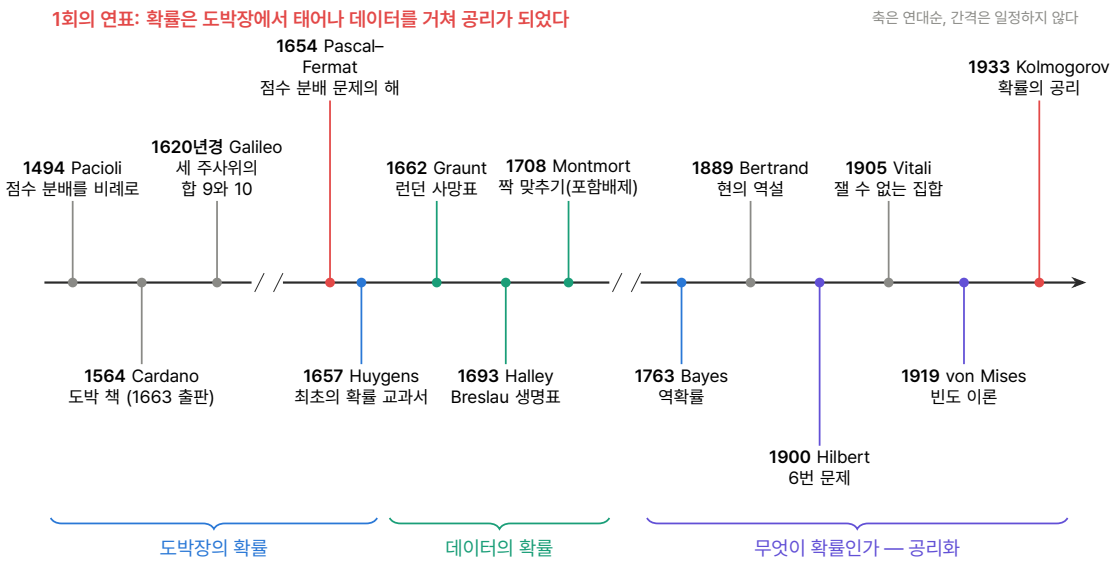
수리통계학 × 통계학의 역사

1회 강의노트

# 확률이란 무엇인가 도박장에서 공리까지

## 1494 → 1933

확률은 멈춘 내기의 판돈을 나누는 문제로 태어났고, 사망 기록을 읽는 기술로 자랐으며, "무작위로"라는 말이 모순을 낳자 1933년에야 공리 위에 섰다. 이 노트는 그 순서를 따라가며 셈의 원리, 확률의 공리와 그 결과, 조건부확률과 독립, 확률변수와 분포함수를 처음부터 세운다.




| | |
|---|---|
| **다루는 것** | 셈의 원리 · 확률의 공리 · 포함배제 · 연속성 · 조건부확률 · 베이즈 · 독립 · 확률변수 · 분포함수 |
| **원자료** | Halley(1693)의 Breslau 생명표 |
| **노트북** | `sessions/01_probability/01_probability.ipynb` — 절·정리 번호가 같다 |
| **워크북** | `01_probability_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf` |
| **선수 지식** | 집합, 수열과 급수의 극한, 정적분 (고등학교·1학년 미적분) |

## 이 노트를 쓰는 법

이 노트는 다른 책 없이 읽도록 썼다. 쓰이는 정의와 정리는 모두 이 노트나 앞 회차의 노트에서 증명한다. 이 과정의 범위를 넘어 증명하지 않는 사실은 **증명 없이 쓰는 사실** 상자에 따로 밝힌다.

각 절은 **역사 → 정의·정리 → 증명 → 그림 → 연습문제** 순서다. 핵심 정리의 증명은 번호 붙은 단계 ① ② ③으로 쓰고, 단계마다 오른쪽 끝에 근거(앞 정리의 번호나 공리)를 단다. 읽을 때는 근거를 가리고 "왜 이 단계가 성립하는가"를 먼저 말해 본다. 나머지 정리는 **증명 전략**만 적고, 증명을 소문항으로 나눈 유도형 연습으로 돌린다. 이 연습은 건너뛰지 않고 손으로 쓴다.

근거에 붙은 [T1]–[T8]은 별책 **증명 도구 상자**의 도구 번호다(원소 추적, 귀류, 귀납, 부등식, 극한, 극한의 교환, 변수변환, Taylor 전개). 이 회차에서는 [T1], [T2], [T3], [T5]를 쓴다.

연습문제는 소절 끝에 모여 있다. 난이도는 ★ 계산·확인, ★★ 증명·응용, ★★★ 도전이다. 모든 문제의 풀이가 부록 A에 있고, 문제 끝의 **풀이 ○○쪽**을 누르면 그 쪽으로 간다. 손으로 푸는 문제는 따로 **워크북**(계산 드릴, 빈칸 증명, 증명 문제, 퀄형)에 모았다.

### 상자의 뜻

> **정의** 대상을 처음 정하는 문장. 왼쪽 선이 보라다.

> **정리** 증명할 주장(정리·보조정리·따름정리·명제). 왼쪽 선이 빨강이다.

> **예제** 숫자를 넣어 끝까지 계산해 보는 곳. 왼쪽 선이 파랑이다.

> **가정 주의** 빼면 정리가 무너지는 가정. 왼쪽 선이 주황이다.

> **증명 없이 쓰는 사실** (예) 이 과정에서 증명하지 않고 쓰는 결과. 무엇이 참인지, 왜 그럴듯한지, 어디서 증명되는지를 밝힌다.

> **역사** 그 개념이 태어난 사람과 문제. 회색 선 옆의 작은 글씨다.

> **연습** 소절 끝의 연습문제. 가는 청록 선이고, 끝에 풀이가 있는 쪽 번호가 붙는다.

**색의 뜻** 수식과 그림에서 보라는 모수, 청록은 데이터·확률변수, 파랑은 통계량, 빨강은 핵심 결론이다. 노트북과 같은 색이다.

# 차례


**0 들어가며 1**

0.1 집합의 언어 . . . 2

**1 도박장에서 태어난 확률 4**

1.1 점수 분배 문제와 Pacioli의 답 . . . 4

1.2 셈의 원리 . . . 4

1.3 de Méré의 두 내기 . . . 7

1.4 Pascal과 Fermat의 해 (1654) . . . 8

1.5 Huygens (1657): 기대 몫 . . . 10

**2 데이터에서 온 확률 12**

2.1 Graunt의 런던 사망표 (1662) . . . 12

2.2 Halley의 Breslau 생명표 (1693) . . . 12

**3 "무작위로"의 위기 15**

3.1 Bertrand의 현 역설 (1889) . . . 15

3.2 Bertrand의 상자 역설 . . . 16

3.3 Hilbert의 6번 문제 (1900) . . . 17

**4 Kolmogorov의 공리 (1933) 18**

4.1 표본공간과 사건 . . . 18

4.2 확률의 공리 . . . 19

4.3 공리에서 나오는 성질 . . . 19

4.4 포함배제의 원리 . . . 20

4.5 확률의 연속성 . . . 22

4.6 이산 확률공간과 고전적 정의 . . . 24

**5 조건부확률과 독립 28**

5.1 조건부확률 28

5.2 곱셈법칙 · 전확률 공식 · 베이즈 정리 28

5.3 베이즈 정리로 푸는 두 문제 30

5.4 독립 30

5.5 점수 분배, 다시: Pascal의 재귀 33

**6 확률변수와 분포함수 36**

6.1 확률변수 36

6.2 분포함수의 성질 37

6.3 이산 확률변수와 연속 확률변수 38

**7 부쉬보기 — 공리의 가정을 빼면 42**

7.1 자연수 위에는 "균등한" 확률이 없다 42

7.2 모든 부분집합에 확률을 줄 수는 없다 (Vitali, 1905) 43

7.3 확률 0인 사건에 조건을 걸면 43

**8 Beautiful insights — 1회의 수학은 어디에 있나 45**

8.1 시뮬레이션은 상대도수 해석이다 45

8.2 기저율 오류는 분류기의 precision이다 45

8.3 베이즈 정리와 naive Bayes 분류기 46

8.4 생일 문제와 해시 충돌 46

**9 정리와 역사 노트 47**

9.1 한 장 요약 47

9.2 Stigler의 명명 법칙 48

9.3 7 Pillars 지도 48

9.4 다음 회 — 기댓값·부등식·대수의 법칙 (1657 → 1867) 48

**A 연습문제 풀이 50**

# 0

# 들어가며

> **여는 질문.** 두 사람이 먼저 여섯 판을 이기는 쪽이 판돈을 모두 갖기로 하고 내기를 하다가, A가 5판, B가 3판을 이긴 상태에서 그만두어야 했다. 판돈을 어떻게 나눠야 공평한가?

1494년 이탈리아의 수학자 Luca Pacioli는 이긴 판 수에 비례해 5 : 3으로 나누라고 답했다. 그럴듯하지만 틀렸다. 바른 답은 7 : 1이고(1.4절), 이 답에 이르는 데 160년이 걸렸다. 1654년 Pascal과 Fermat가 편지를 주고받으며 이 문제를 푼 해를 사람들은 확률론이 태어난 해로 친다.

그런데 그들이 쓴 "확률"이란 정확히 무엇이었나? 같은 경우의 수 가운데 유리한 경우의 비율이었다. 이 정의는 주사위와 카드에서는 잘 작동했지만, 경우가 무한히 많아지자 무너졌다. 1889년 Bertrand는 "원에 무작위로 그은 현"이라는 한 문장에서 서로 다른 세 답을 끌어냈다(3절). 확률이 무엇인지에 대한 답은 1933년 Kolmogorov의 공리(4절)로 나왔다. 확률론은 쓰이기 시작하고 거의 300년이 지나서야 기초를 얻었다.

이 노트는 그 순서를 따른다. 1–3절은 문제가 생긴 역사이고, 4–6절은 그 문제를 푸는 엄밀한 언어다. 4–6절의 정의와 정리가 이 과정 전체의 출발점이다.

## 학습 목표

1. 셈의 원리(곱의 법칙, 순열, 조합, 이항정리)를 증명하고 고전적 확률 계산에 쓴다.
2. 점수 분배 문제를 Fermat의 열거와 Pascal의 재귀 두 가지로 풀고, 두 풀이가 같음을 보인다.
3. 확률의 공리에서 여사건·단조성·포함배제·연속성 정리를 증명한다.
4. 조건부확률, 곱셈법칙, 전확률 공식, 베이즈 정리, 독립을 정의하고 증명한다. 쌍별 독립과 상호 독립을 구별한다.
5. 확률변수와 분포함수를 정의하고, 분포함수의 세 성질을 연속성 정리로 증명한다.
6. 공리의 가정 하나하나가 왜 필요한지 반례로 보인다.

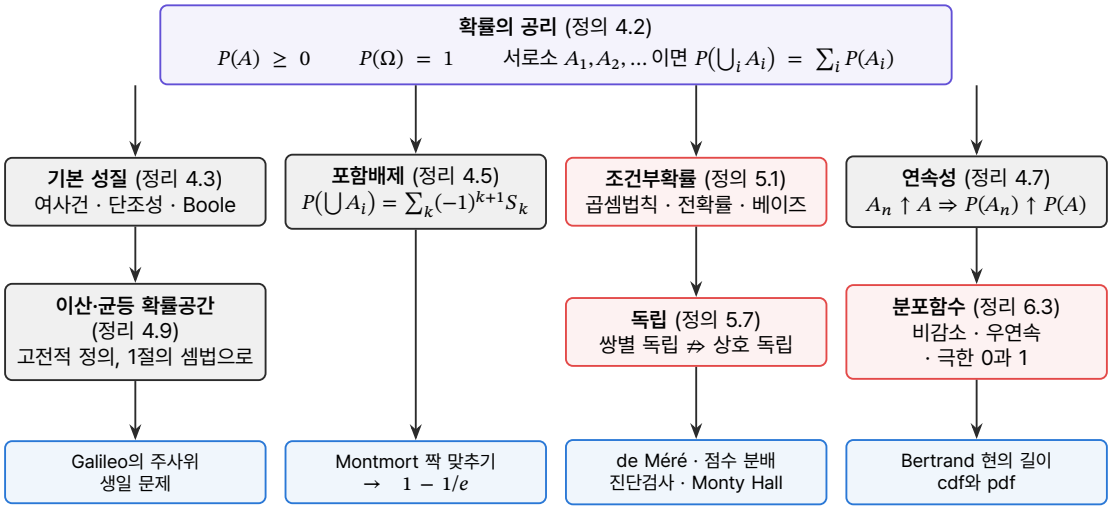



**그림 0.1** 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

## 0.1 집합의 언어

확률은 **사건**에 붙는 수이고, 사건은 가능한 결과들의 모임, 곧 집합이다. 이 노트에서 쓰는 집합 기호를 정리한다.

| 기호 | 뜻 | 기호 | 뜻 |
|---|---|---|---|
| $\omega \in A$ | $\omega$는 $A$의 원소 | $A \subset B$ | $A$의 원소는 모두 $B$의 원소 |
| $A \cup B$ | 합집합: $A$ 또는 $B$ | $A \cap B$, $AB$ | 교집합: $A$이고 $B$ |
| $A^c$ | 여집합: $A$가 아닌 것 | $A \setminus B$ | 차집합 $A \cap B^c$ |
| $\varnothing$ | 공집합 | $\lvert A \rvert$ | 유한집합 $A$의 원소 수 |
| $\bigcup_i A_i$ | 적어도 하나의 $A_i$ | $\bigcap_i A_i$ | 모든 $A_i$ |

$A \cap B = \varnothing$이면 $A$와 $B$는 **서로소**라 한다. $A_1, A_2, \ldots$가 쌍마다 서로소이고 합집합이 $\Omega$이면 $\{A_i\}$를 $\Omega$의 **분할**이라 한다.

**보조정리 0.1** (De Morgan의 법칙) 임의의(유한 또는 무한) 집합족 $\{A_i\}_{i\in I}$에 대해

$$\Big(\bigcup_{i\in I} A_i\Big)^c = \bigcap_{i\in I} A_i^c, \qquad \Big(\bigcap_{i\in I} A_i\Big)^c = \bigcup_{i\in I} A_i^c.$$

**증명.** **전략** 두 집합이 같다는 것은 서로 포함한다는 것이다. 왼쪽 집합의 원소를 하나 잡아 오른쪽 집합에 속함을 보이고, 거꾸로도 보인다 [T1]. 둘째 식은 첫째 식을 여집합들에 쓴다. 단계별 증명은 연습 0.1(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 50쪽에 있다.

이 노트에서 증명은 두 가지 모양으로 나온다. 핵심 정리는 번호 붙은 단계 ① ② ③으로 증명하고, 단계마다 오른쪽 끝에 근거를 단다(예: 정리 4.7). 나머지 정리는 위처럼 전략만 적고, 증명을 소문항으로 나눈 유도형 연습으로 돌린다. 근거에 붙은 [T1] 같은 번호는 별책 **증명 도구 상자**의 도구 번호다.

### 연습문제

**연습 0.1** (De Morgan의 법칙 증명 ★ 유도형) 보조정리 0.1을 다음 순서로 증명하라.

(a) $\omega \in (\bigcup_i A_i)^c$이면 모든 $i$에서 $\omega \in A_i^c$임을 보여라. 따라서 $(\bigcup_i A_i)^c \subset \bigcap_i A_i^c$이다.

(b) 거꾸로 $\omega \in \bigcap_i A_i^c$이면 $\omega \in (\bigcup_i A_i)^c$임을 보여라.

(c) 둘째 식은 첫째 식을 집합족 $\{A_i^c\}_{i\in I}$에 쓰고 양변의 여집합을 취해 얻어라. $(A^c)^c = A$를 어디서 쓰는지 밝혀라.

풀이 50쪽

# 1

# 도박장에서 태어난 확률

확률의 첫 계산은 모두 같은 모양이었다. 주사위나 카드에서 일어날 수 있는 경우를 빠짐없이 세고, 그 가운데 원하는 경우의 비율을 구한다. 그러려면 먼저 잘 세는 법이 필요하다. 이 절은 셈의 원리를 증명하고, 그것으로 17세기의 세 문제(Galileo의 주사위, de Méré의 내기, 점수 분배)를 푼다.

## 1.1 점수 분배 문제와 Pacioli의 답

> **역사** Luca Pacioli는 1494년 *Summa de arithmetica*에서 멈춘 내기의 판돈을 이긴 판 수에 비례해 나누라고 했다. 1556년 Tartaglia는 이 규칙이 이상한 답을 낸다고 지적했다. 먼저 6판을 이기는 내기에서 1:0으로 멈추면 Pacioli의 규칙은 판돈 전부를 A에게 준다. 겨우 한 판을 앞섰을 뿐인데도 그렇다. Tartaglia와 Cardano도 다른 규칙을 내놓았지만 맞는 답은 아니었다.

문제를 정확히 쓰자. A와 B가 여러 판을 하고, 먼저 정해진 수의 판을 이기는 쪽이 판돈을 모두 갖는다. 판마다 A가 이길 확률은 $p$, B가 이길 확률은 $q = 1 - p$이고 판들의 결과는 서로 영향을 주지 않는다. 내기가 멈춘 순간 A가 이기려면 $a$판, B가 이기려면 $b$판을 더 이겨야 한다. 이 상태를 $(a, b)$로 쓴다.

Pascal과 Fermat가 받아들인 **공평함의 기준**은 다음과 같다. 판돈을 **계속했다면 각자가 이겼을 확률**에 비례해 나눈다. 여는 질문의 상태는 $(a, b) = (1, 3)$이다. Pacioli의 답 5 : 3이 이 기준과 어떻게 다른지는 1.4절에서 본다.

## 1.2 셈의 원리

> **정리 1.1** (곱의 법칙) $k$단계로 이루어진 선택에서 1단계의 방법이 $n_1$가지이고, 앞 단계에서 무엇을 골랐든 $j$단계의 방법이 $n_j$가지이면, 전체 선택의 방법은 $n_1 n_2 \cdots n_k$가지다.

**증명.** $k$에 대한 귀납법. $k = 1$이면 자명하다. $k - 1$단계까지의 선택이 $n_1 \cdots n_{k-1}$가지라고 하자. 그 각각에 $k$단계의 $n_k$가지가 이어 붙으므로, 전체는 $n_k$개씩 묶인 $n_1 \cdots n_{k-1}$개의 묶음이고 원소 수는 $n_1 \cdots n_{k-1} \cdot n_k$다. ∎

**정리 1.2** (순열) 서로 다른 $n$개에서 $k$개를 골라 한 줄로 늘어놓는 방법의 수는

$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

이다. 특히 $n$개 전부를 늘어놓는 방법은 $n!$가지다($0! = 1$로 둔다).

**증명.** 첫 자리에 $n$가지, 둘째 자리에는 남은 $n-1$가지, $\cdots$, $k$째 자리에는 $n-k+1$가지가 있다. 정리 1.1로 곱한다. ∎

**정의 1.3** (조합) 서로 다른 $n$개에서 순서를 생각하지 않고 $k$개를 고르는 방법의 수, 곧 $n$개짜리 집합의 $k$개짜리 부분집합의 수를 $\binom{n}{k}$로 쓴다. $k < 0$이거나 $k > n$이면 $\binom{n}{k} = 0$으로 둔다.

**정리 1.4** (조합의 수) $0 \le k \le n$이면

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

**증명.** $k$개를 골라 늘어놓는 방법(정리 1.2)을 두 단계로 센다. 먼저 $k$개짜리 부분집합을 고르고($\binom{n}{k}$가지), 고른 $k$개를 늘어놓는다($k!$가지). 곱의 법칙으로 $\binom{n}{k}\,k! = \frac{n!}{(n-k)!}$이다. ∎

**정리 1.5** (Pascal의 항등식) $n \ge 1$이면

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

**증명.** **전략** $\{1, \dots, n\}$의 $k$개짜리 부분집합을 원소 $n$을 포함하는 것과 포함하지 않는 것으로 나눠 센다. 단계별 증명은 연습 1.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 51쪽에 있다.

**정리 1.6** (이항정리) $n \ge 0$이고 $x, y$가 실수이면

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

**증명.** **전략** $n$에 대한 귀납법 [T3]. $(x+y)^{n+1} = (x+y)(x+y)^n$을 전개하고, 같은 거듭제곱의 계수를 Pascal의 항등식(정리 1.5)으로 합친다. 단계별 증명은 연습 1.9(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 52쪽에 있다.

**따름정리 1.7** (이항계수의 합) 모든 $n \geq 0$에서 첫 식이, $n \geq 1$이면 둘째 식이 성립한다.

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \qquad \sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0.$$

**증명.** 정리 1.6에 $x = y = 1$, 그리고 $x = -1,\ y = 1$을 넣는다. ∎

**예제 1.8** (Galileo의 세 주사위 (1620년 무렵)) 도박꾼들은 세 주사위의 합이 9가 되는 경우와 10이 되는 경우가 모두 여섯 가지라고 셌다.

$$9: \quad \{1,2,6\},\ \{1,3,5\},\ \{1,4,4\},\ \{2,2,5\},\ \{2,3,4\},\ \{3,3,3\},$$
$$10: \quad \{1,3,6\},\ \{1,4,5\},\ \{2,2,6\},\ \{2,3,5\},\ \{2,4,4\},\ \{3,3,4\}.$$

그런데 경험상 10이 더 자주 나왔다. Galileo의 설명은 이렇다. 같은 확률로 일어나는 것은 이런 **조합**이 아니라 순서까지 구별한 $6^3 = 216$가지 결과다(주사위에 빨강·파랑·초록 색을 칠했다고 생각하면 된다). $\{1,2,6\}$처럼 세 수가 다르면 순서가 $3! = 6$가지, $\{1,4,4\}$처럼 둘이 같으면 3가지, $\{3,3,3\}$은 1가지다. 따라서

$$\#\{\text{합} = 9\} = 6+6+3+3+6+1 = 25, \qquad \#\{\text{합} = 10\} = 6+6+3+6+3+3 = 27,$$

이고 확률은 $\frac{25}{216} \approx 0.116$과 $\frac{27}{216} = 0.125$다.

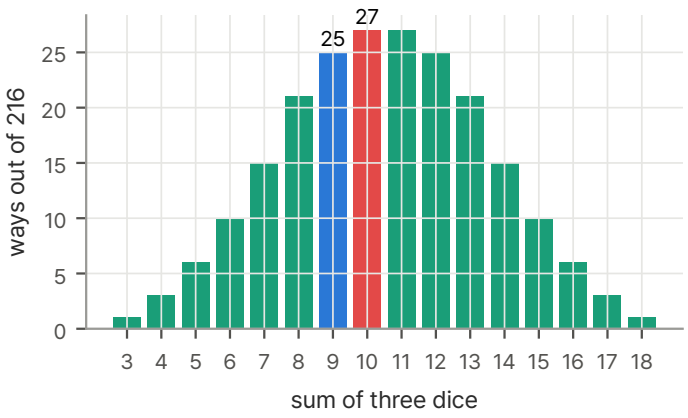



**그림 1.1** 세 주사위의 합. 216가지 순서쌍 가운데 합이 9인 것(파랑)은 25가지, 10인 것(빨강)은 27가지다.

**교훈** 확률을 "유리한 경우 ÷ 전체 경우"로 계산하려면 **같은 확률로 일어나는 경우**를 세야 한다. 무엇이 같은 확률인지는 수학이 정해 주지 않는다. 4.6절에서 이것을 "균등 확률공간"이라는 가정으로 명시하고, 3절에서 이 가정이 무너지는 경우를 본다.

## 1.3 de Méré의 두 내기

**역사** 파리의 사교계 인사 Chevalier de Méré(본명 Antoine Gombaud)는 1654년 Pascal에게 두 문제를 냈다. 하나는 점수 분배 문제였고, 다른 하나는 아래의 주사위 내기였다. de Méré는 두 내기가 같은 정도로 유리하다고 믿었는데 경험은 그렇지 않았다.

**예제 1.9** (주사위 4번 vs 주사위 두 개 24번) **내기 1**: 주사위 하나를 4번 던져 6이 적어도 한 번 나오면 이긴다. **내기 2**: 주사위 두 개를 24번 던져 (6,6)이 적어도 한 번 나오면 이긴다. de Méré의 추론은 비례였다. 주사위 하나는 6가지, 두 개는 36가지이므로, 4번이 하나짜리에 맞다면 두 개짜리에는 $4 \times 6 = 24$번이 맞다. 그러나 정확한 값은 "한 번도 나오지 않는다"의 확률을 빼서 얻는다.

$$P(\text{내기 1}) = 1 - \left(\frac{5}{6}\right)^4 = \frac{671}{1296} \approx 0.5177, \qquad P(\text{내기 2}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914.$$

내기 1은 유리하고 내기 2는 불리하다. 여기서 쓴 두 가지, 곧 "적어도 한 번"의 확률이 $1 - P(\text{한 번도 없음})$라는 것(정리 4.3(c))과 던질 때마다 확률을 곱해도 된다는 것(독립, 5.4절)은 4·5절에서 증명한다.

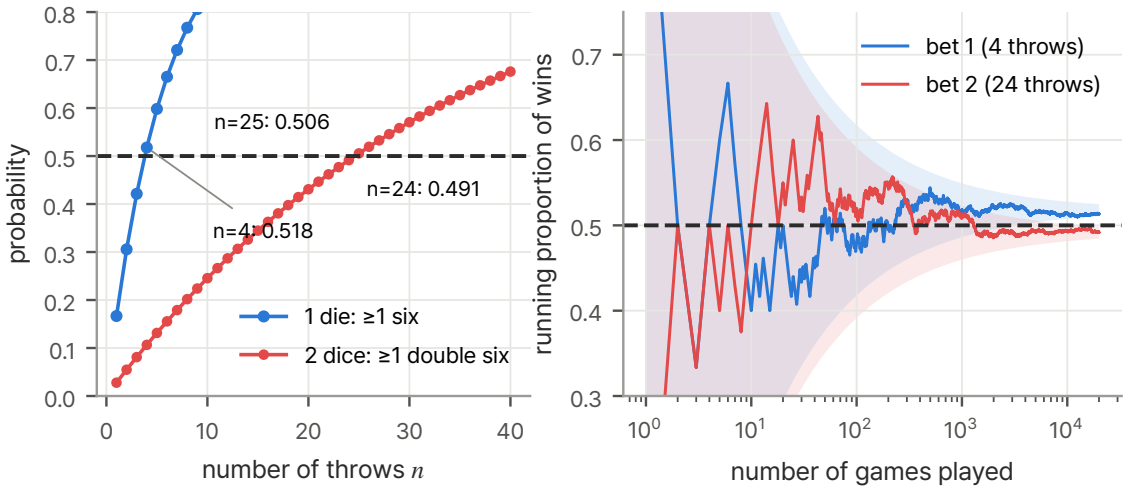



**그림 1.2** 왼쪽: 던지는 횟수에 따른 두 내기의 승률. 두 개짜리는 25번부터 $\frac{1}{2}$을 넘는다. 오른쪽: 각 내기를 2만 번 시뮬레이션한 누적 승률과 $p \pm 2\sqrt{p(1-p)/n}$ 띠(8.1절).

**비례 추론은 왜 거의 맞았나** 한 번에 성공할 확률이 $p$일 때 $n$번 안에 한 번이라도 성공할 확률이 $\frac{1}{2}$을 넘는 최소의 $n$은 $n \geq \log 2/(-\log(1-p))$이다. $p$가 작으면 $-\log(1-p) \approx p$이므로 $n \approx 0.693/p$, 곧 $n$은 $1/p$에 비례한다. 이것이 de Méré의 비례 규칙이다. $p = \frac{1}{36}$이면 근사 24.95, 정확한 값 24.6이라 규칙이 거의 맞지만, $p = \frac{1}{6}$이면 근사 4.16, 정확한 값 3.80이라 어긋난다. 두 내기의 기준을 같은 비례로 정한 것이 문제였다.

## 1.4 Pascal과 Fermat의 해 (1654)

**역사** 1654년 여름 Pascal과 Fermat는 편지로 점수 분배 문제를 풀었다. 7월 29일 Pascal의 편지에는 판돈이 한 사람당 32피스톨(합 64)인 예가 있다. 먼저 세 판을 이기는 내기에서 2:1이면 앞선 사람이 48, 2:0이면 56, 1:0이면 44를 가져야 한다. Fermat는 경우를 모두 세는 방법으로, Pascal은 한 판씩 거꾸로 따라가는 방법으로 같은 답에 이르렀다.

**정리 1.10** (점수 분배 — Fermat의 공식) 상태 $(a, b)$에서 A가 결국 이길 확률은

$$f(a,b) = \sum_{k=a}^{a+b-1} \binom{a+b-1}{k} p^{k}\, q^{a+b-1-k}.$$

**증명.** **1단계: 가상의 연장.** $m = a + b - 1$로 둔다. 승부가 이미 났더라도 정확히 $m$판을 끝까지 한다고 상상한다. 이렇게 해도 승자는 바뀌지 않는다. $m$판 가운데 A가 이긴 판 수를 $K$라 하면 B는 $m-K$판을 이긴다. $K \geq a$이면 $m-K \leq b-1$이므로 A가 B보다 먼저 $a$승에 이른다. $K \leq a-1$이면 $m-K \geq b$이므로 B가 먼저 $b$승에 이른다. 실제 승부는 $m$판 안에 반드시 끝나고, 그 뒤의 판은 결과를 바꾸지 않는다. 따라서 "A가 이긴다"는 "$K \geq a$"와 같다.

**2단계: 수열을 센다.** $m$판의 결과를 A/B로 쓴 길이 $m$의 수열 가운데 A가 정확히 $k$번 나오는 것은 $\binom{m}{k}$개다(A가 나올 자리를 고른다). 판들이 서로 영향을 주지 않으므로 그런 수열 하나의 확률은 $p^k q^{m-k}$다(이 곱셈은 5.4절의 독립으로 정당화된다).

**3단계: 더한다.** $K \geq a$인 경우를 모두 더하면 결론이 나온다. ∎

**예제 1.11** (Pacioli와 Pascal의 차이) 여는 질문에서 A는 1판, B는 3판이 더 필요하다($p = \frac{1}{2}$). B가 이기려면 남은 판을 세 번 내리 이겨야 하므로

$$f(1,3) = 1 - \left(\frac{1}{2}\right)^{3} = \frac{7}{8}.$$

판돈은 Pacioli의 5 : 3이 아니라 7 : 1로 나눠야 한다. Pascal의 편지의 수도 같은 식으로 나온다. 2:1이면 $(a,b) = (1,2)$이고 $64 \cdot \frac{3}{4} = 48$, 2:0이면 $(1,3)$이고 $64 \cdot \frac{7}{8} = 56$, 1:0이면 $(2,3)$이고 $64 \cdot \frac{11}{16} = 44$다.

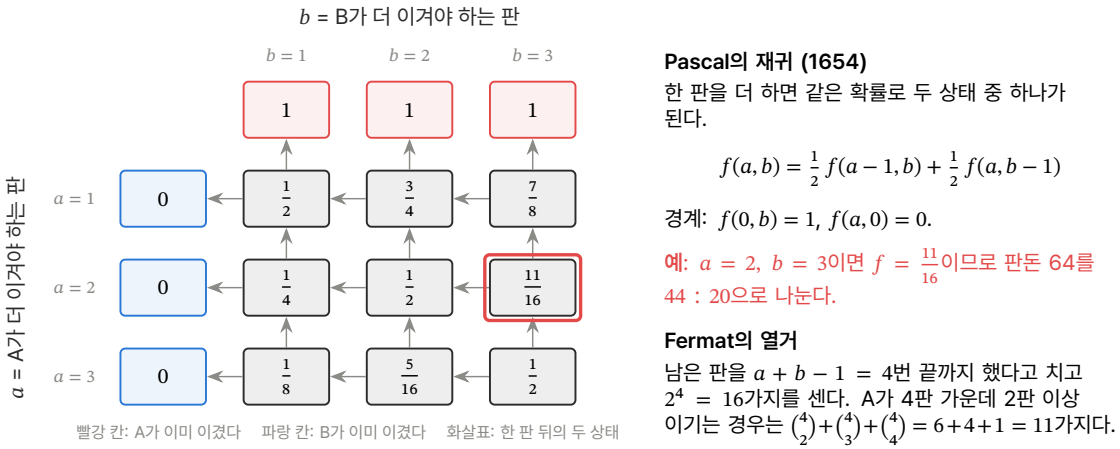



**그림 1.3** $p = \frac{1}{2}$일 때 상태 $(a, b)$에서 A가 이길 확률. 칸마다 바로 위 칸과 바로 왼쪽 칸의 평균이다 (Pascal의 재귀, 정리 5.13).

Pascal은 경우를 세지 않고 한 판씩 거꾸로 따라갔다. 다음 판에서 A가 이기면 상태는 $(a-1, b)$, 지면 $(a, b-1)$이 된다. 그러므로

$$f(a,b) = p\,f(a-1,b) + q\,f(a,b-1), \qquad f(0,b) = 1, \quad f(a,0) = 0.$$

그림 1.3의 칸들이 이 재귀로 채워진다. 이 식은 5.5절에서 **전확률 공식**의 한 예로 증명되고, Fermat의 공식과 같은 답을 준다는 것은 연습 1.6에서 보인다.

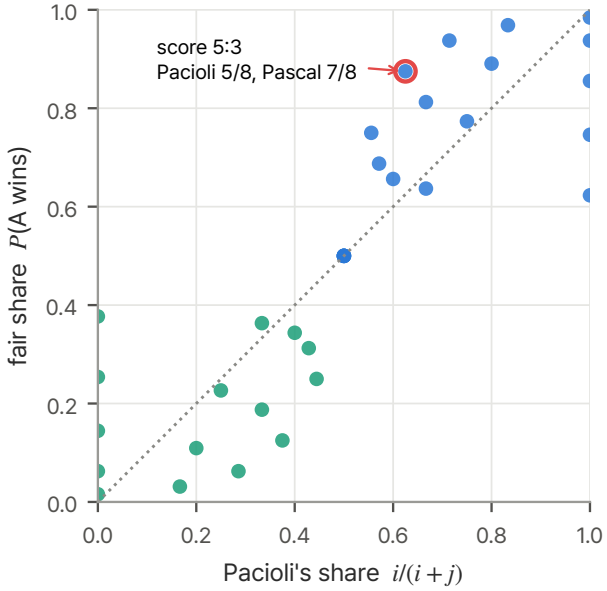



**그림 1.4** 먼저 6판을 이기는 내기의 모든 중간 점수 $(i, j)$에서 Pacioli의 몫 $i/(i+j)$와 공평한 몫. 점선 위의 점은 Pacioli가 앞선 쪽을 과소평가한 경우다.

## 1.5 Huygens (1657): 기대 몫

> **역사** 1655년 파리를 방문한 네덜란드의 Christiaan Huygens는 Pascal–Fermat의 해법을 전해 듣고, 1657년 *De ratiociniis in ludo aleae*(주사위 놀이의 계산에 대하여)를 냈다. 처음으로 인쇄된 확률 교과서다. 첫 명제는 이렇다. "$a$를 얻을 기회가 $p$번, $b$를 얻을 기회가 $q$번이고 모두 같은 정도로 일어날 수 있다면, 그 내기의 값어치는 $\frac{pa+qb}{p+q}$이다."

점수 분배의 공평한 몫은 Huygens의 언어로 "판돈 × 이길 확률", 곧 내기의 **기대 몫**이다. 이 생각을 확률변수로 일반화한 것이 기댓값이고, 2회의 첫 주제다.

### 연습문제

**연습 1.1** (포커의 풀하우스 ★) 52장에서 5장을 뽑는 방법은 $\binom{52}{5} = 2{,}598{,}960$가지다. 풀하우스(같은 숫자 3장 + 다른 같은 숫자 2장)의 확률을 구하라. 풀이 50쪽

**연습 1.2** (로또 ★) 1부터 45까지에서 6개를 고르는 로또에서 1등(6개 모두 일치)의 확률과, 정확히 5개가 맞을 확률(보너스 번호 무시)을 구하라. 풀이 50쪽

**연습 1.3** (Vandermonde 항등식의 특수형 ★★) 다음 항등식을 두 가지로 보여라.

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$$

(a) 남자 $n$명, 여자 $n$명에서 $n$명을 고르는 방법을 센다. (b) $(1+x)^n(1+x)^n = (1+x)^{2n}$에서 $x^n$의 계수를 비교한다. 풀이 51쪽

**연습 1.4** (Galileo의 11과 12 ★) 세 주사위의 합이 11인 경우와 12인 경우를 예제 1.8처럼 조합과 순서쌍으로 각각 세어라. 풀이 51쪽

**연습 1.5** (한 개짜리 내기의 기준 ★) 주사위 하나를 $n$번 던져 6이 적어도 한 번 나올 확률이 $\frac{1}{2}$을 넘는 최소의 $n$을 구하라. 비례 근사 $0.693/p$와 비교하라. 풀이 51쪽

**연습 1.6** (Fermat의 공식은 Pascal의 재귀를 만족한다 ★★) 정리 1.10의 $f(a,b)$가 $a,b \ge 1$에서 $f(a,b) = p\,f(a-1,b) + q\,f(a,b-1)$을 만족함을 보여라. (힌트: $\binom{m}{k} = \binom{m-1}{k-1} + \binom{m-1}{k}$를 쓰고, $f(a-1,b)$와 $f(a,b-1)$이 모두 $m-1 = a+b-2$판짜리 합임에 주목한다.) 풀이 51쪽

**연습 1.7** (실력이 다른 두 사람 ★) A가 판마다 0.6의 확률로 이긴다. 상태 $(a,b) = (2,3)$에서 A가 이길 확률을 구하라. $p = \frac{1}{2}$일 때의 $\frac{11}{16}$과 비교하라. 풀이 51쪽

**연습 1.8** (Pascal의 항등식 증명 ★ 유도형) 정리 1.5을 두 가지로 증명하라.

(a) (세기) $\{1, \dots, n\}$의 $k$개짜리 부분집합 가운데 원소 $n$을 포함하는 것의 수와 포함하지 않는 것의 수를 각각 정의 1.3로 세어라. 두 묶음이 겹치지 않고 합치면 전체가 됨을 확인하라.

(b) (계산) $1 \le k \le n-1$이면 정리 1.4의 식을 오른쪽에 넣고 통분해 왼쪽을 얻어라.

(c) $k = 0$과 $k = n$, 그리고 $k < 0$이나 $k > n$에서 식이 성립함을 정의 1.3의 약속으로 확인하라. (a)와 (b) 가운데 어느 증명이 이 경우들을 따로 다룰 필요가 없는가?

풀이 51쪽

**연습 1.9** (이항정리 증명 ★★ 유도형) 정리 1.6을 $n$에 대한 귀납법으로 증명하라.

(a) $n = 0$에서 양변이 1임을 확인하라.

(b) $n$에서 식이 성립한다고 하자. $(x+y)^{n+1} = x(x+y)^n + y(x+y)^n$에 귀납 가정을 넣고, 첫째 합의 첨자를 $j = k+1$로 바꿔 두 합을 모두 $x^j y^{n+1-j}$의 합으로 써라.

(c) $1 \le j \le n$인 항의 계수를 Pascal의 항등식으로 합치고, $j = 0$과 $j = n+1$인 항을 따로 확인해 $n+1$에서의 식을 얻어라.

(d) 귀납법 대신 세기로도 증명할 수 있다. $(x+y)^n$을 곱셈 그대로 전개한 $2^n$개의 항 가운데 $x^k y^{n-k}$인 항의 수를 정의 1.3로 세어 같은 결론을 얻어라.

풀이 52쪽

# 2

# 데이터에서 온 확률

도박장의 확률은 같은 정도로 일어나는 경우를 세서 얻었다. 그러나 사람이 몇 살까지 사는지에는 셀 수 있는 "같은 경우"가 없다. 17세기 후반 사람들은 다른 길을 찾았다. 실제로 일어난 일의 **상대도수**를 확률로 읽는 것이다. 이 절에서 조건부확률이 처음으로 자연스럽게 나타난다.

## 2.1 Graunt의 런던 사망표 (1662)

> **역사** 런던의 상인 John Graunt는 1662년 *Natural and Political Observations made upon the Bills of Mortality*를 냈다. 교구마다 매주 발표하던 세례·장례 기록을 수십 년 치 모아 표로 만들고, 그 안에서 규칙을 찾았다. 그는 해마다 남아가 여아보다 조금 더 많이 태어나 그 비가 대략 14 : 13으로 일정하다는 것을 알아챘다. 같은 기록을 1710년 Arbuthnot이 최초의 유의성 검정에 쓴다(8회).

Graunt의 방법은 오늘날의 말로 **상대도수 해석**이다. 사건 $A$를 같은 조건에서 여러 번 관찰할 수 있다면, $A$가 일어난 횟수의 비율이 시행을 거듭할수록 일정한 값에 가까워지고, 그 값을 $A$의 확률로 본다. 이 해석이 수학적으로 정당하다는 것, 곧 상대도수가 실제로 확률에 가까워진다는 것은 2회의 **대수의 법칙**이 증명한다.

## 2.2 Halley의 Breslau 생명표 (1693)

> **역사** 혜성으로 유명한 천문학자 Edmond Halley는 1693년 왕립학회지에 *An Estimate of the Degrees of the Mortality of Mankind*를 냈다. 런던은 인구가 드나들어 기록이 불완전했으므로, 그는 Caspar Neumann이 보내 온 독일(오늘날 폴란드) Breslau 시의 1687–1691년 출생·사망 기록을 썼다. 1세 아이 1,000명이 해마다 몇 명씩 남는지를 표로 만들고, 그 표로 종신연금의 값을 계산했다. 최초의 과학적 생명표다.

Halley의 표는 나이 $x$마다 생존자 수 $\ell(x)$를 준다($\ell(1) = 1000$). 1세 아이를 무작위로 하나 골랐을 때 그 아이가 $x$세까지 살 확률을 상대도수로 읽으면 $\ell(x)/1000$이다.

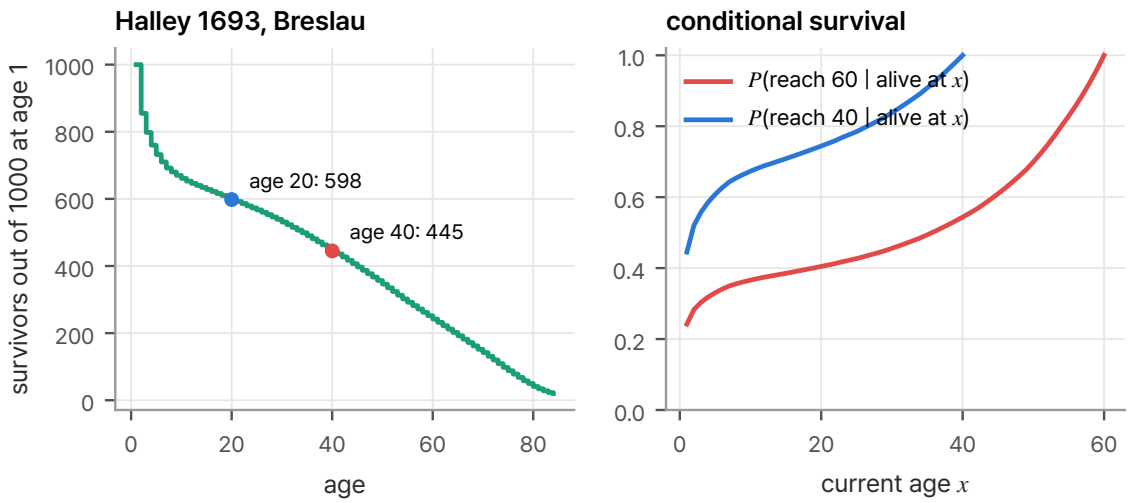



**그림 2.1** Halley(1693)의 생명표. 왼쪽: 1세 1,000명 가운데 각 나이까지의 생존자 수. 오른쪽: 지금 $x$세인 사람이 40세, 60세까지 살 비율.

**예제 2.1** (20세인 사람이 40세까지 살 비율) 표에서 $\ell(20) = 598$, $\ell(40) = 445$, $\ell(60) = 242$이다. 20세까지 산 598명 가운데 40세까지 산 사람은 445명이므로, 이미 20세인 사람이 40세까지 살 비율은

$$\frac{\ell(40)}{\ell(20)} = \frac{445}{598} \approx 0.744.$$

이것을 확률의 언어로 고쳐 보자. $A = \{\text{40세까지 산다}\}$, $B = \{\text{20세까지 산다}\}$라 하면 $A \subset B$이고

$$\frac{\ell(40)}{\ell(20)} = \frac{\ell(40)/1000}{\ell(20)/1000} = \frac{P(A \cap B)}{P(B)}.$$

"$B$가 일어났다는 것을 알 때 $A$의 확률"은 "$B$ 안에서 $A$가 차지하는 비율"이다. 이것이 5절의 조건부확률의 정의다.

같은 표에서 $\ell(60)/\ell(20) = 242/598 \approx 0.405$이고 $\ell(60)/\ell(40) = 242/445 \approx 0.544$이다. 그런데

$$\frac{\ell(60)}{\ell(20)} = \frac{\ell(40)}{\ell(20)} \cdot \frac{\ell(60)}{\ell(40)}, \qquad 0.405 \approx 0.744 \times 0.544.$$

"60세까지 산다"는 "40세까지 살고, 거기서 다시 60세까지 산다"이므로 조건부 비율이 곱해진다. 5절의 **곱셈법칙**이 생명표에서는 이렇게 보인다.

### 연습문제

**연습 2.1** (30대에 죽을 비율 ★) Halley의 표에서 $\ell(30) = 531$이다. 30세인 사람이 40세가 되기 전에 죽을 비율을 구하라. 풀이 52쪽

**연습 2.2** (생존의 중앙값 ★) 1세 아이 1,000명 가운데 절반 이하만 남는 첫 나이는 34세다 ($\ell(34) = 499$). 이 수가 무엇을 뜻하는지 확률의 말로 설명하라. 이것이 "평균 수명"과 다른 이유는? 풀이 52쪽

**연습 2.3** (생명표의 곱셈법칙 ★★) $x < y < z$에 대해 $\frac{\ell(z)}{\ell(x)} = \frac{\ell(y)}{\ell(x)} \cdot \frac{\ell(z)}{\ell(y)}$임을 보이고, 이를 사건 $A_x = \{x\text{세까지 산다}\}$와 $P(A_z \mid A_x) = P(A_y \mid A_x)P(A_z \mid A_y)$의 말로 바꿔 써라. $A_z \subset A_y \subset A_x$가 어디에 쓰였는가? 풀이 53쪽

# 3

# "무작위로" 의 위기

경우가 유한하고 모두 같은 정도로 일어나면 확률은 세면 된다. 경우가 무한히 많으면 어떻게 되는가? "원 위의 무작위 점" 이나 "무작위로 그은 현" 은 직관적으로는 분명해 보인다. 19세기 말 이 직관이 모순을 낳으면서 확률이 무엇인지를 다시 묻게 되었다.

## 3.1 Bertrand의 현 역설 (1889)

> **역사** 프랑스의 Joseph Bertrand는 1889년 *Calcul des probabilités*에서 다음 질문을 냈다. 원에 현을 무작위로 하나 그을 때, 그 현이 내접 정삼각형의 한 변보다 길 확률은 얼마인가? 그는 그럴듯한 세 가지 풀이가 서로 다른 답을 낸다는 것을 보이고, 무한히 많은 경우에 대한 "무작위" 라는 말이 정의되지 않았다고 결론지었다.

반지름을 $r$로 둔다. 내접 정삼각형의 한 변은 $\sqrt{3}\,r$이고, 원의 중심에서 그 변까지의 거리는 $r/2$이다. 현의 길이를 $L$이라 하자.

**예제 3.1** (세 가지 "무작위 현")

**(1) 두 끝점을 원 위에서 고른다.** 회전대칭이므로 한 끝점을 고정해도 된다. 고정한 끝점을 한 꼭짓점으로 하는 내접 정삼각형을 그리면, 다른 끝점이 맞은편 호(나머지 두 꼭짓점 사이, 원둘레의 $\frac{1}{3}$)에 떨어질 때 그리고 그때만 $L > \sqrt{3}\,r$이다. 다른 끝점이 원둘레 위에 고르게 퍼져 있다면 확률은 $\frac{1}{3}$이다.

**(2) 반지름 위의 한 점을 지나는 수직선.** 회전대칭이므로 반지름 하나를 고정한다. 그 위에서 중심과의 거리가 $d$인 점을 고르고, 반지름에 수직인 현을 긋는다. 길이는 $L = 2\sqrt{r^2 - d^2}$이므로 $L > \sqrt{3}\,r \iff d < r/2$다. $d$가 $[0, r]$에 고르게 퍼져 있다면 확률은 $\frac{1}{2}$이다.

**(3) 원 안의 한 점을 중점으로.** 원 안의 점을 하나 고르고 그 점을 중점으로 하는 현을 긋는다(중심이 아니면 유일하다). 중점과 중심의 거리를 $d$라 하면 역시 $L = 2\sqrt{r^2 - d^2}$이므로, 중점이 반지름 $r/2$인 동심원 안에 있어야 한다. 점이 원판의 넓이에 고르게 퍼져 있다면 확률은 넓이의 비 $(r/2)^2/r^2 = \frac{1}{4}$이다.

세 풀이 모두 계산에는 틀림이 없다. 다른 것은 "고르게 퍼져 있다" 의 대상이다. (1)은 원둘레의 **길이**, (2)는 반지름의 **길이**, (3)은 원판의 **넓이**에 대해 고르다. 같은 "무작위 현" 이 서로 다른 세 확률을 뜻한다. 질문이 **무엇에 대해 균등한가**를 정하지 않았으므로 답이 하나로 정해지지 않는다.

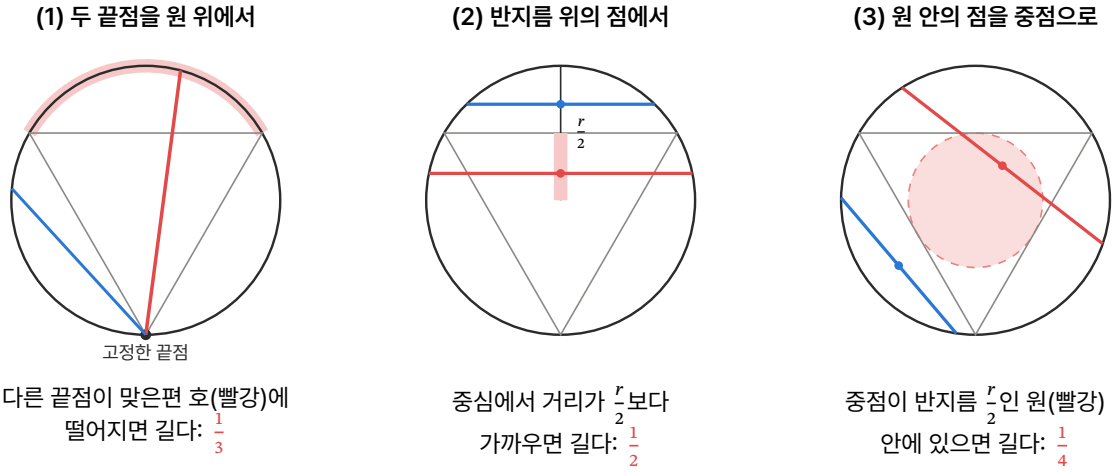



**그림 3.1** Bertrand(1889)의 세 가지 "무작위 현". 같은 질문에 $\frac{1}{3}$, $\frac{1}{2}$, $\frac{1}{4}$의 세 답이 나온다.

6.3절에서 세 방법이 만드는 현의 길이 분포를 구하고, 노트북에서 시뮬레이션으로 확인한다.

> **그러면 어느 것이 옳은가** 질문 자체에는 답이 없다. 다만 문제에 대칭을 더 요구하면 답이 정해질 수 있다. Jaynes(1973)는 원의 위치를 조금 옮기거나 크기를 바꿔도 답이 같아야 한다고 요구하면 (2)가 남는다고 논증했다. 이것도 "무엇이 같은 정도로 일어나는가"에 대한 **추가 가정**이다.

## 3.2 Bertrand의 상자 역설

같은 책에 Bertrand는 짧은 문제를 하나 더 실었다.

**예제 3.2** (세 상자) 상자가 셋 있다. 하나에는 금화 두 개, 하나에는 은화 두 개, 하나에는 금화와 은화가 하나씩 들어 있다. 상자를 하나 골라 동전을 하나 꺼냈더니 금화였다. 같은 상자의 다른 동전도 금화일 확률은?
흔한 추론은 이렇다. 금화가 나왔으니 상자는 금금이거나 금은이고, 둘 중 하나이므로 $\frac{1}{2}$이다. 그러나 답은 $\frac{2}{3}$이다. 금화가 나올 수 있는 경우는 **동전 단위로** 셋(금금 상자의 두 금화, 금은 상자의 금화 하나)이고, 그 가운데 둘이 금금 상자다. 5.3절에서 베이즈 정리로 다시 푼다. 같은 함정이 1975년 Selvin이 낸 Monty Hall 문제(연습 5.2)에 그대로 있다.

## 3.3 Hilbert의 6번 문제 (1900)

**역사** 1900년 파리 국제수학자대회에서 David Hilbert는 20세기의 과제 23개를 제시했다. 6번 문제는 "수학이 중요한 역할을 하는 물리학, 특히 확률론과 역학을 공리로 세우라"는 것이었다. 같은 시기 Borel과 Lebesgue는 길이·넓이·부피를 일반화한 **측도**의 이론을 세우고 있었다. 이 두 흐름이 1933년 Kolmogorov의 **확률론의 기초 개념**(Grundbegriffe der Wahrscheinlichkeitsrechnung)에서 만났다.

Bertrand의 역설이 알려 준 것은 두 가지다. 첫째, 확률을 말하려면 **가능한 결과의 집합**과 **확률을 줄 사건들의 모임**을 먼저 정해야 한다. 둘째, 확률이 따라야 할 **규칙**을 정하고, "균등" 처럼 구체적인 확률은 그 규칙을 만족하는 여러 선택 가운데 하나로 따로 가정해야 한다. 4절이 정확히 이 일을 한다.

### 연습문제

**연습 3.1** (반지름보다 긴 현 ★) Bertrand의 세 방법 각각에서 현이 반지름 $r$보다 길 확률을 구하라. 풀이 53쪽

**연습 3.2** (정사각형 안의 점과 $\pi$ ★) 점이 정사각형 $[0,1]^2$의 넓이에 고르게 퍼져 있다. 점이 원점에서 거리 1 이내에 있을 확률을 구하고, 이것으로 시뮬레이션에서 $\pi$를 추정하는 방법을 써라(7회의 Monte Carlo 예고). 풀이 53쪽

**연습 3.3** (구 위의 "무작위 점" ★★) 지구 표면의 "무작위 점"을 두 가지로 정한다. (a) 위도를 $[-90°, 90°]$, 경도를 $[0°, 360°)$에서 각각 고르게 고른다. (b) 점이 구면의 넓이에 고르게 퍼져 있다. 각각에서 점이 위도 $60°$ 이북에 떨어질 확률을 구하라. (힌트: 반지름 1인 구에서 위도 $\phi$ 이북의 넓이는 $2\pi(1-\sin\phi)$이다.) 풀이 53쪽

# 4

# Kolmogorov의 공리 (1933)

> **역사** Andrei Kolmogorov는 1933년 독일어로 낸 얇은 책 *Grundbegriffe der Wahrscheinlichkeitsrechnung*에서 확률을 **측도**의 한 종류로 정의했다. 확률이 "무엇인가"(같은 경우의 비율? 상대도수의 극한? 믿음의 정도?)는 묻지 않고, 확률이 "어떤 규칙을 따르는가"만 정했다. 1919년 Richard von Mises는 상대도수의 극한으로 확률을 정의하려 했지만 그 극한이 존재하는 수열을 정확히 정하는 데서 어려움을 겪었다. Kolmogorov의 공리는 해석과 무관하게 받아들여졌고, 오늘날 확률론의 표준이 되었다.

## 4.1 표본공간과 사건

**정의 4.1** (표본공간 · 사건 · σ-대수) **표본공간** $\Omega$는 실험에서 일어날 수 있는 모든 결과 $\omega$의 집합이다. $\Omega$의 부분집합들의 모임 $\mathcal{F}$가 다음 세 조건을 만족하면 $\mathcal{F}$를 **σ-대수**라 하고, $\mathcal{F}$의 원소를 **사건**이라 한다.

(i) $\Omega \in \mathcal{F}$.

(ii) $A \in \mathcal{F}$이면 $A^c \in \mathcal{F}$.

(iii) $A_1, A_2, \dots \in \mathcal{F}$이면 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

σ-대수는 사건의 모임이 확률을 말하는 데 필요한 연산에 닫혀 있도록 한다. De Morgan의 법칙(보조정리 0.1)으로 $\bigcap_i A_i = (\bigcup_i A_i^c)^c$이므로 가산 교집합에도 닫혀 있고, $A_3 = A_4 = \cdots = \varnothing$을 넣으면 유한 합집합에도 닫혀 있다(연습 4.2).

$\Omega$가 유한하거나 셀 수 있으면 보통 $\Omega$의 **모든** 부분집합을 사건으로 삼는다. $\Omega$가 원이나 구간처럼 셀 수 없을 때는 그럴 수 없는 경우가 있다(7.2절). 그래서 사건의 모임을 따로 정한다.

## 4.2 확률의 공리

**정의 4.2** (확률) $\sigma$-대수 $\mathcal{F}$ 위의 함수 $P : \mathcal{F} \to \mathbb{R}$가 다음을 만족하면 $P$를 **확률**이라 하고, $(\Omega, \mathcal{F}, P)$를 **확률공간**이라 한다.

(P1) 모든 사건 $A$에 대해 $P(A) \geq 0$.

(P2) $P(\Omega) = 1$.

(P3) (**가산가법성**) $A_1, A_2, \ldots$ 가 쌍마다 서로소인 사건이면

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

세 공리는 상대도수가 갖는 성질을 옮긴 것이다. 비율은 음수가 아니고, 전체의 비율은 1이며, 겹치지 않는 경우들의 비율은 더해진다. (P3)에서 **무한히 많은** 사건을 허용한 것이 중요하다. 이 덕분에 극한을 다룰 수 있고(정리 4.7), 이것을 유한한 경우로 줄이면 무엇이 무너지는지는 7.1절에서 본다.

## 4.3 공리에서 나오는 성질

**정리 4.3** (기본 성질) $(\Omega, \mathcal{F}, P)$가 확률공간이고 $A, B, A_1, A_2, \ldots$ 가 사건이면

(a) $P(\varnothing) = 0$.

(b) (유한가법성) $A_1, \ldots, A_n$이 쌍마다 서로소이면 $P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$.

(c) (여사건) $P(A^c) = 1 - P(A)$.

(d) (단조성) $A \subset B$이면 $P(B \setminus A) = P(B) - P(A)$이고, 따라서 $P(A) \leq P(B)$.

(e) $0 \leq P(A) \leq 1$.

(f) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(g) (Boole 부등식) $P\big(\bigcup_{i=1}^{\infty} A_i\big) \leq \sum_{i=1}^{\infty} P(A_i)$.

**증명.** **전략** 구하려는 사건을 서로소인 조각들의 합집합으로 쓰고 (P3)을 쓴다. 조각이 유한개이면 나머지 자리에 공집합을 채운다. (a)를 먼저 증명해야 이 채우기를 쓸 수 있다. 단계별 증명은 연습 4.12(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 55쪽에 있다.

**예제 4.4** (적어도 한 장의 에이스) 잘 섞은 52장에서 5장을 뽑는다(모든 5장 조합이 같은 확률이라는 **균등 모형**, 4.6절). 에이스가 적어도 한 장 나올 확률은 여사건(에이스가 한 장도 없다)으로 계산하는 편이 쉽다. 에이스가 아닌 48장에서 5장을 고르는 방법은 $\binom{48}{5} = 1{,}712{,}304$ 가지이므로, 정리 4.3(c)로

$$P(\text{에이스} \geq 1) = 1 - \frac{\binom{48}{5}}{\binom{52}{5}} = 1 - \frac{1{,}712{,}304}{2{,}598{,}960} \approx 0.341.$$

## 4.4 포함배제의 원리

정리 4.3(f)를 사건이 여럿인 경우로 넓힌다. 사건 $A_1, \dots, A_n$에 대해

$$S_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \qquad (k = 1, \dots, n)$$

로 둔다. $S_1 = \sum_i P(A_i)$, $S_2 = \sum_{i<j} P(A_i A_j)$이고, $S_k$는 $\binom{n}{k}$개 항의 합이다.

**정리 4.5** (포함배제) 사건 $A_1, \dots, A_n$에 대해

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} S_k = S_1 - S_2 + S_3 - \cdots + (-1)^{n+1} S_n.$$

**증명 1 (귀납법).** **전략** $n$에 대한 귀납법 [T3]. 사건을 하나 더하는 단계에서 두 사건의 공식(정리 4.3(f))을 쓰고, 그때 새로 생기는 교집합들에 귀납 가정을 한 번 더 쓴다.

① $n = 1$이면 양변이 $P(A_1)$이고, $n = 2$이면 정리 4.3(f)이다. — 기저

② $n$개의 사건에서 식이 성립한다고 하고 $B = A_1 \cup \cdots \cup A_n$으로 둔다.

$$P(B \cup A_{n+1}) = P(B) + P(A_{n+1}) - P(B \cap A_{n+1}).$$ — 정리 4.3(f)

③ $B \cap A_{n+1} = \bigcup_{i=1}^{n} (A_i A_{n+1})$이다. — 분배법칙 [T1]

④ ③의 $n$개 사건 $A_1 A_{n+1}, \dots, A_n A_{n+1}$에 귀납 가정을 쓴다.

$$P(B \cap A_{n+1}) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \cdots < i_k \leq n} P(A_{i_1} \cdots A_{i_k} A_{n+1}).$$ — 귀납 가정

⑤ $P(B)$에도 귀납 가정을 쓴다. 그러면 ②의 오른쪽은 $A_1, \dots, A_{n+1}$ 가운데 몇 개를 고른 교집합의 확률들에 부호를 붙여 더한 것이다. — 귀납 가정

⑥ $j$개짜리 교집합 하나가 어디서 어떤 부호로 나오는지 센다.

- $A_{n+1}$을 포함하지 않으면 $P(B)$의 전개에서만 부호 $(-1)^{j+1}$로 한 번 나온다.

- $A_{n+1}$ 하나뿐이면($j=1$) $+P(A_{n+1})$로 한 번 나오고, 부호는 $(-1)^{1+1}$이다.
- $A_{n+1}$과 다른 $k=j-1\geq 1$개의 교집합이면 ④를 거쳐 $-P(B\cap A_{n+1})$에서 부호 $-(-1)^{k+1}=(-1)^{j+1}$로 한 번 나온다.

⑦ 어느 경우든 $j$개짜리 교집합은 부호 $(-1)^{j+1}$로 한 번 나오므로 $n+1$개에 대한 식이 성립한다. ⑥, 귀납 [T3]

■

**증명 2 (세기, $\Omega$가 유한하고 균등할 때).** 양변에 $|\Omega|$를 곱하면 왼쪽은 합집합의 원소 수, 오른쪽은 $\sum_k (-1)^{k+1}\sum |A_{i_1}\cdots A_{i_k}|$이다. 정확히 $m\geq 1$개의 사건에 속한 점 $\omega$는 $k$개짜리 교집합 가운데 $\binom{m}{k}$개에 속하므로 오른쪽에서

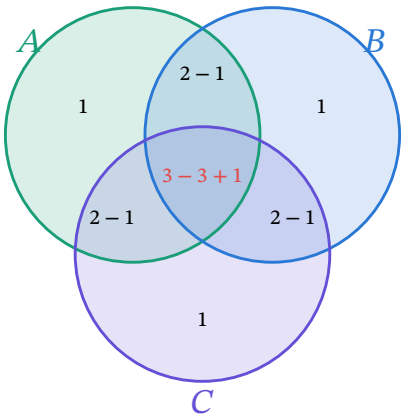
$$\sum_{k=1}^{m}(-1)^{k+1}\binom{m}{k}=1-\sum_{k=0}^{m}(-1)^{k}\binom{m}{k}=1$$

번 세어진다(따름정리 1.7). 어느 사건에도 속하지 않는 점은 0번 세어진다. 따라서 양변이 같다. 이 증명은 2회의 지시함수와 기댓값으로 일반적인 확률공간에도 그대로 옮겨진다. ■

**예제 4.6** (Montmort의 짝 맞추기 (1708)) **문제.** 1부터 $n$까지 적힌 카드 $n$장을 잘 섞어(모든 $n!$가지 순서가 같은 확률) 한 장씩 뒤집으며 $1,2,\dots,n$을 센다. 센 수와 카드의 수가 적어도 한 번 일치할 확률 $p_n$은?

$A_i=\{i$번째 카드가 $i\}$로 두면 구하는 것은 $P(A_1\cup\cdots\cup A_n)$이다. $i_1<\cdots<i_k$를 고정하면 그 $k$자리는 정해지고 나머지 $n-k$장은 아무렇게나 놓이므로

$$P(A_{i_1}\cdots A_{i_k})=\frac{(n-k)!}{n!},\qquad S_k=\binom{n}{k}\frac{(n-k)!}{n!}=\frac{1}{k!}.$$




**포함배제 ($n=3$)**

$$P(A\cup B\cup C)=\underbrace{\sum P(A)}_{S_1}-\underbrace{\sum P(AB)}_{S_2}+\underbrace{P(ABC)}_{S_3}$$

영역 안의 숫자는 그 영역이 $S_1$, $S_2$, $S_3$에서 각각 몇 번 세어지는지를 뜻한다. 세 사건에 모두 속한 점은 $3-3+1=1$번, 두 사건에 속한 점은 $2-1=1$번 — 어느 점이든 정확히 한 번 세어진다.

일반적으로 $m$개의 사건에 속한 점은 $\binom{m}{1}-\binom{m}{2}+\cdots\pm\binom{m}{m}=1$번 세어진다 (정리 4.5의 증명 2).

**그림 4.1** 포함배제($n=3$). 각 영역은 $S_1$, $S_2$, $S_3$을 더하고 빼는 과정에서 정확히 한 번 세어진다.

포함배제로

$$p_n = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{n+1}}{n!} = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!}.$$

$e^{-1} = \sum_{k\ge 0}(-1)^k/k!$이므로 $1 - p_n = \sum_{k=0}^{n}(-1)^k/k! \to e^{-1}$, 곧 $p_n \to 1 - 1/e \approx 0.6321$이다. 교대급수의 꼬리는 첫 항보다 작으므로 $|p_n - (1-1/e)| \le 1/(n+1)!$이다. 카드가 13장이면 $p_{13} = 0.632120558\ldots$로 극한과 $10^{-11}$ 정도만 다르다. 놀랍게도 답은 $n$에 거의 의존하지 않는다.

**역사** Pierre Rémond de Montmort는 1708년 *Essay d'analyse sur les jeux de hazard*에서 13장으로 하는 카드놀이 "Treize"의 이 문제를 풀었다. 1713년 2판에는 Nicolaus Bernoulli와 주고받은 일반해가 실렸다.

## 4.5 확률의 연속성

사건의 열 $A_1, A_2, \ldots$가 $A_1 \subset A_2 \subset \cdots$이면 **증가한다**고 하고 $A_n \uparrow A = \bigcup_n A_n$으로 쓴다. $A_1 \supset A_2 \supset \cdots$이면 **감소한다**고 하고 $A_n \downarrow A = \bigcap_n A_n$으로 쓴다.

**정리 4.7** (확률의 연속성) (a) $A_n \uparrow A$이면 $P(A_n) \uparrow P(A)$. (b) $A_n \downarrow A$이면 $P(A_n) \downarrow P(A)$.

**증명.** **전략** 증가하는 열을 서로소인 조각으로 쪼갠다. 그러면 $P(A_n)$은 급수의 부분합이 되고, $P(A)$는 그 급수의 합이 된다. 감소하는 열은 여사건을 취해 증가하는 열로 바꾼다.

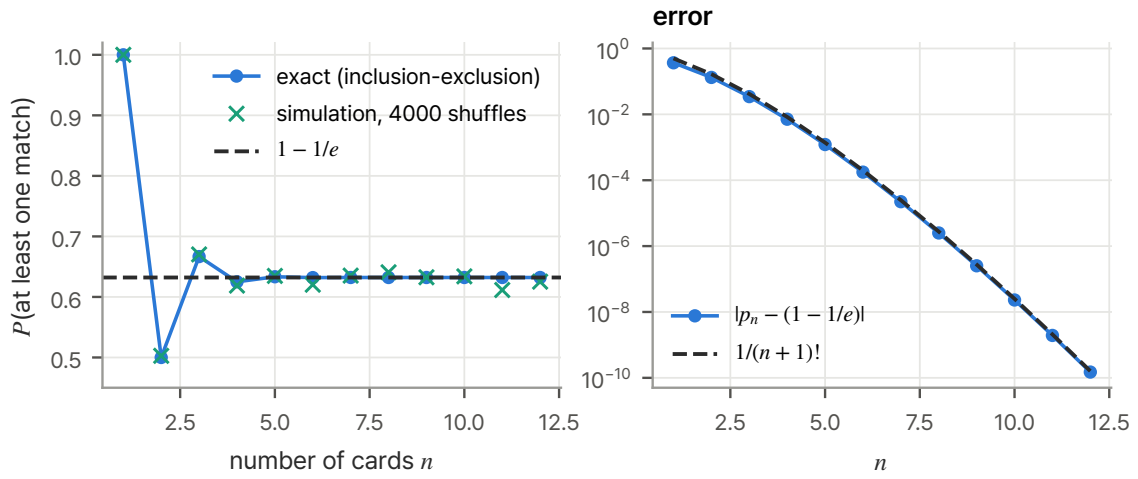



**그림 4.2** Montmort의 짝 맞추기. 왼쪽: 포함배제로 구한 정확한 값과 4,000번 섞은 시뮬레이션. 오른쪽: 극한 $1 - 1/e$와의 차이는 $1/(n+1)!$보다 작다.

**(a)**

① $B_1 = A_1$, $B_n = A_n \setminus A_{n-1}$ $(n \geq 2)$로 둔다. $B_n$은 "$n$번째에 새로 들어온 부분"이다. 정의

② $B_n$들은 쌍마다 서로소다. $m < n$이면 $B_m \subset A_m \subset A_{n-1}$이고 $B_n \cap A_{n-1} = \varnothing$이기 때문이다. 열이 증가한다 [T1]

③ $A_n = B_1 \cup \cdots \cup B_n$이고 $A = \bigcup_{k=1}^{\infty} B_k$이다. $\omega \in A$이면 $\omega \in A_n$인 가장 작은 $n$이 있고, 그때 $\omega \in B_n$이다. 원소 추적 [T1]

④ ①–③을 공리에 넣는다.

$$\begin{aligned} P(A) &= \sum_{k=1}^{\infty} P(B_k) && \text{②③, (P3)} \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) && \text{급수의 합의 정의 [T5]} \\ &= \lim_{n\to\infty} P(A_n) && \text{③, 유한가법성} \end{aligned}$$

⑤ $A_n \subset A_{n+1}$이므로 $P(A_n) \leq P(A_{n+1})$이다. 극한은 증가하며 다가간다. 단조성, 정리 4.3(d)

**(b)**

① $A_n \downarrow A$이면 $A_n^c \uparrow A^c$이다. $A_n^c \subset A_{n+1}^c$이고 $\bigcup_n A_n^c = (\bigcap_n A_n)^c$이기 때문이다. De Morgan, 보조정리 0.1

② (a)를 $A_n^c$에 쓰면 $P(A_n^c) \to P(A^c)$이다. (a)

③ 여사건으로 바꾸면 $1 - P(A_n) \to 1 - P(A)$, 곧 $P(A_n) \to P(A)$이다. 정리 4.3(c)

■

연속성은 가산가법성의 다른 얼굴이다. 유한가법성에 연속성을 더하면 거꾸로 가산가법성이 나온다(연습 4.11). 6.2절에서 분포함수의 성질을 모두 이 정리로 증명한다.

**예제 4.8** (언젠가는 앞면이 나온다) 공정한 동전을 끝없이 던진다. $T_n$을 처음 $n$번이 모두 뒷면인 사건이라 하면 $P(T_n) = 2^{-n}$이고(5.4절의 독립), $T_1 \supset T_2 \supset \cdots$이며 $\bigcap_n T_n$은 한 번도 앞면이 나오지 않는 사건이다. 정리 4.7(b)로

$$P(\text{한 번도 앞면이 나오지 않는다}) = \lim_{n\to\infty} 2^{-n} = 0.$$

확률이 0인 사건이 반드시 공집합인 것은 아니다. 뒷면만 끝없이 나오는 결과 $\omega = (\text{뒤}, \text{뒤}, \dots)$은 $\Omega$의 원소다.

**증명 없이 쓰는 사실** (무한히 던지는 동전의 확률공간) 처음 몇 번의 결과에 대해서는 독립인 공정한 동전처럼 행동하는 확률 $P$가 무한 수열의 공간 $\Omega = \{\text{앞}, \text{뒤}\}^{\mathbb{N}}$ 위에 (적절한 $\sigma$-대수와 함께) 존재한다. 이것은 Kolmogorov(1933)의 확장 정리의 한 경우이고, 측도론에서 증명한다. 직관적으로는 $[0,1]$에서 고르게 고른 수를 이진법으로 쓴 자릿수가 바로 그런 동전이다.

## 4.6 이산 확률공간과 고전적 정의

**정리 4.9** (이산 확률공간) $\Omega = \{\omega_1, \omega_2, \dots\}$가 유한하거나 셀 수 있고 $p_i \ge 0$, $\sum_i p_i = 1$이라 하자. $\Omega$의 모든 부분집합 $A$에 대해

$$P(A) = \sum_{i:\ \omega_i \in A} p_i$$

로 정하면 $P$는 ($\mathcal{F}$ = 모든 부분집합 위의) 확률이다.

**증명.** 모든 부분집합의 모임은 $\sigma$-대수다. (P1)은 $p_i \ge 0$에서, (P2)는 $\sum p_i = 1$에서 나온다. (P3): $A_1, A_2, \dots$ 가 쌍마다 서로소이면 $\bigcup_j A_j$에 속한 $\omega_i$는 정확히 하나의 $A_j$에 속하므로, 합 $\sum_{\omega_i \in \bigcup A_j} p_i$를 $A_j$ 별로 묶으면 $\sum_j \sum_{\omega_i \in A_j} p_i = \sum_j P(A_j)$이다. 음이 아닌 항의 급수는 묶는 방법과 순서에 상관없이 같은 값을 갖는다(아래 상자). ∎

**증명 없이 쓰는 사실** (음이 아닌 급수의 재배열) 음이 아닌 항들의 급수는 항의 순서를 바꾸거나 항들을 (유한하거나 무한한) 묶음으로 나눠 묶음별로 먼저 더해도 같은 값(무한대일 수도 있다)을 갖는다. 해석학 교재의 표준 결과이고, 2회 이후 기댓값의 계산에서도 계속 쓴다.

**따름정리 4.10** (고전적 정의) $\Omega$가 $N$개의 원소를 갖고 $p_i = 1/N$이면 $P(A) = \dfrac{|A|}{|\Omega|}$.

**증명.** 정리 4.9에서 $P(A) = \sum_{\omega_i \in A} 1/N = |A|/N$. ∎

Cardano부터 Laplace까지의 "유리한 경우 ÷ 전체 경우"는 이렇게 공리의 한 특수한 경우가 된다. 그것을 쓸 수 있는가, 곧 결과들이 정말 같은 정도로 일어나는가는 공리가 아니라 **모형의 가정**이다 (예제 1.8).

**예제 4.11** (생일 문제) $n$명의 생일이 365일 가운데 고르고 서로 무관하다고 하자(윤일과 계절 차이는 무시한다). 모형은 $\Omega = \{1, \dots, 365\}^n$ 위의 균등 확률이다. 생일이 모두 다른 결과의 수는 정리 1.2의 순열 수 $365 \cdot 364 \cdots (365 - n + 1)$이므로

$$P(\text{같은 생일인 쌍이 있다}) = 1 - \prod_{i=0}^{n-1}\left(1 - \frac{i}{365}\right).$$

$n = 23$이면 0.507로 처음 $\frac{1}{2}$을 넘는다. $1 - x \le e^{-x}$이므로 곱은 $\exp(-\sum_{i<n} i/365) = \exp(-\frac{n(n-1)}{730})$보다 작고, 따라서

$$P(\text{같은 생일인 쌍이 있다}) \ge 1 - e^{-n(n-1)/730}.$$

쌍의 수 $\binom{n}{2}$이 $n^2$으로 자라기 때문에 생각보다 빨리 1에 가까워진다.

**Bertrand의 역설, 다시** $\Omega$가 원이나 원판이면 결과가 셀 수 없이 많아 정리 4.9를 쓸 수 없다. "고르게 퍼져 있다"는 "길이(또는 넓이)에 비례하는 확률"이라는 뜻이고, 이런 확률이 존재한다는 것은 측도론의 결과다(Lebesgue 측도). Bertrand의 세 방법은 서로 다른 $\Omega$와 서로 다른 $P$를 고른 것이다. 공리는 셋을 모두 허락하고, 어느 것을 쓸지는 모형을 세우는 사람이 정한다.

### 연습문제

**연습 4.1** (De Morgan 연습 ★) $(A \cup B \cup C)^c = A^c B^c C^c$을 보이고, $A \cup B = A \cup (B \setminus A)$가 서로소인 합집합임을 보여라. 풀이 53쪽

**연습 4.2** (σ-대수의 성질 ★) σ-대수 $\mathcal{F}$가 $\varnothing$, 유한 합집합, 가산 교집합, 차집합 $A \setminus B$에 닫혀 있음을 보여라. 풀이 54쪽

**연습 4.3** (두 사건의 계산 ★) $P(A) = 0.6$, $P(B) = 0.5$, $P(A \cup B) = 0.8$이다. $P(AB)$, $P(A^c B^c)$, 그리고 정확히 하나만 일어날 확률을 구하라. 풀이 54쪽

**연습 4.4** (하트가 적어도 한 장 ★) 52장에서 5장을 뽑을 때 하트가 적어도 한 장 나올 확률을 구하라. 풀이 54쪽

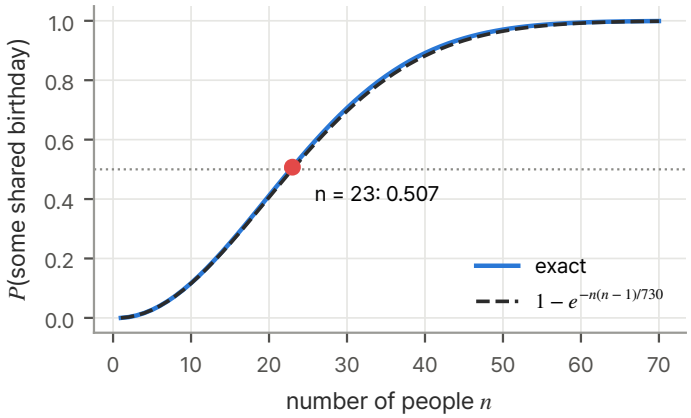



**그림 4.3** 생일 문제의 정확한 확률과 하한 $1 - e^{-n(n-1)/730}$. 23명이면 $\frac{1}{2}$을 넘는다.

**연습 4.5** (Bonferroni 부등식 ★★) $P\left(\bigcup_{i=1}^{n} A_i\right) \geq S_1 - S_2$ 임을 $n$에 대한 귀납법으로 보여라. (포함배제를 $S_1$에서 자르면 위로, $S_2$에서 자르면 아래로 어긋난다.) 풀이 54쪽

**연습 4.6** (네 장의 짝 맞추기 ★) 카드 4장일 때 하나도 맞지 않는 순서를 모두 적어 그 수가 $4!\left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}\right) = 9$임을 확인하라. 풀이 54쪽

**연습 4.7** (정확히 $k$개가 맞을 확률 ★★) Montmort의 문제에서 정확히 $k$장이 맞을 확률이 다음과 같음을 보이고, $n \to \infty$일 때 $e^{-1}/k!$로 감을 보여라. (이 극한은 4회의 Poisson 분포다.)

$$P(\text{정확히 } k\text{장}) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}$$

풀이 54쪽

**연습 4.8** (99%가 되려면 ★) 생일 문제에서 같은 생일인 쌍이 있을 확률이 처음으로 0.99를 넘는 $n$을 구하라. 풀이 54쪽

**연습 4.9** (확률 1인 사건들의 교집합 ★★) 모든 $n$에 대해 $P(A_n) = 1$이면 $P\left(\bigcap_n A_n\right) = 1$임을 보여라. 풀이 55쪽

**연습 4.10** (합집합의 확률은 극한 ★★) 증가하지 않는 임의의 사건열에 대해서도 다음이 성립함을 보여라.

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_1 \cup \cdots \cup A_n)$$

풀이 55쪽

**연습 4.11** (연속성 ⇒ 가산가법성 ★★★) $P$가 (P1), (P2), 유한가법성을 만족하고, $A_n \downarrow \varnothing$이면 $P(A_n) \to 0$이라 하자. $P$가 가산가법성 (P3)을 만족함을 보여라. 풀이 55쪽

**연습 4.12** (공리에서 기본 성질로 ★★ 유도형) 정리 4.3을 다음 순서로 증명하라. 각 단계에서 쓴 공리나 앞 단계를 밝혀라.

(a) $A_1 = \Omega, A_2 = A_3 = \cdots = \varnothing$에 (P3)을 써서 $\sum_{i\geq 2} P(\varnothing) = 0$을 얻고, $P(\varnothing) = 0$을 보여라. $P(\varnothing) > 0$이면 무엇이 틀리는가?

(b) 유한개 $A_1, \ldots, A_n$ 뒤에 공집합을 채워 (P3)과 (a)에서 유한가법성을 얻어라.

(c) $A$와 $A^c$의 관계에서 여사건의 공식을 얻어라.

(d) $A \subset B$이면 $B = A \cup (B \setminus A)$가 서로소인 합집합임을 보이고 단조성을 얻어라.

(e) (d)와 (P2)에서 $P(A) \leq 1$을 얻어라.

(f) $A \cup B = A \cup (B \setminus AB)$가 서로소인 합집합임을 보이고 (d)를 $AB \subset B$에 써서 (f)를 얻어라.

(g) $B_1 = A_1$, $B_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1})$로 두고, $B_n$들이 쌍마다 서로소이며 $\bigcup_n B_n = \bigcup_n A_n$, $B_n \subset A_n$임을 보여 Boole 부등식을 얻어라.

풀이 55쪽

# 5

# 조건부확률과 독립

2절의 생명표에서 "이미 20세인 사람이 40세까지 살 비율"은 20세 생존자 가운데 40세 생존자가 차지하는 비율이었다. 정보가 주어지면 확률을 다시 매긴다. 이것을 공리의 언어로 정의하고, 이 과정 전체에서 가장 많이 쓰는 세 도구(곱셈법칙, 전확률 공식, 베이즈 정리)를 증명한다.

## 5.1 조건부확률

**정의 5.1** (조건부확률) $P(B) > 0$인 사건 $B$에 대해, 사건 $A$의 **$B$가 주어졌을 때의 조건부확률**은

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

$B$가 일어났다는 것을 알면 가능한 결과는 $B$ 안으로 줄어든다. 그 안에서 $A$가 차지하는 몫을 $B$의 확률로 나눠 다시 잰 것이 $P(A \mid B)$다. 정의가 공리와 어긋나지 않는다는 것, 곧 조건부확률도 확률이라는 것이 먼저 확인할 일이다.

**정리 5.2** (조건부확률은 확률이다) $P(B) > 0$이면 $Q(A) = P(A \mid B)$는 $(\Omega, \mathcal{F})$ 위의 확률이다. 따라서 4절의 모든 성질(여사건, 포함배제, 연속성 등)이 $P(\cdot \mid B)$에 대해서도 성립한다.

**증명.** **전략** $Q$가 공리 (P1)–(P3)을 하나씩 만족함을 확인한다. (P3)에서는 $B$와의 교집합이 서로소성과 합집합을 보존한다는 것을 쓴다 [T1]. 단계별 증명은 연습 5.10(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 57쪽에 있다.

## 5.2 곱셈법칙 · 전확률 공식 · 베이즈 정리

**정리 5.3** (곱셈법칙) $P(A_1 \cap \cdots \cap A_{n-1}) > 0$이면

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1 A_2) \cdots P(A_n \mid A_1 \cdots A_{n-1}).$$

**증명.** **전략** 먼저 오른쪽의 조건부확률이 모두 정의됨을 단조성으로 확인한다. 그다음 $n$에 대한 귀납법으로, 조건부확률의 정의를 한 번 쓸 때마다 사건이 하나씩 늘어남을 보인다 [T3]. 단계별 증명은 연습 5.11(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 57쪽에 있다.

**정리 5.4** (전확률 공식) $B_1, B_2, \dots$ 가 $\Omega$의 분할이고 모든 $i$에서 $P(B_i) > 0$이면, 모든 사건 $A$에 대해

$$P(A) = \sum_i P(A \mid B_i)\, P(B_i).$$

**증명.** **전략** $A$를 분할의 조각들 위로 나눠 서로소인 합집합으로 쓰고 가산가법성을 쓴다.

① $A = A \cap \Omega = \bigcup_i (A \cap B_i)$이다. — $\bigcup_i B_i = \Omega$, 분배법칙 [T1]

② $A \cap B_i$들은 쌍마다 서로소다. $A \cap B_i \subset B_i$이고 $B_i$들이 쌍마다 서로소이기 때문이다. — 분할의 정의

③ $P(A) = \sum_i P(A \cap B_i)$이다. 분할이 유한하면 유한가법성을 쓴다. — ①②, (P3)

④ $P(A \cap B_i) = P(A \mid B_i)\, P(B_i)$를 ③에 넣는다. — 정의 5.1, $P(B_i) > 0$

■

**정리 5.5** (베이즈 정리) 정리 5.4의 가정에서 $P(A) > 0$이면

$$P(B_j \mid A) = \frac{P(A \mid B_j)\, P(B_j)}{\sum_i P(A \mid B_i)\, P(B_i)}.$$

**증명.** **전략** 조건부확률의 정의에서 출발해 분자와 분모를 각각 "원인이 결과를 낼 확률"로 바꾼다.

① $P(B_j \mid A) = P(AB_j)/P(A)$이다. — 정의 5.1, $P(A) > 0$

② 분자는 $P(AB_j) = P(A \mid B_j)\, P(B_j)$이다. — 정의 5.1, $P(B_j) > 0$

③ 분모는 $P(A) = \sum_i P(A \mid B_i)\, P(B_i)$이다. — 정리 5.4

■

> **역사** Thomas Bayes는 이 정리를 내놓지 않은 채 1761년에 죽었다. 친구 Richard Price가 원고를 정리해 1763년 왕립학회지에 실었다(*An Essay towards solving a Problem in the Doctrine of Chances*). Bayes가 실제로 다룬 것은 이항분포의 성공률에 대한 추론이었고(4회, 16회), 위와 같은 일반적인 형태로 쓰고 널리 쓴 사람은 Laplace(1774)였다. 베이즈 정리는 "원인의 확률"을 결과로부터 거꾸로 계산한다고 해서 오랫동안 "역확률"이라 불렸다.

베이즈 정리의 각 항에는 이름이 있다. $P(B_j)$는 정보를 얻기 전의 **사전확률**, $P(A \mid B_j)$는 원인 $B_j$가 결과 $A$를 낼 **가능도**, $P(B_j \mid A)$는 정보를 얻은 뒤의 **사후확률**이다. 두 원인 $B$와 $B^c$만 있을 때는 **오즈**$(O(B) = P(B)/P(B^c))$로 쓰면 더 간단하다.

$$\underbrace{\frac{P(B \mid A)}{P(B^c \mid A)}}_{\text{사후 오즈}} = \underbrace{\frac{P(A \mid B)}{P(A \mid B^c)}}_{\text{가능도비}} \times \underbrace{\frac{P(B)}{P(B^c)}}_{\text{사전 오즈}}.$$

정보가 하나 들어올 때마다 오즈에 가능도비가 곱해진다.

## 5.3 베이즈 정리로 푸는 두 문제

**예제 5.6** (진단검사와 세 상자)

**(1) 진단검사.** 어떤 병의 유병률이 1%이다. 검사는 병이 있으면 90%의 확률로 양성(민감도)이고, 병이 없어도 9%의 확률로 양성(위양성률)이다. 한 사람이 양성이 나왔다. 실제로 병이 있을 확률은?

$D=\{\text{병이 있다}\}$, $+=\{\text{양성}\}$으로 두면 베이즈 정리로

$$P(D\mid +)=\frac{P(+\mid D)P(D)}{P(+\mid D)P(D)+P(+\mid D^c)P(D^c)}=\frac{0.9\times 0.01}{0.9\times 0.01+0.09\times 0.99}=\frac{0.009}{0.0981}\approx 0.092.$$

양성인 사람 가운데 실제 환자는 열에 하나도 안 된다. 1,000명 가운데 환자 10명에서 양성 9명, 건강한 990명에서 양성 약 89명이 나오기 때문이다. 이것을 **기저율 오류**라 한다. 같은 검사를 독립적으로 한 번 더 해서 또 양성이면, 사전확률을 0.092로 바꿔 다시 계산해 0.503이 된다.

**(2) Bertrand의 세 상자.** 3.2절의 상자를 $G$(금금), $S$(은은), $M$(금은)이라 하자. 상자는 같은 확률 $\frac{1}{3}$로 고르고, 상자 안에서 동전도 같은 확률로 꺼낸다. 금화가 나왔다는 사건을 $g$라 하면 $P(g\mid G)=1$, $P(g\mid S)=0$, $P(g\mid M)=\frac{1}{2}$이므로

$$P(G\mid g)=\frac{1\cdot\frac{1}{3}}{1\cdot\frac{1}{3}+0\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{3}}=\frac{2}{3}.$$

"둘 중 하나이므로 $\frac{1}{2}$"이라는 추론은 두 상자가 금화를 낼 가능도(1과 $\frac{1}{2}$)가 다르다는 것을 무시했다.

## 5.4 독립

**정의 5.7** (두 사건의 독립) $P(A\cap B)=P(A)\,P(B)$이면 사건 $A$와 $B$는 **독립**이라 한다.

$P(B)>0$이면 독립은 $P(A\mid B)=P(A)$, 곧 $B$를 알아도 $A$의 확률이 바뀌지 않는다는 뜻이다. 정의를 곱의 꼴로 쓰는 이유는 $P(B)=0$인 경우까지 포함하고 $A$와 $B$에 대칭이기 때문이다.

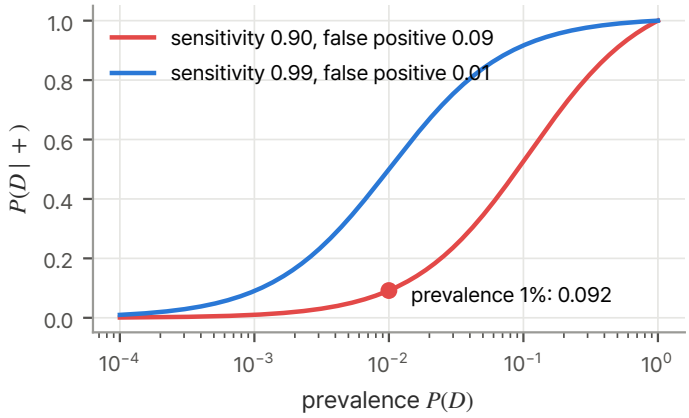



**그림 5.1** 양성일 때 실제로 병이 있을 확률은 유병률에 크게 의존한다. 같은 검사라도 드문 병에서는 양성의 대부분이 위양성이다.

**정리 5.8** (여사건과의 독립) $A$와 $B$가 독립이면 $A$와 $B^c$, $A^c$와 $B$, $A^c$와 $B^c$도 각각 독립이다.

**증명.** $AB^c = A \setminus AB$이므로 $P(AB^c) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)\big(1 - P(B)\big) = P(A)P(B^c)$. 나머지 둘은 이 결과를 $(B, A)$와 $(A, B^c)$에 다시 쓰면 나온다. ∎

**정의 5.9** (여러 사건의 상호 독립) 사건 $A_1, \dots, A_n$이 **상호 독립**이라는 것은, 서로 다른 첨자 $i_1 < \cdots < i_k (k \geq 2)$를 어떻게 골라도

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

가 성립한다는 것이다. 무한히 많은 사건이 상호 독립이라는 것은 그 가운데 유한개가 모두 상호 독립이라는 것이다.

모든 **쌍**에서 독립인 것만으로는 부족하다.

**예제 5.10** (Bernstein의 정사면체) 정사면체의 네 면 가운데 세 면을 각각 빨강, 초록, 파랑으로 칠하고 네 번째 면에는 세 색을 모두 칠한다. 정사면체를 던져 바닥에 닿은 면을 본다(네 면이 같은 확률). $R$, $G$, $B$를 그 면에 각 색이 있는 사건이라 하자. 각 색은 두 면에 있으므로 $P(R) = P(G) = P(B) = \frac{1}{2}$이고, 두 색이 함께 있는 면은 하나뿐이므로

$$P(RG) = P(RB) = P(GB) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}.$$

세 쌍 모두 독립이다. 그러나 $P(RGB) = \frac{1}{4} \neq \frac{1}{8} = P(R)P(G)P(B)$이다. 두 색을 알면 세 번째 색이 있는지가 정해진다. **쌍별 독립이 상호 독립을 주지는 않는다.**

**정리 5.11** (독립인 사건들과 여사건) $A_1, \dots, A_n$이 상호 독립이면, 그 가운데 몇 개를 여사건으로 바꾼 것도 상호 독립이다. 특히

$$P(A_1^c \cap \cdots \cap A_n^c) = \prod_{i=1}^{n}(1 - P(A_i)), \qquad P(\text{적어도 하나}) = 1 - \prod_{i=1}^{n}(1 - P(A_i)).$$

**증명.** **전략** 여사건으로 바꾼 사건의 개수 $m$에 대한 귀납법 [T3]. 한 단계에서는 사건 하나만 바꾸고, 차 공식 $P(A^cC) = P(C) - P(AC)$로 곱의 꼴을 되살린다.

① "상호 독립인 $A_1, \dots, A_n$ 가운데 어느 $m$개를 여사건으로 바꿔도 상호 독립이다"를 $\mathcal{S}(m)$이라 하자. $\mathcal{S}(0)$은 가정 그 자체다. 기저

② $\mathcal{S}(m)$을 가정하고 $m+1$개를 바꾼 모임을 생각한다. 바뀐 것 가운데 하나를 첨자를 바꿔 $A_1^c$이라 하자. $A_1$은 그대로 두고 나머지 $m$개만 바꾼 모임 $E_1 = A_1, E_2, \dots, E_n$은 상호 독립이다. 귀납 가정 $\mathcal{S}(m)$

③ $A_1^c, E_2, \dots, E_n$이 상호 독립임을 보이면 된다. 첨자 $i_1 < \cdots < i_k$ $(k \ge 2)$를 고른다. 1이 들어 있지 않으면 ②에서 이미 성립한다. 1이 들어 있으면$(i_1 = 1)$ $C = E_{i_2} \cap \cdots \cap E_{i_k}$로 둔다. 정의 5.9

④ 곱의 꼴을 확인한다.

$$\begin{aligned} P(A_1^c \cap C) &= P(C) - P(A_1 \cap C) && A_1^cC = C \setminus A_1C, \text{ 정리 4.3(d)} \\ &= \textstyle\prod_{l\ge 2} P(E_{i_l}) - P(A_1)\prod_{l\ge 2} P(E_{i_l}) && \text{②의 상호 독립} \\ &= P(A_1^c)\textstyle\prod_{l\ge 2} P(E_{i_l}) && \text{정리 4.3(c)} \end{aligned}$$

⑤ 따라서 $\mathcal{S}(m+1)$이 성립하고, 귀납법으로 모든 $m$에서 성립한다. $m = n$이면 첫째 식이다. 귀납 [T3]

⑥ $\{\text{적어도 하나}\} = \bigcup_i A_i$의 여사건은 $\bigcap_i A_i^c$이므로 둘째 식이 나온다. De Morgan, 보조정리 0.1

■

**예제 5.12** (de Méré의 계산, 다시) 주사위 하나를 4번 던지는 모형은 $\Omega = \{1, \dots, 6\}^4$ 위의 균등 확률이다. $A_i = \{i\text{번째가 } 6\}$이라 하자. 첨자 $i_1 < \cdots < i_k$를 고르면 그 $k$자리가 6이고 나머지 $4-k$자리는 아무 수인 결과가 $6^{4-k}$개이므로

$$P(A_{i_1} \cdots A_{i_k}) = \frac{6^{4-k}}{6^4} = \left(\frac{1}{6}\right)^k = P(A_{i_1}) \cdots P(A_{i_k}).$$

$A_1, \dots, A_4$는 상호 독립이고, 정리 5.11로 $P(6\text{이 적어도 한 번}) = 1 - (5/6)^4$이다. 1.3절의 계산이 이렇게 정당화된다.

판마다 A가 이길 확률이 $p$인 **독립 시행**도 같은 식으로 만든다. $\Omega = \{\mathrm{A}, \mathrm{B}\}^m$에서 A가 $k$

번 나오는 결과에 $p^k q^{m-k}$를 준다. 이항정리로 합이 $(p+q)^m = 1$이므로 정리 4.9에 의해 확률이고, "$i$번째 판은 A가 이긴다"들은 상호 독립이며 확률이 $p$이다(연습 5.8). 정리 1.10의 2단계가 이 모형에서 정당하다.

## 5.5 점수 분배, 다시: Pascal의 재귀

**정리 5.13** (Pascal의 재귀) 독립 시행 모형에서 상태 $(a,b)(a,b \geq 1)$에서 A가 이길 확률 $f(a,b)$는

$$f(a,b) = p\,f(a-1,b) + q\,f(a,b-1), \qquad f(0,b) = 1, \quad f(a,0) = 0$$

을 만족하고, 이 재귀와 경계조건이 $f$를 하나로 정한다. 따라서 Pascal의 풀이와 Fermat의 공식(정리 1.10)은 같은 답을 준다.

**증명.** **재귀.** $W$를 "A가 이긴다", $H$를 "다음 판은 A가 이긴다"라 하자. $\{H, H^c\}$는 분할이므로 전확률 공식으로

$$P(W) = P(W \mid H)P(H) + P(W \mid H^c)P(H^c).$$

$P(H) = p$이고, $H$가 일어나면 남은 승부는 상태 $(a-1,b)$에서 새로 시작하는 것과 같으므로 $P(W \mid H) = f(a-1,b)$이다(다음 판 이후의 판들은 다음 판과 독립이다). 마찬가지로 $P(W \mid H^c) = f(a,b-1)$이다.

**유일성.** $a+b$에 대한 귀납법. $a+b = 1$이면 경계 상태뿐이다. $a+b = s$인 모든 상태의 값이 정해졌다면, $a+b = s+1$인 상태의 값은 재귀식으로 $a+b = s$인 두 상태의 값에서 하나로 정해진다.

**같은 답.** Fermat의 공식도 같은 재귀와 경계조건을 만족하므로(연습 1.6) 유일성에 의해 두 답은 같다. ∎

### 연습문제

**연습 5.1** (두 아이 문제 ★) 아이가 둘인 가정에서 각 아이가 아들일 확률은 $\frac{1}{2}$이고 서로 독립이다. (a) 적어도 하나가 아들일 때 둘 다 아들일 확률은? (b) 첫째가 아들일 때 둘 다 아들일 확률은? 풀이 55쪽

**연습 5.2** (Monty Hall 문제 ★★) 문 세 개 가운데 하나 뒤에 자동차가 있다(같은 확률). 참가자가 1번 문을 고르면, 자동차 위치를 아는 사회자가 나머지 두 문 가운데 염소가 있는 문 하나를 연다(둘 다 염소면 같은 확률로 고른다). 사회자가 3번 문을 열었을 때 (a) 1번을

고수하면 자동차를 얻을 확률과 (b) 2번으로 바꾸면 얻을 확률을 베이즈 정리로 구하라. (c) 사회자가 위치를 모르고 $2\cdot 3$번 가운데 아무 문이나 열었는데 우연히 염소가 나왔다면 답은 어떻게 바뀌는가? (Selvin 1975) 풀이 56쪽

**연습 5.3** (세 상자에서 은화 ★) Bertrand의 세 상자에서 처음 꺼낸 동전이 은화이면 다른 동전도 은화일 확률은? 풀이 56쪽

**연습 5.4** (양의 연관은 대칭이다 ★★) $P(A), P(B) > 0$이면 $P(A \mid B) > P(A) \iff P(B \mid A) > P(B)$임을 보여라. 풀이 56쪽

**연습 5.5** (몇 번 양성이어야 하나 ★★) 유병률이 0.1%인 병에 예제 5.6의 검사(민감도 0.9, 위양성률 0.09)를 독립적으로 거듭 한다. 연속으로 몇 번 양성이어야 사후확률이 $\frac{1}{2}$을 넘는가? 오즈의 식으로 풀어라. 풀이 56쪽

**연습 5.6** (합집합과의 독립 ★★) $A, B, C$가 상호 독립이면 $A$와 $B \cup C$가 독립임을 보여라. 풀이 56쪽

**연습 5.7** (확률 0과 1인 사건 ★) $P(A) = 0$이거나 $P(A) = 1$이면 $A$는 모든 사건과 독립임을 보여라. 풀이 56쪽

**연습 5.8** (독립 시행 모형 ★★) 예제 5.12의 모형 $\Omega = \{\mathrm{A}, \mathrm{B}\}^m$, $P(\omega) = p^{k(\omega)} q^{m-k(\omega)}$에서 $H_i = \{i\text{번째 판은 A}\}$들이 상호 독립이고 $P(H_i) = p$임을 보여라. 풀이 57쪽

**연습 5.9** (Simpson의 역설 ★★) 신장결석 치료 연구(Charig 외 1986)에서 치료 A는 작은 결석 87명 중 81명, 큰 결석 263명 중 192명을 치료했고, 치료 B는 작은 결석 270명 중 234명, 큰 결석 80명 중 55명을 치료했다. (a) 결석 크기마다 성공률을 비교하라. (b) 전체 성공률을 비교하라. (c) 전확률 공식으로 역전이 왜 일어나는지 설명하라. (13회에서 회귀로 다시 다룬다.) 풀이 57쪽

**연습 5.10** (조건부확률은 확률이다: 증명 ★ 유도형) $P(B) > 0$이고 $Q(A) = P(A \mid B)$이다. 정리 5.2를 다음 순서로 증명하라.

(a) (P1)과 (P2)를 확인하라.

(b) $A_1, A_2, \ldots$가 쌍마다 서로소이면 $A_1 B, A_2 B, \ldots$도 쌍마다 서로소이고 $\left(\bigcup_i A_i\right) \cap B = \bigcup_i (A_i B)$임을 원소 추적으로 보여라.

(c) (b)와 $P$의 (P3)에서 $Q$의 (P3)을 얻어라. 급수를 $P(B)$로 나눌 때 무엇을 쓰는가?

풀이 57쪽

**연습 5.11** (곱셈법칙 증명 ★★ 유도형) 정리 5.3을 다음 순서로 증명하라.

(a) $A_1 \supset A_1A_2 \supset \cdots \supset A_1 \cdots A_{n-1}$과 단조성에서 오른쪽의 조건부확률이 모두 정의됨을 보여라.

(b) $n = 2$에서 식은 조건부확률의 정의임을 확인하라.

(c) $n$에서 성립한다고 하고 $C = A_1 \cdots A_n$으로 두어 $P(C \cap A_{n+1}) = P(C)\,P(A_{n+1} \mid C)$에 귀납 가정을 넣어라.

(d) 오른쪽에 정의를 넣으면 분자와 분모가 차례로 소거된다는 한 줄 증명과 (c)를 비교하라. 소거 증명에서 "…"가 숨긴 것은 무엇인가?

풀이 57쪽

# 6

# 확률변수와 분포함수

실험의 결과 그 자체보다 결과에 붙은 **수**가 관심인 경우가 많다. 주사위 4번에서 나온 6의 개수, 현의 길이, 사람이 죽은 나이. 결과에 수를 붙이는 함수가 확률변수이고, 그 수가 어떻게 퍼져 있는지를 한 함수로 요약한 것이 분포함수다. 2회부터의 모든 내용이 이 두 정의 위에 선다.

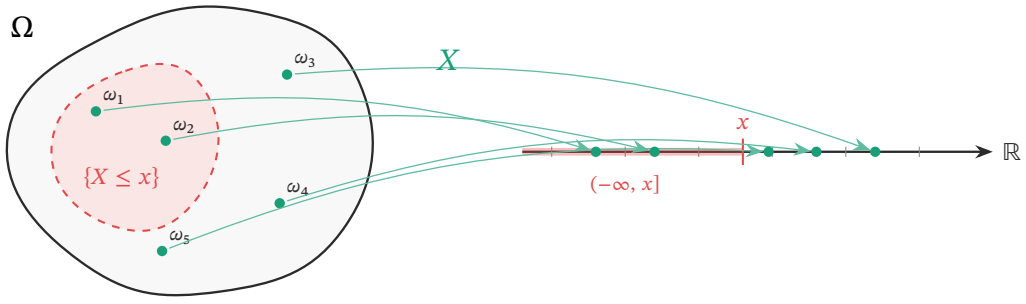
## 6.1 확률변수

**정의 6.1** (확률변수) 확률공간 $(\Omega, \mathcal{F}, P)$ 위의 함수 $X : \Omega \to \mathbb{R}$가 모든 실수 $x$에 대해

$$\{X \le x\} := \{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}$$

를 만족하면 $X$를 **확률변수**라 한다.

조건은 "$X$가 $x$ 이하다"가 언제나 확률을 말할 수 있는 사건이라는 뜻이다. 이것만 요구해도 $\{X < x\} = \bigcup_n \{X \le x - \tfrac{1}{n}\}$, $\{X = x\} = \{X \le x\} \setminus \{X < x\}$, $\{a < X \le b\} = \{X \le b\} \setminus \{X \le a\}$가 모두 사건이 된다($\sigma$-대수의 성질, 연습 4.2). 확률변수는 이름과 달리 "변수"가 아니라 **함수**다. 무작위성은 $\omega$에 있고, $X$는 정해진 규칙이다.




확률변수는 결과 $\omega$마다 실수 $X(\omega)$를 붙이는 **함수**다. $\{X \le x\} = \{\omega : X(\omega) \le x\}$가 언제나 사건(정의 4.1의 $\mathcal{F}$의 원소)이어야 확률 $F(x) = P(X \le x)$를 말할 수 있다. 그림에서는 $\omega_1, \omega_2$가 이 사건에 속하고 $\omega_3, \omega_4, \omega_5$는 속하지 않는다.

**그림 6.1** 확률변수 $X$는 결과 $\omega$에 실수를 붙이는 함수다. $\{X \le x\}$는 $(-\infty, x]$로 보내지는 결과들의 집합이다.

**정의 6.2** (분포함수) 확률변수 $X$의 **분포함수**(cumulative distribution function, cdf)는 $F_X(x) = P(X \le x)$, $x \in \mathbb{R}$이다. 혼동이 없으면 $F$로 쓴다.

## 6.2 분포함수의 성질

**정리 6.3** (분포함수의 성질) 분포함수 $F$는 다음을 만족한다.

(a) (비감소) $x < y$이면 $F(x) \le F(y)$.

(b) (양 끝) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$.

(c) (오른쪽 연속) 모든 $x$에서 $\lim_{y\downarrow x} F(y) = F(x)$.

(d) (왼쪽 극한과 점의 확률) $F(x-) := \lim_{y\uparrow x} F(y) = P(X < x)$이고, 따라서 $P(X = x) = F(x) - F(x-)$, $P(a < X \le b) = F(b) - F(a)$.

**증명.** **전략** $F$의 극한을 사건열의 극한으로 바꾸고 연속성(정리 4.7)을 쓴다. (a)로 $F$가 비감소이므로, 한쪽 극한은 존재하고 단조롭게 다가가는 수열 하나를 따라 구한 값과 같다 [T5]. 그래서 (b)–(d)에서는 수열 하나만 따라간다.

**(a)**

① $x < y$이면 $\{X \le x\} \subset \{X \le y\}$이다. $X(\omega) \le x$이면 $X(\omega) \le y$이기 때문이다. 원소 추적 [T1]

② $F(x) = P(X \le x) \le P(X \le y) = F(y)$이다. 단조성, 정리 4.3(d)

**(b)**

① $\{X \le -n\}$은 감소하고 $\bigcap_n \{X \le -n\} = \varnothing$이다. $X(\omega)$는 실수이므로 모든 $n$에서 $X(\omega) \le -n$일 수는 없다. (a)의 ①, 아르키메데스 성질

② $F(-n) \to P(\varnothing) = 0$이다. 연속성, 정리 4.7(b)

③ $\{X \le n\} \uparrow \Omega$이므로 같은 방법으로 $F(n) \to P(\Omega) = 1$이다. 정리 4.7(a)

**(c)**

① $x_n = x + \frac{1}{n}$이면 $\{X \le x_n\} \downarrow \{X \le x\}$이다. $X(\omega) \le x + \frac{1}{n}$이 모든 $n$에서 성립하면 $X(\omega) \le x$이기 때문이다. 극한과 부등식 [T5]

② $F(x + \frac{1}{n}) \to F(x)$이다. 정리 4.7(b)

**(d)**

① $x_n = x - \frac{1}{n}$이면 $\{X \le x_n\} \uparrow \{X < x\}$이다. $X(\omega) < x$이면 충분히 큰 $n$에서 $X(\omega) \le x - \frac{1}{n}$이기 때문이다. [T1], [T5]

② $F(x-) = \lim_n F(x - \frac{1}{n}) = P(X < x)$이다. 정리 4.7(a)

③ $\{X = x\} = \{X \le x\} \setminus \{X < x\}$이므로 $P(X = x) = F(x) - F(x-)$이다. 정리 4.3(d)

④ $\{a < X \le b\} = \{X \le b\} \setminus \{X \le a\}$이므로 $P(a < X \le b) = F(b) - F(a)$이다. 정리 4.3(d)

■

분포함수는 왼쪽에서는 끊길 수 있지만 오른쪽에서는 늘 이어져 있고, 끊긴 높이가 그 점의 확률이다. 오른쪽 연속인 이유는 정의에서 "$\le$"를 썼기 때문이다.

> **증명 없이 쓰는 사실** (분포함수가 분포를 정한다) 두 확률변수의 분포함수가 같으면, 직선의 모든 "보통의" 집합 $B$(구간들로부터 합집합·여집합을 셀 수 있는 번 반복해 만든 Borel 집합)에서 $P(X \in B) = P(Y \in B)$이다. 거꾸로 (a)–(c)를 만족하는 함수는 모두 어떤 확률변수의 분포함수다. 둘 다 측도론의 Carathéodory 확장 정리에서 나온다. 이 과정에서는 "분포"를 분포함수(또는 아래의 pmf·pdf)로 다룬다.

## 6.3 이산 확률변수와 연속 확률변수

**정의 6.4** (이산 확률변수와 확률질량함수) 확률변수 $X$가 유한하거나 셀 수 있는 집합 $\{x_1, x_2, \dots\}$ 안의 값만 확률 1로 가지면 $X$를 **이산**이라 하고, $p(x) = P(X = x)$를 **확률질량함수**(pmf)라 한다. 이때 모든 $B \subset \mathbb{R}$에 대해 $P(X \in B) = \sum_{x_i \in B} p(x_i)$이고, 분포함수는 계단 $F(x) = \sum_{x_i \le x} p(x_i)$이다.

식 $P(X \in B) = \sum_{x_i \in B} p(x_i)$는 $\{X \in B\}$를 서로소인 사건 $\{X = x_i\}$, $x_i \in B$로 나누고 (P3)을 쓰면 나온다.

**예제 6.5** (주사위 4번에서 나온 6의 개수) 예제 5.12의 모형에서 $K = 6$의 개수라 하자. $K = k$인 결과는 6이 나올 $k$자리를 고르고($\binom{4}{k}$가지) 나머지 $4-k$자리에 6이 아닌 수를 넣는($5^{4-k}$가지) 것이므로

$$P(K = k) = \binom{4}{k}\frac{5^{4-k}}{6^4} = \binom{4}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{4-k}, \qquad k = 0, 1, 2, 3, 4.$$

값은 차례로 0.482, 0.386, 0.116, 0.015, 0.001이고, $P(K \ge 1) = 1 - 0.482 = 0.518$이 de Méré의 내기 1이다. 이 꼴의 분포가 4회의 **이항분포**다.

**정의 6.6** (연속 확률변수와 확률밀도함수) 음이 아니고 적분 가능한 함수 $f$가 있어 모든 $x$에서

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

이면 $X$를 **연속**이라 하고, $f$를 **확률밀도함수**(pdf)라 한다.

**정리 6.7** (밀도의 성질) $X$가 밀도 $f$를 가지면

(a) $\int_{-\infty}^{\infty} f(t)\,dt = 1$.

(b) 모든 $x$에서 $P(X = x) = 0$.

(c) $P(a < X \le b) = P(a \le X \le b) = P(a < X < b) = \int_a^b f(t)\,dt$.

(d) $f$가 $x$에서 연속이면 $F'(x) = f(x)$.

**증명.** **전략** 모든 주장을 정리 6.3의 성질과 적분의 성질로 옮긴다. (b)에서는 적분으로 정의된 함수가 연속이라는 것을 쓴다. 단계별 증명은 연습 6.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 59쪽에 있다.

밀도 $f(x)$ 자체는 확률이 아니다. 1보다 클 수도 있고 무한대로 갈 수도 있다. 확률은 넓이 $\int_a^b f$이고, 작은 $h$에 대해 $P(x < X \le x+h) \approx f(x)\,h$로 읽는다. 가장 간단한 예는 $(0,1)$ 위의 **균등분포** $\mathrm{U}(0,1)$이다. $f(x) = 1$ $(0 < x < 1)$, 그 밖에서 0이므로 구간의 확률이 구간의 길이다.

**예제 6.8** (Bertrand의 현의 길이) 반지름 $r = 1$인 원에서 현의 길이 $L$의 분포함수를 세 방법 각각에서 구한다$(0 \le \ell \le 2)$.

**(1) 두 끝점.** 다른 끝점의 중심각 $\Phi$가 $\mathrm{U}(0, 2\pi)$이고 $L = 2\sin(\Phi/2)$이다. $\Psi = \Phi/2 \sim \mathrm{U}(0,\pi)$에 대해 $\sin\Psi \le s$는 $\Psi \le \arcsin s$ 또는 $\Psi \ge \pi - \arcsin s$와 같으므로

$$F_1(\ell) = P\bigl(\sin\Psi \le \tfrac{\ell}{2}\bigr) = \frac{2}{\pi}\arcsin\frac{\ell}{2}, \qquad f_1(\ell) = \frac{2}{\pi\sqrt{4-\ell^2}}.$$

**(2) 반지름 위의 점.** $D \sim \mathrm{U}(0,1)$, $L = 2\sqrt{1-D^2}$이므로 $L \le \ell \iff D \ge \sqrt{1-\ell^2/4}$이고

$$F_2(\ell) = 1 - \sqrt{1-\ell^2/4}, \qquad f_2(\ell) = \frac{\ell}{2\sqrt{4-\ell^2}}.$$

**(3) 원 안의 중점.** 중점이 원판에 고르게 퍼져 있으면 중심과의 거리 $R$은 $P(R \le \rho) = \rho^2$이다(반지름 $\rho$인 원의 넓이 비). $L \le \ell \iff R^2 \ge 1 - \ell^2/4$이므로

$$F_3(\ell) = \frac{\ell^2}{4}, \qquad f_3(\ell) = \frac{\ell}{2}.$$

세 경우 모두 $1 - F(\sqrt{3})$을 계산하면 $\frac{1}{3}$, $\frac{1}{2}$, $\frac{1}{4}$로 3절의 답이 다시 나온다. $f_1$과 $f_2$는 $\ell \to 2$에서 무한대로 가지만 적분은 1이다.

이산도 연속도 아닌 확률변수도 있다. 보험금처럼 확률 0.3으로 정확히 0이고, 나머지 0.7의 확률

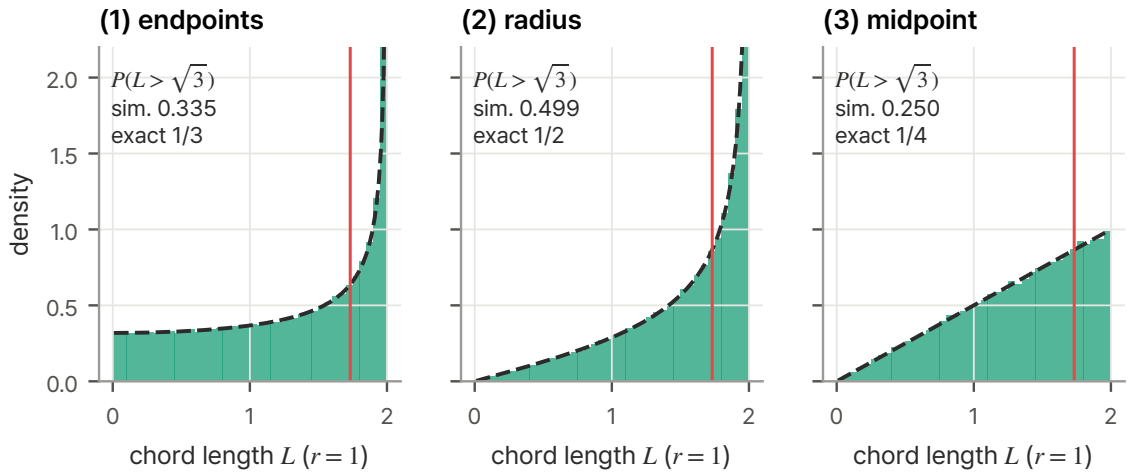



**그림 6.2** 세 방법으로 12만 개씩 그은 현의 길이 히스토그램과 예제 6.8의 밀도(점선). 빨간 선은 $\sqrt{3}$이다.

로는 밀도 $e^{-x}(x>0)$를 따라 퍼지는 양이 그렇다. 분포함수는 $x \ge 0$에서 $F(x) = 0.3 + 0.7(1-e^{-x})$이고 $x=0$에서 높이 0.3만큼 뛴다(그림 6.3의 오른쪽).

### 연습문제

**연습 6.1** (분포함수에서 확률 읽기 ★) 방법 (3)의 현의 길이 $L$에 대해 $P(0.5 < L \le 1.5)$, 중앙값($F(m)=\frac{1}{2}$인 $m$), 밀도를 구하라. 풀이 58쪽

**연습 6.2** (뛰는 점은 셀 수 있다 ★★) 분포함수의 불연속점은 기껏해야 셀 수 있음을 보여라. (힌트: 높이가 $1/n$보다 큰 점프는 $n$개를 넘을 수 없다.) 풀이 58쪽

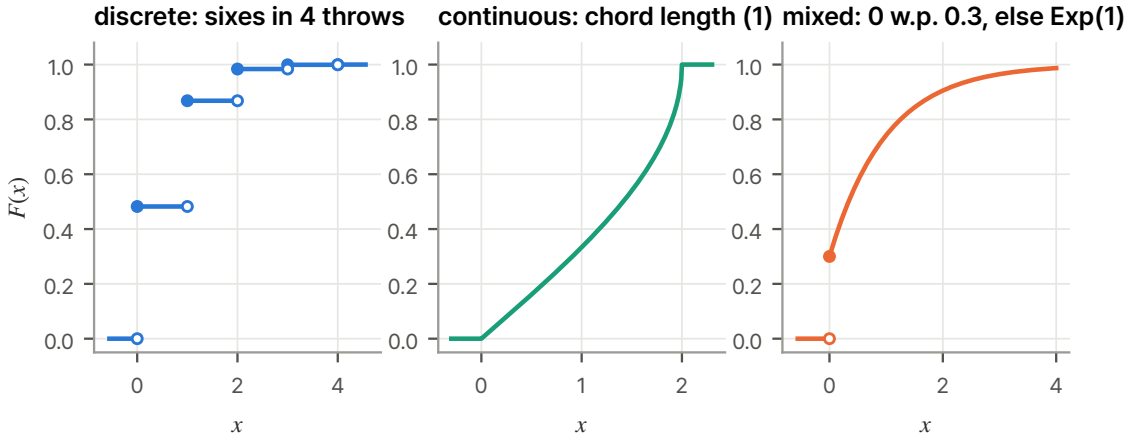



**그림 6.3** 세 종류의 분포함수. 왼쪽: 계단(이산). 가운데: 연속. 오른쪽: 0에서 뛰고 그 뒤로 이어지는 혼합형. 채운 점이 함숫값, 빈 점이 왼쪽 극한이다.

**연습 6.3** (혼합형의 확률 ★) 본문의 혼합형 $X$(확률 0.3으로 0, 나머지는 밀도 $e^{-x}$)에 대해 $P(X=0)$, $P(X\le 1)$, $P(0<X\le 1)$을 구하라. 풀이 58쪽

**연습 6.4** (죽는 나이의 분포 ★) Halley의 1세 아이 한 명이 죽는 나이를 $X$라 하자(표의 상대도수를 확률로 읽는다). $P(X\ge 40)$과 $P(20\le X<40)$을 구하라. 풀이 58쪽

**연습 6.5** (세 방법의 중앙값 ★★) 예제 6.8의 세 분포에서 현의 길이의 중앙값을 구하라. (1)과 (3)은 분포가 다른데 중앙값이 같다. 분포를 비교하는 데 중앙값 하나로 충분하지 않은 이유를 그림 6.2으로 설명하라. 풀이 58쪽

**연습 6.6** (분포함수의 거듭제곱 ★★) $F$가 분포함수이면 $G(x)=F(x)^n$도 분포함수의 세 성질을 만족함을 보여라. (3회에서 이것이 서로 독립인 $n$개의 최댓값의 분포함수임을 본다.) 풀이 58쪽

**연습 6.7** (제곱의 분포 ★★) $X\sim \mathrm{U}(0,1)$, $Y=X^2$이다. $F_Y(y)=P(X\le\sqrt{y})$에서 $Y$의 분포함수와 밀도를 구하라. (2회의 변수변환 예고) 풀이 58쪽

**연습 6.8** (밀도의 성질 증명 ★★ 유도형) $X$가 밀도 $f$를 갖는다. 정리 6.7를 다음 순서로 증명하라.

(a) 정리 6.3(b)에서 $\int_{-\infty}^{\infty} f = 1$을 얻어라. 이상적분 $\int_{-\infty}^{\infty} f$의 정의를 먼저 적어라.

(b) $h>0$이면 $0\le F(x)-F(x-h)=\int_{x-h}^{x} f(t)\,dt$이다. 이것이 $h\downarrow 0$에서 0으로 감을 보이고, 정리 6.3(d)로 $P(X=x)=0$을 얻어라. ($f$가 유계가 아닐 수 있으므로, $\int_{-\infty}^{x} f<\infty$를 쓴다.)

(c) (b)와 정리 4.3(d)로 네 구간의 확률이 모두 $\int_a^b f$임을 보여라.

(d) $f$가 $x$에서 연속이면 $\left|\frac{F(x+h)-F(x)}{h}-f(x)\right|\le \sup_{|t-x|\le|h|}|f(t)-f(x)|$임을 보이고 $F'(x)=f(x)$를 얻어라.

풀이 59쪽

# 7

# 부쉬보기 — 공리의 가정을 빼면

공리의 조건 하나하나는 장식이 아니다. 가산가법성이나 $\sigma$-대수를 빼거나, 조건부확률의 조건 $P(B) > 0$을 무시하면 무엇이 무너지는지 본다.

## 7.1 자연수 위에는 "균등한" 확률이 없다

"자연수 하나를 무작위로 고른다"는 말은 흔히 쓰인다. 짝수일 확률은 $\frac{1}{2}$이라고 말하고 싶다. 그러나 모든 자연수가 같은 확률을 갖는 확률은 존재하지 않는다.

> **정리 7.1** (자연수 위의 균등 확률은 없다) $\Omega = \mathbb{N} = \{1, 2, 3, \dots\}$ 위에 모든 $n$에서 $P(\{n\}) = c$인 확률 $P$는 없다.

**증명.** **전략** 그런 $P$가 있다고 가정하고 가산가법성에서 모순을 끌어낸다 [T2].

① 모든 $n$에서 $P(\{n\}) = c$인 확률 $P$가 있다고 하자. — 귀류 가정

② $\mathbb{N} = \bigcup_{n=1}^{\infty} \{n\}$이고 한 점 집합 $\{n\}$들은 쌍마다 서로소다. — [T1]

③ $1 = P(\mathbb{N}) = \sum_{n=1}^{\infty} P(\{n\}) = \sum_{n=1}^{\infty} c$이다. — (P2), (P3)

④ $c = 0$이면 오른쪽은 0이다. $c > 0$이면 부분합 $Nc$가 $N \to \infty$에서 무한대로 가므로 급수는 발산한다. — 급수의 정의 [T5]

⑤ 어느 경우에도 ③이 성립하지 않으므로 ①의 가정은 틀렸다. — 모순

■

**가정 주의** 가산가법성을 유한가법성으로 약하게 하면 "균등한" 값을 줄 수는 있다. 예를 들어 **자연 밀도** $d(A)=\lim_{n\to\infty}|A\cap\{1,\dots,n\}|/n$은 짝수에 $\frac{1}{2}$, 3의 배수에 $\frac{1}{3}$을 준다. 그러나 $d$는 극한이 없는 집합에는 정의되지 않고, 한 점의 밀도는 0인데 그 가산 합집합인 $\mathbb{N}$의 밀도는 1이므로 가산가법성을 어긴다(연습 7.2). 그러면 연속성(정리 4.7)도 무너지고, 분포함수의 성질(정리 6.3)도 증명할 수 없다.

## 7.2 모든 부분집합에 확률을 줄 수는 없다 (Vitali, 1905)

원 위의 "고른" 점에는 원둘레의 모든 부분집합에 확률을 주고 싶다. 원을 $[0,1)$로 펴고 끝을 이어 붙여(덧셈은 1을 법으로 한다) 생각하자. "고르다"의 최소한의 요구는 회전해도 확률이 같다는 것이다.

**정리 7.2** (Vitali의 집합) $[0,1)$의 **모든** 부분집합에 정의되고 회전에 불변인($P(A+t \bmod 1)=P(A)$) 확률 $P$는 없다.

**증명.** **1단계: 묶기.** $x-y$가 유리수이면 $x\sim y$로 두면 이것은 동치관계이고, $[0,1)$은 서로소인 동치류로 나뉜다. 선택공리로 각 동치류에서 대표를 하나씩 골라 그 집합을 $V$라 한다.

**2단계: 회전한 복사본.** 유리수 $q\in\mathbb{Q}\cap[0,1)$마다 $V_q=V+q \bmod 1$로 둔다. 이 $V_q$들은 쌍마다 서로소다. $x\in V_q\cap V_{q'}$이면 $x=v+q=v'+q' \pmod 1$인 대표 $v,v'$가 있어 $v-v'$가 유리수이고, 같은 동치류의 대표는 하나뿐이므로 $v=v'$, 따라서 $q=q'$이다. 또 모든 $x\in[0,1)$은 어떤 대표 $v$와 유리수만큼 차이 나므로 어떤 $V_q$에 속한다. 즉 $\{V_q\}$는 $[0,1)$의 셀 수 있는 분할이다.

**3단계: 모순.** 회전 불변성으로 모든 $q$에서 $P(V_q)=P(V)$이다. 가산가법성으로 $1=\sum_q P(V_q)=\sum_q P(V)$이고, 이 합은 $P(V)=0$이면 0, $P(V)>0$이면 무한대다. ∎

그래서 원이나 구간 위에서는 사건을 모든 부분집합이 아니라 구간에서 출발해 만든 $\sigma$-대수(Borel 집합)로 제한한다. 정의 4.1에서 사건의 모임 $\mathcal{F}$를 따로 둔 이유가 이것이다. Vitali의 $V$처럼 확률을 줄 수 없는 집합은 선택공리 없이는 만들 수 없을 만큼 기이해서, 통계에서 실제로 다루는 사건은 모두 Borel 집합이다.

## 7.3 확률 0인 사건에 조건을 걸면

정의 5.1는 $P(B)>0$을 요구한다. 그런데 연속 확률변수에서는 $P(X=x)=0$이므로(정리 6.7), "$X=x$일 때 $Y$의 확률"을 정의대로는 말할 수 없다. 3회에서 이것을 밀도의 비로 정의하는데,

그 정의는 "$X = x$"에 어떻게 가까이 가느냐에 따라 달라질 수 있다(Borel–Kolmogorov 역설). Bertrand의 역설이 "무작위로"의 모호함이었다면, 이것은 "확률 0인 조건"의 모호함이다.

### 연습문제

**연습 7.1** (정수와 반직선 ★) (a) 정수 $\mathbb{Z}$ 위에 모든 점의 확률이 같은 확률은 없음을 보여라. (b) $[0,\infty)$ 위에, 길이가 1인 모든 구간 $[n, n+1)$에 같은 확률을 주는 확률은 없음을 보여라. 풀이 59쪽

**연습 7.2** (자연 밀도 ★★) 자연 밀도 $d(A) = \lim_n |A \cap \{1,\dots,n\}|/n$에 대해 (a) $d(\text{짝수}) = \frac{1}{2}$, $d(3\text{의 배수}) = \frac{1}{3}$을 보이고, (b) $d$가 서로소인 두 집합에서는 더해지지만(극한이 있을 때) 가산가법성은 어김을 보여라. 풀이 59쪽

# 8

# Beautiful insights — 1회의 수학은 어디에 있나

## 8.1 시뮬레이션은 상대도수 해석이다

노트북은 거의 모든 확률을 정확한 값과 시뮬레이션으로 두 번 구한다. 시뮬레이션은 2절의 상대도수 해석을 컴퓨터로 옮긴 것이다. 같은 실험을 독립적으로 $n$번 되풀이해 사건이 일어난 비율 $\hat{p}$를 확률 $p$의 추정으로 쓴다.

얼마나 정확한가? 2회에서 Chebyshev 부등식으로, 7회에서 중심극한정리로 증명하겠지만, $\hat{p}$는 대략 $p \pm 2\sqrt{p(1-p)/n}$ 안에 떨어진다. 오차는 $1/\sqrt{n}$의 속도로만 준다. 소수 셋째 자리까지 믿으려면(표준오차 0.0005) de Méré의 내기 1에서 $n \approx p(1-p)/0.0005^2 \approx 10^6$번이 필요하다. 시뮬레이션은 정확한 계산을 대신하는 것이 아니라 **확인**하는 도구다.

## 8.2 기저율 오류는 분류기의 precision이다

예제 5.6의 진단검사는 이진 분류기와 같다. 민감도 $P(+ \mid D)$는 기계학습의 **재현율**(recall), $P(D \mid +)$는 **정밀도**(precision)다. 베이즈 정리가 말하는 것은 정밀도가 분류기만의 성질이 아니라 **양성 비율(기저율)에 의존한다**는 것이다. 같은 모델이라도 양성이 1%인 데이터에서는 정밀도가 9%, 30%인 데이터에서는 81%가 될 수 있다(그림 5.1). 불균형 데이터에서 정확도나 정밀도를 서로 다른 데이터셋끼리 비교할 때 이 점을 잊으면 안 된다.

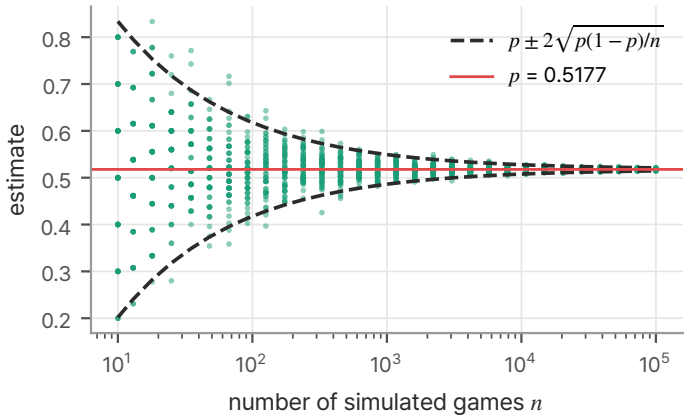



**그림 8.1** de Méré의 내기 1을 $n$번 시뮬레이션한 추정값(크기마다 60번 반복). 점선은 $p \pm 2\sqrt{p(1-p)/n}$이다.

## 8.3 베이즈 정리와 naive Bayes 분류기

스팸 필터의 고전적인 형태는 베이즈 정리에 "클래스가 주어지면 단어들이 독립"이라는 가정(조건부 독립)을 더한 것이다. 메일의 단어가 $w_1, \dots, w_m$이면

$$\frac{P(\text{스팸} \mid w_1,\dots,w_m)}{P(\text{정상} \mid w_1,\dots,w_m)} = \frac{P(\text{스팸})}{P(\text{정상})} \prod_{j=1}^{m} \frac{P(w_j \mid \text{스팸})}{P(w_j \mid \text{정상})}.$$

5.2절의 오즈 공식을 단어마다 되풀이한 것이다. 예를 들어 사전 스팸 비율이 0.4이고 "무료"가 스팸의 30%, 정상 메일의 2%에 나오면 "무료"가 든 메일의 스팸 확률은 0.91이다. 같은 메일에 정상 메일에 흔한 "회의"(스팸 1%, 정상 10%)도 있으면 가능도비 0.1이 곱해져 0.5로 내려간다(연습 8.2). "naive"는 조건부 독립이 대개 사실이 아니라는 뜻이지만, 분류에는 놀랄 만큼 잘 작동한다.

## 8.4 생일 문제와 해시 충돌

$k$비트 해시 함수의 출력이 $N = 2^k$개의 값에 고르게 퍼진다면, 서로 다른 입력 $n$개 가운데 출력이 같은 쌍이 생길 확률은 생일 문제와 같은 계산으로 대략 $1 - e^{-n^2/(2N)}$이다. 이것이 $\frac{1}{2}$이 되는 $n$은 $\sqrt{2 \ln 2}\, 2^{k/2} \approx 1.18 \cdot 2^{k/2}$이다. 64비트 해시라도 약 50억 개면 충돌이 반반이다. 암호에서는 이것을 **생일 공격**이라 하고, 그래서 충돌을 막아야 하는 해시는 원하는 안전 비트의 두 배 길이로 만든다.

### 연습문제

**연습 8.1** (몇 번 시뮬레이션해야 하나 ★) 내기 2($p \approx 0.4914$)가 불리하다는 것을 시뮬레이션으로 보이려면, $\hat{p}$의 두 표준오차가 $p$와 $\frac{1}{2}$의 차이보다 작아야 한다. 대략 몇 번이 필요한가? 풀이 59쪽

**연습 8.2** (naive Bayes 계산 ★) 본문의 수(사전 스팸 비율 0.4, "무료" 0.3/0.02, "회의" 0.01/0.1)로 두 확률 0.91과 0.5를 직접 계산하라. 풀이 59쪽

# 9

# 정리와 역사 노트

## 9.1 한 장 요약

| | 내용 | 증명의 핵심 |
|---|---|---|
| **셈의 원리** | $\frac{n!}{(n-k)!}$ 순열, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, Pascal 항등식, 이항정리 | 곱의 법칙, 원소 하나로 나누기 |
| **점수 분배** | $f(a,b) = \sum_{k\geq a}\binom{a+b-1}{k}p^k q^{a+b-1-k}$ | 가상의 연장 (Fermat) |
| **공리** | $P \geq 0, P(\Omega) = 1$, 가산가법성 | — |
| **기본 성질** | 여사건, 단조성, $P(A \cup B)$, Boole | 서로소로 쪼개기 |
| **포함배제** | $P(\bigcup A_i) = \sum_k (-1)^{k+1} S_k$ | 귀납법 / 각 점이 한 번 |
| **Montmort** | $p_n = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} \to 1 - \frac{1}{e}$ | $S_k = \frac{1}{k!}$ |
| **연속성** | $A_n \uparrow A \Rightarrow P(A_n) \to P(A)$ | $B_n = A_n \setminus A_{n-1}$ |
| **고전적 정의** | 균등 확률공간에서 $P(A) = \lvert A\rvert/\lvert\Omega\rvert$ | 이산 확률공간 |
| **조건부확률** | $P(A \mid B) = P(AB)/P(B)$, 그 자체로 확률 | 공리 확인 |
| **세 법칙** | 곱셈법칙, 전확률 공식, 베이즈 정리 | 소거, 분할 |
| **독립** | $P(AB) = P(A)P(B)$, 상호 독립은 모든 부분 모임 | 여사건으로 바꾸기 |
| **Pascal의 재귀** | $f(a,b) = pf(a-1,b) + qf(a,b-1)$ | 전확률 + 유일성 |
| **분포함수** | 비감소, 양 끝 0과 1, 오른쪽 연속, $P(X = x) = F(x) - F(x-)$ | 연속성 정리 |
| **부숴보기** | $\mathbb{N}$ 위의 균등 확률 없음, Vitali 집합 | 가산가법성 |

## 9.2 Stigler의 명명 법칙

> **Stigler's law of eponymy** "어떤 과학 법칙도 처음 발견한 사람의 이름을 달고 있지 않다."
>
> - **Pascal의 삼각형**: 중국의 가헌(Jia Xian, 11세기)과 이를 전한 양휘(Yang Hui, 1261), 페르시아의 al-Karaji(1000년 무렵)가 먼저 알았다. 이탈리아에서는 Tartaglia의 삼각형이라고도 부른다.
> - **베이즈 정리**: 일반적인 형태로 쓰고 널리 퍼뜨린 사람은 Laplace(1774)였다. Stigler(1983)는 맹인 수학자 Nicholas Saunderson이 Bayes보다 먼저 알았을 가능성을 제기했다.
> - **Monty Hall 문제**: 같은 함정이 Bertrand(1889)의 상자 역설에 이미 있다.

## 9.3 7 Pillars 지도

Stigler(2016)의 일곱 기둥(Aggregation, Information, Likelihood, Intercomparison, Regression, Design, Residual)은 모두 이번 회차의 언어 위에 선다. 1회는 기둥을 세울 땅, 곧 **확률의 언어**다. 첫 기둥 **Aggregation**(관측을 모으면 정보가 늘어난다)은 2회에서 Bernoulli의 대수의 법칙과 함께 세운다.

## 9.4 다음 회 — 기댓값 · 부등식 · 대수의 법칙 (1657 → 1867)

Huygens의 "기대 몫"(1.5절)을 확률변수의 **기댓값**으로 정의하고, 분산과 적률생성함수를 세운다. 그리고 Jacob Bernoulli가 20년을 들여 증명한 "황금정리"로 간다. 동전을 몇 번 던지면 앞면의 비율을 "거의 확실하게" 알 수 있는가? Bernoulli의 답은 25,550번이었다. 2절의 상대도수 해석이 왜 정당한지, 8.1절의 $1/\sqrt{n}$이 어디서 오는지가 그 답 안에 있다.

## 참고문헌


- Stigler, S. M. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard, 1986.
- Stigler, S. M. Who discovered Bayes's theorem? *The American Statistician* 37 (1983).
- Stigler, S. M. *The Seven Pillars of Statistical Wisdom*. Harvard, 2016.
- Hald, A. *A History of Probability and Statistics and Their Applications before 1750*. Wiley, 1990.
- Hacking, I. *The Emergence of Probability*. Cambridge, 1975.
- Kolmogorov, A. N. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, 1933.
- Bertrand, J. *Calcul des probabilités*. Gauthier-Villars, 1889.
- Halley, E. An estimate of the degrees of the mortality of mankind. *Philosophical Transactions* 17 (1693).
- Jaynes, E. T. The well-posed problem. *Foundations of Physics* 3 (1973).
- Selvin, S. A problem in probability (letter). *The American Statistician* 29 (1975).

- Charig, C. R. 외. Comparison of treatment of renal calculi. *British Medical Journal* 292 (1986).

# A

# 연습문제 풀이

각 풀이의 머리에 있는 문제 ○○쪽을 누르면 문제로 돌아간다. 계산 문제의 수치는 노트북에서 같은 계산으로 확인했다. 유도형 연습의 풀이는 본문에서 연습으로 돌린 정리의 완성된 증명이다.

## 0절

### 연습 0.1 문제 3쪽

(a) $\omega \in (\bigcup_i A_i)^c$이면 $\omega \notin \bigcup_i A_i$이므로 어떤 $i$에서도 $\omega \in A_i$가 아니다. 곧 모든 $i$에서 $\omega \in A_i^c$이고 $\omega \in \bigcap_i A_i^c$이다. (b) $\omega \in \bigcap_i A_i^c$이면 모든 $i$에서 $\omega \notin A_i$이므로 $\omega$는 어느 $A_i$에도 속하지 않고, $\omega \notin \bigcup_i A_i$, 곧 $\omega \in (\bigcup_i A_i)^c$이다. (a)와 (b)로 두 집합은 서로 포함하므로 같다. (c) 첫째 식을 집합족 $\{A_i^c\}$에 쓰면 $(\bigcup_i A_i^c)^c = \bigcap_i (A_i^c)^c = \bigcap_i A_i$이다. 여기서 $(A_i^c)^c = A_i$를 썼다. 양변의 여집합을 취하고 다시 $(B^c)^c = B$를 쓰면 $\bigcup_i A_i^c = (\bigcap_i A_i)^c$이다. 첨자 집합 $I$가 유한한지 무한한지는 어디에도 쓰이지 않았다.

## 1절

### 연습 1.1 문제 10쪽

같은 숫자 3장의 숫자를 고르고(13가지) 그 숫자의 무늬 3개를 고른다($\binom{4}{3} = 4$). 나머지 2장의 숫자는 남은 12개 가운데 하나이고 무늬 2개를 고른다($\binom{4}{2} = 6$). 곱의 법칙으로 $13 \cdot 4 \cdot 12 \cdot 6 = 3744$가지이므로

$$P(\text{풀하우스}) = \frac{3744}{2{,}598{,}960} \approx 0.00144.$$

### 연습 1.2 문제 10쪽

6개를 고르는 방법은 $\binom{45}{6} = 8{,}145{,}060$가지이고 1등은 그 가운데 하나다. 정확히 5개가 맞으려면 당첨 번호 6개에서 5개($\binom{6}{5} = 6$), 나머지 39개에서 1개(39)를 고른다.

$$P(1\text{등}) = \frac{1}{8{,}145{,}060} \approx 1.23 \times 10^{-7}, \qquad P(\text{정확히 } 5\text{개}) = \frac{6 \cdot 39}{8{,}145{,}060} \approx 2.87 \times 10^{-5}.$$

### 연습 1.3 문제 10쪽

(a) $2n$명에서 $n$명을 고르는 방법은 $\binom{2n}{n}$이다. 이것을 고른 남자의 수 $k$로 나누면 남자 $k$명($\binom{n}{k}$)과 여자 $n-k$명($\binom{n}{n-k}=\binom{n}{k}$)을 고르는 방법이므로 $\sum_k \binom{n}{k}^2$이다. (b) 이항정리로 $(1+x)^n(1+x)^n$에서 $x^n$의 계수는 $\sum_k \binom{n}{k}\binom{n}{n-k}=\sum_k \binom{n}{k}^2$이고, $(1+x)^{2n}$에서는 $\binom{2n}{n}$이다.

### 연습 1.4 문제 10쪽

합 11: $\{1,4,6\},\{1,5,5\},\{2,3,6\},\{2,4,5\},\{3,3,5\},\{3,4,4\}$의 6가지 조합이고, 순서쌍은 $6+3+6+6+3+3=27$가지. 합 12: $\{1,5,6\},\{2,4,6\},\{2,5,5\},\{3,3,6\},\{3,4,5\},\{4,4,4\}$이고, 순서쌍은 $6+6+3+3+6+1=25$가지. 분포가 10.5를 중심으로 대칭이므로 11과 10, 12와 9의 수가 같다(그림 1.1).

### 연습 1.5 문제 10쪽

$1-(5/6)^n>\frac{1}{2} \iff n>\log 2/\log(6/5)=3.80$이므로 $n=4$다($n=3$이면 0.421, $n=4$이면 0.518). 비례 근사 $0.693\times 6=4.16$은 정확한 경계 3.80보다 크다. $p=\frac{1}{6}$은 "작은 $p$"가 아니어서 $-\log(1-p)\approx p$가 잘 맞지 않는다.

### 연습 1.6 문제 10쪽

$m=a+b-1$로 두면 $f(a-1,b)$와 $f(a,b-1)$은 모두 $m-1$판짜리 합이다.

$$p\,f(a-1,b)+q\,f(a,b-1)=\sum_{k\ge a-1}\binom{m-1}{k}p^{k+1}q^{m-1-k}+\sum_{k\ge a}\binom{m-1}{k}p^kq^{m-k}$$
$$=\sum_{j\ge a}\left[\binom{m-1}{j-1}+\binom{m-1}{j}\right]p^jq^{m-j}=\sum_{j\ge a}\binom{m}{j}p^jq^{m-j}=f(a,b).$$

첫 합에서 $j=k+1$로 바꿨고, 마지막에 Pascal 항등식(정리 1.5)을 썼다. 경계도 맞는다. $f(0,b)$는 $m=b-1$판의 이항 확률을 모두 더한 것이므로 이항정리로 1이고, $f(a,0)$은 $k\ge a>m=a-1$인 항이 없으므로 0이다.

### 연습 1.7 문제 10쪽

$m=4$판에서 A가 2판 이상 이길 확률이다.

$$\binom{4}{2}(0.6)^2(0.4)^2+\binom{4}{3}(0.6)^3(0.4)+(0.6)^4=0.3456+0.3456+0.1296=0.8208.$$

판마다 0.6의 우위가 승부에서는 0.69에서 0.82로 커진다.

### 연습 1.8 문제 11쪽

(a) 원소 $n$을 포함하는 $k$개짜리 부분집합은 나머지 $k-1$개를 $\{1,\dots,n-1\}$에서 고른 것과 하나씩 대응하므로 $\binom{n-1}{k-1}$개다. 포함하지 않는 것은 $k$개를 모두 $\{1,\dots,n-1\}$에서 고른 것이므로 $\binom{n-1}{k}$

개다. 모든 부분집합은 $n$을 포함하거나 포함하지 않으며 둘 다일 수는 없으므로, 두 묶음은 겹치지 않고 합치면 전체다. (b) $1 \le k \le n-1$이면

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \frac{(n-1)!}{(k-1)!\,(n-k)!} + \frac{(n-1)!}{k!\,(n-1-k)!} = \frac{(n-1)!}{k!\,(n-k)!}\big(k + (n-k)\big) = \binom{n}{k}.$$

통분할 때 $k! = k\cdot(k-1)!$, $(n-k)! = (n-k)\cdot(n-1-k)!$을 썼다. (c) $k=0$: 왼쪽은 1, 오른쪽은 $\binom{n-1}{-1} + \binom{n-1}{0} = 0 + 1$. $k = n$: 왼쪽은 1, 오른쪽은 $\binom{n-1}{n-1} + \binom{n-1}{n} = 1 + 0$. $k < 0$이나 $k > n$이면 양변이 0이다. (a)의 세기는 $k$의 범위를 가리지 않는다. $k = 0$이면 "$n$을 포함하는 0개짜리 부분집합" 이 없다는 것이 바로 $\binom{n-1}{-1} = 0$이다. (b)는 계승의 식이 $0 \le k \le n$에서만 정의되므로 경계를 따로 봐야 한다.

### 연습 1.9 문제 11쪽

(a) $(x+y)^0 = 1$이고 오른쪽은 $\binom{0}{0}x^0y^0 = 1$이다. (b) 귀납 가정을 넣으면

$$(x+y)^{n+1} = \sum_{k=0}^{n}\binom{n}{k}x^{k+1}y^{n-k} + \sum_{k=0}^{n}\binom{n}{k}x^{k}y^{n+1-k} = \sum_{j=1}^{n+1}\binom{n}{j-1}x^{j}y^{n+1-j} + \sum_{j=0}^{n}\binom{n}{j}x^{j}y^{n+1-j}.$$

첫째 합에서 $j = k+1$로 바꿨다. (c) $1 \le j \le n$인 항은 두 합에 모두 있고 계수는 $\binom{n}{j-1}+\binom{n}{j} = \binom{n+1}{j}$이다(정리 1.5). $j = 0$인 항은 둘째 합에만 있고 계수 $\binom{n}{0} = 1 = \binom{n+1}{0}$이다. $j = n+1$인 항은 첫째 합에만 있고 계수 $\binom{n}{n} = 1 = \binom{n+1}{n+1}$이다. 따라서 $(x+y)^{n+1} = \sum_{j=0}^{n+1}\binom{n+1}{j}x^jy^{n+1-j}$이다. (d) $(x+y)^n = (x+y)(x+y)\cdots(x+y)$을 분배법칙만으로 전개하면, 각 인수에서 $x$나 $y$를 하나씩 골라 곱한 $2^n$개의 항이 나온다. $x$를 고른 인수가 $k$개인 항은 $x^ky^{n-k}$이고, $x$를 고를 $k$개의 인수를 정하는 방법은 $n$개 인수의 $k$개짜리 부분집합의 수 $\binom{n}{k}$다. 귀납법은 Pascal의 항등식을 쓰고, 세기는 그것을 쓰지 않는다.

## 2절

### 연습 2.1 문제 14쪽

30세 생존자 531명 가운데 40세까지 산 사람은 445명이므로 그 사이에 죽은 비율은 $(531-445)/531 \approx 0.162$다.

### 연습 2.2 문제 14쪽

1세 아이가 34세 이후까지 살 비율이 처음으로 절반 아래로 내려간다는 뜻이다. 죽는 나이를 확률변수 $X$로 보면 34세는 그 분포의 (대략의) **중앙값**이다. 평균 수명은 죽는 나이를 확률로 가중평균한 값(2회의 기댓값)이라 중앙값과 다르다. 분포가 대칭이 아니면, 예를 들어 노년까지 사는 소수가 있으면 평균은 중앙값과 떨어진다.

**연습 2.3** 문제 14쪽

$\frac{\ell(z)}{\ell(x)} = \frac{\ell(y)}{\ell(x)} \cdot \frac{\ell(z)}{\ell(y)}$는 대수적으로 자명하다. 확률로 읽으면 $P(A_x) = \ell(x)/1000$이고, $A_z \subset A_y \subset A_x$이므로 $A_z \cap A_x = A_z$, $A_y \cap A_x = A_y$, $A_z \cap A_y = A_z$이다. 따라서

$$P(A_z \mid A_x) = \frac{P(A_z)}{P(A_x)}, \quad P(A_y \mid A_x) = \frac{P(A_y)}{P(A_x)}, \quad P(A_z \mid A_y) = \frac{P(A_z)}{P(A_y)}$$

이고 앞의 식이 곧 $P(A_z \mid A_x) = P(A_y \mid A_x) P(A_z \mid A_y)$이다. 사건이 포함관계로 겹쳐 있다는 것이 교집합을 작은 사건 하나로 바꾸는 데 쓰였다. 일반적인 경우는 곱셈법칙(정리 5.3)이다.

## 3절

**연습 3.1** 문제 17쪽

반지름 $r = 1$로 둔다. (1) 중심각을 $\Phi$라 하면 $L = 2\sin(\Phi/2) > 1 \iff \Phi/2 \in (\pi/6, 5\pi/6)$이고 $\Phi/2 \sim \mathrm{U}(0,\pi)$이므로 확률은 $\frac{2}{3}$. (2) $2\sqrt{1-d^2} > 1 \iff d < \sqrt{3}/2$이므로 $\sqrt{3}/2 \approx 0.866$. (3) 중점이 반지름 $\sqrt{3}/2$인 원 안에 있어야 하므로 3/4.

**연습 3.2** 문제 17쪽

사분원의 넓이는 $\pi/4$이고 정사각형의 넓이는 1이므로 확률은 $\pi/4$다. 점 $n$개를 뽑아 원 안에 든 비율 $\hat{p}$를 세면 $4\hat{p}$가 $\pi$의 추정이다.

**연습 3.3** 문제 17쪽

(a) 위도가 $[-90°, 90°]$에서 고르면 60° 이북의 확률은 $30/180 = \frac{1}{6} \approx 0.167$. (b) 반지름 1인 구의 넓이는 $4\pi$이고 위도 60° 이북의 넓이는 $2\pi(1-\sin 60°)$이므로 확률은 $(1-\sqrt{3}/2)/2 \approx 0.067$. 위도·경도를 고르게 고르면 극 근처에 점이 몰린다. 극 근처에서는 같은 위도 폭에 해당하는 넓이가 작기 때문이다. "무작위 점"은 무엇에 대해 고른지를 정해야 뜻이 생긴다.

## 4절

**연습 4.1** 문제 25쪽

보조정리 0.1을 $\{A, B, C\}$에 쓰면 $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$이다. $A \cup B = A \cup (B \cap A^c)$이고, $A \cap (B \cap A^c) = \varnothing$이므로 서로소다. 같은 식으로 $B = (B \cap A) \cup (B \cap A^c)$도 서로소인 합집합이다.

**연습 4.2** 문제 25쪽

$\varnothing = \Omega^c \in \mathcal{F}$. 유한 합집합은 $A_{n+1} = A_{n+2} = \cdots = \varnothing$을 넣은 가산 합집합이다. 가산 교집합은 $\bigcap A_i = (\bigcup A_i^c)^c$(De Morgan)이고, 차집합은 $A \setminus B = A \cap B^c$이다.

**연습 4.3** 문제 25쪽

$P(AB) = P(A) + P(B) - P(A \cup B) = 0.3$. $P(A^cB^c) = P((A \cup B)^c) = 0.2$. 정확히 하나만 일어날 확률은 $P(A \cup B) - P(AB) = 0.5$.

**연습 4.4** 문제 25쪽

하트가 아닌 39장에서 5장을 뽑는 경우의 여사건이므로

$$1 - \frac{\binom{39}{5}}{\binom{52}{5}} = 1 - \frac{575{,}757}{2{,}598{,}960} \approx 0.778.$$

**연습 4.5** 문제 26쪽

$n = 2$이면 등호다. $n$개에서 성립한다고 하고 $B = A_1 \cup \cdots \cup A_n$으로 두자. 정리 4.3(f)와 귀납 가정, 그리고 Boole 부등식 $P(BA_{n+1}) = P(\bigcup_{i\le n} A_iA_{n+1}) \le \sum_{i\le n} P(A_iA_{n+1})$을 쓰면

$$P(B\cup A_{n+1}) = P(B)+P(A_{n+1})-P(BA_{n+1}) \ge \Big[\sum_{i\le n} P(A_i) - \sum_{i<j\le n} P(A_iA_j)\Big] + P(A_{n+1}) - \sum_{i\le n} P(A_iA_{n+1}),$$

이고 오른쪽이 $n+1$개에 대한 $S_1 - S_2$이다.

**연습 4.6** 문제 26쪽

하나도 맞지 않는 순서는 2143, 2341, 2413, 3142, 3412, 3421, 4123, 4312, 4321의 9가지다. 공식은 $24\big(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24}\big) = 12 - 4 + 1 = 9$.

**연습 4.7** 문제 26쪽

맞는 자리가 정확히 $I(|I| = k)$인 순서는 $I$의 자리를 고정하고 나머지 $n-k$장을 하나도 맞지 않게 놓는 것이다. Montmort의 결과로 $m$장을 하나도 맞지 않게 놓는 방법은 $D_m = m!\sum_{j=0}^{m}(-1)^j/j!$ 가지이므로

$$P(\text{정확히 } k) = \binom{n}{k}\frac{D_{n-k}}{n!} = \binom{n}{k}\frac{(n-k)!}{n!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!} = \frac{1}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!} \xrightarrow{n\to\infty} \frac{e^{-1}}{k!}.$$

**연습 4.8** 문제 26쪽

$n = 56$이면 0.9883, $n = 57$이면 0.9901이므로 $n = 57$이다. 1년이 365일인데 57명이면 거의 확실하다.

**연습 4.9** 문제 26쪽

여사건과 Boole 부등식으로 $P\big((\bigcap A_n)^c\big) = P\big(\bigcup A_n^c\big) \le \sum P(A_n^c) = 0$이다.

**연습 4.10** 문제 26쪽

$C_n = A_1 \cup \cdots \cup A_n$은 증가하고 $\bigcup_n C_n = \bigcup_n A_n$이다. 연속성(정리 4.7(a))을 $C_n$에 쓴다.

**연습 4.11** 문제 26쪽

$A_i$가 쌍마다 서로소이고 $A = \bigcup A_i$라 하자. $B_n = \bigcup_{i>n} A_i$로 두면 $B_n \downarrow \varnothing$이다(어떤 $\omega$도 모든 $B_n$에 속할 수는 없다. $\omega \in A_i$이면 $\omega \notin B_i$). $A = A_1 \cup \cdots \cup A_n \cup B_n$은 서로소인 유한 합집합이므로 유한가법성으로

$$P(A) = \sum_{i=1}^{n} P(A_i) + P(B_n).$$

$n \to \infty$에서 $P(B_n) \to 0$이므로 $P(A) = \sum_{i=1}^{\infty} P(A_i)$.

**연습 4.12** 문제 26쪽

(a) $A_1 = \Omega$, $A_2 = A_3 = \cdots = \varnothing$은 쌍마다 서로소이고 합집합이 $\Omega$이므로 (P3)에서 $P(\Omega) = P(\Omega) + \sum_{i\ge 2} P(\varnothing)$, 곧 $\sum_{i\ge 2} P(\varnothing) = 0$이다. (P1)로 $c = P(\varnothing) \ge 0$이다. $c > 0$이면 부분합 $(n-1)c$가 무한대로 가서 급수가 발산하므로 $c = 0$이다. (b) $A_{n+1} = A_{n+2} = \cdots = \varnothing$으로 두면 $A_1, A_2, \ldots$는 쌍마다 서로소이고 합집합은 $A_1 \cup \cdots \cup A_n$이다. (P3)과 (a)로 $P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) + 0$이다. (c) $A$와 $A^c$는 서로소이고 합집합이 $\Omega$이므로 (b)와 (P2)로 $P(A) + P(A^c) = 1$이다. (d) $A \subset B$이면 $B = A \cup (B \setminus A)$이고, $A \cap (B \setminus A) = \varnothing$이다. (b)로 $P(B) = P(A) + P(B \setminus A)$이고, (P1)로 $P(B \setminus A) \ge 0$이므로 $P(A) \le P(B)$이다. (e) (P1)로 $P(A) \ge 0$이고, $A \subset \Omega$에 (d)와 (P2)를 쓰면 $P(A) \le 1$이다. (f) $A \cup B = A \cup (B \setminus AB)$이고 $A \cap (B \setminus AB) = \varnothing$이다. (b)로 $P(A \cup B) = P(A) + P(B \setminus AB)$이고, $AB \subset B$에 (d)를 쓰면 $P(B \setminus AB) = P(B) - P(AB)$이다. (g) $m < n$이면 $B_m \subset A_m$이고 $B_n \cap A_m = \varnothing$이므로 $B_m \cap B_n = \varnothing$이다. $B_n \subset A_n$은 정의에서 바로 나온다. $\omega \in \bigcup_n A_n$이면 $\omega \in A_n$인 가장 작은 $n$에서 $\omega \in B_n$이므로 $\bigcup_n B_n = \bigcup_n A_n$이다. (P3)과 (d)로 $P(\bigcup_n A_n) = \sum_n P(B_n) \le \sum_n P(A_n)$이다.

## 5절

**연습 5.1** 문제 33쪽

결과 아아, 아딸, 딸아, 딸딸은 같은 확률 $\frac{1}{4}$이다. (a) $P(\text{둘 다 아들} \mid \text{적어도 하나}) = \frac{1/4}{3/4} = \frac{1}{3}$. (b) $P(\text{둘 다} \mid \text{첫째가 아들}) = \frac{1/4}{1/2} = \frac{1}{2}$. 조건이 "적어도 하나"인가 "특정한 하나"인가에 따라 답이

다르다.

### 연습 5.2 문제 33쪽

$C_i$를 자동차가 $i$번 뒤에 있는 사건, $H_3$을 사회자가 3번을 여는 사건이라 하자. $P(C_i)=\frac{1}{3}$이고, 사회자의 규칙으로 $P(H_3\mid C_1)=\frac{1}{2}$, $P(H_3\mid C_2)=1$, $P(H_3\mid C_3)=0$이다. 베이즈 정리로

$$P(C_1\mid H_3)=\frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3}+1\cdot\frac{1}{3}+0}=\frac{1}{3},\qquad P(C_2\mid H_3)=\frac{2}{3}.$$

(a) $\frac{1}{3}$, (b) $\frac{2}{3}$이므로 바꾸는 편이 낫다. (c) 사회자가 모르고 $2\cdot3$번 가운데 하나를 연다면 $E=\{$3번을 열고 염소$\}$에 대해 $P(E\mid C_1)=\frac{1}{2}$, $P(E\mid C_2)=\frac{1}{2}$, $P(E\mid C_3)=0$이므로 $P(C_1\mid E)=P(C_2\mid E)=\frac{1}{2}$이다. 같은 장면이라도 정보가 **어떻게** 만들어졌는가(가능도)가 답을 바꾼다.

### 연습 5.3 문제 34쪽

대칭으로 금과 은의 역할을 바꾸면 예제 5.6(2)와 같은 계산이므로 $\frac{2}{3}$이다.

### 연습 5.4 문제 34쪽

$P(A\mid B)>P(A)\iff P(AB)>P(A)P(B)\iff P(B\mid A)>P(B)$. 가운데 부등식이 $A$와 $B$에 대칭이다.

### 연습 5.5 문제 34쪽

사전 오즈는 $0.001/0.999$이고 양성 한 번의 가능도비는 $0.9/0.09=10$이다. 독립인 양성 $k$번 뒤의 오즈는 $\frac{0.001}{0.999}10^k$이고, 이것이 1을 넘으려면 $10^k>999$, 곧 $k=3$이다. 그때 사후확률은 $\frac{1000/999}{1+1000/999}\approx0.5003$이다($k=1$: 0.0099, $k=2$: 0.091).

### 연습 5.6 문제 34쪽

분배법칙과 정리 4.3(f), 상호 독립으로

$$P\big(A\cap(B\cup C)\big)=P(AB)+P(AC)-P(ABC)=P(A)\big[P(B)+P(C)-P(B)P(C)\big]=P(A)P(B\cup C).$$

마지막 등호에서 $P(B\cup C)=P(B)+P(C)-P(BC)$와 $P(BC)=P(B)P(C)$를 썼다.

### 연습 5.7 문제 34쪽

$P(A)=0$이면 $0\le P(AB)\le P(A)=0=P(A)P(B)$. $P(A)=1$이면 $P(A^cB)\le P(A^c)=0$이므로 $P(AB)=P(B)-P(A^cB)=P(B)=P(A)P(B)$.

### 연습 5.8 문제 34쪽

첨자 $i_1 < \cdots < i_k$를 고정하면 $H_{i_1}\cdots H_{i_k}$는 그 $k$자리가 A인 결과들이다. 나머지 $m-k$자리에서 A가 $j$번 나오는 결과는 $\binom{m-k}{j}$개이고 각각의 확률은 $p^{k+j}q^{m-k-j}$이므로

$$P(H_{i_1}\cdots H_{i_k}) = p^k \sum_{j=0}^{m-k} \binom{m-k}{j} p^j q^{m-k-j} = p^k (p+q)^{m-k} = p^k.$$

$k=1$이면 $P(H_i)=p$이므로 이것은 $P(H_{i_1})\cdots P(H_{i_k})$와 같다.

### 연습 5.9 문제 34쪽

(a) 작은 결석: A 81/87 = 0.931, B 234/270 = 0.867. 큰 결석: A 192/263 = 0.730, B 55/80 = 0.688. 두 경우 모두 A가 낫다. (b) 전체: A 273/350 = 0.780, B 289/350 = 0.826로 B가 낫다. (c) 전확률 공식으로 $P(\text{성공}\mid \text{A}) = \sum_s P(\text{성공}\mid \text{A}, s)\,P(s\mid \text{A})$이다. A는 환자의 263/350 = 75%가 치료가 어려운 큰 결석이었고, B는 270/350 = 77%가 작은 결석이었다. 전체 성공률은 크기별 성공률을 **서로 다른 가중치**로 평균한 것이어서 순서가 뒤집힌다. 결석 크기는 치료 선택과 결과 모두에 영향을 주는 **교란변수**다.

### 연습 5.10 문제 34쪽

(a) (P1): $Q(A) = P(AB)/P(B)$에서 분자는 (P1)로 0 이상이고 분모는 양수다. (P2): $\Omega \cap B = B$이므로 $Q(\Omega) = P(B)/P(B) = 1$이다. (b) $i \neq j$이면 $(A_iB)\cap(A_jB) = (A_i\cap A_j)\cap B = \varnothing$이다. $\omega \in (\bigcup_i A_i)\cap B \iff$ 어떤 $i$에서 $\omega\in A_i$이고 $\omega\in B \iff$ 어떤 $i$에서 $\omega \in A_iB \iff \omega \in \bigcup_i (A_iB)$이다. (c) (b)와 $P$의 (P3)으로

$$Q\Big(\bigcup_i A_i\Big) = \frac{P\big(\bigcup_i (A_iB)\big)}{P(B)} = \frac{\sum_i P(A_iB)}{P(B)} = \sum_i \frac{P(A_iB)}{P(B)} = \sum_i Q(A_i).$$

셋째 등호에서 수렴하는 급수에 상수를 곱해도 된다는 것($\sum_i ca_i = c\sum_i a_i$, 부분합에 $c$를 곱하고 극한을 취한다)을 썼다.

### 연습 5.11 문제 35쪽

(a) $A_1\cdots A_k \supset A_1\cdots A_{n-1}$ $(k \le n-1)$이므로 단조성으로 $P(A_1\cdots A_k) \ge P(A_1\cdots A_{n-1}) > 0$이다. 따라서 오른쪽의 모든 조건부확률의 조건이 양의 확률을 갖는다. (b) $n=2$이면 오른쪽은 $P(A_1)\cdot P(A_1A_2)/P(A_1) = P(A_1A_2)$이다. (c) $n$에서 성립하고 $P(A_1\cdots A_n)>0$이라 하자. $C = A_1\cdots A_n$이면 정의로 $P(C\cap A_{n+1}) = P(C)\,P(A_{n+1}\mid C)$이고, $P(C)$에 귀납 가정을 넣으면 $n+1$에서의 식이 된다. (d) 소거 증명의 "…"는 "이 소거가 $n$이 몇이든 끝까지 이어진다"는 주장이다. (c)의 귀납법이 그 주장을 증명한 것이다. 또 소거 증명은 분모가 모두 0이 아님을 쓰는데, 그것이 (a)다.

## 6절

**연습 6.1** 문제 40쪽

$F(\ell) = \ell^2/4 (0 \le \ell \le 2)$이므로 $P(0.5 < L \le 1.5) = (2.25 - 0.25)/4 = 0.5$. 중앙값은 $m^2/4 = \frac{1}{2}$에서 $m = \sqrt{2} \approx 1.414$. 밀도는 $f(\ell) = \ell/2 (0 < \ell < 2)$.

**연습 6.2** 문제 40쪽

서로 다른 점 $x_1, \dots, x_k$에서의 점프는 서로소인 사건 $\{X = x_i\}$의 확률이므로 합이 1을 넘지 않는다. 따라서 높이가 $1/n$보다 큰 점프는 $n$개보다 적다. 불연속점의 집합은 $n = 1, 2, \dots$에 대한 이 유한집합들의 합집합이므로 기껏해야 셀 수 있다.

**연습 6.3** 문제 41쪽

$P(X = 0) = 0.3$(점프의 높이). $P(X \le 1) = 0.3 + 0.7(1 - e^{-1}) \approx 0.742$. $P(0 < X \le 1) = F(1) - F(0) = 0.7(1 - e^{-1}) \approx 0.442$.

**연습 6.4** 문제 41쪽

$X \ge 40$은 "40세에 살아 있다"와 같으므로 $P(X \ge 40) = \ell(40)/1000 = 0.445$. $P(20 \le X < 40) = (\ell(20) - \ell(40))/1000 = (598 - 445)/1000 = 0.153$.

**연습 6.5** 문제 41쪽

$F(m) = \frac{1}{2}$을 풀면 (1) $\frac{2}{\pi}\arcsin\frac{m}{2} = \frac{1}{2}$에서 $m = 2\sin\frac{\pi}{4} = \sqrt{2}$, (2) $\sqrt{1 - m^2/4} = \frac{1}{2}$에서 $m = \sqrt{3}$, (3) $m = \sqrt{2}$이다. (1)의 밀도는 짧은 현에서도 $1/\pi$의 높이가 있고 길이 2 근처에서 무한대로 치솟는 반면, (3)의 밀도는 0에서 출발해 직선으로 올라간다. 중앙값은 분포의 한 점만 요약하므로 모양의 차이를 보지 못한다.

**연습 6.6** 문제 41쪽

$t \mapsto t^n$은 $[0,1]$에서 증가하고 연속이다. 따라서 $G = F^n$은 비감소이고, $\lim_{x\to-\infty} G = 0^n = 0$, $\lim_{x\to\infty} G = 1$이며, $F$가 오른쪽 연속이므로 $G$도 오른쪽 연속이다.

**연습 6.7** 문제 41쪽

$0 \le y \le 1$에서 $F_Y(y) = P(X^2 \le y) = P(X \le \sqrt{y}) = \sqrt{y}$이고, $y < 0$이면 0, $y > 1$이면 1이다. 미분하면 $f_Y(y) = \frac{1}{2\sqrt{y}} (0 < y < 1)$. $y \to 0$에서 밀도가 무한대로 가지만 적분은 1이다.

### 연습 6.8 문제 41쪽

(a) 이상적분은 $\int_{-\infty}^{\infty} f = \lim_{b\to\infty}\int_{-\infty}^{b} f = \lim_{b\to\infty} F(b)$이고, 정리 6.3(b)로 이 극한은 1이다. (b) $h>0$이면 $F(x)-F(x-h)=\int_{x-h}^{x} f(t)\,dt \geq 0$이다. $\int_{-\infty}^{x} f = F(x)$가 유한하므로 이상적분의 정의로 $\int_{-\infty}^{x-h} f \to \int_{-\infty}^{x} f$ $(h\downarrow 0)$, 곧 $F(x-h)\to F(x)$이다. 따라서 $F(x-)=F(x)$이고, 정리 6.3(d)로 $P(X=x)=F(x)-F(x-)=0$이다. (c) $P(a<X\leq b)=F(b)-F(a)=\int_a^b f$이다. $\{a\leq X\leq b\}=\{a<X\leq b\}\cup\{X=a\}$이고 (b)로 $P(X=a)=0$이므로 확률이 같다. 나머지 두 구간도 끝점 하나를 빼거나 더한 것이므로 같다. (d) $h\neq 0$이면 $F(x+h)-F(x)=\int_x^{x+h} f(t)\,dt$이고 $hf(x)=\int_x^{x+h} f(x)\,dt$이므로

$$\left|\frac{F(x+h)-F(x)}{h}-f(x)\right| = \left|\frac{1}{h}\int_x^{x+h}\big(f(t)-f(x)\big)dt\right| \leq \sup_{|t-x|\leq|h|}|f(t)-f(x)|.$$

$f$가 $x$에서 연속이면 오른쪽은 $h\to 0$에서 0으로 가므로 $F'(x)=f(x)$이다.

## 7절

### 연습 7.1 문제 44쪽

(a) $\mathbb{Z}$는 셀 수 있으므로 정리 7.1의 증명이 그대로 된다. (b) $[n,n+1)(n=0,1,2,\dots)$은 $[0,\infty)$의 셀 수 있는 분할이다. 모두 같은 확률 $c$라면 $1=\sum_n c$인데, 이는 $c=0$이면 0, $c>0$이면 무한대다.

### 연습 7.2 문제 44쪽

(a) $\{1,\dots,n\}$의 짝수는 $\lfloor n/2\rfloor$개이므로 비율이 $\frac{1}{2}$로, 3의 배수는 $\lfloor n/3\rfloor$개이므로 $\frac{1}{3}$로 간다. (b) $A,B$가 서로소이면 $|(A\cup B)\cap\{1..n\}| = |A\cap\{1..n\}|+|B\cap\{1..n\}|$이므로 극한이 있으면 더해진다. 그러나 한 점 $\{k\}$의 밀도는 $1/n\to 0$이고 $\mathbb{N}=\bigcup_k\{k\}$의 밀도는 1이므로 $1\neq\sum_k 0$이다.

## 8절

### 연습 8.1 문제 46쪽

$\frac{1}{2}-p\approx 0.0086$이다. $2\sqrt{p(1-p)/n}<0.0086$을 풀면 $n>4p(1-p)/0.0086^2\approx 1.35\times 10^4$이다. de Méré가 경험으로 차이를 알아챘다면 수천 판 이상을 기억에 담았다는 뜻이다.

### 연습 8.2 문제 46쪽

"무료"만 있을 때 사후 오즈는 $\frac{0.4}{0.6}\cdot\frac{0.3}{0.02}=10$이므로 확률은 $10/11\approx 0.909$. "회의"가 더해지면 가능도비 $0.01/0.1=0.1$이 곱해져 오즈가 1, 확률이 0.5가 된다.