)"은 양방향으로 작동한다. 여기서는 확률이 계승의 추정값을 주었고, 에르되시(Erdős)는 이 방법으로 그래프(graph)의 존재를 증명했다(Section 9.2). 대학에서는 확률을 이용해 해석학, 정수론, 조합론의 정리를 증명하는 것을 보게 될 것이다.

더 도전하기

1 **공식 확인하기.** 스털링 공식으로 10!과 20!을 추정하고, 참값 3 628 800, 2 432 902 008 176 640 000과 비교하여라.

2 **(iii)에서 (ii)로.** 소문항 (iii)의 스털링 공식으로부터 소문항 (ii)의 근사가 나옴을 보여라.

3 **100!의 자릿수.** $\log_{10} n! \approx n\log_{10}\frac{n}{e} + \frac{1}{2}\log_{10}(2\pi n)$을 이용하여 100!이 158자리 수임을 보여라.

4 동전 던지기에서의 동점. 소문항 (ii)를 이용하여, 공정한 동전을 1000번 던질 때 앞면이 정확히 500번 나올 확률이 약 0.025임을 보여라.

더 읽을거리

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1** (Wiley), 제II장. 스털링 공식을 증명한다.
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics*** (Addison-Wesley), 제9장. 점근적 방법을 다룬다.
- **Stephen Stigler, *The History of Statistics*** (Harvard University Press). 드무아브르와 정규곡선의 탄생을 다룬다.

12.24 중심극한정리

문제

STEP III 2008, 12번 문제

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

`step.maths.org/questions/08-s3-q12`

도입

서로 독립인 작은 무작위 효과를 많이 더하면, 각각이 어떤 분포를 따르든 그 합은 종 모양의 정규분포(normal distribution) 곡선을 따른다. 이것이 **중심극한정리**(*central limit theorem*)로, 아마도 확률론에서 가장 중요한 정리일 것이다. 이 정리는 키, 측정 오차, 시험 점수가 대체로 정규분포를 따르는 이유를 설명하고, 또 통계학자들이 바탕에 깔린 분포에 대해 아무것도 모르면서도 평균이 얼마나 믿을 만한지 말할 수 있는 이유를 설명한다. 단, 그 분포의 분산(variance)이 유한해야 한다(Section 12.22와 비교해 보라).

이 문제는 특수한 경우에 이 정리를 증명한다. 개구리 한 마리가 양쪽 지수분포(two-sided exponential distribution)를 따르는 무작위 점프를 한다. 이 분포의 밀도함수(density)는 뾰족한 봉우리를 가지며 종 모양과는 전혀 닮지 않았다. 그런데 점프를 많이 하고 나면 개구리의 전체 변위는 정규분포를 따른다. 여기서 쓰는 도구는 **적률생성함수**(*moment generating function*)로, 이 함수는 독립인 확률변수의 합을 곱으로 바꾸고, 로그를 취하면 곱은 다시 합으로 바뀐다.

출제 범위에 관한 참고

적률생성함수는 2019년에 STEP 출제 범위에서 빠졌다. 확률변수 X의 적률생성함수는 $\mathrm{M}_X(\theta) = \mathrm{E}(e^{\theta X})$이다. 이 함수의 0에서의 도함수들이 적률(moment)을 준다: $\mathrm{M}'(0) = \mathrm{E}(X)$, $\mathrm{M}''(0) = \mathrm{E}(X^2)$. 서로 독립인 X와 Y에 대하여 $\mathrm{M}_{X+Y} = \mathrm{M}_X\mathrm{M}_Y$이고, $\mathrm{M}_{cX}(\theta) = \mathrm{M}_X(c\theta)$이다. 적률생성함수들이 수렴하면 (적절한 의미에서) 분포들도 수렴한다.

시작하기 전에

생각해 보기 밀도함수 $\frac{1}{2}e^{-|x|}$는 대칭이다. 그래프는 어떤 모양인가? 평균은 얼마인가?

이 그래프는 0에서 높이 $\frac{1}{2}$인 뾰족한 봉우리를 가지고 양쪽으로 지수적으로 감소하는 천막 모양의 곡선이다(그림 12.24.1). 대칭성에 의해 평균은 0이다.

차근차근 풀기

1단계 적률생성함수

$|\theta| < 1$일 때, 적분을 0에서 나누면

$$\mathrm{M}_X(\theta) = \frac{1}{2}\int_{-\infty}^{0} e^{(1+\theta)x}\,\mathrm{d}x + \frac{1}{2}\int_{0}^{\infty} e^{-(1-\theta)x}\,\mathrm{d}x = \frac{1}{2}\left(\frac{1}{1+\theta} + \frac{1}{1-\theta}\right) = \frac{1}{1-\theta^2}. \qquad \blacksquare$$

전개하면 $\mathrm{M}_X(\theta) = 1 + \theta^2 + \theta^4 + \cdots$이므로 $\mathrm{E}(X) = 0$이고 $\mathrm{E}(X^2) = 2$이다($\frac{\theta^2}{2!}$의 계수). 따라서 분산은 2이다.

2단계 개구리의 변위

생각해 보기 $Y = X_1 + \cdots + X_n$의 적률생성함수는 무엇인가? $Y/\sqrt{2n}$의 적률생성함수는?

독립성에 의해 $\mathrm{M}_Y(\theta) = (1-\theta^2)^{-n}$이고, 척도를 바꾸면

$$\mathrm{M}_{Y/\sqrt{2n}}(\theta) = \mathrm{M}_Y\left(\frac{\theta}{\sqrt{2n}}\right) = \left(1 - \frac{\theta^2}{2n}\right)^{-n}.$$

(Y의 표준편차(standard deviation)인 $\sqrt{2n}$으로 나누면 분산이 1이 된다.)

3단계 극한

생각해 보기 로그를 취하고, 작은 u에 대하여 $\ln(1-u) \approx -u$임을 이용하라.

$\ln \mathrm{M}_{Y/\sqrt{2n}}(\theta) = -n\ln\left(1 - \frac{\theta^2}{2n}\right) = -n\left(-\frac{\theta^2}{2n} - \frac{\theta^4}{8n^2} - \cdots\right) \to \frac{\theta^2}{2}$이므로

$$\mathrm{M}_{Y/\sqrt{2n}}(\theta) \to e^{\theta^2/2}, \qquad \blacksquare$$

이고, 이것은 표준정규분포의 적률생성함수이다.

4단계 개구리의 확률

> **생각해 보기** n이 크면 Y는 근사적으로 $\mathrm{N}(0, 2n)$을 따른다. 언제 $\mathrm{P}(|Y| \geq 25) = 0.05$가 되는가?

표준정규분포를 따르는 Z에 대하여 $\mathrm{P}(|Z| \geq 1.96) = 0.05$이다. 따라서 $\frac{25}{\sqrt{2n}} = 1.96$이어야 하고, 이로부터 $2n = \left(\frac{25}{1.96}\right)^2 \approx 162.7$, 즉 $n \approx 81.3$을 얻는다. 멀리 떨어진 곳에 도착할 확률은 n이 커질수록 커지므로, 확률이 5% 이상이 되게 하는 최소 점프 횟수는 약 82이다. ■

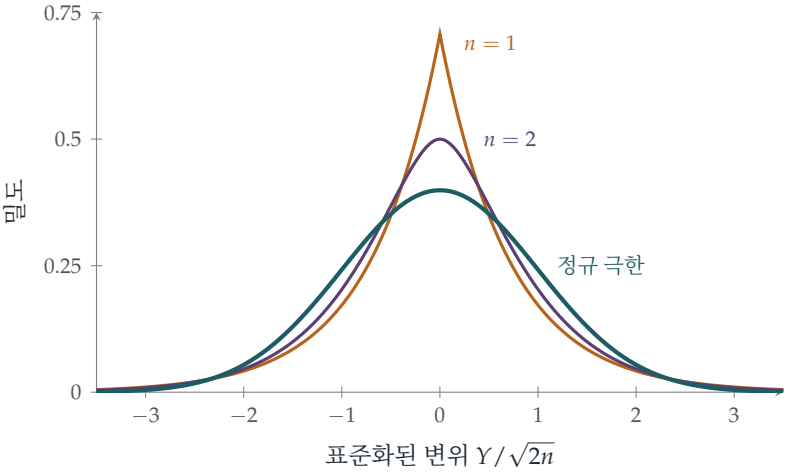


그림 12.24.1 $n = 1$번 점프한 뒤(호박색: 뾰족한 라플라스 봉우리)와 $n = 2$번 점프한 뒤(자주색) 개구리의 표준화된 변위의 밀도함수를 표준정규분포의 밀도함수(청록색)와 비교한 것. 점프를 두 번만 해도 봉우리가 이미 둥글게 다듬어진다. 중심극한정리에 따르면 $n \to \infty$일 때 이 곡선들은 정규분포로 수렴한다.

돌아보기

- **변환은 합을 곱으로 바꾼다.** 독립인 확률변수들의 합의 적률생성함수는 각 적률생성함수의 곱이다.
- **로그는 곱을 합으로 바꾼다.** $\ln\left(1 - \frac{\theta^2}{2n}\right)^{-n}$은 극한을 구하기 쉬운 합이 되었다.
- **표준화하라.** 표준편차 $\sqrt{2n}$으로 나누어 분산을 1로 유지했기 때문에 극한이 0도 무한대도 아니게 되었다.

숨은 수학

일반적인 정리

같은 논증은 평균이 0이고 분산 σ^2이 유한한 임의의 분포에 대해서도 성립한다(적률생성함수가 존재해야 하지만, 이 조건은 피할 수 있다). 이때 $\mathrm{M}(\theta) = 1 + \tfrac{1}{2}\sigma^2\theta^2 + o(\theta^2)$이고,

$$\left(\mathrm{M}\left(\frac{\theta}{\sigma\sqrt{n}}\right)\right)^n = \left(1 + \frac{\theta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{\theta^2/2}.$$

극한에서 살아남는 것은 분산뿐이고, 원래 분포의 모양은 잊힌다. 정규분포가 보편적인 이유가 바로 이것이다. 대학에서는 항상 존재하는 **특성함수**(*characteristic function*) $\mathrm{E}(e^{i\theta X})$를 이용하여 이 정리를 증명하고, 분포의 수렴은 레비(Lévy)의 연속성 정리로 엄밀하게 다룬다.

얼마나 빠른가?

중심극한정리는 극한에서 무슨 일이 일어나는지 말해 줄 뿐, 얼마나 빨리 일어나는지는 말해 주지 않는다. **베리-에센 정리**(*Berry–Esseen theorem*)는 오차를 상수 곱하기 $\frac{\mathrm{E}|X|^3}{\sigma^3\sqrt{n}}$으로 제한한다. 개구리의 분포처럼 대칭인 분포에서는 수십 번만 점프해도 근사가 좋다. 치우친(skewed) 분포이거나 꼬리의 극단적인 부분에서는 n이 커도 근사가 나쁠 수 있다. 정규근사는 매우 드문 사건이 일어날 확률을 과소평가하며, 이 사실은 보험과 금융에 중대한 영향을 미친다.

드무아브르에서 린데베리까지

드무아브르(de Moivre)는 1733년에 이항분포의 정규근사를 발견했고(Section 12.23), 라플라스(Laplace)는 1810년에 이를 더 일반적인 확률변수들의 합으로 확장하면서 측정 오차가 많은 작은 효과의 합이라는 생각을 처음 내놓았다. 엄밀한 증명은 훨씬 뒤에 나왔다. 1901년에 알렉산드르 랴푸노프(Aleksandr Lyapunov)가, 그리고 1922년에 야를 린데베리(Jarl Lindeberg)가 증명했는데, 린데베리는 정리가 성립하는 본질적으로 가장 약한 조건을 찾아냈다. "중심극한정리"라는 이름은 1920년에 조지 포여(George Pólya)가 붙인 것으로, 이 정리가 확률론의 중심에 있다는 뜻이다.

사고 기술

변환하고, 단순화하고, 되돌려 변환하라

어려운 연산(독립인 확률변수를 더하는 것, 즉 밀도함수들의 합성곱(*convolution*)을 구하는 것)도 변환을 거치면 쉬운 연산(곱셈)이 된다. 그곳에서 계산을 하고, 그 결과를 해석하라.

같은 전략이 생성함수(Section 9.4), 해석학과 공학에서 쓰이는 푸리에(Fourier) 변환과 라플라스 변환, 그리고 곱셈을 덧셈으로 바꾸는 로그의 바탕에 깔려 있다.

더 도전하기

1 **점프 두 번의 정확한 분포.** $Y = X_1 + X_2$의 밀도함수가 $\frac{1}{4}(1 + |y|)e^{-|y|}$임을 보이고, 분산이 4임을 확인하라.

2 **이항분포에서 정규분포로.** 평균이 0이 되도록 평행이동한(즉 $\pm\frac{1}{2}$의 값을 갖는) $X \sim \mathrm{B}(1, \frac{1}{2})$에 대하여 $\mathrm{M}_X(\theta) = \cosh\frac{\theta}{2}$임을 구하고, 이러한 확률변수 n개의 합을 표준화한 것의 적률생성함수가 $e^{\theta^2/2}$로 수렴함을 보여라.

3 **개구리, 정확하게.** 컴퓨터를 이용하여(또는 검산용으로 $n = 2$일 때의 정확한 밀도함수를 이용하여) $n = 82$일 때 정규근사가 얼마나 정확한지 알아보라.

4 **코시 분포는 왜 안 되는가?** 코시 분포(Cauchy distribution)(Section 12.22)는 적률생성함수가 없다. 논증의 어느 단계가 실패하는가? 그리고 코시 분포를 따르는 확률변수 n개의 평균은 어떻게 되는가?

더 읽을거리

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University Press), 생성함수와 극한정리를 다룬 장.
- **Hans Fischer, *A History of the Central Limit Theorem*** (Springer). 라플라스부터 린데베리까지를 다룬다.
- **David Spiegelhalter, *The Art of Statistics*** (Pelican), 정규분포가 어디에나 나타나는 이유를 다룬다.

12.25 여럿 가운데 가장 큰 값

문제

STEP III 2025, 11번 문제

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

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

도입

다음 백 년 동안 일어날 가장 큰 홍수는 얼마나 높을까? 전구 천 개 가운데 가장 오래가는 전구는 얼마나 버틸까? n명이 뛰는 마라톤에서 가장 빠른 주자는 얼마나 빠를까? 여러 확률적 값 가운데 가장 큰 값이나 가장 작은 값, 곧 **극값**(*extremes*)에 관한 질문은 공학, 보험, 기후과학에서 매우 중요하며, 평균과는 전혀 다르게 움직인다.

이 문제에서는 지수분포(exponential distribution)를 따르는 수명 n개 가운데 가장 큰 값의 분포를 구한다. 그 전형적인 크기는 $\ln n$처럼 커진다. 느리지만 한없이 커진다. 그러나 그 퍼짐은 전혀 커지지 않는다. 수명 천 개의 최댓값은 수명 열 개의 최댓값과 대략 같은 정도로 예측 가능하다. 극한에서 그 분포의 모양은 **검벨 분포**(*Gumbel distribution*)이며, 이는 정규분포 곡선에 대응하는 극값의 분포이다.

시작하기 전에

생각해 보기 $X_1, \ldots, X_n$ 가운데 가장 큰 값이 y 이하인 것은 정확히 어떤 일이 일어날 때인가?

$Y \le y$인 것은 정확히 모든 $X_i \le y$일 때이다. 그러면 독립성(independence)에 의해 이것이 곱으로 바뀐다.

차근차근 풀기

1단계 최댓값의 분포

각 X_i는 $y \ge 0$에서 $\mathrm{F}(y) = 1 - e^{-\lambda y}$이고, 독립성에 의해

$$\mathrm{G}(y) = \mathrm{P}(X_1 \le y, \dots, X_n \le y) = (1 - e^{-\lambda y})^n. \qquad \blacksquare$$

2단계 분위수

생각해 보기 $\mathrm{G}(L) = \alpha$를 L에 대하여 풀어라. 그다음 $\mathrm{G}(U) = 1 - \alpha$를 풀어라.

$(1 - e^{-\lambda L})^n = \alpha$에서 $e^{-\lambda L} = 1 - \alpha^{1/n}$을 얻으므로

$$L(\alpha) = -\frac{1}{\lambda}\ln\bigl(1 - \alpha^{1/n}\bigr), \qquad U(\alpha) = -\frac{1}{\lambda}\ln\bigl(1 - (1-\alpha)^{1/n}\bigr). \qquad \blacksquare$$

3단계 큰 n

생각해 보기 $\alpha^{1/n} = e^{(\ln\alpha)/n}$으로 써라. n이 크면 지수가 작다. $1 - \alpha^{1/n}$은 근사적으로 얼마인가?

$\alpha^{1/n} = e^{(\ln\alpha)/n} \approx 1 + \frac{\ln\alpha}{n}$이므로 $1 - \alpha^{1/n} \approx \frac{-\ln\alpha}{n} = \frac{\ln(1/\alpha)}{n}$이다. 따라서

$$\lambda L(\alpha) \approx -\ln\frac{\ln(1/\alpha)}{n} = \ln n - \ln\Bigl(\ln\frac{1}{\alpha}\Bigr). \qquad \blacksquare$$

마찬가지로 $\lambda U(\alpha) \approx \ln n - \ln\bigl(\ln\frac{1}{1-\alpha}\bigr)$이다.

4단계 증가와 퍼짐

중앙값(median)은 $L(\frac{1}{2})$이고, $\lambda L(\frac{1}{2}) \approx \ln n - \ln\ln 2 \to \infty$이다. 그러나

$$\lambda\bigl(U(\alpha) - L(\alpha)\bigr) \approx \ln\Bigl(\ln\frac{1}{\alpha}\Bigr) - \ln\Bigl(\ln\frac{1}{1-\alpha}\Bigr),$$

이고, 여기서 $\ln n$ 항들은 서로 상쇄되었다. 즉 폭은 n에 의존하지 않는다. $\blacksquare$

5단계 90% 구간

생각해 보기 $\alpha = 0.05$로 놓아라. $\ln\frac{1}{\alpha} = \ln 20$과 $\ln\frac{1}{0.95}$는 근사적으로 얼마인가?

$\alpha = 0.05$일 때 $\ln 20 \approx 3$이므로 $\ln\ln\frac{1}{\alpha} \approx \ln 3$이다. 또 $\ln\frac{1}{0.95} = -\ln(1 - 0.05) \approx 0.05 = \frac{1}{20}$이므로 $\ln\ln\frac{1}{1-\alpha} \approx \ln\frac{1}{20} \approx -3$이다. 따라서 폭은 약 $\frac{1}{\lambda}(\ln 3 + 3) \approx \frac{4.1}{\lambda}$이고, Y는 확률 0.9로 폭이 약 $4\lambda^{-1}$인 구간 $[L(0.05), U(0.05)]$에 놓인다. $\blacksquare$

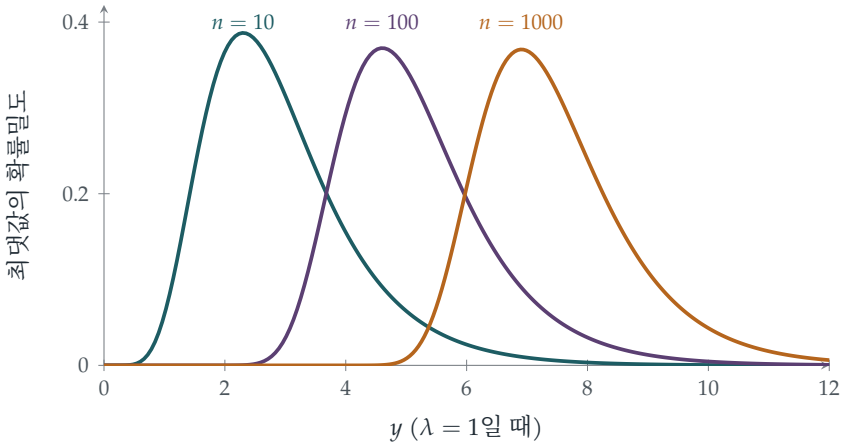


그림 12.25.1 $n = 10, 100, 1000$일 때 지수분포를 따르는 수명(평균 1) n개 가운데 가장 큰 값의 확률밀도. n이 10배가 될 때마다 곡선은 오른쪽으로 $\ln 10 \approx 2.3$만큼 이동하지만, 그 모양과 폭은 사실상 그대로이다. 이것이 $\ln n$만큼 평행이동한 검벨 분포이다.

돌아보기

- **최댓값이 y 이하인 것은 모두가 그럴 때이다.** 독립성 덕분에 최댓값의 분포가 거듭제곱이 되었다.
- **역으로 풀어 분위수 구하기.** $G(y) = \alpha$를 풀어 구간 양 끝의 정확한 공식을 얻었다.
- **근사한 다음 비교하기.** L과 U의 근사식에 모두 $\ln n$이 들어 있었고, 폭에서는 이것이 상쇄되었다.

숨은 수학

검벨 분포

$P(\lambda Y - \ln n \le x) = \left(1 - \frac{e^{-x}}{n}\right)^n \to \exp(-e^{-x})$이므로, 평행이동한 최댓값 $\lambda Y - \ln n$은 극한에서 분포함수(distribution function)가 $\exp(-e^{-x})$인 **검벨 분포**를 따른다. 그 평균은 오일러(Euler) 상수 $\gamma \approx 0.577$이다. 이 분포는 이미 만난 적이 있다. 쿠폰 n종을 모두 모으는 데 걸리는 시간(Section 9.11)은 (거의 독립인) 대기 시간 n개의 최댓값이며, 그래서 그 값이 약 $n \ln n + \gamma n$이고 n 정도 크기의 검벨 요동을 보이는 것이다.

극값 이론

피셔(Fisher), Tippett, 그네덴코(Gnedenko)의 놀라운 정리(1928–1943)에 따르면, 독립이고 동일한 분포를 따르는 많은 확률변수의 최댓값은 적절히 평행이동하고 축척을 조정하면 극한에서 가능한 모양이 세 가지뿐이다. 검벨 분포(지수분포나 정규분포처럼 꼬리가 가벼운 경우), 프레셰(Fréchet) 분포(코시(Cauchy) 분포처럼 꼬리가 무거운 경우), 바이불(Weibull) 분포(엄격한 상한이 있는 확률변수의 경우)이다. 이는 중심극한정리(central limit theorem)(Section 12.24)의 극값판이라 할 수 있다. 공학자들은 이를 이용해 "백 년 만의 파도"에 대비한 홍수 방어 시설과 해상 구조물을 설계하고, 보험사는 재해 보험료를 책정하며, 기후과학자들은 폭염이 어떻게 변하고 있는지를 추정한다.

검벨과 홍수

에밀 율리우스 검벨(Emil Julius Gumbel)은 독일의 통계학자로, 나치에 공공연히 반대하다가 1932년 대학 교수직에서 쫓겨났고 결국 미국으로 망명했다. 그는 1930년대와 1940년대에 극값 이론을 강의 범람에 적용했다. 그의 책 *Statistics of Extremes*(극값의 통계학, 1958)는 이 분야를 실용적인 도구로 만들었다. 1953년 북해 대홍수 이후 네덜란드를 보호하기 위해 건설된 델타 공사(Delta Works)는 약 10 000년에 한 번 일어날 것으로 예상되는 폭풍 해일을 견디도록 극값 통계를 이용해 설계되었다.

사고 기술

최댓값과 최솟값의 분포

독립인 확률변수들의 최댓값에 대해서는 $\mathrm{P}(\max \le y) = \prod \mathrm{P}(X_i \le y)$이고, 최솟값에 대해서는 $\mathrm{P}(\min > y) = \prod \mathrm{P}(X_i > y)$이다.

이 한 가지 아이디어로 가장 가까운 슈퍼마켓(Section 12.19), 가장 먼저 오는 버스(Section 12.18), 가장 오래가는 전구를 모두 다룰 수 있다. 대학에서는 이것이 순서통계량(order statistics)과 극값 이론의 출발점이 된다.

더 도전하기

1 **최댓값의 기댓값.** $\mathrm{E}(Y) = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)$임을 보여라. (힌트: 무기억성(memorylessness)에 의해, n개의 수명 가운데 첫 번째 것이 끝난 뒤 남은 $n-1$개는 새로 시작하는 것과 같다.)

2 **최솟값.** $X_1, \ldots, X_n$ 가운데 가장 작은 값이 모수 $n\lambda$인 지수분포를 따름을 보여라.

3 **정확한 중앙값.** Y의 정확한 중앙값을 구하고, $n = 10$일 때 이를 근삿값 $\frac{1}{\lambda}(\ln n - \ln\ln 2)$와 비교하여라.

4 **균등분포의 최댓값.** $X_1, \dots, X_n$이 $[0,1]$에서 균등분포(uniform distribution)를 따를 때, 그 최댓값 M의 분포를 구하고, n이 클 때 $n(1-M)$이 근사적으로 지수분포를 따름을 보여라. 이는 세 가지 극값 모양 가운데 어느 것인가?

더 읽을거리

- **Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values*** (Springer). 홍수와 해수면에 관한 실제 자료를 담은 표준 입문서이다.
- **Emil Julius Gumbel, *Statistics of Extremes*** (Dover). 고전이다.
- **David Spiegelhalter, *The Art of Uncertainty*** (Pelican). 드문 사건과 그것을 어떻게 생각해야 하는지에 관한 책이다.

우리가 배운 것

이 장은 불확실성도 정확하게 추론할 수 있지만 직관으로는 그렇게 할 수 없다는 말로 시작했다. 그리고 세 가지 물음을 던졌다. 문제가 정확히 무엇을 묻는가, 기발한 접근법은 무엇인가, 그리고 장기적으로 어떤 일이 일어나는가? 이 장의 스물다섯 문제가 보여 준 것은 다음과 같다.

1. 문제가 정확히 무엇을 묻는가?

처음 네 절은 문제를 올바르게 이해하는 일에 관한 것이었다. "번호가 적힌 원판을 하나 뽑았다는 조건에서"와 "3번 원판을 뽑았다는 조건에서"는 서로 다른 답을 주었다(Section 12.1). 희귀병에 대한 정확도 90%의 검사에서 양성이 나온 환자가 실제로 병에 걸렸을 확률은 9%에 불과했는데, 이는 기저율(base rate)이 중요하기 때문이다(Section 12.2). "무작위로 그은 현"은 그것을 어떻게 고르는지 정하기 전까지는 기댓값(expected value)으로서의 길이를 갖지 않았고(Section 12.3), 부러진 막대 문제(Section 12.21)도 같은 점을 보여 주었다. 또 "독립(independent)"이라는 말에는 여러 의미가 있음이 드러났다. 사건들이 두 개씩은 독립이지만 전체로는 독립이 아닐 수 있고(Section 12.4), 상관관계가 없는(uncorrelated) 확률변수라고 해서 반드시 독립인 것은 아니다. 어느 경우든 해결책은 가능한 결과가 정확히 무엇이며 그것들이 어떻게 생겨나는지를 적어 보는 것이었다.

2. 기발한 접근법은 무엇인가?

많은 문제에는 긴 계산과 짧은 통찰이 함께 있었다.

- **대칭(symmetry).** 과정이 서로 구별할 수 없는 두 좌석(Section 12.5), 서로 여사건이 되는 거울상으로서의 "앞면이 더 많음"과 "뒷면이 더 많음"(Section 12.6), 어느 경기에서든 모든 선수 쌍이 맞붙을 가능성이 똑같다는 사실(Section 12.7), 어떤 동전이든 HT와 TH가 나올 가능성이 같다는 사실(Section 12.9), 그리고 방향이 균등하게 분포하는 원점 대칭인 점(Section 12.22)이 그 예이다.
- **지시변수(indicator)와 선형성(linearity).** 쌍(Section 12.4), 매듭(Section 12.10), 경기(Section 12.7)의 개수를 지시변수의 합으로 세면, 전체 분포를 구하지 않고도 평균과 분산을 얻을 수 있었다.
- **상태(state)와 첫걸음 분석(first-step analysis).** Penney의 게임(Section 12.8), 토끼 굴과 Ecks 교수(Section 12.16), 그리고 성씨의 소멸(Section 12.15)은 모두 다음 단계에 대한

조건부 확률을 이용해 풀었다.

- **전략(strategy).** 삼자 결투(truel)에서 사격 실력이 가장 나쁜 사람은 일부러 빗맞혀야 하고(Section 12.13), 무작위화 전략(randomised strategy)은 상대가 역이용할 수 없으며(Section 12.14), 후보자 중 처음 $\frac{1}{e}$를 거르는 것이 최적이다(Section 12.12).

3. 장기적으로 어떤 일이 일어나는가?

이 장의 마지막 부분에서는 시간에 따라, 그리고 수많은 시행에 걸쳐 무작위성이 어떻게 전개되는지 따라가 보았다. 분기 과정(branching process)은 소멸하거나 폭발적으로 커지며, 폭발적으로 커지는 과정도 우연히 소멸할 수 있다(Section 12.15). 마르코프(Markov) 연쇄(Markov chain)는 정상 분포(stationary distribution)에 안착하고 자신의 과거를 잊는다(Section 12.16). 지수분포를 따르는 대기 시간(exponential waiting time)은 기억이 전혀 없다(Section 12.18). 평균은 대개 일정한 값에 안정되지만, 평균이 존재하지 않는 코시(Cauchy) 분포(Cauchy distribution)는 예외이다(Section 12.22). 서로 독립인 수많은 효과의 합은 중심극한정리(central limit theorem)에 따라 정규분포가 되며(Section 12.24), 이 정규성 속에는 $n!$에 대한 스털링(Stirling) 공식(Stirling's formula)이 숨어 있다(Section 12.23). 최댓값은 또 다르게 행동한다. 최댓값은 $\ln n$처럼 커지지만 그 퍼진 정도는 전혀 커지지 않는다(Section 12.25). 그리고 상수 e, π, γ는 컵과 우표, 바늘과 고리에서 거듭 나타났다.

4. 확률을 위한 도구 모음

방법	사용한 곳
표본공간(sample space) 적기	조건부 확률(12.1); 베르트랑(Bertrand)(12.3)
베이즈(Bayes) 정리와 오즈(odds)	기저율(12.2); 마르코프 역전(12.16)
대칭	12.5, 12.6, 12.7, 12.22
지시변수와 선형성	12.4, 12.7, 12.10
첫걸음 분석	Penney(12.8); 토끼 굴(12.16); 분기 과정(12.15)
여사건(complement)과 부등식으로 범위 정하기	공정한 결정(12.9)
기댓값을 위한 급수	우표와 e(12.11)
매개변수화와 최적화	비서 문제(secretary problem)(12.12)
게임 이론적 추론	삼자 결투(12.13); 혼합 전략(mixed strategy)(12.14)
끝에 이르는 경로 세기	성냥갑(12.17)
분포함수(distribution function)부터 구하기	12.18, 12.19, 12.25
넓이나 적분으로서의 확률	뷔퐁(Buffon)(12.20); 부러진 막대(12.21)
정규근사(normal approximation)	스털링(12.23); 중심극한정리(12.24)

다음으로

- 13장에서는 확률에서 통계로 넘어가, Section 13.1부터 시작하여 데이터로부터 미지의 양을 추정하는 방법을 다룬다.
- Section 9.8, Section 9.9, Section 9.11의 세기 문제는 이 장의 조합론적 측면에 해당한다.
- 마르코프 연쇄는 7장의 선형대수를 이용하고, 분기 과정의 고정점은 Section 8.7의 동역학을 이용한다.

대학에 가기 전에 더 나아가고 싶다면

- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***(확률론 입문): 훌륭한 첫 강좌로, 온라인에서 무료로 볼 수 있으며 동영상 강의도 있다.
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***(확률: 입문): 간결하고 엄밀하며, 대학 1학년 수준이다.
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***(확률의 도전적인 문제 50선): 완전한 풀이가 실린 고전적인 퍼즐 모음이다.
- **David Spiegelhalter, *The Art of Statistics***(통계의 기술)와 ***The Art of Uncertainty***(불확실성의 기술): 현실 세계의 확률과 위험을 다룬다.

라플라스(Laplace)는 확률을 계산으로 환원된 상식이라고 불렀다. 이 장에서 보았듯이, 계산이 필요한 것은 바로 상식이 실패하는 경우가 많기 때문이며, 가장 좋은 계산은 숫자를 내놓는 데 그치지 않고 무엇인가를 설명해 준다. 대칭, 지시변수, 잘 고른 상태, 또는 극한 정리는 퍼즐을 통찰로 바꾸어 놓으며, 대학의 확률론 강좌는 바로 이러한 통찰 위에 세워진다.

더 도전하기 해답

여기에 실은 것은 완전한 풀이가 아니라 간단한 답과 힌트이다. 결과와 핵심 단계만 제시하므로 자신의 풀이를 확인하는 데 쓰면 된다. 답을 보기 전에 각 문제를 제대로 풀어 보아야 한다.

12.1 조건을 정확하게 걸기

1. (iv)와 (v) 모두에서 나머지 $k-1$개의 원판은 남은 $b+n-1$개 가운데에서 무작위로 고른 것이고, 그중 $n-1$개에 번호가 적혀 있으므로 둘 다 $\frac{(n-1)\binom{b}{k-2}}{\binom{b+n-1}{k-1}}$와 같다. (vi)의 답은 $\frac{\binom{n}{2}\binom{b}{k-2}}{\binom{b+n}{k}-\binom{b}{k}}$이다. $N=b+n$이라 하면 두 값은 $(2N-nk)\binom{N-1}{k-1}=2k\binom{b}{k}$일 때에만 일치한다. 이런 일은 $(b,n,k)=(2,4,3),(3,3,4),(6,11,3)$과 같은 특이한 경우나, $n=2$이고 모든 원판을 뽑는 자명한 경우에만 일어난다. 문제에서 주어진 값에서는 성립하지 않는다.

2. 두 아이에 대한 같은 정도로 가능한 $14^2=196$개의 (성별, 요일) 쌍 가운데 화요일에 태어난 남자아이가 적어도 한 명 있는 것은 $196-13^2=27$개이다. 그중 두 아이가 모두 남자아이인 것은 $7^2-6^2=13$개이다. 따라서 확률(probability)은 $\frac{13}{27}$이다.

3. 1번 문을 골랐다고 하자. 자동차가 1번 문 뒤에 있으면(확률 $\frac{1}{3}$) 바꾸면 진다. 자동차가 2번이나 3번 문 뒤에 있으면 진행자는 염소가 있는 나머지 문을 열 수밖에 없으므로 바꾸면 이긴다. 그 확률은 $\frac{2}{3}$이다. 진행자가 문을 무작위로 열었는데 마침 염소가 나왔다면, 본 것이 주어졌을 때 "자동차가 내가 고른 문 뒤에 있다"와 "자동차가 닫힌 다른 문 뒤에 있다"는 두 결과는 같은 정도로 가능하므로 답은 $\frac{1}{2}$이다.

4. $\mathrm{P}(\geq 2 \mid \geq 1)=\frac{1-\mathrm{P}(0)-\mathrm{P}(1)}{1-\mathrm{P}(0)}\approx 0.370$이지만 $\mathrm{P}(\geq 2 \mid$ 스페이드 에이스$)=1-\binom{48}{12}/\binom{51}{12}\approx 0.561$이다. 어느 에이스인지 알면 두 번째 에이스가 있을 가능성이 더 커진다.

12.2 기저율 오류

1. 두 사람이 모두 "시신이 있다"고 말할 확률은 시신이 실제로 있으면 pq, 없으면 $(1-p)(1-q)$이다. 사전 오즈(prior odds)가 1 : 1이면 사후 확률(posterior probability)은 $\frac{pq}{pq+(1-p)(1-q)}=\frac{pq}{1-p-q+2pq}$이다. 두 번째 부분에서 "$A$에 있다"는 같은 정도로 가능한 세 경우 가운데 두 경우에서 참이므로 $\mathrm{P}($둘 다 $\mid$ 두 사람 모두 지목한 곳 $A)=$

$\frac{pq}{2pq+(1-p)(1-q)}$이다.

2. $0.9\pi > \frac{1}{11}(1-\pi)$, 즉 $\pi > \frac{10}{109} \approx 9.2\%$이어야 한다.

3. AL 검사에서 양성이 한 번 나올 때마다 오즈에 $\frac{0.9}{1/11} = 9.9$가 곱해진다. 사전 오즈 $\frac{1}{99}$에서 시작하면, 양성이 한 번 나온 뒤의 오즈는 0.1(확률 $\frac{1}{11}$), 두 번 나온 뒤에는 0.99(확률 약 0.497), 세 번 나온 뒤에는 9.8(확률 약 0.907)이다.

4. 그 질환이 있는 100명 가운데 99명이 양성 판정을 받고, 질환이 없는 999 900명 가운데 약 9999명이 양성 판정을 받는다. 따라서 약 10 100명이 질환이 있을 수 있다는 말을 듣지만, 그중 실제로 질환이 있는 사람은 99명(약 1%)뿐이다.

12.3 "무작위"란 무슨 뜻인가?

1. 현(chord)이 내접 정삼각형의 한 변 $\sqrt{3}a$보다 길어지는 경우는 다음과 같다. 먼 쪽 끝점이 가까운 쪽 끝점의 반대편에 있는 원의 삼분의 일 부분에 놓일 때($\frac{1}{3}$), O로부터의 거리가 $\frac{a}{2}$보다 작을 때($\frac{1}{2}$), 중점이 반지름 $\frac{a}{2}$인 원 안에 놓일 때($\frac{1}{4}$)이다.

2. 중점이 C인 현은 OC에 수직이므로, C를 지나고 각이 $t = \frac{\pi}{2}$인 현이다. 따라서 $L(\frac{\pi}{2}) = \frac{4a}{3}\left(0 + \frac{1}{2}\sec^2\frac{\pi}{4}\right) = \frac{4a}{3}$이다.

3. 길이는 $2\sqrt{a^2-d^2}$이고, 그 평균은 $\frac{2}{a}\int_0^a \sqrt{a^2-d^2}\,\mathrm{d}d = \frac{2}{a}\cdot\frac{\pi a^2}{4} = \frac{\pi a}{2}$이다.

4. 무작위 끝점: $\frac{2}{\pi\sqrt{4a^2-\ell^2}}$. 무작위 거리: $\frac{\ell}{2a\sqrt{4a^2-\ell^2}}$. 무작위 중점: $\frac{\ell}{2a^2}$.

12.4 둘씩은 독립이지만 한꺼번에는 독립이 아닌

1. 전개하면 $\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = (a-pq)(x_1-x_2)(y_1-y_2)$이므로 $a = pq$이다. 그러면 나머지 세 칸도 따라서 정해지고, 독립(independence)이 된다. 예: $(0,0)$에 확률 $\frac{1}{3}$을, $(\pm1,\pm1)$ 각각에 확률 $\frac{1}{6}$을 준다. 두 주변 분포(marginal distribution)는 모두 균등하고 $\mathrm{E}(AB) = 0$이지만 $\mathrm{P}(A=0, B=1) = 0 \neq \frac{1}{9}$이다.

2. 쌍은 21개이고, 각 쌍이 일치할 확률은 $\frac{1}{6}$이며, 지시변수(indicator)들은 쌍마다 독립(pairwise independent)이다. 따라서 평균은 3.5, 분산(variance)은 $21\cdot\frac{1}{6}\cdot\frac{5}{6} \approx 2.92$, 표준편차(standard deviation)는 약 1.71이다. 점수 7은 평균보다 표준편차의 약 두 배만큼 높다.

3. 평균은 $\binom{k}{2}\frac{1}{365}$이다. 주사위의 경우와 똑같이 지시변수들이 쌍마다 독립이므로 분산은 $\binom{k}{2}\frac{1}{365}\cdot\frac{364}{365}$이다. (상호 독립(mutually independent)은 아니다. 1, 2와 2, 3이 일치하면 1, 3도 일치하기 때문이다.)

4. 같은 정도로 가능한 결과 여덟 개를 잡고 $A = B = \{1,2,3,4\}$, $C = \{1,5,6,7\}$로 둔다. 그러면 $\mathrm{P}(A\cap B\cap C) = \frac{1}{8}$이지만 $\mathrm{P}(A\cap B) = \frac{1}{2} \neq \frac{1}{4}$이다.

12.5 마지막 승객의 좌석

1. 모든 $n \geq 3$에 대하여 $\frac{1}{3}$이다. 이제 마지막 좌석은 S_1, S_2, S_n 가운데 하나이고, 모든 무작위 선택이 이 세 좌석을 대칭적으로 다룬다.

2. T_1이 S_1에 앉으면 아무도 자리를 빼앗기지 않는다. T_1이 S_3에 앉으면 T_1과 T_3이 자리를 빼앗긴다. T_1이 S_2에 앉으면 T_2가 고르는데, S_1을 고르면 2명, S_3을 고르면 3명이 자리를 빼앗긴다. 기댓값(expectation)은 $\frac{1}{3}\left(0+2+\frac{5}{2}\right) = \frac{3}{2}$이고, 이는 공식에서 얻은 $\frac{2}{3}+\frac{1}{3}+\frac{1}{2}$와 일치한다.

3. $r \geq 2$에 대하여 $P_r = \frac{1}{2}$이면 $\frac{1}{n}\left(1+\frac{n-2}{2}\right) = \frac{1}{2}$이므로, 점화식(recurrence)은 $P_n = \frac{1}{2}$와 모순되지 않는다.

4. 마지막으로 비어 있는 좌석은 항상 S_1 또는 S_n이다(다른 S_k는 늦어도 T_k가 탑승할 때에는 차 있다). 따라서 그것이 S_n이 아니면 S_1이다.

12.6 동전 하나 더

1. Bob이 처음 $n+1$번 던진 결과를 H', 마지막으로 던진 결과를 X라 하자. Bob은 $H' > H_A$이거나, $H' = H_A$이고 X가 앞면이면 이긴다. 따라서 확률은 $\mathrm{P}(H' > H_A) + \frac{1}{2}\mathrm{P}(H' = H_A)$이고, 동전 하나를 더 던지는 경우의 결과에 의해 $\mathrm{P}(H' > H_A) = \frac{1}{2}$이다.

2. 아니다. $n = 0$이면 Bob이 이길 확률은 p이다. 간결한 증명은 앞면과 뒷면을 맞바꾸는데, 이는 공정한 동전일 때에만 대칭이 된다.

3. $H_A = H_B$인 것은 정확히 $H_A + (\text{Bob의 뒷면 수}) = n$일 때이고, 이 좌변은 n과 같아지는 $\mathrm{B}(2n, \frac{1}{2})$ 확률변수이다. 값은 $\frac{1}{2}$, $\frac{3}{8}$, $\frac{5}{16}$이다.

4. 아니다. 6이 나오는 것은 $p = \frac{1}{6}$인 동전과 같다. $n = 0$일 때 확률은 $\frac{1}{6}$이고, 일반적으로 $\frac{1}{2}$보다 작다.

12.7 토너먼트에서 누가 누구와 만나는가

1. k라운드에는 2^{n-k}개의 경기가 있고, 대칭성에 의해 각 경기는 $\binom{2^n}{2}$개의 쌍 가운데 어느 쌍이든 같은 정도로 치를 가능성이 있다. 토너먼트 전체의 $2^n - 1$개 경기에 대해 더하면 $\frac{2^n-1}{\binom{2^n}{2}} = \frac{2}{2^n}$이다.

2. 아니다. 고정된 대진표에 무작위로 배치해도 대칭성은 유지되므로 같은 셈이 적용된다.

3. X의 k라운드 상대는 2^{k-1}명으로 이루어진 한 조에서 나오고, 그 조의 각 선수가 조에서 이길 가능성은 같다. 따라서 Y가 그 상대일 확률은 $\frac{2^{k-1}}{2^n-1} \cdot \frac{1}{2^{k-1}} = \frac{1}{2^n-1}$이고, n개 라운드에 대한 합은 $\frac{n}{2^n-1}$이다.

4. 경기는 $\binom{2n}{2} = n(2n-1)$개이고, 모든 쌍이 맞붙으므로 확률은 1이다.

12.8 페니의 게임

1. HHT는 처음 두 번이 HH일 때에만 이긴다. 그렇지 않으면 HH가 완성되기 전에 T가 나오고, 그러면 THH가 먼저 나타날 수밖에 없다. 따라서 HHT가 이길 확률은 $\frac{1}{4}$이다. 네 명이 모두 참가하는 경우: HH 다음에는 HHT가 반드시 이기고, TT 다음에는 TTH가 반드시 이긴다. HT와 TH에서 HTT와 THH가 이길 확률은 $\frac{2}{3}$과 $\frac{1}{3}$(또는 그 반대)이므로, 각 참가자가 이길 확률은 $\frac{1}{4}\cdot 1$ 또는 $\frac{1}{4}\left(\frac{2}{3}+\frac{1}{3}\right)$, 즉 $\frac{1}{4}$이다. TTH는 확률 $\frac{2}{3}$으로 THH를 이긴다(소문항 (i)).

2. HH의 경우: $e_0 = 1 + \frac{1}{2}e_{\mathrm{H}} + \frac{1}{2}e_0$이고 $e_{\mathrm{H}} = 1 + \frac{1}{2}e_0$이므로 $e_0 = 6$이다. HT의 경우: H가 한 번 나오면 T를 기다리면 되고(평균 두 번), 첫 H가 나오는 데에도 두 번이 걸리므로 4이다.

3. HHH와 TTT는 각각 THH와 HTT에게 확률 $\frac{7}{8}$로 지고, HHT와 TTH는 THH와 HTT에게 $\frac{3}{4}$로 지며, 나머지 네 개(HTH, HTT, THH, THT)는 각각 HHT, HHT, TTH, TTH에게 $\frac{2}{3}$으로 진다. HHH와 TTT가 가장 쉽게 진다.

4. $A = \mathrm{HTT}$, $B = \mathrm{HHT}$라 하면 상관값은 $AA = 4$, $AB = 0$, $BB = 4$, $BA = 2$이다. B가 이길 오즈는 $(AA - AB) : (BB - BA) = 4 : 2$, 즉 확률 $\frac{2}{3}$이다.

12.9 치우친 동전으로 공정하게 정하기

1. 라운드 수는 기하 분포(geometric distribution)를 따른다. 평균은 소문항 (i)에서 $\frac{1}{p^2+q^2}$이고, 폰 노이만(von Neumann)의 절차(HT 또는 TH)에서는 $\frac{1}{2pq}$이다. $p^2 + q^2 - 2pq = (p-q)^2$이므로 모든 $p \neq \frac{1}{2}$에 대해 폰 노이만의 방법이 더 느리다. 그러나 공정한 것은 이 방법이다.

2. 폰 노이만의 방법으로 공정한 비트를 만든다. 두 개씩 묶어서 00, 01, 10으로 세 가지 선택지를 정하고, 11이면 다시 시작한다.

3. 던진 결과를 하나씩 이진 전개 $\frac{1}{3} = 0.010101\ldots$와 비교한다(H는 1, T는 0으로 읽는다). 처음으로 다른 자리가 나왔을 때, 던진 결과가 그 자리의 숫자보다 **작으면**(T와 1) "앞면", 그렇지 않으면 "뒷면"이라고 선언한다. 이렇게 하면 앞면이 나올 확률이 정확히 $\frac{1}{3}$이 되고, 평균 2번 던진다.

4. 결정이 나지 않은 라운드 뒤에는 한 면이 두 개, 다른 면이 한 개이고, 이 과정은 이 두 상태 위의 마르코프 연쇄(Markov chain)이다. 이를 풀면 $\mathrm{P}(\text{호텔}) = \frac{p^3}{p^3+q^3}$이고, 이는 $p = \frac{1}{2}$일 때 $\frac{1}{2}$이다. 그때에는 공정하지만 다른 경우에는 공정하지 않다.

12.10 느슨한 끈 끝 묶기

1. 첫 번째 매듭이 고리를 만들 확률은 $\frac{1}{3}$이고(다른 세 끝 가운데 하나가 같은 끈의 끝이다), 마지막 매듭은 항상 고리를 만든다. 고리의 수는 확률 $\frac{1}{3}$으로 2, 확률 $\frac{2}{3}$으로 1이고, 평균은 $\frac{4}{3}$이다.

2. 끈이 k개 남았을 때 매듭이 고리를 만들지 않을 확률은 $1-\frac{1}{2k-1}$이다. 곱 $\prod_{k=2}^{n}\frac{2k-2}{2k-1}$는 $\frac{4^n}{2n\binom{2n}{n}}$과 같고, 이는 약 $\frac{\sqrt{\pi n}}{2n}=\sqrt{\frac{\pi}{4n}}$이다.

3. 원소 $1,2,\ldots,n$을 차례로 끼워 넣는다. 원소 k는 새 순환(cycle)을 시작하거나(확률 $\frac{1}{k}$), 앞서 넣은 $k-1$개 원소 가운데 하나 뒤에 놓인다. 지시변수들을 더하면 H_n이다.

4. 연이은 매듭들의 고리 지시변수들은 확률이 $\frac{1}{2k-1}$인 독립 변수이므로 분산은 $\sum\frac{1}{2k-1}\left(1-\frac{1}{2k-1}\right)$이다. 그리고 $\sum_{k\ge1}\frac{1}{(2k-1)^2}=\frac{\pi^2}{6}-\frac{1}{4}\cdot\frac{\pi^2}{6}=\frac{\pi^2}{8}$이다.

12.11 우표, 컵, 그리고 e

1. $\sum_{k\ge2}\frac{k-1}{k!}t^k=te^t-(e^t-1)=1+(t-1)e^t$이다. 미분(differentiation)하면 $\mathrm{E}(N)=G'(1)=e$이고 $\mathrm{E}(N(N-1))=G''(1)=2e$이므로 $\mathrm{Var}(N)=3e-e^2$이다.

2. $\mathrm{P}(N=j)$는 $k=0,\ldots,j-1$에 대한 $\mathrm{P}(N>k)$에 나타나므로 j번 나타난다. 여기서 $\mathrm{P}(N>k)=\frac{1}{k!}$이므로 $\mathrm{E}(N)=\sum\frac{1}{k!}=e$이다.

3. 전혀 달라지지 않는다. 컵이 하나 남으면 Jim은 반드시 올바르게 "추측"한다. 가격의 기댓값은 여전히 e이다.

4. $\int_0^1\frac{(1-u)^2}{2}\,\mathrm{d}u=\frac{1}{6}$이다. 일반적으로 $u_k=u$에서의 단면은 $1-u$배로 축소한 $(k-1)$차원 단체(simplex)이므로 $V_k=V_{k-1}\int_0^1(1-u)^{k-1}\,\mathrm{d}u=\frac{V_{k-1}}{k}$이다.

12.12 언제 그만 찾아야 하는가

1. 가장 큰 것이 $k\ge2$번째 자리에 있으면, 첫 번째 아이스크림이 처음 $k-1$개 가운데 가장 클 때에만 그것을 얻으며, 그 확률은 $\frac{1}{k-1}$이다. k에 대해 평균하면 $\frac{1}{n}\sum_{k=2}^{n}\frac{1}{k-1}$이고, $n=4$이면 이는 $\frac{11}{24}$이다. 24가지 순서를 모두 나열하면 두 번째, 세 번째, 네 번째로 큰 것을 얻을 확률은 $\frac{7}{24}$, $\frac{1}{6}$, $\frac{1}{12}$이다.

2. $r=0,1,2,3$명을 거절하면 확률은 $\frac{1}{4}$, $\frac{11}{24}$, $\frac{5}{12}$, $\frac{1}{4}$이다. 따라서 한 명을 거절해야 하고, 성공 확률은 $\frac{11}{24}$이다.

3. 처음 n명 가운데 가장 나은 사람이 상위 두 명 안에 들 확률은 $\frac{n}{N}+\frac{n}{N}\cdot\frac{N-n}{N-1}$이고, n이 작을 때 이는 $\frac{n}{N}$의 거의 두 배이다. 상위 두 명 가운데 한 명이면 된다면, 결정하기 전에 살펴볼 후보를 더 적게 잡아도 된다. (이 목표에 대한 최선의 전략은 약 0.57의 확률로 성공한다.)

4. $\mathrm{P}(r) = \frac{r}{N}\sum_{k=r}^{N-1}\frac{1}{k}$를 쓰면, $N = 10$일 때 $r = 2, 3, 4$에서 0.366, 0.399, 0.398이고, $N = 100$일 때 $r = 36, 37, 38$에서 0.3710, 0.3710, 0.3708이며 $r = 37$이 근소하게 앞선다.

12.13 삼자 결투

1. Algernon이 빗나가면 Bertie는 Cuthbert를 향해 던진다. Bertie가 맞히면($\frac{3}{5}$) Algernon이 먼저 던지는 Algernon 대 Bertie의 대결이 되고, Bertie가 이길 확률은 $1 - \frac{2/5}{2/5+3/5-6/25} = \frac{9}{19}$이다. Bertie가 빗나가면 Cuthbert가 Bertie를 맞히고, 그다음 Algernon이 먼저 맞히지 않는 한($\frac{3}{5}$) Cuthbert가 이긴다. 따라서 Bertie가 이길 확률은 $\frac{3}{5}\cdot\frac{9}{19} = \frac{135}{475}$이고 Cuthbert가 이길 확률은 $\frac{2}{5}\cdot\frac{3}{5} = \frac{114}{475}$이다.

2. 모든 $a < 1$에 대해서이다. 빗나가면 $\frac{3}{5}\cdot\frac{a}{a+\frac{3}{5}-\frac{3}{5}a} + \frac{2}{5}a$이고, Cuthbert를 맞히면 $1 - \frac{3/5}{3/5+a-\frac{3}{5}a}$인데(그러면 Bertie가 먼저 Algernon을 향해 던진다) 이는 항상 더 작으며, Bertie를 맞히면 0이다.

3. 아니다. Bertie가 Algernon을 향해 던져 맞히면 그다음 Cuthbert가 Bertie를 맞히고, 빗나가도 Cuthbert가 어차피 Bertie를 맞힌다. 어느 쪽이든 Bertie는 아무것도 얻지 못하지만, Cuthbert를 향해 던지면 $\frac{27}{95}$을 얻는다.

4. $p + (1-p)(1-r)p$이고, 이는 기하급수(geometric series) 전체의 합인 $\frac{p}{p+r-pr}$보다 작다.

12.14 혼합 전략

1. $b = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$이면 $X = Y = Z = 0$이므로 B는 어떤 전략에 대해서든 0을 얻는다. $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$에 대해서는 A가 1, 2, 3을 추측할 때의 기대 이득이 각각 $-\frac{1}{3}$, 0, $\frac{1}{3}$이므로 A는 항상 3을 추측해야 한다.

2. 이득 행렬(payoff matrix)은 대각선에 0, 나머지에 ± 1이 있고, 세 전략을 순환적으로 치환해도 변하지 않으므로 최적 전략은 $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$이다. 문제에서는 차이가 1일 때와 2일 때의 벌점이 다르므로 전략들을 서로 바꿀 수 없다.

3. $X, Y, Z \geq 0$이고 $X + 2Y + Z = 0$이면 세 값이 모두 0이다. 그러면 $b_1 = b_3, b_2 = 2b_3$이어야 한다.

4. 이제 $X = -b_2 + 3b_3$, $Y = b_1 - b_3$, $Z = -3b_1 + b_2$이고 $X + 3Y + Z = 0$이다. 세 값을 모두 0으로 놓으면 $(\frac{1}{5}, \frac{3}{5}, \frac{1}{5})$를 얻는다.

12.15 멸종인가 폭발인가

1. 소멸 확률(extinction probability)은 $s = (q + ps)^2$의 $[0, 1]$에 있는 가장 작은 근이다. 근은 1과 $\left(\frac{q}{p}\right)^2$이고, 두 번째 근이 1보다 작은 것은 정확히 $p > \frac{1}{2}$일 때이다.

2. X_{n-1}이 주어졌을 때 X_n의 평균은 $2pX_{n-1}$, 분산은 $2pqX_{n-1}$이므로 $\mathrm{Var}(X_n) = \mathrm{E}(2pqX_{n-1}) + (2p)^2\mathrm{Var}(X_{n-1})$이다. $p = \frac{1}{2}$이면 $\mathrm{Var}(X_n) = \mathrm{Var}(X_{n-1}) + \frac{1}{2}$이므로 $\mathrm{Var}(X_n) = \frac{n}{2}$이다.

3. n세대까지 소멸한다는 것은 첫 세대 각 입자의 계보가 $n-1$세대 안에 끊긴다는 뜻이다. $p = \frac{3}{4}$이면 $e_1 = 0.0625$, $e_2 \approx 0.0881$, $e_3 \approx 0.0999$이고, $\frac{1}{9} \approx 0.111$을 향해 증가한다.

4. $s = \frac{1}{4}(1 + s + s^2 + s^3)$은 $(s-1)(s^2 + 2s - 1) = 0$으로 인수분해되므로 소멸 확률은 $\sqrt{2} - 1 \approx 0.414$이다.

12.16 앞으로 보고 거꾸로 보는 마르코프 연쇄

1. 대칭성에 의해 $p_B = p_D$이고, 또 $p_A = \frac{1}{3} + \frac{2}{3}p_B$, $p_B = \frac{1}{3}p_A + \frac{1}{3}p_C$, $p_C = \frac{2}{3}p_B$이다. 따라서 $p_A = \frac{7}{15}$이다. A에서 출발하여 C를 거치지 않고 A에서 떠나는 경우: $x_A = \frac{1}{3} + \frac{2}{3}x_B$, $x_B = \frac{1}{3}x_A$이므로 $x_A = \frac{3}{7}$이다. 조건부 확률(conditional probability)은 $\frac{3/7}{7/15} = \frac{45}{49}$이다.

2. 첫 번째 행에 P를 곱하면 $(0.19, 0.33, 0.48)$이다.

3. 대각합(trace)은 1.3, 행렬식(determinant)은 0.02이고, 1은 고윳값(eigenvalue)이다(각 행의 합이 1). 나머지 두 고윳값은 합이 0.3, 곱이 0.02이므로 0.2와 0.1이다.

4. $a_1 p_{12} = \frac{1}{20}$이지만 $a_2 p_{21} = \frac{1}{15}$이므로 이 연쇄는 가역(reversible)이 아니다. 거꾸로 돌리면 전이 확률(transition probability)이 달라진다(소문항 (iii)의 q_{ij}가 p_{ij}와 다르다).

12.17 바나흐의 성냥갑

1. 두 번째로 손을 뻗을 때 첫 번째와 같은 종류를 고르면 그때 봉지가 비었음을 알게 된다($N = 2$). 그렇지 않으면 다른 사탕을 꺼내고, 세 번째로 손을 뻗을 때 빈 봉지를 발견한다. 따라서 $N = 2$ 또는 3이고, 각각의 확률은 $\frac{1}{2}$이다.

2. $n = 1$: $\frac{1}{2} + \frac{1}{2} = 1$. $n = 2$: $\frac{6}{16} + \frac{6}{16} + \frac{4}{16} = 1$.

3. $\binom{2n-1}{n} 2^{-(2n-1)} = \frac{1}{2}\binom{2n}{n} \cdot 2 \cdot 4^{-n} = \binom{2n}{n} 4^{-n}$.

4. 남은 토피와 민트의 개수에 대해 첫 단계 점화식을 세우면 $\frac{64}{81} \approx 0.79$를 얻는다.

12.18 무기억성

1. $\mathrm{P}(\min > t) = \mathrm{e}^{-\lambda_1 t}\mathrm{e}^{-\lambda_2 t}$이다. 다음 버스를 기다리는 시간은 분당 비율이 $\frac{1}{15} + \frac{1}{30} = \frac{1}{10}$인 지수 분포(exponential distribution)를 따르므로, 그 중앙값(median)은 $10\ln 2 \approx 6.9$분이다.
2. $\mathrm{P}(Y \geq k) = (1-p)^k$이므로 비는 $(1-p)^n$이다.
3. t를 포함하는 간격은 직전 버스까지 거슬러 올라가는 시간과 다음 버스까지의 시간으로 이루어지며, (t가 클 때) 둘 다 평균이 $\frac{1}{\lambda}$인 지수 분포를 따르므로 그 길이의 기댓값은 $\frac{2}{\lambda}$이다. 긴 간격일수록 주어진 시각을 포함할 가능성이 크다.
4. $[0,1]$에서 $\frac{\lambda \mathrm{e}^{-\lambda z}}{1-\mathrm{e}^{-\lambda}}$이다. $\lambda = 1$이면 약 1.58에서 0.58로 완만하게 감소하고, $\lambda = 5$이면 0 근처에 몰려 있으며 약 5.03에서 0.03으로 감소한다.

12.19 평면 위의 무작위 점

1. $1 - \mathrm{e}^{-\pi k y^2} = \frac{1}{2}$에서 중앙값 $\sqrt{\frac{\ln 2}{\pi k}} \approx \frac{0.470}{\sqrt{k}}$를 얻는데, 이는 평균 $\frac{1}{2\sqrt{k}}$보다 조금 작다.
2. $\mathrm{P}(Y_2 > y) = \mathrm{e}^{-\pi k y^2}(1 + \pi k y^2)$이므로 확률밀도(probability density)는 $2\pi^2 k^2 y^3 \mathrm{e}^{-\pi k y^2}$이다.
3. 극좌표(polar coordinates)에서 R의 확률밀도는 $\frac{1}{2\pi\sigma^2}\mathrm{e}^{-r^2/(2\sigma^2)} \cdot 2\pi r$이다. 이는 $\sigma^2 = \frac{1}{2\pi k}$일 때 슈퍼마켓 분포와 일치한다.
4. m^2당 $k = 4 \times 10^{-4}$이면 완전한 무작위성에서 예측되는 평균은 $\frac{1}{2\sqrt{k}} = 25\,\mathrm{m}$이다. 관측값 35 m는 이보다 크므로, 나무들이 모여 있다기보다는 (경쟁 때문에) 서로 떨어져 있음을 시사한다.

12.20 뷔퐁의 바늘

1. 바늘을 짧은 조각들로 자른다. 교차 횟수의 기댓값은 더할 수 있고, 각 짧은 조각은 길이에 비례하여 기여하므로, a가 무엇이든 전체는 $\frac{2a}{\pi b}$이다.
2. 같은 가법성(뷔퐁(Buffon)의 국수 문제)에 의해 여전히 $\frac{2a}{\pi b}$이다.
3. 중심이 모든 변에서 적어도 r만큼 떨어져 있어야 하므로 $\frac{(s-2r)^2}{s^2}$이다.
4. $\frac{1808}{3408} \approx \frac{5}{3\pi}$에서 $\pi \approx \frac{5\times 3408}{3\times 1808} = \frac{17\,040}{5424} = \frac{355}{113}$을 얻는다. 의심스러울 만큼 완벽하다.

12.21 막대 부러뜨리기

1. 삼각형이 되려면 모든 조각이 $\frac{1}{2}$보다 짧아야 한다. 부러지는 점들의 단위 정사각형에서 이 조건은 넓이가 $\frac{1}{4}$인 영역을 잘라 낸다.

2. 첫 번째로 부러지는 점이 $x < \frac{1}{2}$에 있고(다른 경우는 대칭이다), 더 긴 조각 $1-x$를 균등하게 부러뜨린다고 하자. 새로 생긴 두 조각이 모두 $\frac{1}{2}$보다 짧아야 하며, 그 확률은 $\frac{x}{1-x}$이다. 따라서 답은 $2\int_0^{1/2} \frac{x}{1-x}\,\mathrm{d}x = 2\ln 2 - 1$이다.

3. C에서의 각이 둔각인 것은 $(2-x-y)^2 > x^2+y^2$, 즉 $y < \frac{2-2x}{2-x}$일 때이다. 이 곡선과 $x+y=1$ 사이의 넓이는 $\frac{3}{2} - 2\ln 2$이므로 확률은 $3-4\ln 2$이고, B에서와 같다. 이는 우연이 아니다. 세 변은 각각 1 이하이고 합이 2이며 균등하게 분포하므로 서로 바꿀 수 있다. 따라서 $\mathrm{P}(\text{둔각}) = 9 - 12\ln 2 \approx 0.68$이다.

4. 포함배제 원리(inclusion–exclusion principle)에 의해 가장 긴 조각 M은 $\mathrm{P}(M > m) = 3(1-m)^2 - 3(1-2m)_+^2 + (1-3m)_+^2$을 만족한다. 여기서 $x_+ = \max(x,0)$이다. $[0,1]$에서 적분(integration)하면 $\mathrm{E}(M) = 1 - \frac{1}{2} + \frac{1}{9} = \frac{11}{18}$이다.

12.22 평균이 없는 분포

1. $\mathrm{P}(\tan\Theta \le k) = \frac{1}{\pi}\left(\tan^{-1}k + \frac{\pi}{2}\right)$이고, 미분하면 $\frac{1}{\pi(1+k^2)}$이다.

2. U와 V를 맞바꾸어도 방사 대칭인 확률밀도는 보존되므로 $\frac{V}{U}$는 $\frac{U}{V}$와 같은 분포를 따른다.

3. 확률밀도가 대칭이므로 중앙값은 0이고, $\mathrm{P}(C \le 1) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}1 = \frac{3}{4}$이다.

4. $\frac{1}{2\pi}\mathrm{e}^{-r^2/2}\,r\,\mathrm{d}r\,\mathrm{d}\theta$이고, $s = r^2$으로 놓으면 이는 $\frac{1}{2\pi}\,\mathrm{d}\theta \cdot \frac{1}{2}\mathrm{e}^{-s/2}\,\mathrm{d}s$이다. 즉 각은 균등 분포, R^2은 평균이 2인 지수 분포를 따르며, 둘은 독립이다.

12.23 확률로 얻는 스털링 공식

1. $\sqrt{20\pi}(10/\mathrm{e})^{10} \approx 3\,598\,696$(약 0.8% 작음)이고 $\sqrt{40\pi}(20/\mathrm{e})^{20} \approx 2.4228 \times 10^{18}$(약 0.4% 작음)이다.

2. $\frac{(2n)!}{(n!)^2} \approx \frac{\sqrt{4\pi n}(2n/\mathrm{e})^{2n}}{2\pi n(n/\mathrm{e})^{2n}} = \frac{4^n}{\sqrt{\pi n}}$.

3. $100\log_{10}\frac{100}{\mathrm{e}} + \frac{1}{2}\log_{10}(200\pi) \approx 157.97$이므로 100!은 158자리 수이다.

4. $\binom{1000}{500}2^{-1000} \approx \frac{1}{\sqrt{500\pi}} \approx 0.0252$.

12.24 중심극한정리

1. 합성곱(convolution)을 하면 $\frac{1}{4}\int \mathrm{e}^{-|x|-|y-x|}\,\mathrm{d}x = \frac{1}{4}(1+|y|)\mathrm{e}^{-|y|}$이다. 분산은 $2+2=4$이다.

2. $\mathrm{E}(\mathrm{e}^{\theta X}) = \frac{1}{2}(\mathrm{e}^{\theta/2} + \mathrm{e}^{-\theta/2}) = \cosh\frac{\theta}{2}$이다. 표준화한 합 $\frac{2}{\sqrt{n}}\sum X_i$의 생성함수는 $\cosh^n\frac{\theta}{\sqrt{n}}$이고, 그 로그는 $n\ln\left(1 + \frac{\theta^2}{2n} + \cdots\right) \to \frac{\theta^2}{2}$이다.

3. Y를 서로 독립인 두 $\Gamma(82,1)$ 확률변수의 차로 쓰고 수치적으로 적분하면 $\mathrm{P}(|Y| \geq 25) \approx 0.051$을 얻는다. 정규 근사(normal approximation)가 매우 좋다.

4. 적률생성함수(moment generating function)가 없고, 평균조차 없다. 사실 서로 독립인 n개의 코시(Cauchy) 확률변수의 평균은 그 자체로 코시 분포를 따른다. 평균을 내도 전혀 한곳으로 모이지 않는다.

12.25 여럿 가운데 가장 큰 값

1. n개의 수명 가운데 첫 번째 수명은 평균 시간 $\frac{1}{n\lambda}$ 뒤에 끝나고, 그다음에는 $n-1$개가 남아 새로 시작하며, 이런 식으로 계속된다. 따라서 $\mathrm{E}(Y) = \frac{1}{\lambda}\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\right)$이다.

2. $\mathrm{P}(\min > t) = (\mathrm{e}^{-\lambda t})^n = \mathrm{e}^{-n\lambda t}$.

3. $(1-\mathrm{e}^{-\lambda y})^n = \frac{1}{2}$에서 $\lambda y = -\ln(1-2^{-1/n})$을 얻는다. $n = 10$이면 이는 2.704이고, 근삿값은 $\ln 10 - \ln\ln 2 \approx 2.669$이다.

4. $\mathrm{P}(M \leq m) = m^n$이므로 $\mathrm{P}\big(n(1-M) > s\big) = \big(1-\frac{s}{n}\big)^n \to \mathrm{e}^{-s}$이다. 이것은 와이블(Weibull) 유형이다. 확률변수들이 위로 유계(bounded above)이고, 최댓값이 그 상계 쪽으로 몰린다.

Beyond STEP

STEP은 대부분의 학생이 학창 시절에 치르는 수학 시험 가운데 가장 어려운 시험이다. 또한 가장 훌륭한 시험 가운데 하나이기도 하다. 문제는 짧고 낯설며 깊이가 있고, 그중 가장 뛰어난 문제들은 대학 수학의 작은 조각이 모습을 감추고 있는 것이다.

이 책은 1987년 첫 시험에 앞서 나온 예시 시험지(specimen papers)부터 2025년 시험지까지 가운데 가장 흥미로운 STEP 문제들을 골라, 각 문제를 하나의 문으로 삼는다. 모든 문제는 뛰어난 학생이 풀어 가면서 던질 법한 질문들과 함께 꼼꼼하게 풀이하고, 그런 다음 책은 그 문을 지나 한 걸음 더 나아간다. 그 아이디어가 어디에서 왔는지, 대학에서는 무엇으로 발전하는지, 그리고 다음에는 무엇을 읽으면 좋은지를 살펴본다.

이 책은 STEP을 준비하면서 기출 문제 연습 이상의 것을 원하는 재능 있는 예비 수학자를 위해 쓰였다. 수학과 학부생처럼 생각하기 시작하고, 앞으로 무엇이 자신을 기다리고 있는지 알아보고 싶은 학생들을 위한 책이다.

- 순수수학(pure mathematics), 역학(mechanics), 확률(probability)과 통계(statistics)를 아우르는 13개의 장
- 이 36개의 STEP 기출 문제와 각 문제의 안내식 풀이
- 모든 문제 뒤에 숨은 대학 수학
- 답이 실린 후속 연습 문제와 더 읽을거리

전체 13개 장 가운데 제12장. 책 전체와 모든 장,
모든 풀이를 다음 주소에서 무료로 읽고, 내려받고, 공유할 수 있다.

beyondstep.keithmcnulty.org

본문은 퍼블릭 도메인(CC0 1.0)으로 공개한다. STEP 문제 © Cambridge University Press & Assessment.