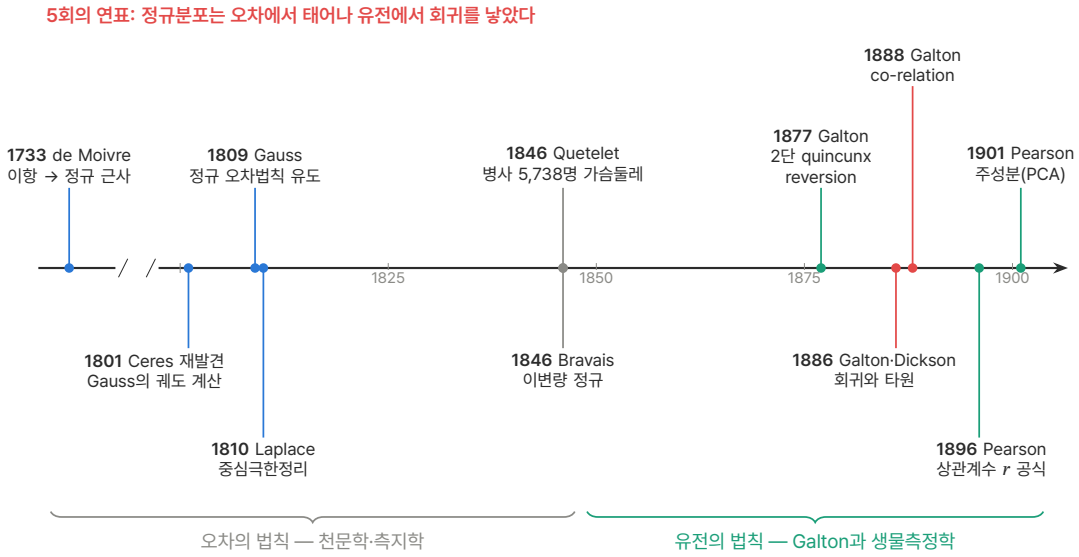
수리통계학 × 통계학의 역사

5회 강의노트

정규분포의 탄생과 Galton의 타원

1733 → 1901

정규분포는 처음에 관측 오차의 분포였다. 이 노트는 de Moivre의 근사와 Gauss의 특성화에서 출발해 Galton이 유전 자료에서 본 동심 타원까지 따라가고, 두 시대를 다변량 정규분포의 조건부분포라는 하나의 정리로 묶는다.



역사	Stigler, The History of Statistics (1986), 4장 · 5장 · 8장
노트북	`sessions/05_normal_galton/05_normal_galton.ipynb` — 절·정리 번호가 같다
워크북	`05_normal_galton_workbook.pdf` (손으로 푸는 34문항) · 풀이판 `_workbook_sol.pdf`
선수 지식	1–4회 (특히 2회 mgf, 3회 야코비안·전분산 공식·가우스 적분, 4회 감마와 χ^2)

이 노트를 쓰는 법

각 절은 **역사 → 정의·정리 → 증명 → 그림 → 연습문제** 순서다. 핵심 정리의 증명은 번호 붙은 단계 ① ② ③으로 쓰고, 단계마다 오른쪽 끝에 근거(앞 정리의 번호나 도구)를 단다. 극한·적분·급수의 순서를 바꾸는 곳에는 **조건 확인** 단락을 붙였다. 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 [T3] 귀납, [T5] 극한, [T6] 적분 순서의 교환, [T7] 변수변환과 야코비안, [T8] Taylor 전개와 O를 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** 모수와 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.

> **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 선수 지식 복습 1

1 정규분포는 어떻게 태어났나 4

1.1 de Moivre (1733): 이항분포의 근사로서 4

1.2 Gauss (1809): “평균이 가장 그럴듯하다” 면 오차는 정규다 7

1.3 Quetelet (1846): 오차의 분포가 사람에게 쓰이다 9

1.4 단변량 정규분포 11

2 Galton의 문제: 변이는 어떻게 유전되는가 15

2.1 Quincunx (1873): 작은 충격의 합이 종을 만든다 15

2.2 2단 quincunx (1877): 정규를 정규로 섞으면 다시 정규 16

2.3 세대 간 분산의 수수께끼와 reversion 18

2.4 Galton (1886): 928명의 표와 동심 타원 19

3 다변량 정규분포 22

3.1 행렬 준비 22

3.2 정의 · 구성 · 밀도 24

3.3 정리 A — 선형변환 26

3.4 정리 B — 독립과 무상관 27

3.5 정리 C — 조건부분포 28

3.6 정리 D — 이차형식과 확률 타원 32

3.7 이변량의 경우 — 회귀식, 그리고 접점 33

4 계산 — 정리를 알고리즘으로 36

5 부숴보기 — 가정을 하나씩 빼면 38

5.1 '결합'을 빼면: 주변은 정규, 결합은 아니다 38

5.2 정규를 약하게: 오염된 정규 39

5.3 한 집단이 아니면: 혼합과 Galton의 ×1.08 40

5.4 Σ가 특이하면: 밀도는 없어도 정규다 41

6 Beautiful insights — 정리 C는 어디에나 있다 42

6.1 Gaussian Process 회귀 = 정리 C 한 번 42

6.2 Pearson (1901): 회귀선 둘, 주성분 하나 42

6.3 백색화와 BatchNorm 43

6.4 Reparameterization trick과 두 정규 사이의 KL 43

7 정리와 역사 노트 45

7.1 한 장 요약 45

7.2 Stigler의 명명 법칙 45

7.3 7 Pillars 지도에 칠할 칸 46

7.4 다음 회 — Student의 t와 신뢰구간 (1876 → 1937) 46

A 연습문제 풀이 47

0

들어가며

> **여는 질문.** 키 큰 부모의 자녀는 왜 부모보다 덜 클까? 그리고 **키 큰 자녀의 부모**는 왜 자녀보다 덜 클까?

두 번째 질문이 이상하게 들린다면 좋은 출발이다. 첫 번째 질문에는 "유전이 평균으로 끌어당긴다"는 생물학적 이야기를 지어낼 수 있지만, 두 번째 질문은 시간을 거꾸로 거슬러 올라간다. 자녀가 부모를 끌어당길 수는 없다. 이 노트가 끝날 때 두 질문은 같은 답을 갖는다. 답은 생물학이 아니라 **타원 하나의 기하**다.

정규분포는 처음에 **오차의 분포**였다. de Moivre는 동전 던지기의 근사로(1733), Gauss는 천문 관측 오차의 법칙으로(1809), Laplace는 작은 오차들의 합의 극한으로(1810) 정규분포에 도달했다. 19세기 후반 Galton은 같은 분포를 **사람의 변이**와 **유전**에 썼고, 그 과정에서 **회귀**와 **상관**이 태어났다(1886, 1888). 이 노트는 두 시대를 하나의 수학, 곧 **다변량 정규분포의 조건부분포**로 묶는다.

학습 목표

1. 정규분포가 태어난 세 경로(de Moivre 근사, Gauss의 특성화, Laplace의 CLT)를 수학으로 설명한다.
2. 다변량 정규분포를 mgf로 정의하고 선형변환·독립·조건부·이차형식 네 정리를 증명한다.
3. 조건부분포 공식에서 **평균으로의 회귀**를 유도하고 Galton의 1886년 원자료로 확인한다.
4. 같은 공식이 GP 회귀, PCA, 백색화, VAE에서 어떻게 다시 나타나는지 설명한다.

0.1 선수 지식 복습

이 노트에서 계속 쓰는 세 가지 도구를 정리한다. 모두 앞 회차에서 증명했다. 보조정리마다 어디서 증명했는지 밝힌다.

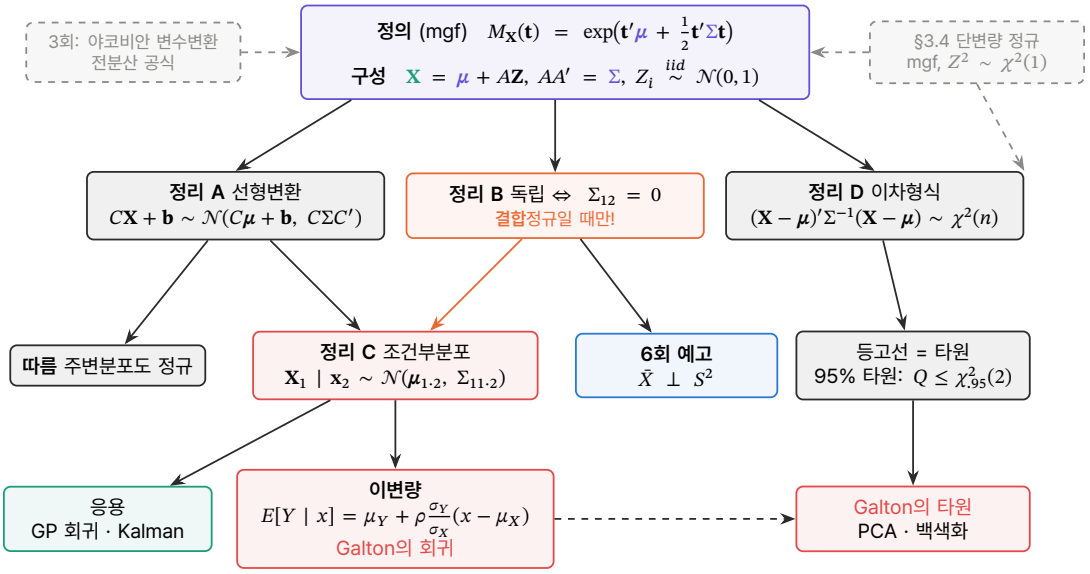


그림 0.1 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

보조정리 0.1 (적률생성함수(mgf)의 성질 — 2회 3절, 3회 정리 3.4) 확률변수 X의 mgf $M_X(t) = E\,e^{tX}$가 0의 근방 $|t| < h$에서 존재한다고 하자.

(a) **유일성.** 두 확률변수의 mgf가 0의 근방에서 같으면 두 분포는 같다.

(b) **적률.** $M_X^{(k)}(0) = EX^k$. 특히 $\psi(t) = \log M_X(t)$라 하면 $\psi'(0) = EX$, $\psi''(0) = \operatorname{Var} X$.

(c) **아핀 변환.** $M_{aX+b}(t) = e^{bt} M_X(at)$.

(d) **독립합.** $X \perp Y$이면 $M_{X+Y}(t) = M_X(t) M_Y(t)$.

확률벡터 $\mathbf{X}$의 mgf는 $M_{\mathbf{X}}(\mathbf{t}) = E\,e^{\mathbf{t}'\mathbf{X}}$이고, (a)의 유일성이 그대로 성립한다. 또한 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$에 대해 $M_{\mathbf{X}}(\mathbf{t}_1, \mathbf{t}_2) = M_{\mathbf{X}_1}(\mathbf{t}_1) M_{\mathbf{X}_2}(\mathbf{t}_2)$가 $\mathbf{0}$의 근방에서 성립하면 $\mathbf{X}_1 \perp \mathbf{X}_2$이다.

(a)는 2회 3.3절의 **증명 없이 쓰는 사실**, (b)는 2회 정리 3.3과 따름정리 3.4, (c)는 2회 정리 3.6(a), (d)는 3회 정리 3.4(d)다. 벡터에 대한 마지막 두 문장은 (a)를 다변수 Laplace 변환으로 넓힌 것이고, 같은 이유로 이 과정에서 증명 없이 쓴다(Billingsley, *Probability and Measure*, 30절).

보조정리 0.2 (변수변환 공식 — 3회 정리 5.1) $\mathbf{X}$가 열린 집합 $S \subset \mathbb{R}^n$ 위에서 밀도 $f_{\mathbf{X}}$를 갖고, $\mathbf{Y} = g(\mathbf{X})$가 S에서 T로 가는 일대일 미분가능 변환이며 역변환 $\mathbf{x} = h(\mathbf{y})$의 야코비안 $J(\mathbf{y}) = \det\left[\partial h_i / \partial y_j\right]$가 0이 아니면

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\big(h(\mathbf{y})\big)\,|J(\mathbf{y})|, \qquad \mathbf{y} \in T.$$

보조정리 0.3 (반복기댓값과 전분산 공식 — 3회 정리 2.5 · 2.6) $E\big[E(Y \mid X)\big] = EY$이고, $\operatorname{Var} Y = E\big[\operatorname{Var}(Y \mid X)\big] + \operatorname{Var}\big(E[Y \mid X]\big)$이다. 벡터에서는 $\operatorname{Cov}\mathbf{Y} = E\big[\operatorname{Cov}(\mathbf{Y} \mid \mathbf{X})\big] + \operatorname{Cov}\big(E[\mathbf{Y} \mid \mathbf{X}]\big)$이다.

스칼라판은 3회에서 증명했다. 벡터판은 같은 증명을 공분산의 각 성분에 쓰면 된다(연습 0.2). 이 밖에 3회 정리 5.3(가우스 적분)과 4회 정리 3.9($Z^2 \sim \chi^2(1)$, 감마함수)를 1.4절에서 다시 쓴다.

전분산 공식은 이 노트에서 세 번 다른 얼굴로 나온다. 2.2절의 2단 quincunx, 3.5절의 조건부분포(피타고라스 정리), 5.3절의 혼합분포다.

연습문제

연습 0.1 (극좌표의 야코비안 ★) $x = r\cos\varphi$, $y = r\sin\varphi$ $(r > 0,\ 0 \le \varphi < 2\pi)$의 야코비안이 r임을 보여라. 이것으로 (X, Y)가 독립 표준정규일 때 (R, Φ)의 결합밀도를 구하고, R과 Φ가 독립임을 보여라. 풀이 47쪽

연습 0.2 (벡터의 전분산 공식 ★★) $\operatorname{Cov}(\mathbf{Y} \mid \mathbf{X}) = E\big[(\mathbf{Y} - E[\mathbf{Y} \mid \mathbf{X}])(\mathbf{Y} - E[\mathbf{Y} \mid \mathbf{X}])' \mid \mathbf{X}\big]$로 정의하고, 3회 정리 2.6의 증명을 행렬로 옮겨 보조정리 0.3의 벡터판을 증명하라. 풀이 47쪽

1

정규분포는 어떻게 태어났나

정규분포에 이르는 길은 셋이다. 이항분포의 근사(de Moivre), "평균이 최선"이라는 요구(Gauss), 작은 오차들의 합의 극한(Laplace). 이 절은 앞의 둘을 증명하고, 셋째는 7회(중심극한정리)로 넘긴다. 마지막 1.4절에서 이후에 계속 쓸 단변량 정규분포의 사실들을 모아 증명한다.

1.1 de Moivre (1733): 이항분포의 근사로서

> **역사** 1733년 11월, 런던의 프랑스 망명 수학자 Abraham de Moivre는 짧은 라틴어 소논문 *Approximatio ad Summam Terminorum Binomii* $(a+b)^n$ *in Seriem expansi*를 지인들에게 돌렸다. 1738년 *The Doctrine of Chances* 2판에 영어로 실렸다. 물음은 동전을 n번 던질 때 앞면 수가 $n/2$ 근처의 구간에 들 확률이었다. n이 수천이면 이항계수를 하나하나 더할 수 없었다. de Moivre는 $\log n!$의 급수 전개를 이용했고(그 상수 $\sqrt{2\pi}$는 Stirling이 찾았다), 주로 $p=\frac{1}{2}$를 다뤘다. 그 답에 오늘날의 정규밀도가 처음 나타났다.

보조정리 1.1 (Stirling 공식) $m \to \infty$일 때

$$m! = \sqrt{2\pi m}\left(\frac{m}{e}\right)^m \left(1 + O(m^{-1})\right).$$

더 정확히는 모든 $m \geq 1$에서 $\sqrt{2\pi m}\,(m/e)^m\, e^{1/(12m+1)} < m! < \sqrt{2\pi m}\,(m/e)^m\, e^{1/(12m)}$이다.

> **감 잡기** $\log m! = \sum_{k=1}^{m} \log k$는 $\int_1^m \log x\,dx = m\log m - m + 1$에 가깝다. 사다리꼴 보정을 하면 $\log m! \approx m\log m - m + \frac{1}{2}\log m + C$가 되고, 상수 C가 $\log\sqrt{2\pi}$다. 흥미롭게도 이 상수는 **정규분포의 적분 상수와 같은 것**이다. 연습 1.3에서 de Moivre 정리를 거꾸로 써서 C를 결정한다. 실제 비 $m!/\left(\sqrt{2\pi m}(m/e)^m\right)$은 $m = 5, 10, 20$에서 1.0168, 1.0084, 1.0042로, 대략 $1 + \frac{1}{12m}$이다.

정리 1.2 (국소 de Moivre–Laplace 정리) $S_n \sim \mathrm{Bin}(n, p)$, $q = 1-p$, $0 < p < 1$이라 하자. 정수 k에 대해 $x_k = \frac{k-np}{\sqrt{npq}}$로 두면, 임의의 $A > 0$에 대해 $|x_k| \le A$인 k들에서 **균등하게**

$$P(S_n = k) = \frac{1}{\sqrt{2\pi npq}}\, e^{-x_k^2/2}\left(1 + O(n^{-1/2})\right) \qquad (n \to \infty).$$

증명. **전략** Stirling 공식으로 이항확률을 "앞인수 × 거듭제곱 부분"으로 나누고, 거듭제곱 부분의 로그를 $\delta = k - np$에 대해 2차까지 Taylor 전개한다 [T8]. 1차 항은 서로 소거되고 2차 항이 정규밀도의 지수가 된다.

$\delta = k - np = x_k\sqrt{npq}$로 둔다. $|x_k| \le A$이므로 $|\delta| \le A\sqrt{npq} = O(\sqrt{n})$이고, $k = np + \delta$와 $n - k = nq - \delta$는 모두 n에 비례해 커진다.

① Stirling 공식(보조정리 1.1)을 $n!$, $k!$, $(n-k)!$에 넣는다. $e^{-n}/(e^{-k}e^{-(n-k)}) = 1$이므로

$$P(S_n = k) = \frac{n!}{k!\,(n-k)!}\, p^k q^{n-k} = \frac{\sqrt{2\pi n}\, n^n\, p^k q^{n-k}}{\sqrt{2\pi k}\, k^k\, \sqrt{2\pi(n-k)}\,(n-k)^{n-k}}\left(1 + O(n^{-1})\right)$$

$$= \underbrace{\sqrt{\frac{n}{2\pi k(n-k)}}}_{\text{(가) 앞인수}} \underbrace{\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}}_{\text{(나) 거듭제곱 부분}} \left(1 + O(n^{-1})\right). \qquad n^n = n^k n^{n-k}$$

② (가)를 δ로 쓴다.

$$k(n-k) = (np+\delta)(nq-\delta) = n^2pq\Big(1 + \frac{\delta}{np}\Big)\Big(1 - \frac{\delta}{nq}\Big) = n^2pq\left(1 + O(n^{-1/2})\right),$$

$\delta = O(\sqrt{n})$ [T8]

$$\sqrt{\frac{n}{2\pi k(n-k)}} = \frac{1}{\sqrt{2\pi npq}}\left(1 + O(n^{-1/2})\right). \qquad (1+h)^{-1/2} = 1 + O(h)$$

③ $u = \delta/(np)$, $v = \delta/(nq)$로 두면 $u, v = O(n^{-1/2})$이고 $\log(\text{나}) = -k\log(1+u) - (n-k)\log(1-v)$이다. 로그의 성질

④ $|w| \le \frac{1}{2}$에서 $\log(1+w) = w - \frac{w^2}{2} + R(w)$, $|R(w)| \le |w|^3$이다. 이것을 넣고 $k = np + \delta$, $n - k = nq - \delta$로 전개한다.

$$-(np+\delta)\Big(\frac{\delta}{np} - \frac{\delta^2}{2n^2p^2}\Big) = -\delta - \frac{\delta^2}{2np} + O\Big(\frac{\delta^3}{n^2}\Big), \qquad \text{Taylor 전개 [T8]}$$

$$-(nq-\delta)\Big(-\frac{\delta}{nq} - \frac{\delta^2}{2n^2q^2}\Big) = +\delta - \frac{\delta^2}{2nq} + O\Big(\frac{\delta^3}{n^2}\Big). \qquad \text{같은 계산}$$

⑤ 나머지항은 $k\,|R(u)| \le k|u|^3 = O(n \cdot n^{-3/2}) = O(n^{-1/2})$이고, $(n-k)|R(-v)|$도 같다. $\delta^3/n^2 = O(n^{-1/2})$이다. O의 대수 [T8]

⑥ 두 줄을 더하면 $\pm\delta$가 소거된다.

$$\log(\text{나}) = -\frac{\delta^2}{2n}\Big(\frac{1}{p} + \frac{1}{q}\Big) + O(n^{-1/2}) = -\frac{\delta^2}{2npq} + O(n^{-1/2}). \qquad \frac{1}{p} + \frac{1}{q} = \frac{1}{pq}$$

⑦ $\delta^2/(npq) = x_k^2$이므로 (나) $= e^{-x_k^2/2}\,e^{O(n^{-1/2})} = e^{-x_k^2/2}(1 + O(n^{-1/2}))$이다. ①, ②와 곱하면 결론이다. $e^h = 1 + O(h)$ [T8]

조건 확인 단계 ④의 전개에는 $|u|, |v| \le \frac{1}{2}$가 필요하다. $|\delta| \le A\sqrt{npq}$이므로 이 조건은 $n \ge 4A^2 \max(q/p,\, p/q)$이면 성립한다. 각 단계의 $O(\cdot)$ 상수는 A와 p에만 의존하고 k에는 의존하지 않으므로 근사는 $|x_k| \le A$에서 균등하다. ∎

한 걸음 더 단계 ④에서 δ^3 항을 끝까지 모으면 계수가 $\frac{q-p}{6n^2p^2q^2}$이다(연습 1.1). $p = \frac{1}{2}$이면 이 항이 사라지므로 근사가 한 차수 더 빨리 좋아진다. 그림 1.1의 오른쪽에서 최대 오차의 기울기가 $p = \frac{1}{2}$일 때 약 $-\frac{3}{2}$, $p = 0.2$일 때 약 -1인 이유다.

따름정리 1.3 (적분형 de Moivre–Laplace 정리) $a < b$이면

$$\lim_{n\to\infty} P\Big(a \le \frac{S_n - np}{\sqrt{npq}} \le b\Big) = \Phi(b) - \Phi(a).$$

증명. **전략** 구하는 확률을 $P(S_n = k)$들의 합으로 쓰고, 정리 1.2의 균등한 근사로 각 항을 $\varphi(x_k)\Delta$로 바꾼다. 그 합은 φ의 리만 합이다. 단계별 증명은 연습 1.13(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 50쪽에 있다.

연속성 보정 정수값 확률변수의 구간 $\{a \le S_n \le b\}$는 실수축에서 $[a - \frac{1}{2},\, b + \frac{1}{2}]$와 같은 칸을 차지한다. 따라서 다음 근사가 훨씬 정확하다.

$$P(a \le S_n \le b) \approx \Phi\Big(\frac{b + \frac{1}{2} - np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{a - \frac{1}{2} - np}{\sqrt{npq}}\Big).$$

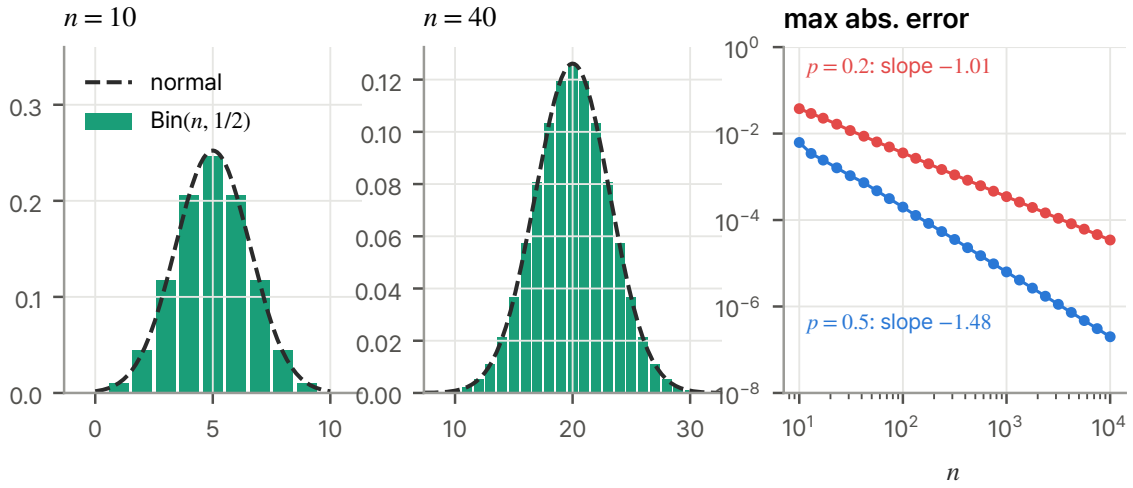


그림 1.1 이항 pmf(막대)와 정규 근사(점선). 오른쪽: $\max_k |P(S_n = k) - \text{근사}|$의 감소 속도.

> 예를 들어 $n=100$, $p=\frac{1}{2}$에서 $P(45\le S_n\le 55)$의 정확한 값은 0.7287이다. 보정 없는 근사는 0.6827, 보정한 근사는 0.7287이다.

연습문제

> **연습 1.1** (3차 항의 계수 ★★) 정리 1.2의 증명 단계 ④에서 $\log(1+w)=w-\frac{w^2}{2}+\frac{w^3}{3}+O(w^4)$까지 써서, $\log(\text{나})=-\frac{\delta^2}{2npq}+\frac{(q-p)\,\delta^3}{6n^2p^2q^2}+O(n^{-1})$임을 보여라. 풀이 48쪽

> **연습 1.2** (수치로 확인하기 ★) $S\sim\mathrm{Bin}(100,\frac{1}{2})$이다. (a) $P(S=50)$의 정확한 값과 정리 1.2의 근사를 비교하라. (b) $P(45\le S\le 55)$를 정확히, 보정 없이, 연속성 보정으로 세 번 계산하라. 풀이 48쪽

> **연습 1.3** (Stirling의 상수를 de Moivre로 구하기 ★★) Stirling 공식의 상수를 모른다고 하자: $m!=C\sqrt{m}\,(m/e)^m(1+o(1))$. (a) 정리 1.2의 증명을 그대로 따라가면 $P(S_n=k)=\frac{1}{C\sqrt{npq}}e^{-x_k^2/2}(1+o(1))$임을 보여라. (b) $\sum_k P(S_n=k)=1$과 따름정리 1.3의 논리를 쓰면 $\frac{1}{C}\int_{-\infty}^{\infty}e^{-x^2/2}dx=1$임을 보이고, 1.4절의 $\int e^{-x^2/2}dx=\sqrt{2\pi}$에서 $C=\sqrt{2\pi}$를 결론지어라. (꼬리의 기여가 작다는 것은 Chebyshev 부등식으로 처리한다.) 풀이 48쪽

1.2 Gauss (1809): "평균이 가장 그럴듯하다"면 오차는 정규다

> **역사** 1801년 1월 1일 Piazzi가 발견한 소행성 Ceres는 6주쯤 관측된 뒤 태양 빛 속으로 사라졌다. 24세의 Gauss는 그 짧은 호로 궤도를 계산했고, 그해 말 Ceres는 Gauss가 예측한 자리 근처에서 다시 발견되었다. Gauss는 1809년 *Theoria Motus Corporum Coelestium*에서 이 방법을 정리하며 오차분포를 거꾸로 이끌어 냈다. 관측자들이 늘 쓰던 산술평균이 "가장 그럴듯한 값"이 되려면 오차는 어떤 분포여야 하는가?

먼저 해석학의 작은 보조정리가 필요하다.

> **보조정리 1.4** (Cauchy 함수방정식) $g:\mathbb{R}\to\mathbb{R}$가 연속이고 모든 a,b에 대해 $g(a+b)=g(a)+g(b)$이면, $c=g(1)$에 대해 $g(u)=cu$이다.

증명. **전략** 정수, 유리수, 실수의 순서로 넓힌다. 정수에서는 귀납법 [T3], 유리수에서는 $k\,g(m/k)=g(m)$, 실수에서는 연속성과 유리수의 조밀성 [T5]을 쓴다. 단계별 증명은 연습 1.14(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 51쪽에 있다.

정리 1.5 (Gauss의 특성화) 위치모형 $X_i = \theta + \varepsilon_i$, $\varepsilon_i \overset{\text{iid}}{\sim} f$를 생각한다. $f > 0$은 미분가능한 밀도이고 $g = f'/f$는 연속이라 하자. 모든 $n \geq 1$과 모든 표본 $(x_1, \dots, x_n)$에 대해 가능도 $L(\theta) = \prod_{i=1}^{n} f(x_i - \theta)$가 $\theta = \bar{x}$에서 최대라면,

$$f(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2/(2\sigma^2)} \qquad \text{인 } \sigma > 0 \text{이 존재한다.}$$

증명. **전략** "$\bar{x}$가 최대점"이라는 가정을 1차 조건으로 바꾸면, $g = (\log f)'$가 합이 0인 수열을 합이 0인 수열로 보낸다는 조건이 된다. 표본을 마음대로 고를 수 있으므로 이 조건에서 Cauchy 함수방정식이 나온다.

① $\ell(\theta) = \sum_i \log f(x_i - \theta)$는 미분가능하고 내부의 점 $\bar{x}$에서 최대이므로 $0 = \ell'(\bar{x}) = -\sum_{i=1}^{n} g(x_i - \bar{x})$이다. — 최대점의 1차 조건

② 합이 0인 임의의 $(u_1, \dots, u_n)$에 대해 표본을 $x_i = u_i$로 잡으면 $\bar{x} = 0$이고 $x_i - \bar{x} = u_i$이다. 따라서 $u_1 + \cdots + u_n = 0$이면 $g(u_1) + \cdots + g(u_n) = 0$이다. — 가정이 모든 표본에서 성립

③ $n = 1$에서 $g(0) = 0$, $n = 2$와 $(u, -u)$에서 $g(-u) = -g(u)$, $n = 3$과 $(a, b, -a-b)$에서 $g(a) + g(b) + g(-(a+b)) = 0$이다. 홀함수이므로 모든 a, b에서 $g(a+b) = g(a) + g(b)$이다. — ②를 $n = 1, 2, 3$에

④ g는 연속이므로 $g(u) = cu$이다. — 보조정리 1.4

⑤ $(\log f)'(u) = cu$를 적분하면 $\log f(u) = \frac{c}{2}u^2 + \text{상수}$, 곧 $f(u) = A\, e^{cu^2/2}$이다. — 미적분의 기본정리

⑥ $c \geq 0$이면 $f \geq A > 0$이어서 $\int f = \infty$이므로 $c < 0$이다. $c = -1/\sigma^2$로 쓰고 $\int f = 1$을 쓰면 $A = 1/(\sigma\sqrt{2\pi})$이다. — 정리 1.8(i)

■

가장 그럴듯한 값 Gauss에게 "가장 그럴듯한 값"은 사전분포가 균등할 때의 사후분포 최빈값이었다. 사후밀도는 가능도에 비례하므로 이는 오늘날의 최대가능도추정량(MLE)이다. 16회 베이즈에서 이 해석으로 돌아온다.

순환을 끊은 Laplace Gauss는 "평균이 최선 ⇒ 정규"를 보였고, 정규를 가정해 최소제곱(13회)을 정당화했다. 그러면 평균이 최선이라는 전제는 어디서 오는가? 1810년 Laplace의 **중심극한정리**가 이 순환을 끊었다. 오차가 작은 독립 원인 여럿의 합이면 그 분포는 **가정하지 않아도** 정규에 가깝다. Stigler(1986, 4장)는 이 결합을 **Gauss–Laplace 종합**이라 부른다.

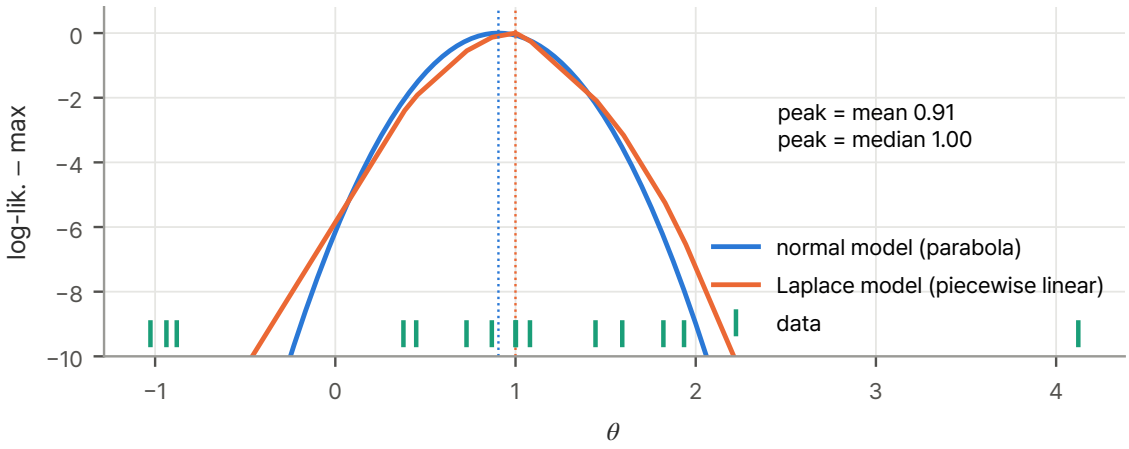


그림 1.2 Laplace 오차로 만든 같은 표본($n = 15$)에 두 모형을 맞춘 로그가능도. 정규 모형의 꼭대기는 평균, Laplace 모형의 꼭대기는 중앙값이다.

연습문제

연습 1.4 (Laplace 오차의 MLE는 중앙값 ★) $f(u) = \frac{1}{2}e^{-|u|}$일 때 $\ell(\theta) = -\sum_i |x_i - \theta| +$ 상수이다. n이 홀수이면 ℓ을 최대로 하는 θ가 표본 중앙값 하나뿐이고, n이 짝수이면 가운데 두 값 사이의 모든 점임을 보여라. 풀이 48쪽

연습 1.5 ($n = 2$만으로는 부족하다 ★★) 로지스틱 밀도 $f(u) = e^{-u}/(1+e^{-u})^2$를 생각한다. (a) $(\log f)'(u) = -\tanh(u/2)$임을 보여라. (b) f가 $n = 2$일 때 정리 1.5의 조건(가능도가 $\bar{x}$에서 최대)을 만족함을 보여라. (c) 표본 $(0, 0, 3)$에서 $\ell'(\bar{x}) \neq 0$임을 계산해, $n = 3$에서는 조건이 깨짐을 보여라. 증명의 어느 단계가 $n = 3$을 꼭 필요로 했는가? 풀이 49쪽

연습 1.6 (균등하지 않은 사전분포 ★★) $X_1, \dots, X_n \mid \theta \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2)$이고 사전분포가 $\theta \sim \mathcal{N}(m, \tau^2)$이면, 사후분포의 최빈값이

$$\hat{\theta} = \frac{(n/\sigma^2)\,\bar{x} + (1/\tau^2)\,m}{n/\sigma^2 + 1/\tau^2}$$

임을 보여라. $\tau \to \infty$와 $n \to \infty$일 때 각각 무엇이 되는가? (2.2절의 "축소"와 16회의 예고다.) 풀이 49쪽

1.3 Quetelet (1846): 오차의 분포가 사람에게 쓰이다

역사 벨기에의 천문학자 Adolphe Quetelet는 1846년 *Lettres sur la théorie des probabilités*에서 스코틀랜드 병사 5,738명의 가슴둘레 표(원자료는 1817년 *Edinburgh Medical and Surgical Journal*)에 정규분포를 맞췄다. 그는 사람의 신체가 **평균인**(l'homme moyen)을 중심으로 관측오차처럼 흩어진다고 주장했다. 천문학의 오차법칙이 처음으로 **사람들 사이의 변이**에 쓰인 순간이다. 이 성공은 동시에 "자연의 모든 것은 정규"라는 과신을 낳았다. 5절에서 그 과신을 부순다.

가슴둘레는 1인치 단위로 기록되었다. 값 x_j의 도수를 f_j, $N=\sum f_j$라 하자.

예제 1.6 (구간화된 자료의 정규 적합)

가슴둘레	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
병사 수	3	18	81	185	420	749	1073	1079	934	658	370	92	50	21	4	1

적률 추정치는 $\hat\mu=\sum f_j x_j/N=39.832$, $\hat\sigma^2=\sum f_j(x_j-\hat\mu)^2/N=4.200$, $\hat\sigma=2.049$인치다. 칸 x_j의 기대도수는

$$e_j = N\Big[\Phi\Big(\frac{x_j+0.5-\hat\mu}{\hat\sigma}\Big)-\Phi\Big(\frac{x_j-0.5-\hat\mu}{\hat\sigma}\Big)\Big].$$

기대도수가 5보다 작은 양 끝의 칸을 이웃과 합치면 칸이 13개 남는다. Pearson의 $\chi^2=\sum(f_j-e_j)^2/e_j$를 계산하면 35.0, 자유도 $13-1-2=10$, $p\approx 0.0001$이다(8회 예고). 그림으로는 정규와 거의 구별되지 않지만, $N=5{,}738$이나 되는 표본은 작은 어긋남도 잡아낸다. 가장 큰 기여는 44인치 칸이다(관측 92, 기대 146).

가정 주의 기각은 "정규가 쓸모없다"는 뜻이 아니다. 표본이 크면 실질적으로 무시할 만한 어긋남도 유의해진다. 거꾸로 기각하지 못함도 참이라는 증명이 아니다. 1인치로 반올림한 자료는 꼬리의 작은 어긋남을 가리고(연습 1.8), 병사라는 선택된 집단의 이야기일 뿐이다. 이 구분은 8회(검정의 논리)와 12회(검정력)의 주제다.

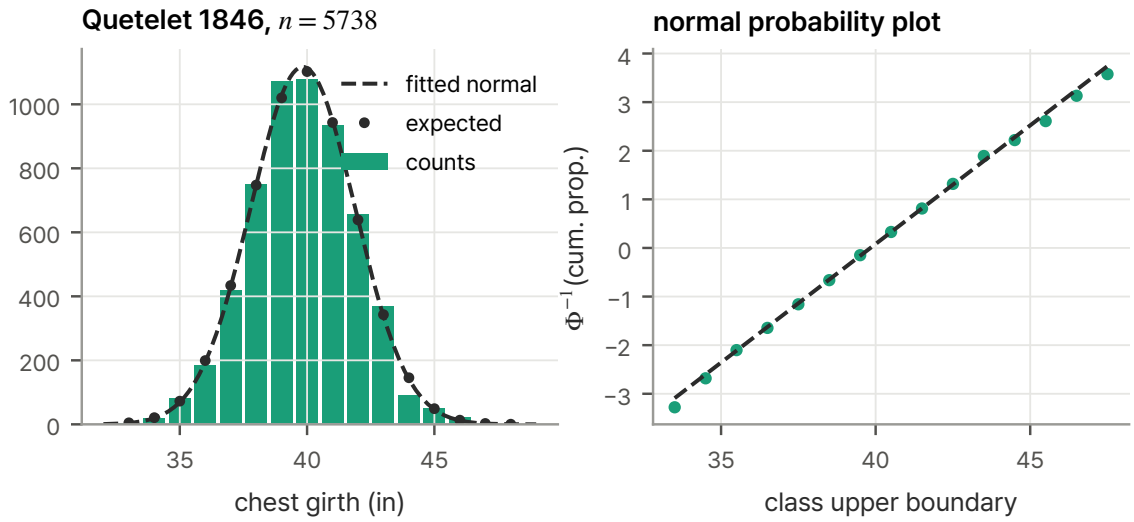


그림 1.3 왼쪽: 관측도수와 적합한 정규. 오른쪽: 누적비율의 정규분위수를 칸 경계에 대해 그린 정규확률그림. 직선이면 정규다.

연습문제

연습 1.7 (Quetelet 표 계산 ★) 위 표로 $\hat{\mu}$와 $\hat{\sigma}$를 직접 계산하라. 또 가슴둘레 44인치 이상의 관측비율과, 경계 43.5를 쓴 정규 근사를 비교하라. 풀이 49쪽

연습 1.8 (Sheppard 보정 ★★) 연속 확률변수 X를 폭 h로 반올림한 값을 $Y = X + U$로 쓰고, 반올림 오차 $U \sim \mathrm{U}(-\frac{h}{2}, \frac{h}{2})$가 X와 독립이라고 **가정**하자. $\operatorname{Var} Y = \operatorname{Var} X + h^2/12$임을 보여라. Quetelet 자료($h = 1$)에서 보정한 $\hat{\sigma}$는 얼마인가? 이 가정은 언제 그럴듯하고 언제 아닌가? 풀이 49쪽

1.4 단변량 정규분포

정의 1.7 (정규분포) 확률변수 X의 밀도가

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad x \in \mathbb{R}$$

이면 $X \sim \mathcal{N}(\mu, \sigma^2)$라 쓴다. $\mu = 0, \sigma = 1$이면 표준정규분포이고, 밀도와 분포함수를 φ, Φ로 쓴다.

정리 1.8 (단변량 정규분포의 네 가지 사실)

(i) $\int_{-\infty}^{\infty} e^{-z^2/2}\,dz = \sqrt{2\pi}$. 따라서 f는 밀도다.

(ii) $M_X(t) = \exp\!\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$, $t \in \mathbb{R}$. 따라서 $EX = \mu$, $\operatorname{Var} X = \sigma^2$이고 $E(X-\mu)^4 = 3\sigma^4$.

(iii) $Z \sim \mathcal{N}(0,1)$이면 $Z^2 \sim \chi^2(1)$.

(iv) $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$가 독립이면 $\sum a_i X_i \sim \mathcal{N}\!\left(\sum a_i\mu_i,\ \sum a_i^2\sigma_i^2\right)$.

증명. **전략** (i)은 적분을 제곱해 극좌표로 바꾼다. (ii)는 지수를 완전제곱해 (i)로 돌아간다. (iii)은 분포함수를 미분하고, (iv)는 mgf의 곱과 유일성을 쓴다.

(i) 가우스 적분 (3회 정리 5.3)

① $I = \int e^{-z^2/2}dz > 0$으로 두면 $I^2 = \int_{\mathbb{R}^2} e^{-(z_1^2+z_2^2)/2}\,dz_1\,dz_2$이다. Tonelli, 피적분함수 ≥ 0 [T6]

② 극좌표로 바꾼다.

$$I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\, r\,dr\,d\varphi = 2\pi\Big[-e^{-r^2/2}\Big]_0^{\infty} = 2\pi. \qquad \text{야코비안 } r\text{, 연습 0.1 [T7]}$$

③ $z = (x-\mu)/\sigma$로 치환하면 $\int f = I/\sqrt{2\pi} = 1$이다. 1차원 치환 [T7]

(ii) mgf와 적률

① 지수를 x에 대해 완전제곱한다.

$$tx-\frac{(x-\mu)^2}{2\sigma^2}=-\frac{\big(x-\mu-\sigma^2 t\big)^2}{2\sigma^2}+\mu t+\frac{\sigma^2 t^2}{2}. \qquad \text{전개해서 비교}$$

② 첫 항은 평균이 $\mu+\sigma^2 t$인 정규밀도의 지수이므로 적분하면 $\sigma\sqrt{2\pi}$이고, $M_X(t)=e^{\mu t+\sigma^2 t^2/2}$이다. (i)

③ $\psi(t)=\log M_X(t)=\mu t+\frac{1}{2}\sigma^2 t^2$에서 $\psi'(0)=\mu$, $\psi''(0)=\sigma^2$이다. 보조정리 0.1(b)

④ $Z=(X-\mu)/\sigma$의 mgf를 두 가지로 전개해 t^4의 계수를 비교한다.

$$e^{t^2/2}=\sum_{k\ge 0}\frac{t^{2k}}{2^k\,k!},\qquad M_Z(t)=\sum_{m\ge 0}\frac{EZ^m}{m!}\,t^m \quad\Longrightarrow\quad EZ^4=\frac{4!}{2^2\cdot 2!}=3.$$

거듭제곱급수 계수의 유일성

조건 확인 단계 ④의 계수 비교는 두 사실에 기댄다. 하나는 모든 t에서 유한한 M_Z가 거듭제곱급수로 전개되고 그 계수가 $M_Z^{(m)}(0)/m!=EZ^m/m!$이라는 것이다(보조정리 0.1(b)). 다른 하나는 두 거듭제곱급수가 0의 근방에서 같으면 계수가 같다는 것이다 [T8].

(iii) $Z^2\sim\chi^2(1)$ (4회 정리 3.9)

① $w>0$에서 $P(Z^2\le w)=P(-\sqrt{w}\le Z\le\sqrt{w})=2\Phi(\sqrt{w})-1$이다. φ의 대칭

② 미분한다.

$$f_{Z^2}(w)=2\varphi(\sqrt{w})\cdot\frac{1}{2\sqrt{w}}=\frac{1}{\sqrt{2\pi}}\,w^{-1/2}e^{-w/2}=\frac{1}{\Gamma(\frac{1}{2})\,2^{1/2}}\,w^{\frac{1}{2}-1}e^{-w/2}. \qquad \text{연쇄법칙, } \Gamma(\tfrac{1}{2})=\sqrt{\pi}$$

③ 이것은 $\Gamma(\alpha=\frac{1}{2},\beta=2)=\chi^2(1)$의 밀도다. 4회 정의 3.8

④ mgf로도 확인된다. $t<\frac{1}{2}$에서

$$E\,e^{tZ^2}=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-(1-2t)z^2/2}\,dz=(1-2t)^{-1/2}. \qquad \text{(i), } z\sqrt{1-2t}\text{로 치환}$$

(iv) 독립 정규의 선형결합

① $M_{\sum a_iX_i}(t)=\prod_i M_{X_i}(a_it)$이다. 독립, 보조정리 0.1(c)(d)

② (ii)를 넣으면 $\prod_i\exp\!\big(a_i\mu_i t+\frac{1}{2}a_i^2\sigma_i^2t^2\big)=\exp\!\big(t\sum_i a_i\mu_i+\frac{1}{2}t^2\sum_i a_i^2\sigma_i^2\big)$이다. (ii)

③ 이것은 $\mathcal{N}\!\big(\sum a_i\mu_i,\sum a_i^2\sigma_i^2\big)$의 mgf이므로 결론이 나온다. mgf의 유일성, 보조정리 0.1(a)

■

따름정리 1.9 ($\chi^2(n)$) $Z_1, \dots, Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$이면 $Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n)$이다.

증명. (iii)과 독립성으로 mgf가 $\left[(1-2t)^{-1/2}\right]^n = (1-2t)^{-n/2}$이고, 이것은 $\Gamma(\frac{n}{2}, 2) = \chi^2(n)$의 mgf다. ∎

연습문제

연습 1.9 (짝수 적률 ★) $Z \sim \mathcal{N}(0,1)$이면 $EZ^{2k} = (2k-1)!! = 1 \cdot 3 \cdots (2k-1)$이고 홀수 적률은 0임을 두 가지로 보여라. (a) mgf 급수의 계수 비교. (b) $\varphi'(z) = -z\varphi(z)$와 부분적분으로 $EZ^{m+1} = m\,EZ^{m-1}$(Stein 항등식의 특수형). 풀이 49쪽

연습 1.10 (로그정규분포 ★) $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = e^X$이다. mgf만 써서 $EY = e^{\mu+\sigma^2/2}$, $\operatorname{Var} Y = e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$을 구하라. $EY \neq e^{EX}$인 이유를 Jensen 부등식으로 설명하라. 풀이 50쪽

연습 1.11 (정규 꼬리의 부등식(Mills 비) ★★) $z > 0$에서 다음을 보여라. $z = 2$에서 세 값을 계산해 비교하라.

$$\left(\frac{1}{z} - \frac{1}{z^3}\right)\varphi(z) \;\le\; 1 - \Phi(z) \;\le\; \frac{\varphi(z)}{z}$$

풀이 50쪽

연습 1.12 (합성곱으로 직접 ★★) $X \sim \mathcal{N}(0, \sigma_1^2)$과 $Y \sim \mathcal{N}(0, \sigma_2^2)$가 독립일 때, mgf를 쓰지 않고 합성곱 $f_{X+Y}(s) = \int f_X(x) f_Y(s-x)\,dx$를 완전제곱으로 계산해 $X+Y \sim \mathcal{N}(0, \sigma_1^2+\sigma_2^2)$임을 보여라. 풀이 50쪽

연습 1.13 (적분형 de Moivre–Laplace 정리의 증명 ★★ 유도형) 따름정리 1.3를 다음 순서로 증명하라.

(a) 구하는 확률이 $\sum_{k:\, x_k \in [a,b]} P(S_n = k)$이고, 인접한 x_k의 간격이 $\Delta = 1/\sqrt{npq}$임을 보여라.

(b) 정리 1.2를 $A = \max(|a|, |b|)$로 써서 이 합이 $\left(1 + O(n^{-1/2})\right) \sum_{k:\, x_k \in [a,b]} \varphi(x_k)\,\Delta$임을 보여라. 근사가 k에 대해 균등해야 하는 이유는 무엇인가?

(c) 칸 $[x_k, x_k + \Delta)$들을 생각해 $\left|\sum_{x_k \in [a,b]} \varphi(x_k)\Delta - \int_a^b \varphi\right| \le (b-a+\Delta)\,\Delta \max|\varphi'| + 2\Delta \max \varphi$임을 보여라. (칸 위에서는 평균값 정리를 쓰고, 칸들의 합집합과 $[a,b]$가 다른 부분은 양 끝의 길이 Δ 이하인 두 조각이다.)

(d) $\Delta \to 0$이고 $\int_a^b \varphi = \Phi(b) - \Phi(a)$이므로 결론이 나옴을 보여라.

풀이 50쪽

연습 1.14 (Cauchy 함수방정식의 증명 ★★ 유도형) 보조정리 1.4를 다음 순서로 증명하라.

(a) $a = b = 0$과 $b = -a$를 넣어 $g(0) = 0$과 $g(-a) = -g(a)$를 얻어라.

(b) 자연수 m에 대해 $g(mu) = m\,g(u)$를 m에 대한 귀납법으로 보이고, (a)로 모든 정수 m에 넓혀 $g(m) = cm$을 얻어라.

(c) $m \in \mathbb{Z}$, $k \in \mathbb{N}$이면 $k\,g(m/k) = g(m)$임을 보이고 $g(m/k) = c\,m/k$를 얻어라.

(d) 실수 u로 수렴하는 유리수열을 잡고 연속성으로 $g(u) = cu$를 얻어라. 연속성 가정은 이 단계에서만 쓰인다는 것을 확인하라.

풀이 51쪽

2

Galton의 문제: 변이는 어떻게 유전되는가

> **역사** Francis Galton(1822–1911)은 Darwin의 사촌이었다. *Hereditary Genius*(1869) 이후 그의 관심은 하나였다. **변이는 어떻게 다음 세대로 전해지는가?** Stigler(1986, 8장)에 따르면 Galton을 괴롭힌 것은 역설이었다. 사람의 키는 부모를 닮는데도 인구 전체의 키 분포는 세대가 지나도 거의 변하지 않는다. 닮음과 안정을 동시에 설명하는 수학이 필요했다.

2.1 Quincunx (1873): 작은 충격의 합이 종을 만든다

공이 i번째 못에서 오른쪽으로 가면 $\xi_i = +\frac{1}{2}$, 왼쪽이면 $\xi_i = -\frac{1}{2}$라 하자. 못마다 독립이고 $P(\xi_i = \pm\frac{1}{2}) = \frac{1}{2}$이면 n줄을 지난 위치는

$$P_n = \sum_{i=1}^{n} \xi_i = K_n - \frac{n}{2}, \qquad K_n = \#\{\text{오른쪽}\} \sim \mathrm{Bin}\left(n, \tfrac{1}{2}\right)$$

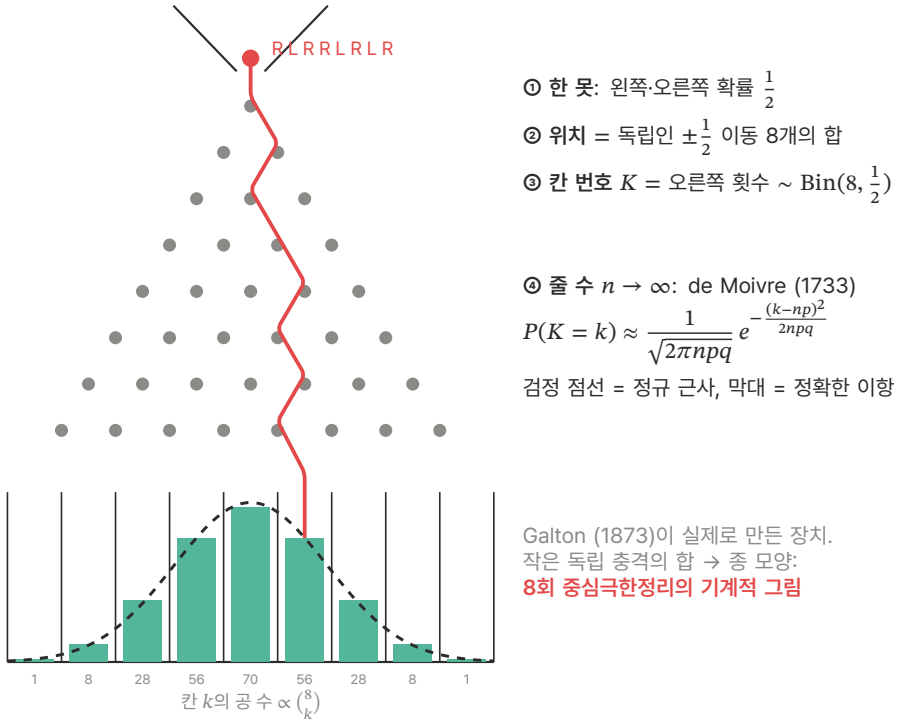


그림 2.1 Galton 판(quincunx). 공 하나의 경로(빨강)와 8줄 판의 칸별 도수.

이다($\xi_i = B_i - \frac{1}{2}$, $B_i \sim$ Bernoulli). $E\xi_i = 0$, $\operatorname{Var}\xi_i = \frac{1}{4}$이므로 $EP_n = 0$, $\operatorname{Var}P_n = \frac{n}{4}$이고, 정리 1.2에 의해

$$P(P_n = j) \approx \frac{1}{\sqrt{2\pi\, n/4}} \exp\!\left(-\frac{j^2}{2 \cdot n/4}\right).$$

장치 자체가 "작은 독립 충격의 합은 정규"라는 중심극한정리(7회)의 기계적 그림이다.

연습문제

> **연습 2.1** (가운데 칸 ★) 16줄 판에서 공이 가운데 칸($P_{16} = 0$)에 떨어질 확률을 정확히 계산하고 정규 근사와 비교하라. 풀이 51쪽

> **연습 2.2** (기울어진 판 ★) 공이 오른쪽으로 갈 확률이 $p \neq \frac{1}{2}$이면 EP_n과 $\operatorname{Var}P_n$은 무엇인가? 정규 근사는 어떻게 바뀌는가? 풀이 51쪽

2.2 2단 quincunx (1877): 정규를 정규로 섞으면 다시 정규

1877년 왕립연구소 강연 *Typical Laws of Heredity*에서 Galton은 중간에 칸을 단 quincunx를 보였다. 1단을 지난 공은 중간 칸에 멈춘다. 칸 하나만 열면 그 칸 아래에 작은 종이 생기고, 모든 칸을 열면 작은 종들이 합쳐져 처음과 같은 모양의 큰 종이 된다.

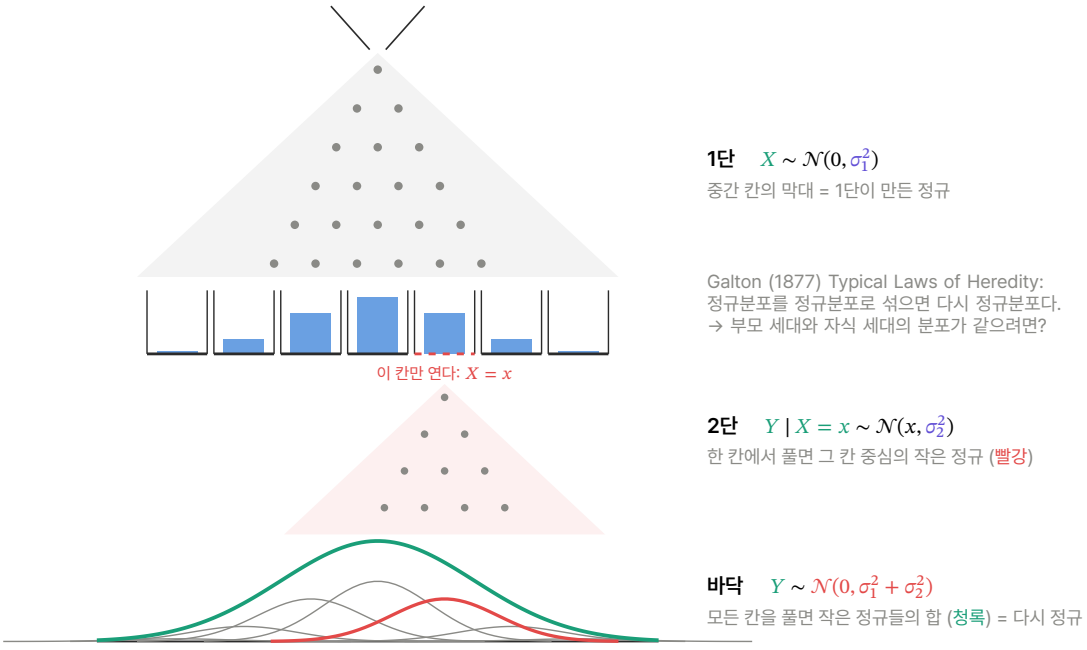


그림 2.2 Galton의 2단 quincunx. 중간 칸 하나만 열면 작은 종(빨강), 모두 열면 작은 종들의 합(청록)이 다시 정규다.

정리 2.1 (정규의 정규 혼합) $X \sim \mathcal{N}(0, \sigma_1^2)$이고 $Y \mid X = x \sim \mathcal{N}(x, \sigma_2^2)$이면

(a) $Y \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$이다.

(b) (X, Y)는 이변량 정규이고 $\mathrm{Cov}(X, Y) = \sigma_1^2$, $\mathrm{Corr}(X, Y) = \sigma_1 / \sqrt{\sigma_1^2 + \sigma_2^2}$이다.

증명. **전략** (a)는 X에 조건을 걸어 mgf를 구한다(반복기댓값). (b)는 $D = Y - X$가 X와 독립인 정규임을 보이고, $(X, Y) = (X, X + D)$의 결합 mgf를 정의 3.4의 꼴로 쓴다. 단계별 증명은 연습 2.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 52쪽에 있다.

분산만 보면 전분산 공식 그대로다: $\mathrm{Var}\, Y = E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(E[Y \mid X]) = \sigma_2^2 + \sigma_1^2$.

이제 2단 quincunx를 **거꾸로** 읽어 보자. 바닥에서 공의 위치 y만 보고, 그 공이 어느 중간 칸에서 왔는지 추측한다면?

따름정리 2.2 (거꾸로 본 조건부분포: 축소) 정리 2.1의 상황에서

$$X \mid Y = y \;\sim\; \mathcal{N}\!\left(B\,y,\; \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right), \qquad B = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} < 1.$$

증명. **전략** 베이즈 정리로 조건부밀도의 로그를 x의 이차식으로 쓴다. x^2의 계수에서 정밀도(분산의 역수)를, 완전제곱에서 평균을 읽는다. 단계별 증명은 연습 2.9(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 52쪽에 있다.

축소(shrinkage) 바닥의 위치 y를 그대로 중간 칸의 추측으로 쓰지 않고, 중심 쪽으로 $B < 1$배 줄여 쓴다. 이것이 Galton의 **회귀**이며 동시에 사전분포 $\mathcal{N}(0, \sigma_1^2)$, 관측 $\mathcal{N}(x, \sigma_2^2)$의 **베이즈 사후평균**이다. 그림 2.2의 수치($\sigma_1 = 1.2$, $\sigma_2 = 0.8$)에서는 $B = 0.692$, 사후분산 0.443이다. 16회에서 Stein의 축소추정량이 바로 이 구조임을 본다.

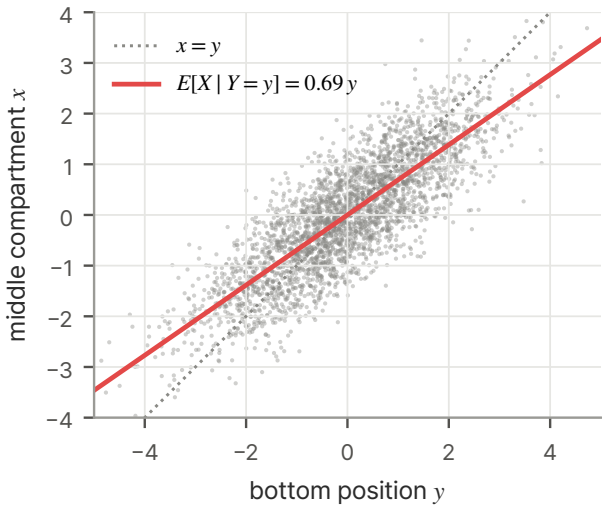


그림 2.3 2단 quincunx 3,000개 공의 (바닥, 중간 칸). 중간 칸의 조건부 평균은 $x = y$가 아니라 기울기 $B = 0.69$의 직선이다.

연습문제

연습 2.3 (기울기로 확인 ★) 정리 2.1(b)의 공분산행렬만 써서 $E[X \mid Y = y]$의 기울기가 $\mathrm{Cov}(X,Y)/\operatorname{Var} Y = B$임을 확인하라(3.7절의 이변량 회귀식을 미리 써도 좋다). 풀이 51쪽

연습 2.4 (측정을 늘리면 ★★) $X \sim \mathcal{N}(m, \tau^2)$이고, X를 독립적으로 n번 측정한 평균 $\bar{Y} \mid X \sim \mathcal{N}(X, s^2/n)$을 관측한다. $E[X \mid \bar{Y} = \bar{y}]$와 $\operatorname{Var}(X \mid \bar{Y})$를 구하고 $n \to \infty$일 때를 해석하라. 연습 1.6와 같은 답인 이유는? 풀이 51쪽

2.3 세대 간 분산의 수수께끼와 reversion

부모의 편차가 **전부** 물려지고 독립 잡음 E가 더해진다고 해 보자($C = P + E$). 그러면 $\operatorname{Var} C = \operatorname{Var} P + \sigma_E^2$이므로 분산은 세대마다 커져야 한다. 실제로는 그렇지 않다.

명제 2.3 (안정은 회귀를 강요한다) $C - \mu = r(P - \mu) + E$, $E \perp P$, $\operatorname{Var} P = \sigma^2$, $\operatorname{Var} E = \sigma_E^2 < \sigma^2$, $r \geq 0$이라 하자. 분포가 안정하다는 조건 $\operatorname{Var} C = \operatorname{Var} P$가 성립하면

$$r = \sqrt{1 - \sigma_E^2/\sigma^2} < 1 \qquad \text{이고} \qquad r = \mathrm{Corr}(P, C).$$

증명. **전략** 독립인 두 조각의 분산을 더해 안정 조건을 r의 방정식으로 바꾸고, 공분산의 쌍선형성으로 상관을 구한다. 단계별 증명은 연습 2.10(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 53쪽에 있다.

회귀는 생물학적인 힘이 아니라, 분포가 안정하다는 사실의 대수적 귀결이다. Galton은 1877년 스위트피 씨앗 실험에서 이 계수를 *reversion*이라 불렀고, 1886년 사람의 키에서 *regression*이라 다시 불렀다.

Galton의 실제 자료는 한 부모가 아니라 **중간부모**(아버지와 1.08배 한 어머니 키의 평균)를 썼다.

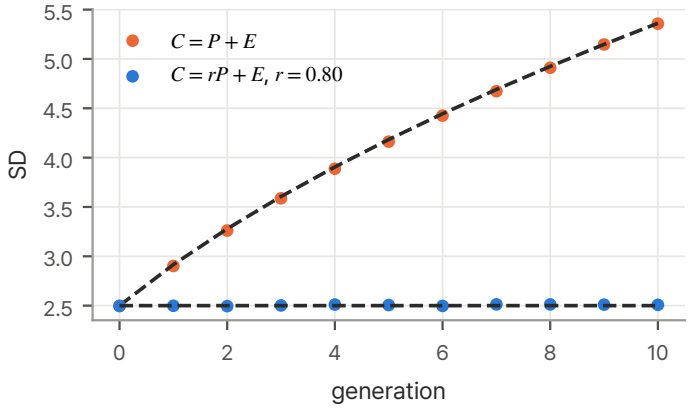


그림 2.4 회귀가 없으면($C = P + E$) 표준편차가 $\sqrt{\sigma^2 + g\sigma_E^2}$로 자라고, $r = \sqrt{1 - \sigma_E^2/\sigma^2}$이면 그대로다.

이 차이가 Galton의 유명한 두 숫자 2/3과 1/3을 설명한다.

예제 2.4 (중간부모 모형) 아버지 F와 (1.08배 한) 어머니 M_o가 독립 $\mathcal{N}(\mu, \sigma^2)$이고, 중간부모 $M = (F + M_o)/2$, 자녀

$$C - \mu = \beta\,(M - \mu) + E, \qquad E \perp (F, M_o)$$

라 하자. $\operatorname{Var} M = \sigma^2/2$이다. 안정 조건 $\operatorname{Var} C = \sigma^2$에서 $\beta^2\sigma^2/2 + \sigma_E^2 = \sigma^2$. 이때

$$\begin{aligned} \text{중간부모에 대한 기울기} &= \frac{\operatorname{Cov}(C, M)}{\operatorname{Var} M} = \beta, \\ \operatorname{Corr}(C, M) &= \frac{\beta \operatorname{Var} M}{\sigma\sqrt{\operatorname{Var} M}} = \beta\sqrt{\tfrac{1}{2}} = \frac{\beta}{\sqrt{2}}, \\ \text{한 부모에 대한 기울기} &= \frac{\operatorname{Cov}(C, F)}{\operatorname{Var} F} = \frac{\beta \operatorname{Cov}(M, F)}{\sigma^2} = \frac{\beta\,(\sigma^2/2)}{\sigma^2} = \frac{\beta}{2}. \end{aligned}$$

Galton의 $\beta \approx \frac{2}{3}$를 넣으면 상관은 0.471, 한 부모 기울기는 $\frac{1}{3}$이다. Galton(1886)이 보고한 두 숫자가 정확히 이것이고, 1886년 자료의 실제 상관은 0.459, 중간부모 분산 대 자녀 분산의 비는 $3.19/6.33 = 0.504 \approx \frac{1}{2}$이다.

연습문제

연습 2.5 (설명되는 분산 ★) 예제 2.4에서 $\beta = \frac{2}{3}$이면 $\sigma_E^2/\sigma^2 = \frac{7}{9}$임을 보여라. 중간부모는 자녀 키 분산의 몇 퍼센트를 설명하는가? 풀이 52쪽

연습 2.6 (부부의 키가 상관되면 ★★) 예제 2.4에서 $\operatorname{Corr}(F, M_o) = a$(동류 교배)라 하자. $\operatorname{Var} M$, 안정 조건, $\operatorname{Corr}(C, M)$, 한 부모에 대한 기울기를 β와 a로 나타내라. 풀이 52쪽

2.4 Galton (1886): 928명의 표와 동심 타원

1886년 논문 *Regression towards Mediocrity in Hereditary Stature*에서 Galton은 205가족의 성인 자녀 928명의 키를 중간부모 키와 함께 표로 만들었다. 남녀를 한 표에 넣으려고 **여성의 키에 1.08을 곱했다**(5.3절에서 검토한다). Galton은 표에서 같은 도수의 칸을 이으면 **비슷한 모양의 동심 타원**이 된다는 것을 알아챘다. 케임브리지의 수학자 J. D. Hamilton Dickson에게 이유를 묻자, Dickson은 이변량 정규분포의 식(예제 3.7)을 써서 답했다.

표 2.1 Galton(1886) 원자료 요약. $n = 928$, 단위 인치, 분산은 분모 n.

	평균	표준편차	상관	회귀 기울기
중간부모 x	68.31	1.786	$r = 0.459$	$b_{y\mid x} = 0.646$
자녀 y	68.09	2.517		$b_{x\mid y} = 0.326$

예제 2.5 (예측과 거꾸로 예측) 적합한 이변량 정규의 조건부 평균(3.7절)은 $\hat{E}[y \mid x] = 68.09 + 0.646\,(x - 68.31) = 23.94 + 0.646\,x$이다.

- 중간부모가 72인치(평균보다 3.7인치 위)이면 자녀의 기대 키는 70.47인치(평균보다 2.4인치 위)다.
- 중간부모가 64인치이면 65.30인치다. 아래쪽도 평균 쪽으로 당겨진다.
- 거꾸로, 자녀가 73인치이면 중간부모의 기대 키는 $68.31 + 0.326(73 - 68.09) = 69.91$인치다.
- 조건부 표준편차는 $\hat{\sigma}_y\sqrt{1 - r^2} = 2.236$이므로, 중간부모 72인치 가족 자녀의 95% 예측구간은 $70.47 \pm 1.96 \times 2.236 = [66.09,\ 74.86]$이다.

연습문제

연습 2.7 (예측구간 ★) 표 2.1로 중간부모 66인치 가족 자녀의 기대 키와 95% 예측구간을 구하라. 이 구간은 어떤 가정 아래에서 정당한가? 풀이 52쪽

연습 2.8 (정규의 정규 혼합: 증명 ★★ 유도형) 정리 2.1을 다음 순서로 증명하라.

(a) 반복기댓값(보조정리 0.3)과 정리 1.8(ii)로 $M_Y(t) = E\big[E(e^{tY} \mid X)\big] = e^{(\sigma_1^2+\sigma_2^2)t^2/2}$을 보여라.

(b) $D = Y - X$로 두면 $X = x$가 주어질 때 $D \sim \mathcal{N}(0, \sigma_2^2)$이다. 이 조건부분포가 x에 의존하지 않으므로 $D \perp X$임을 보여라. ($E[h(D)\,k(X)] = E\big[k(X)\,E(h(D) \mid X)\big]$를 쓴다.)

(c) $Y = X + D$로 결합 mgf $E\,e^{sX+tY}$를 계산하고, 정의 3.4의 꼴 $\exp\{\tfrac{1}{2}(s,t)\Sigma(s,t)'\}$로 써서 Σ를 읽어라.

(d) $\mathrm{Cov}(X,Y)$와 $\mathrm{Corr}(X,Y)$를 Σ에서 읽고, 분산만은 전분산 공식으로도 나옴을 확인하라.

풀이 52쪽

연습 2.9 (축소: 증명 ★ 유도형) 따름정리 2.2를 다음 순서로 증명하라.

(a) 베이즈 정리로 $f(x \mid y) \propto f_X(x)\,f_{Y|X}(y \mid x)$임을 쓰고, $\log f(x \mid y)$에서 x에 의존하는 부분을 모아라.

(b) 그 이차식을 $-\frac{\tau}{2}x^2 + \frac{y}{\sigma_2^2}x +$ 상수 꼴로 정리해 τ를 구하라.

(c) 완전제곱으로 $X \mid Y = y$가 정규이고 평균이 By, 분산이 $1/\tau$임을 보여라.

(d) $\sigma_1^2 \to \infty$와 $\sigma_2^2 \to 0$일 때 B는 어떻게 되는가? 각각 무엇을 뜻하는가?

풀이 52쪽

연습 2.10 (안정은 회귀를 강요한다: 증명 ★ 유도형) 명제 2.3을 다음 순서로 증명하라.

(a) $E \perp P$에서 $\mathrm{Var}\,C = r^2\sigma^2 + \sigma_E^2$임을 보여라.

(b) 안정 조건과 $r \ge 0$에서 $r = \sqrt{1 - \sigma_E^2/\sigma^2}$을 얻고, $\sigma_E^2 < \sigma^2$이 왜 필요한지 밝혀라.

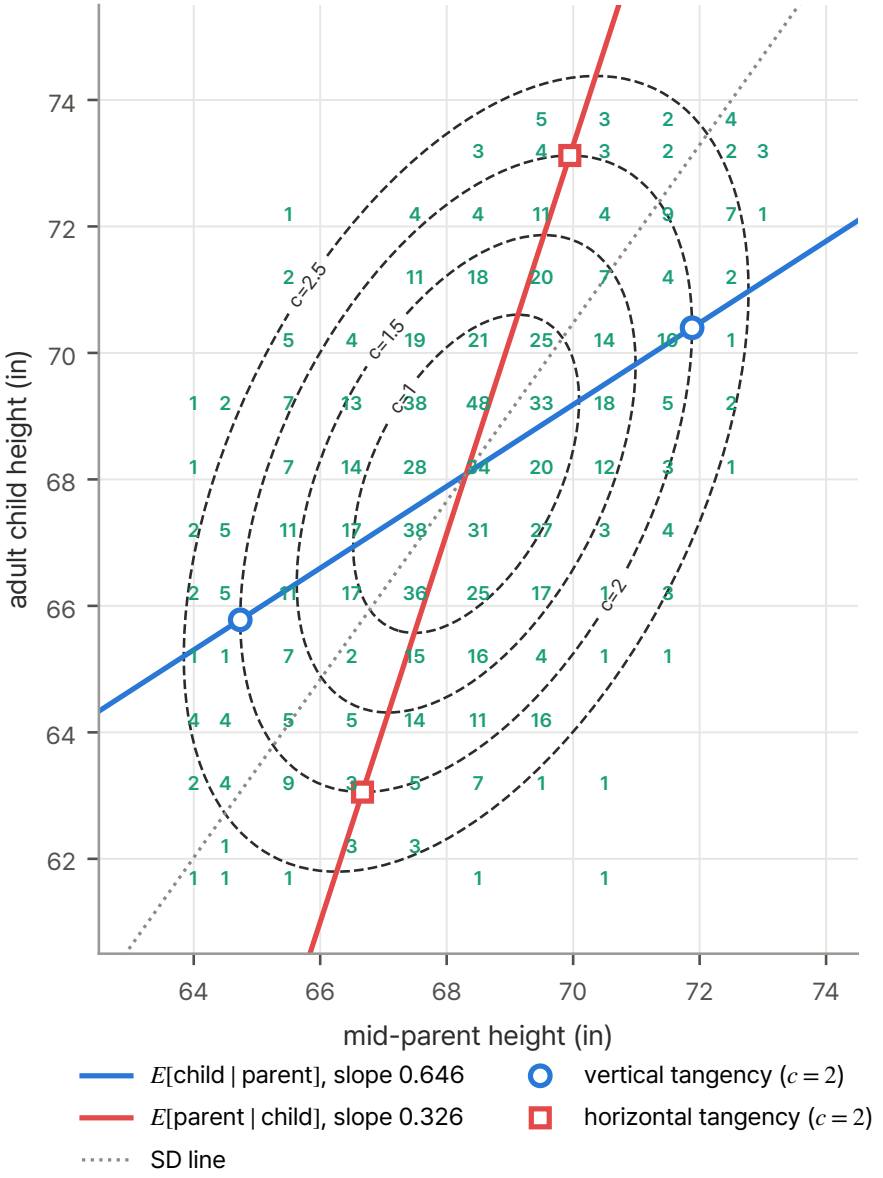


그림 2.5 Galton(1886) 표의 재구성. 청록 숫자는 칸 도수, 점선은 적합한 이변량 정규의 등고선 $Q = c^2$이다. 두 회귀선은 타원의 수직 접점(○)과 수평 접점(□)을 지난다(정리 3.20).

(c) $\mathrm{Cov}(P, C) = r\sigma^2$을 보이고 $\mathrm{Corr}(P, C) = r$을 얻어라.

풀이 53쪽

3

다변량 정규분포

Galton의 타원을 설명하려면 정규분포를 여러 변수로 넓혀야 한다. 이 절의 목표는 정리 A–D(그림 0.1)를 엄밀하게 세우는 것이다. 먼저 필요한 행렬 도구를 증명과 함께 정리한다.

3.1 행렬 준비

확률벡터 $\mathbf{X} = (X_1, \dots, X_n)'$의 평균은 $\boldsymbol{\mu} = E\mathbf{X}$, 공분산행렬은 $\Sigma = \operatorname{Cov}\mathbf{X} = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'$, 곧 $\Sigma_{ij} = \operatorname{Cov}(X_i, X_j)$이다.

보조정리 3.1 (선형변환의 평균과 공분산) $m \times n$ 상수행렬 A와 $\mathbf{b} \in \mathbb{R}^m$에 대해 $E(A\mathbf{X}+\mathbf{b}) = A\boldsymbol{\mu}+\mathbf{b}$, $\operatorname{Cov}(A\mathbf{X}+\mathbf{b}) = A\Sigma A'$이다. 특히 $\mathbf{a}'\Sigma\mathbf{a} = \operatorname{Var}(\mathbf{a}'\mathbf{X}) \geq 0$이므로 Σ는 대칭 반정치(positive semidefinite)다.

증명. 기댓값의 선형성으로 첫 식이 나온다. 그러면 $\operatorname{Cov}(A\mathbf{X}+\mathbf{b}) = E\big[A(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'A'\big] = A\Sigma A'$. ∎

보조정리 3.2 (스펙트럼 분해, 제곱근, Cholesky) 대칭 반정치 Σ는 직교행렬 V와 $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_i \geq 0$으로 $\Sigma = V\Lambda V'$라 쓸 수 있다.

(a) $\Sigma^{1/2} := V\Lambda^{1/2}V'$는 대칭이고 $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$이다. Σ가 양정치이면 $\Sigma^{-1/2} := V\Lambda^{-1/2}V' = (\Sigma^{1/2})^{-1}$이다.

(b) Σ가 양정치이면 대각원소가 양수인 하삼각행렬 L이 유일하게 존재해 $\Sigma = LL'$이다(Cholesky 분해).

(a)는 $V'V = I$로 바로 확인되고, (b)는 선형대수 교재의 표준 결과다. 정규분포를 **구성**할 때는 $\Sigma = AA'$인 행렬 A 하나만 있으면 되고, (a)의 $\Sigma^{1/2}$와 (b)의 L이 모두 그런 A다.

이 절에서 가장 중요한 도구는 블록 행렬이다. 다음 보조정리 하나가 정리 C의 두 번째 증명, 정밀도행렬 표현, 행렬식 공식을 모두 준다.

보조정리 3.3 (블록 행렬과 Schur 보수) 대칭 행렬 Σ를 블록으로 나누고 Σ_{22}가 가역이라 하자. 다음과 같이 둔다.

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \qquad T = \begin{pmatrix} I & -B \\ 0 & I \end{pmatrix},$$

$$B = \Sigma_{12}\Sigma_{22}^{-1}, \qquad \Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad \text{(Schur 보수)}.$$

(a) (블록 대각화) $T\,\Sigma\,T' = \operatorname{diag}(\Sigma_{11\cdot 2},\ \Sigma_{22})$.

(b) (행렬식) $|\Sigma| = |\Sigma_{22}|\,|\Sigma_{11\cdot 2}|$.

(c) (양정치) $\Sigma \succ 0 \iff \Sigma_{22} \succ 0$이고 $\Sigma_{11\cdot 2} \succ 0$.

(d) (역행렬) Σ가 가역이면

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11\cdot 2}^{-1} & -\Sigma_{11\cdot 2}^{-1}B \\ -B'\Sigma_{11\cdot 2}^{-1} & \Sigma_{22}^{-1} + B'\Sigma_{11\cdot 2}^{-1}B \end{pmatrix}.$$

(e) (이차형식 분해) $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$에 대해

$$\mathbf{x}'\Sigma^{-1}\mathbf{x} = (\mathbf{x}_1 - B\mathbf{x}_2)'\,\Sigma_{11\cdot 2}^{-1}\,(\mathbf{x}_1 - B\mathbf{x}_2) + \mathbf{x}_2'\Sigma_{22}^{-1}\mathbf{x}_2.$$

증명. Σ가 대칭이므로 $\Sigma_{21} = \Sigma_{12}'$이고 $B' = \Sigma_{22}^{-1}\Sigma_{21}$이다.

(a) 먼저 왼쪽에 T를 곱하면

$$T\Sigma = \begin{pmatrix} \Sigma_{11} - B\Sigma_{21} & \Sigma_{12} - B\Sigma_{22} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

이다($B\Sigma_{22} = \Sigma_{12}$). 여기에 오른쪽으로 T'를 곱하면

$$T\Sigma T' = \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} I & 0 \\ -B' & I \end{pmatrix} = \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ \Sigma_{21} - \Sigma_{22}B' & \Sigma_{22} \end{pmatrix},$$

이고 $\Sigma_{22}B' = \Sigma_{21}$이므로 (2,1) 블록이 0이 된다.

(b) $\det T = 1$이므로 (a)의 양변에 행렬식을 취한다.

(c) T는 가역이므로 $\Sigma \succ 0 \iff T\Sigma T' \succ 0$이고, 블록 대각행렬은 각 블록이 양정치일 때 그리고 그때만 양정치다.

(d) (a)에서 $\Sigma = T^{-1}DT^{-\prime}$, $D = \operatorname{diag}(\Sigma_{11\cdot 2}, \Sigma_{22})$이므로 $\Sigma^{-1} = T'D^{-1}T$. 블록 곱을 계산하면

$$\begin{pmatrix} I & 0 \\ -B' & I \end{pmatrix}\begin{pmatrix} \Sigma_{11\cdot 2}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} = \begin{pmatrix} \Sigma_{11\cdot 2}^{-1} & -\Sigma_{11\cdot 2}^{-1}B \\ -B'\Sigma_{11\cdot 2}^{-1} & B'\Sigma_{11\cdot 2}^{-1}B + \Sigma_{22}^{-1} \end{pmatrix}.$$

(e) $\mathbf{x}'\Sigma^{-1}\mathbf{x} = \mathbf{x}'T'D^{-1}T\mathbf{x} = (T\mathbf{x})'D^{-1}(T\mathbf{x})$이고 $T\mathbf{x} = (\mathbf{x}_1 - B\mathbf{x}_2,\ \mathbf{x}_2)$이다. ∎

연습문제

연습 3.1 (2×2에서 확인 ★) 다음 $\Sigma(|\rho|<1)$에서 $\Sigma_{11\cdot 2}$, $|\Sigma|$, Σ^{-1}을 보조정리 3.3로 구하고 직접 계산과 비교하라.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

풀이 53쪽

연습 3.2 (정밀도행렬의 블록 ★★) $\Lambda=\Sigma^{-1}$을 같은 크기로 나눠 $\Lambda_{11},\Lambda_{12},\dots$ 라 하자. 보조정리 3.3(d)에서 $\Lambda_{11}=\Sigma_{11\cdot 2}^{-1}$이고 $-\Lambda_{11}^{-1}\Lambda_{12}=B$임을 보여라. 풀이 53쪽

3.2 정의 · 구성 · 밀도

정의 3.4 (다변량 정규분포 (mgf에 의한 정의)) n차원 확률벡터 $\mathbf{X}$의 mgf가 모든 $\mathbf{t}\in\mathbb{R}^n$에서

$$M_{\mathbf{X}}(\mathbf{t}) = E\,e^{\mathbf{t}'\mathbf{X}} = \exp\!\left(\mathbf{t}'\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right)$$

이면($\boldsymbol{\mu}\in\mathbb{R}^n$, Σ는 $n\times n$ 대칭 반정치) $\mathbf{X}\sim\mathcal{N}_n(\boldsymbol{\mu},\Sigma)$라 쓴다.

왜 mgf로 정의하는가 (a) Σ가 특이해 밀도가 없는 경우까지 포함한다(5.4절). (b) 정리 A가 한 줄로 끝난다. (c) mgf가 분포를 결정한다(보조정리 0.1(a)). 단점은 정의가 비어 있지 않음, 곧 그런 분포가 존재함을 따로 보여야 한다는 것이다. 다음 정리가 그 일을 한다.

정리 3.5 (존재와 모수의 의미) $\boldsymbol{\mu}\in\mathbb{R}^n$, Σ가 대칭 반정치이고 $AA'=\Sigma$라 하자. $Z_1,\dots,Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, $\mathbf{Z}=(Z_1,\dots,Z_n)'$이면 $\mathbf{X}=\boldsymbol{\mu}+A\mathbf{Z}\sim\mathcal{N}_n(\boldsymbol{\mu},\Sigma)$이다. 또한 $E\mathbf{X}=\boldsymbol{\mu}$, $\operatorname{Cov}\mathbf{X}=\Sigma$이다.

증명. Z_i의 독립성과 정리 1.8(ii)($M_{Z_i}(s)=e^{s^2/2}$)로

$$M_{\mathbf{X}}(\mathbf{t}) = E\,e^{\mathbf{t}'\boldsymbol{\mu}+\mathbf{t}'A\mathbf{Z}} = e^{\mathbf{t}'\boldsymbol{\mu}}\,E\,e^{(A'\mathbf{t})'\mathbf{Z}} = e^{\mathbf{t}'\boldsymbol{\mu}}\prod_{i=1}^{n} e^{(A'\mathbf{t})_i^2/2} = \exp\!\left(\mathbf{t}'\boldsymbol{\mu}+\tfrac{1}{2}\|A'\mathbf{t}\|^2\right),$$

그리고 $\|A'\mathbf{t}\|^2=\mathbf{t}'AA'\mathbf{t}=\mathbf{t}'\Sigma\mathbf{t}$이다. 보조정리 3.2로 그런 A는 항상 존재한다. 평균과 공분산은 보조정리 3.1에 $E\mathbf{Z}=\mathbf{0}$, $\operatorname{Cov}\mathbf{Z}=I$를 넣으면 나온다. ∎

분포는 A가 아니라 $AA'=\Sigma$에만 의존한다. 대칭 제곱근을 쓰든 Cholesky 인수를 쓰든 같은 분포가 나온다. 이 구성은 그대로 **샘플러**다(4절 알고리즘 4.1).

정리 3.6 (밀도) $\Sigma \succ 0$이면 $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$는 밀도

$$\boxed{f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}} \exp\!\left\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}, \qquad \mathbf{x} \in \mathbb{R}^n}$$

를 갖는다.

증명. **전략** 밀도를 아는 $\mathbf{Z}$에서 출발해 가역 선형변환 $\mathbf{x} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{z}$로 옮기고, 변수변환 공식을 쓴다 [T7].

① 정리 3.5에서 $A = \Sigma^{1/2}$로 잡으면 $\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$이다. — 정리 3.5

② $\Sigma \succ 0$이면 $\Sigma^{1/2}$는 가역이므로 $\mathbf{x} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{z}$는 $\mathbb{R}^n$에서 $\mathbb{R}^n$으로의 일대일 변환이고, 역변환은 $\mathbf{z} = \Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$이다. — 보조정리 3.2(a)

③ 역변환의 야코비안은 $\det \Sigma^{-1/2} = |\Sigma|^{-1/2}$이다. — $\det \Sigma^{-1/2} = \prod \lambda_i^{-1/2}$

④ $\mathbf{Z}$의 밀도는 $\prod_i \varphi(z_i) = (2\pi)^{-n/2} e^{-\mathbf{z}'\mathbf{z}/2}$이다. — Z_i 독립

⑤ $\mathbf{z}'\mathbf{z} = (\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1/2}\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu}) = (\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$이다. — $\Sigma^{-1/2}$는 대칭

⑥ ③–⑤를 변수변환 공식에 넣으면 결론이다. — 보조정리 0.2 [T7]

■

예제 3.7 (이변량 정규 밀도 — Dickson의 식) $n = 2$이고 $|\rho| < 1$이면

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}, \qquad |\Sigma| = \sigma_X^2\sigma_Y^2(1-\rho^2),$$

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1/\sigma_X^2 & -\rho/(\sigma_X\sigma_Y) \\ -\rho/(\sigma_X\sigma_Y) & 1/\sigma_Y^2 \end{pmatrix}.$$

표준화 변수 $u = \frac{x-\mu_X}{\sigma_X}$, $w = \frac{y-\mu_Y}{\sigma_Y}$로 쓰면

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\!\left\{-\frac{u^2 - 2\rho uw + w^2}{2(1-\rho^2)}\right\}.$$

지수의 이차형식 $Q = \frac{u^2 - 2\rho uw + w^2}{1-\rho^2}$이 상수인 곡선이 Galton이 본 타원이다.

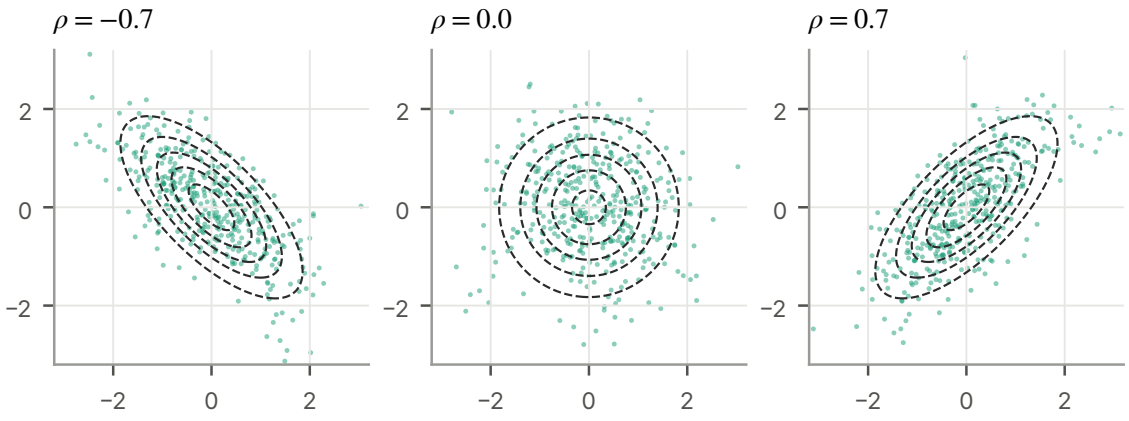


그림 3.1 표준화 이변량 정규. ρ에 따라 등고 타원이 기운다.

연습문제

연습 3.3 (인수분해와 $\rho = 0$ ★) 예제 3.7의 $f(x,y)$가 x만의 함수와 y만의 함수의 곱으로 쓰일 필요충분조건이 $\rho = 0$임을 보여라. 풀이 54쪽

연습 3.4 (사분면 확률 ★★) (X_1, X_2)가 표준화 이변량 정규($\sigma = 1$, 상관 ρ)이면 $P(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{\arcsin \rho}{2\pi}$ 임을 보여라. (힌트: $X_2 = \rho X_1 + \sqrt{1-\rho^2}\, Z$, $Z \perp X_1$로 쓰고 (X_1, Z)의 회전대칭을 이용한다.) 풀이 54쪽

3.3 정리 A — 선형변환

정리 3.8 (정리 A: 선형변환) $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, C는 $m \times n$ 상수행렬, $\mathbf{b} \in \mathbb{R}^m$이면

$$\mathbf{Y} = C\mathbf{X} + \mathbf{b} \;\sim\; \mathcal{N}_m\!\left(C\boldsymbol{\mu} + \mathbf{b},\; C\Sigma C'\right).$$

증명. $\mathbf{t}'(C\mathbf{X}) = (C'\mathbf{t})'\mathbf{X}$이므로 $\mathbf{X}$의 mgf를 $C'\mathbf{t}$에서 읽으면 된다.

$$M_{\mathbf{Y}}(\mathbf{t}) = E\, e^{\mathbf{t}'(C\mathbf{X}+\mathbf{b})} = e^{\mathbf{t}'\mathbf{b}}\, M_{\mathbf{X}}(C'\mathbf{t}) = \exp\!\Big(\mathbf{t}'(C\boldsymbol{\mu}+\mathbf{b}) + \tfrac{1}{2}\,\mathbf{t}'C\Sigma C'\mathbf{t}\Big).$$

$C\Sigma C'$는 대칭 반정치이므로 정의 3.4에 맞는다. ∎

따름정리 3.9 (주변분포) $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$로 나누면 $\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_{11})$이다. 특히 각 성분은 $X_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$이다.

증명. $C = \begin{bmatrix} I & 0 \end{bmatrix}$에 정리 3.8를 쓴다. ∎

결합정규의 주변분포는 모두 정규다. 그러나 **역은 거짓이다**(5.1절). 무엇이 결합정규를 판정하는가? 답은 "모든 선형결합"이다.

정리 3.10 (Cramér–Wold 판정) $\mathbf{X}$가 결합정규일 필요충분조건은 모든 $\mathbf{a} \in \mathbb{R}^n$에 대해 $\mathbf{a}'\mathbf{X}$가 (분산이 0인 경우를 포함한) 단변량 정규인 것이다.

증명. **전략** ($\Rightarrow$)는 정리 A의 특수한 경우다. ($\Leftarrow$)는 $\mathbf{X}$의 mgf를 $\mathbf{t}$마다 한 변수 $\mathbf{t}'\mathbf{X}$의 mgf의 값 $M_{\mathbf{t}'\mathbf{X}}(1)$로 읽는다. 단계별 증명은 연습 3.16(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 56쪽에 있다.

연습문제

연습 3.5 (선형결합 하나 ★) $\mathbf{X} \sim \mathcal{N}_3(\boldsymbol{\mu}, \Sigma)$일 때 $X_1 - 2X_2 + X_3$의 분포를 구하라.

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 4 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

풀이 54쪽

연습 3.6 (특이 공분산과 상수 선형결합 ★★) $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$이고 $\operatorname{rank}\Sigma < n$이면, $P(\mathbf{a}'\mathbf{X} = \mathbf{a}'\boldsymbol{\mu}) = 1$인 $\mathbf{a} \neq \mathbf{0}$이 존재함을 보여라. 풀이 54쪽

3.4 정리 B — 독립과 무상관

정리 3.11 (정리 B: 독립 ⇔ 무상관) $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')' \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$이면 $\mathbf{X}_1 \perp \mathbf{X}_2 \iff \Sigma_{12} = 0$.

증명. **전략** ($\Leftarrow$)에서 결합 mgf의 지수를 블록으로 펼치면 Σ_{12}가 들어 있는 교차항 하나만 두 블록을 잇는다. 그 항이 사라지면 mgf가 두 주변 mgf의 곱으로 쪼개진다.

($\Rightarrow$)

① 독립이면 $\mathbf{X}_1$의 성분과 $\mathbf{X}_2$의 성분은 쌍마다 독립이므로 공분산이 0이고, $\Sigma_{12} = 0$이다. 독립이면 무상관

($\Leftarrow$)

① $\mathbf{t} = (\mathbf{t}_1', \mathbf{t}_2')'$로 나누면 $\mathbf{t}'\Sigma\mathbf{t} = \mathbf{t}_1'\Sigma_{11}\mathbf{t}_1 + 2\,\mathbf{t}_1'\Sigma_{12}\mathbf{t}_2 + \mathbf{t}_2'\Sigma_{22}\mathbf{t}_2$이다. 블록 곱, $\Sigma_{21} = \Sigma_{12}'$

② $\Sigma_{12} = 0$이면 가운데 항이 사라진다.

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\!\Big(\mathbf{t}_1'\boldsymbol{\mu}_1 + \tfrac{1}{2}\,\mathbf{t}_1'\Sigma_{11}\mathbf{t}_1\Big)\,\exp\!\Big(\mathbf{t}_2'\boldsymbol{\mu}_2 + \tfrac{1}{2}\,\mathbf{t}_2'\Sigma_{22}\mathbf{t}_2\Big). \qquad \text{정의 3.4}$$

③ 두 인수는 각각 $M_{\mathbf{X}_1}(\mathbf{t}_1)$, $M_{\mathbf{X}_2}(\mathbf{t}_2)$이다. 따름정리 3.9

④ 결합 mgf가 $\mathbf{0}$의 근방에서 주변 mgf의 곱이므로 $\mathbf{X}_1 \perp \mathbf{X}_2$이다. 보조정리 0.1

■

가정 주의 가정은 "$\mathbf{X}_1$과 $\mathbf{X}_2$가 각각 정규"가 아니라 "**$(\mathbf{X}_1, \mathbf{X}_2)$가 결합정규**"다. 증명에서 **결합** mgf의 식을 쓴 곳이 바로 이 가정이다. 5.1절에서 이 가정을 빼면 무상관이지만 종속인 정규쌍이 생긴다.

따름정리 3.12 (성분의 상호 독립) $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$의 성분 $X_1, \dots, X_n$이 상호 독립일 필요충분조건은 Σ가 대각행렬인 것이다. 결합정규에서는 **쌍별 무상관이 상호 독립을 준다.**

증명. Σ가 대각이면 mgf가 n개 인수로 쪼개진다. ■

6회에서는 정리 3.11로 정규표본의 $\bar{X} \perp S^2$를 증명한다. $(\bar{X}, X_1 - \bar{X}, \dots, X_n - \bar{X})$가 결합정규이고 공분산이 0이기 때문이다.

연습문제

연습 3.7 (합과 차 ★) (X_1, X_2)가 결합정규이고 $\operatorname{Var} X_1 = \operatorname{Var} X_2$이면 $X_1 + X_2 \perp X_1 - X_2$임을 보여라. 분산이 다르면 어떻게 되는가? 풀이 54쪽

3.5 정리 C — 조건부분포

정리 3.13 (정리 C: 조건부분포) $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')' \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$이고 $\Sigma_{22} \succ 0$이면

$$\boxed{\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2 \;\sim\; \mathcal{N}\Big(\underbrace{\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)}_{\boldsymbol{\mu}_{1\cdot 2}},\; \underbrace{\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}_{\Sigma_{11\cdot 2}} \Big)}$$

증명 1 (무상관 분해). **전략** $\mathbf{X}_1$에서 $\mathbf{X}_2$로 설명되는 부분 $B\mathbf{X}_2$를 빼 잔차 $\mathbf{W} = \mathbf{X}_1 - B\mathbf{X}_2$를 만든다. B를 잘 고르면 $\mathbf{W}$와 $\mathbf{X}_2$가 무상관이고, 결합정규이므로 독립이다. 그러면 $\mathbf{X}_2$에 조건을 걸어도 $\mathbf{W}$의 분포는 바뀌지 않는다.

$B = \Sigma_{12}\Sigma_{22}^{-1}$, $\mathbf{W} = \mathbf{X}_1 - B\mathbf{X}_2$로 둔다.

① $(\mathbf{W}, \mathbf{X}_2)$는 결합정규다.

$$\begin{pmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = T\mathbf{X}.$$ 보조정리 3.3의 T, 정리 3.8

② $\operatorname{Cov}(\mathbf{W}, \mathbf{X}_2) = \operatorname{Cov}(\mathbf{X}_1, \mathbf{X}_2) - B \operatorname{Cov}(\mathbf{X}_2, \mathbf{X}_2) = \Sigma_{12} - B\Sigma_{22} = 0$이다. 공분산의 쌍선형성

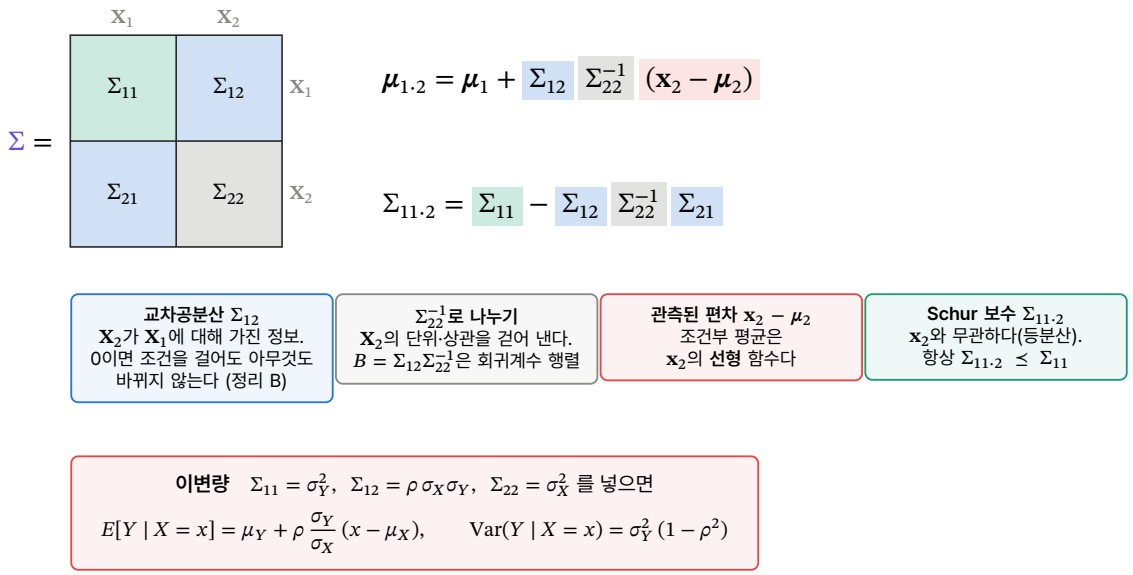


그림 3.2 정리 C 공식의 해부. 블록마다 한 색을 쓴다.

③ $\mathbf{W} \perp \mathbf{X}_2$이다. ①②, 정리 3.11

④ $E\mathbf{W} = \boldsymbol{\mu}_1 - B\boldsymbol{\mu}_2$이고, 공분산은 다음과 같다.

$$\operatorname{Cov}\mathbf{W} = \Sigma_{11} - B\Sigma_{21} - \Sigma_{12}B' + B\Sigma_{22}B' = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11\cdot 2}. \qquad \text{보조정리 3.1}$$

뒤의 세 항은 모두 $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$이다.

⑤ $\mathbf{X}_1 = \mathbf{W} + B\mathbf{X}_2$이고 $\mathbf{W}$는 $\mathbf{X}_2$와 독립이므로, $\mathbf{X}_2 = \mathbf{x}_2$로 조건을 걸어도 $\mathbf{W}$의 분포는 그대로다. 따라서 $\mathbf{X}_1 \mid \mathbf{x}_2$는 $\mathbf{W} + B\mathbf{x}_2$의 분포, 곧 $\mathcal{N}(\boldsymbol{\mu}_1 - B\boldsymbol{\mu}_2 + B\mathbf{x}_2,\ \Sigma_{11\cdot 2})$이다. ③④

■

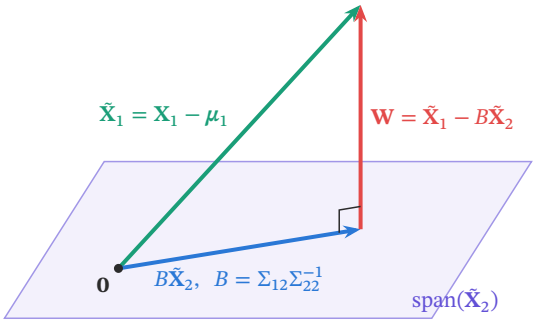
증명 2 (밀도의 비, $\Sigma > 0$일 때). $\tilde{\mathbf{x}} = \mathbf{x} - \boldsymbol{\mu}$로 쓰고 보조정리 3.3(e)를 쓰면

$$\tilde{\mathbf{x}}'\Sigma^{-1}\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2)'\Sigma_{11\cdot 2}^{-1}(\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2) + \tilde{\mathbf{x}}_2'\Sigma_{22}^{-1}\tilde{\mathbf{x}}_2,$$

(b)로 $|\Sigma|^{1/2} = |\Sigma_{11\cdot 2}|^{1/2}|\Sigma_{22}|^{1/2}$이다. 따라서 결합밀도는 두 정규밀도의 곱으로 쪼개진다.

$$f(\mathbf{x}) = \underbrace{\frac{\exp\{-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_{1\cdot 2})'\Sigma_{11\cdot 2}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_{1\cdot 2})\}}{(2\pi)^{n_1/2}|\Sigma_{11\cdot 2}|^{1/2}}}_{f(\mathbf{x}_1 \mid \mathbf{x}_2)} \cdot \underbrace{\frac{\exp\{-\frac{1}{2}\tilde{\mathbf{x}}_2'\Sigma_{22}^{-1}\tilde{\mathbf{x}}_2\}}{(2\pi)^{n_2/2}|\Sigma_{22}|^{1/2}}}_{f_2(\mathbf{x}_2)}$$

($\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2 = \mathbf{x}_1 - \boldsymbol{\mu}_{1\cdot 2}$). 둘째 인수는 따름정리 3.9의 주변밀도이므로 첫째 인수가 조건부밀도 $f(\mathbf{x})/f_2(\mathbf{x}_2)$다. ■



같은 그림이 13회 최소제곱(hat matrix)과 11회 Rao–Blackwell에서 다시 나온다.

① **직교** $\mathrm{Cov}(\mathbf{W}, \mathbf{X}_2) = \Sigma_{12} - B\Sigma_{22} = \mathbf{0}$
② **결합정규** $(\mathbf{W}, \mathbf{X}_2)$는 $\mathbf{X}$의 선형변환 (정리 A)
③ **독립** ①+② ⇒ $\mathbf{W} \perp \mathbf{X}_2$ (정리 B)
④ **조건부** $\mathbf{X}_1 = \mathbf{W} + B\mathbf{X}_2$ 에서 $\mathbf{X}_2 = \mathbf{x}_2$를 넣으면
$\mathbf{X}_1 \mid \mathbf{x}_2 \sim \mathcal{N}\big(\boldsymbol{\mu}_1 + B(\mathbf{x}_2 - \boldsymbol{\mu}_2),\ \mathrm{Cov}(\mathbf{W})\big)$

피타고라스 (= 전분산 공식)
$\Sigma_{11} = \underbrace{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}_{\mathrm{Cov}(E[\mathbf{X}_1\mid\mathbf{X}_2])} + \underbrace{(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}_{E[\mathrm{Cov}(\mathbf{X}_1\mid\mathbf{X}_2)] = \mathrm{Cov}(\mathbf{W})}$

그림 3.3 증명 1의 기하. 내적 $\langle U, V\rangle = \mathrm{Cov}(U, V)$에서 $B\tilde{\mathbf{X}}_2$는 $\tilde{\mathbf{X}}_1$의 직교사영이고 $\mathbf{W}$는 그 잔차다.

따름정리 3.14 (조건부분포의 세 성질) 정리 3.13의 상황에서

(a) (선형성) $E[\mathbf{X}_1 \mid \mathbf{X}_2] = \boldsymbol{\mu}_1 + B(\mathbf{X}_2 - \boldsymbol{\mu}_2)$는 $\mathbf{X}_2$의 선형 함수다.

(b) (등분산) $\mathrm{Cov}(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2) = \Sigma_{11\cdot 2}$는 $\mathbf{x}_2$에 의존하지 않는다.

(c) (피타고라스) 공분산이 두 조각으로 나뉜다.

$$\Sigma_{11} = \underbrace{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}_{\mathrm{Cov}(E[\mathbf{X}_1\mid\mathbf{X}_2])} + \underbrace{\Sigma_{11\cdot 2}}_{E[\mathrm{Cov}(\mathbf{X}_1\mid\mathbf{X}_2)]}.$$

따라서 $\Sigma_{11\cdot 2} \preceq \Sigma_{11}$이다. 조건을 걸면 불확실성이 줄어든다.

증명. (a)(b)는 정리에서 바로 읽힌다. (c)는 $\mathrm{Cov}(B\mathbf{X}_2) = B\Sigma_{22}B' = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$이므로 정리의 Schur 보수 정의를 옮겨 쓴 것이고, 전분산 공식(보조정리 0.3)과 같은 식이다. $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \succeq 0$이므로 부등식이 나온다. ∎

조건부 평균과 분산은 공분산행렬 대신 **정밀도행렬** $\Lambda = \Sigma^{-1}$로도 쓸 수 있다. 이 표현이 노트북의 독립 검산과 가우시안 그래프 모형의 출발점이다.

정리 3.15 (정밀도행렬 표현) $\Sigma \succ 0$, $\Lambda = \Sigma^{-1}$을 같은 크기로 나누면

$$\Sigma_{11\cdot 2} = \Lambda_{11}^{-1}, \qquad \boldsymbol{\mu}_{1\cdot 2} = \boldsymbol{\mu}_1 - \Lambda_{11}^{-1}\Lambda_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2).$$

증명. 보조정리 3.3(d)에서 $\Lambda_{11} = \Sigma_{11\cdot 2}^{-1}$, $\Lambda_{12} = -\Sigma_{11\cdot 2}^{-1}B$이므로 $-\Lambda_{11}^{-1}\Lambda_{12} = B$이다(연습 3.2). ∎

조건부 독립과 정밀도의 0 $\mathbf{X}_1 = (X_i, X_j)$, $\mathbf{X}_2 =$ 나머지 전부로 두면 정리 3.15에서 $\mathrm{Cov}\big((X_i, X_j) \mid \text{나머지}\big) = \Lambda_{\{ij\}}^{-1}$이고, 2×2 행렬의 역의 비대각원소는 $-\Lambda_{ij}/\det$이다. 따라서 $X_i \perp X_j \mid$ **나머지** $\iff \Lambda_{ij} = 0$이다(연습 3.11). Σ의 0은 **주변** 무상관, Λ의 0은 **조건부** 무상관이다. 가우시안 그래프 모형과 graphical lasso가 이 사실 위에 서 있다.

예제 3.16 (3 × 3 손계산) 다음 분포에서 $X_1 \mid (X_2, X_3) = (x_2, x_3)$를 구하자.

$$\mu = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 4 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

$$\Sigma_{22} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad |\Sigma_{22}| = 3, \quad \Sigma_{22}^{-1} = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \qquad \Sigma_{12} = (2,\ 0),$$

$$B = \Sigma_{12}\Sigma_{22}^{-1} = \tfrac{1}{3}(4, -2) = \left(\tfrac{4}{3}, -\tfrac{2}{3}\right),$$

$$\mu_{1\cdot 2} = 1 + \tfrac{4}{3}(x_2 - 0) - \tfrac{2}{3}(x_3 + 1), \qquad \Sigma_{11\cdot 2} = 4 - B\Sigma_{21} = 4 - \left(\tfrac{4}{3}\cdot 2 + (-\tfrac{2}{3})\cdot 0\right) = \tfrac{4}{3}.$$

예를 들어 $(x_2, x_3) = (1, 0)$이면 $X_1 \mid (1,0) \sim \mathcal{N}(\frac{5}{3}, \frac{4}{3})$이다.

검산. $|\Sigma| = 4$이고

$$\Lambda = \Sigma^{-1} = \frac{1}{4}\begin{pmatrix} 3 & -4 & 2 \\ -4 & 8 & -4 \\ 2 & -4 & 4 \end{pmatrix}.$$

정리 3.15으로 $1/\Lambda_{11} = \frac{4}{3}$, $-\Lambda_{12}/\Lambda_{11} = \frac{4}{3}$, $-\Lambda_{13}/\Lambda_{11} = -\frac{2}{3}$로 일치한다. 보조정리 3.3(b)로도 $|\Sigma| = |\Sigma_{22}| \cdot \Sigma_{11\cdot 2} = 3 \cdot \frac{4}{3} = 4$.

읽을거리. $\Sigma_{13} = 0$이므로 X_1과 X_3은 무상관이고, 결합정규이므로 독립이다. 그런데 X_2를 알면 X_3의 계수가 $-\frac{2}{3} \neq 0$이다($\Lambda_{13} = \frac{1}{2} \neq 0$). **주변으로는 독립인데 조건부로는 종속**이다. 공통의 자손 X_2를 통해 연결된 구조다.

연습문제

연습 3.8 (다른 조건 걸기 ★) 예제 3.16의 분포에서 $X_3 \mid (X_1, X_2) = (x_1, x_2)$를 구하고 Λ로 검산하라. 풀이 54쪽

연습 3.9 (조건을 더 걸면 ★) 같은 분포에서 $X_2 \mid X_1 = x_1$과 $X_2 \mid (X_1, X_3) = (x_1, x_3)$을 구하라. 조건부 분산을 비교하고 따름정리 3.14(c)와 연결하라. 풀이 55쪽

연습 3.10 (최선의 예측과 최선의 선형 예측 ★★) (a) 분산이 유한한 임의의 (X, Y)에서 모든 함수 g에 대해 $E[(Y - g(X))^2] \geq E[(Y - E[Y \mid X])^2]$임을 보여라. (b) $E[(Y - a - bX)^2]$를 최소로 하는 (a, b)가 $b = \mathrm{Cov}(X,Y)/\mathrm{Var}\,X$, $a = EY - b\,EX$임을 보여라. (c) (X, Y)가 결합정규이면 (a)와 (b)의 답이 같음을 보여라. 풀이 55쪽

연습 3.11 (조건부 독립 $\Leftrightarrow$ $\Lambda_{ij}=0$ ★★★) $\mathbf{X}\sim\mathcal{N}_n(\boldsymbol{\mu},\Sigma)$, $\Sigma\succ 0$, $i\neq j$에 대해 $X_i \perp X_j \mid (X_k)_{k\neq i,j}$ 일 필요충분조건이 $\Lambda_{ij}=0$임을 증명하라. 풀이 55쪽

3.6 정리 D — 이차형식과 확률 타원

정리 3.17 (정리 D: 이차형식은 χ^2) $\mathbf{X}\sim\mathcal{N}_n(\boldsymbol{\mu},\Sigma)$, $\Sigma\succ 0$이면 $Q=(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})\sim\chi^2(n)$.

증명. **전략** $\Sigma^{-1/2}$로 백색화하면 Q는 독립 표준정규의 제곱합이 된다.

① $\mathbf{Z}=\Sigma^{-1/2}(\mathbf{X}-\boldsymbol{\mu})\sim\mathcal{N}_n(\mathbf{0},\ \Sigma^{-1/2}\Sigma\Sigma^{-1/2})=\mathcal{N}_n(\mathbf{0},I)$이다. — 정리 3.8, 보조정리 3.2(a)

② $Z_1,\dots,Z_n$은 독립 $\mathcal{N}(0,1)$이다. — 따름정리 3.12

③ $Q=(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-1/2}\Sigma^{-1/2}(\mathbf{X}-\boldsymbol{\mu})=\mathbf{Z}'\mathbf{Z}=\sum_i Z_i^2$이다. — $\Sigma^{-1}=\Sigma^{-1/2}\Sigma^{-1/2}$

④ $Q\sim\chi^2(n)$이다. — 따름정리 1.9

■

따름정리 3.18 (확률 타원) $E_c=\{\mathbf{x}:(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\le c^2\}$에 대해 $P(\mathbf{X}\in E_c)=F_{\chi^2(n)}(c^2)$이다. E_c는 Σ의 고유벡터 $\mathbf{v}_i$ 방향으로 반지름 $c\sqrt{\lambda_i}$를 갖는 타원체이고, $n=2$이면

$$P(\mathbf{X}\in E_c)=1-e^{-c^2/2},\qquad \text{넓이}(E_c)=\pi c^2\sqrt{|\Sigma|}.$$

증명. **전략** 확률은 정리 D에서 바로 읽는다. 모양은 Σ의 고유벡터로 좌표를 회전해 축에 평행한 타원체로 만들어 읽는다. 단계별 증명은 연습 3.17(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 56쪽에 있다.

$n=2$의 95% 타원은 $c^2=-2\log 0.05=5.991$이다. 밀도는 Q만의 함수이므로 등고선이 곧 확률 타원이다.

연습문제

연습 3.12 (Galton의 95% 타원 ★) 표 2.1의 적합값으로 95% 타원의 두 반지름과 넓이를 구하라. (고유값은 7.355와 2.170이다.)

$$\hat{\Sigma}=\begin{pmatrix}3.191 & 2.062\\ 2.062 & 6.333\end{pmatrix}$$

풀이 55쪽

연습 3.13 (모든 방향을 동시에 ★★) $\Sigma \succ 0$일 때 첫 식을 보이고, 이로부터 둘째 식을 결론지어라. (14회 Scheffé 동시구간의 원형이다.)

$$\sup_{\mathbf{a}\neq\mathbf{0}} \frac{\big(\mathbf{a}'(\mathbf{X}-\boldsymbol{\mu})\big)^2}{\mathbf{a}'\Sigma\mathbf{a}} = Q,$$

$$P\Big(\text{모든 } \mathbf{a}\text{에 대해 } |\mathbf{a}'(\mathbf{X}-\boldsymbol{\mu})| \le c\sqrt{\mathbf{a}'\Sigma\mathbf{a}}\Big) = F_{\chi^2(n)}(c^2).$$

풀이 56쪽

3.7 이변량의 경우 — 회귀식, 그리고 접점

따름정리 3.19 (이변량 회귀식) (X,Y)가 예제 3.7의 이변량 정규이면

$$\boxed{E[Y\mid X=x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \qquad \mathrm{Var}(Y\mid X=x) = \sigma_Y^2(1-\rho^2)}$$

이고, X와 Y의 자리를 바꾼 식도 성립한다. 표준화하면 $E[Z_Y\mid Z_X=z]=\rho z$, $E[Z_X\mid Z_Y=z]=\rho z$.

증명. 정리 3.13에 $\Sigma_{11}=\sigma_Y^2$, $\Sigma_{12}=\rho\sigma_X\sigma_Y$, $\Sigma_{22}=\sigma_X^2$를 넣으면 $B=\rho\sigma_Y/\sigma_X$, $\Sigma_{11\cdot 2}=\sigma_Y^2-\rho^2\sigma_Y^2$. ∎

$|\rho|<1$이면 **양쪽 방향 모두** 평균 쪽으로 당겨진다. ρ^2은 Y의 분산 가운데 X가 설명하는 비율이다 ($\mathrm{Var}(E[Y\mid X])/\mathrm{Var}\,Y=\rho^2$). 회귀는 인과의 방향이 아니라 대칭적인 기하다. 이것이 여는 질문의 답이다.

정리 3.20 (접점의 자취) 회귀선 $y=m(x)=E[Y\mid X=x]$는 등고 타원들이 **수직 접선**을 갖는 점들의 자취다. 마찬가지로 $E[X\mid Y=y]$는 **수평 접선**을 갖는 점들의 자취다.

증명 1 (조건부밀도). 고정된 x에서 $f(x,y)=f_X(x)\,f_{Y|X}(y\mid x)$이고, $f_{Y|X}(\cdot\mid x)$는 $m(x)$에서 유일하게 최대인 정규밀도다. 따라서 수직선 $\{X=x\}$ 위에서 f는 $(x,m(x))$에서만 최대다. $c=f(x,m(x))$라 하면 수직선은 볼록집합 $\{f\ge c\}$(타원의 내부)와 정확히 한 점에서 만난다. 즉 수직선은 등고선 $\{f=c\}$의 받침선이고, 매끄러운 타원에서 받침선은 접선이다. ∎

증명 2 (음함수 미분). 표준화 변수로 등고선은 $Q(u,w)=\frac{u^2-2\rho uw+w^2}{1-\rho^2}=c^2$이다. 접선이 수직인 점은 기울기벡터 ∇Q가 수평인 점, 곧 $\partial Q/\partial w=0$인 점이다.

$$\frac{\partial Q}{\partial w}=\frac{2w-2\rho u}{1-\rho^2}=0 \iff w=\rho u \iff y=\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X).$$

마찬가지로 $\partial Q/\partial u=0 \iff u=\rho w$가 수평 접점의 자취이고, 이는 $E[X\mid Y]$의 식이다. ∎

$\sigma_X = \sigma_Y$이면 $\Sigma \propto \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$의 고유벡터가 $(1,1)/\sqrt{2}$(고유값 $1+\rho$)와 $(1,-1)/\sqrt{2}$(고유값 $1-\rho$)이므로, $\rho > 0$일 때 타원의 장축은 기울기 1의 'SD 선'이다. 두 회귀선은 장축을 사이에 두고 하나는 완만하고(기울기 ρ) 하나는 가파르다(그려진 기울기 $1/\rho$).

> **가정 주의** 회귀의 오류(regression fallacy). Kahneman의 비행교관 일화가 유명하다. 교관들은 "칭찬하면 다음 비행이 나빠지고 질책하면 좋아진다"고 믿었다. 그러나 두 번의 비행 성적이 같은 실력에 잡음이 더해진 값이면, 극단적인 성적 다음에는 **아무 개입이 없어도** 평균 쪽 성적이 나온다(연습 3.15). 회귀는 원인이 필요 없는 현상이다.

연습문제

연습 3.14 (기울기의 순서 ★) $0 < \rho < 1$이면 x–y 평면에서 $E[Y \mid X]$의 기울기 $\rho\sigma_Y/\sigma_X$는 SD 선의 기울기 σ_Y/σ_X보다 작고, $E[X \mid Y]$를 그린 직선의 기울기 $\sigma_Y/(\rho\sigma_X)$는 더 큼을 보여라. Galton의 자료(표 2.1)에서 세 기울기는? 풀이 56쪽

연습 3.15 (시험–재시험 모형 ★★) 실력 $T \sim \mathcal{N}(\mu, \tau^2)$, 두 번의 성적 $X = T + E_1$, $Y = T + E_2$, $E_1, E_2 \overset{\text{iid}}{\sim} \mathcal{N}(0, s^2)$이 T와 독립이다. (X, Y)가 이변량 정규임을 보이고 $E[Y \mid X = x]$를 구하라. 신뢰도 $\tau^2/(\tau^2 + s^2)$의 역할을 설명하고, 이것으로 비행교관 일화를 설명하라. 풀이 56쪽

연습 3.16 (Cramér–Wold 판정의 증명 ★★ 유도형) 정리 3.10를 다음 순서로 증명하라.

(a) ($\Rightarrow$): 정리 3.8에 $C = \mathbf{a}'$를 넣어라.

(b) ($\Leftarrow$): 각 $X_i = \mathbf{e}_i'\mathbf{X}$가 정규이므로 $\boldsymbol{\mu} = E\mathbf{X}$와 $\Sigma = \mathrm{Cov}\,\mathbf{X}$가 존재함을 보여라. 공분산의

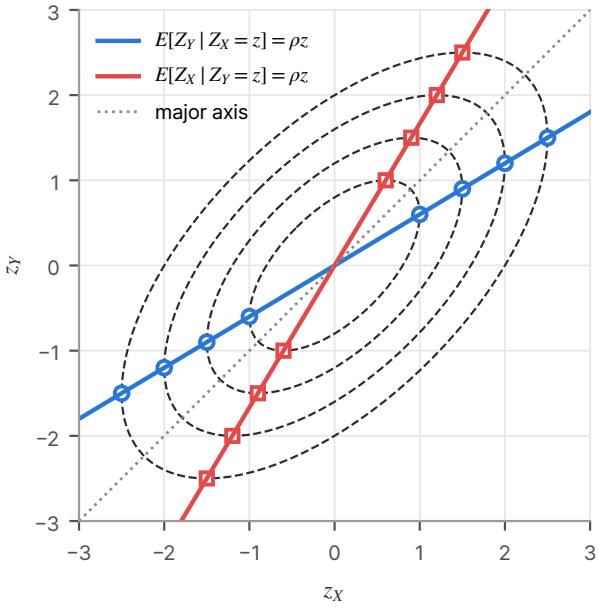


그림 3.4 $\rho = 0.6$. 회귀선 $E[Z_Y \mid Z_X]$(파랑)은 수직 접점(○)을, $E[Z_X \mid Z_Y]$(빨강)은 수평 접점(□)을 지난다.

존재에는 무엇이 필요한가?

(c) 보조정리 3.1와 가정으로 $\mathbf{t}'\mathbf{X} \sim \mathcal{N}(\mathbf{t}'\boldsymbol{\mu},\ \mathbf{t}'\Sigma\mathbf{t})$임을 보여라.

(d) $M_{\mathbf{X}}(\mathbf{t}) = E\,e^{1\cdot(\mathbf{t}'\mathbf{X})} = M_{\mathbf{t}'\mathbf{X}}(1)$을 계산해 정의 3.4의 꼴임을 보여라. "모든 $\mathbf{a}$" 라는 가정을 어디서 썼는가?

풀이 56쪽

연습 3.17 (확률 타원의 증명 ★ 유도형) 따름정리 3.18를 다음 순서로 증명하라.

(a) 정리 3.17로 $P(\mathbf{X} \in E_c) = F_{\chi^2(n)}(c^2)$를 보여라.

(b) $\Sigma = V\Lambda V'$ 에서 $\mathbf{y} = V'(\mathbf{x}-\boldsymbol{\mu})$로 두면 $(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \sum_i y_i^2/\lambda_i$ 임을 보이고, E_c가 $\mathbf{v}_i$ 방향의 반지름이 $c\sqrt{\lambda_i}$인 타원체임을 설명하라.

(c) $n=2$이면 $\chi^2(2)$가 평균 2인 지수분포임을 밀도로 확인하고 $F(c^2) = 1 - e^{-c^2/2}$를 얻어라.

(d) 반지름이 a, b인 타원의 넓이 πab로 넓이 $\pi c^2\sqrt{|\Sigma|}$를 얻어라.

풀이 56쪽

4

계산 — 정리를 알고리즘으로

노트북 4절의 `MVN` 클래스는 아래 세 알고리즘을 그대로 옮긴 것이다. 원칙은 하나다. **역행렬을 만들지 않는다.** Σ^{-1}을 직접 계산하면 비용이 들고, Σ가 특이에 가까울 때 수치 오차가 커진다. Cholesky 인수 L로 삼각 연립방정식을 푸는 편이 빠르고 안정적이다.

알고리즘 4.1 (표본 추출) **입력** $\boldsymbol{\mu}$, $\Sigma \succ 0$, 표본 수 N. **출력** $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)} \overset{\text{iid}}{\sim} \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$.

1. Cholesky 분해 $\Sigma = LL'$ (한 번, 비용 $O(n^3)$).
2. 각 k마다 $\mathbf{z}^{(k)} \sim \mathcal{N}_n(\mathbf{0}, I)$를 뽑고 $\mathbf{x}^{(k)} = \boldsymbol{\mu} + L\mathbf{z}^{(k)}$ (각 $O(n^2)$).

정당성: 정리 3.5에서 $A = L$.

알고리즘 4.2 (로그밀도) **입력** $\mathbf{x}$, $\boldsymbol{\mu}$, L. **출력** $\log f(\mathbf{x})$.

1. 전진대입으로 $L\mathbf{u} = \mathbf{x} - \boldsymbol{\mu}$를 푼다 ($O(n^2)$).
2. $Q = \mathbf{u}'\mathbf{u}$. (이유: $\Sigma^{-1} = L^{-\prime}L^{-1}$이므로 $Q = \|L^{-1}(\mathbf{x}-\boldsymbol{\mu})\|^2$.)
3. $\log|\Sigma| = 2\sum_i \log L_{ii}$. (이유: $|\Sigma| = |L|^2 = \prod L_{ii}^2$.)
4. $\log f(\mathbf{x}) = -\frac{1}{2}\big(n\log 2\pi + \log|\Sigma| + Q\big)$.

알고리즘 4.3 (조건부분포) **입력** 블록 $\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$, $\boldsymbol{\mu}$, 관측 $\mathbf{x}_2$. **출력** $\boldsymbol{\mu}_{1\cdot2}$, $\Sigma_{11\cdot2}$.

1. 연립방정식 $\Sigma_{22}\,G = \Sigma_{21}$을 풀어 $B = G' = \Sigma_{12}\Sigma_{22}^{-1}$을 얻는다.
2. $\boldsymbol{\mu}_{1\cdot2} = \boldsymbol{\mu}_1 + B(\mathbf{x}_2 - \boldsymbol{\mu}_2)$, $\Sigma_{11\cdot2} = \Sigma_{11} - B\Sigma_{21}$.
3. 반올림으로 생긴 비대칭을 없애려고 $\Sigma_{11\cdot2} \leftarrow \frac{1}{2}(\Sigma_{11\cdot2} + \Sigma_{11\cdot2}')$.

반정치인 경우 Cholesky는 $\Sigma \succ 0$에서만 작동한다. Σ가 특이하거나 수치적으로 거의 특이하면 (6.1절의 GP 공분산이 그렇다) 고유분해 $\Sigma = V\Lambda V'$로 $A = V\Lambda^{1/2}$를 만들어 알고리즘 4.1을 그대로 쓴다. 이때 밀도는 없으므로 알고리즘 4.2는 쓸 수 없다(명제 5.6).

연습문제

연습 4.1 (예제 3.16을 코드로 ★) 노트북의 `MVN.conditional`로 예제 3.16의 $\boldsymbol{\mu}_{1\cdot 2}$와 $\Sigma_{11\cdot 2}$를 $(x_2, x_3) = (1, 0)$에서 계산하고, 10^6개 표본에서 $|X_2 - 1| < 0.05$, $|X_3| < 0.05$인 점들의 X_1 평균·분산과 비교하라. 풀이 57쪽

연습 4.2 (KL 발산 구현 ★★) `MVN`에 `kl(self, other)`를 추가하라. 식은 명제 6.2이다. 역행렬 없이 $\operatorname{tr}(\Sigma_1^{-1}\Sigma_0)$와 이차형식을 계산하는 방법을 설명하고, Monte Carlo 추정 $\frac{1}{N}\sum_k [\log f_0(\mathbf{x}^{(k)}) - \log f_1(\mathbf{x}^{(k)})]$과 대조하라. 풀이 57쪽

연습 4.3 (아들과 딸 ★★) `GaltonFamilies`에서 아들과 딸을 나눠 `midparentHeight`에 대한 회귀를 각각 구하라. 기울기와 잔차 표준편차를 비교해, Galton의 $\times 1.08$ 손질이 평균뿐 아니라 회귀 구조까지 맞추는지 논하라. 풀이 57쪽

5

부숴보기 — 가정을 하나씩 빼면

Legend13의 ablation study처럼, 정리의 가정을 하나씩 빼서 무엇이 무너지는지 본다.

5.1 '결합'을 빼면: 주변은 정규, 결합은 아니다

예제 5.1 (부호를 섞은 쌍) $X \sim \mathcal{N}(0,1)$, S는 X와 독립이고 $P(S=1)=P(S=-1)=\frac{1}{2}$, $Y=SX$라 하자.

(i) $Y \sim \mathcal{N}(0,1)$: $P(Y \le y) = \frac{1}{2}P(X \le y) + \frac{1}{2}P(-X \le y) = \frac{1}{2}\Phi(y) + \frac{1}{2}\Phi(y) = \Phi(y)$ ($-X \sim \mathcal{N}(0,1)$).

(ii) 무상관: $\mathrm{Cov}(X,Y) = E[SX^2] = E[S]\,E[X^2] = 0$.

(iii) 종속: $|Y| = |X|$이므로 $P(|X|>2,\ |Y|\le 1) = 0 \neq P(|X|>2)\,P(|Y|\le 1)$.

(iv) 결합정규가 아니다: $X+Y = (1+S)X$는 확률 $\frac{1}{2}$로 정확히 0이므로 정규가 아니다. 결합정규라면 정리 3.10($\mathbf{a}=(1,1)$)에 의해 $X+Y$가 정규여야 한다.

예제 5.2 (스위치 쌍) $c>0$에 대해 $Y = X\ (|X|\le c)$, $Y=-X\ (|X|>c)$로 두자. 대칭성으로 $Y\sim\mathcal{N}(0,1)$이다. $x\varphi(x) = -\varphi'(x)$와 부분적분으로

$$E\big[X^2;\,|X|\le c\big] = \int_{-c}^{c} x\cdot x\varphi(x)\,dx = \Big[-x\varphi(x)\Big]_{-c}^{c} + \int_{-c}^{c}\varphi(x)\,dx = 2\Phi(c) - 1 - 2c\,\varphi(c).$$

따라서 $\mathrm{Cov}(X,Y) = E[X^2;|X|\le c] - E[X^2;|X|>c] = 2E[X^2;|X|\le c] - 1$이고, 이 값이 0이 되는 $c^* = 1.5382$에서 X와 Y는 무상관이다. 그런데 Y는 X의 **함수**다. 결합정규였다면 정리 3.11로 독립이어야 하지만, 예제 5.1(iii)과 같은 논리로 독립이 아니다.

가정 주의 실제 데이터에서 "변수마다 히스토그램이 종 모양이니 다변량 정규"라고 말하는 것은 **결합**정규 가정을 건너뛰는 일이다. 정리 B·C·D는 모두 이 가정 위에 서 있다.

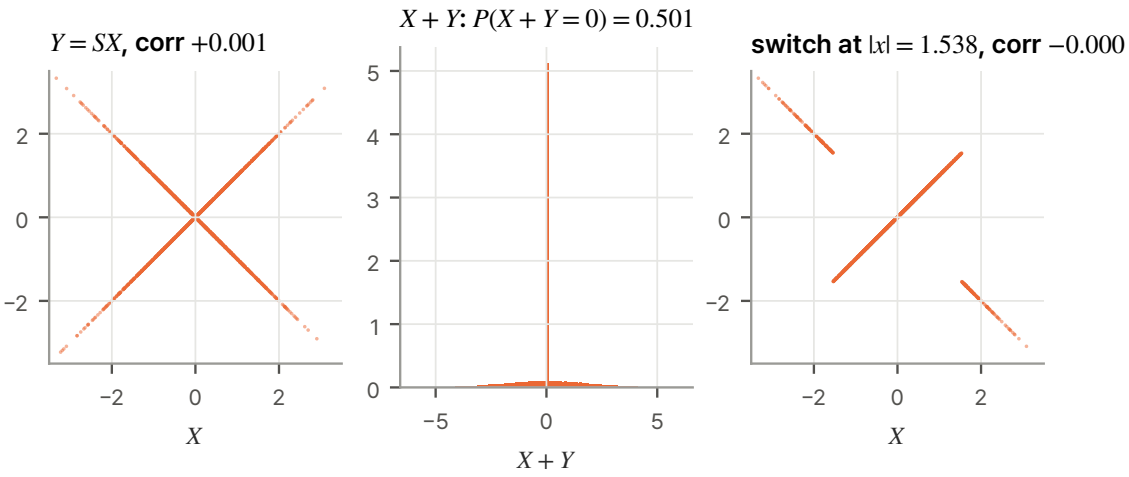


그림 5.1 두 반례. 주변분포는 완벽한 표준정규, 상관은 0인데 점들은 선 위에 있다. 가운데: $X+Y$의 절반이 정확히 0이다.

연습문제

연습 5.1 (제곱의 공분산 ★) 예제 5.1에서 $\operatorname{Cov}(X^2, Y^2) = 2$임을 보여라. 상관이 0이라도 종속을 드러내는 함수가 있다. 풀이 57쪽

5.2 정규를 약하게: 오염된 정규

명제 5.3 (오염된 정규의 적률) $0 < \varepsilon < 1$, $\sigma_c > 1$에 대해 밀도 $f(x) = (1-\varepsilon)\varphi(x) + \frac{\varepsilon}{\sigma_c}\varphi(x/\sigma_c)$를 갖는 X는

$$EX = 0, \qquad \operatorname{Var} X = 1 + \varepsilon(\sigma_c^2 - 1), \qquad \frac{EX^4}{(\operatorname{Var} X)^2} = \frac{3\left[(1-\varepsilon) + \varepsilon\sigma_c^4\right]}{\left[(1-\varepsilon) + \varepsilon\sigma_c^2\right]^2}.$$

증명. 숨은 지시변수 $I \sim \text{Bernoulli}(\varepsilon)$를 두고 $X \mid I=0 \sim \mathcal{N}(0,1)$, $X \mid I=1 \sim \mathcal{N}(0,\sigma_c^2)$로 쓰면 f가 X의 밀도다. 반복기댓값으로 $EX^k = (1-\varepsilon)EZ^k + \varepsilon\,\sigma_c^k EZ^k$이고, 정리 1.8(ii)의 $EZ^2 = 1$, $EZ^4 = 3$을 넣는다. ∎

$\varepsilon = 0.1$, $\sigma_c = 3$이면 분산 1.8, 첨도 $3 \cdot 9/3.24 = 8.33$이다(정규는 3). 밀도 그림으로는 거의 구별되지 않지만 꼬리는 전혀 다르다(그림 5.2). 같은 분산의 이변량 정규라고 착각하고 99% 타원을 그리면 실제로 밖에 떨어지는 비율은 1%가 아니라 4.0%다(연습 5.2). 15회에서 Tukey(1960)가 보인 대로, 이 정도 오염만으로도 표준편차보다 평균절대편차가 나은 추정량이 된다.

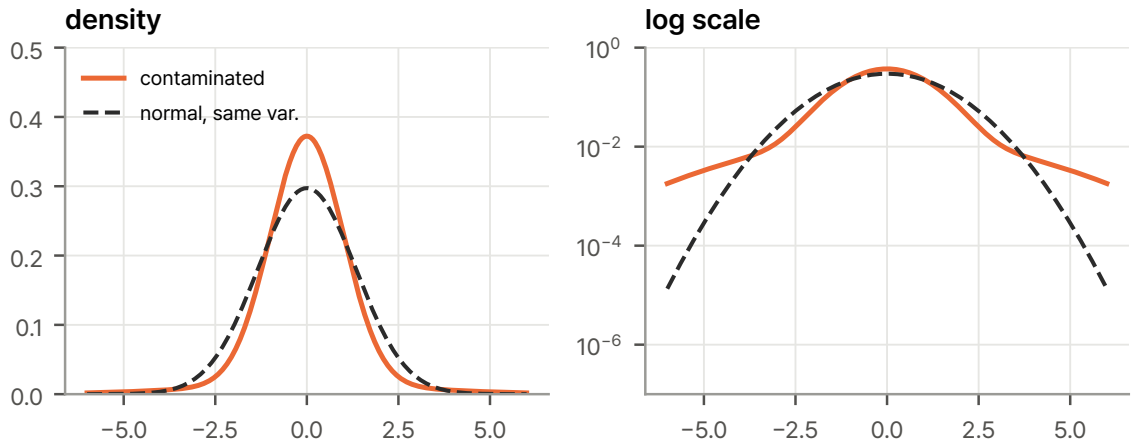


그림 5.2 오염된 정규($\varepsilon = 0.1, \sigma_c = 3$)와 같은 분산의 정규. 로그 척도에서 꼬리의 차이가 드러난다.

연습문제

연습 5.2 (오염된 이변량 정규의 꼬리 ★★) $\mathbf{X} \mid I = 0 \sim \mathcal{N}_2(\mathbf{0}, I)$, $\mathbf{X} \mid I = 1 \sim \mathcal{N}_2(\mathbf{0}, 9I)$, $P(I = 1) = 0.1$이다. 분석가가 이를 $\mathcal{N}_2(\mathbf{0}, 1.8I)$로 착각해 $Q = \|\mathbf{X}\|^2/1.8$을 계산한다. $P(Q > \chi^2_{0.99}(2))$를 정확히 구하라($\chi^2_{0.99}(2) = 9.210$). (힌트: $\|\mathbf{X}\|^2/\sigma^2 \sim \chi^2(2)$의 꼬리는 $e^{-x/2}$.) 풀이 57쪽

5.3 한 집단이 아니면: 혼합과 Galton의 ×1.08

Galton이 딸의 키에 1.08을 곱하지 않았다면 자녀 키는 **두 정규의 혼합**(4회 5절)이다.

명제 5.4 (두 정규 혼합의 평균과 분산) 확률 π로 $\mathcal{N}(\mu_1, \sigma_1^2)$, 확률 $1-\pi$로 $\mathcal{N}(\mu_2, \sigma_2^2)$에서 뽑은 X는

$$EX = \pi\mu_1 + (1-\pi)\mu_2, \qquad \operatorname{Var} X = \underbrace{\pi\sigma_1^2 + (1-\pi)\sigma_2^2}_{E[\operatorname{Var}(X\mid I)]} + \underbrace{\pi(1-\pi)(\mu_1-\mu_2)^2}_{\operatorname{Var}(E[X\mid I])}.$$

증명. 집단 지시변수 I에 대한 전분산 공식이다. $E[X \mid I]$는 확률 $\pi, 1-\pi$로 μ_1, μ_2를 갖는 이점분포이므로 분산이 $\pi(1-\pi)(\mu_1-\mu_2)^2$. ∎

예제 5.5 (GaltonFamilies의 아들과 딸) 아들 481명(평균 69.23, 표준편차 2.62), 딸 453명(평균 64.10, 표준편차 2.35)이다. 명제 5.4로 합친 분산은 약 12.8(표준편차 3.58)이고, 그 가운데 절반 가까이($\pi(1-\pi)(\mu_1-\mu_2)^2 \approx 6.6$)가 **성별의 평균 차이**에서 온다. 딸에 1.08을 곱하면 평균이 69.23으로 같아지고 이 항이 사라진다. 평균의 비는 정확히 $69.23/64.10 = 1.080$이고, 표준편차는 $2.35 \times 1.08 = 2.54$로 아들의 2.62와 비슷해진다. 합친 분포의 초과첨도(첨도−3)는 −0.50(정규보다 납작)에서 +0.25로 0에 가까워진다.

연습문제

연습 5.3 (혼합분산의 수치 ★) 예제 5.5의 수치로 명제 5.4의 세 항을 계산하라. 딸에 1.08을 곱한 뒤에는 어떻게 되는가? 풀이 58쪽

5.4 Σ가 특이하면: 밀도는 없어도 정규다

명제 5.6 (특이 정규분포의 받침) $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$이고 $\operatorname{rank}\Sigma = r < n$이면 $P\big(\mathbf{X} - \boldsymbol{\mu} \in \operatorname{col}(\Sigma)\big) = 1$이다. $\operatorname{col}(\Sigma)$는 r차원 부분공간이므로 $\mathbf{X}$는 $\mathbb{R}^n$에서 밀도를 갖지 않는다.

증명. 정리 3.5의 구성 $\mathbf{X} - \boldsymbol{\mu} = A\mathbf{Z}$에서 $A\mathbf{Z} \in \operatorname{col}(A)$이고, $\operatorname{col}(A) = \operatorname{col}(AA') = \operatorname{col}(\Sigma)$이다(둘의 영공간이 같다: $AA'\mathbf{v} = 0 \Rightarrow \|A'\mathbf{v}\|^2 = 0$). 차원이 $r < n$인 부분공간은 르베그 측도가 0인데 확률은 1이므로, 밀도가 있다면 모순이다. ∎

연습문제

연습 5.4 (직선 위의 정규 ★) $\mathbf{X} \sim \mathcal{N}_2(\mathbf{0}, \Sigma)$일 때 $\operatorname{Var}(\mathbf{a}'\mathbf{X}) = 0$인 $\mathbf{a}$를 찾고, $\mathbf{X}$가 놓이는 직선을 구하라. $\mathbf{X}$의 mgf는 무엇인가?

$$\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

풀이 58쪽

6

Beautiful insights — 정리 C는 어디에나 있다

6.1 Gaussian Process 회귀 = 정리 C 한 번

격자 위의 함수값 $\mathbf{f}_g = (f(s_1), \dots, f(s_m))'$ 와 관측 $\mathbf{y} = f(\mathbf{s}_o) + \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 I)$를 함께 결합정규로 놓는다. 공분산은 커널 $k(s,s')$로 준다. 예를 들어 $k(s,s') = a^2 \exp\{-(s-s')^2/(2\ell^2)\}$.

$$\begin{pmatrix} \mathbf{f}_g \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} K_{gg} & K_{go} \\ K_{og} & K_{oo} + \sigma_n^2 I \end{pmatrix} \right).$$

정리 3.13을 한 번 쓰면 사후분포가 나온다.

$$\mathbf{f}_g \mid \mathbf{y} \sim \mathcal{N}\Big(K_{go}(K_{oo}+\sigma_n^2 I)^{-1}\mathbf{y},\;\; K_{gg} - K_{go}(K_{oo}+\sigma_n^2 I)^{-1}K_{og} \Big).$$

$\Sigma_{12}\Sigma_{22}^{-1}$ 와 Schur 보수가 그대로 들어 있다. 관측점에서 멀어질수록 K_{go}가 작아져 사후평균은 사전평균 0으로 **회귀**하고 사후분산은 사전분산으로 돌아간다. 노트북 6.1절은 `MVN.conditional`을 고치지 않고 이것을 그린다.

6.2 Pearson (1901): 회귀선 둘, 주성분 하나

Pearson은 1901년 *On lines and planes of closest fit to systems of points in space*에서 물었다. 점들에 "가장 가까운" 직선은 무엇인가? 세로 거리를 최소화하면 $E[Y \mid X]$, 가로 거리를 최소화하면 $E[X \mid Y]$다. **수직 거리**를 최소화하면 다음 명제의 답이 나온다.

> **명제 6.1** (제1 주성분은 수직거리를 최소화한다) $E\mathbf{X} = \boldsymbol{\mu}$, $\operatorname{Cov}\mathbf{X} = \Sigma$인 $\mathbf{X} \in \mathbb{R}^n$과 단위벡터 $\mathbf{v}$가 정하는 직선 $\ell_{\mathbf{v}} = \{\boldsymbol{\mu} + t\mathbf{v}\}$에 대해
>
> $$E\big[\operatorname{dist}(\mathbf{X}, \ell_{\mathbf{v}})^2\big] = \operatorname{tr}\Sigma - \mathbf{v}'\Sigma\mathbf{v}.$$
>
> 따라서 이 값을 최소로 하는 $\mathbf{v}$는 Σ의 가장 큰 고유값의 고유벡터이고, 최솟값은 나머지 고유값의 합이다.

증명. $\mathbf{D} = \mathbf{X} - \boldsymbol{\mu}$로 두면 직선까지의 거리는 $\mathbf{D}$에서 $\mathbf{v}$ 방향 성분을 뺀 길이이므로 $\operatorname{dist}^2 = \|\mathbf{D}\|^2 - (\mathbf{v}'\mathbf{D})^2$. 기댓값을 취하면 $E\|\mathbf{D}\|^2 = \operatorname{tr}\Sigma$, $E(\mathbf{v}'\mathbf{D})^2 = \mathbf{v}'\Sigma\mathbf{v}$. Rayleigh 몫 $\mathbf{v}'\Sigma\mathbf{v}$의 단위벡터 위 최댓값은 $\lambda_{\max}$이고 그 고유벡터에서 얻는다. ∎

주성분 축은 정리 D의 타원의 장축이다(따름정리 3.18). Galton의 자료에서 세 직선의 기울기는 0.65($E[\text{자녀}|\text{부모}]$), 2.02(제1 주성분), 3.07($E[\text{부모}|\text{자녀}]$)이다. 주의할 점이 있다. **단위를 바꾸면 주성분은 바뀌지만 회귀선은 바뀌지 않는다.**

6.3 백색화와 BatchNorm

정리 3.8로 $\Sigma^{-1/2}(\mathbf{X}-\boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, I)$, 곧 타원을 원으로 만들 수 있다(**백색화**). BatchNorm은 좌표마다 평균과 표준편차만 나눈다. $D = \operatorname{diag}(\Sigma)$로 쓰면 $D^{-1/2}(\mathbf{X}-\boldsymbol{\mu})$이고, 그 공분산은 상관행렬이다. 기울어진 타원은 그대로 남는다. 상관까지 없애는 Decorrelated BatchNorm(Huang et al., 2018)이 바로 $\Sigma^{-1/2}$ 백색화다.

6.4 Reparameterization trick과 두 정규 사이의 KL

VAE의 인코더가 $\boldsymbol{\mu}_\phi(\mathbf{x})$와 $L_\phi(\mathbf{x})$를 내면 잠재변수는 정리 3.5의 구성 그대로 $\mathbf{z} = \boldsymbol{\mu}_\phi + L_\phi \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, I)$로 뽑는다. 무작위성이 $\boldsymbol{\epsilon}$에만 있으므로 $\partial \mathbf{z}/\partial \boldsymbol{\mu}_\phi = I$이고 역전파가 흐른다. VAE 손실의 둘째 항은 두 정규분포 사이의 KL이다.

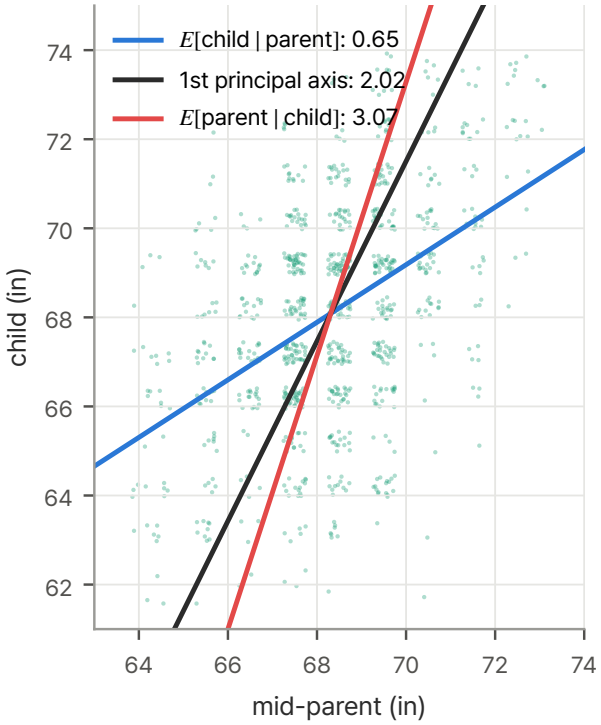


그림 6.1 Galton 자료를 지나는 세 "최선의" 직선.

명제 6.2 (두 다변량 정규 사이의 KL 발산) $\mathcal{N}_0 = \mathcal{N}_n(\boldsymbol{\mu}_0, \Sigma_0)$, $\mathcal{N}_1 = \mathcal{N}_n(\boldsymbol{\mu}_1, \Sigma_1)$, 둘 다 양정치이면

$$\mathrm{KL}(\mathcal{N}_0 \,\|\, \mathcal{N}_1) = \frac{1}{2}\Big[\operatorname{tr}(\Sigma_1^{-1}\Sigma_0) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'\Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - n + \log\frac{|\Sigma_1|}{|\Sigma_0|}\Big].$$

증명. $\mathrm{KL} = E_0[\log f_0(\mathbf{X}) - \log f_1(\mathbf{X})]$이고 정리 3.6로

$$\log f_0 - \log f_1 = \frac{1}{2}\log\frac{|\Sigma_1|}{|\Sigma_0|} + \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_1)'\Sigma_1^{-1}(\mathbf{X}-\boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_0)'\Sigma_0^{-1}(\mathbf{X}-\boldsymbol{\mu}_0).$$

$E[\mathbf{Y}'A\mathbf{Y}] = \operatorname{tr}(A\operatorname{Cov}\mathbf{Y}) + (E\mathbf{Y})'A(E\mathbf{Y})$(연습 6.1)를 $\mathbf{X} \sim \mathcal{N}_0$에 쓰면 셋째 항의 기댓값은 $\operatorname{tr}(\Sigma_0^{-1}\Sigma_0) = n$이고, 둘째 항의 기댓값은 $\operatorname{tr}(\Sigma_1^{-1}\Sigma_0) + (\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)$이다. ∎

연습문제

연습 6.1 (이차형식의 기댓값 ★) $E\mathbf{Y} = \mathbf{m}$, $\operatorname{Cov}\mathbf{Y} = V$이면 $E[\mathbf{Y}'A\mathbf{Y}] = \operatorname{tr}(AV) + \mathbf{m}'A\mathbf{m}$임을 보여라. (힌트: $\mathbf{Y}'A\mathbf{Y} = \operatorname{tr}(A\mathbf{Y}\mathbf{Y}')$.) 풀이 58쪽

연습 6.2 (VAE의 KL 항 ★) 명제 6.2에서 $\mathcal{N}_0 = \mathcal{N}(\boldsymbol{\mu}, \operatorname{diag}(\sigma_1^2, \dots, \sigma_n^2))$, $\mathcal{N}_1 = \mathcal{N}(\mathbf{0}, I)$이면 $\mathrm{KL} = \frac{1}{2}\sum_i(\sigma_i^2 + \mu_i^2 - 1 - \log\sigma_i^2)$임을 보여라. 풀이 58쪽

연습 6.3 (주성분의 기울기 ★★) 2×2 공분산 Σ의 가장 큰 고유값을 λ_1이라 하면 제1 주성분 방향의 기울기가 $(\lambda_1 - \Sigma_{11})/\Sigma_{12}$임을 보이고, 표 2.1의 $\hat{\Sigma}$로 2.02를 확인하라. y의 단위를 센티미터로 바꾸면($\times 2.54$) 주성분의 기울기와 두 회귀선의 기울기는 각각 어떻게 되는가? 풀이 58쪽

7

정리와 역사 노트

7.1 한 장 요약

	식	증명의 핵심
de Moivre	$P(S_n = k) \approx \frac{1}{\sqrt{2\pi npq}}\, e^{-(k-np)^2/(2npq)}$	Stirling + $\log(1+u)$ 전개
Gauss	평균이 언제나 MLE $\Rightarrow$ 오차는 정규	Cauchy 함수방정식
정규 혼합	$X \sim \mathcal{N}(0, \sigma_1^2),\ Y \mid X \sim \mathcal{N}(X, \sigma_2^2) \Rightarrow Y \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$	반복기댓값
안정 $\Rightarrow$ 회귀	$\mathrm{Var}\, C = \mathrm{Var}\, P \Rightarrow r = \sqrt{1 - \sigma_E^2/\sigma^2} < 1$	분산의 대수
정의	$M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}$	구성 $\boldsymbol{\mu} + A\mathbf{Z}$
정리 A	$C\mathbf{X} + \mathbf{b} \sim \mathcal{N}(C\boldsymbol{\mu} + \mathbf{b},\ C\Sigma C')$	mgf에 $C'\mathbf{t}$
Cramér–Wold	결합정규 $\iff$ 모든 $\mathbf{a}'\mathbf{X}$가 정규	$M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{t}'\mathbf{X}}(1)$
정리 B	결합정규에서 $\mathbf{X}_1 \perp \mathbf{X}_2 \iff \Sigma_{12} = 0$	mgf 분해
정리 C	$\mathbf{X}_1 \mid \mathbf{x}_2 \sim \mathcal{N}\big(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\big)$	$\mathbf{W} = \mathbf{X}_1 - B\mathbf{X}_2 \perp \mathbf{X}_2$ 또는 블록 분해
정밀도	$\Sigma_{11\cdot 2} = \Lambda_{11}^{-1},\ X_i \perp X_j \mid$ 나머지 $\iff \Lambda_{ij} = 0$	블록 역행렬
정리 D	$(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(n)$	백색화 + $Z^2 \sim \chi^2(1)$
회귀	$E[Z_Y \mid Z_X = z] = \rho z,\ E[Z_X \mid Z_Y = z] = \rho z$	정리 C, $n = 2$
접점	회귀선 = 수직 접점의 자취	받침선 또는 $\partial Q/\partial w = 0$

7.2 Stigler의 명명 법칙

> **Stigler's law of eponymy** "어떤 과학 법칙도 처음 발견한 사람의 이름을 달고 있지 않다." Stigler(1980)는 이 법칙조차 사회학자 Merton이 먼저 말했다고 덧붙였다.
>
> - **가우스(정규) 분포**: de Moivre(1733)가 이항의 근사로 먼저 얻었다.
> - **Pearson 상관계수 r**: 개념과 이름은 Galton(1888), 이변량 정규의 그 모수는 Bravais(1846). Pear-

son(1896)도 Bravais를 인용했다.

- **이변량 정규분포**: Galton의 질문에 Dickson이 답하기 전에 오차론에서 Laplace, Bravais 등이 이미 다뤘다.

7.3 7 Pillars 지도에 칠할 칸

Stigler(2016)의 일곱 기둥 가운데 이번 회차는 **Regression**이다. 그 장의 요지는 Galton의 통찰이 **회귀는 두 방향**이라는 것, 그리고 조건부분포로 생각하는 습관이 다변량 분석과 베이즈 추론의 출발점이 되었다는 것이다. 2.2절의 축소(따름정리 2.2)가 그 씨앗이고, 16회에서 Stein의 축소추정량이 바로 이 회귀임을 본다(Stigler 1990).

7.4 다음 회 — Student의 t와 신뢰구간 (1876 → 1937)

Guinness 양조장의 Gosset은 표본 4개로, σ도 모른 채 평균의 오차를 말해야 했다. 그 열쇠인 "정규표본에서 $\bar{X}$와 S^2는 독립"은 오늘의 정리 B로 증명된다. $(\bar{X}, X_1-\bar{X}, \dots, X_n-\bar{X})$는 정리 A에 의해 결합정규이고, $\mathrm{Cov}(\bar{X}, X_i-\bar{X}) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$이기 때문이다.

참고문헌

- Hogg, R. V., McKean, J. W., Craig, A. T. *Introduction to Mathematical Statistics*, 6th ed. Pearson, 2005. (다룰 내용의 범위를 정할 때 참고)
- Stigler, S. M. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard, 1986. Ch. 4, 5, 8.
- Stigler, S. M. *The Seven Pillars of Statistical Wisdom*. Harvard, 2016. Ch. 5.
- Stigler, S. M. A Galtonian perspective on shrinkage estimators. *Statistical Science* 5 (1990).
- de Moivre, A. *The Doctrine of Chances*, 2nd ed. 1738.
- Gauss, C. F. *Theoria Motus Corporum Coelestium*. 1809.
- Quetelet, A. *Lettres sur la théorie des probabilités*. 1846.
- Galton, F. Regression towards mediocrity in hereditary stature. *J. Anthropological Institute* 15 (1886).
- Galton, F. Co-relations and their measurement. *Proc. Royal Society* 45 (1888).
- Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2 (1901).

A

연습문제 풀이

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

0절

연습 0.1 문제 3쪽

야코비안은

$$\frac{\partial(x,y)}{\partial(r,\varphi)} = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r(\cos^2\varphi + \sin^2\varphi) = r.$$

(X,Y)의 밀도 $\frac{1}{2\pi}e^{-(x^2+y^2)/2}$에 변수변환 공식을 쓰면

$$f_{R,\Phi}(r,\varphi) = \frac{1}{2\pi}\, e^{-r^2/2}\, r = \underbrace{r\, e^{-r^2/2}}_{r>0} \cdot \underbrace{\frac{1}{2\pi}}_{0\le\varphi<2\pi}.$$

결합밀도가 r만의 함수와 φ만의 함수의 곱이므로 $R \perp \Phi$이다. R은 Rayleigh 분포, Φ는 균등분포다. 이것을 거꾸로 읽으면 Box–Muller 생성기다(3회).

연습 0.2 문제 3쪽

$\mathbf{m}(\mathbf{X}) = E[\mathbf{Y}\mid\mathbf{X}]$로 쓰고 $\mathbf{Y} - E\mathbf{Y} = \mathbf{a} + \mathbf{b}$, $\mathbf{a} = \mathbf{Y} - \mathbf{m}(\mathbf{X})$, $\mathbf{b} = \mathbf{m}(\mathbf{X}) - E\mathbf{Y}$로 나눈다.

$$\operatorname{Cov}\mathbf{Y} = E\,\mathbf{a}\mathbf{a}' + E\,\mathbf{b}\mathbf{b}' + E\,\mathbf{a}\mathbf{b}' + E\,\mathbf{b}\mathbf{a}'.$$

교차항은 성분마다 3회 정리 2.5의 둘째 식(g(X)를 곱해도 반복기댓값이 성립한다)으로 $E\big[\mathbf{b}\,E(\mathbf{a}\mid\mathbf{X})'\big] = \mathbf{0}$이다. 첫 항은 $E\big[E(\mathbf{a}\mathbf{a}'\mid\mathbf{X})\big] = E[\operatorname{Cov}(\mathbf{Y}\mid\mathbf{X})]$이고, $E\,\mathbf{m}(\mathbf{X}) = E\mathbf{Y}$이므로 둘째 항은 $\operatorname{Cov}(\mathbf{m}(\mathbf{X}))$이다. 스칼라(Y가 한 성분)이면 3회 정리 2.6이다.

1절

연습 1.1 문제 7쪽

$u=\delta/(np)$, $v=\delta/(nq)$로 두고 3차 항만 모은다.

$$-(np+\delta)\Big(u-\frac{u^2}{2}+\frac{u^3}{3}\Big):\quad -np\cdot\frac{u^3}{3}+\delta\cdot\frac{u^2}{2}=-\frac{\delta^3}{3n^2p^2}+\frac{\delta^3}{2n^2p^2}=\frac{\delta^3}{6n^2p^2},$$
$$(nq-\delta)\Big(v+\frac{v^2}{2}+\frac{v^3}{3}\Big):\quad nq\cdot\frac{v^3}{3}-\delta\cdot\frac{v^2}{2}=\frac{\delta^3}{3n^2q^2}-\frac{\delta^3}{2n^2q^2}=-\frac{\delta^3}{6n^2q^2}.$$

둘째 줄은 $-(n-k)\log(1-v)=(nq-\delta)(v+\frac{v^2}{2}+\frac{v^3}{3}+\cdots)$에서 왔다. 합은

$$\frac{\delta^3}{6n^2}\Big(\frac{1}{p^2}-\frac{1}{q^2}\Big)=\frac{\delta^3\,(q^2-p^2)}{6n^2p^2q^2}=\frac{(q-p)\,\delta^3}{6n^2p^2q^2}\qquad(q+p=1)$$

이고, 남는 항은 $O(\delta^4/n^3)=O(n^{-1})$이다.

연습 1.2 문제 7쪽

$npq=25$이다.

(a) $P(S=50)=\binom{100}{50}2^{-100}=0.07959$, 근사 $1/\sqrt{50\pi}=0.07979$.

(b) 정확한 값 0.72875. 보정 없이 $2\Phi(5/5)-1=0.68269$. 연속성 보정 $2\Phi(5.5/5)-1=0.72867$.

연습 1.3 문제 7쪽

(a) 증명의 단계 ①에서 세 계승의 $\sqrt{2\pi}$ 자리에 C를 쓰면 앞인수가

$$\frac{C\sqrt{n}}{C\sqrt{k}\cdot C\sqrt{n-k}}=\frac{1}{C}\sqrt{\frac{n}{k(n-k)}}=\frac{1}{C\sqrt{npq}}\big(1+O(n^{-1/2})\big)$$

가 되고, 나머지 단계는 그대로다.

(b) $A>0$을 고정하고 $\{|x_k|\le A\}$와 나머지로 나눈다. Chebyshev 부등식으로 $P(|x|>A)\le 1/A^2$이고, 가운데 부분은 따름정리 1.3의 논리로 $\frac{1}{C}\int_{-A}^{A}e^{-x^2/2}dx$로 수렴한다. 따라서

$$1-\frac{1}{A^2}\;\le\;\frac{1}{C}\int_{-A}^{A}e^{-x^2/2}\,dx\;\le\;1,$$

이고 $A\to\infty$로 $\sqrt{2\pi}/C=1$이다.

연습 1.4 문제 9쪽

$\ell(\theta)=-\sum|x_i-\theta|$는 꺾은선이고, 꺾인 점 밖에서 $\ell'(\theta)=\#\{x_i>\theta\}-\#\{x_i<\theta\}$이다. 중앙값보다 왼쪽에서는 양수, 오른쪽에서는 음수다. n이 홀수이면 중앙값 하나에서 부호가 바뀌고, 짝수이면 가운데 두 값 사이에서 $\ell'=0$이라 그 구간 전체가 최대점이다.

연습 1.5 문제 9쪽

(a) $\log f(u) = -u - 2\log(1+e^{-u})$이므로 $(\log f)'(u) = -1 + \frac{2e^{-u}}{1+e^{-u}} = \frac{e^{-u}-1}{1+e^{-u}} = -\tanh(u/2)$.

(b) $(\log f)''(u) = -\frac{1}{2}\operatorname{sech}^2(u/2) < 0$이므로 $\ell(\theta) = \log f(x_1-\theta) + \log f(x_2-\theta)$는 순오목이다. f가 짝함수이므로 ℓ은 $\theta \mapsto x_1 + x_2 - \theta$에 대해 대칭이다. 순오목이면서 대칭인 함수의 유일한 최대점은 대칭축 $\bar{x}$다.

(c) $\ell'(\theta) = \sum_i \tanh\frac{x_i-\theta}{2}$. 표본 $(0,0,3)$, $\bar{x}=1$에서 $2\tanh(-0.5)+\tanh(1) = -0.1626 \neq 0$이다. 증명에서 $n=3$이 필요한 곳은 단계 ③, 곧 $g(a+b)=g(a)+g(b)$를 얻는 곳이다. $n=2$만으로는 g가 홀함수라는 것만 나온다.

연습 1.6 문제 9쪽

로그 사후밀도는 (충분통계량 $\bar{x}$로 정리하면)

$$-\frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-m)^2}{2\tau^2} + \text{상수}.$$

θ의 이차식이므로 최빈값과 평균이 같다. θ로 미분해 0으로 두면 주어진 가중평균이 나온다. $\tau \to \infty$(평평한 사전분포)이면 $\bar{x}$로 가고, $n \to \infty$이면 역시 $\bar{x}$로 간다. 데이터가 많을수록 사전분포의 영향이 사라진다.

연습 1.7 문제 11쪽

$\sum f_j x_j = 228{,}555$이므로 $\hat{\mu} = 228555/5738 = 39.832$. $\hat{\sigma}^2 = 4.2002$, $\hat{\sigma} = 2.049$. 44인치 이상은 $92+50+21+4+1 = 168$명, 비율 0.0293. 정규 근사는

$$1 - \Phi\Big(\frac{43.5 - 39.832}{2.049}\Big) = 0.0367$$

이다. 오른쪽 꼬리는 정규가 조금 과대평가한다.

연습 1.8 문제 11쪽

독립이므로 $\operatorname{Var} Y = \operatorname{Var} X + \operatorname{Var} U$이고 $\operatorname{Var} U = \int_{-h/2}^{h/2} \frac{u^2}{h}\,du = \frac{h^2}{12}$. 보정한 표준편차는 $\sqrt{4.2002 - 1/12} = 2.029$. 이 가정은 X의 밀도가 폭 h 안에서 거의 평평할 때(표준편차가 h보다 충분히 클 때) 그럴듯하다. 여기서는 $\sigma \approx 2h$라 대략 맞지만, 꼬리 칸에서는 밀도의 기울기가 커서 반올림 오차와 X가 독립이 아니다.

연습 1.9 문제 13쪽

(a) $e^{t^2/2} = \sum_k \frac{t^{2k}}{2^k k!}$와 $\sum_m \frac{EZ^m}{m!}t^m$을 비교하면 홀수 적률은 0이고 $EZ^{2k} = \frac{(2k)!}{2^k k!} = (2k-1)!!$.

(b) $x\varphi(x) = -\varphi'(x)$이므로

$$EZ^{m+1} = \int z^m \cdot z\varphi(z)\,dz = \int z^m\big(-\varphi'(z)\big)dz = \Big[-z^m\varphi\Big]_{-\infty}^{\infty} + m\int z^{m-1}\varphi\,dz = m\,EZ^{m-1}.$$

$EZ^0 = 1$, $EZ^1 = 0$에서 귀납한다.

연습 1.10 문제 13쪽

$EY = M_X(1) = e^{\mu+\sigma^2/2}$, $EY^2 = M_X(2) = e^{2\mu+2\sigma^2}$이므로

$$\operatorname{Var} Y = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2}\bigl(e^{\sigma^2}-1\bigr).$$

e^x는 볼록하므로 Jensen에 의해 $Ee^X \geq e^{EX}$이고, 그 차이를 만드는 인수가 $e^{\sigma^2/2}$다.

연습 1.11 문제 13쪽

위쪽. $x \geq z$이면 $1 \leq x/z$이므로

$$1-\Phi(z) = \int_z^\infty \varphi(x)\,dx \leq \int_z^\infty \frac{x}{z}\varphi(x)\,dx = \frac{\varphi(z)}{z} \qquad (x\varphi = -\varphi').$$

아래쪽. 미분하면

$$\frac{d}{dx}\Bigl[-\Bigl(\frac{1}{x}-\frac{1}{x^3}\Bigr)\varphi(x)\Bigr] = \Bigl(1-\frac{3}{x^4}\Bigr)\varphi(x) \leq \varphi(x),$$

이므로 z부터 ∞까지 적분하면 $\bigl(\frac{1}{z}-\frac{1}{z^3}\bigr)\varphi(z) \leq 1-\Phi(z)$. $z=2$에서 $0.02025 \leq 0.02275 \leq 0.02700$이다.

연습 1.12 문제 13쪽

$\tau = \frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}$로 두면 피적분함수의 지수는

$$-\frac{1}{2}\Bigl[\tau x^2 - \frac{2sx}{\sigma_2^2} + \frac{s^2}{\sigma_2^2}\Bigr] = -\frac{\tau}{2}\Bigl(x - \frac{s}{\sigma_2^2\tau}\Bigr)^2 - \frac{s^2}{2(\sigma_1^2+\sigma_2^2)},$$

이다($\frac{s^2}{\sigma_2^2} - \frac{s^2}{\sigma_2^4\tau} = \frac{s^2}{\sigma_1^2+\sigma_2^2}$). x로 적분하면 $\sqrt{2\pi/\tau} = \sqrt{2\pi}\,\sigma_1\sigma_2/\sqrt{\sigma_1^2+\sigma_2^2}$이므로

$$f_{X+Y}(s) = \frac{1}{2\pi\sigma_1\sigma_2}\cdot\frac{\sqrt{2\pi}\,\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\,e^{-s^2/(2(\sigma_1^2+\sigma_2^2))} = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\,e^{-s^2/(2(\sigma_1^2+\sigma_2^2))}.$$

연습 1.13 문제 13쪽

(a) $\{a \leq (S_n-np)/\sqrt{npq} \leq b\}$는 $x_k \in [a,b]$인 정수 k들에 대한 사건 $\{S_n = k\}$의 서로소인 합집합이다. $x_{k+1}-x_k = 1/\sqrt{npq} = \Delta$이다. (b) $|x_k| \leq A = \max(|a|,|b|)$인 모든 k에서 $P(S_n=k) = \varphi(x_k)\Delta\,(1+\epsilon_{n,k})$이고, 정리 1.2의 균등성으로 $\max_k|\epsilon_{n,k}| \leq C/\sqrt{n}$이다. 따라서 합은 $\bigl(1+O(n^{-1/2})\bigr)\sum\varphi(x_k)\Delta$다. 합의 항 수가 약 $(b-a)\sqrt{npq}$로 늘어나므로, 오차가 k마다 다른 속도로 줄면 합에서 오차를 하나로 묶을 수 없다. 균등성이 그것을 막는다. (c) $x_k \in [a,b]$인 칸 $I_k = [x_k, x_k+\Delta)$에서 평균값 정리로 $|\varphi(x)-\varphi(x_k)| \leq \Delta\max|\varphi'|$이므로 $\bigl|\varphi(x_k)\Delta - \int_{I_k}\varphi\bigr| \leq \Delta^2\max|\varphi'|$이다. 칸은 많아야 $(b-a)/\Delta+1$개이므로 이 오차의 합은 $(b-a+\Delta)\Delta\max|\varphi'|$ 이하다.

칸들의 합집합은 $[a,b]$와 양 끝에서 길이 Δ 이하인 두 조각만큼 다르고, 그 위의 적분은 $2\Delta\max\varphi$ 이하다. (d) $n\to\infty$에서 $\Delta\to 0$이므로 (c)의 오차는 0으로 가고, (b)의 인수 $1+O(n^{-1/2})$는 1로 간다. 따라서 확률은 $\int_a^b \varphi = \Phi(b)-\Phi(a)$로 수렴한다.

연습 1.14 문제 14쪽

(a) $a=b=0$이면 $g(0)=2g(0)$이므로 $g(0)=0$이다. $b=-a$이면 $0=g(0)=g(a)+g(-a)$이다. (b) $m=1$이면 자명하다. $g(mu)=mg(u)$이면 $g((m+1)u)=g(mu)+g(u)=(m+1)g(u)$이다. $u=1$을 넣으면 자연수 m에서 $g(m)=cm$이고, (a)로 $g(0)=0$, $g(-m)=-g(m)=c(-m)$이다. (c) (b)를 $u=m/k$와 자연수 k에 쓰면 $k\,g(m/k)=g(k\cdot m/k)=g(m)=cm$이므로 $g(m/k)=cm/k$이다. (d) 유리수는 실수에서 조밀하므로 $r_j\to u$인 유리수열이 있다. 연속성으로 $g(u)=\lim_j g(r_j)=\lim_j cr_j=cu$이다. (a)–(c)는 덧셈 공식만 썼고 연속성은 (d)에서만 썼다. 연속성이 없으면 일차함수가 아닌 해가 있다(선택공리로 만든다).

2절

연습 2.1 문제 16쪽

$P(K_{16}=8)=\binom{16}{8}2^{-16}=12870/65536=0.1964$. 근사는 분산이 $16/4=4$이므로 $1/\sqrt{8\pi}=0.1995$.

연습 2.2 문제 16쪽

$\xi_i=B_i-\tfrac12$, $B_i\sim\mathrm{Bernoulli}(p)$이므로 $E\xi_i=p-\tfrac12$, $\operatorname{Var}\xi_i=p(1-p)$. 따라서 $EP_n=n(p-\tfrac12)$, $\operatorname{Var}P_n=np(1-p)$이고 근사는 $\mathcal{N}\big(n(p-\tfrac12),\,np(1-p)\big)$이다. 종의 중심이 옮겨 가고 폭은 조금 좁아진다.

연습 2.3 문제 18쪽

정리 2.1(b)에서 $\operatorname{Cov}(X,Y)=\sigma_1^2$, $\operatorname{Var}Y=\sigma_1^2+\sigma_2^2$. 따름정리 3.19에서 X와 Y를 바꾼 식으로 기울기는 $\operatorname{Cov}(X,Y)/\operatorname{Var}Y=B$이다.

연습 2.4 문제 18쪽

따름정리 2.2의 계산에서 $\sigma_1^2\to\tau^2$, $\sigma_2^2\to s^2/n$으로 바꾸고 평균 m을 더하면

$$E[X\mid\bar y]=\frac{(n/s^2)\,\bar y+(1/\tau^2)\,m}{n/s^2+1/\tau^2},\qquad \operatorname{Var}(X\mid\bar Y)=\Big(\frac{n}{s^2}+\frac{1}{\tau^2}\Big)^{-1}.$$

$n\to\infty$이면 $\bar y$로 가고 분산은 0이 된다. 연습 1.6에서 $\bar x$의 분산이 σ^2/n이었으므로 같은 식이다. 정밀도(분산의 역수)가 더해지고, 평균은 정밀도로 가중된다.

연습 2.5 문제 19쪽

$\sigma_E^2/\sigma^2 = 1-\beta^2/2 = 1-\frac{4/9}{2} = \frac{7}{9}$. 설명되는 비율은 $\frac{2}{9} \approx 22\%$이고, 이는 $\mathrm{Corr}(C,M)^2 = \beta^2/2$와 같다.

연습 2.6 문제 19쪽

중간부모의 분산부터 차례로 구한다.

$$\mathrm{Var}\,M = \frac{\sigma^2}{4}(2+2a) = \frac{\sigma^2(1+a)}{2}, \qquad \text{안정 조건} \quad \beta^2\frac{\sigma^2(1+a)}{2} + \sigma_E^2 = \sigma^2,$$

$$\mathrm{Corr}(C,M) = \frac{\beta\,\mathrm{Var}\,M}{\sigma\sqrt{\mathrm{Var}\,M}} = \beta\sqrt{\frac{1+a}{2}}, \qquad \frac{\mathrm{Cov}(C,F)}{\mathrm{Var}\,F} = \frac{\beta\cdot\frac{1}{2}(\sigma^2+a\sigma^2)}{\sigma^2} = \beta\,\frac{1+a}{2}.$$

연습 2.7 문제 20쪽

$68.088 + 0.6463\,(66-68.308) = 66.60$인치. 구간은 $66.60 \pm 1.96\times 2.236 = [62.21,\ 70.98]$. 이 구간을 정당화하려면 이변량 정규 가정(조건부분포가 정규이고 등분산)이 필요하다. 또 모수를 추정값으로 바꿔 넣은 오차를 무시하는 근사이기도 하다(후자는 13회에서 다룬다).

연습 2.8 문제 20쪽

(a) $X=x$가 주어지면 $Y \sim \mathcal{N}(x,\sigma_2^2)$이므로 정리 1.8(ii)로 $E(e^{tY}\mid X) = e^{tX+\sigma_2^2t^2/2}$이다. 반복기댓값으로

$$M_Y(t) = E\big[e^{tX+\sigma_2^2t^2/2}\big] = e^{\sigma_2^2t^2/2}\,M_X(t) = e^{(\sigma_1^2+\sigma_2^2)t^2/2}.$$

(b) $X=x$가 주어지면 $D = Y-x \sim \mathcal{N}(0,\sigma_2^2)$이다. 유계인 h,k에 대해 $E\big[h(D)k(X)\big] = E\big[k(X)E(h(D)\mid X)\big]$이고, $E(h(D)\mid X)$는 X에 의존하지 않는 상수 $E\,h(N)$ $(N\sim\mathcal{N}(0,\sigma_2^2))$이다. 따라서 $E[h(D)k(X)] = E\,h(N)\,E\,k(X)$이고, $k\equiv 1$을 넣으면 $D\sim\mathcal{N}(0,\sigma_2^2)$, 일반의 h,k에서 $D\perp X$이다. (c) $Y=X+D$이므로

$$E\,e^{sX+tY} = E\,e^{(s+t)X}\,E\,e^{tD} = \exp\Big\{\tfrac{1}{2}\big[(s+t)^2\sigma_1^2 + t^2\sigma_2^2\big]\Big\}$$
$$= \exp\Big\{\tfrac{1}{2}\,(s,t)\begin{pmatrix}\sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma_1^2+\sigma_2^2\end{pmatrix}\begin{pmatrix}s\\ t\end{pmatrix}\Big\},$$

이고 이는 정의 3.4의 이변량 정규 mgf다. (d) $\mathrm{Cov}(X,Y)=\sigma_1^2$, $\mathrm{Corr}(X,Y) = \sigma_1^2/(\sigma_1\sqrt{\sigma_1^2+\sigma_2^2}) = \sigma_1/\sqrt{\sigma_1^2+\sigma_2^2}$이다. 분산은 전분산 공식으로 $E[\mathrm{Var}(Y\mid X)] + \mathrm{Var}(E[Y\mid X]) = \sigma_2^2+\sigma_1^2$이다.

연습 2.9 문제 20쪽

(a) $f(x\mid y) = f_X(x)f_{Y|X}(y\mid x)/f_Y(y)$이고 분모는 x에 의존하지 않는다. x에 의존하는 부분만

모으면

$$\log f(x \mid y) = -\frac{x^2}{2\sigma_1^2} - \frac{(y-x)^2}{2\sigma_2^2} + \text{상수} = -\frac{1}{2}\Big(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\Big)x^2 + \frac{y}{\sigma_2^2}\,x + \text{상수}.$$

(b) $\tau = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}$ 이다. (c) $-\frac{\tau}{2}x^2 + \frac{y}{\sigma_2^2}x = -\frac{\tau}{2}\big(x - \frac{y}{\sigma_2^2\tau}\big)^2 + \text{상수}$ 이므로 $X \mid Y = y$는 평균 $\frac{y/\sigma_2^2}{\tau} = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}y = By$, 분산 $1/\tau = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ 인 정규다. (d) $\sigma_1^2 \to \infty$(중간 칸에 대한 사전 정보가 없다)이면 $B \to 1$이어서 y를 그대로 쓴다. $\sigma_2^2 \to 0$(아래 단의 잡음이 없다)이면 역시 $B \to 1$이다. 축소는 사전분포의 분산이 유한하고 관측에 잡음이 있을 때 일어난다.

연습 2.10 문제 20쪽

(a) $C - \mu = r(P-\mu) + E$에서 $E \perp P$이므로 $\operatorname{Var} C = r^2 \operatorname{Var} P + \operatorname{Var} E = r^2\sigma^2 + \sigma_E^2$ 이다. (b) $\operatorname{Var} C = \sigma^2$ 이면 $r^2 = 1 - \sigma_E^2/\sigma^2$ 이다. $r \geq 0$이므로 $r = \sqrt{1 - \sigma_E^2/\sigma^2}$ 이다. $\sigma_E^2 < \sigma^2$ 이어야 오른쪽이 양수이고 r이 실수다. $\sigma_E^2 \geq \sigma^2$ 이면 잡음만으로 분산이 σ^2을 넘어 안정할 수 없다. (c) $\operatorname{Cov}(P, C) = \operatorname{Cov}(P,\, r(P-\mu)+E) = r \operatorname{Var} P + \operatorname{Cov}(P,E) = r\sigma^2$ 이고 $\sqrt{\operatorname{Var} P \operatorname{Var} C} = \sigma^2$ 이므로 $\operatorname{Corr}(P, C) = r$ 이다.

3절

연습 3.1 문제 24쪽

$$\Sigma_{11\cdot 2} = \sigma_1^2 - \frac{\rho^2\sigma_1^2\sigma_2^2}{\sigma_2^2} = \sigma_1^2(1-\rho^2), \qquad |\Sigma| = \sigma_2^2 \cdot \sigma_1^2(1-\rho^2), \qquad B = \frac{\rho\sigma_1}{\sigma_2}.$$

보조정리 3.3(d)로

$$\Sigma^{-1} = \begin{pmatrix} \dfrac{1}{\sigma_1^2(1-\rho^2)} & -\dfrac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ -\dfrac{\rho}{\sigma_1\sigma_2(1-\rho^2)} & \dfrac{1}{\sigma_2^2} + \dfrac{\rho^2}{\sigma_2^2(1-\rho^2)} \end{pmatrix} = \frac{1}{|\Sigma|}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix},$$

이고 마지막 식은 수반행렬로 직접 구한 역행렬과 같다.

연습 3.2 문제 24쪽

보조정리 3.3(d)의 (1,1) 블록과 (1,2) 블록이 $\Lambda_{11} = \Sigma_{11\cdot 2}^{-1}$, $\Lambda_{12} = -\Sigma_{11\cdot 2}^{-1}B$ 이므로 $-\Lambda_{11}^{-1}\Lambda_{12} = \Sigma_{11\cdot 2}\Sigma_{11\cdot 2}^{-1}B = B$.

연습 3.3 문제 26쪽

$\rho = 0$이면 지수가 $-\frac{u^2}{2} - \frac{w^2}{2}$로 갈라진다. 거꾸로 $f = g(x)h(y)$이면 $\frac{\partial^2}{\partial x\,\partial y}\log f = 0$이어야 하는데

$$\frac{\partial^2}{\partial x\,\partial y}\Big(-\frac{u^2 - 2\rho uw + w^2}{2(1-\rho^2)}\Big) = \frac{\rho}{(1-\rho^2)\,\sigma_X\sigma_Y}$$

이므로 $\rho = 0$이다.

연습 3.4 문제 26쪽

$X_2 = \rho X_1 + \sqrt{1-\rho^2}\,Z$로 두면 (X_1, Z)는 독립 표준정규이므로 회전대칭이다. (X_1, Z)의 편각을 θ, $\alpha = \arcsin\rho$로 쓰면

$$X_2 > 0 \iff \sin\alpha\cos\theta + \cos\alpha\sin\theta > 0 \iff \sin(\theta+\alpha) > 0.$$

$X_1 > 0$은 $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, $X_2 > 0$은 $\theta \in (-\alpha, \pi-\alpha)$이고, 교집합은 폭 $\frac{\pi}{2}+\alpha$의 부채꼴이다. 편각이 균등하므로 확률은 $\frac{\pi/2+\alpha}{2\pi} = \frac{1}{4} + \frac{\arcsin\rho}{2\pi}$.

연습 3.5 문제 27쪽

$\mathbf{a} = (1,-2,1)'$이면 $\mathbf{a}'\boldsymbol{\mu} = 1-0-1 = 0$, $\Sigma\mathbf{a} = (0,-1,0)'$이므로 $\mathbf{a}'\Sigma\mathbf{a} = 2$. 따라서 $X_1 - 2X_2 + X_3 \sim \mathcal{N}(0, 2)$.

연습 3.6 문제 27쪽

$\operatorname{rank}\Sigma < n$이면 $\Sigma\mathbf{a} = \mathbf{0}$인 $\mathbf{a} \neq \mathbf{0}$이 있다. 정리 3.8로 $\mathbf{a}'\mathbf{X} \sim \mathcal{N}(\mathbf{a}'\boldsymbol{\mu},\ \mathbf{a}'\Sigma\mathbf{a}) = \mathcal{N}(\mathbf{a}'\boldsymbol{\mu}, 0)$, 곧 점질량이다(mgf가 $e^{t\,\mathbf{a}'\boldsymbol{\mu}}$).

연습 3.7 문제 28쪽

$(X_1 + X_2,\ X_1 - X_2)$는 정리 3.8로 결합정규이고

$$\operatorname{Cov}(X_1 + X_2,\ X_1 - X_2) = \operatorname{Var}X_1 - \operatorname{Var}X_2.$$

분산이 같으면 0이므로 정리 3.11로 독립이다. 분산이 다르면 상관이 있어 종속이다.

연습 3.8 문제 31쪽

조건 블록과 교차공분산은

$$\Sigma_{22} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}, \qquad \Sigma_{22}^{-1} = \frac{1}{4}\begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}, \qquad \Sigma_{12} = (\Sigma_{31}, \Sigma_{32}) = (0,\ 1)$$

이다. $B = (0,1)\,\Sigma_{22}^{-1} = (-\frac{1}{2},\ 1)$이므로

$$X_3 \mid (x_1, x_2) \sim \mathcal{N}\Big(-1 - \tfrac{1}{2}(x_1 - 1) + x_2,\ \ 2 - \big(0\cdot(-\tfrac{1}{2}) + 1\cdot 1\big)\Big) = \mathcal{N}\Big(-\tfrac{1}{2} - \tfrac{1}{2}x_1 + x_2,\ 1\Big).$$

검산: $1/\Lambda_{33} = 1$, $-\Lambda_{31}/\Lambda_{33} = -\tfrac{1}{2}$, $-\Lambda_{32}/\Lambda_{33} = 1$.

연습 3.9 문제 31쪽

$X_2 \mid X_1$: $B = \Sigma_{21}/\Sigma_{11} = \tfrac{2}{4}$이므로 평균 $\tfrac{1}{2}(x_1 - 1)$, 분산 $2 - \tfrac{2^2}{4} = 1$.

$X_2 \mid (X_1, X_3)$: 조건 블록 $\mathrm{diag}(4,2)$, $\Sigma_{12} = (2,1)$, $B = (\tfrac{1}{2}, \tfrac{1}{2})$이므로 평균 $\tfrac{1}{2}(x_1 - 1) + \tfrac{1}{2}(x_3 + 1)$, 분산 $2 - (\tfrac{1}{2}\cdot 2 + \tfrac{1}{2}\cdot 1) = \tfrac{1}{2}$. 검산: $1/\Lambda_{22} = \tfrac{1}{2}$.

조건을 하나 더 걸자 분산이 1에서 $\tfrac{1}{2}$로 줄었다. 따름정리 3.14(c)를 X_1이 주어진 조건부분포 안에서 다시 쓴 것이다. X_3은 X_1과 독립이지만, X_1이 주어진 뒤에도 X_2에 대한 정보를 더한다.

연습 3.10 문제 31쪽

(a) $m(X) = E[Y \mid X]$로 쓰면

$$E(Y-g)^2 = E(Y-m)^2 + E(m-g)^2 + 2E\big[(Y-m)(m-g)\big],$$

이고 마지막 항은 반복기댓값으로 0이다. 따라서 $g = m$에서 최소다.

(b) a로 미분하면 $a = EY - b\,EX$이고, 이를 넣고 b로 미분하면 $b = \mathrm{Cov}(X,Y)/\mathrm{Var}\,X$.

(c) 따름정리 3.19로 $E[Y \mid X] = \mu_Y + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}\,X}(X - \mu_X)$이고 이는 (b)의 직선이다.

연습 3.11 문제 32쪽

$\mathbf{X}_1 = (X_i, X_j)$, $\mathbf{X}_2 =$ 나머지로 나눈다. 정리 3.13로 $\mathbf{X}_1 \mid \mathbf{X}_2$는 이변량 정규이고, 정리 3.15으로 그 공분산은 $\Lambda^{-1}_{[ij]}$이다. 2×2 역행렬을 직접 쓰면

$$\Lambda^{-1}_{[ij]} = \begin{pmatrix} \Lambda_{ii} & \Lambda_{ij} \\ \Lambda_{ji} & \Lambda_{jj} \end{pmatrix}^{-1} = \frac{1}{\Lambda_{ii}\Lambda_{jj} - \Lambda_{ij}^2} \begin{pmatrix} \Lambda_{jj} & -\Lambda_{ij} \\ -\Lambda_{ji} & \Lambda_{ii} \end{pmatrix}$$

이므로 조건부 공분산은 $-\Lambda_{ij}/(\Lambda_{ii}\Lambda_{jj} - \Lambda_{ij}^2)$이다. 조건부분포가 결합정규이므로 정리 3.11에 의해

$$X_i \perp X_j \mid \text{나머지} \iff \text{조건부 공분산} = 0 \iff \Lambda_{ij} = 0.$$

연습 3.12 문제 32쪽

$c^2 = 5.991$이다. 두 반지름은 $\sqrt{5.991 \times 7.355} = 6.64$인치와 $\sqrt{5.991 \times 2.170} = 3.61$인치. $|\hat{\Sigma}| = 15.96$이므로 넓이는 $\pi \times 5.991 \times \sqrt{15.96} = 75.2$제곱인치다.

연습 3.13 문제 33쪽

$\mathbf{d} = \mathbf{X} - \boldsymbol{\mu}$로 쓰면 Cauchy–Schwarz로

$$(\mathbf{a}'\mathbf{d})^2 = \big((\Sigma^{1/2}\mathbf{a})'(\Sigma^{-1/2}\mathbf{d})\big)^2 \le (\mathbf{a}'\Sigma\mathbf{a})\,(\mathbf{d}'\Sigma^{-1}\mathbf{d}),$$

이고 등호는 $\mathbf{a} \propto \Sigma^{-1}\mathbf{d}$에서 성립한다. 따라서 상한이 Q이고, 사건 "모든 $\mathbf{a}$에 대해 $|\mathbf{a}'\mathbf{d}| \le c\sqrt{\mathbf{a}'\Sigma\mathbf{a}}$"는 $\{Q \le c^2\}$와 같다. 확률은 $F_{\chi^2(n)}(c^2)$이다.

연습 3.14 문제 34쪽

$0 < \rho < 1$이면 $\rho\dfrac{\sigma_Y}{\sigma_X} < \dfrac{\sigma_Y}{\sigma_X} < \dfrac{\sigma_Y}{\rho\sigma_X}$. Galton의 자료에서는 $0.646 < 1.409 < 3.071$이다.

연습 3.15 문제 34쪽

(X, Y)는 독립 정규 (T, E_1, E_2)의 선형변환이므로 결합정규다. $\operatorname{Var} X = \operatorname{Var} Y = \tau^2 + s^2$, $\operatorname{Cov}(X, Y) = \tau^2$이므로

$$E[Y \mid X = x] = \mu + R\,(x - \mu), \qquad R = \frac{\tau^2}{\tau^2 + s^2}\ \text{(신뢰도)}.$$

첫 비행이 극단적으로 좋았다면 그 일부는 운(E_1)이고, 운은 다음 비행에 이어지지 않는다. 그래서 둘째 비행은 평균 쪽으로 돌아온다. 칭찬이나 질책과 무관하게 일어나는 일이다.

연습 3.16 문제 34쪽

(a) $\mathbf{a}'\mathbf{X} = C\mathbf{X}$ ($C = \mathbf{a}'$, $1 \times n$)이므로 정리 3.8로 $\mathbf{a}'\mathbf{X} \sim \mathcal{N}(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a})$이다. (b) 가정에 $\mathbf{a} = \mathbf{e}_i$를 넣으면 X_i가 정규이므로 $EX_i^2 < \infty$이다. 그러면 $|X_iX_j| \le \frac{1}{2}(X_i^2 + X_j^2)$로 모든 공분산이 존재한다. 공분산의 존재에는 2차 적률이 유한하다는 것이 필요하다. (c) 보조정리 3.1로 $E\mathbf{t}'\mathbf{X} = \mathbf{t}'\boldsymbol{\mu}$, $\operatorname{Var}(\mathbf{t}'\mathbf{X}) = \mathbf{t}'\Sigma\mathbf{t}$이고, 가정으로 $\mathbf{t}'\mathbf{X}$가 정규이므로 그 평균과 분산을 가진 정규다. (d) $M_{\mathbf{X}}(\mathbf{t}) = E\,e^{\mathbf{t}'\mathbf{X}} = M_{\mathbf{t}'\mathbf{X}}(1) = \exp\!\big(\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\big)$이다. $\mathbf{t}$마다 다른 선형결합 $\mathbf{t}'\mathbf{X}$의 정규성을 썼으므로 "모든 $\mathbf{a}$"가 필요하다. 성분마다($\mathbf{a} = \mathbf{e}_i$) 정규인 것만으로는 부족하다는 것이 5.1절의 반례다.

연습 3.17 문제 35쪽

(a) $\{\mathbf{X} \in E_c\} = \{Q \le c^2\}$이고 정리 3.17로 $Q \sim \chi^2(n)$이다. (b) $\Sigma^{-1} = V\Lambda^{-1}V'$이므로 $(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{y}'\Lambda^{-1}\mathbf{y} = \sum_i y_i^2/\lambda_i$이다. $\mathbf{y}$ 좌표에서 E_c는 $\sum_i (y_i/(c\sqrt{\lambda_i}))^2 \le 1$, 곧 축이 좌표축이고 반지름이 $c\sqrt{\lambda_i}$인 타원체다. $\mathbf{x} = \boldsymbol{\mu} + V\mathbf{y}$이므로 $\mathbf{x}$ 좌표에서 축은 $\mathbf{v}_i$ 방향이다(회전은 길이를 바꾸지 않는다). (c) $\chi^2(2) = \Gamma(1, 2)$의 밀도는 $\frac{1}{2}e^{-w/2}$ ($w > 0$)이고, 이것은 평균 2인 지수분포다. $F(c^2) = \int_0^{c^2} \frac{1}{2}e^{-w/2}dw = 1 - e^{-c^2/2}$이다. (d) 반지름이 $c\sqrt{\lambda_1}$, $c\sqrt{\lambda_2}$이므로 넓이는 $\pi c^2\sqrt{\lambda_1\lambda_2} = \pi c^2\sqrt{|\Sigma|}$이다($|\Sigma| = \lambda_1\lambda_2$).

4절

연습 4.1 문제 37쪽

이론값은 평균 $5/3 = 1.667$, 분산 $4/3 = 1.333$이다. 10^6개 가운데 상자에 들어오는 점은 600여 개라서 Monte Carlo 평균의 표준오차가 $\sqrt{1.33/600} \approx 0.05$ 정도다. 노트북의 한 실행에서는 1.688 ± 0.089(2 표준오차)였다.

연습 4.2 문제 37쪽

$\Sigma_1 = L_1 L_1'$, $\Sigma_0 = L_0 L_0'$로 쓰면

$$\operatorname{tr}(\Sigma_1^{-1}\Sigma_0) = \left\| L_1^{-1} L_0 \right\|_F^2, \qquad (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \Sigma_1^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \left\| L_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) \right\|^2,$$

$$\log|\Sigma_k| = 2\sum_i \log (L_k)_{ii}.$$

$L_1^{-1}L_0$은 삼각방정식 $L_1 M = L_0$을 풀어 얻는다. 노트북 4.6절의 두 분포에서 정답은 0.8621이고, Monte Carlo 추정(2×10^5개)은 0.860이었다.

연습 4.3 문제 37쪽

아들: 기울기 0.713, 잔차 표준편차 2.30. 딸: 기울기 0.661, 잔차 표준편차 2.02. 딸에 1.08을 곱하면 기울기 $0.661 \times 1.08 = 0.714$, 잔차 표준편차 2.18로 아들과 거의 같아진다. 1.08은 평균만이 아니라 회귀 구조까지 꽤 잘 맞춘다. 잔차 분산은 아직 조금 다르다.

5절

연습 5.1 문제 39쪽

$Y^2 = X^2$이므로 $\operatorname{Cov}(X^2, Y^2) = \operatorname{Var} X^2 = EX^4 - (EX^2)^2 = 3 - 1 = 2$.

연습 5.2 문제 40쪽

$I = 0$이면 $\|\mathbf{X}\|^2 \sim \chi^2(2)$, $I = 1$이면 $\|\mathbf{X}\|^2/9 \sim \chi^2(2)$이고, $P(\chi^2(2) > x) = e^{-x/2}$이다. $q = 9.210$에서 $\|\mathbf{X}\|^2 > 1.8q = 16.58$일 확률은

$$0.9\, e^{-16.58/2} + 0.1\, e^{-16.58/18} = 0.9 \times 0.00025 + 0.1 \times 0.398 = 0.0400,$$

으로 명목 1%의 네 배다.

연습 5.3 문제 41쪽

$\pi = 481/934 = 0.515$이다.

$$\text{집단 내} = 0.515\,(2.62)^2 + 0.485\,(2.35)^2 = 6.22, \quad \text{집단 간} = 0.515 \times 0.485 \times (5.13)^2 = 6.57,$$

합은 12.80(표준편차 3.58). 딸에 1.08을 곱한 뒤에는 집단 내 6.67, 집단 간 거의 0, 합 6.67(표준편차 2.58)이다.

연습 5.4 문제 41쪽

$\mathbf{a} = (2,-1)'$이면 $\Sigma\mathbf{a} = (2-2,\,4-4)' = \mathbf{0}$이므로 $\mathbf{X}$는 직선 $x_2 = 2x_1$ 위에 있다. $\mathbf{t}'\Sigma\mathbf{t} = t_1^2 + 4t_1t_2 + 4t_2^2 = (t_1+2t_2)^2$이므로 $M_{\mathbf{X}}(\mathbf{t}) = \exp\{\tfrac{1}{2}(t_1+2t_2)^2\}$이고, 이는 $X_1 \sim \mathcal{N}(0,1)$, $X_2 = 2X_1$의 mgf다.

6절

연습 6.1 문제 44쪽

$\mathbf{Y}'A\mathbf{Y} = \operatorname{tr}(\mathbf{Y}'A\mathbf{Y}) = \operatorname{tr}(A\mathbf{Y}\mathbf{Y}')$이고 대각합과 기댓값은 순서를 바꿀 수 있으므로

$$E[\mathbf{Y}'A\mathbf{Y}] = \operatorname{tr}\big(A\,E[\mathbf{Y}\mathbf{Y}']\big) = \operatorname{tr}\big(A(V + \mathbf{m}\mathbf{m}')\big) = \operatorname{tr}(AV) + \mathbf{m}'A\mathbf{m}.$$

연습 6.2 문제 44쪽

$\Sigma_1 = I$, $\boldsymbol{\mu}_1 = \mathbf{0}$을 넣으면 $\operatorname{tr}\Sigma_0 = \sum \sigma_i^2$, 이차형식은 $\sum \mu_i^2$, $\log\frac{|I|}{|\Sigma_0|} = -\sum \log\sigma_i^2$이다.

연습 6.3 문제 44쪽

$(\Sigma - \lambda_1 I)\mathbf{v} = \mathbf{0}$의 첫 행 $(\Sigma_{11} - \lambda_1)v_1 + \Sigma_{12}v_2 = 0$에서

$$\frac{v_2}{v_1} = \frac{\lambda_1 - \Sigma_{11}}{\Sigma_{12}} = \frac{7.355 - 3.191}{2.062} = 2.02.$$

y를 센티미터로 바꾸면 두 회귀선의 기울기는 정확히 2.54배(1.64와 7.80)가 된다. 그런데 주성분의 기울기는 $2.02 \times 2.54 = 5.13$이 아니라 7.33이 된다. 주성분은 척도에 따라 방향 자체가 바뀐다.