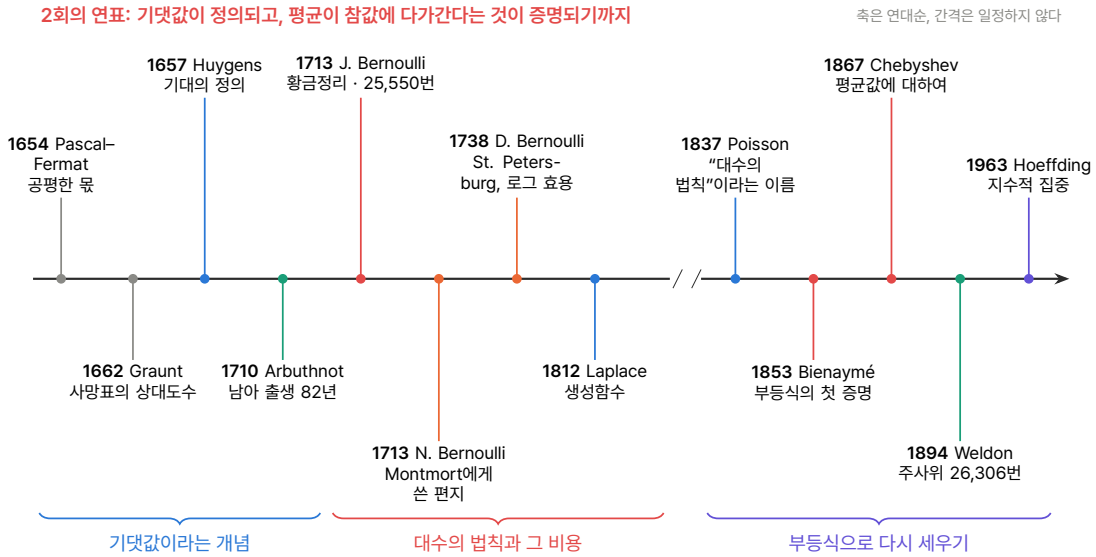
수리통계학 × 통계학의 역사

2회 강의노트

기댓값과 대수의 법칙 몇 번 던져야 아는가

1657 → 1867

Huygens는 확률보다 먼저 “기대”를 정의했고, Jacob Bernoulli는 20년을 들여 관측한 비율이 참값에 다가간다는 것을 증명했다. 그러나 그가 얻은 시행 횟수 25,550은 실제로 필요한 수의 네 배였다. 그 차이를 메운 것은 150년 뒤 Bienaymé와 Chebyshev의 부등식, 그리고 다시 100년 뒤의 지수적 한계였다. 이 노트는 기댓값, 분산과 적률, 적률생성함수, 변수변환을 처음부터 세우고, Markov·Chebyshev·Jensen 부등식으로 대수의 법칙을 증명한다.



다루는 것	기댓값 · 꼬리 공식 · 분산과 적률 · 적률생성함수 · 변수변환 · 확률적분변환 · Markov · Chebyshev · Jensen · 대수의 법칙 · Hoeffding
원자료	Weldon(1894)의 주사위 26,306번 · Arbuthnot(1710)의 런던 세례 기록 82년 · Halley(1693)의 생명표
노트북	`sessions/02_expectation/02_expectation.ipynb` — 절·정리 번호가 같다
워크북	`02_expectation_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf`
선수 지식	1회 전체 (확률의 공리, 독립, 확률변수와 분포함수), 1학년 미적분 (급수, 이상적분, 테일러 급수)

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 ①②③으로 쓰고, 단계마다 오른쪽 끝에 근거(앞 정리의 번호나 공리)를 단다. 극한·합·적분의 순서를 바꾸는 곳에는 **조건 확인** 단락을 둔다. 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. 이 회차에서는 [T1], [T2], [T4], [T5], [T6](합과 기댓값의 순서 바꾸기), [T7](변수변환), [T8]을 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 1회에서 가져오는 것 2

0.2 절대수렴 3

1 Huygens의 기대 — 기댓값 5

1.1 공평한 내기의 값 5

1.2 이산 확률변수의 기댓값 5

1.3 연속 확률변수의 기댓값 7

1.4 함수의 기댓값과 선형성 8

1.5 꼬리 공식 10

2 분산과 적률 — 흩어짐을 재는 수 14

2.1 분산 14

2.2 이항분포의 평균과 분산 15

2.3 독립인 이산 확률변수 16

2.4 적률 18

3 적률생성함수 21

3.1 정의와 예 21

3.2 mgf에서 적률로 22

3.3 변환과 독립합, 그리고 유일성 23

4 변수변환 — 확률변수의 함수의 분포 26

4.1 분포함수 방법 26

4.2 단조변환의 밀도 27

4.3 확률적분변환과 역변환 표집 28

5 부등식 — 기댓값으로 꼬리를 누르기 32

5.1 Markov와 Chebyshev . . . 32

5.2 Jensen . . . 34

6 대수의 법칙 — Bernoulli의 황금정리 39

6.1 Ars Conjectandi . . . 39

6.2 확률수렴과 Bernoulli의 정리 . . . 39

6.3 Bernoulli의 계산 . . . 40

6.4 같은 질문의 다섯 답 . . . 41

6.5 일반적인 대수의 법칙 . . . 42

6.6 상대도수와 확률 — 다시 데이터로 . . . 43

7 부숴보기 — 가정을 빼면 45

7.1 기댓값이 무한대 — St. Petersburg 역설 . . . 45

7.2 기댓값이 없다 — Cauchy . . . 46

7.3 적률이 분포를 정하지 못한다 . . . 46

7.4 독립이 아니면 . . . 47

8 Beautiful insights — 같은 도구, 오늘의 문제 50

8.1 지수적 집중: Chernoff와 Hoeffding . . . 50

8.2 Jensen과 정보 . . . 52

8.3 기댓값으로서의 손실, 그리고 Kelly . . . 52

9 정리와 역사 노트 54

9.1 한 장 요약 . . . 54

9.2 Stigler의 명명 법칙 . . . 55

9.3 7 Pillars 지도 . . . 55

9.4 다음 회 — 왜 평균을 내는가: $\sqrt{n}$의 발견 (1755 → 1958) . . . 55

A 연습문제 풀이 57

0

들어가며

> **여는 질문.** 항아리에 흰 돌 3,000개와 검은 돌 2,000개가 들어 있다. 우리는 그 수를 모르고, 돌을 하나 꺼내 색을 보고 다시 넣기를 되풀이할 수 있을 뿐이다. 몇 번 꺼내면 "흰 돌의 비율이 $\frac{29}{50}$과 $\frac{31}{50}$ 사이에 든다"는 쪽에 1,000 대 1로 걸 수 있는가?

Jacob Bernoulli가 *Ars Conjectandi*(추측의 기술)의 마지막 장에서 스스로 던지고 스스로 답한 질문이다. 그는 이 질문의 답이 유한하다는 것, 곧 **시행을 충분히 많이 하면 관측한 비율이 참 비율에 얼마든지 가까워진다**는 것을 증명하는 데 20년을 썼다. 그리고 실제로 수를 넣어 25,550번이라는 답을 얻었다. 책은 그 계산 바로 뒤에서 끝난다. Bernoulli는 1705년에 죽었고 책은 1713년에야 조카 Nicolaus의 손으로 출판되었다.

오늘의 답은 약 6,500번이다(6절). Bernoulli의 증명은 옳았지만 어림이 거칠었다. 그 차이를 메운 도구가 이 노트의 중심이다. 확률변수의 **기댓값**과 **분산**, 그리고 그 둘만으로 꼬리 확률을 누르는 **부등식**이다. 1853년 Bienaymé와 1867년 Chebyshev는 이 부등식으로 Bernoulli의 정리를 몇 줄로 다시 증명했다. 이 짧은 증명이 대수의 법칙의 표준적인 증명이다.

기댓값 자체는 더 오래되었다. 1657년 Huygens는 "확률"이라는 말을 쓰기 전에 도박의 "기대"(expectatio)를 정의했다(1절). 이 노트는 그 기대를 확률변수의 기댓값으로 세우고(1절), 분산과 적률(2절), 적률을 한 함수에 담는 적률생성함수(3절), 확률변수의 함수의 분포(4절)를 차례로 다룬 뒤, 부등식(5절)으로 대수의 법칙(6절)을 증명한다. 7절은 기댓값이 없거나 적률이 분포를 정하지 못하는 경우를 보고, 8절은 같은 도구가 오늘의 학습 이론과 정보 이론에서 어떻게 쓰이는지 본다.

학습 목표

1. 이산·연속 확률변수의 기댓값을 정의하고, 존재 조건(절대수렴)이 왜 필요한지 설명한다.
2. 함수의 기댓값, 선형성, 꼬리 공식을 증명하고 지시변수로 기댓값을 계산한다.
3. 분산과 적률, 독립인 이산 확률변수의 합의 분산을 증명하고 이항분포의 평균과 분산을 구한다.
4. 적률생성함수로 적률을 구하고, 독립합의 분포를 적률생성함수의 곱으로 찾는다.
5. 단조변환의 밀도 공식과 확률적분변환을 증명하고, 역변환 표집을 구현한다.

6. Markov·Chebyshev·Jensen 부등식을 증명하고, 대수의 법칙을 Chebyshev로 증명한다.
7. Bernoulli의 질문에 다섯 가지로 답하고(Bernoulli · Chebyshev · Hoeffding · 정규근사 · 정확한 계산), 차이가 어디서 오는지 설명한다.

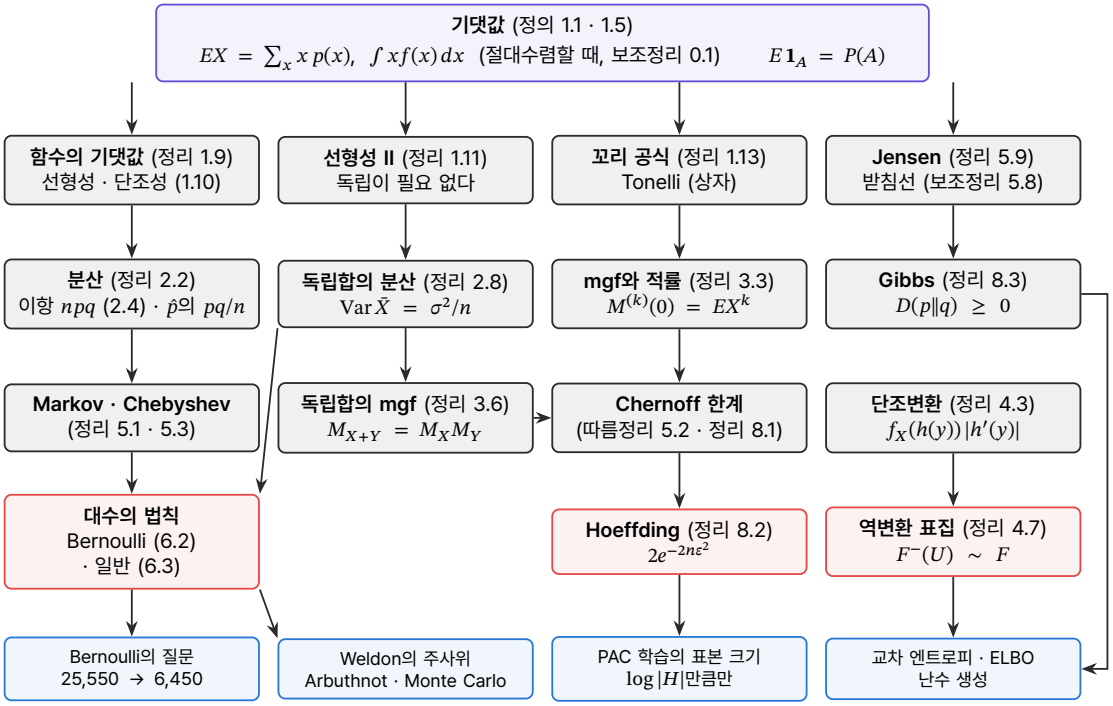


그림 0.1 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

0.1 1회에서 가져오는 것

이 노트는 1회의 정의와 정리 위에 선다. 쓰는 것을 모아 둔다. 번호는 1회 노트의 번호다.

1회	내용
정의 4.2 · 정리 4.3	확률의 공리, 여사건 $P(A^c) = 1 - P(A)$, 단조성, Boole 부등식 $P(\bigcup A_i) \le \sum P(A_i)$
정리 4.7	연속성: $A_n \uparrow A$이면 $P(A_n) \uparrow P(A)$, $A_n \downarrow A$이면 $P(A_n) \downarrow P(A)$
정리 4.9	이산 확률공간: $\Omega = \{\omega_1, \omega_2, \dots\}$, $P(A) = \sum_{\omega_i \in A} p_i$
정의 5.7 · 5.9	두 사건의 독립 $P(AB) = P(A)P(B)$, 여러 사건의 상호 독립
정의 6.1 · 6.2 · 정리 6.3	확률변수, 분포함수 $F(x) = P(X \le x)$와 그 성질 (비감소, 양 끝, 오른쪽 연속)
정의 6.4 · 6.6 · 정리 6.7	확률질량함수 $p(x)$, 확률밀도함수 $f(x)$, $P(a < X \le b) = \int_a^b f$
정리 1.6	이항정리 $(a+b)^n = \sum_k \binom{n}{k} a^k b^{n-k}$
상자 (4.6절)	음이 아닌 급수는 순서를 바꾸거나 묶어서 더해도 값이 같다

0.2 절대수렴

기댓값은 "값 × 확률"의 합이다. 값이 무한히 많으면 이 합은 급수이고, 급수는 더하는 순서에 따라 값이 달라질 수 있다. 확률변수의 값에는 정해진 순서가 없으므로, 순서에 따라 달라지는 합은 기댓값이 될 자격이 없다. 순서에 상관없는 급수가 바로 절대수렴하는 급수다.

> **보조정리 0.1** (절대수렴하는 급수의 재배열) $\sum_i |a_i| < \infty$이면 $\sum_i a_i$는 항의 순서를 어떻게 바꾸거나 어떻게 묶어 더해도 같은 값으로 수렴한다.

증명. **전략** 항을 양의 부분과 음의 부분으로 쪼갠다. 두 부분은 음이 아닌 급수이므로 1회 4.6절의 상자를 쓸 수 있다 [T5]. 단계별 증명은 연습 0.2(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 57쪽에 있다.

절대수렴하지 않으면 사정이 다르다. $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$는 $\log 2$로 수렴하지만, 양수 항 둘에 음수 항 하나씩 더하는 순서로 바꾸면 $\frac{3}{2}\log 2$로 수렴한다. Riemann은 조건수렴하는 급수를 재배열하면 **어떤** 실수로도 수렴하게 만들 수 있음을 보였다(연습 0.1). 그래서 기댓값의 정의에는 언제나 "절대수렴하면"이라는 조건이 붙는다. 적분도 같다. 이상적분 $\int_{-\infty}^{\infty} g$는 $\int |g| < \infty$일 때만 자르는 방법에 상관없는 값을 갖는다.

연습문제

연습 0.1 (재배열하면 값이 바뀐다 ★★) $s = 1 - \frac{1}{2} + \frac{1}{3} - \cdots = \log 2$를 받아들이고, 양수 항 둘 다음에 음수 항 하나를 놓는 재배열 $1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \cdots$의 합이 $\frac{3}{2}\log 2$임을 보여라. (힌트: $\frac{1}{2}s = \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \cdots$에 0을 끼워 s와 더한다.)

풀이 57쪽

연습 0.2 (절대수렴하는 급수의 재배열: 증명 ★★ 유도형) 보조정리 0.1를 다음 순서로 증명하라.

(a) $a_i^+ = \max(a_i, 0)$, $a_i^- = \max(-a_i, 0)$으로 두면 $a_i = a_i^+ - a_i^-$, $|a_i| = a_i^+ + a_i^-$, $0 \le a_i^{\pm} \le |a_i|$임을 확인하라.

(b) $\sum_i a_i^+$와 $\sum_i a_i^-$가 유한함을 보여라. 음이 아닌 급수의 부분합이 위로 유계이면 수렴한다는 것을 어디서 쓰는가?

(c) 순서를 바꾼 급수 $\sum_k a_{\sigma(k)}$의 부분합이 $\sum_{k\le K} a^+_{\sigma(k)} - \sum_{k\le K} a^-_{\sigma(k)}$임을 쓰고, 1회 4.6절의 상자로 $\sum_k a_{\sigma(k)} = \sum_i a_i^+ - \sum_i a_i^-$를 결론지어라. 묶어 더하는 경우도 같은 논리임을 설명하라.

풀이 57쪽

1

Huygens의 기대 — 기댓값

1회의 점수 분배 문제에서 Pascal과 Fermat가 구한 것은 사실 확률이 아니라 **몫**이었다. 판돈 64 가운데 A의 공평한 몫은 56이다. 몫은 "얻을 수 있는 금액"과 "그것을 얻을 확률"을 함께 담은 수다. 이 수를 처음으로 정의하고 계산 규칙을 세운 사람이 Huygens였다. 이 절은 그 정의를 확률변수의 언어로 옮긴다.

1.1 공평한 내기의 값

> **역사** Christiaan Huygens는 1655년 파리를 방문해 Pascal과 Fermat의 문제를 전해 듣고, 1657년 *De ratiociniis in ludo aleae*(주사위 놀이의 계산에 대하여)를 냈다. 열네 개의 명제로 된 이 짧은 책이 최초의 확률 교과서이고, 이후 50년 동안 유일한 교과서였다. 책의 출발점은 확률이 아니라 "기대"다. 명제 I은 이렇다. "a나 b를 같은 정도로 쉽게 얻을 수 있다면, 그 기대는 $\frac{a+b}{2}$의 값어치가 있다." 라틴어 번역자 van Schooten이 이 값을 *expectatio*라 불렀다.

Huygens의 논증은 확률을 쓰지 않는다. **공평한 내기**만 쓴다. 나와 상대가 각각 $\frac{a+b}{2}$씩 걸고(판돈 $a+b$), 이긴 쪽이 진 쪽에게 a를 주기로 하고 공평한 동전을 던진다. 내가 지면 a를 받고, 이기면 판돈에서 a를 주고 남은 b를 갖는다. 그러니 $\frac{a+b}{2}$를 내고 이 내기에 들어가는 것은 "a나 b를 같은 정도로 얻는 처지"를 사는 것과 같고, 그 처지의 값은 낸 돈 $\frac{a+b}{2}$다. 같은 논증으로 명제 III을 얻는다. a를 얻을 경우가 p가지, b를 얻을 경우가 q가지이면 기대는 $\frac{pa+qb}{p+q}$다(연습 1.2).

오늘의 말로 하면 "값 × 확률의 합"이다. 기댓값은 이렇게 확률보다 먼저, 공평함이라는 직관에서 나왔다.

1.2 이산 확률변수의 기댓값

정의 1.1 (이산 확률변수의 기댓값) 이산 확률변수 X가 값 $x_1, x_2, \dots$를 확률질량함수 p로 가질 때

$$\sum_{i} |x_i|\, p(x_i) < \infty$$

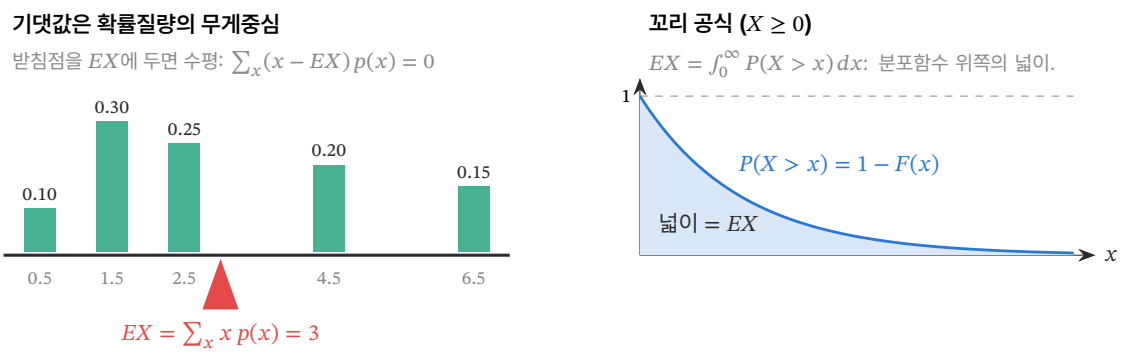


그림 1.1 왼쪽: 기댓값은 확률질량의 무게중심이다. 오른쪽: 음이 아닌 확률변수의 기댓값은 $P(X>x)$ 곡선 아래의 넓이다(정리 1.13).

이면 X의 **기댓값**(평균)을

$$EX=\sum_i x_i\,p(x_i)$$

로 정의한다. 절대수렴하지 않으면 EX는 존재하지 않는다고 한다. 다만 $X\ge 0$이고 합이 발산하면 $EX=\infty$로 쓴다.

절대수렴 조건 때문에 합의 순서는 상관없다(보조정리 0.1). EX를 μ나 μ_X로도 쓴다. 그림 1.1의 왼쪽처럼, 막대 위의 점 x_i에 질량 $p(x_i)$를 놓으면 EX는 막대가 수평을 이루는 받침점이다. $\sum_i (x_i-EX)\,p(x_i)=EX-EX=0$이기 때문이다.

예제 1.2 (주사위 · 지시변수 · de Méré의 기대 이익)

주사위. X가 공평한 주사위의 눈이면 $EX=\frac{1}{6}(1+2+\cdots+6)=3.5$이다. 기댓값은 X가 가질 수 없는 값일 수 있다.

지시변수. 사건 A의 **지시변수** $\mathbf{1}_A$는 A가 일어나면 1, 아니면 0인 확률변수다. 값이 0과 1뿐이므로

$$E\,\mathbf{1}_A=1\cdot P(A)+0\cdot P(A^c)=P(A).$$

확률은 지시변수의 기댓값이다. 이 한 줄이 기댓값과 확률을 잇는 다리이고, 5절과 6절의 증명은 모두 이 다리를 건넌다.

de Méré의 내기. 1회 예제 1.9의 내기 1에 1을 걸어 이기면 1을 받고 지면 1을 잃는다고 하자. 이익 G는 확률 $\frac{671}{1296}$로 $+1$, 나머지 확률로 -1이므로 $EG=2\cdot\frac{671}{1296}-1\approx 0.0355$이다. 판마다 판돈의 3.5%를 기대할 수 있다. 내기 2는 $2\cdot 0.4914-1\approx -0.0172$로 불리하다. 두 내기의 차이는 확률로는 0.026이지만 기대 이익으로는 부호가 갈린다.

예제 1.3 (점수 분배는 기댓값이다) 1회 예제 1.11의 상태 $(a,b)=(1,3)$에서 A가 받을 금액 W는 내기를 계속하면 확률 $\frac{7}{8}$로 64, $\frac{1}{8}$로 0이다. $EW = 64\cdot\frac{7}{8} = 56$이 Pascal이 정한 A의 몫이다. Huygens의 말로는 "56의 기대"이고, 오늘의 말로는 "공평한 몫 = 받을 금액의 기댓값"이다.

예제 1.4 (처음 성공할 때까지의 시행 수) 성공 확률이 p인 독립 시행을 처음 성공할 때까지 할 때 시행 수를 X라 하자. 처음 $k-1$번 실패하고 k번째에 성공해야 하므로 1회 정리 5.11로

$$P(X=k) = q^{k-1}p, \qquad k=1,2,\dots, \quad q=1-p.$$

이 분포를 **기하분포**라 한다. $S=\sum_{k\ge1} kq^{k-1}$은 비판정법으로 수렴하고(항의 비가 $\frac{k+1}{k}q \to q<1$),

$$S - qS = \sum_{k\ge1} kq^{k-1} - \sum_{k\ge1} kq^{k} = \sum_{k\ge1} q^{k-1} = \frac{1}{1-q}$$

이므로 $S = 1/(1-q)^2 = 1/p^2$이고 $EX = pS = 1/p$이다. 6이 나올 때까지는 평균 6번, 6이 두 개 나올 때까지는 평균 36번 던진다.

1회의 de Méré는 "반 이상의 확률로 성공하는 횟수"(중앙값)를 따졌다. 그 수는 $\log 2/p \approx 0.693/p$ 근처이고 평균 $1/p$보다 작다. 기하분포는 오른쪽 꼬리가 길어서, 드물게 아주 오래 기다리는 경우가 평균을 끌어올린다. 평균과 중앙값은 다른 질문의 답이다.

1.3 연속 확률변수의 기댓값

정의 1.5 (연속 확률변수의 기댓값) 연속 확률변수 X의 밀도가 f이고 $\int_{-\infty}^{\infty} |x|\,f(x)\,dx < \infty$이면

$$EX = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

로 정의한다. $X\ge 0$이고 적분이 발산하면 $EX=\infty$로 쓴다.

이산의 합을 적분으로 바꾼 것이다. 작은 h에 대해 $P(x<X\le x+h)\approx f(x)h$이므로, 실수축을 폭 h의 칸으로 나눈 이산 근사의 기댓값 $\sum x\,f(x)h$가 $h\to 0$에서 이 적분이 된다.

예제 1.6 (균등분포와 지수분포)

균등분포. $X\sim \mathrm{U}(a,b)$, 곧 $f(x)=\frac{1}{b-a}$ $(a<x<b)$이면 $EX = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2}$. 구간의 가운데다.

지수분포. $\lambda>0$에 대해 밀도 $f(x)=\lambda e^{-\lambda x}$ $(x>0)$를 갖는 분포를 **지수분포** $\mathrm{Exp}(\lambda)$라 한다.

분포함수는 $F(x) = 1 - e^{-\lambda x}$이다. 부분적분으로

$$EX = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \Big[-xe^{-\lambda x}\Big]_0^\infty + \int_0^\infty e^{-\lambda x}\,dx = 0 + \frac{1}{\lambda}.$$

경계항이 0인 것은 $xe^{-\lambda x} \to 0$ $(x \to \infty)$이기 때문이다. 지수분포는 기하분포의 연속판이다. 기다리는 시간의 평균이 "단위시간당 성공률"의 역수다.

예제 1.7 (Bertrand의 세 현, 세 평균) 1회 예제 6.8의 세 방법에서 현의 길이 L의 평균을 구한다($r = 1$).

(1) $L = 2\sin\Psi$, $\Psi \sim \mathrm{U}(0,\pi)$이므로 (아래 정리 1.9를 미리 쓰면) $EL = \frac{1}{\pi}\int_0^\pi 2\sin\psi\,d\psi = \frac{4}{\pi} \approx 1.273$.

(2) $L = 2\sqrt{1-D^2}$, $D \sim \mathrm{U}(0,1)$이므로 $EL = 2\int_0^1 \sqrt{1-d^2}\,dd = 2\cdot\frac{\pi}{4} = \frac{\pi}{2} \approx 1.571$. 적분은 사분원의 넓이다.

(3) $f_3(\ell) = \frac{\ell}{2}$ $(0 < \ell < 2)$이므로 $EL = \int_0^2 \frac{\ell^2}{2}\,d\ell = \frac{4}{3} \approx 1.333$.

1회에서 "무작위 현"이 세 개의 서로 다른 분포였으므로 평균도 셋이다. 길이의 평균만 보면 (2)가 가장 긴 현을 만든다.

예제 1.8 (기댓값이 없는 분포 — Cauchy) 밀도 $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$를 갖는 분포를 **Cauchy 분포**라 한다. $\int f = \frac{1}{\pi}[\arctan x]_{-\infty}^{\infty} = 1$이므로 밀도이고, 0에 대해 대칭이다. 그런데

$$\int_{-\infty}^\infty |x|\,f(x)\,dx = \frac{2}{\pi}\int_0^\infty \frac{x}{1+x^2}\,dx = \frac{1}{\pi}\Big[\log(1+x^2)\Big]_0^\infty = \infty$$

이므로 기댓값이 없다. 대칭이니 "평균은 0"이라고 말하고 싶지만, $\int_{-A}^{2A} x f(x)\,dx \to \frac{1}{\pi}\log 2$ $(A \to \infty)$처럼 자르는 방법에 따라 극한이 달라진다. 0이라는 답은 우리가 자르는 방법을 골랐기 때문에 나온 것이다. 7절에서 Cauchy 표본의 평균이 대수의 법칙을 따르지 않는 것을 본다.

1.4 함수의 기댓값과 선형성

X의 기댓값을 알면 X^2, e^{tX}, $(X-\mu)^2$ 같은 $g(X)$의 기댓값이 필요해진다. $Y = g(X)$의 분포를 먼저 구하지 않고 X의 분포로 바로 계산할 수 있다.

정리 1.9 (함수의 기댓값)

(a) X가 이산이고 pmf가 p이면, $\sum_i |g(x_i)|\,p(x_i) < \infty$일 때 $Y = g(X)$의 기댓값이 존재하고

$$E\,g(X) = \sum_i g(x_i)\,p(x_i).$$

(b) X가 연속이고 밀도가 f이면, $\int |g(x)| f(x)\,dx < \infty$일 때

$$E\,g(X) = \int_{-\infty}^{\infty} g(x)\,f(x)\,dx.$$

증명. **전략** (a) $Y = g(X)$의 값 y_j마다 그 값으로 가는 x_i들을 한 묶음으로 모은다. 먼저 절댓값으로 합이 유한함을 보이고, 그다음 같은 묶음을 절댓값 없이 한다 [T5]. (b)는 g가 미분가능한 단조함수인 경우를 4절의 따름정리 4.4에서 증명한다. 일반적인 경우는 아래 상자의 사실이다. 단계별 증명은 연습 1.9(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 59쪽에 있다.

증명 없이 쓰는 사실 (연속 확률변수의 함수의 기댓값) X가 연속이어도 $Y = g(X)$는 이산일 수도(예: $g(x) = \mathbf{1}\{x > 0\}$), 혼합형일 수도(예: $g(x) = \max(x, 0)$) 있다. 그런 Y까지 포함해 "기댓값"을 하나로 정의하려면 측도론의 Lebesgue 적분이 필요하고, 그 이론에서 (b)는 정의를 바꿔 쓴 것에 지나지 않는다. 이 과정에서는 (b)를 함수의 기댓값을 계산하는 규칙으로 쓴다. 증명은 Billingsley, *Probability and Measure*, 16–17절.

정리 1.10 (기댓값의 선형성과 단조성 I) 기댓값이 존재하는 $g(X)$, $h(X)$와 상수 a, b에 대해

(a) $E[a\,g(X) + b\,h(X)] = a\,E\,g(X) + b\,E\,h(X)$. 특히 $E(aX + b) = a\,EX + b$.

(b) 모든 x에서 $g(x) \le h(x)$이면 $E\,g(X) \le E\,h(X)$.

(c) $|EX| \le E|X|$.

증명. 정리 1.9로 기댓값이 p나 f에 대한 합·적분이 되므로, (a)는 절대수렴하는 합과 적분의 선형성이다($|ag + bh| \le |a||g| + |b||h|$이므로 좌변도 절대수렴한다). (b)는 $p, f \ge 0$이므로 합과 적분이 순서를 보존한다는 것이다. (c)는 (b)를 $-|x| \le x \le |x|$에 쓴 것이다. ∎

선형성의 진짜 힘은 **서로 다른** 확률변수의 합에서 나온다. 같은 실험에서 정의된 X와 Y가 있으면 $E(X + Y) = EX + EY$인가? 표본공간이 이산이면 바로 보일 수 있다.

정리 1.11 (선형성 II — 이산 확률공간) (Ω, P)가 이산 확률공간(1회 정리 4.9, $p_\omega = P(\{\omega\})$)이고 X가 그 위의 확률변수이면, $\sum_\omega |X(\omega)| p_\omega < \infty$일 때

$$EX = \sum_{\omega \in \Omega} X(\omega)\,p_\omega.$$

따라서 기댓값이 존재하는 $X_1, \dots, X_n$과 상수 $c_1, \dots, c_n$에 대해 $E\left(\sum_i c_i X_i\right) = \sum_i c_i\,EX_i$이다.

증명. ω들을 $X(\omega)$의 값으로 묶는다. 값이 x인 ω들의 확률의 합이 $P(X=x)=p(x)$이므로, 정리 1.9(a)의 증명(연습 1.9)과 같은 묶음으로 $\sum_\omega X(\omega)p_\omega = \sum_x x\,p(x) = EX$이다(절댓값으로 먼저 확인하고 보조정리 0.1). 그러면 $\sum_i c_iX_i$의 기댓값은 $\sum_\omega \sum_i c_iX_i(\omega)p_\omega$이고, 유한 합과 절대수렴하는 급수는 순서를 바꿀 수 있으므로 $\sum_i c_i \sum_\omega X_i(\omega)p_\omega = \sum_i c_i EX_i$이다. ∎

연속 확률변수를 포함한 일반적인 경우는 3회에서 결합밀도로 증명한다. 그 전까지는 이 정리를 이산 확률공간에서만 쓴다. 이 정리에는 **독립의 가정이 없다.** 아래 예제가 그 힘을 보여 준다.

예제 1.12 (선형성의 힘 — Montmort, 생일, 이항)

Montmort의 짝 맞추기. n장의 카드를 섞을 때 일치하는 카드의 수를 N이라 하면 $N=\sum_{i=1}^{n}\mathbf{1}_{A_i}$, $A_i=\{i\text{번째 카드가 맞는다}\}$, $P(A_i)=\frac{1}{n}$이다. 사건 A_i들은 독립이 아니지만 선형성에는 상관없다.

$$EN=\sum_{i=1}^{n}P(A_i)=n\cdot\frac{1}{n}=1.$$

카드가 몇 장이든 평균 한 장이 맞는다. 1회 예제 4.6처럼 포함배제로 N의 분포를 구할 필요가 없다.

생일. n명 가운데 생일이 같은 쌍의 수는 $\binom{n}{2}$개 쌍의 지시변수의 합이고, 쌍마다 확률은 $\frac{1}{365}$이므로 기댓값은 $\binom{n}{2}/365$이다. $n=23$이면 $253/365\approx 0.69$다. 같은 생일의 쌍이 평균 한 개 이상이 되려면 $n(n-1)\ge 730$, 곧 $n\ge 28$이다.

이항. n번의 독립 시행에서 성공 수 $K=\sum_i \mathbf{1}_{\{i\text{번째 성공}\}}$이므로 $EK=np$. 이항분포의 pmf를 쓰지 않고 얻었다.

1.5 꼬리 공식

정리 1.13 (꼬리 공식) (a) X가 음이 아닌 정수값을 가지면, (b) $X\ge 0$이 연속이면 각각

$$\text{(a)}\quad EX=\sum_{k=1}^{\infty}P(X\ge k),\qquad \text{(b)}\quad EX=\int_0^\infty P(X>x)\,dx=\int_0^\infty\big(1-F(x)\big)\,dx.$$

둘 다 양변이 ∞일 수 있다.

증명. **전략** 값 j를 $\sum_{k=1}^{j}1$로 펼쳐 EX를 이중합(이중적분)으로 만든다. 음이 아닌 항의 이중합은 더하는 순서를 바꿀 수 있다 [T6].

(a)

① $P(X \ge k) = \sum_{j \ge k} p(j)$이다. $\{X \ge k\} = \bigcup_{j \ge k} \{X = j\}$, (P3)

② 따라서 $\sum_{k \ge 1} P(X \ge k) = \sum_{k \ge 1} \sum_{j \ge k} p(j)$이고, 항 $p(j)$는 $1 \le k \le j$인 k마다 한 번씩, 모두 j번 나온다. 셈

③ 음이 아닌 항의 이중합이므로 j를 먼저 고정해 더해도 된다.

$$\sum_{k=1}^{\infty} \sum_{j=k}^{\infty} p(j) = \sum_{j=1}^{\infty} \Big(\sum_{k=1}^{j} 1 \Big) p(j) = \sum_{j=1}^{\infty} j\, p(j) = EX. \qquad \text{1회 4.6절의 상자}$$

(b)

① $P(X > x) = \int_x^{\infty} f(t)\, dt$이다. 1회 정리 6.7(c)

② 영역 $\{(x,t) : 0 \le x < t\}$ 위에서 음이 아닌 함수 $f(t)$를 두 순서로 적분한다.

$$\begin{aligned} \int_0^{\infty} P(X > x)\, dx &= \int_0^{\infty} \int_x^{\infty} f(t)\, dt\, dx \\ &= \int_0^{\infty} \Big(\int_0^{t} dx \Big) f(t)\, dt \qquad \text{Tonelli, 아래 상자 [T6]} \\ &= \int_0^{\infty} t\, f(t)\, dt = EX. \qquad \text{정의 1.5} \end{aligned}$$

③ $P(X > x) = 1 - F(x)$이므로 둘째 식도 같다. 1회 정리 4.3(c)

조건 확인 순서를 바꿀 때 쓴 것은 항(피적분함수)이 음이 아니라는 것뿐이다. 그래서 양변이 ∞여도 등식이 성립한다. 부호가 섞인 확률변수에는 그대로 쓸 수 없고, $X = X^+ - X^-$로 나눠 각각에 쓴다. ■

> **증명 없이 쓰는 사실** (음이 아닌 함수의 적분 순서 (Tonelli)) $h(x,t) \ge 0$이면 $\int \big(\int h\, dt \big) dx = \int \big(\int h\, dx \big) dt$이다(양변이 ∞일 수도 있다). 합과 적분, 합과 합 사이에서도 같다. 특히 음이 아닌 확률변수 Y_k에 대해 $E \sum_k Y_k = \sum_k EY_k$. 이것과 $Y = Y^+ - Y^-$로 쪼개기를 합하면, $\sum_k E|Y_k| < \infty$일 때 $E \sum_k Y_k = \sum_k EY_k$이다. 측도론의 Tonelli·Fubini 정리이고, 이 과정에서는 이 형태로만 쓴다.

> **예제 1.14** (꼬리 공식으로 다시)
>
> **기하분포.** $P(X \ge k) = P(\text{처음 } k-1\text{번 실패}) = q^{k-1}$이므로 $EX = \sum_{k \ge 1} q^{k-1} = \frac{1}{p}$. 예제 1.4보다 짧다.
>
> **지수분포.** $P(X > x) = e^{-\lambda x}$이므로 $EX = \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$. 부분적분이 필요 없다.
>
> **혼합형.** 1회 6.3절의 보험금 X(확률 0.3으로 0, 나머지는 밀도 $0.7e^{-x}$)는 이산도 연속도 아니어서 정의 1.1도 1.5도 쓸 수 없다. 그러나 꼬리 공식의 우변은 계산할 수 있다. $P(X >$

$x) = 0.7e^{-x}$ 이므로 $\int_0^\infty 0.7e^{-x}dx = 0.7$이다. 이처럼 꼬리 공식은 음이 아닌 **모든** 확률변수의 기댓값을 정의하는 방법이 된다.

연습문제

연습 1.1 (룰렛 ★) 미국식 룰렛에는 0, 00과 1–36의 38칸이 있고, 빨강은 18칸이다. 빨강에 1을 걸어 맞으면 1을 받고 틀리면 1을 잃는다. 기대 이익을 구하라. 100판 동안의 기대 이익은 얼마인가? 풀이 58쪽

연습 1.2 (Huygens의 명제 III ★★) a를 얻을 경우가 p가지, b를 얻을 경우가 q가지이고 모든 경우가 같은 정도로 쉽게 일어난다(p, q는 자연수). $p+q$명이 각각 x씩 걸고 제비를 뽑는 공평한 내기를 만들어, 기대가 $x = \frac{pa+qb}{p+q}$ 임을 Huygens의 방식으로 보여라. (힌트: 나는 다른 사람 가운데 q명과 "당신이 이기면 내게 b를 주고, 내가 이기면 내가 당신에게 b를 준다"고 약속하고, 나머지 $p-1$명과는 같은 약속을 a로 한다.) 풀이 58쪽

연습 1.3 (기댓값이 무한대 ★) $P(X=k) = \frac{1}{k(k+1)}$, $k = 1, 2, \dots$ 가 확률질량함수임을 보이고, $EX = \infty$ 임을 보여라. 풀이 58쪽

연습 1.4 (조건수렴은 기댓값이 아니다 ★★) X가 값 $(-1)^{k+1}2^k/k$를 확률 2^{-k}로 갖는다($k = 1, 2, \dots$). $\sum_k x_k p(x_k)$는 수렴하지만 X의 기댓값은 존재하지 않음을 보여라. 연습 0.1과 연결해, 이 합을 "기댓값 $\log 2$"라 부르면 안 되는 이유를 말하라. 풀이 58쪽

연습 1.5 (쿠폰 모으기 ★★) 과자 봉지마다 n종류의 쿠폰 가운데 하나가 같은 확률로 독립적으로 들어 있다. n종류를 모두 모을 때까지 사는 봉지 수를 T라 하자. $T = G_1 + G_2 + \cdots + G_n$, G_j는 $j-1$종류를 가진 뒤 새 종류를 얻을 때까지의 봉지 수로 쓰고, $ET = n\sum_{j=1}^{n} \frac{1}{j}$ 임을 보여라. 주사위의 여섯 눈을 모두 보려면 평균 몇 번 던져야 하나? 풀이 58쪽

연습 1.6 (제곱의 꼬리 공식 ★★) X가 음이 아닌 정수값을 가지면 $EX^2 = \sum_{k\ge 1}(2k-1)P(X \ge k)$ 임을 보여라. 기하분포에서 $EX^2 = \frac{2-p}{p^2}$를 이것으로 구하라. 풀이 59쪽

연습 1.7 (Halley의 기대여명 ★) Halley의 생명표에서 $\ell(x)$를 1세 1,000명 가운데 x세에 살아 있는 수라 하자. 이미 x세인 사람이 앞으로 **완전히** 살 햇수 K는 $P(K \ge k) = \ell(x+k)/\ell(x)$를 만족한다. 꼬리 공식으로 기대여명 $EK = \sum_{k\ge 1} \ell(x+k)/\ell(x)$를 쓰고, 노트북으로 $x = 20$일 때 값을 구하라(표는 84세에서 끝나므로 그 뒤는 0으로 둔다). 풀이 59쪽

연습 1.8 (대칭이어도 ★★) 예제 1.8의 Cauchy 밀도에서 $\int_{-A}^{2A} x\,f(x)\,dx \to \frac{1}{\pi}\log 2$ $(A \to \infty)$ 임을 보여라. 풀이 59쪽

연습 1.9 (함수의 기댓값: 증명 ★★ 유도형) 정리 1.9(a)를 다음 순서로 증명하라. $Y = g(X)$가 갖는 서로 다른 값을 $y_1, y_2, \dots$ 라 하자.

(a) $\{Y = y_j\} = \bigcup_{i:\, g(x_i)=y_j} \{X = x_i\}$가 서로소인 합집합임을 보이고 $P(Y = y_j) = \sum_{i:\, g(x_i)=y_j} p(x_i)$를 얻어라.

(b) 음이 아닌 급수를 묶어 더해도 된다는 것(1회 4.6절의 상자)으로 $\sum_j |y_j| P(Y = y_j) = \sum_i |g(x_i)| p(x_i)$를 보여라. 따라서 EY가 존재한다.

(c) 같은 묶음을 절댓값 없이 해서 $EY = \sum_i g(x_i) p(x_i)$를 얻어라. 이 단계에서 보조정리 0.1가 왜 필요한가?

풀이 59쪽

2

분산과 적률 — 흩어짐을 재는 수

기댓값은 분포의 위치를 하나의 수로 말한다. Bernoulli의 질문에 답하려면 한 가지가 더 필요하다. 관측한 비율이 참값 **둘레에 얼마나 흩어지는가**다. 이 절은 흩어짐의 표준 척도인 분산을 정의하고, 이항분포의 분산 npq와 독립합의 분산 공식을 증명한다. 6절의 대수의 법칙은 이 두 결과와 5절의 Chebyshev 부등식을 합친 것이다.

> **이름의 역사** 흩어짐을 재는 일은 천문학의 관측오차에서 시작했다. Gauss(1809, 1821)는 오차 제곱의 평균을 기준으로 삼았고, Karl Pearson(1894)이 그 제곱근에 *standard deviation*이라는 이름을 붙였다. *variance*라는 말은 Fisher(1918)가 처음 썼다. 개념은 이름보다 150년 앞섰다. de Moivre(1733)는 이항분포의 흩어짐이 $\sqrt{n}$에 비례한다는 것을 이미 알고 있었다(5회).

2.1 분산

정의 2.1 (분산과 표준편차) $EX^2 < \infty$인 확률변수 X의 평균을 μ라 할 때

$$\operatorname{Var} X = E(X-\mu)^2$$

을 **분산**이라 하고, $\sigma = \sqrt{\operatorname{Var} X}$를 **표준편차**라 한다. σ^2, σ_X^2로도 쓴다.

$EX^2 < \infty$이면 $|x| \le 1 + x^2$이므로 $E|X| < \infty$이고, $(x-\mu)^2 \le 2x^2 + 2\mu^2$이므로 분산도 유한하다. 분산은 무게중심 둘레의 "관성 모멘트"다. 평균에서 멀리 떨어진 질량이 제곱으로 무겁게 센다.

정리 2.2 (분산의 계산) $EX^2 < \infty$이고 $\mu = EX$일 때

(a) $\operatorname{Var} X = EX^2 - \mu^2$.

(b) $\operatorname{Var}(aX+b) = a^2 \operatorname{Var} X$. 따라서 $\sigma_{aX+b} = |a|\,\sigma_X$.

(c) 모든 실수 c에 대해 $E(X-c)^2 = \operatorname{Var} X + (\mu - c)^2$. 따라서 $E(X-c)^2$은 $c = \mu$에서 가장 작다.

증명. **전략** (c)를 먼저 보인다. $X - c = (X-\mu) + (\mu - c)$를 제곱하고 선형성을 쓰면 가운데 항이 사라진다. (a)와 (b)는 (c)와 정의에서 바로 나온다. 단계별 증명은 연습 2.10(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 61쪽에 있다.

(c)는 한 줄이지만 깊다. X를 상수 하나로 예측해야 하고 손실이 제곱오차이면, 가장 좋은 예측은

평균이고 그때의 손실이 분산이다. 13회의 최소제곱법은 이 사실을 여러 변수로 넓힌 것이다. 손실이 절대오차 $E|X-c|$이면 가장 좋은 예측은 중앙값이 된다(연습 2.2).

예제 2.3 (기본적인 분산들)

Bernoulli. $X=\mathbf{1}_A$, $P(A)=p$이면 $X^2=X$이므로 $\operatorname{Var}X=p-p^2=pq$. 이 값은 $p=\tfrac12$에서 가장 크고, 언제나 $pq\le\tfrac14$이다.

주사위. $EX^2=\tfrac16(1+4+\cdots+36)=\tfrac{91}{6}$이므로 $\operatorname{Var}X=\tfrac{91}{6}-\tfrac{49}{4}=\tfrac{35}{12}\approx 2.92$.

균등분포. $X\sim \mathrm{U}(a,b)$이면 $X=a+(b-a)U$, $U\sim\mathrm{U}(0,1)$이고 $\operatorname{Var}U=\tfrac13-\tfrac14=\tfrac{1}{12}$이므로 정리 2.2(b)로 $\operatorname{Var}X=\tfrac{(b-a)^2}{12}$.

지수분포. 부분적분을 두 번 하면 $EX^2=\tfrac{2}{\lambda^2}$이므로 $\operatorname{Var}X=\tfrac{1}{\lambda^2}$. 표준편차가 평균과 같다.

2.2 이항분포의 평균과 분산

정리 2.4 (이항분포) K가 성공 확률 p인 독립 시행 n번의 성공 수이면, 곧

$$P(K=k)=\binom{n}{k}p^k q^{n-k},\qquad k=0,1,\dots,n$$

이면 ($K\sim \mathrm{b}(n,p)$로 쓴다)

$$EK=np,\qquad \operatorname{Var}K=npq.$$

증명. **전략** pmf의 합이 1이라는 것은 이항정리(1회 정리 1.6)다. 같은 정리를 한 번 더 쓴다. $k\binom{n}{k}=n\binom{n-1}{k-1}$로 k를 계수 안으로 흡수하고, 분산은 $E\,K(K-1)$을 같은 방법으로 구해 얻는다. 단계별 증명은 연습 2.11(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 61쪽에 있다.

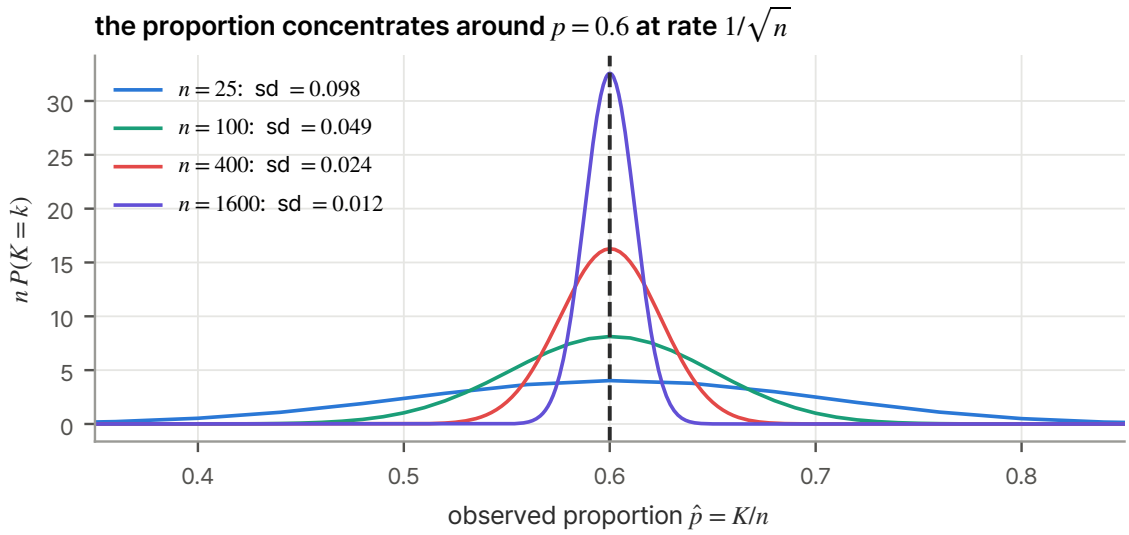


그림 2.1 $p = 0.6$일 때 관측 비율 $\hat{p} = K/n$의 분포. n이 4배가 될 때마다 폭이 절반이 된다. 세로축은 $n\,P(K = k)$로 맞춰 모양을 비교할 수 있게 했다.

따름정리 2.5 (관측 비율의 평균과 분산) $\hat{p} = K/n$이면

$$E\hat{p} = p, \qquad \operatorname{Var}\hat{p} = \frac{pq}{n} \le \frac{1}{4n}, \qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}.$$

증명. 정리 2.4과 정리 2.2(b)에 $a = \frac{1}{n}$을 쓴다. ∎

관측 비율은 평균적으로 참값이고, 흩어짐의 표준편차는 $1/\sqrt{n}$에 비례해 준다. 시행을 4배로 늘려야 흩어짐이 절반이 된다(그림 2.1). 이것이 1회 8.1절에서 시뮬레이션으로 본 $1/\sqrt{n}$이고, Bernoulli의 정리의 씨앗이다.

2.3 독립인 이산 확률변수

이항분포의 분산은 "n번 시행의 분산이 한 번의 분산의 n배"라는 것을 말해 준다. 이것은 독립이면 언제나 참이다.

정의 2.6 (독립인 이산 확률변수) 이산 확률변수 $X_1, \dots, X_n$이 모든 값 $x_1, \dots, x_n$에 대해

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

을 만족하면 **독립**이라 한다.

이 정의는 1회 정의 5.9의 사건의 상호 독립과 같은 말이다. 모든 값의 조합에서 사건 $\{X_1 = x_1\}, \dots, \{X_n = x_n\}$이 상호 독립이라는 것이다. 정의에는 n개를 모두 곱하는 식만 있지만, 일부만

고른 식은 나머지 변수의 값에 대해 더하면 나온다(연습 2.6). 같은 이유로 독립인 X_i들의 함수 $g_i(X_i)$도 독립이다.

정리 2.7 (독립이면 곱의 기댓값은 기댓값의 곱) X, Y가 독립인 이산 확률변수이고 $E|X|, E|Y| < \infty$이면 $E|XY| < \infty$이고

$$E(XY) = EX \cdot EY.$$

증명. **전략** (X, Y)를 하나의 이산 확률변수로 보고 정리 1.9(a)를 쓴다. 독립이면 이중합이 두 합의 곱으로 갈라진다. 먼저 절댓값으로 합이 유한함을 보이고, 그다음 절댓값 없이 계산한다 [T5]. 단계별 증명은 연습 2.12(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 61쪽에 있다.

정리 2.8 (독립합의 분산) $X_1, \dots, X_n$이 독립인 이산 확률변수이고 분산이 유한하면

$$\operatorname{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \operatorname{Var} X_i.$$

특히 모두 분산이 σ^2이면 표본평균 $\bar{X} = \frac{1}{n}\sum X_i$에 대해 $\operatorname{Var}\bar{X} = \sigma^2/n$.

증명. **전략** 중심을 뺀 $D_i = X_i - \mu_i$로 분산을 "합의 제곱의 기댓값"으로 쓴다. 제곱을 전개하면 교차항이 생기고, 독립이 교차항을 0으로 만든다.

① $\mu_i = EX_i$, $D_i = X_i - \mu_i$로 두면 $E\sum_i X_i = \sum_i \mu_i$이고 $\sum_i X_i - \sum_i \mu_i = \sum_i D_i$이다. — 정리 1.11

② 분산의 정의에 넣고 제곱을 전개한다.

$$\operatorname{Var}\Big(\sum_i X_i\Big) = E\Big(\sum_i D_i\Big)^2 \qquad \text{정의 2.1, ①}$$

$$= \sum_i E\,D_i^2 + \sum_{i \neq j} E(D_i D_j) \qquad \text{전개, 정리 1.11}$$

③ $E|D_iD_j| < \infty$이다. $2|D_iD_j| \le D_i^2 + D_j^2$이고 오른쪽의 기댓값이 유한하기 때문이다. — $2|ab| \le a^2 + b^2$ [T4]

④ $i \neq j$이면 D_i와 D_j는 독립이다. — 독립인 변수의 함수 (2.3절)

⑤ $E(D_iD_j) = ED_i \cdot ED_j = 0 \cdot 0 = 0$이다. — ③ ④, 정리 2.7

⑥ 남는 것은 $\sum_i E\,D_i^2 = \sum_i \operatorname{Var} X_i$이다. — 정의 2.1

⑦ 모두 분산이 σ^2이면 $\operatorname{Var}\bar{X} = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$이다. — 정리 2.2(b), $a = \frac{1}{n}$

조건 확인 전개한 합의 기댓값을 항마다 나눠 쓰려면(선형성) 각 항의 기댓값이 존재해야 한다. 제곱항은 분산이 유한하다는 가정으로, 교차항은 ③으로 확인했다. ∎

증명에서 쓴 것은 $i \neq j$인 **쌍**마다 $E(D_iD_j) = 0$이라는 것뿐이다. 그러니 쌍별 독립이면 충분하다(연습 2.5). 연속 확률변수에 대한 같은 정리는 3회에서 증명하고, 그때 $E(D_iD_j)$에 **공분산**이라는

이름을 붙인다.

예제 2.9 (지시변수로 다시 — 이항과 Montmort)

이항. $K=\sum_i \mathbf{1}_i$에서 $\mathbf{1}_i$들이 독립이고 각각 분산이 pq이므로 $\operatorname{Var}K=npq$. 정리 2.4의 계산 없이 얻었다.

Montmort. 예제 1.12의 일치 수 $N=\sum \mathbf{1}_{A_i}$은 독립이 아니어서 정리 2.8을 쓸 수 없다. 대신 증명의 첫 줄로 돌아가 교차항을 직접 계산한다. $i\neq j$이면 $P(A_iA_j)=\frac{(n-2)!}{n!}=\frac{1}{n(n-1)}$이므로 $E\,N(N-1)=\sum_{i\neq j}P(A_iA_j)=1$이고, $n\geq 2$이면

$$\operatorname{Var}N=E\,N(N-1)+EN-(EN)^2=1+1-1=1.$$

평균도 1, 분산도 1이다. 1회 연습 4.7에서 N의 분포가 평균 1인 Poisson 분포에 다가간다는 것을 보았다. 4회에서 Poisson 분포의 평균과 분산이 같다는 것을 증명한다.

2.4 적률

정의 2.10 (적률 · 왜도 · 첨도) $E|X|^k<\infty$일 때 EX^k를 X의 k차 **적률**, $E(X-\mu)^k$를 k차 **중심적률**이라 한다. 분산이 $\sigma^2>0$이면

$$\gamma_1=\frac{E(X-\mu)^3}{\sigma^3}\ \text{(왜도)},\qquad \gamma_2=\frac{E(X-\mu)^4}{\sigma^4}-3\ \text{(초과첨도)}$$

로 정의한다.

k차 적률이 있으면 그보다 낮은 차수의 적률도 모두 있다. $j<k$이면 $|x|^j\leq 1+|x|^k$이기 때문이다. 왜도와 첨도는 $aX+b$ $(a>0)$로 바꿔도 변하지 않는 **모양**의 수다. 왜도는 오른쪽 꼬리가 길면 양수다. 초과첨도의 -3은 5회의 정규분포에서 0이 되도록 맞춘 것이다. 지수분포는 $\gamma_1=2,\ \gamma_2=6$이고, 균등분포는 $\gamma_1=0,\ \gamma_2=-\frac{6}{5}$이다(연습 2.7).

> **역사** 동물학자 W. F. R. Weldon은 1894년 주사위 12개를 한꺼번에 굴려 5나 6이 나온 주사위의 수를 세는 일을 26,306번 되풀이하고, 그 기록을 Galton과 Karl Pearson에게 보냈다. Pearson은 1900년 이 자료로 χ^2 적합도 검정을 처음 보였다(8회).

예제 2.11 (Weldon의 주사위 (1894)) 공평한 주사위라면 한 번 굴릴 때 5·6이 나온 수 K는 $\mathrm{b}(12,\frac{1}{3})$이므로 $EK=4$, $\operatorname{Var}K=12\cdot\frac{1}{3}\cdot\frac{2}{3}=\frac{8}{3}\approx 2.667$이다. Weldon의 기록에서 계산한 평균과 분산은 다음과 같다(자료는 `vcd::WeldonDice`, 10개 이상은 10으로 묶여 있다).

$$\bar{k}=\frac{1}{26306}\sum_j k_jn_j=4.0522,\qquad \frac{1}{26306}\sum_j(k_j-\bar{k})^2n_j=2.696.$$

평균이 4보다 0.052 크다. 이 차이가 우연인지는 평균의 흩어짐에 달려 있다. 굴림들이 독

립이면 정리 2.8으로 $\bar{K}$의 표준편차는 $\sqrt{(8/3)/26306} \approx 0.0101$이므로, 차이는 표준편차의 5배다. 5절의 Chebyshev 부등식이 이것을 확률의 말로 바꾼다(예제 5.5).

연습문제

연습 2.1 (두 주사위의 합 ★) 두 주사위 눈의 합 S의 평균과 분산을 (a) pmf로 직접, (b) 정리 2.8으로 구하라. 풀이 59쪽

연습 2.2 (중앙값은 절대오차를 최소로 한다 ★★) X가 연속이고 $E|X| < \infty$이며 중앙값 m($F(m) = \frac{1}{2}$)이 있다. $c > m$일 때 $E|X-c| - E|X-m| = \int_m^c (2F(x)-1)\,dx \geq 0$임을 보이고, $E|X-c|$가 $c = m$에서 가장 작음을 결론지어라. (힌트: $|x-c| - |x-m|$을 $x \leq m$, $m < x < c$, $x \geq c$로 나눈다.) 풀이 60쪽

연습 2.3 (기하분포의 분산 ★★) 기하분포에서 $\operatorname{Var} X = \frac{q}{p^2}$임을 보여라(연습 1.6을 써도 좋다). 6이 나올 때까지 던지는 횟수의 표준편차는? 풀이 60쪽

연습 2.4 (포커의 에이스 ★★) 52장에서 5장을 뽑을 때 에이스의 수 K의 평균과 분산을 지시변수로 구하라. 이항분포 $\mathrm{b}(5, \frac{1}{13})$의 분산과 비교하고, 왜 더 작은지 설명하라. 풀이 60쪽

연습 2.5 (쌍별 독립이면 충분하다 ★★) 1회 예제 5.10의 Bernstein 정사면체에서 R, G, B를 각 색이 바닥에 있는 사건의 지시변수라 하자. 셋은 쌍별 독립이지만 상호 독립이 아니다. $\operatorname{Var}(R+G+B)$를 직접 계산해 $\operatorname{Var} R + \operatorname{Var} G + \operatorname{Var} B$와 같음을 확인하라. 풀이 60쪽

연습 2.6 (독립의 정의에서 일부만 ★) 정의 2.6가 성립하면 $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$임을 보여라. (힌트: $x_3, \dots, x_n$의 모든 값에 대해 더한다.) 풀이 60쪽

연습 2.7 (모양의 수 ★) $\mathrm{U}(0,1)$의 왜도와 초과첨도가 $0, -\frac{6}{5}$임을 보이고, $\mathrm{Exp}(1)$에서 $EX^k = k!$를 써서 $\gamma_1 = 2$, $\gamma_2 = 6$을 확인하라. Bernoulli(p)의 왜도는 $\frac{1-2p}{\sqrt{pq}}$임을 보여라. 풀이 60쪽

연습 2.8 (두 독립 복사본 ★★) X와 X'이 독립이고 분포가 같은 이산 확률변수이면 $\operatorname{Var} X = \frac{1}{2}E(X-X')^2$임을 보여라. 분산은 "무작위로 뽑은 두 값의 차이"로도 정의할 수 있다. 풀이 61쪽

연습 2.9 (Weldon의 분산 ★) Weldon의 자료에서 주사위 하나가 5·6을 낼 확률을 $\hat{p} = \bar{k}/12$로 추정하면 이항분포의 분산 $12\hat{p}(1-\hat{p})$은 얼마인가? 관측한 분산 2.696과 비교하라. 풀이 61쪽

연습 2.10 (분산의 계산: 증명 ★ 유도형) $EX^2 < \infty$, $\mu = EX$이다. 정리 2.2를 다음 순서로 증명하라.

(a) $(X-c)^2 = (X-\mu)^2 + 2(\mu-c)(X-\mu) + (\mu-c)^2$을 확인하고, 세 항의 기댓값이 모두 존재함을 보여라.

(b) 선형성(정리 1.10)으로 가운데 항의 기댓값이 0임을 보이고 (c)를 얻어라.

(c) (c)에 $c = 0$을 넣어 (a)를 얻어라.

(d) $aX + b$의 평균이 $a\mu + b$임을 쓰고 (b)를 얻어라.

풀이 61쪽

연습 2.11 (이항분포의 평균과 분산: 증명 ★★ 유도형) 정리 2.4을 다음 순서로 증명하라.

(a) 이항정리로 $\sum_k P(K=k) = 1$을 확인하라.

(b) $k \ge 1$이면 $k\binom{n}{k} = n\binom{n-1}{k-1}$임을 계승으로 보여라.

(c) (b)를 EK의 합에 넣고 첨자를 $j = k-1$로 바꿔 $EK = np(p+q)^{n-1} = np$를 얻어라.

(d) 같은 방법으로 $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$와 $E\,K(K-1) = n(n-1)p^2$을 보여라.

(e) $\operatorname{Var} K = E\,K(K-1) + EK - (EK)^2$을 쓰고 npq를 얻어라. 왜 EK^2 대신 $E\,K(K-1)$을 먼저 구하는가?

풀이 61쪽

연습 2.12 (곱의 기댓값: 증명 ★★ 유도형) 정리 2.7를 다음 순서로 증명하라.

(a) (X, Y)를 값 (x, y)를 갖는 이산 확률변수로 보고 정리 1.9(a)를 $g(x,y) = xy$에 써라. 어떤 합이 유한해야 하는가?

(b) 독립(정의 2.6)과 음이 아닌 이중급수의 순서 바꾸기로 $\sum_{x,y} |xy| P(X=x, Y=y) = E|X|\,E|Y|$를 보여라.

(c) 같은 계산을 절댓값 없이 해서 $E(XY) = EX \cdot EY$를 얻어라. 절댓값 없는 이중급수의 순서를 바꿀 수 있는 근거는 무엇인가?

풀이 61쪽

3

적률생성함수

적률을 하나씩 계산하는 대신 모두를 한 함수에 담을 수 있다. 그 함수를 미분하면 적률이 나오고, 두 독립 확률변수의 합에서는 그 함수가 곱이 된다. 합의 분포를 구하는 일이 곱셈으로 바뀌는 것이다. 이 도구는 4회의 분포들, 5회의 정규분포, 7회의 중심극한정리에서 계속 쓰인다.

> **역사** de Moivre는 *Miscellanea Analytica*(1730)에서 주사위 여러 개의 합의 분포를 다항식 $(z+z^2+\cdots+z^6)^n$의 계수로 구했다. Laplace는 *Théorie analytique des probabilités*(1812)의 첫 권 전체를 이 "생성함수"(fonctions génératrices)의 계산법에 바쳤다. 오늘의 적률생성함수 $E\,e^{tX}$는 분포의 Laplace 변환이다.

3.1 정의와 예

정의 3.1 (적률생성함수) 어떤 $h>0$에 대해 $-h<t<h$인 모든 t에서 $E\,e^{tX}<\infty$이면

$$M_X(t)=E\,e^{tX},\qquad -h<t<h$$

를 X의 **적률생성함수**(moment generating function, mgf)라 한다.

$M(0)=1$은 언제나 성립한다. 중요한 것은 0을 포함하는 열린 구간에서 유한하다는 조건이다. 이 조건은 꼬리가 지수적으로 가볍다는 뜻이고(연습 3.4), 바로 이 조건에서 아래의 정리들이 성립한다.

예제 3.2 (기본적인 적률생성함수)

Bernoulli와 이항. $X=\mathbf{1}_A$이면 $M(t)=q+pe^t$. $K\sim \mathrm{b}(n,p)$이면 이항정리(1회 정리 1.6)로

$$M_K(t)=\sum_{k=0}^{n}e^{tk}\binom{n}{k}p^kq^{n-k}=\sum_{k=0}^{n}\binom{n}{k}(pe^t)^kq^{n-k}=(q+pe^t)^n,\qquad t\in\mathbb{R}.$$

기하분포. $qe^t<1$, 곧 $t<-\log q$이면 등비급수로 $M(t)=\sum_{k\ge1}e^{tk}q^{k-1}p=\dfrac{pe^t}{1-qe^t}$.

지수분포. $t<\lambda$이면 $M(t)=\int_0^\infty e^{tx}\lambda e^{-\lambda x}dx=\dfrac{\lambda}{\lambda-t}$. $t\ge\lambda$이면 적분이 발산한다.

균등분포. $X\sim\mathrm{U}(0,1)$이면 $M(t)=\int_0^1 e^{tx}dx=\dfrac{e^t-1}{t}$ $(t\neq0)$, $M(0)=1$.

Cauchy. $t>0$이면 $e^{tx}\ge\dfrac{t^2x^2}{2}$ $(x>0)$이므로 $\int_0^\infty e^{tx}f(x)\,dx\ge\dfrac{t^2}{2\pi}\int_0^\infty\dfrac{x^2}{1+x^2}dx=\infty$. $t<0$도 같다. Cauchy 분포에는 mgf가 없다.

3.2 mgf에서 적률로

정리 3.3 (적률생성함수와 적률) X의 mgf가 $|t|<h$에서 존재하면 모든 차수의 적률이 유한하고, $|t|<h$에서

$$M(t)=\sum_{k=0}^{\infty}\frac{EX^k}{k!}t^k.$$

따라서 M은 $|t|<h$에서 무한 번 미분가능하고 $M^{(k)}(0)=EX^k$이다.

증명. **전략** e^{tX}를 거듭제곱급수 $\sum_k t^kX^k/k!$로 펼치고, 합과 기댓값의 순서를 바꾼다. 순서를 바꿔도 되는 근거는 절댓값의 급수의 기댓값이 유한하다는 것이고, 그것을 mgf가 0의 양쪽에서 존재한다는 가정에서 얻는다 [T6].

① $0\le s<h$이면 $e^{s|x|}\le e^{sx}+e^{-sx}$이므로 $Ee^{s|X|}\le M(s)+M(-s)<\infty$이다. 정리 1.10(b)

② $0<s<h$를 하나 고르면 $e^{s|x|}=\sum_j\frac{s^j|x|^j}{j!}\ge\frac{s^k|x|^k}{k!}$이므로 $E|X|^k\le k!\,s^{-k}\,Ee^{s|X|}<\infty$이다. 모든 k에서 성립한다. 지수함수의 급수 [T8], ①

③ $|t|<h$에서 $e^{tX}=\sum_k Y_k$, $Y_k=\frac{t^kX^k}{k!}$로 쓰면

$$\sum_k E|Y_k|=E\sum_k\frac{|t|^k|X|^k}{k!}=Ee^{|t||X|}<\infty.$$

Tonelli (1.5절 상자), ①

④ 따라서 $Ee^{tX}=\sum_k EY_k=\sum_k\frac{t^kEX^k}{k!}$이다. 1.5절 상자의 마지막 문장

⑤ 수렴반경이 h 이상인 거듭제곱급수는 그 안에서 항별로 미분할 수 있다. k번 미분해 $t=0$을 넣으면 t^k의 계수의 $k!$배, 곧 EX^k만 남는다. 거듭제곱급수의 항별 미분

조건 확인 ④에서 합과 기댓값의 순서를 바꾼 근거는 $\sum_k E|Y_k|<\infty$이고, 이것을 ③에서 확인했다. $M(s)$와 $M(-s)$가 모두 유한해야 ①이 성립하므로, mgf가 0의 한쪽에서만 존재하면 이 증명은 성립하지 않는다. ∎

따름정리 3.4 (로그 mgf) $\psi(t)=\log M(t)$라 하면 $\psi'(0)=EX$, $\psi''(0)=\operatorname{Var}X$이다.

증명. $\psi'=M'/M$이고 $\psi''=\big(M''M-(M')^2\big)/M^2$이다. $M(0)=1$, $M'(0)=EX$, $M''(0)=EX^2$을 넣는다. ∎

ψ를 **누율생성함수**라 한다. 평균과 분산이 ψ의 처음 두 미분계수라는 사실은 5회(정규분포의 $\psi(t)=\mu t+\frac12\sigma^2t^2$)와 7회(중심극한정리)에서 중심이 된다. 그림 3.1에서 보듯 0 근처의 ψ는 이차식과 거의 같다.

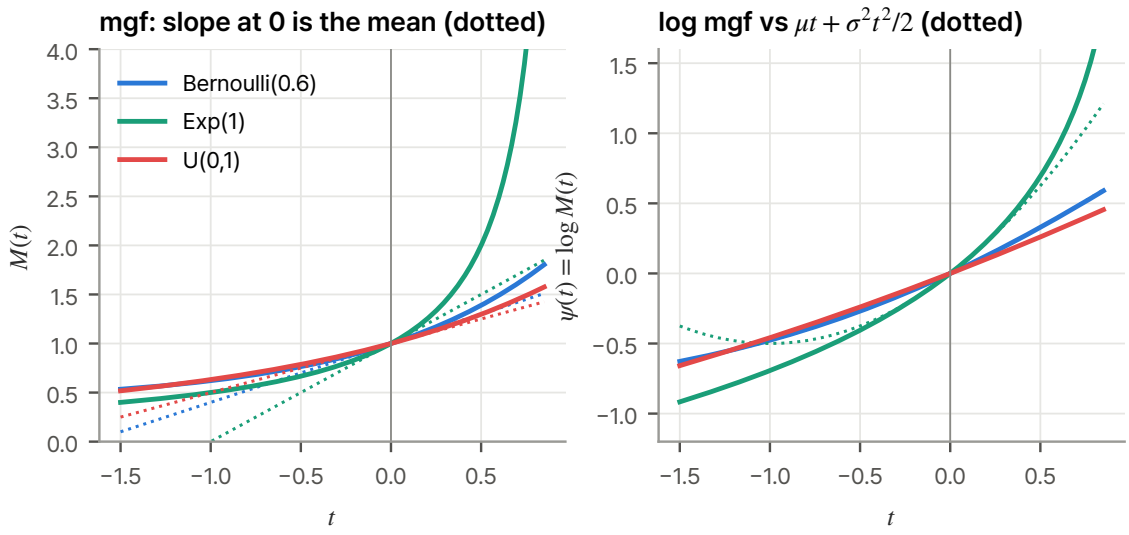


그림 3.1 세 분포의 mgf(왼쪽)와 로그 mgf(오른쪽). 점선은 0에서의 접선 $1 + \mu t$와 이차 근사 $\mu t + \frac{1}{2}\sigma^2 t^2$이다. 지수분포의 mgf는 $t \to 1$에서 무한대로 간다.

예제 3.5 (mgf로 계산하기)

이항. $M = (q + pe^t)^n$에서 $\psi(t) = n\log(q + pe^t)$, $\psi'(t) = \frac{npe^t}{q+pe^t}$, $\psi''(t) = \frac{npqe^t}{(q+pe^t)^2}$이므로 $\psi'(0) = np$, $\psi''(0) = npq$. 정리 2.4을 미분 두 번으로 다시 얻었다.

지수분포의 모든 적률. $|t| < \lambda$에서 $M(t) = \frac{1}{1-t/\lambda} = \sum_k \frac{t^k}{\lambda^k}$이다. 정리 3.3의 급수와 계수를 비교하면 $\frac{EX^k}{k!} = \frac{1}{\lambda^k}$, 곧 $EX^k = \frac{k!}{\lambda^k}$. 미분을 한 번도 하지 않았다.

3.3 변환과 독립합, 그리고 유일성

정리 3.6 (아핀 변환과 독립합)

(a) $M_{aX+b}(t) = e^{bt} M_X(at)$.

(b) X, Y가 독립인 이산 확률변수이고 둘 다 mgf가 있으면 $M_{X+Y}(t) = M_X(t)\,M_Y(t)$.

증명. (a) $E\,e^{t(aX+b)} = e^{bt} E\,e^{(at)X}$. (b) e^{tX}와 e^{tY}는 독립인 변수의 함수이므로 독립이고(2.3절), 정리 2.7로 $E\,e^{t(X+Y)} = E\,e^{tX}e^{tY} = E\,e^{tX}\,E\,e^{tY}$. ∎

(b)는 연속 확률변수에서도 성립한다. 3회에서 결합밀도로 정리 2.7를 넓히면 증명이 그대로 옮겨 간다. mgf가 쓸모 있는 것은 다음 사실 때문이다.

증명 없이 쓰는 사실 (mgf의 유일성) 두 확률변수 X, Y의 mgf가 0을 포함하는 어떤 열린 구간에서 같으면 X와 Y의 분포함수가 같다. mgf는 분포의 Laplace 변환이고, 이 변환을 거꾸로 푸는 공식이 있다는 것이 증명의 핵심이다(복소해석이 필요하다). Billingsley, *Probability and Measure*, 30절. 조건 "0의 근방에서 존재"는 빼면 안 된다. 7절에서 mgf가 없는 분포는 모든 적률이 같아도 다를 수 있음을 본다.

예제 3.7 (독립인 시행의 합) $X_1, \dots, X_n$이 독립이고 각각 Bernoulli(p)이면 정리 3.6(b)를 되풀이해 $M_{\sum X_i}(t) = (q + pe^t)^n$이다. 이것은 b($n, p$)의 mgf이므로 유일성으로 $\sum X_i \sim \mathrm{b}(n, p)$. 같은 방법으로 $K_1 \sim \mathrm{b}(n_1, p)$와 $K_2 \sim \mathrm{b}(n_2, p)$가 독립이면 $K_1 + K_2 \sim \mathrm{b}(n_1 + n_2, p)$이다. 분포를 합성곱으로 계산하지 않았다.

예제 3.8 (de Moivre의 방법 — 세 주사위) 음이 아닌 정수값의 확률변수에서는 $z = e^t$로 바꾼 $G(z) = E\,z^X = \sum_k P(X = k) z^k$(**확률생성함수**)가 편하다. 주사위 하나는 $G(z) = \frac{1}{6}(z + z^2 + \cdots + z^6)$이고, 독립인 세 주사위의 합은 정리 3.6(b)로

$$G_S(z) = \frac{1}{216}\left(z + z^2 + \cdots + z^6\right)^3$$

이다. 전개하면 z^9의 계수는 25, z^{10}의 계수는 27이다. 1회 예제 1.8에서 Galileo가 경우를 하나하나 세어 얻은 수가 다항식의 곱셈 한 번으로 나온다. 노트북에서 `np.convolve`로 확인한다.

연습문제

연습 3.1 (주사위의 mgf ★) 주사위 눈 X의 mgf를 쓰고, $M'(0)$과 $M''(0)$에서 평균과 분산을 확인하라. 풀이 62쪽

연습 3.2 (기하분포 ★) 예제 3.2의 기하분포 mgf에 따름정리 3.4를 써서 $EX = \frac{1}{p}$, $\mathrm{Var}\,X = \frac{q}{p^2}$을 다시 구하라. 풀이 62쪽

연습 3.3 (mgf에서 분포로 ★) $M(t) = \frac{1}{4}e^{-t} + \frac{1}{2} + \frac{1}{4}e^{t}$인 확률변수의 분포를 말하라. 근거는 무엇인가? 이 분포는 공평한 동전 두 개로 어떻게 만들 수 있나? 풀이 62쪽

연습 3.4 (mgf가 있으면 꼬리가 가볍다 ★★) mgf가 $|t| < h$에서 존재하면 $0 < s < h$에 대해 $P(X \ge x) \le M(s)\,e^{-sx}$임을 보여라. (힌트: $\mathbf{1}\{X \ge x\} \le e^{s(X-x)}$.) 꼬리가 다항식처럼 줄어드는 분포(예: $P(X \ge x) = x^{-2}$, $x \ge 1$)에는 mgf가 없음을 결론지어라. 풀이 62쪽

연습 3.5 (누율의 덧셈 ★★) $\psi_X = \log M_X$라 하자. 독립인 이산 확률변수에서 $\psi_{X+Y} = \psi_X + \psi_Y$이므로, 분산뿐 아니라 $\psi'''(0)$도 독립합에서 더해진다. $\psi'''(0) = E(X-\mu)^3$임을 보이고, n개의 독립인 Bernoulli(p)의 합의 왜도가 $\frac{1-2p}{\sqrt{npq}}$임을 결론지어라. 풀이 62쪽

연습 3.6 (Sicherman의 주사위 ★★★) 면에 $1,2,2,3,3,4$와 $1,3,4,5,6,8$이 적힌 두 주사위를 굴린 합은 보통 주사위 두 개의 합과 분포가 같음을 확률생성함수로 보여라. (힌트: $z + z^2 + \cdots + z^6 = z(1+z)(1+z+z^2)(1-z+z^2)$이고, 두 주사위는 이 인수들을 다르게 나눠 가진 것이다.) 풀이 63쪽

4

변수변환 — 확률변수의 함수의 분포

1절은 $g(X)$의 **기댓값**을 X의 분포로 구했다. 이 절은 $g(X)$의 **분포** 자체를 구한다. 방법은 둘이다. 분포함수에서 출발하는 방법은 언제나 쓸 수 있고, 단조함수이면 밀도를 바로 주는 공식이 있다. 마지막으로 이 공식을 거꾸로 써서, 균등 난수 하나로 어떤 분포의 난수든 만드는 방법을 증명한다. 컴퓨터의 모든 확률 시뮬레이션이 이 방법에서 시작한다.

4.1 분포함수 방법

$Y = g(X)$의 분포함수는 정의 그대로 $F_Y(y) = P(g(X) \le y)$이다. 사건 $\{g(X) \le y\}$를 X에 대한 사건으로 다시 쓰고 X의 분포로 확률을 계산한다. 그다음 미분하면 밀도가 나온다.

예제 4.1 (제곱) X가 밀도 f_X를 갖고 $Y = X^2$이라 하자. $y > 0$이면 $\{X^2 \le y\} = \{-\sqrt{y} \le X \le \sqrt{y}\}$이므로

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}), \qquad f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}} \quad (y > 0).$$

미분에서 연쇄법칙 $\frac{d}{dy}\sqrt{y} = \frac{1}{2\sqrt{y}}$를 썼다. 두 항은 Y의 한 값 y로 가는 X의 두 값 $\pm\sqrt{y}$에서 온다. $X \sim \mathrm{U}(-1,1)$이면 $f_Y(y) = \frac{1}{2\sqrt{y}}$ $(0 < y < 1)$이고, 이것은 1회 연습 6.7의 $\mathrm{U}(0,1)$의 제곱과 같은 분포다. 5회에서 X가 표준정규이면 같은 공식이 $\chi^2(1)$ 분포를 준다.

예제 4.2 (지수분포는 균등 난수의 로그) $U \sim \mathrm{U}(0,1)$, $Y = -\frac{1}{\lambda}\log U$라 하자. $y > 0$이면 $\{Y \le y\} = \{U \ge e^{-\lambda y}\}$이므로 $F_Y(y) = 1 - e^{-\lambda y}$, 곧 $Y \sim \mathrm{Exp}(\lambda)$이다. 균등 난수의 로그가 기다리는 시간이 된다.

4.2 단조변환의 밀도

정리 4.3 (단조변환) X의 밀도 f_X가 열린 구간 S 밖에서 0이고, g가 S에서 미분가능하며 $g'(x) \neq 0$ (따라서 순증가이거나 순감소)이라 하자. $T = g(S)$, $h = g^{-1} : T \to S$로 두면 $Y = g(X)$의 밀도는

$$f_Y(y) = f_X\big(h(y)\big)\,|h'(y)|, \qquad y \in T$$

이고, T 밖에서 0이다.

증명. **전략** 분포함수 방법이다. 사건 $\{g(X) \le y\}$를 X에 대한 사건으로 바꿔 F_Y를 F_X와 h로 쓰고, 연쇄법칙으로 미분한다 [T7]. $P(X \in S) = 1$이므로 S 밖의 결과는 확률에 기여하지 않는다.

증가하는 경우

① g가 증가하면 h도 증가하고, $y \in T$에서 $\{g(X) \le y\} = \{X \le h(y)\}$이다($S$ 밖의 확률 0인 부분을 빼고). — 순증가 함수와 그 역함수 [T1]

② $F_Y(y) = P(X \le h(y)) = F_X(h(y))$이다. — 1회 정의 6.2

③ h는 미분가능하고 $h'(y) = 1/g'(h(y)) > 0$이다. — 역함수의 미분

④ 연쇄법칙으로 $f_Y(y) = f_X(h(y))\,h'(y) = f_X(h(y))\,|h'(y)|$이다. — 1회 정리 6.7(d), ③

감소하는 경우

① $\{g(X) \le y\} = \{X \ge h(y)\}$이므로 $F_Y(y) = 1 - F_X(h(y))$이다. — 1회 정리 4.3(c), $P(X = h(y)) = 0$

② 미분하면 $f_Y(y) = -f_X(h(y))\,h'(y) = f_X(h(y))\,|h'(y)|$이다. — $h' < 0$

조건 확인 미분 $F_X' = f_X$는 f_X가 연속인 점에서만 보장된다(1회 정리 6.7(d)). 그래서 거꾸로 확인한다. 위의 f_Y를 $(-\infty, y]$에서 적분하고 $x = h(s)$로 치환하면 $F_X(h(y))$(감소하면 $1 - F_X(h(y))$)가 나온다 [T7]. 밀도는 적분이 분포함수가 되는 함수이므로 이것으로 충분하다. ∎

공식의 $|h'(y)|$를 **야코비안**이라 한다. 확률은 변환해도 보존된다. X가 폭 dx의 칸에 들 확률 $f_X(x)\,dx$는 Y가 그 칸의 상인 폭 $dy = |g'(x)|\,dx$의 칸에 들 확률 $f_Y(y)\,dy$와 같아야 한다. g가

늘이는 곳에서는 밀도가 그만큼 엷어지고 줄이는 곳에서는 짙어진다. 3회에서 다변수의 야코비안 행렬식으로 넓힌다.

따름정리 4.4 (함수의 기댓값 — 단조인 경우) 정리 4.3의 조건에서 $\int_S |g(x)| f_X(x)\,dx < \infty$이면 $EY = \int_S g(x) f_X(x)\,dx$이다. 이것으로 정리 1.9(b)가 단조인 g에서 증명된다.

증명. **전략** 정리 4.3의 밀도로 EY를 쓰고 $y = g(x)$로 치환한다 [T7]. 감소하는 경우 적분 구간의 방향이 바뀌는 것을 절댓값이 맞춘다. 단계별 증명은 연습 4.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 64쪽에 있다.

예제 4.5 (위치와 척도) $Y = a + bX$ $(b \neq 0)$이면 $h(y) = (y-a)/b$, $|h'| = 1/|b|$이므로

$$f_Y(y) = \frac{1}{|b|} f_X\!\left(\frac{y-a}{b}\right).$$

하나의 밀도 f에서 이렇게 만든 분포의 모임을 **위치–척도족**이라 한다. 균등분포 $\mathrm{U}(a,b)$는 $\mathrm{U}(0,1)$의 위치–척도족이고, $\mathrm{Exp}(\lambda)$는 $\mathrm{Exp}(1)$을 $\frac{1}{\lambda}$배 한 것이다. 거꾸로 평균 μ, 표준편차 σ인 X를 $Z = (X-\mu)/\sigma$로 **표준화**하면 $EZ = 0$, $\operatorname{Var} Z = 1$이다(정리 2.2).

4.3 확률적분변환과 역변환 표집

분포함수 F는 일반적으로 역함수가 없다(평평한 구간과 뛰는 점이 있다). 그래서 일반화된 역함수를 쓴다.

$$F^{-}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \qquad 0 < u < 1.$$

F가 연속이고 순증가이면 $F^{-} = F^{-1}$이다. $F^{-}(u)$는 F의 **u-분위수**이고 $F^{-}(\frac{1}{2})$가 중앙값이다.

보조정리 4.6 (분위수 함수의 기본 성질) $0 < u < 1$과 실수 x에 대해 $F^{-}(u) \leq x \iff u \leq F(x)$.

증명. ($\Leftarrow$) $u \leq F(x)$이면 x는 집합 $\{x' : F(x') \geq u\}$의 원소이므로 하한 $F^{-}(u)$는 x 이하다. ($\Rightarrow$) 먼저 집합 $\{x' : F(x') \geq u\}$가 비어 있지 않고 아래로 유계라는 것을 확인한다. $F \to 1 > u$ $(x \to \infty)$이므로 비어 있지 않고, $F \to 0 < u$ $(x \to -\infty)$이므로 아래로 유계다(1회 정리 6.3(b)). 하한의 정의로 $F(x_n) \geq u$이고 $x_n \downarrow F^{-}(u)$인 수열이 있다. F는 오른쪽 연속이므로(1회 정리 6.3(c)) $F(F^{-}(u)) = \lim F(x_n) \geq u$이다. $F^{-}(u) \leq x$이면 F가 비감소이므로 $F(x) \geq F(F^{-}(u)) \geq u$. ∎

정리 4.7 (역변환 표집과 확률적분변환)

(a) (역변환 표집) F가 **임의의** 분포함수이고 $U \sim \mathrm{U}(0,1)$이면 $X = F^{-}(U)$의 분포함수는 F이다.

(b) (확률적분변환) X의 분포함수 F가 연속이면 $F(X) \sim \mathrm{U}(0,1)$이다.

증명. **전략** (a)는 보조정리 4.6로 사건을 바꿔 쓰는 한 줄이다. (b)는 같은 보조정리의 대우로 $\{F(X) < u\}$를 X에 대한 사건으로 바꾸고, F의 연속성으로 $F(F^{-}(u)) = u$를 보인다 [T2].

(a)

① $\{F^{-}(U) \le x\} = \{U \le F(x)\}$이다. — 보조정리 4.6

② $P(U \le F(x)) = F(x)$이다. — $0 \le F(x) \le 1$, $U \sim \mathrm{U}(0,1)$

(b)

① $0 < u < 1$을 고정한다. 보조정리 4.6의 대우로 $F(x) < u \iff x < F^{-}(u)$이므로 $\{F(X) < u\} = \{X < F^{-}(u)\}$이다. — 대우 [T2]

② $P(F(X) < u) = P(X < F^{-}(u)) = F(F^{-}(u)-) = F(F^{-}(u))$이다. — 1회 정리 6.3(d), F 연속

③ $F(F^{-}(u)) \ge u$이다. — 보조정리 4.6의 증명

④ $x < F^{-}(u)$이면 $F(x) < u$이고, $x \uparrow F^{-}(u)$로 보내면 $F(F^{-}(u)) \le u$이다. ③과 합하면 $F(F^{-}(u)) = u$이다. — ①, F 연속 [T5]

⑤ $\{F(X) < u + \tfrac{1}{n}\} \downarrow \{F(X) \le u\}$이므로 $P(F(X) \le u) = \lim_n (u + \tfrac{1}{n}) = u$이다. — 1회 정리 4.7(b), ②④

⑥ $F(X)$의 분포함수는 $(0,1)$에서 u이므로 $F(X) \sim \mathrm{U}(0,1)$이다. — 1회 정리 6.3

■

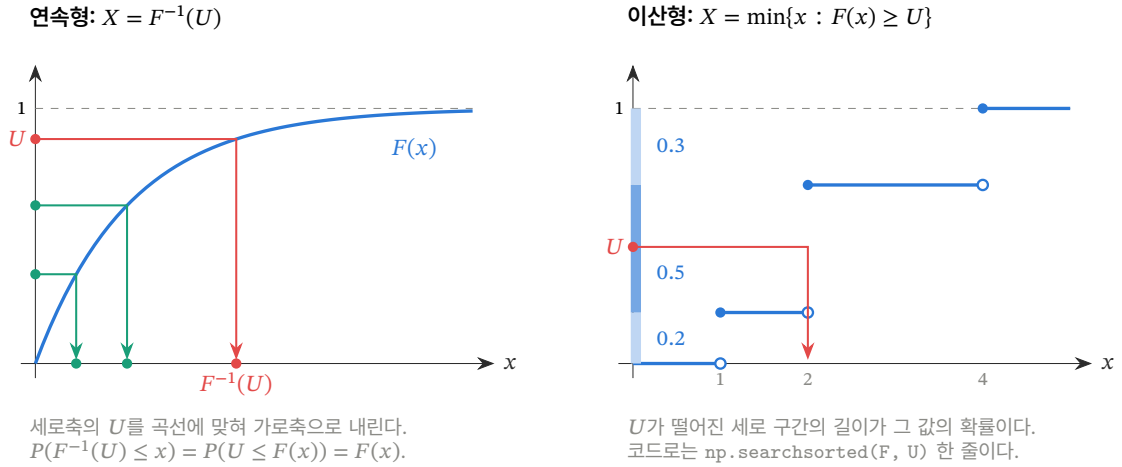


그림 4.1 역변환 표집(정리 4.7(a)). 왼쪽: 연속형은 세로축의 균등 난수를 곡선에 맞혀 가로축으로 내린다. 오른쪽: 이산형은 균등 난수가 떨어진 세로 구간의 길이가 그 값의 확률이다.

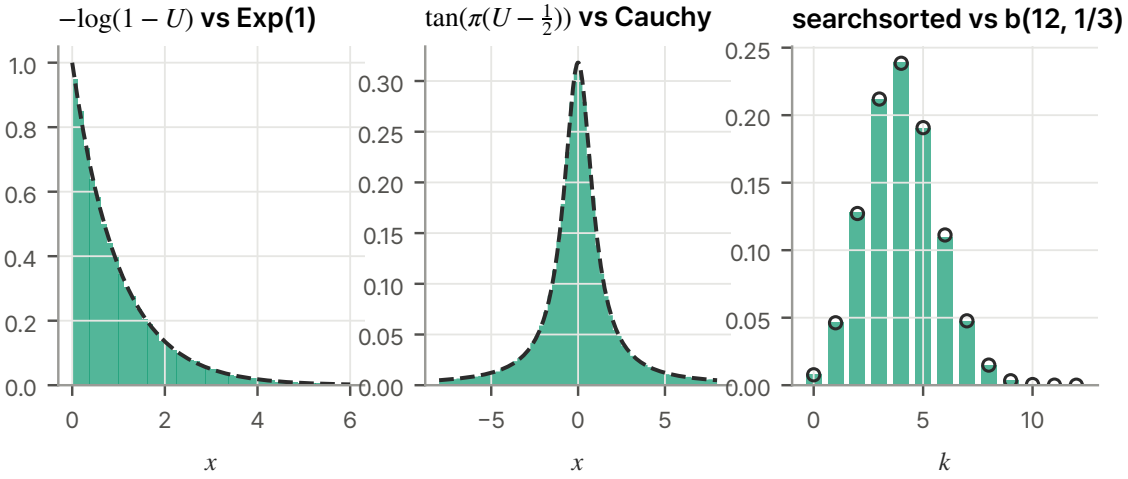


그림 4.2 균등 난수 10만 개를 역변환한 표본(청록)과 목표 분포(점선·원). 오른쪽은 공평한 주사위 12개 가운데 5·6이 나온 수의 분포 $\mathrm{b}(12, \frac{1}{3})$이다.

> **알고리즘 4.8** (역변환 표집) 분포함수 F를 따르는 난수를 만든다.
>
> 1. 균등 난수 $U \sim \mathrm{U}(0,1)$을 뽑는다.
> 2. $X = F^{-}(U)$를 돌려준다.
>
> 예. 지수분포 $X = -\frac{1}{\lambda}\log(1-U)$. Cauchy 분포 $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x$에서 $X = \tan\!\big(\pi(U-\tfrac{1}{2})\big)$. 기하분포 $X = \lceil \log(1-U)/\log q \rceil$. 값 $x_1 < x_2 < \cdots$와 누적확률 $c_j = \sum_{i \le j} p_i$를 갖는 이산분포는 $c_{j-1} < U \le c_j$인 j를 찾아 x_j를 돌려준다(`np.searchsorted`).

(b)는 거꾸로 쓰인다. 모형이 맞다면 관측값을 모형의 분포함수에 넣은 $F(X_i)$들이 $\mathrm{U}(0,1)$처럼 퍼져 있어야 한다. 1회 노트북에서 현의 길이에 쓴 Kolmogorov–Smirnov 비교, 일기예보와 기계

학습 모형의 보정(calibration) 검사가 이 원리다.

연습문제

연습 4.1 (지수분포를 두 가지로 ★) $U \sim \mathrm{U}(0,1)$일 때 $Y = -\frac{1}{\lambda}\log(1-U)$의 밀도를 정리 4.3으로 구하라. 예제 4.2의 $-\frac{1}{\lambda}\log U$와 같은 분포인 이유를 한 줄로 말하라. 풀이 63쪽

연습 4.2 (일대일이 아닌 제곱 ★★) $X \sim \mathrm{U}(-1,2)$이고 $Y = X^2$이다. $0 < y < 1$과 $1 \le y < 4$로 나눠 F_Y와 f_Y를 구하라. 풀이 63쪽

연습 4.3 (Benford의 법칙 ★★) 양수 X에 대해 $\log_{10} X$의 소수 부분이 $\mathrm{U}(0,1)$이라 하자. X의 첫 자리 숫자가 d일 확률이 $\log_{10}(1+\frac{1}{d})$임을 보여라. 1이 첫 자리일 확률은? 풀이 63쪽

연습 4.4 (Cauchy의 역수 ★★) X가 Cauchy 분포이면 $1/X$도 Cauchy 분포임을 보여라. (힌트: $g(x) = 1/x$는 $(-\infty,0)$과 $(0,\infty)$에서 각각 단조다.) 또 $\Theta \sim \mathrm{U}(-\frac{\pi}{2},\frac{\pi}{2})$이면 $\tan\Theta$가 Cauchy 분포임을 보여라. 원점에서 무작위 방향으로 쏜 빛이 벽 $x=1$에 닿는 높이다. 풀이 63쪽

연습 4.5 (혼합형 표집 ★) 1회의 보험금 X(확률 0.3으로 0, 나머지는 밀도 $0.7e^{-x}$)의 분포함수 F에 대해 $F^{-}(u)$를 구하라. $u \le 0.3$이면 얼마인가? 풀이 63쪽

연습 4.6 (Pareto 분포 ★) $\alpha > 0$에 대해 $F(x) = 1 - x^{-\alpha}$ $(x \ge 1)$인 분포를 Pareto 분포라 한다. 역변환 표집 공식을 쓰고, EX가 $\alpha > 1$일 때만 유한하며 그때 $\frac{\alpha}{\alpha-1}$임을, EX^2이 $\alpha > 2$일 때만 유한함을 보여라. 풀이 64쪽

연습 4.7 (확률적분변환은 연속이어야 ★) $X \sim \mathrm{Bernoulli}(\frac{1}{2})$이면 $F(X)$는 어떤 분포인가? 정리 4.7(b)의 연속성 가정이 왜 필요한가? 풀이 64쪽

연습 4.8 (함수의 기댓값, 단조인 경우: 증명 ★★ 유도형) 따름정리 4.4를 다음 순서로 증명하라. g가 증가하는 경우부터 한다.

(a) 정리 4.3의 밀도로 EY를 y에 대한 적분으로 쓰고, $y = g(x)$로 치환해 $\int_S g(x) f_X(x)\,dx$를 얻어라. $dy = g'(x)\,dx$와 $h'(g(x))\,g'(x) = 1$을 쓴다.

(b) 같은 치환을 $|y|$에 해서 $\int_T |y| f_Y(y)\,dy = \int_S |g(x)| f_X(x)\,dx < \infty$, 곧 EY가 존재함을 보여라.

(c) g가 감소하면 치환 뒤 적분 구간의 위끝과 아래끝이 바뀐다. $|h'| = -h'$이 그 방향을 되돌림을 보여라.

풀이 64쪽

5

부등식 — 기댓값으로 꼬리를 누르기

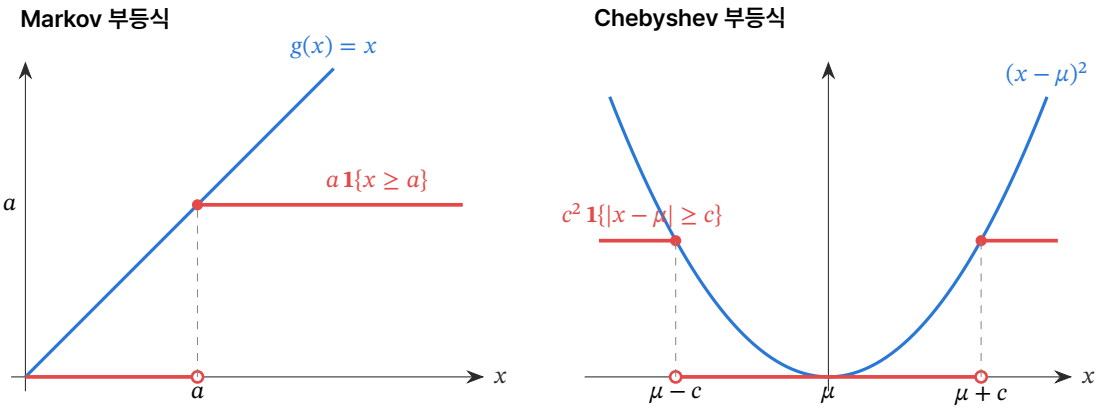
분포를 모르고 평균과 분산만 알 때, 확률변수가 평균에서 멀리 떨어질 확률에 대해 무엇을 말할 수 있나? 놀랍게도 꽤 많이 말할 수 있다. 이 절의 세 부등식은 모두 한 그림에서 나온다. 함수 하나를 다른 함수로 누르고, 양변의 기댓값을 취하는 것이다.

> **역사** Irénée-Jules Bienaymé는 1853년 Laplace의 최소제곱법을 옹호하는 논문에서, 독립인 오차의 합이 평균에서 벗어날 확률을 분산으로 누르는 부등식을 증명했다. Pafnuty Chebyshev는 1867년 "평균값에 대하여"에서 같은 부등식을 독립적으로 증명하고, 그것으로 대수의 법칙을 몇 줄에 증명했다. 같은 해 Liouville의 학술지는 Bienaymé의 1853년 논문을 Chebyshev의 논문 바로 앞에 다시 실었고, Chebyshev는 뒤에 Bienaymé의 우선권을 인정했다. 더 단순한 첫 부등식은 Chebyshev의 제자 Andrei Markov의 이름으로 불린다.

5.1 Markov와 Chebyshev

정리 5.1 (Markov 부등식) $X \geq 0$이고 $a > 0$이면

$$P(X \geq a) \leq \frac{EX}{a}.$$



$x \geq 0$에서 빨간 계단은 파란 직선 아래에 있다. 양변의 기댓값을 취하면 $a\,P(X \geq a) \leq EX$.

같은 그림을 $(X-\mu)^2$에 쓴다. 기댓값을 취하면 $c^2 P(|X-\mu| \geq c) \leq \operatorname{Var} X$.

그림 5.1 Markov와 Chebyshev 부등식의 그림. 빨간 계단 함수를 파란 함수가 위에서 누른다. 양변의 기댓값을 취하면 부등식이 된다.

증명. **전략** 지시함수를 x로 위에서 누르고(그림 5.1의 왼쪽), 양변의 기댓값을 취한다 [T4].

① 모든 $x \ge 0$에서 $a\,\mathbf{1}\{x \ge a\} \le x$이다. $x \ge a$이면 좌변이 $a \le x$이고, $x < a$이면 좌변이 $0 \le x$이다. — 경우 나누기

② $E\big[a\,\mathbf{1}\{X \ge a\}\big] \le EX$이다. — 단조성, 정리 1.10(b)

③ $E\big[a\,\mathbf{1}\{X \ge a\}\big] = a\,P(X \ge a)$이다. — 선형성, $E\,\mathbf{1}_A = P(A)$ (예제 1.2)

④ ②③을 $a > 0$으로 나누면 결론이다.

■

따름정리 5.2 (일반화된 Markov 부등식) $\varphi \ge 0$이 증가함수이고 $\varphi(a) > 0$이면 $P(X \ge a) \le E\varphi(X)/\varphi(a)$. 특히 mgf가 있으면 $t > 0$에 대해

$$P(X \ge a) \le e^{-ta} M(t) \qquad \text{(Chernoff 한계)}.$$

증명. $\{X \ge a\} \subset \{\varphi(X) \ge \varphi(a)\}$이므로 단조성과 Markov 부등식을 $\varphi(X)$에 쓴다. 둘째 식은 $\varphi(x) = e^{tx}$이다. ■

정리 5.3 (Chebyshev 부등식) $\operatorname{Var} X = \sigma^2 < \infty$이고 $\mu = EX$이면 모든 $c > 0$에 대해

$$P\big(|X - \mu| \ge c\big) \le \frac{\sigma^2}{c^2}, \qquad \text{곧} \qquad P\big(|X - \mu| \ge k\sigma\big) \le \frac{1}{k^2}.$$

증명. **전략** 거리의 사건을 제곱의 사건으로 바꾸고, 음이 아닌 확률변수 $(X-\mu)^2$에 Markov 부등식을 쓴다.

① $c > 0$이므로 $\{|X - \mu| \ge c\} = \{(X-\mu)^2 \ge c^2\}$이다. — $t \mapsto t^2$은 $t \ge 0$에서 증가 [T1]

② $P\big((X-\mu)^2 \ge c^2\big) \le E(X-\mu)^2/c^2 = \sigma^2/c^2$이다. — 정리 5.1, 정의 2.1

③ $c = k\sigma$를 넣으면 둘째 식이다.

■

분포가 무엇이든 평균에서 표준편차의 2배 이상 떨어질 확률은 $\frac{1}{4}$ 이하, 3배 이상은 $\frac{1}{9}$ 이하다. 대부분의 분포에서 실제 확률은 이보다 훨씬 작다(지수분포에서 2배 이상은 $e^{-3} \approx 0.050$). 그러나 이 한계는 더 줄일 수 없다.

예제 5.4 (Chebyshev 한계는 정확하다) $k > 1$에 대해 X가 $-k$와 k를 각각 확률 $\frac{1}{2k^2}$로, 0을 확률 $1 - \frac{1}{k^2}$로 가지면 $\mu = 0$, $\sigma^2 = k^2 \cdot \frac{1}{k^2} = 1$이고 $P(|X| \geq k\sigma) = \frac{1}{k^2}$이다. 등호가 성립한다. 평균과 분산만으로는 이보다 나은 한계를 말할 수 없다. 더 나은 한계를 원하면 더 많은 정보가 필요하다. 8절의 Hoeffding 부등식은 "값이 $[0,1]$ 안에 있다"는 정보를 쓴다.

예제 5.5 (Weldon의 주사위는 공평했나) 예제 2.11에서 공평한 주사위 모형 아래 $\bar{K}$는 평균 4, 분산 $\frac{8/3}{26306}$을 갖는다. Chebyshev 부등식으로

$$P\big(|\bar{K} - 4| \geq 0.0522\big) \leq \frac{(8/3)/26306}{0.0522^2} \approx 0.037.$$

공평한 주사위라면 이만큼 벗어나는 일은 많아야 100번에 4번이다. 분포를 전혀 가정하지 않고 얻은 결론이다. 7회의 정규근사로는 이 확률이 2×10^{-7} 정도다. Weldon의 주사위는 5·6이 조금 더 잘 나왔다. 흔히 드는 설명은 이렇다. 눈을 파서 새긴 주사위는 눈이 많은 면이 조금 가벼워서, 그 면이 위로 오기 쉽다.

따름정리 5.6 (분산이 0이면 상수) $\operatorname{Var} X = 0$이면 $P(X = \mu) = 1$이다.

증명. **전략** Chebyshev 부등식을 $c = \frac{1}{n}$마다 쓰고, 사건열의 극한에 1회 정리 4.7(연속성)을 쓴다 [T5]. 단계별 증명은 연습 5.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 65쪽에 있다.

5.2 Jensen

정의 5.7 (볼록함수) 구간 I 위의 함수 φ가 모든 $x, y \in I$와 $0 \leq \lambda \leq 1$에 대해

$$\varphi\big(\lambda x + (1-\lambda)y\big) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y)$$

를 만족하면 **볼록**이라 한다. 그래프 위의 두 점을 잇는 현이 그래프 위에 있다는 뜻이다. $-\varphi$가 볼록이면 φ를 **오목**이라 한다.

φ가 두 번 미분가능하면 $\varphi'' \geq 0$일 때 볼록이다(연습 5.6). x^2, e^x, $|x|$, $x\log x$는 볼록이고, $\log x$와 $\sqrt{x}$는 오목이다.

보조정리 5.8 (받침선) φ가 열린 구간 I에서 볼록이고 $\mu \in I$이면, 어떤 기울기 b에 대해

$$\varphi(x) \geq \varphi(\mu) + b\,(x-\mu) \qquad \text{모든 } x \in I.$$

증명. **1단계: 기울기는 오른쪽으로 갈수록 크다.** $x < \mu < y$이면 $\mu = \lambda x + (1-\lambda)y$, $\lambda = \frac{y-\mu}{y-x}$이다. 볼록성에 넣고 정리하면

$$\frac{\varphi(\mu)-\varphi(x)}{\mu - x} \leq \frac{\varphi(y)-\varphi(\mu)}{y-\mu}.$$

(양변에 $(\mu-x)(y-\mu)>0$을 곱하면 볼록성의 식과 같아진다. 연습 5.5.)

2단계: 사이의 기울기. 왼쪽 기울기들 $\frac{\varphi(\mu)-\varphi(x)}{\mu-x}$ $(x<\mu)$는 어떤 오른쪽 기울기보다도 작으므로 위로 유계다. 그 상한을 b라 하자. 그러면 모든 $x<\mu<y$에 대해

$$\frac{\varphi(\mu)-\varphi(x)}{\mu-x} \leq b \leq \frac{\varphi(y)-\varphi(\mu)}{y-\mu}.$$

왼쪽 부등식에 $\mu - x > 0$을 곱하면 $\varphi(x) \geq \varphi(\mu) + b(x-\mu)$, 오른쪽 부등식에 $y-\mu>0$을 곱하면 $\varphi(y) \geq \varphi(\mu)+b(y-\mu)$이다. $x=\mu$에서는 등호다. (I가 열린 구간이어서 μ의 양쪽에 점이 있다.) ∎

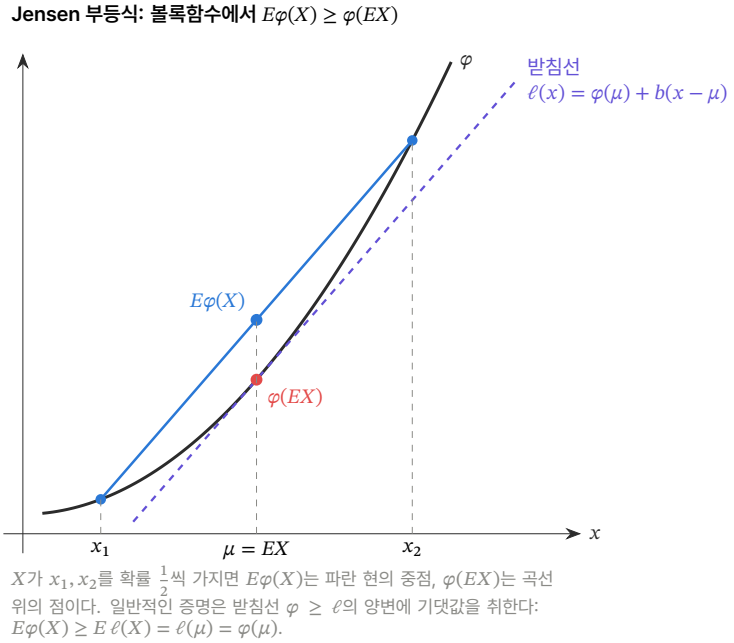


그림 5.2 Jensen 부등식. 두 점 분포에서 $E\varphi(X)$는 현의 중점이고 $\varphi(EX)$는 그 아래 곡선 위의 점이다. 일반적인 증명은 받침선 ℓ(보조정리 5.8)을 쓴다.

정리 5.9 (Jensen 부등식) φ가 열린 구간 I에서 볼록이고, $P(X \in I) = 1$, $E|X| < \infty$, $E|\varphi(X)| < \infty$이면

$$\varphi(EX) \le E\,\varphi(X).$$

φ가 오목이면 부등호가 반대다.

증명. **전략** 볼록함수의 그래프는 한 점에서의 받침선보다 위에 있다(보조정리 5.8). 그 점을 $\mu = EX$로 잡고 양변의 기댓값을 취하면, 받침선은 일차식이라 기댓값이 $\varphi(\mu)$가 된다.

① $\mu = EX$는 $I = (\alpha, \beta)$ 안에 있다(α, β는 $\pm\infty$일 수 있다). 확률 1로 $\alpha < X < \beta$이므로 $\alpha \le \mu \le \beta$이고, $\mu = \alpha$이면 $X - \alpha > 0$인데 $E(X-\alpha) = 0$이다. Markov 부등식으로 모든 n에서 $P(X - \alpha \ge \frac{1}{n}) \le 0$이고, 연속성으로 $P(X > \alpha) = 0$이 되어 모순이다. $\mu = \beta$도 같다. 정리 5.1, 1회 정리 4.7

② 받침선 $\ell(x) = \varphi(\mu) + b(x-\mu)$가 있어 I에서 $\varphi \ge \ell$이다. 보조정리 5.8, ①

③ $P(X \in I) = 1$이므로 $E\,\varphi(X) \ge E\,\ell(X)$이다. 단조성, 정리 1.10(b)

④ $E\,\ell(X) = \varphi(\mu) + b\,(EX - \mu) = \varphi(\mu)$이다. 선형성, 정리 1.10(a)

⑤ 오목이면 $-\varphi$에 ①–④를 쓴다. 정의 5.7

조건 확인 ③에서 단조성을 쓰려면 두 기댓값이 존재해야 한다. $E|\varphi(X)| < \infty$는 가정이고, $E|\ell(X)| \le |\varphi(\mu)| + |b|\big(E|X| + |\mu|\big) < \infty$이다. $E|X| < \infty$라는 가정이 여기서 쓰인다. ∎

따름정리 5.10 (Jensen의 따름정리들)

(a) $(EX)^2 \le EX^2$, $|EX| \le E|X|$.

(b) $X > 0$이면 $E\log X \le \log EX$, $E\frac{1}{X} \ge \frac{1}{EX}$.

(c) (산술–기하 평균) $a_1, \dots, a_n > 0$이면 $\sqrt[n]{a_1 \cdots a_n} \le \frac{1}{n}(a_1 + \cdots + a_n)$.

(d) (Lyapunov) $0 < r < s$이면 $\big(E|X|^r\big)^{1/r} \le \big(E|X|^s\big)^{1/s}$.

증명. **전략** 식마다 알맞은 볼록함수나 오목함수와 확률변수를 골라 정리 5.9에 넣는다. (c)는 값 a_i들을 같은 확률로 갖는 확률변수를 만든다. 단계별 증명은 연습 5.9(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 65쪽에 있다.

예제 5.11 (오목한 효용 — 위험 회피) 돈 w의 "쓸모"를 증가하는 오목함수 $u(w)$로 재는 사람은 불확실한 금액 W보다 그 기댓값을 확실하게 받는 쪽을 언제나 좋아한다. Jensen으로 $E\,u(W) \le u(EW)$이기 때문이다. $E\,u(W) = u(c)$인 확실한 금액 c를 **확실성 등가**라 하면 $c \le EW$이고, 그 차이가 **위험 프리미엄**이다. 7절에서 Daniel Bernoulli가 이 생각으로 St. Petersburg 역설을 푸는 것을 본다.

연습문제

연습 5.1 (Markov도 정확하다 ★) $a>0$에 대해 등호 $P(X\ge a)=EX/a$가 성립하는 음이 아닌 X를 하나 만들어라. 언제 등호가 성립하는가? 풀이 64쪽

연습 5.2 (한계와 실제 ★) $k=1.5,2,3$에 대해 $P(|X-\mu|\ge k\sigma)$를 $\mathrm{U}(0,1)$과 $\mathrm{Exp}(1)$에서 정확히 구하고 $1/k^2$과 비교하라. 풀이 64쪽

연습 5.3 (한쪽 꼬리 — Cantelli ★★★) $c>0$이면 $P(X-\mu\ge c)\le \dfrac{\sigma^2}{\sigma^2+c^2}$임을 보여라. (힌트: 모든 $s\ge 0$에서 $P(X-\mu\ge c)\le P\big((X-\mu+s)^2\ge (c+s)^2\big)$에 Markov를 쓰고 s에 대해 최소화한다.) $k=2$이면 한쪽 꼬리는 $\frac{1}{5}$ 이하다. 풀이 64쪽

연습 5.4 (평균과 중앙값은 σ 이내 ★★) 연속 확률변수의 중앙값 m과 평균 μ에 대해 $|\mu-m|\le\sigma$임을 보여라. (힌트: $|\mu-m|=|E(X-m)|\le E|X-m|\le E|X-\mu|\le\sigma$. 각 부등호의 이유를 대라. 연습 2.2을 쓴다.) 풀이 65쪽

연습 5.5 (기울기의 단조성 ★★) 보조정리 5.8의 1단계 부등식을 볼록성의 정의에서 직접 유도하라. 풀이 65쪽

연습 5.6 (두 번 미분가능한 볼록함수 ★★) $\varphi''\ge 0$이면 φ가 볼록임을 보여라. (힌트: $z=\lambda x+(1-\lambda)y$에서의 접선 $\varphi(z)+\varphi'(z)(t-z)$가 테일러 정리로 그래프 아래에 있음을 보이고, $t=x,y$를 넣어 $\lambda,1-\lambda$를 곱해 더한다.) 풀이 65쪽

연습 5.7 (평균 속력 ★) 같은 거리를 n번 가는데 i번째에 속력 v_i로 갔다. 전체의 평균 속력은 v_i들의 조화평균임을 보이고, 따름정리 5.10(b)로 산술평균보다 크지 않음을 보여라. 60km/h로 가고 40km/h로 돌아오면 평균 속력은? 풀이 65쪽

연습 5.8 (분산이 0이면 상수: 증명 ★ 유도형) 따름정리 5.6를 다음 순서로 증명하라.

(a) Chebyshev 부등식으로 모든 n에서 $P(|X-\mu|\ge\frac{1}{n})=0$임을 보여라.

(b) 사건 $A_n=\{|X-\mu|\ge\frac{1}{n}\}$이 증가하고 $\bigcup_n A_n=\{X\ne\mu\}$임을 원소 추적으로 보여라.

(c) 1회 정리 4.7로 $P(X\ne\mu)=0$을 결론지어라. 같은 논리로 "$X\ge 0$이고 $EX=0$이면 $P(X=0)=1$"도 보여라.

풀이 65쪽

연습 5.9 (Jensen의 따름정리들: 증명 ★★ 유도형) 따름정리 5.10를 다음 순서로 증명하라. 각 단계에서 쓰는 볼록함수(또는 오목함수)와 구간 I를 밝혀라.

(a) $\varphi(x)=x^2$과 $\varphi(x)=|x|$로 (a)를 얻어라.

(b) $\log$가 $(0,\infty)$에서 오목이고 $1/x$가 볼록임을 이계도함수로 확인하고(연습 5.6) (b)를 얻어라.

(c) X가 값 $a_1, \dots, a_n$을 각각 확률 $\frac{1}{n}$로 갖는다고 하고 (b)의 첫 식에서 (c)를 얻어라. a_i들 가운데 같은 값이 있어도 되는가?

(d) $Y = |X|^r$, $\varphi(y) = |y|^{s/r}$로 (d)를 얻어라. $E|X|^r < \infty$는 어디서 나오는가?

풀이 65쪽

6

대수의 법칙 — Bernoulli의 황금정리

이제 여는 질문으로 돌아간다. 필요한 도구는 모두 갖추었다. 관측 비율의 평균과 분산(따름정리 2.5), 그리고 분산으로 꼬리를 누르는 Chebyshev 부등식(정리 5.3)이다.

6.1 Ars Conjectandi

> **역사** Jacob Bernoulli(1654–1705)는 바젤 대학의 수학 교수였다. 그는 *Ars Conjectandi*의 4부에서 확률을 도박 밖으로 끌어내려 했다. 질병, 날씨, 사람의 수명처럼 "같은 정도로 일어나는 경우"를 셀 수 없는 문제에서는 확률을 **사전에** 알 수 없다. 그는 이렇게 썼다. "사전에 얻을 수 없는 것을 적어도 사후에, 곧 비슷한 많은 예에서 관측한 결과로부터 얻을 수 있다." 이것이 정당하려면 관측을 늘릴수록 관측한 비율이 참 비율에 다가간다는 것을 증명해야 했다. 그는 이 정리를 "황금정리"라 불렀고, 이 문제를 20년 동안 붙들고 있었다고 책에 썼다. "대수의 법칙"(la loi des grands nombres)이라는 이름은 1837년 Poisson이 붙였다.

Bernoulli가 원한 것은 "도덕적 확실성"이었다. 1,000 대 1 이상으로 확실하면, 곧 확률이 $\frac{1000}{1001}$ 이상이면 사람은 그것을 확실한 것으로 여기고 행동한다는 것이다. 그의 질문은 이것이다. 흰 돌의 참 비율이 $p = \frac{3}{5}$일 때, 몇 번 꺼내면

$$P\Big(\frac{29}{50} \le \hat{p}_n \le \frac{31}{50}\Big) \ge \frac{1000}{1001}$$

인가? 오차의 폭은 $\varepsilon = \frac{1}{50} = 0.02$이다.

6.2 확률수렴과 Bernoulli의 정리

정의 6.1 (확률수렴) 확률변수의 수열 $Y_1, Y_2, \dots$와 상수 c가 모든 $\varepsilon > 0$에 대해

$$\lim_{n\to\infty} P\big(|Y_n - c| \ge \varepsilon\big) = 0$$

을 만족하면 Y_n이 c로 **확률수렴**한다고 하고 $Y_n \xrightarrow{P} c$로 쓴다.

Y_n이 확률변수이므로 "$Y_n \to c$"를 수열의 극한으로 말할 수 없다. 대신 c에서 ε 이상 떨어질 **확률**이 0으로 간다고 말한다. 7회에서 다른 수렴들과 비교한다.

정리 6.2 (Bernoulli의 대수의 법칙) $\hat{p}_n$이 성공 확률 p인 독립 시행 n번의 성공 비율이면 모든 $\varepsilon > 0$에 대해

$$P\big(|\hat{p}_n - p| \ge \varepsilon\big) \le \frac{pq}{n\varepsilon^2} \le \frac{1}{4n\varepsilon^2}.$$

따라서 $\hat{p}_n \xrightarrow{P} p$이다.

증명. 따름정리 2.5로 $E\hat{p}_n = p$, $\operatorname{Var}\hat{p}_n = \frac{pq}{n}$이다. Chebyshev 부등식에 $c = \varepsilon$을 넣고, $pq \le \frac{1}{4}$(예제 2.3)를 쓴다. ∎

두 줄이다. Bernoulli가 20년을 들인 정리가 분산이라는 개념 하나로 두 줄이 되었다. 그러나 이 짧은 증명이 주는 시행 횟수는 Bernoulli의 것보다 **나쁘다**. Chebyshev 한계로 오른쪽이 $\frac{1}{1001}$ 이하가 되려면

$$n \ \ge\ \frac{pq}{\varepsilon^2}\cdot 1001 = \frac{0.24}{0.0004}\cdot 1001 = 600{,}600.$$

6.3 Bernoulli의 계산

Bernoulli는 분산을 몰랐다. 그는 이항분포의 항 $T_k = \binom{N}{k}p^k q^{N-k}$를 직접 다뤘다. 핵심은 이웃한 두 항의 비

$$\frac{T_{k+1}}{T_k} = \frac{N-k}{k+1}\cdot\frac{p}{q}$$

가 k에 대해 감소한다는 것이다. 따라서 항들은 최빈값 근처에서 가장 크고, 거기서 멀어지면 **등비급수보다 빨리** 줄어든다. 그는 중심에서 εN 이상 떨어진 꼬리의 합을 중심 근처 항들의 합과 비교해, 그 비를 등비급수로 눌렀다. 연습 6.7에서 이 생각으로 정리 6.2의 결론을 Chebyshev 없이 증명한다.

$p = \frac{r}{t}$, $r = 30$, $s = 20$, $t = r + s = 50$, $\varepsilon = \frac{1}{t}$로 놓은 그의 결론은 다음 꼴이다. 한쪽 꼬리에 대해

$$m_1 \ge \frac{\log\big(c(s-1)\big)}{\log(r+1) - \log r}, \qquad N_1 = m_1 t + \frac{m_1 s t - s t}{r+1},$$

다른 쪽 꼬리에 대해 r과 s를 바꾼 m_2, N_2를 구하면, $N \ge \max(N_1, N_2)$번 시행에서 관측 비율이 $\frac{r-1}{t}$와 $\frac{r+1}{t}$ 사이에 들 확률이 밖에 들 확률의 c배를 넘는다. $c = 1000$이면

$$m_1 = 301,\ \ N_1 = 24{,}728; \qquad m_2 = 211,\ \ N_2 = 25{,}550.$$

이것이 25,550이다. 그는 이어서 $c = 10^4$이면 31,258번, $c = 10^5$이면 36,966번이라고 쓰고, c가 10배 될 때마다 5,708번씩 더하면 된다고 덧붙였다. 확실성을 높이는 비용은 **로그**로만 늘어난다. 이 관찰은 옳았고, 8절의 지수적 한계가 그 이유를 말해 준다.

6.4 같은 질문의 다섯 답

방법	시행 수	쓰는 정보
Chebyshev (1867)	600,600	평균과 분산 (정리 6.2)
Bernoulli (1713)	25,550	이항분포의 항의 비
Hoeffding (1963)	9,503	값이 $[0,1]$ 안에 있다는 것 (8절, 정리 8.2)
정규근사	6,498	중심극한정리 (7회): $n = z^2 pq/\varepsilon^2$, $z = 3.29$
정확한 계산	6,450	이항분포의 합. 6,520 이상이면 언제나 성립

표는 이 과정 전체의 지도이기도 하다. 위로 갈수록 가정이 적고 한계가 느슨하다. Chebyshev는 분포에 대해 평균과 분산만 알면 되고, 그래서 가장 나쁜 분포(예제 5.4)까지 견뎌야 한다. 아래로 갈수록 분포를 더 많이 쓰고 답이 정확해진다. 정확한 계산은 분포를 완전히 알아야 하고, 정규근사는 "n이 크다"는 근사를 받아들인다. 정확한 계산이 주는 6,450은 Bernoulli의 답의 약 4분의 1이다.

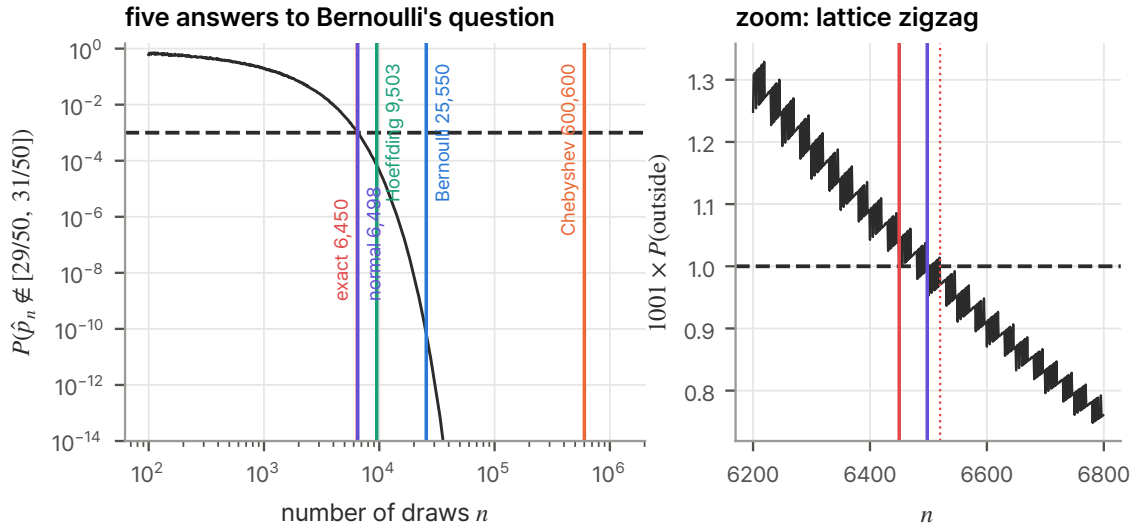


그림 6.1 여는 질문의 정확한 답. 곡선은 $P\left(\frac{29}{50} \le \hat{p}_n \le \frac{31}{50}\right)$을 이항분포로 정확히 계산한 것이다. n이 바뀌면 구간에 드는 정수의 개수가 달라져서 곡선이 톱니 모양이다. 점선은 $\frac{1000}{1001}$, 세로선은 다섯 가지 답이다.

6.5 일반적인 대수의 법칙

정리 6.3 (약한 대수의 법칙) $X_1, X_2, \dots$ 가 독립이고 평균이 모두 μ, 분산이 모두 $\sigma^2 < \infty$ 이면, 표본평균 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 에 대해

$$P\left(|\bar{X}_n - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0, \qquad \text{곧} \qquad \bar{X}_n \xrightarrow{P} \mu.$$

증명. **전략** 표본평균의 평균과 분산을 구하고 Chebyshev 부등식에 넣는다. 오른쪽이 n에 반비례하므로 0으로 간다.

① $E\bar{X}_n = \frac{1}{n}\sum_i EX_i = \mu$ 이다. — 선형성, 정리 1.11

② $\operatorname{Var}\bar{X}_n = \frac{\sigma^2}{n}$ 이다. — 정리 2.8

③ $P\left(|\bar{X}_n - \mu| \geq \varepsilon\right) \leq \frac{\operatorname{Var}\bar{X}_n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$ 이다. — 정리 5.3, ① ②

④ ε을 고정하고 $n \to \infty$ 이면 오른쪽은 0으로 가므로 $\bar{X}_n \xrightarrow{P} \mu$ 이다. — 정의 6.1 [T5]

■

증명은 독립합의 분산 공식(정리 2.8)만 쓴다. 이 노트에서 그 공식은 이산 확률변수에서 증명했고, 연속 확률변수에서는 3회에서 증명한다. 그러면 이 정리와 증명이 그대로 연속 확률변수에도 성립한다. 아래 예제의 연속 확률변수는 그 결과를 미리 쓴다. 분산이 없어도 평균만 있으면 결론이 성립한다는 것(Khinchin, 1929)은 7회에서 증명한다.

예제 6.4 (Monte Carlo 적분) $\int_0^1 g(x)\,dx$를 구하고 싶다. $U \sim \mathrm{U}(0,1)$ 이면 정리 1.9로 $E\,g(U) = \int_0^1 g$ 이므로, 독립인 균등 난수 $U_1, \dots, U_n$ 으로

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} g(U_i) \xrightarrow{P} \int_0^1 g(x)\,dx$$

이고, 오차의 표준편차는 $\sigma_g/\sqrt{n}$, $\sigma_g^2 = \operatorname{Var} g(U)$ 이다. 차원이 높아져도 오차의 차수 $1/\sqrt{n}$은 그대로다. 격자 위의 수치적분은 차원이 높아지면 격자점이 기하급수적으로 늘지만, Monte Carlo는 그렇지 않다. 1940년대 Los Alamos에서 이 방법에 Monte Carlo라는 이름이 붙었다(7회).

예로 $g(x) = e^{x^2}$ 이면 $\int_0^1 g \approx 1.4627$ 이고 $\sigma_g^2 \approx 0.225$ 이다. 오차가 0.01 이상일 확률을 1% 이하로 하려면 Chebyshev로 $n \geq \frac{0.225}{0.01^2 \cdot 0.01} \approx 2.3 \times 10^5$ 이면 충분하다.

6.6 상대도수와 확률 — 다시 데이터로

1회 2절에서 Halley의 생명표의 상대도수를 확률로 읽었다. 대수의 법칙이 그 근거다. 사건 A에 대해 독립인 반복에서 $\mathbf{1}_A$의 표본평균이 상대도수이고, 그 평균이 $P(A)$이므로 상대도수는 $P(A)$로 확률수렴한다. 그러나 이 법칙에는 가정이 있다. **독립이고 같은 확률**이라는 가정이다. 실제 자료가 이 가정을 따르는지는 따로 확인해야 한다.

> **예제 6.5** (Arbuthnot의 남아 출생 비율) John Arbuthnot은 1710년 런던의 1629–1710년 82년간 세례 기록에서 해마다 남자아이가 여자아이보다 많다는 것을 보였다(그는 이것을 신의 섭리의 증거로 읽었다. 8회에서 그의 논증이 최초의 유의성 검정이었음을 본다). 82년 전체의 남아 비율은 $\hat{p} = 484{,}382/938{,}223 = 0.5163$이다.
>
> 해마다 약 11,400명이 태어났으므로, 출생마다 독립이고 남아 확률이 해마다 같다면 한 해의 남아 비율의 표준편차는 따름정리 2.5로 $\sqrt{\hat{p}(1-\hat{p})/11{,}400} \approx 0.0048$이어야 한다. 실제로 82개 연도 비율의 표준편차는 0.0072로 1.5배 크다. 1879년 Wilhelm Lexis는 이 비를 **분산비**(Lexis ratio)라 부르고, 1보다 크면 확률이 해마다 일정하지 않다는 증거로 읽었다. 대수의 법칙은 평균이 어디로 가는지를 말하고, 분산은 그 모형이 맞는지를 시험하게 해 준다.

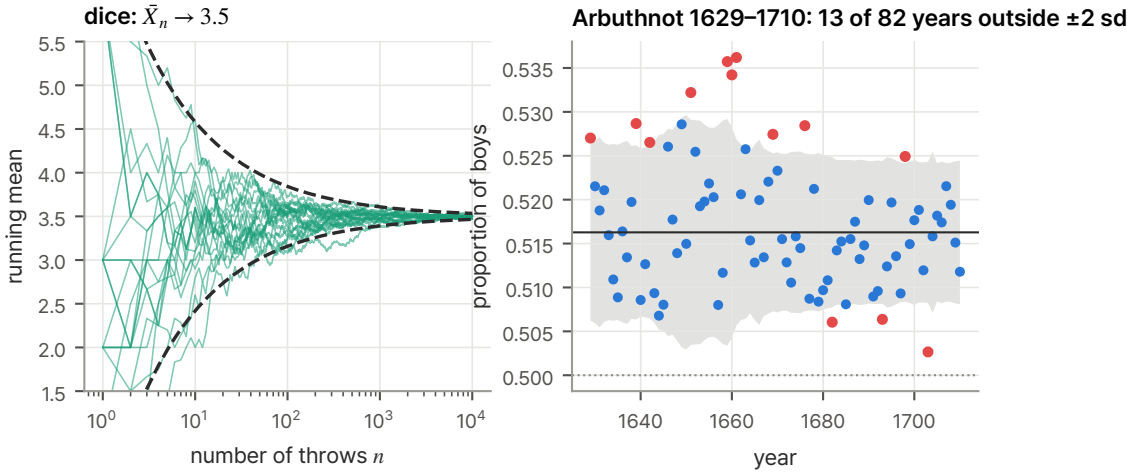


그림 6.2 왼쪽: 주사위를 던진 누적 평균 20개 경로와 $3.5 \pm 2\sigma/\sqrt{n}$ (점선). 오른쪽: Arbuthnot 자료의 해마다 남아 비율. 회색 띠는 출생이 독립이고 확률이 일정하다는 모형 아래 ±2 표준편차다. 띠를 벗어나는 해가 모형이 예측하는 5%보다 훨씬 많다.

연습문제

연습 6.1 (여론조사의 표본 크기 ★) 지지율을 ±3%포인트 안으로, 틀릴 확률 5% 이하로 알고 싶다. 정리 6.2의 한계 $\frac{1}{4n\varepsilon^2}$로 필요한 n을 구하라. (정규근사로는 1,068명이다.) 풀이 66쪽

연습 6.2 (Bernoulli의 공식 ★) 본문의 공식으로 $c = 10{,}000$일 때 m_1, N_1, m_2, N_2를 계산하고 Bernoulli의 31,258과 비교하라. c가 10배 될 때 m_2는 얼마씩 늘어나는가? 풀이 66쪽

연습 6.3 (주사위 평균 ★) 주사위를 n번 던진 평균이 3.5에서 0.1 이상 벗어날 확률을 5% 이하로 하려면 Chebyshev로 n이 얼마면 충분한가? 풀이 66쪽

연습 6.4 (쌍별 독립이면 충분하다 ★★) 정리 6.3에서 "독립"을 "쌍별 독립"으로 바꿔도 결론이 성립함을 보여라. 증명의 어느 줄을 바꿔야 하나? 풀이 66쪽

연습 6.5 (Poisson의 대수의 법칙 ★★) 독립 시행의 성공 확률이 시행마다 다르다(i번째는 p_i). 성공 수 K_n에 대해 $\frac{1}{n}\big(K_n - \sum_{i\le n} p_i\big) \xrightarrow{P} 0$임을 보여라. Poisson(1837)이 "대수의 법칙"이라는 이름으로 부른 것은 이 형태였다. 풀이 66쪽

연습 6.6 (Lexis의 분산비 ★) 노트북에서 Arbuthnot 자료의 해마다 남아 비율 $\hat p_t$와 출생 수 n_t로 $Q^2 = \frac{1}{81}\sum_t \frac{(\hat p_t - \hat p)^2}{\hat p(1-\hat p)/n_t}$를 계산하라. 모형이 맞으면 Q^2은 1 근처여야 한다. 값은? 어떤 이유들로 1보다 클 수 있나? 풀이 67쪽

연습 6.7 (Bernoulli식 증명 ★★★) $K \sim \mathrm{b}(N, p)$, $T_k = P(K = k)$, $0 < \varepsilon < q$이다. 분산을 쓰지 않고 $P(\hat p_N \ge p + \varepsilon) \to 0$을 보인다. (a) $\frac{T_{k+1}}{T_k} = \frac{(N-k)p}{(k+1)q}$임을 보이고, $k \ge Np$이면 이 비가 1보다 작음을 보여라. 따라서 T_k는 $k \ge Np$에서 감소한다. (b) $j = \lceil N(p+\varepsilon)\rceil$이면 $k \ge j$에서 $\frac{T_{k+1}}{T_k} \le \rho := \frac{(q-\varepsilon)p}{(p+\varepsilon)q} < 1$임을 보이고, $P(K \ge j) \le \frac{T_j}{1-\rho}$를 결론지어라. (c) $Np \le k \le j$인 정수 k는 적어도 $N\varepsilon - 1$개이고 그 T_k는 모두 T_j 이상이다. $\sum_k T_k \le 1$에서 $T_j \le \frac{1}{N\varepsilon - 1}$을 얻고 결론을 내려라. (힌트: (b)에서 $k+1 > N(p+\varepsilon)$와 $N-k \le N(q-\varepsilon)$을 쓴다.) 풀이 67쪽

7

부쉬보기 — 가정을 빼면

이 회차의 정리들에는 모두 "기댓값이 존재하면", "분산이 유한하면", "mgf가 0의 근방에서 존재하면", "독립이면" 이라는 조건이 붙어 있다. 조건을 하나씩 빼 본다.

7.1 기댓값이 무한대 — St. Petersburg 역설

> **역사** Nicolaus Bernoulli(Jacob의 조카)는 1713년 Montmort에게 보낸 편지에서 다음 내기를 물었다. 공평한 동전을 앞면이 나올 때까지 던져, k번째에 처음 앞면이 나오면 2^k 두카트를 받는다. 이 내기에 들어가는 값으로 얼마를 내겠는가? 25년 뒤 그의 사촌 Daniel Bernoulli가 상트페테르부르크 학술원의 논문집(1738)에 해법을 실었고, 문제는 그 도시의 이름을 얻었다.

상금 $X = 2^K$, $P(K=k) = 2^{-k}$ 이므로

$$E X = \sum_{k=1}^{\infty} 2^k \cdot 2^{-k} = 1 + 1 + 1 + \cdots = \infty.$$

기댓값으로 값을 매기면(Huygens의 규칙) 어떤 금액을 내도 싸다. 그러나 실제로 20두카트를 내겠다는 사람도 드물다. 상금의 중앙값은 2두카트이고, 4두카트 이하일 확률이 $\frac{3}{4}$이다.

Daniel Bernoulli의 해법은 돈의 **쓸모**가 금액에 비례하지 않는다는 것이었다. 가진 돈이 늘수록 같은 1두카트의 쓸모는 줄어든다. 그는 쓸모의 증가가 가진 돈에 반비례한다고 놓아 $u(w) = \log w$를 얻었다. 그러면

$$E \log_2 X = \sum_{k=1}^{\infty} k\, 2^{-k} = 2$$

이므로 확실성 등가(예제 5.11)는 $2^2 = 4$두카트다. 기댓값은 무한대인데 이 내기의 값은 4두카트라는 것이다.

대수의 법칙은 어떻게 되나? 평균이 무한대이므로 정리 6.3을 쓸 수 없고, 실제로 $\bar X_n$은 어떤 상수로도 수렴하지 않는다. 그림 7.1에서 평균 상금은 $\log_2 n$처럼 자란다. Feller(1945)는 이것을 정확히 했다. $\bar X_n / \log_2 n \xrightarrow{P} 1$이다(연습 7.3). 1,000판의 공평한 참가비는 판마다 약 10두카트, 100만 판이면 약 20두카트다. 공평한 값이 판의 수에 따라 달라지는 내기다.

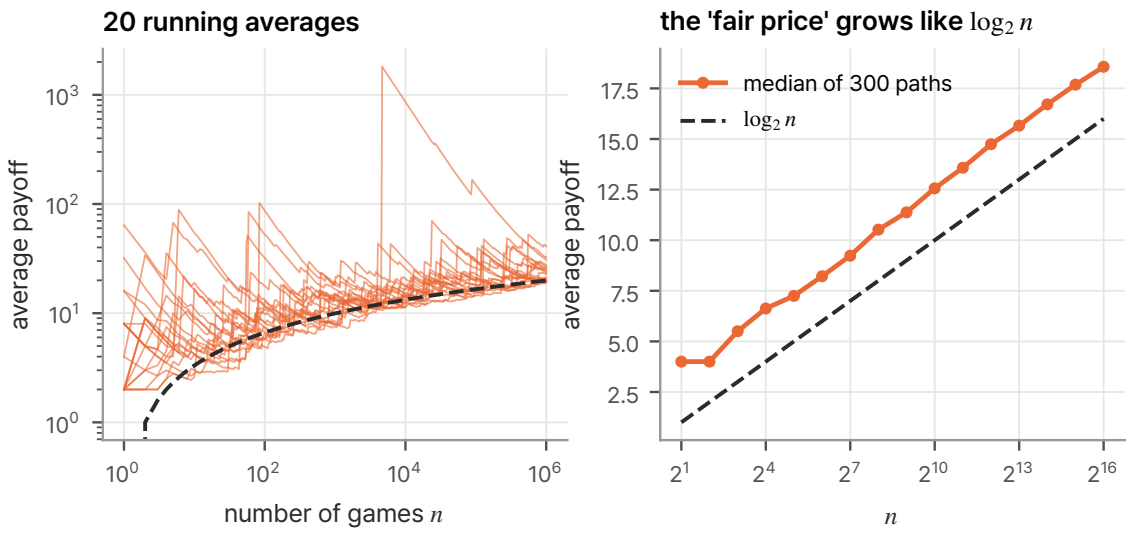


그림 7.1 St. Petersburg 내기. 왼쪽: n판의 평균 상금 경로 20개. 큰 상금이 가끔 경로를 위로 튕겨 올리고, 그 사이에는 천천히 내려온다. 오른쪽: 경로들의 중앙값은 $\log_2 n$(점선)을 따라 자란다.

7.2 기댓값이 없다 — Cauchy

Cauchy 분포에는 기댓값이 없다(예제 1.8). 이 경우 표본평균은 더 극적으로 행동한다. 독립인 Cauchy 확률변수 n개의 평균 $\bar{X}_n$은 **다시 같은 *Cauchy* 분포**를 따른다(7회에서 증명한다). 1,000개를 평균해도 한 개를 볼 때보다 나아지는 것이 없다. 그림 7.2의 경로는 몇 번이고 큰 값에 끌려가고 어디에도 정착하지 않는다.

분산은 무한대이지만 평균은 있는 경우도 있다. Pareto 분포(연습 4.6)에서 $\alpha = 1.5$이면 $EX = 3$이고 $EX^2 = \infty$이다. 이때는 Chebyshev 부등식을 쓸 수 없지만, 대수의 법칙의 결론 $\bar{X}_n \xrightarrow{P} 3$은 여전히 참이다(Khinchin, 7회). 다만 수렴이 $1/\sqrt{n}$보다 훨씬 느리다. 오른쪽 꼬리가 무거운 손실, 소득, 파일 크기 같은 자료에서 평균이 오래 흔들리는 이유다.

7.3 적률이 분포를 정하지 못한다

mgf의 유일성(3.3절)에는 "0의 근방에서 존재하면"이라는 조건이 있다. 이 조건을 빼면, 곧 모든 적률이 있어도 mgf가 없으면 적률이 분포를 정하지 못할 수 있다.

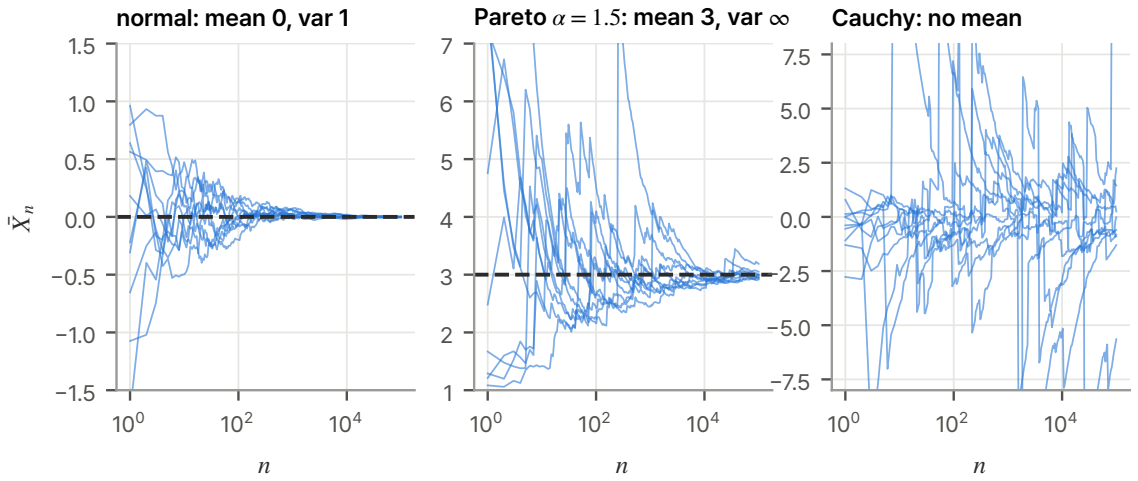


그림 7.2 누적 표본평균의 경로 각 10개. 왼쪽: 정규분포(분산 유한)는 $1/\sqrt{n}$로 정착한다. 가운데: Pareto $\alpha = 1.5$(평균 3, 분산 무한대)는 정착하지만 느리다. 오른쪽: Cauchy(평균 없음)는 정착하지 않는다.

정리 7.1 (적률이 모두 같은 두 분포) $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$로 두고 $x > 0$에서

$$f(x) = \frac{1}{x}\,\phi(\log x), \qquad f_a(x) = f(x)\big[1 + a\sin(2\pi \log x)\big], \quad |a| \le 1$$

라 하자. 그러면 모든 $k = 0, 1, 2, \ldots$ 에 대해 $\int_0^\infty x^k f_a(x)\,dx = \int_0^\infty x^k f(x)\,dx = e^{k^2/2}$이다.

증명. $\int_0^\infty x^k f(x)\sin(2\pi \log x)\,dx = 0$을 보이면 된다. $x = e^y$로 치환하면 $dx = e^y dy$, $f(e^y)e^y = \phi(y)$이므로 적분은 $\int_{-\infty}^{\infty} e^{ky}\phi(y)\sin(2\pi y)\,dy$이다. 지수를 완전제곱으로 쓰면 $ky - \frac{y^2}{2} = \frac{k^2}{2} - \frac{(y-k)^2}{2}$이므로 $y = k + z$로 다시 치환해

$$e^{k^2/2}\int_{-\infty}^{\infty} \phi(z)\sin\big(2\pi(z+k)\big)\,dz = e^{k^2/2}\int_{-\infty}^{\infty} \phi(z)\sin(2\pi z)\,dz = 0.$$

k가 정수이므로 $\sin(2\pi(z+k)) = \sin(2\pi z)$이고, ϕ는 짝함수, sin은 홀함수이므로 적분은 0이다. 같은 계산에서 sin을 빼면 $\int x^k f = e^{k^2/2}\int \phi$이다. ∎

$\int \phi = 1$은 5회에서 증명한다(가우스 적분). 그러면 f는 **로그정규분포**의 밀도($\log X$가 표준정규)이고, f_a도 음이 아니고 적분이 1인 밀도다. 두 분포는 다르지만($a \neq 0$이면 f_a는 물결친다) 모든 적률이 같다(Heyde, 1963). 로그정규분포에는 mgf가 없다. $t > 0$이면 $E\,e^{tX} = \infty$이다(연습 7.4). 적률의 수열 $e^{k^2/2}$이 너무 빨리 자라서 분포를 묶어 두지 못하는 것이다.

7.4 독립이 아니면

X_1이 Bernoulli($\frac{1}{2}$)이고 $X_2 = X_3 = \cdots = X_1$이면, 각 X_i의 평균은 $\frac{1}{2}$이지만 $\bar{X}_n = X_1$은 0 아니면 1이다. 대수의 법칙이 무너진다. 정리 2.8의 교차항 $E(D_iD_j)$가 0이 아니라 $\frac{1}{4}$이기 때문이다. 관측들

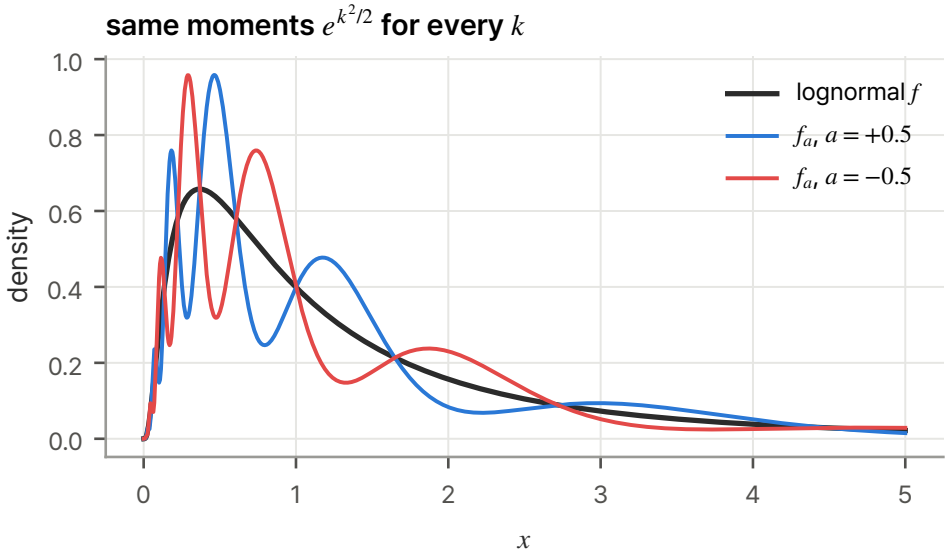


그림 7.3 로그정규 밀도 f(검정)와 f_a, $a = \pm\frac{1}{2}$. 세 밀도의 1차, 2차, 3차, … 모든 적률이 같다.

이 서로 닮아 있으면 n개의 관측이 담은 정보는 n개보다 적다. 3회에서 이것을 "유효 표본 크기"로 잰다.

연습문제

연습 7.1 (은행의 한도 ★) St. Petersburg 내기를 여는 카지노의 돈이 2^{20}두카트(약 100만)뿐이어서, 상금이 그보다 크면 2^{20}만 준다. 기대 상금이 21두카트임을 보여라. 기댓값을 무한대로 만든 것은 확률이 2^{-20}보다 작은 사건들이었다. 풀이 67쪽

연습 7.2 (가진 돈이 있으면 ★★) 처음 가진 돈이 w이고 효용이 $\log$이면, 참가비 c를 낼 의향이 있는 최대 c는 $\sum_k 2^{-k}\log(w-c+2^k) = \log w$의 해다. 노트북으로 $w = 100$과 $w = 10^6$일 때 c를 구하라. 풀이 67쪽

연습 7.3 (Feller의 정규화 ★★★) St. Petersburg 상금 $X_1, \dots, X_n$이 독립이다. $b_n = n\log_2 n$, $Y_i = X_i \mathbf{1}\{X_i \le b_n\}$으로 자른다. (a) $P(\text{어떤 } X_i \neq Y_i) \le nP(X_1 > b_n) \le \frac{2n}{b_n} \to 0$. (b) $EY_i = \lfloor \log_2 b_n \rfloor$, $\operatorname{Var} Y_i \le EY_i^2 \le 2b_n$. (c) Chebyshev로 $\left(\sum Y_i - nEY_1\right)/b_n \xrightarrow{P} 0$이고, $nEY_1/b_n \to 1$이다. 이것으로 $\sum X_i/(n\log_2 n) \xrightarrow{P} 1$을 결론지어라. 풀이 67쪽

연습 7.4 (로그정규에는 mgf가 없다 ★★) $X = e^Y$, Y가 표준정규이면 $t > 0$에 대해 $E\,e^{tX} = \infty$임을 보여라. (힌트: $e^{tx} \ge \frac{(tx)^k}{k!}$이므로 $E\,e^{tX} \ge \frac{t^k}{k!}e^{k^2/2}$이고, 이것은 $k \to \infty$에서 발산한다.) 풀이 68쪽

연습 7.5 (어떤 적률이 있나 ★) Pareto 분포 $F(x) = 1 - x^{-\alpha}$ $(x \geq 1)$에서 $EX^k < \infty$일 필요충분조건을 구하라. 이 분포에 mgf가 있나? 풀이 68쪽

8

Beautiful insights — 같은 도구, 오늘의 문제

8.1 지수적 집중: Chernoff와 Hoeffding

Chebyshev 한계는 $1/n$으로 준다. 실제 꼬리 확률은 지수적으로 준다. Bernoulli가 "확실성을 10배 높이는 데 5,708번이면 된다"고 쓴 것이 그 흔적이다. 따름정리 5.2의 Chernoff 한계를 이항분포에 쓰면 이 지수적 감소를 증명할 수 있다.

정리 8.1 (이항분포의 Chernoff 한계) $K \sim \mathrm{b}(n,p)$, $\hat{p} = K/n$, $p < a < 1$이면

$$P(\hat{p} \ge a) \le e^{-n\,D(a\|p)}, \qquad D(a\|p) = a\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p}.$$

증명. $t > 0$에 대해 따름정리 5.2와 예제 3.2로

$$P(K \ge na) \le e^{-tna}(q + pe^t)^n = \exp\bigl(-n\,[ta - \log(q+pe^t)]\bigr).$$

대괄호를 t에 대해 최대로 한다. 미분해 0으로 놓으면 $a = \frac{pe^t}{q+pe^t}$, 곧 $e^t = \frac{aq}{(1-a)p}$이고, $a > p$이므로 $t > 0$이다. 이 값을 넣으면 $q + pe^t = \frac{q}{1-a}$이므로

$$ta - \log(q+pe^t) = a\log\frac{aq}{(1-a)p} - \log\frac{q}{1-a} = a\log\frac{a}{p} + (1-a)\log\frac{1-a}{q} = D(a\|p).$$

■

$D(a\|p)$는 Bernoulli(a)와 Bernoulli(p) 사이의 **Kullback–Leibler 발산**이다(8.2절). 꼬리 확률의 지수가 정보 이론의 양으로 나온다. 이 지수를 다루기 쉬운 이차식으로 누르면 다음을 얻는다.

정리 8.2 (Hoeffding 부등식 (Bernoulli 시행)) $\varepsilon > 0$이면

$$P\bigl(|\hat{p} - p| \ge \varepsilon\bigr) \le 2\,e^{-2n\varepsilon^2}.$$

증명. $g(\varepsilon) = D(p+\varepsilon\|p) - 2\varepsilon^2$ $(0 \le \varepsilon < q)$이 음이 아님을 보이면 정리 8.1로 $P(\hat{p} \ge p+\varepsilon) \le e^{-2n\varepsilon^2}$이다. 직접 미분하면 $g(0) = 0$, $g'(0) = 0$이고

$$g''(\varepsilon) = \frac{1}{(p+\varepsilon)(q-\varepsilon)} - 4 \;\ge\; 0,$$

$a(1-a) \le \frac{1}{4}$이기 때문이다. g'은 증가하고 $g'(0) = 0$이므로 $g' \ge 0$, 따라서 g는 증가하고 $g \ge g(0) = 0$이다. 왼쪽 꼬리는 $\hat{p}$ 대신 $1-\hat{p}$에 같은 논증을 쓰고, 두 꼬리를 Boole 부등식으로 더한다. ∎

Bernoulli의 질문에 쓰면 $2e^{-2n\varepsilon^2} \le \frac{1}{1001}$, 곧

$$n \ \ge \ \frac{\log(2 \cdot 1001)}{2\varepsilon^2} = \frac{\log 2002}{0.0008} \approx 9{,}503.$$

Chebyshev의 600,600보다 60배 작다. 정리 8.1를 그대로 쓰면 9,121이다. 결정적인 차이는 $\delta = \frac{1}{1001}$이 $\log\frac{1}{\delta}$로 들어간다는 것이다. Chebyshev에서는 $\frac{1}{\delta}$로 들어갔다. 확실성을 10배 높이는 비용이 Chebyshev로는 시행 수의 10배이지만, Hoeffding으로는 $\frac{\log 10}{2\varepsilon^2} \approx 2{,}878$번을 **더하는** 것이다. 정확한 계산에서는 약 2,600번이고, Bernoulli의 공식은 5,708번이었다. 그의 어림은 상수가 두 배쯤 컸을 뿐 모양은 옳았다. 일반적인 Hoeffding 부등식(1963)은 $[0,1]$ 안의 값을 갖는 **임의의** 독립 확률변수의 평균에 대해 같은 한계를 준다.

> **학습 이론의 첫 정리** 분류기 h의 시험 오류율 $R(h)$를 n개의 독립 표본에서 잰 오류율 $\hat{R}(h)$로 추정한다. h가 정해져 있으면 오류 지시변수들의 평균이므로 Hoeffding으로 $P(|\hat{R}(h) - R(h)| \ge \varepsilon) \le 2e^{-2n\varepsilon^2}$이다. 그런데 학습은 $|H|$개의 후보 가운데 $\hat{R}$이 가장 작은 것을 **고른다**. 그러면 하나가 아니라 모든 후보가 동시에 맞아야 하고, Boole 부등식(1회 정리 4.3)으로
>
> $$P\big(\text{어떤 } h \in H \text{에서 } |\hat{R}(h) - R(h)| \ge \varepsilon\big) \le 2|H|e^{-2n\varepsilon^2}.$$
>
> 오른쪽이 δ 이하가 되는 $n \ge \frac{\log(2|H|/\delta)}{2\varepsilon^2}$이 PAC 학습의 표본 크기다. 후보가 100만 개여도 $\log|H| \approx 14$만큼 늘 뿐이다. Bernoulli의 질문에 Boole 부등식 하나를 더한 것이 일반화 이론의 출발점이다.

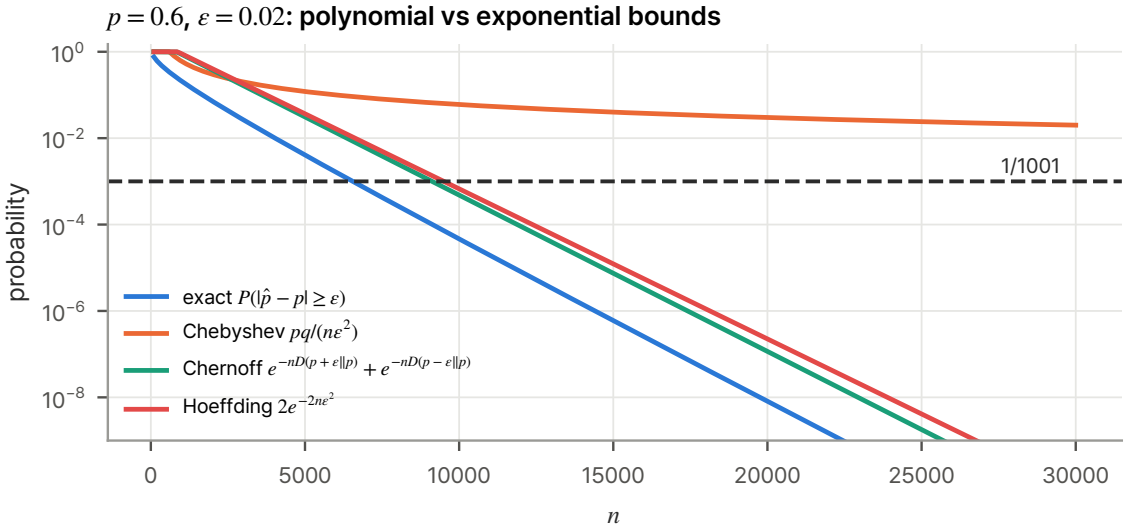


그림 8.1 $p = 0.6, \varepsilon = 0.02$에서 $P(|\hat{p}_n - p| \ge \varepsilon)$의 정확한 값(파랑)과 세 한계. Chebyshev는 $1/n$로, Chernoff·Hoeffding은 지수적으로 준다. 세로축은 로그 눈금이다.

8.2 Jensen과 정보

정리 8.3 (Gibbs 부등식) $p=(p_1,\dots,p_m)$과 $q=(q_1,\dots,q_m)$이 확률질량함수이고 $p_i>0$이면 $q_i>0$이라 하자. 그러면

$$D(p\|q)=\sum_i p_i\log\frac{p_i}{q_i}\ \ge\ 0$$

이고, 등호는 $p=q$일 때뿐이다.

증명. X가 값 $\frac{q_i}{p_i}$를 확률 p_i로 갖는다고 하자($p_i>0$인 i만). log는 오목이므로 Jensen으로

$$-D(p\|q)=E\log X\le\log EX=\log\sum_{i:p_i>0}q_i\le\log 1=0.$$

log는 순오목이므로 등호는 X가 상수일 때, 곧 q_i/p_i가 일정해 $q=p$일 때뿐이다. ∎

이것이 기계학습의 교차 엔트로피 손실이 작동하는 이유다. 참 분포 p와 모형 q의 교차 엔트로피는 $-\sum p_i\log q_i=H(p)+D(p\|q)$이고, $H(p)=-\sum p_i\log p_i$는 모형과 상관없으므로, 교차 엔트로피를 줄이는 것은 $D(p\|q)$를 줄이는 것이다. Gibbs 부등식으로 최솟값은 $q=p$에서 얻는다. 9회에서 최대가능도 추정이 바로 이 원리임을 본다.

변분 추론과 VAE의 ELBO도 Jensen 한 줄이다. 관측 x와 숨은 변수 z의 모형 $p(x,z)$와 아무 분포 $r(z)$에 대해

$$\log p(x)=\log\sum_z r(z)\frac{p(x,z)}{r(z)}\ \ge\ \sum_z r(z)\log\frac{p(x,z)}{r(z)}\qquad(\log\text{가 오목}).$$

계산할 수 없는 왼쪽 대신 오른쪽을 키운다. 10회의 EM 알고리즘도 같은 부등식에서 나온다.

8.3 기댓값으로서의 손실, 그리고 Kelly

모형을 훈련할 때 줄이는 "훈련 손실" $\frac{1}{n}\sum_i \ell(\theta;x_i)$는 손실의 표본평균이고, 우리가 정말 줄이고 싶은 것은 그 기댓값 $E\,\ell(\theta;X)$다. 정해진 θ에서 둘을 잇는 것이 대수의 법칙이고, 고른 θ에서 둘을 잇는 것이 앞의 학습 이론이다. 미니배치 확률적 경사하강법은 기울기의 기댓값을 작은 표본평균으로 추정하고, 그 오차가 배치 크기의 $1/\sqrt{b}$로 준다는 것을 이용한다.

Daniel Bernoulli의 로그 효용은 1956년 Bell 연구소의 John Kelly에게서 다시 태어났다. 이길 확률이 $p>\frac{1}{2}$인 같은 배당의 내기를 되풀이할 때 가진 돈의 비율 f를 건다면, n판 뒤의 돈은 $W_n=W_0\prod_i(1\pm f)$이고

$$\frac{1}{n}\log\frac{W_n}{W_0}=\frac{1}{n}\sum_{i=1}^{n}\log(1+fR_i)\ \xrightarrow{P}\ G(f)=p\log(1+f)+q\log(1-f)$$

이다($R_i = \pm 1$, 대수의 법칙). 장기 성장률 G를 최대로 하는 비율은 $f^* = p - q$이다(연습 8.3). 기댓값 EW_n을 최대로 하려면 전부를 걸어야 하지만($f = 1$), 그러면 한 번만 져도 파산한다. 기댓값을 최대로 하는 것과 **거의 확실하게** 잘 되는 것은 다르다. St. Petersburg 역설과 같은 교훈이다.

연습문제

연습 8.1 (Hoeffding의 여론조사 ★) 연습 6.1의 여론조사(±3%포인트, 틀릴 확률 5%)에 필요한 n을 Hoeffding 부등식으로 구하고, Chebyshev(5,556명)·정규근사(1,068명)와 비교하라. 풀이 68쪽

연습 8.2 (두 주사위의 KL ★) 공평한 주사위 p와, Weldon의 자료에서 추정한 주사위 q(면 1–4에 $\frac{1-\hat{p}}{4}$, 면 5·6에 $\frac{\hat{p}}{2}$, $\hat{p} = 0.33769$, 예제 2.11) 사이의 $D(p\|q)$와 $D(q\|p)$를 구하라. 대칭인가? 풀이 68쪽

연습 8.3 (Kelly 비율 ★★) $G(f) = p\log(1+f) + q\log(1-f)$를 최대로 하는 $f^* = p - q$를 구하고, $p = 0.6$일 때 $G(f^*)$를 계산하라. $f = 2f^*$이면 G는 어떻게 되나? 풀이 68쪽

연습 8.4 (후보가 많을 때 ★) $|H| = 10^6$개의 분류기 가운데 하나를 고를 때, 모든 후보의 오류율을 동시에 ±0.02 안으로 99% 확신하려면 n이 얼마면 되나? $|H| = 1$일 때와 비교하라. 풀이 69쪽

9

정리와 역사 노트

9.1 한 장 요약

	내용	증명의 핵심
기댓값	$EX = \sum x\,p(x)$, $\int x f(x)dx$ (절대수렴할 때)	순서와 상관없는 합
함수의 기댓값	$E\,g(X) = \sum g(x)p(x)$, $\int gf$	값으로 묶기
선형성	$E\sum c_iX_i = \sum c_iEX_i$ (독립 불필요)	표본공간 위의 합
꼬리 공식	$EX = \sum_k P(X \geq k) = \int_0^\infty P(X > x)dx$	이중합의 순서
분산	$\mathrm{Var}\,X = EX^2 - \mu^2$, $E(X-c)^2 = \mathrm{Var}\,X + (\mu - c)^2$	제곱 전개
이항분포	$EK = np$, $\mathrm{Var}\,K = npq$	$k\binom{n}{k} = n\binom{n-1}{k-1}$
독립합	$E(XY) = EX\,EY$, $\mathrm{Var}\sum X_i = \sum \mathrm{Var}\,X_i$	교차항이 0
mgf	$M^{(k)}(0) = EX^k$, $M_{X+Y} = M_XM_Y$, 유일성	거듭제곱급수
단조변환	$f_Y(y) = f_X(h(y))\,\lvert h'(y)\rvert$	분포함수 + 연쇄법칙
역변환 표집	$F^-(U) \sim F$, 연속이면 $F(X) \sim \mathrm{U}(0,1)$	$F^-(u) \leq x \iff u \leq F(x)$
Markov · Chebyshev	$P(X \geq a) \leq \frac{EX}{a}$, $P(\lvert X-\mu\rvert \geq c) \leq \frac{\sigma^2}{c^2}$	함수를 누르고 기댓값
Jensen	볼록이면 $\varphi(EX) \leq E\varphi(X)$	받침선
대수의 법칙	$P(\lvert\bar{X}_n - \mu\rvert \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$	Chebyshev + 독립합
Hoeffding	$P(\lvert\hat{p} - p\rvert \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$	Chernoff + $D \geq 2\varepsilon^2$
부숴보기	St. Petersburg, Cauchy, 로그정규의 적률, 독립이 없는 평균	조건을 하나씩 뺀다

9.2 Stigler의 명명 법칙

> **Stigler's law of eponymy** "어떤 과학 법칙도 처음 발견한 사람의 이름을 달고 있지 않다."
>
> - **Chebyshev 부등식**: Bienaymé가 1853년에 먼저 증명했다. 프랑스 문헌에서는 지금도 Bienaymé–Tchebychev 부등식이라 부른다.
> - **Markov 부등식**: Markov의 스승 Chebyshev의 연구에 이미 나온다. 해석학에서는 지금도 이것을 Chebyshev 부등식이라 부르곤 한다.
> - **Daniel Bernoulli의 효용**: 1728년 Gabriel Cramer가 Nicolaus Bernoulli에게 보낸 편지에서 먼저 오목한 효용(제곱근)으로 St. Petersburg 역설을 풀었다. Daniel은 1738년 논문에 이 편지를 인용했다.
> - **Bernoulli의 정리**가 "대수의 법칙"이 된 것은 Poisson의 이름 짓기 때문이다. 그가 붙인 이름은 Bernoulli의 정리보다 일반적인 자신의 정리(연습 6.5)를 가리켰다.

9.3 7 Pillars 지도

Stigler(2016)의 첫째 기둥 **Aggregation**(모으기)은 "관측을 모으면, 개별 관측의 정보를 버리는 데도 오히려 더 많이 알게 된다"는 역설이다. 평균은 각 관측이 무엇이었는지를 버린다. 그런데도 대수의 법칙은 평균이 참값으로 간다고 말한다. 이번 회차가 그 기둥의 수학이다. 그러나 **얼마나 빨리** 가는가, 곧 $1/\sqrt{n}$은 둘째 기둥 **Information**의 주제이고 3회에서 세운다. 18세기의 천문학자들에게 "여러 번 재서 평균을 내는 것이 한 번을 잘 재는 것보다 낫다"는 것은 자명하지 않았다.

9.4 다음 회 — 왜 평균을 내는가: $\sqrt{n}$의 발견 (1755 → 1958)

이번 회차의 여러 정리에는 "이산 확률변수에서 증명하고, 연속은 3회"라는 꼬리표가 붙어 있었다. 3회는 두 개 이상의 확률변수를 함께 다루는 언어, 곧 결합분포·주변분포·조건부분포·공분산을 세우고 그 꼬리표를 뗀다. 역사의 질문은 Thomas Simpson(1755)의 것이다. 천문 관측을 여러 번 해서 평균내는 것이 정말 나은가? 그리고 런던 조폐국은 수백 년 동안 금화의 무게 허용오차를 $\sqrt{n}$이 아니라 n에 비례하게 잡았다. 그 틈을 누가 이용했을까?

참고문헌

- Stigler, S. M. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard, 1986. 2장 (Bernoulli와 de Moivre).
- Stigler, S. M. *The Seven Pillars of Statistical Wisdom*. Harvard, 2016. 1장 (Aggregation).
- Hald, A. *A History of Probability and Statistics and Their Applications before 1750*. Wiley, 1990.

- Bernoulli, J. *Ars Conjectandi*. Basel, 1713. 4부 영어 번역: Sung, B. *Translations from James Bernoulli*, Harvard Univ. Dept. of Statistics Tech. Report 2, 1966.
- Bernoulli, D. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5 (1738). 영어 번역: *Econometrica* 22 (1954).
- Huygens, C. *De ratiociniis in ludo aleae*. 1657.
- Chebyshev, P. L. Des valeurs moyennes. *Journal de mathématiques pures et appliquées* (2) 12 (1867).
- Heyde, C. C., Seneta, E. *I. J. Bienaymé: Statistical Theory Anticipated*. Springer, 1977.
- Feller, W. Note on the law of large numbers and "fair" games. *Annals of Mathematical Statistics* 16 (1945).
- Heyde, C. C. On a property of the lognormal distribution. *JRSS B* 25 (1963).
- Hoeffding, W. Probability inequalities for sums of bounded random variables. *JASA* 58 (1963).
- Kelly, J. L. A new interpretation of information rate. *Bell System Technical Journal* 35 (1956).
- Weldon의 주사위 자료: R 패키지 `vcd`의 `WeldonDice`. Arbuthnot 자료: `HistData`의 `Arbuthnot`.

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 4쪽

$\frac{1}{2}s = \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \cdots$의 항 사이에 0을 끼워 $0 + \frac{1}{2} + 0 - \frac{1}{4} + 0 + \frac{1}{6} + 0 - \frac{1}{8} + \cdots$로 쓴다(값은 그대로 $\frac{1}{2}s$). 이것을 $s = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots$와 항별로 더하면(수렴하는 두 급수의 합)

$$1 + 0 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + 0 + \frac{1}{7} - \frac{1}{4} + \cdots = \frac{3}{2}s.$$

0을 지우면 양수 항 둘 다음에 음수 항 하나인 재배열이 되고, 그 합은 $\frac{3}{2}\log 2$이다. 같은 항들을 다른 순서로 더해 다른 값을 얻었다.

연습 0.2 문제 4쪽

(a) $a_i \geq 0$이면 $a_i^+ = a_i$, $a_i^- = 0$이고, $a_i < 0$이면 $a_i^+ = 0$, $a_i^- = -a_i$이다. 두 경우 모두 $a_i = a_i^+ - a_i^-$, $|a_i| = a_i^+ + a_i^-$이고, 따라서 $0 \leq a_i^{\pm} \leq |a_i|$이다. (b) $\sum_{i\leq n} a_i^+ \leq \sum_{i\leq n} |a_i| \leq \sum_i |a_i| < \infty$이므로 음이 아닌 급수 $\sum a_i^+$의 부분합은 증가하고 위로 유계이다. 유계인 단조수열은 수렴하므로 $\sum a_i^+$는 유한하다 [T5]. $\sum a_i^-$도 같다. (c) 재배열 σ에 대해 부분합은 $\sum_{k\leq K} a_{\sigma(k)} = \sum_{k\leq K} a^+_{\sigma(k)} - \sum_{k\leq K} a^-_{\sigma(k)}$이다. 1회 4.6절의 상자로 음이 아닌 급수 $\sum_k a^+_{\sigma(k)}$는 순서와 상관없이 $\sum_i a_i^+$로 수렴하고, $\sum_k a^-_{\sigma(k)}$는 $\sum_i a_i^-$로 수렴한다. 두 수렴하는 수열의 차는 차로 수렴하므로 $\sum_k a_{\sigma(k)} = \sum_i a_i^+ - \sum_i a_i^-$이다. 이 값은 σ에 의존하지 않는다. 묶어 더하는 경우도 각 묶음 안에서 같은 분해를 하고, 음이 아닌 급수는 묶어 더해도 값이 같다는 것(같은 상자)을 a^+와 a^-에 따로 쓰면 된다.

1절

연습 1.1 문제 12쪽

이길 확률 $\frac{18}{38}$, 질 확률 $\frac{20}{38}$이므로 $EG = \frac{18-20}{38} = -\frac{1}{19} \approx -0.0526$. 선형성으로 100판의 기대 이익은 -5.26이다.

연습 1.2 문제 12쪽

나와 다른 $p+q-1$명이 각각 x씩 걸어 판돈은 $(p+q)x$이고, 공평한 제비로 한 명이 이긴다. 나는 다른 사람 가운데 q명과 "당신이 이기면 내게 b를 주고, 내가 이기면 당신에게 b를 준다" 고, 나머지 $p-1$명과는 같은 약속을 a로 한다. 그 q명 가운데 누가 이기면 나는 b를, $p-1$명 가운데 누가 이기면 a를 받는다. 내가 이기면 판돈에서 $qb+(p-1)a$를 주고 $(p+q)x-qb-(p-1)a$를 갖는다. 이것이 a가 되도록 $x=\frac{pa+qb}{p+q}$로 잡으면, 나는 p가지 경우(나 자신과 $p-1$명)에 a, q가지 경우에 b를 받는다. 공평한 내기에 x를 내고 이 처지를 얻었으므로 그 처지의 값은 x다.

연습 1.3 문제 12쪽

$\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$이므로 부분합이 $1 - \frac{1}{K+1} \to 1$. $EX = \sum_k k \cdot \frac{1}{k(k+1)} = \sum_k \frac{1}{k+1} = \infty$ (조화급수).

연습 1.4 문제 12쪽

$\sum_k x_k p(x_k) = \sum_k \frac{(-1)^{k+1}}{k}$는 교대급수로 $\log 2$에 수렴한다. 그러나 $\sum_k |x_k| p(x_k) = \sum_k \frac{1}{k} = \infty$이므로 정의 1.1의 조건을 어긴다. 값 x_k들에는 정해진 순서가 없다. 값을 다른 순서로 나열하면 같은 확률분포에서 연습 0.1처럼 $\frac{3}{2}\log 2$나 다른 어떤 값도 얻는다. 분포에서 순서와 상관없이 정해지는 수가 아니므로 기댓값이라 부를 수 없다.

연습 1.5 문제 12쪽

$j-1$종류를 가졌을 때 새 종류가 나올 확률은 $p_j = \frac{n-j+1}{n}$이고 봉지마다 독립이므로 G_j는 기하분포, $EG_j = \frac{n}{n-j+1}$이다(예제 1.4). 선형성으로

$$ET = \sum_{j=1}^{n} \frac{n}{n-j+1} = n \sum_{i=1}^{n} \frac{1}{i}.$$

$n=6$이면 $6(1+\frac{1}{2}+\cdots+\frac{1}{6}) = 14.7$번. ($G_j$들은 서로 다른 시점의 시행이지만, 선형성에는 독립도 필요 없다.)

연습 1.6 문제 12쪽

$P(X \ge k) = \sum_{j \ge k} p(j)$를 넣고 음이 아닌 이중합의 순서를 바꾸면

$$\sum_{k \ge 1} (2k-1) \sum_{j \ge k} p(j) = \sum_{j \ge 1} p(j) \sum_{k=1}^{j} (2k-1) = \sum_{j \ge 1} j^2 p(j) = EX^2,$$

$\sum_{k=1}^{j} (2k-1) = j^2$이기 때문이다. 기하분포에서 $P(X \ge k) = q^{k-1}$이므로 $EX^2 = 2\sum_k k q^{k-1} - \sum_k q^{k-1} = \frac{2}{p^2} - \frac{1}{p} = \frac{2-p}{p^2}$.

연습 1.7 문제 12쪽

$\{K \ge k\}$는 $x+k$세까지 산다는 사건이고 그 확률이 $\ell(x+k)/\ell(x)$이므로 꼬리 공식(정리 1.13(a))으로 $EK = \sum_{k \ge 1} \ell(x+k)/\ell(x)$. Halley의 표로 $x = 20$이면 약 33.6년이다(완전한 햇수만 센 값이고, 죽는 해의 반년을 더하면 약 34.1년). 1세의 기대여명 32.9년보다 길다. 어린 나이의 높은 사망률을 이미 지나왔기 때문이다. 1671년 Johan de Witt가 종신연금의 값을 이런 기댓값으로 매겼다.

연습 1.8 문제 12쪽

$\int_{-A}^{2A} \frac{x}{\pi(1+x^2)} dx = \frac{1}{2\pi}\big[\log(1+x^2)\big]_{-A}^{2A} = \frac{1}{2\pi} \log \frac{1+4A^2}{1+A^2} \to \frac{1}{2\pi} \log 4 = \frac{1}{\pi} \log 2.$

연습 1.9 문제 13쪽

(a) $Y = y_j$인 것은 $g(X) = y_j$인 것, 곧 $g(x_i) = y_j$인 어떤 x_i에서 $X = x_i$인 것이다. 서로 다른 x_i에 대한 사건 $\{X = x_i\}$는 서로소이므로 (P3)으로 $P(Y = y_j) = \sum_{i:\, g(x_i) = y_j} p(x_i)$이다. (b) 첨자 i들을 $g(x_i)$의 값으로 나누면 모든 i가 정확히 한 묶음에 들어간다. 음이 아닌 급수는 묶어 더해도 같으므로

$$\sum_j |y_j| \, P(Y = y_j) = \sum_j \sum_{i:\, g(x_i) = y_j} |g(x_i)| \, p(x_i) = \sum_i |g(x_i)| \, p(x_i) < \infty.$$

첫 등호에서 묶음 안에서는 $|g(x_i)| = |y_j|$임을 썼다. 따라서 정의 1.1로 EY가 존재한다. (c) 같은 묶음을 절댓값 없이 하면 $EY = \sum_j y_j P(Y = y_j) = \sum_i g(x_i) p(x_i)$이다. 부호가 섞인 급수를 묶어 더해도 값이 같다는 것은 절대수렴할 때만 보장되므로 보조정리 0.1가 필요하다. 그 조건을 (b)에서 확인했다.

2절

연습 2.1 문제 19쪽

(a) $P(S = s) = \frac{6 - |s-7|}{36}$ $(s = 2, \dots, 12)$에서 $ES = 7$, $ES^2 = \frac{1974}{36}$, $\operatorname{Var} S = \frac{1974}{36} - 49 = \frac{35}{6} \approx 5.83$.

(b) 두 눈이 독립이므로 $\operatorname{Var} S = 2 \cdot \frac{35}{12} = \frac{35}{6}$.

연습 2.2 문제 19쪽

$m < c$이면 $t \in (m, c)$에 대해 $\frac{d}{dt}|x-t| = \operatorname{sgn}(t-x)$ $(t \neq x)$이므로 $|x-c| - |x-m| = \int_m^c \operatorname{sgn}(t-x)\,dt$ 이다. 기댓값을 취하고 (유계인 함수의 유한 구간 적분이므로) 순서를 바꾸면, X가 연속이어서 $E\operatorname{sgn}(t-X) = P(X<t) - P(X>t) = 2F(t) - 1$이므로

$$E|X-c| - E|X-m| = \int_m^c \big(2F(t)-1\big)dt \;\geq\; 0,$$

$t \geq m$이면 $F(t) \geq F(m) = \frac{1}{2}$이기 때문이다. $c < m$도 같다. 따라서 $E|X-c|$는 $c = m$에서 가장 작다.

연습 2.3 문제 19쪽

연습 1.6로 $EX^2 = \frac{2-p}{p^2}$이므로 $\operatorname{Var} X = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}$. $p = \frac{1}{6}$이면 $\operatorname{Var} X = 30$, 표준편차 $\sqrt{30} \approx 5.48$번. 평균 6번에 표준편차 5.5번이니 기다리는 시간은 크게 흔들린다.

연습 2.4 문제 19쪽

$K = \sum_{i=1}^{5} \mathbf{1}_i$, $\mathbf{1}_i = \{i$번째 카드가 에이스$\}$, $P = \frac{4}{52} = \frac{1}{13}$이므로 $EK = \frac{5}{13} \approx 0.385$. $i \neq j$이면 $P(\text{둘 다 에이스}) = \frac{4\cdot 3}{52\cdot 51}$이고 순서쌍 (i,j)가 20개이므로 $E\,K(K-1) = 20 \cdot \frac{12}{2652} \approx 0.0905$. 따라서

$$\operatorname{Var} K = 0.0905 + 0.3846 - 0.3846^2 \approx 0.327.$$

$\mathrm{b}(5, \frac{1}{13})$의 분산은 $5 \cdot \frac{1}{13} \cdot \frac{12}{13} \approx 0.355$로 더 크다. 되돌려 넣지 않으므로 에이스 하나가 나오면 다음 에이스가 나올 확률이 줄어든다($\frac{12}{2652} < \frac{1}{169}$). 교차항이 음수가 되어 분산이 줄어든다.

연습 2.5 문제 19쪽

각 지시변수는 $P = \frac{1}{2}$이므로 분산 $\frac{1}{4}$이고 합은 $\frac{3}{4}$. 직접 계산: 면 R, G, B에서 $R+G+B = 1$, 면 RGB에서 3이므로 $R+G+B$는 확률 $\frac{3}{4}$로 1, $\frac{1}{4}$로 3이다. 평균 $\frac{3}{2}$, $E(\cdot)^2 = \frac{3}{4} + \frac{9}{4} = 3$, 분산 $3 - \frac{9}{4} = \frac{3}{4}$. 같다.

연습 2.6 문제 19쪽

$(x_3, \ldots, x_n)$의 모든 값에 걸친 사건 $\{X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n\}$들은 서로소이고 합집합이 $\{X_1 = x_1, X_2 = x_2\}$이므로 (P3)으로 좌변의 합은 $P(X_1 = x_1, X_2 = x_2)$이다. 우변의 합은 $P(X_1 = x_1)P(X_2 = x_2)\prod_{i\geq 3}\sum_{x_i} P(X_i = x_i) = P(X_1 = x_1)P(X_2 = x_2)$.

연습 2.7 문제 19쪽

$\mathrm{U}(0,1)$: 대칭이므로 3차 중심적률은 0. $E(U - \frac{1}{2})^4 = \int_{-1/2}^{1/2} x^4 dx = \frac{1}{80}$, $\sigma^4 = \frac{1}{144}$이므로 $\gamma_2 =$

$\frac{144}{80} - 3 = -\frac{6}{5}$. Exp(1): $\mu = \sigma = 1$. $E(X-1)^3 = EX^3 - 3EX^2 + 3EX - 1 = 6 - 6 + 3 - 1 = 2$, $E(X-1)^4 = 24 - 24 + 12 - 4 + 1 = 9$이므로 $\gamma_1 = 2$, $\gamma_2 = 6$. Bernoulli: $E(X-p)^3 = pq^3 - qp^3 = pq(q-p)$이므로 $\gamma_1 = \frac{pq(q-p)}{(pq)^{3/2}} = \frac{1-2p}{\sqrt{pq}}$.

연습 2.8 문제 19쪽

$E(X-X')^2 = EX^2 - 2E(XX') + EX'^2 = 2EX^2 - 2(EX)^2 = 2\operatorname{Var} X$. 둘째 등호에서 독립이므로 $E(XX') = EX \cdot EX'$(정리 2.7)이고 분포가 같다.

연습 2.9 문제 19쪽

$\hat{p} = 4.0522/12 = 0.3377$이므로 $12\hat{p}(1-\hat{p}) = 2.684$. 관측한 분산 2.696과 매우 가깝다. 공평하지 않을 뿐, 주사위마다 $5 \cdot 6$이 나올 확률이 약 0.338인 이항 모형은 자료의 흩어짐을 잘 설명한다.

연습 2.10 문제 19쪽

(a) 전개는 $a = X - \mu$, $b = \mu - c$에 대한 $(a+b)^2 = a^2 + 2ab + b^2$이다. $E(X-\mu)^2 = \operatorname{Var} X < \infty$, $E|X-\mu| \le E|X| + |\mu| < \infty$, 상수의 기댓값은 그 상수이므로 세 항의 기댓값이 모두 존재한다. (b) 선형성(정리 1.10(a))으로 $E(X-c)^2 = E(X-\mu)^2 + 2(\mu-c)E(X-\mu) + (\mu-c)^2$이고, $E(X-\mu) = EX - \mu = 0$이므로 $E(X-c)^2 = \operatorname{Var} X + (\mu-c)^2$이다. 오른쪽 둘째 항이 $c = \mu$에서만 0이므로 $c = \mu$에서 가장 작다. (c) $c = 0$이면 $EX^2 = \operatorname{Var} X + \mu^2$이다. (d) 선형성으로 $E(aX+b) = a\mu + b$이므로 $\operatorname{Var}(aX+b) = E\big(aX + b - (a\mu + b)\big)^2 = E\,a^2(X-\mu)^2 = a^2 \operatorname{Var} X$이다. 제곱근을 취하면 $\sigma_{aX+b} = |a|\sigma_X$이다.

연습 2.11 문제 20쪽

(a) 이항정리(1회 정리 1.6)로 $\sum_k \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1$이다. (b) $k\binom{n}{k} = k \cdot \frac{n!}{k!\,(n-k)!} = \frac{n!}{(k-1)!\,(n-k)!} = n \cdot \frac{(n-1)!}{(k-1)!\big((n-1)-(k-1)\big)!} = n\binom{n-1}{k-1}$. (c) $k = 0$인 항은 0이므로

$$EK = \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k q^{n-k} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} = np\,(p+q)^{n-1} = np.$$

(d) (b)를 두 번 쓰면 $k(k-1)\binom{n}{k} = n(k-1)\binom{n-1}{k-1} = n(n-1)\binom{n-2}{k-2}$이고, 같은 계산으로 $E\,K(K-1) = n(n-1)p^2(p+q)^{n-2} = n(n-1)p^2$이다. (e) $K^2 = K(K-1) + K$이므로 $\operatorname{Var} K = E\,K(K-1) + EK - (EK)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$이다. $k^2\binom{n}{k}$는 $\binom{n-2}{k-2}$ 하나로 줄지 않지만 $k(k-1)\binom{n}{k}$는 줄기 때문에 $E\,K(K-1)$을 먼저 구한다.

연습 2.12 문제 20쪽

(a) (X, Y)는 셀 수 있는 값 (x, y)들을 갖는 이산 확률변수이고, 정리 1.9(a)를 $g(x,y) = xy$에 쓰면 $\sum_{x,y} |xy| P(X=x, Y=y) < \infty$일 때 $E(XY) = \sum_{x,y} xy\,P(X=x, Y=y)$이다. (b) 독립이므로

$P(X=x,Y=y)=P(X=x)P(Y=y)$이고, 음이 아닌 이중급수는 x에 대해 먼저 더해도 되므로

$$\sum_{x,y}|x||y|P(X=x)P(Y=y)=\sum_x|x|P(X=x)\sum_y|y|P(Y=y)=E|X|\,E|Y|<\infty.$$

(c) (b)로 이중급수가 절대수렴하므로 보조정리 0.1로 어떤 순서로 더해도 같다. x에 대해 먼저 더하면 $\sum_{x,y}xy\,P(X=x)P(Y=y)=EX\cdot EY$이다.

3절

연습 3.1 문제 24쪽

$M(t)=\frac{1}{6}\sum_{k=1}^{6}e^{kt}$. $M'(0)=\frac{1}{6}\sum k=\frac{21}{6}=3.5$, $M''(0)=\frac{1}{6}\sum k^2=\frac{91}{6}$이므로 $\operatorname{Var}X=\frac{91}{6}-3.5^2=\frac{35}{12}$.

연습 3.2 문제 24쪽

$\psi(t)=\log p+t-\log(1-qe^t)$. $\psi'(t)=1+\frac{qe^t}{1-qe^t}$이므로 $\psi'(0)=1+\frac{q}{p}=\frac{1}{p}$. $\psi''(t)=\frac{qe^t}{(1-qe^t)^2}$이므로 $\psi''(0)=\frac{q}{p^2}$.

연습 3.3 문제 24쪽

값 $-1,0,1$을 확률 $\frac{1}{4},\frac{1}{2},\frac{1}{4}$로 갖는 분포의 mgf가 정확히 이 식이다. mgf가 0의 근방(사실 모든 t)에서 같으므로 유일성(3.3절의 상자)으로 분포가 이것이다. 독립인 공평한 동전 두 개의 앞면 지시변수를 B_1,B_2라 하면 B_1-B_2가 이 분포를 따른다. 실제로 $M_{B_1-B_2}(t)=\frac{1}{2}(1+e^t)\cdot\frac{1}{2}(1+e^{-t})$이다.

연습 3.4 문제 24쪽

$X\ge x$이면 $e^{s(X-x)}\ge 1$이고, 아니면 $e^{s(X-x)}>0$이므로 $\mathbf{1}\{X\ge x\}\le e^{s(X-x)}$. 기댓값을 취하면 $P(X\ge x)\le e^{-sx}M(s)$. $P(X\ge x)=x^{-2}$이면 mgf가 있을 때 $x^{-2}\le M(s)e^{-sx}$, 곧 $e^{sx}\le M(s)x^2$이 모든 $x\ge 1$에서 성립해야 하는데 지수함수가 다항식보다 빨리 자라므로 불가능하다.

연습 3.5 문제 25쪽

$\psi_{X+Y}=\log(M_XM_Y)=\psi_X+\psi_Y$ (정리 3.6(b)). $\psi'''=\big(M'''M^2-3M''M'M+2(M')^3\big)/M^3$에 $t=0$을 넣으면 $EX^3-3EX^2\cdot\mu+2\mu^3$이고, 이것은 $E(X-\mu)^3=EX^3-3\mu EX^2+3\mu^2EX-\mu^3$과 같다. Bernoulli 하나의 3차 누율은 $pq(q-p)$이므로(연습 2.7) n개의 합은 $npq(q-p)$, 분산은 npq이고 왜도는 $\frac{npq(q-p)}{(npq)^{3/2}}=\frac{1-2p}{\sqrt{npq}}$. n이 커지면 0으로 간다. 이항분포가 대칭에 가까워진다는 뜻이다(5회 de Moivre).

연습 3.6 문제 25쪽

보통 주사위 두 개의 합의 확률생성함수는 $\frac{1}{36}z^2(1+z)^2(1+z+z^2)^2(1-z+z^2)^2$ 이다. 첫 주사위에 $z(1+z)(1+z+z^2) = z+2z^2+2z^3+z^4$, 둘째에 $z(1+z)(1+z+z^2)(1-z+z^2)^2 = z+z^3+z^4+z^5+z^6+z^8$ 을 주면 곱이 같다. 각 다항식의 계수의 합(=$z=1$에서의 값)이 6이므로 각각 면이 6개인 주사위다. 확률생성함수가 같으면 계수, 곧 분포가 같다.

4절

연습 4.1 문제 31쪽

$g(u) = -\frac{1}{\lambda}\log(1-u)$는 $(0,1)$에서 증가, $h(y) = 1-e^{-\lambda y}$, $h'(y) = \lambda e^{-\lambda y}$. 정리 4.3으로 $f_Y(y) = 1\cdot\lambda e^{-\lambda y}$ $(y>0)$. $1-U$도 $\mathrm{U}(0,1)$이므로 $-\frac{1}{\lambda}\log U$와 분포가 같다.

연습 4.2 문제 31쪽

$f_X = \frac{1}{3}$ $(-1<x<2)$. $0<y<1$이면 $\{X^2\le y\} = \{-\sqrt{y}\le X\le\sqrt{y}\}$이고 이 구간은 $(-1,2)$ 안이므로 $F_Y(y) = \frac{2\sqrt{y}}{3}$, $f_Y(y) = \frac{1}{3\sqrt{y}}$. $1\le y<4$이면 $-\sqrt{y}\le -1$이므로 $\{X^2\le y\} = \{-1<X\le\sqrt{y}\}$, $F_Y(y) = \frac{\sqrt{y}+1}{3}$, $f_Y(y) = \frac{1}{6\sqrt{y}}$. $F_Y(1) = \frac{2}{3}$으로 이어지고 $F_Y(4) = 1$이다.

연습 4.3 문제 31쪽

$X = 10^{N+V}$, N은 정수, $V\in[0,1)$이 $\log_{10}X$의 소수 부분이다. 첫 자리가 d인 것은 $d\le 10^V < d+1$, 곧 $\log_{10}d\le V<\log_{10}(d+1)$과 같으므로 확률은 $\log_{10}(d+1)-\log_{10}d = \log_{10}(1+\frac{1}{d})$. $d=1$이면 $\log_{10}2\approx 0.301$이다.

연습 4.4 문제 31쪽

$(0,\infty)$에서 $g(x) = \frac{1}{x}$, $h(y) = \frac{1}{y}$, $|h'(y)| = \frac{1}{y^2}$이므로 $f_Y(y) = \frac{1}{\pi(1+y^{-2})}\cdot\frac{1}{y^2} = \frac{1}{\pi(1+y^2)}$ $(y>0)$. $(-\infty,0)$도 같다. 두 조각을 합치면(사건 $\{Y\le y\}$를 X의 부호로 나눠 분포함수 방법을 쓰면) Cauchy 밀도다. $\tan\Theta$: $P(\tan\Theta\le x) = P(\Theta\le\arctan x) = \frac{\arctan x+\pi/2}{\pi}$, 곧 Cauchy의 분포함수.

연습 4.5 문제 31쪽

$F(x) = 0$ $(x<0)$, $F(x) = 0.3+0.7(1-e^{-x})$ $(x\ge 0)$. $u\le 0.3$이면 $F(0) = 0.3\ge u$이고 $x<0$에서 $F = 0<u$이므로 $F^-(u) = 0$. $u>0.3$이면 $F(x) = u$를 풀어 $F^-(u) = \log\frac{0.7}{1-u}$. 균등 난수의 30%가 0으로 가고 나머지가 지수분포의 꼬리로 간다.

연습 4.6 문제 31쪽

$F^{-}(u) = (1-u)^{-1/\alpha}$. 밀도 $f(x) = \alpha x^{-\alpha-1}$ $(x \ge 1)$이므로 $EX = \alpha \int_1^\infty x^{-\alpha} dx$는 $\alpha > 1$에서만 유한하고 $\frac{\alpha}{\alpha-1}$이다. $EX^2 = \alpha \int_1^\infty x^{1-\alpha} dx$는 $\alpha > 2$에서만 유한하다.

연습 4.7 문제 31쪽

$F(0) = \frac{1}{2}$, $F(1) = 1$이므로 $F(X)$는 $\frac{1}{2}$과 1을 확률 $\frac{1}{2}$씩 갖는다. $\mathrm{U}(0,1)$이 아니다. 정리의 증명에서 $P(X < x) = F(x)$를 썼는데, F가 뛰는 점에서는 성립하지 않는다.

연습 4.8 문제 31쪽

(a) g가 증가하면 $h = g^{-1}$도 증가하고 정리 4.3으로 $f_Y(y) = f_X(h(y))h'(y)$ $(y \in T)$이다. $y = g(x)$로 치환하면 x가 S를 지날 때 y는 T를 같은 방향으로 지나고, $h(y) = x$, $dy = g'(x)\,dx$이며 $h'(g(x))g'(x) = 1$(역함수의 미분)이므로

$$\int_T y\, f_X(h(y))\, h'(y)\, dy = \int_S g(x)\, f_X(x)\, h'(g(x))\, g'(x)\, dx = \int_S g(x)\, f_X(x)\, dx.$$

(b) 같은 치환을 $|y|$에 하면 $\int_T |y| f_Y(y)\,dy = \int_S |g(x)| f_X(x)\,dx$이고, 가정으로 유한하다. 따라서 정의 1.5로 EY가 존재하고 (a)의 값과 같다. (c) g가 감소하면 x가 S를 왼쪽에서 오른쪽으로 지날 때 y는 T를 오른쪽에서 왼쪽으로 지난다. 치환한 적분은 $\int_{\sup T}^{\inf T}$가 되어 부호가 하나 바뀌고, $|h'| = -h'$의 부호가 그것을 되돌린다. 결과는 (a)와 같다.

5절

연습 5.1 문제 37쪽

X가 a를 확률 π로, 0을 확률 $1-\pi$로 가지면 $P(X \ge a) = \pi = EX/a$. 등호는 증명의 부등식 $a\mathbf{1}\{X \ge a\} \le X$가 확률 1로 등식일 때, 곧 $X \in \{0, a\}$일 때뿐이다.

연습 5.2 문제 37쪽

$\mathrm{U}(0,1)$: $\sigma = 0.2887$. $k = 1.5$이면 $P = 1 - 2(0.433) = 0.134$, $k = 2, 3$이면 $k\sigma > \frac{1}{2}$이어서 0이다. $\mathrm{Exp}(1)$: $\mu = \sigma = 1$이고 $X \ge 0$이므로 $P(|X-1| \ge k) = P(X \ge 1+k) = e^{-(1+k)}$: 0.082, 0.050, 0.018. 한계 $1/k^2$은 0.444, 0.25, 0.111이다.

연습 5.3 문제 37쪽

$s \ge 0$이면 $X - \mu \ge c$에서 $X - \mu + s \ge c + s > 0$이므로 $\{X - \mu \ge c\} \subset \{(X-\mu+s)^2 \ge (c+s)^2\}$

이고, Markov로

$$P(X-\mu \ge c) \le \frac{E(X-\mu+s)^2}{(c+s)^2} = \frac{\sigma^2+s^2}{(c+s)^2}.$$

오른쪽을 s로 미분하면 $s=\sigma^2/c$에서 최소이고, 그 값은 $\frac{\sigma^2(c^2+\sigma^2)/c^2}{(c^2+\sigma^2)^2/c^2} = \frac{\sigma^2}{\sigma^2+c^2}$. $c=2\sigma$이면 $\frac{1}{5}$이다. 양쪽 Chebyshev($\frac{1}{4}$)보다 작다.

연습 5.4 문제 37쪽

$|\mu - m| = |E(X-m)| \le E|X-m|$ (정리 1.10(c)) $\le E|X-\mu|$ (연습 2.2: 중앙값이 절대오차를 최소로 한다) $\le \sqrt{E(X-\mu)^2} = \sigma$ (따름정리 5.10(a)를 $|X-\mu|$에 쓴다).

연습 5.5 문제 37쪽

$\lambda = \frac{y-\mu}{y-x}$이면 $\lambda x + (1-\lambda)y = \mu$이다. 볼록성 $\varphi(\mu) \le \lambda\varphi(x) + (1-\lambda)\varphi(y)$에 $y-x$를 곱하면 $(y-x)\varphi(\mu) \le (y-\mu)\varphi(x) + (\mu-x)\varphi(y)$. 왼쪽을 $(y-\mu)\varphi(\mu) + (\mu-x)\varphi(\mu)$로 쓰고 정리하면 $(y-\mu)\big(\varphi(\mu)-\varphi(x)\big) \le (\mu-x)\big(\varphi(y)-\varphi(\mu)\big)$이고, 양변을 $(y-\mu)(\mu-x)>0$으로 나누면 된다.

연습 5.6 문제 37쪽

테일러 정리로 $\varphi(t) = \varphi(z) + \varphi'(z)(t-z) + \frac{1}{2}\varphi''(\xi)(t-z)^2 \ge \varphi(z) + \varphi'(z)(t-z)$. $t=x$에 λ, $t=y$에 $1-\lambda$를 곱해 더하면 $\lambda\varphi(x) + (1-\lambda)\varphi(y) \ge \varphi(z) + \varphi'(z)\big(\lambda x + (1-\lambda)y - z\big) = \varphi(z)$.

연습 5.7 문제 37쪽

거리 d를 n번 가는 데 걸린 시간은 $\sum d/v_i$이므로 평균 속력은 $\frac{nd}{\sum d/v_i} = \big(\frac{1}{n}\sum \frac{1}{v_i}\big)^{-1}$, 조화평균이다. X가 v_i들을 확률 $\frac{1}{n}$씩 가지면 $E\frac{1}{X} \ge \frac{1}{EX}$이므로 조화평균 $\le$ 산술평균. 60과 40이면 $\frac{2}{1/60+1/40} = 48$km/h. 50이 아니다. 느린 쪽에서 더 오래 머물기 때문이다.

연습 5.8 문제 37쪽

(a) Chebyshev 부등식(정리 5.3)에 $c=\frac{1}{n}$을 넣으면 $P(|X-\mu| \ge \frac{1}{n}) \le n^2 \operatorname{Var} X = 0$이다. (b) $|X(\omega)-\mu| \ge \frac{1}{n}$이면 $|X(\omega)-\mu| \ge \frac{1}{n+1}$이므로 $A_n \subset A_{n+1}$이다. $\omega \in \bigcup_n A_n$이면 $X(\omega) \ne \mu$이고, 거꾸로 $X(\omega)\ne\mu$이면 $|X(\omega)-\mu|>0$이므로 $\frac{1}{n} \le |X(\omega)-\mu|$인 n이 있다(아르키메데스 성질). (c) 1회 정리 4.7(a)로 $P(X\ne\mu) = \lim_n P(A_n) = 0$이다. $X\ge 0$, $EX=0$이면 Markov 부등식으로 $P(X \ge \frac{1}{n}) \le n\,EX = 0$이고, $\{X \ge \frac{1}{n}\} \uparrow \{X>0\}$에 같은 논리를 쓰면 $P(X>0)=0$이다.

연습 5.9 문제 37쪽

(a) $\varphi(x)=x^2$은 $\mathbb{R}$에서 볼록이므로($\varphi''=2\ge 0$) $(EX)^2 \le EX^2$이다. $|x|$는 삼각부등식 $|\lambda x + (1-\lambda)y| \le \lambda|x| + (1-\lambda)|y|$로 볼록이므로 $|EX| \le E|X|$이다. 첫 식은 $EX^2<\infty$일 때, 둘째 식은

$E|X|<\infty$일 때 쓴다. (b) $I=(0,\infty)$에서 $(\log x)''=-1/x^2<0$이므로 log는 오목이고(연습 5.6를 $-\log$에), $(1/x)''=2/x^3>0$이므로 $1/x$는 볼록이다. 정리 5.9으로 $E\log X\le\log EX$, $E\frac{1}{X}\ge\frac{1}{EX}$이다(두 기댓값이 존재할 때). (c) X가 값 a_i를 확률 $\frac{1}{n}$로 갖게 하면(같은 값이 여러 번 나오면 그 확률을 더한다) $EX=\frac{1}{n}\sum a_i$, $E\log X=\frac{1}{n}\sum\log a_i$이다. (b)로 $\frac{1}{n}\sum\log a_i\le\log\!\left(\frac{1}{n}\sum a_i\right)$이고, 지수함수가 증가하므로 양변에 지수를 취하면 $(a_1\cdots a_n)^{1/n}\le\frac{1}{n}\sum a_i$이다. 같은 값이 있어도 된다. (d) $Y=|X|^r\ge0$이고 $\varphi(y)=|y|^{s/r}$은 $s/r>1$이므로 $\mathbb{R}$에서 볼록이다. $E|X|^s<\infty$이면 2.4절의 $|x|^r\le1+|x|^s$로 $E|X|^r<\infty$이다. Jensen으로 $(EY)^{s/r}\le E\,Y^{s/r}=E|X|^s$이고, 양변을 $\frac{1}{s}$제곱하면(증가함수) $(E|X|^r)^{1/r}\le(E|X|^s)^{1/s}$이다.

6절

연습 6.1 문제 44쪽

$\frac{1}{4n(0.03)^2}\le0.05$에서 $n\ge\frac{1}{4\cdot0.0009\cdot0.05}\approx5{,}556$명. 정규근사의 1,068명보다 5배 많다.

연습 6.2 문제 44쪽

$m_1=\lceil\log(19\cdot10^4)/\log\frac{31}{30}\rceil=\lceil370.7\rceil=371$, $N_1=371\cdot50+\frac{371\cdot1000-1000}{31}\approx30{,}486$. $m_2=\lceil\log(29\cdot10^4)/\log\frac{21}{20}\rceil=\lceil257.8\rceil=258$, $N_2=258\cdot50+\frac{258\cdot1500-1500}{21}\approx31{,}257.1$이므로 31,258번. c가 10배 되면 m_2의 하한은 $\log10/\log1.05\approx47.2$만큼 늘고, Bernoulli는 47을 더해 N_2가 $47\cdot\left(50+\frac{1500}{21}\right)\approx5{,}708$씩 는다고 썼다.

연습 6.3 문제 44쪽

$\sigma^2=\frac{35}{12}$이므로 $\frac{35/12}{n(0.1)^2}\le0.05$에서 $n\ge5{,}834$번.

연습 6.4 문제 44쪽

정리 6.3의 증명이 쓰는 것은 $\operatorname{Var}\bar X_n=\sigma^2/n$뿐이고, 이것은 정리 2.8에서 왔다. 그 증명에서 독립을 쓴 곳은 $i\ne j$인 각 쌍에서 $E(D_iD_j)=ED_i\,ED_j$를 쓴 한 줄이고, 이것은 정리 2.7를 **두 변수** D_i,D_j에 쓴 것이다. 쌍별 독립이면 그대로 성립한다. 바꿀 줄은 없다.

연습 6.5 문제 44쪽

$K_n=\sum_{i\le n}\mathbf{1}_i$, $\mathbf{1}_i$들은 독립이고 $E\mathbf{1}_i=p_i$, $\operatorname{Var}\mathbf{1}_i=p_iq_i\le\frac{1}{4}$. 정리 2.8으로 $\operatorname{Var}K_n\le\frac{n}{4}$이고 Chebyshev로

$$P\!\left(\left|\frac{K_n-\sum p_i}{n}\right|\ge\varepsilon\right)\le\frac{n/4}{n^2\varepsilon^2}=\frac{1}{4n\varepsilon^2}\to0.$$

연습 6.6 문제 44쪽

노트북의 계산으로 $Q^2 \approx 2.09$ ($Q \approx 1.45$)이다. 해마다 남아 비율이 이항 모형이 허용하는 것보다 두 배 가까운 분산으로 흔들린다. 가능한 이유는 여럿이다. 남아 출생 확률 자체가 해마다(지역, 계층 구성에 따라) 달랐을 수 있고, 자료가 출생이 아니라 국교회의 **세례** 기록이어서 기록 관행(내전과 공화정 시기 1640–1660년대의 혼란, 비국교도의 누락)이 해마다 달랐을 수 있다. 쌍둥이처럼 독립이 아닌 출생도 있다. 대수의 법칙은 82년 평균 0.516을 믿게 하지만, 이 평균이 "하나의 고정된 p" 라는 해석은 분산이 거부한다.

연습 6.7 문제 44쪽

(a) 이항계수의 정의로

$$\frac{T_{k+1}}{T_k} = \frac{\binom{N}{k+1}}{\binom{N}{k}} \cdot \frac{p}{q} = \frac{N-k}{k+1} \cdot \frac{p}{q}.$$

$k \ge Np$이면 $(N-k)p \le (N-Np)p = Npq < (k+1)q$이므로 비는 1보다 작다. (b) $k \ge j \ge N(p+\varepsilon)$이면 $N-k \le N(q-\varepsilon)$, $k+1 > N(p+\varepsilon)$이므로 비 $\le \frac{N(q-\varepsilon)p}{N(p+\varepsilon)q} = \rho$. $(q-\varepsilon)p < (p+\varepsilon)q$이므로 $\rho < 1$. 그러면 $T_{j+i} \le \rho^i T_j$이고 $P(K \ge j) \le T_j \sum_i \rho^i = \frac{T_j}{1-\rho}$. (c) $\lceil Np \rceil \le k \le j$인 정수는 $j - \lceil Np \rceil + 1 \ge N(p+\varepsilon) - (Np+1) + 1 = N\varepsilon$개 이상이고, (a)로 T_k는 이 범위에서 감소하므로 모두 T_j 이상이다. 확률의 합은 1 이하이므로 $N\varepsilon\, T_j \le 1$. 따라서 $P(\hat{p}_N \ge p+\varepsilon) \le \frac{1}{N\varepsilon(1-\rho)} \to 0$. 아래쪽 꼬리는 성공과 실패를 바꿔 같은 논증을 쓴다. 분산은 한 번도 쓰지 않았다.

7절

연습 7.1 문제 48쪽

$K \le 20$이면 상금 2^K, $K > 20$이면 2^{20}이므로 기대 상금은 $\sum_{k=1}^{20} 2^k 2^{-k} + 2^{20} P(K > 20) = 20 + 2^{20} \cdot 2^{-20} = 21$두카트.

연습 7.2 문제 48쪽

참가비 c를 내고 참가하면 가진 돈이 $w - c + 2^K$이므로, 기대 효용이 참가하지 않을 때의 $\log w$와 같아지는 c가 최대 참가비다. 노트북의 수치 풀이로 $w = 100$이면 $c \approx 7.8$, $w = 10^6$이면 $c \approx 20.9$두카트. 부자일수록 더 많이 낼 의향이 있지만, 백만장자도 21두카트를 넘지 않는다.

연습 7.3 문제 48쪽

(a) $P(X_1 > b) = P(K > \log_2 b) = 2^{-\lfloor \log_2 b \rfloor} \le \frac{2}{b}$이고 Boole 부등식으로 $P(\text{어떤 } X_i \ne Y_i) \le \frac{2n}{b_n} = \frac{2}{\log_2 n} \to 0$. (b) $EY_1 = \sum_{k:2^k \le b_n} 2^k 2^{-k} = \lfloor \log_2 b_n \rfloor$, $EY_1^2 = \sum_{k:2^k \le b_n} 2^k \le 2 \cdot 2^{\lfloor \log_2 b_n \rfloor} \le 2b_n$. (c)

Y_i들은 독립이므로 $\operatorname{Var}\sum Y_i \le 2nb_n$이고 Chebyshev로

$$P\Big(\Big|\sum_i Y_i - nEY_1\Big| \ge \varepsilon b_n\Big) \le \frac{2nb_n}{\varepsilon^2 b_n^2} = \frac{2}{\varepsilon^2 \log_2 n} \to 0.$$

또 $\dfrac{nEY_1}{b_n} = \dfrac{\lfloor \log_2(n\log_2 n)\rfloor}{\log_2 n} \to 1$. (a)로 $\sum X_i = \sum Y_i$일 확률이 1로 가므로 $\sum X_i/b_n \xrightarrow{P} 1$.

연습 7.4 문제 48쪽

$x > 0$이면 $e^{tx} \ge \dfrac{(tx)^k}{k!}$이므로 $E\,e^{tX} \ge \dfrac{t^k}{k!}EX^k = \dfrac{t^k e^{k^2/2}}{k!}$. $k! \le k^k = e^{k\log k}$이므로 오른쪽은 $\exp\!\Big(\dfrac{k^2}{2} - k\log k + k\log t\Big) \to \infty$.

연습 7.5 문제 49쪽

$EX^k = \alpha\int_1^\infty x^{k-\alpha-1}dx < \infty \iff k < \alpha$, 그때 $\dfrac{\alpha}{\alpha-k}$. 차수가 α 이상인 적률은 없으므로 mgf도 없다($t>0$이면 $e^{tx} \ge (tx)^k/k!$를 $k > \alpha$로 쓴다).

8절

연습 8.1 문제 53쪽

$2e^{-2n(0.03)^2} \le 0.05$에서 $n \ge \dfrac{\log 40}{2\cdot 0.0009} \approx 2{,}050$명. Chebyshev의 5,556명보다 훨씬 작고 정규근사의 1,068명의 두 배쯤이다. Hoeffding은 정규근사처럼 "n이 크면"이라는 근사 없이 모든 n에서 참이다.

연습 8.2 문제 53쪽

노트북의 계산으로 $D(p\|q) \approx 4.24\times10^{-5}$, $D(q\|p) \approx 4.25\times10^{-5}$이다. 거의 같지만 같지 않다. KL 발산은 일반적으로 대칭이 아니다. 아주 작은 값이라는 것은 주사위 한 번으로는 두 분포를 거의 구별할 수 없다는 뜻이다. Weldon의 주사위 315,672개 굴림에서는 $nD(q\|p) \approx 13.4$가 되어 구별할 수 있다. 가능도비로는 $e^{13.4} \approx 7\times10^5$배다(9회).

연습 8.3 문제 53쪽

$G'(f) = \dfrac{p}{1+f} - \dfrac{q}{1-f} = 0$에서 $p(1-f) = q(1+f)$, $f^* = p - q$. $G'' < 0$이므로 최대다. $p = 0.6$이면 $f^* = 0.2$, $G(f^*) = 0.6\log 1.2 + 0.4\log 0.8 \approx 0.0201$. 판마다 약 2%씩 복리로 자라 약 34판에 돈이 두 배가 된다. $f = 0.4$이면 $G \approx -0.0024 < 0$이다. 판마다 유리한 내기인데도 너무 많이 걸면 돈은 확률 1에 가깝게 0으로 간다.

연습 8.4 문제 53쪽

$n \geq \frac{\log(2\cdot 10^6/0.01)}{2(0.02)^2} \approx 23{,}892$. $|H| = 1$이면 $\frac{\log 200}{0.0008} \approx 6{,}623$. 후보가 100만 배 늘어도 표본은 3.6배만 늘어난다.