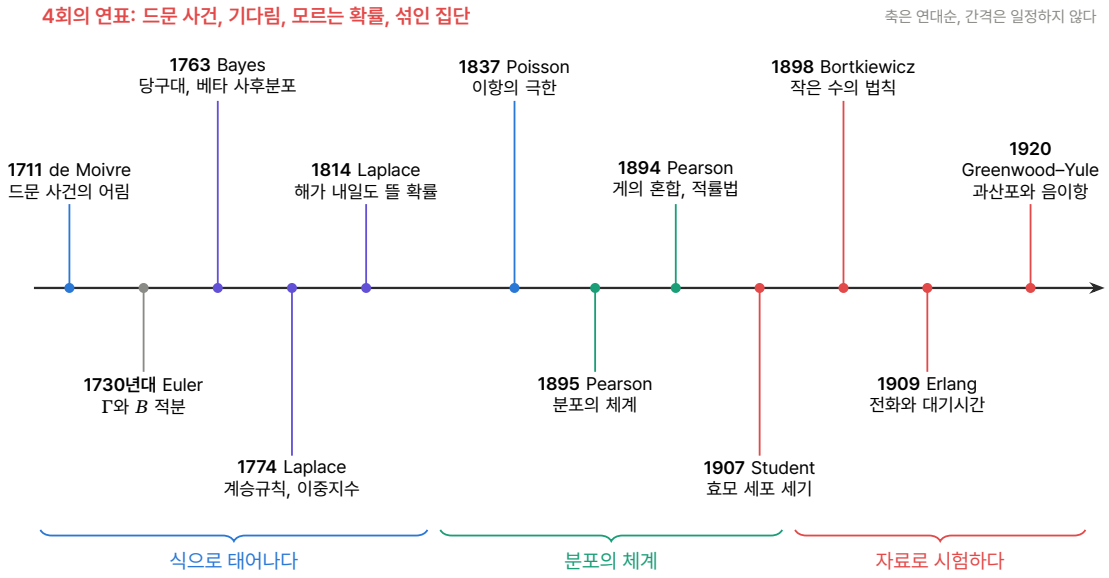
수리통계학 × 통계학의 역사

4회 강의노트

셈하는 분포와 섞인 분포
말에 차여 죽은 병사들

1711 → 1898

드문 사건을 세면 Poisson 분포가, 그 사건을 기다리면 감마분포가, 확률 자체를 모르면 베타분포가 나온다. 여러 집단이 섞이면 분산이 평균보다 커진다. 이 노트는 이 분포들을 한 가족으로 묶어 서로의 극한과 변환으로 증명하고, Bortkiewicz의 기병대, Student의 효모, Madison의 *may*로 확인한다.



다루는 것 이항·음이항·초기하·다항 · Poisson과 그 극한 · Poisson 과정 · 감마와 χ^2 · 베타와 계승규칙 · 혼합과 과산포

원자료 Bortkiewicz(1898)의 기병대 · Student(1907)의 효모 · The Federalist의 may

노트북 `sessions/04_counting/04_counting.ipynb` — 절·정리 번호가 같다

워크북 `04_counting_workbook.pdf` (손으로 푸는 35문항) · 풀이판 `_workbook_sol.pdf`

선수 지식 1–3회 전체 (특히 mgf, 야코비안, 전분산 공식)

이 노트를 쓰는 법

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

각 절은 **역사 → 정의·정리 → 증명 → 그림 → 연습문제** 순서다. 핵심 정리의 증명은 번호 붙은 단계 ①②③으로 쓰고, 단계마다 오른쪽 끝에 근거(앞 정리의 번호나 도구)를 단다. 읽을 때는 근거를 가리고 "왜 이 단계가 성립하는가"를 먼저 말해 본다. 극한, 적분, 미분의 순서를 바꾸는 곳에는 **조건 확인** 단락을 둔다. 나머지 정리는 **증명 전략**만 적고, 증명을 소문항으로 나눈 유도형 연습으로 돌린다. 이 연습은 건너뛰지 않고 손으로 쓴다.

근거에 붙은 [T1]–[T8]은 별책 **증명 도구 상자**의 도구 번호다. 이 회차에서는 [T1] 지시함수, [T2] 귀류, [T3] 귀납, [T4] 망원 합, [T5] 극한과 급수, [T6] 합과 적분의 교환, [T7] 변수변환, [T8] 급수 전개를 쓴다.

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

상자의 뜻

> **정의** 대상을 처음 정하는 문장. 왼쪽 선이 보라다.

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

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

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

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

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

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

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

차례

0 들어가며 1

0.1 1–3회에서 가져오는 것 3

0.2 분포의 가족과 모수 3

1 Bernoulli 시행의 가족 4

1.1 이항분포, 다시 4

1.2 음이항분포 4

1.3 초기하분포 6

1.4 다항분포 7

2 Poisson 분포 — 드문 사건의 법칙 9

2.1 이항분포의 극한 9

2.2 Poisson의 합과 나누기 10

2.3 Bortkiewicz의 기병대 (1898) 11

3 Poisson 과정과 감마분포 14

3.1 Poisson 과정 14

3.2 감마함수와 감마분포 16

3.3 χ^2 분포 17

4 베타분포 — Bayes의 당구대 20

4.1 베타함수와 베타분포 20

4.2 Bayes의 당구대 (1763) 21

5 혼합분포 — 섞인 집단 25

5.1 정의와 적률 25

5.2 Poisson–감마 혼합은 음이항 26

5.3 정규의 혼합 — Pearson의 게 (1894) . . . 27

6 데이터로 — 말발굽, 효모, 그리고 may 29

6.1 적률을 맞추기 . . . 29

6.2 세 자료 . . . 30

7 부숴보기 — 가족의 경계 32

7.1 Poisson인 척하는 음이항 . . . 32

7.2 Poisson 근사가 틀릴 때 . . . 32

7.3 무기억성은 지수와 기하뿐 . . . 32

7.4 혼합은 이름표를 잃는다 . . . 32

8 Beautiful insights — 같은 가족, 오늘의 문제 34

8.1 계승규칙과 add-one smoothing . . . 34

8.2 베타–이항과 Thompson 샘플링 . . . 34

8.3 셈 자료의 회귀 — Poisson과 음이항 . . . 34

8.4 혼합 모형 — GMM과 전문가 혼합 . . . 35

9 정리와 역사 노트 36

9.1 한 장 요약 . . . 36

9.2 Stigler의 명명 법칙 . . . 36

9.3 7 Pillars 지도 . . . 37

9.4 다음 회 — 정규분포의 탄생과 Galton의 타원 (1733 → 1901) . . . 37

A 연습문제 풀이 39

0

들어가며

> **여는 질문.** 1875년부터 1894년까지 20년 동안, 프로이센 군대의 군단마다 해마다 말에 차여 죽은 병사의 수가 기록되었다. 어떤 해, 어떤 군단에서는 한 명도 죽지 않았고, 어떤 곳에서는 네 명이 죽었다. 이 수는 어떤 법칙을 따르는가? 법칙이 있다면, 한 군단에서 한 해에 네 명 이상이 죽는 일은 얼마나 드물어야 하는가?

러시아 출신의 통계학자 Ladislaus von Bortkiewicz는 1898년 *Das Gesetz der kleinen Zahlen*(작은 수의 법칙)에서 이 기록을 분석했다. 말에 차여 죽는 일은 병사 한 명에게는 아주 드문 일이지만 한 군단에는 병사가 많다. 드물고 서로 무관한 사건을 많이 모아 세면, 그 수는 하나의 분포를 따른다. 그 분포를 1837년에 이름 붙인 사람이 Poisson이고, 그보다 120여 년 앞서 그 식을 처음 쓴 사람이 de Moivre였다(2절).

Poisson 분포는 혼자가 아니다. 그 사건이 일어나기를 기다리는 시간은 지수분포와 감마분포를 따르고(3절), 성공 확률 자체를 모를 때 그 확률에 대한 믿음은 베타분포가 된다(4절). 이 분포들은 서로의 극한, 합, 비로 이어진 하나의 가족이다(그림 0.1). 그리고 가족의 서로 다른 구성원이 섞이면, 셈한 수의 분산이 평균보다 커진다(5절). Bortkiewicz의 자료에서도, 18세기 영어 문헌의 단어 수에서도 그 흔적이 보인다(6절).

학습 목표

1. 이항·음이항·초기하·다항 분포를 Bernoulli 시행에서 유도하고 평균·분산·mgf를 구한다.
2. 이항분포의 Poisson 극한을 증명하고, Poisson 분포의 합과 나누기(thinning)를 mgf로 다룬다.
3. Poisson 과정의 대기시간이 지수·감마분포임을 보이고, 감마함수와 χ^2 분포를 세운다.
4. 두 감마의 비에서 베타분포를 얻고, Bayes의 당구대와 Laplace의 계승규칙을 유도한다.
5. 혼합분포의 평균과 분산을 전분산 공식으로 구하고, Poisson–감마 혼합이 음이항임을 증명한다.
6. 원자료에서 분산/평균 비로 Poisson 모형을 시험하고, 과산포를 음이항으로 설명한다.

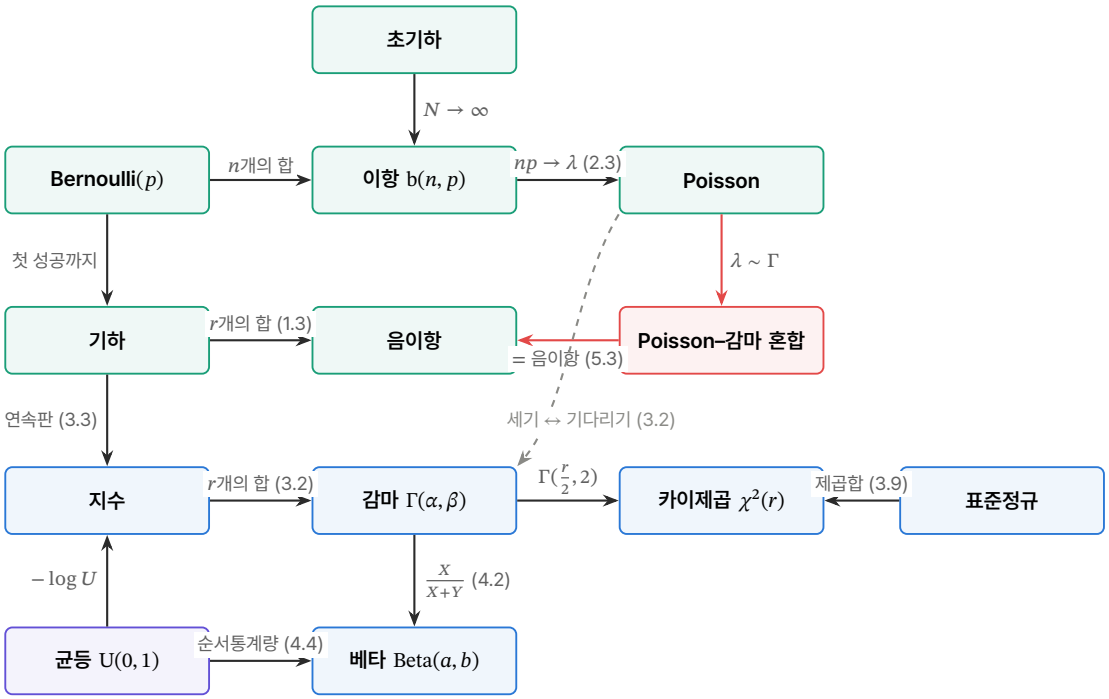


그림 0.1 이 회차의 분포 가족. 화살표 위의 글은 한 분포에서 다른 분포를 얻는 방법(극한, 독립합, 비, 섞기)이고 괄호 안은 증명하는 정리다.

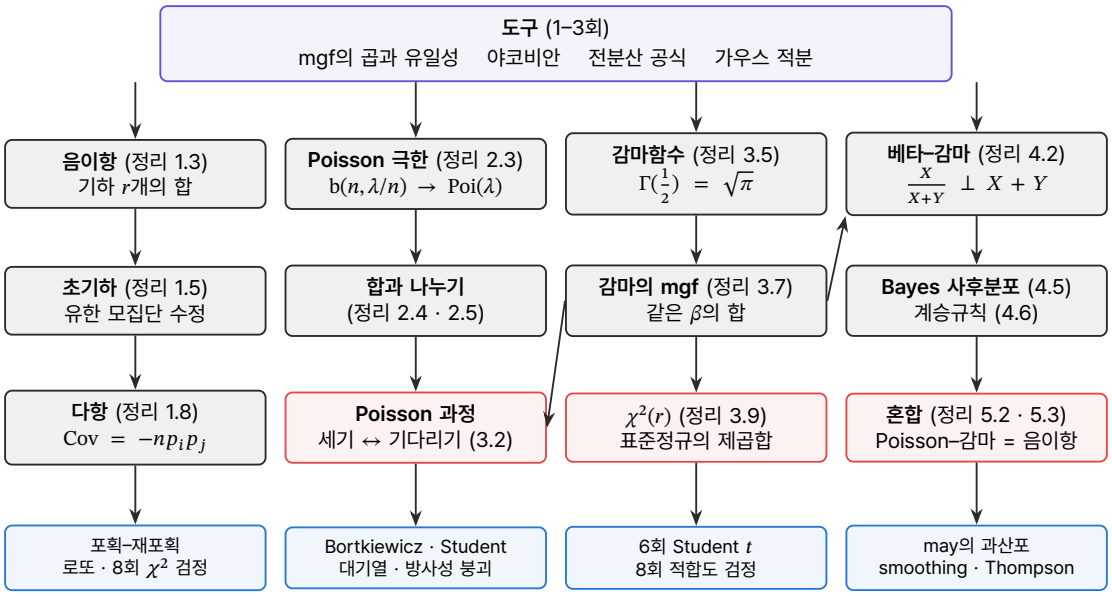


그림 0.2 이번 회차 정리들의 의존 관계. 화살표는 "증명에 쓰인다"는 뜻이다.

0.1 1–3회에서 가져오는 것

앞 회차	내용
1회 정리 1.4 · 1.6	조합의 수 $\binom{n}{k}$, 이항정리
2회 정리 2.4 · 예제 1.4	이항분포 $\mathrm{b}(n,p)$의 평균 np·분산 npq, 기하분포의 평균 $1/p$
2회 정리 3.3 · 3.6 · 상자	mgf에서 적률, 아핀 변환, mgf의 유일성
2회 정리 4.3 · 4.7	단조변환의 밀도, 역변환 표집
3회 정리 2.5 · 2.6	반복기댓값, 전분산 공식 $\operatorname{Var} Y = E\operatorname{Var}(Y \mid X) + \operatorname{Var} E(Y \mid X)$
3회 정리 3.4 · 4.2	독립합의 분산과 mgf의 곱, 공분산의 계산 규칙
3회 정리 5.1 · 5.3 · 5.5	야코비안 변수변환, 가우스 적분 $\int e^{-x^2/2} = \sqrt{2\pi}$, 합성곱
3회 예제 3.5 · 5.7	iid 최댓값의 분포, 지수분포의 합의 밀도

0.2 분포의 가족과 모수

이 회차에서 다루는 것은 하나하나의 분포가 아니라 **모수로 이름 붙인 분포의 가족**이다. $\mathrm{b}(n,p)$는 n과 p를 정하면 하나의 분포가 되는, 분포들의 모임이다. 통계학의 일은 대개 "자료가 이 가족의 한 구성원에서 나왔다"고 보고 모수를 거꾸로 추정하는 것이다(8회 이후). 그러려면 먼저 각 가족이 어떤 상황에서 나오는지, 평균과 분산이 모수와 어떻게 이어지는지 알아야 한다. 이 노트의 각 절은 같은 순서로 쓴다. **어떤 실험에서 나오는가** → pmf 또는 pdf → 평균·분산·mgf → 다른 가족과의 관계.

1

Bernoulli 시행의 가족

성공 확률이 p인 독립 시행을 되풀이한다. 무엇을 세느냐에 따라 서로 다른 분포가 나온다. 정해진 n번 안의 성공 수(이항), 정해진 수의 성공까지의 실패 수(음이항), 그리고 되돌려 넣지 않고 뽑을 때의 성공 수(초기하)다. 결과가 둘이 아니라 여럿이면 다항분포가 된다.

1.1 이항분포, 다시

2회 정리 2.4에서 $K \sim \mathrm{b}(n,p)$이면 $P(K=k)=\binom{n}{k}p^k q^{n-k}$, $EK=np$, $\operatorname{Var}K=npq$, $\operatorname{mgf}(q+pe^t)^n$이었다. 여기에 한 가지를 더한다.

> **정리 1.1** (이항분포의 합) $K_1 \sim \mathrm{b}(n_1,p)$와 $K_2 \sim \mathrm{b}(n_2,p)$가 독립이면 $K_1+K_2 \sim \mathrm{b}(n_1+n_2,p)$.

증명. **전략** 독립합의 mgf는 mgf의 곱이다. 곱이 다시 이항의 mgf 꼴이 됨을 보이고 mgf의 유일성을 쓴다. 합성곱으로 직접 셀 수도 있다. 단계별 증명은 연습 1.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 40쪽에 있다.

p가 같아야 한다. 확률이 다른 두 이항의 합은 이항이 아니다(연습 1.1).

1.2 음이항분포

> **정의 1.2** (음이항분포) 성공 확률 p인 독립 시행에서 r번째 성공이 나오기 전까지의 실패 수 Y의 분포를 **음이항분포** $\mathrm{NB}(r,p)$라 한다. $r=1$이면 **기하분포**다(2회 예제 1.4의 "시행 수"에서 1을 뺀 것).

정리 1.3 (음이항분포의 pmf와 적률) $Y \sim \mathrm{NB}(r, p)$, $q = 1 - p$이면

$$P(Y = k) = \binom{k + r - 1}{k} p^r q^k, \qquad k = 0, 1, 2, \dots,$$

$$EY = \frac{rq}{p}, \qquad \operatorname{Var} Y = \frac{rq}{p^2}, \qquad M_Y(t) = \left(\frac{p}{1 - qe^t}\right)^r \quad (qe^t < 1).$$

증명. **전략** pmf는 "마지막 시행이 r번째 성공"이라는 사건을 세어 구한다. 적률은 Y를 독립인 기하분포 r개의 합으로 쪼개고 독립합의 성질을 쓴다.

pmf

① $\{Y = k\}$는 "처음 $k + r - 1$번 가운데 성공이 정확히 $r - 1$번이고 $(k + r)$번째가 성공"인 사건과 같다. — r번째 성공의 정의

② 앞부분의 확률은 $\binom{k+r-1}{r-1} p^{r-1} q^k$이다. 성공이 올 $r - 1$개의 자리를 고르고, 자리가 정해진 결과 하나의 확률은 $p^{r-1} q^k$이다. — 2회 정리 2.4

③ $(k + r)$번째 시행은 앞의 시행들과 독립이므로 ②에 p를 곱한다. $\binom{k+r-1}{r-1} = \binom{k+r-1}{k}$이다. — 독립, 1회 정리 5.11

적률

① G_1을 첫 성공까지의 실패 수, G_j를 $(j - 1)$번째와 j번째 성공 사이의 실패 수라 하면 $Y = G_1 + \cdots + G_r$이다. — 정의

② $G_1, \dots, G_r$은 독립이고 모두 $P(G = k) = q^k p$를 따른다. 각 G_j는 서로 겹치지 않는 시행들로 정해지고, 성공 하나 뒤에서 시행은 처음과 같은 꼴로 다시 시작한다. — 시행의 독립

③ 기하분포의 mgf는 $\sum_{k \ge 0} q^k p e^{tk} = \frac{p}{1 - qe^t}$ $(qe^t < 1)$이다. — 기하급수 [T5]

④ $EG = \frac{q}{p}$, $\operatorname{Var} G = \frac{q}{p^2}$이다. — 2회 예제 1.4, 연습 2.3

⑤ 독립합이므로 평균과 분산은 r배, mgf는 r제곱이다. — 3회 정리 3.4

■

이름의 "음"은 pmf를 $\binom{-r}{k} p^r (-q)^k$로 쓸 수 있어서, 음수 지수의 이항급수 $(1 - q)^{-r} = \sum_k \binom{-r}{k} (-q)^k$의 항이기 때문이다. 이 표현으로 r이 정수가 아니어도 pmf가 정의된다. 이때 $\binom{k+r-1}{k}$는 $\frac{\Gamma(k+r)}{k!\,\Gamma(r)}$로 읽는다(3절). 5절에서 실수 r의 음이항이 Poisson 분포를 섞어 나오는 것을 본다. 음이항의 분산 $\frac{rq}{p^2} = EY \cdot \frac{1}{p}$는 언제나 평균보다 크다. **과산포**의 표준 모형이 된 이유다.

1.3 초기하분포

정의 1.4 (초기하분포) N개 가운데 K개가 "성공"인 모집단에서 n개를 **되돌려 넣지 않고** 뽑을 때 성공의 수 X의 분포를 **초기하분포**라 한다.

$$P(X=x)=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}},\qquad \max(0,n-N+K)\le x\le \min(n,K).$$

n개짜리 부분집합이 모두 같은 확률이므로 1회 따름정리 4.10의 고전적 정의다. 성공 x개와 실패 $n-x$개를 고르는 방법을 센다.

정리 1.5 (초기하분포의 평균과 분산) $p=K/N$으로 두면

$$EX=np,\qquad \operatorname{Var}X=np(1-p)\cdot\frac{N-n}{N-1}.$$

증명. **전략** X를 지시변수의 합으로 쪼갠다 [T1]. 평균은 선형성으로 구한다. 지시변수들은 독립이 아니므로 분산에는 공분산 항이 남는다.

① $\mathbf{1}_i$를 i번째로 뽑은 것이 성공이라는 지시변수라 하면 $X=\sum_{i=1}^{n}\mathbf{1}_i$이다. — 지시함수 [T1]

② 뽑는 순서의 대칭으로 i번째로 뽑은 것은 N개 가운데 고르게 하나다. 따라서 $P(\mathbf{1}_i=1)=\frac{K}{N}=p$이고 $EX=np$이다. — 대칭, 기댓값의 선형성

③ $i\neq j$이면 $P(\mathbf{1}_i\mathbf{1}_j=1)=\frac{K}{N}\cdot\frac{K-1}{N-1}$이다. — 곱셈법칙, 1회 정리 5.3

④ 공분산을 계산한다.

$$\operatorname{Cov}(\mathbf{1}_i,\mathbf{1}_j)=\frac{K(K-1)}{N(N-1)}-p^2=\frac{K}{N}\cdot\frac{N(K-1)-K(N-1)}{N(N-1)}=-\frac{p(1-p)}{N-1}.$$ ②③

⑤ 분산은 공분산의 합이다.

$$\begin{aligned}\operatorname{Var}X&=\sum_i\operatorname{Var}\mathbf{1}_i+\sum_{i\neq j}\operatorname{Cov}(\mathbf{1}_i,\mathbf{1}_j) && \text{3회 정리 4.2(c)}\\ &=np(1-p)-n(n-1)\frac{p(1-p)}{N-1}=np(1-p)\frac{N-n}{N-1}. && \text{④, } \operatorname{Var}\mathbf{1}_i=p(1-p)\end{aligned}$$

■

$\frac{N-n}{N-1}$을 **유한 모집단 수정**이라 한다. 되돌려 넣지 않으므로 뽑힌 것들이 서로 음의 상관을 가져 분산이 이항보다 작다(2회 연습 2.4의 에이스). N이 n보다 아주 크면 이 비는 1에 가깝고 초기하는 이항과 거의 같다. 여론조사에서 5,000만 명 가운데 1,000명을 뽑을 때 되돌려 넣는지를 걱정하지 않는 이유다. 반대로 $n=N$이면 분산이 0이다. 전수조사에는 표본 오차가 없다.

예제 1.6 (물고기 세기 — 포획–재포획) 호수의 물고기 N마리 가운데 $K = 200$마리를 잡아 표시하고 놓아준다. 나중에 $n = 100$마리를 잡았더니 $x = 16$마리가 표시되어 있었다. X는 초기하분포를 따르고 $EX = nK/N$이므로, $x = nK/N$을 풀면 $\hat{N} = \dfrac{nK}{x} = 1250$이다(Petersen 1896). 8회 이후의 "추정"의 첫 예다. 연습 1.5에서 이것이 가능도를 최대로 하는 N과 거의 같음을 본다.

1.4 다항분포

정의 1.7 (다항분포) 한 번의 시행 결과가 m개의 범주 가운데 하나이고 범주 j의 확률이 p_j($\sum p_j = 1$)일 때, 독립 시행 n번에서 범주별 횟수 $\mathbf{X} = (X_1, \dots, X_m)$의 분포를 **다항분포** $\mathrm{Mult}(n; p_1, \dots, p_m)$라 한다.

$$P(X_1 = x_1, \dots, X_m = x_m) = \frac{n!}{x_1! \cdots x_m!}\, p_1^{x_1} \cdots p_m^{x_m}, \qquad \sum_{j} x_j = n.$$

계수는 n번의 시행에 범주 이름을 붙이는 방법의 수다. x_1개 자리를 고르고($\binom{n}{x_1}$), 남은 자리에서 x_2개를 고르고, $\cdots$ 곱하면 이 다항계수가 된다.

정리 1.8 (다항분포의 주변과 공분산) $X_j \sim \mathrm{b}(n, p_j)$이고, $i \neq j$이면 $\mathrm{Cov}(X_i, X_j) = -np_ip_j$이다.

증명. **전략** 범주를 "j와 나머지"로 묶으면 각 시행이 Bernoulli가 된다. 공분산은 두 범주를 하나로 묶은 합의 분산에서 꺼낸다. 단계별 증명은 연습 1.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 40쪽에 있다.

합이 n으로 고정되어 있으므로 한 범주가 많으면 다른 범주는 적다. 그래서 공분산이 음이다. 8회의 χ^2 적합도 검정(Weldon의 주사위, Pearson 1900)이 이 분포 위에 선다.

연습문제

연습 1.1 (확률이 다르면 ★) $K_1 \sim \mathrm{b}(1, 0.1)$과 $K_2 \sim \mathrm{b}(1, 0.9)$가 독립이다. $K_1 + K_2$의 분포를 구하고, 같은 평균의 $\mathrm{b}(2, 0.5)$와 비교하라. 분산은 어느 쪽이 작은가? 풀이 39쪽

연습 1.2 (Banach의 성냥갑 ★★) 성냥갑 두 개에 성냥이 N개씩 있다. 매번 공평한 동전으로 한 갑을 골라 성냥을 하나 쓴다. 처음으로 빈 갑을 골랐을 때 다른 갑에 k개가 남아 있을 확률이 $\binom{2N-k}{N}2^{-(2N-k)}$임을 음이항분포로 보여라. 풀이 39쪽

연습 1.3 (음이항의 적률 ★) 정리 1.3의 mgf를 미분해 평균과 분산을 다시 구하라. 풀이 39쪽

연습 1.4 (로또와 초기하 ★) 45개 가운데 6개를 고르는 로또에서 내가 고른 6개 가운데 당첨 번호의 수 X는 초기하분포다. EX와 $\operatorname{Var} X$를 구하고, 정확히 3개가 맞을 확률을 구하라. 풀이 39쪽

연습 1.5 (포획–재포획의 가능도 ★★) 예제 1.6에서 $L(N) = P_N(X = 16)$을 N의 함수로 보자. $\frac{L(N)}{L(N-1)} \geq 1$일 필요충분조건이 $N \leq \frac{nK}{x}$임을 보이고, L을 최대로 하는 N이 $\lfloor nK/x \rfloor = 1250$임을 결론지어라. 풀이 39쪽

연습 1.6 (다항의 조건부 ★★) $(X_1, X_2, X_3) \sim \mathrm{Mult}(n; p_1, p_2, p_3)$일 때 $X_1 = x$가 주어지면 X_2의 조건부분포가 $\mathrm{b}\big(n - x, \frac{p_2}{p_2+p_3}\big)$임을 보여라. 풀이 40쪽

연습 1.7 (이항분포의 합: 증명 ★ 유도형) 정리 1.1을 두 가지로 증명하라.

(a) (mgf) $K_1 + K_2$의 mgf가 $(q + pe^t)^{n_1+n_2}$임을 보여라. 어느 정리로 mgf를 곱하는가?

(b) mgf의 유일성을 쓰려면 두 mgf가 0을 포함하는 열린구간에서 같아야 한다. 여기서는 어느 구간에서 같은가?

(c) (합성곱) $P(K_1 + K_2 = m) = \sum_j P(K_1 = j)P(K_2 = m - j)$에 pmf를 넣고, Vandermonde 항등식 $\sum_j \binom{n_1}{j}\binom{n_2}{m-j} = \binom{n_1+n_2}{m}$을 세기로 증명해 결론을 얻어라.

(d) 두 증명이 각각 p가 같다는 가정을 어디서 쓰는지 밝혀라.

풀이 40쪽

연습 1.8 (다항분포의 주변과 공분산: 증명 ★★ 유도형) $\mathbf{X} \sim \mathrm{Mult}(n; p_1, \dots, p_m)$이다. 정리 1.8를 다음 순서로 증명하라.

(a) 각 시행을 "범주 j" 와 "나머지" 로 나눠 $X_j \sim \mathrm{b}(n, p_j)$를 보여라.

(b) $i \neq j$이면 같은 방법으로 $X_i + X_j \sim \mathrm{b}(n, p_i + p_j)$를 보여라.

(c) $\operatorname{Var}(X_i + X_j) = \operatorname{Var} X_i + \operatorname{Var} X_j + 2\operatorname{Cov}(X_i, X_j)$에 (a)(b)를 넣어 $\operatorname{Cov}(X_i, X_j) = -np_ip_j$를 얻어라.

(d) (다른 증명) t번째 시행이 범주 i이면 1인 지시변수 $\mathbf{1}_{t,i}$로 $X_i = \sum_t \mathbf{1}_{t,i}$로 쓰고, 같은 시행에서 $\operatorname{Cov}(\mathbf{1}_{t,i}, \mathbf{1}_{t,j}) = -p_ip_j$, 다른 시행에서 0임을 써서 같은 결론을 얻어라.

풀이 40쪽

2

Poisson 분포 — 드문 사건의 법칙

2.1 이항분포의 극한

> **역사** Abraham de Moivre는 1711년 *De Mensura Sortis*에서, 성공 확률이 아주 작은 시행을 많이 할 때 성공이 k번일 확률을 $e^{-\lambda}\lambda^k/k!$ 꼴로 어림했다. 1837년 Siméon-Denis Poisson은 배심원 재판의 오판 확률을 다룬 *Recherches sur la probabilité des jugements*에서 같은 극한을 유도했고, 분포는 그의 이름을 얻었다. 그러나 이 분포가 실제 자료에 쓰이기 시작한 것은 60년 뒤 Bortkiewicz(1898)부터다.

정의 2.1 (Poisson 분포) $\lambda > 0$에 대해 $P(X=k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$, $k = 0, 1, 2, \dots$ 인 분포를 **Poisson 분포** $\mathrm{Poi}(\lambda)$라 한다.

$e^{\lambda} = \sum_{k} \lambda^k/k!$이므로 확률의 합은 1이다.

정리 2.2 (Poisson 분포의 적률) $X \sim \mathrm{Poi}(\lambda)$이면 $M_X(t) = \exp\!\big(\lambda(e^t - 1)\big)$ $(t \in \mathbb{R})$이고 $EX = \operatorname{Var} X = \lambda$.

증명. **전략** mgf의 급수를 지수함수의 급수로 알아보고, 누율생성함수 $\log M$을 미분한다. 단계별 증명은 연습 2.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 41쪽에 있다.

평균과 분산이 같다. 이것이 Poisson 모형을 자료로 시험하는 가장 간단한 기준이 된다(6절).

정리 2.3 (이항분포의 Poisson 극한) $n \to \infty$에서 $np_n \to \lambda > 0$이면, 모든 k에 대해

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \longrightarrow e^{-\lambda}\frac{\lambda^k}{k!}.$$

증명. **전략** pmf를 극한이 따로 보이는 네 인수로 쪼개고 인수마다 극한을 구한다 [T5]. 지수 꼴의 인수는 로그를 취해 급수 전개로 다룬다 [T8].

① $\lambda_n = np_n$으로 쓰고 $\binom{n}{k} = \dfrac{n(n-1)\cdots(n-k+1)}{k!}$을 넣으면 다음과 같다.

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{\lambda_n^k}{k!} \cdot \Big(1-\frac{\lambda_n}{n}\Big)^{n} \cdot \Big(1-\frac{\lambda_n}{n}\Big)^{-k}.$$

② 첫째 인수는 k개의 인수 $\dfrac{n-j}{n}$ $(j = 0, \dots, k-1)$의 곱이다. k는 고정이고 각 인수가 1로 가므로 곱도 1로 간다. 극한의 곱 [T5]

③ 둘째 인수는 $\lambda^k/k!$로 간다. $\lambda_n \to \lambda$

④ $0 \le x \le \frac{1}{2}$이면 $\log(1-x) = -x - \sum_{m\ge 2} \frac{x^m}{m}$이고 $\sum_{m\ge 2} \frac{x^m}{m} \le \frac{x^2}{2(1-x)} \le x^2$이다. n이 커서 $\lambda_n/n \le \frac{1}{2}$이면

$$\left| n \log\left(1 - \frac{\lambda_n}{n}\right) + \lambda_n \right| \le \frac{\lambda_n^2}{n} \to 0$$

이므로 셋째 인수는 $e^{-\lambda}$로 간다. 급수 전개 [T8], exp의 연속성

⑤ 넷째 인수는 1로 가는 수 $1 - \lambda_n/n$의 고정된 거듭제곱이므로 1로 간다. [T5]

⑥ 네 극한을 곱하면 $e^{-\lambda}\lambda^k/k!$이다. 극한의 곱 [T5]

■

"드물고(p 작음), 기회가 많고(n 큼), 서로 무관한(독립)" 사건의 수가 Poisson이다. 얼마나 가까운지도 말할 수 있다. Le Cam(1960)은 $\mathrm{b}(n,p)$와 $\mathrm{Poi}(np)$의 차이(모든 사건에 대한 확률 차이의 최댓값)가 np^2 이하임을 보였다. $n = 100$, $p = 0.02$이면 0.04 이하이고 실제로는 0.005쯤이다. $n = 20$, $p = 0.4$처럼 p가 크면 이 어림은 쓸모가 없다(7절, 그림 2.1).

2.2 Poisson의 합과 나누기

정리 2.4 (독립인 Poisson의 합) $X_i \sim \mathrm{Poi}(\lambda_i)$가 독립이면 $\sum_i X_i \sim \mathrm{Poi}\left(\sum_i \lambda_i\right)$.

증명. **전략** 두 개의 합을 합성곱과 이항정리로 직접 구하고, 개수에 대한 귀납법으로 넓힌다 [T3]. mgf의 곱으로도 증명한다. 단계별 증명은 연습 2.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 42쪽에 있다.

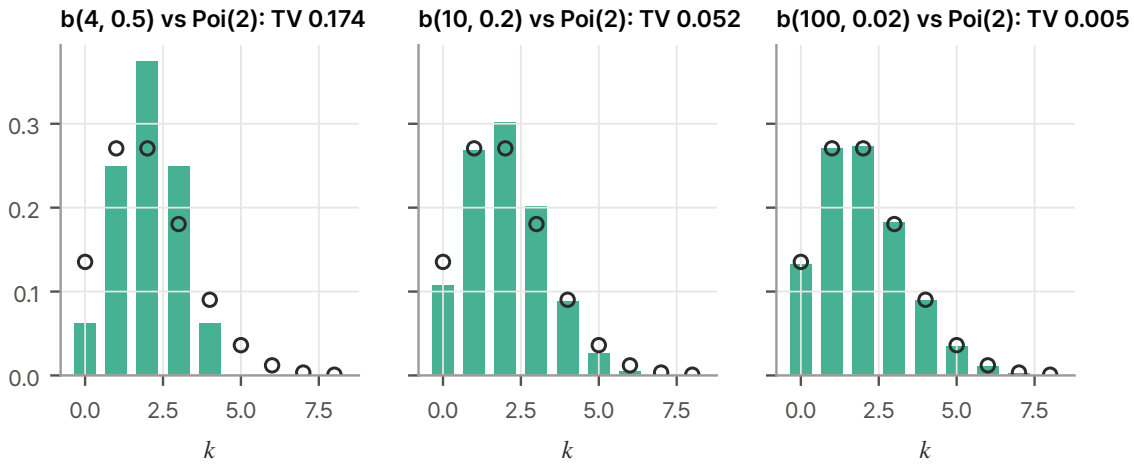


그림 2.1 같은 평균 $\lambda = 2$의 이항분포(막대)와 Poisson 분포(원). p가 작을수록(오른쪽으로 갈수록) 가까워진다.

이항과 달리 모수가 달라도 된다. 14개 군단 각각의 사망 수가 독립인 Poisson이면 군대 전체의 사망 수도 Poisson이다.

정리 2.5 (Poisson 나누기 (thinning)) $N \sim \mathrm{Poi}(\lambda)$이고, $N = n$일 때 각 사건이 독립적으로 확률 p로 "표시" 된다고 하자. 표시된 수 N_1과 표시되지 않은 수 $N_2 = N - N_1$은 **독립**이고 각각 $\mathrm{Poi}(\lambda p)$, $\mathrm{Poi}(\lambda q)$이다.

증명. **전략** 결합 pmf를 곱셈법칙으로 쓰고, 그것이 두 Poisson pmf의 곱으로 인수분해됨을 보인다. 인수분해가 곧 독립이다.

① $\{N_1 = a,\ N_2 = b\} = \{N = a + b,\ N_1 = a\}$이다. $N_2 = N - N_1$

② $N = n$이 주어지면 $N_1 \sim \mathrm{b}(n, p)$이므로 $P(N_1 = a, N_2 = b) = P(N = a + b)\binom{a+b}{a} p^a q^b$이다. 곱셈법칙, 1회 정리 5.3

③ pmf를 넣고 정리한다.

$$P(N_1 = a, N_2 = b) = e^{-\lambda}\frac{\lambda^{a+b}}{(a+b)!} \cdot \frac{(a+b)!}{a!\,b!} p^a q^b \qquad \text{정의 2.1}$$

$$= e^{-\lambda p}\frac{(\lambda p)^a}{a!} \cdot e^{-\lambda q}\frac{(\lambda q)^b}{b!}. \qquad e^{-\lambda} = e^{-\lambda p}e^{-\lambda q}$$

④ 결합 pmf가 a만의 함수와 b만의 함수의 곱이고, 두 인수가 각각 pmf이므로 $N_1 \sim \mathrm{Poi}(\lambda p)$, $N_2 \sim \mathrm{Poi}(\lambda q)$이고 둘은 독립이다. 3회 정리 3.2

■

놀라운 결론이다. 전체 수가 정해져 있지 않고 Poisson이면, 두 종류의 수는 서로에 대해 아무 정보도 주지 않는다. 전체가 n으로 고정된 다항분포에서는 공분산이 $-np_ip_j$였던 것과 대조된다 (정리 1.8).

2.3 Bortkiewicz의 기병대 (1898)

예제 2.6 (말에 차여 죽은 병사) Bortkiewicz는 14개 군단 20년, 곧 280개의 "군단–해" 의 사망 수를 모았다. 그 가운데 조직이 다른 네 군단(근위대와 I, VI, XI 군단)을 뺀 10개 군단 200개의 군단–해가 R. A. Fisher의 *Statistical Methods for Research Workers*(1925)를 거쳐 유명해진 자료다(`vcd::HorseKicks`).

한 해 사망 수 k	0	1	2	3	4
관측한 군단–해	109	65	22	3	1
$200 \cdot e^{-0.61} 0.61^k / k!$	108.7	66.3	20.2	4.1	0.6

평균은 $\frac{122}{200} = 0.61$이고, 그것을 λ로 넣은 Poisson의 기대 빈도가 아래 줄이다. 놀랄 만큼 맞는다. 분산은 0.611로 평균과 거의 같다(정리 2.2). 한 군단에서 한 해에 네 명 이상이 죽을 확률은 $P(X \geq 4) \approx 0.0036$, 200개 군단–해에 0.7번 정도이고, 실제로 한 번 있었다.

예제 2.7 (Student의 효모 (1907)) 맥주 회사 Guinness의 화학자 W. S. Gosset(필명 Student)은 1907년, 혈구계(작은 칸으로 나뉜 유리판)의 400칸에서 효모 세포를 세었다. 세포가 액체 속에 고르게, 서로 무관하게 퍼져 있으면 한 칸의 세포 수는 Poisson이어야 한다(작은 칸 = 많은 작은 부피 가운데 세포가 드는 드문 사건). 네 농도의 분산/평균 비는 1.19, 0.97, 1.09, 0.95로 모두 1 가까이 있다(`HistData::Yeast`). Student는 이 분포를 이용해 "몇 칸을 세야 농도를 원하는 정밀도로 알 수 있는가"를 계산했다. 2회의 Bernoulli의 질문과 같은 질문이다.

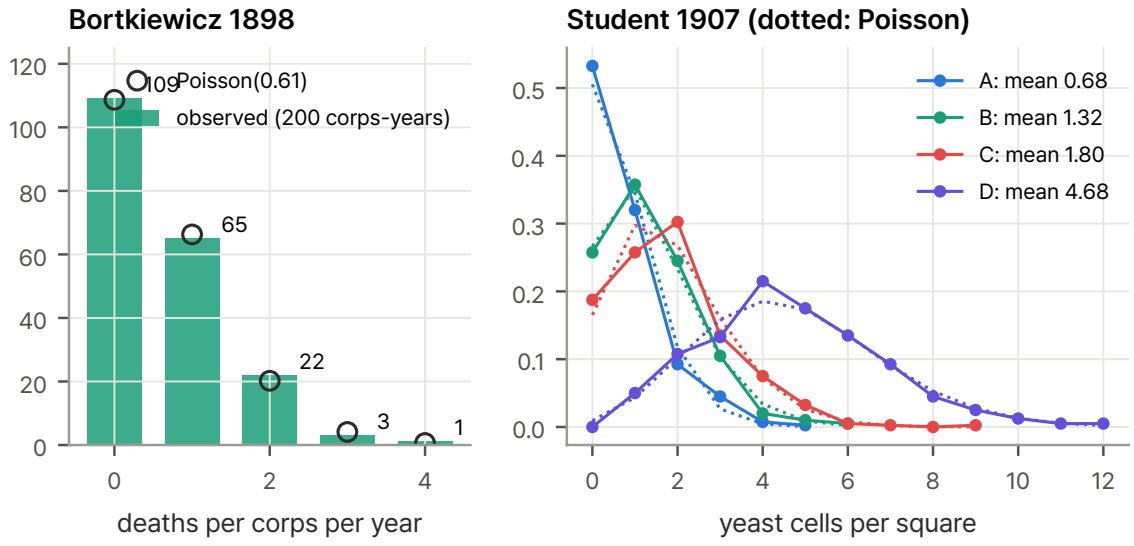


그림 2.2 왼쪽: Bortkiewicz–Fisher의 200개 군단–해(막대)와 Poi(0.61)의 기대 빈도(원). 오른쪽: Student(1907)의 효모 세포 수, 네 농도마다 400칸.

연습문제

연습 2.1 (Poisson의 최빈값 ★) $\frac{P(X=k)}{P(X=k-1)} = \frac{\lambda}{k}$을 보이고, $\mathrm{Poi}(\lambda)$의 최빈값이 $\lfloor \lambda \rfloor$임을 결론지어라(λ가 정수이면 $\lambda - 1$과 λ 둘). 풀이 41쪽

연습 2.2 (mgf로 극한을 ★★) $\mathrm{b}(n, \lambda/n)$의 mgf $\left(1 + \frac{\lambda}{n}(e^t - 1)\right)^n$이 $\exp(\lambda(e^t - 1))$로 감을 보여라. (7회에서 "mgf가 수렴하면 분포도 수렴한다"는 정리로 이것이 정리 2.3의 다른 증명이 됨을 본다.) 풀이 41쪽

연습 2.3 (오탈자 ★) 한 쪽의 오탈자 수가 $\mathrm{Poi}(0.3)$이고 쪽마다 독립이다. 300쪽짜리 책에 오탈자가 하나도 없는 쪽은 몇 쪽쯤인가? 책 전체의 오탈자 수의 분포는? 풀이 41쪽

연습 2.4 (조건부는 이항 ★★) $X_1 \sim \mathrm{Poi}(\lambda_1)$과 $X_2 \sim \mathrm{Poi}(\lambda_2)$가 독립이면, $X_1 + X_2 = n$일 때 X_1의 조건부분포가 $\mathrm{b}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$임을 보여라. 정리 2.5의 거꾸로다. 풀이 41쪽

연습 2.5 (네 명 이상 ★) 예제 2.6에서 $P(X \ge 4)$와 $P(X \ge 3)$을 $\mathrm{Poi}(0.61)$으로 계산하고, 200개 군단–해에서의 기대 횟수를 관측과 비교하라. 풀이 41쪽

연습 2.6 (Poisson 분포의 적률: 증명 ★ 유도형) $X \sim \mathrm{Poi}(\lambda)$이다. 정리 2.2를 다음 순서로 증명하라.

(a) $M(t) = \sum_k e^{tk} e^{-\lambda} \frac{\lambda^k}{k!}$을 $e^x = \sum_k x^k/k!$로 정리해 $M(t) = \exp(\lambda(e^t - 1))$을 얻어라. 이 급수는 어느 t에서 수렴하는가?

(b) 누율생성함수 $\psi = \log M$을 두 번 미분해 2회 따름정리 3.4로 평균과 분산을 얻어라.

(c) (직접 계산) $EX = \sum_k k\, e^{-\lambda} \frac{\lambda^k}{k!}$와 $E[X(X-1)]$을 첨자를 옮겨 계산하고, $\mathrm{Var}\, X = E[X(X-1)] + EX - (EX)^2$로 같은 답을 얻어라.

풀이 41쪽

연습 2.7 (독립인 Poisson의 합: 증명 ★★ 유도형) 정리 2.4을 다음 순서로 증명하라.

(a) 두 개일 때 $P(X_1 + X_2 = n) = \sum_{k=0}^{n} P(X_1 = k) P(X_2 = n - k)$에 pmf를 넣고, 이항정리(1회 정리 1.6)로 $e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}$을 얻어라.

(b) 개수 m에 대한 귀납법으로 m개의 합으로 넓혀라. $X_1 + \cdots + X_m$과 X_{m+1}이 독립이라는 것은 어디서 나오는가?

(c) mgf의 곱으로 같은 결론을 한 줄로 얻어라. (a)–(b)와 비교해 mgf의 유일성이 무엇을 대신하는지 설명하라.

풀이 42쪽

3

Poisson 과정과 감마분포

드문 사건을 **세는** 대신 **기다리면** 어떻게 되나? 시간 위에 사건이 무작위로 흩어져 있을 때, 정해진 시간 동안의 사건 수가 Poisson이면 다음 사건까지의 시간은 지수분포이고, r번째 사건까지의 시간은 감마분포다. 이 절은 그 관계를 증명하고, 감마분포의 한 구성원인 χ^2 분포를 세운다.

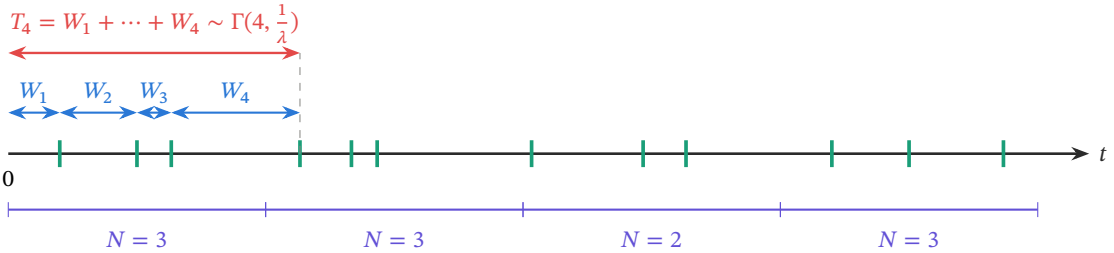
3.1 Poisson 과정

정의 3.1 (Poisson 과정) 시각 $t \geq 0$까지 일어난 사건의 수 $N(t)$가 다음을 만족하면 비율 λ의 **Poisson 과정**이라 한다.

(i) $N(0) = 0$이고, 겹치지 않는 시간 구간들의 사건 수는 독립이다.

(ii) 길이 t인 구간의 사건 수는 $\mathrm{Poi}(\lambda t)$를 따른다.

"왜 Poisson인가"는 정리 2.3가 답한다. 구간 $[0, t]$를 길이 t/n인 작은 칸 n개로 나누면, 칸마다 사건이 하나 있을 확률이 약 $\lambda t/n$이고 두 개 이상일 확률은 무시할 만하며 칸들은 독립이다. 그러면 사건 수는 $\mathrm{b}(n, \lambda t/n)$에 가깝고, $n \to \infty$에서 $\mathrm{Poi}(\lambda t)$가 된다. 방사성 붕괴(Rutherford–Geiger 1910), 전화 교환대에 걸려 오는 전화(Erlang 1909), 서버에 들어오는 요청이 이 모형을 따른다.



사건 사이의 시간 W_i는 독립인 $\mathrm{Exp}(\lambda)$이고, 길이 3인 구간마다 사건 수는 독립인 $\mathrm{Poi}(3\lambda)$이다. $P(T_r \leq t) = P(N(t) \geq r)$: r번째 사건을 t까지 기다렸다는 것은 t까지 r개 이상을 셌다는 것이다.

그림 3.1 Poisson 과정의 한 경로. 사건 사이의 시간 $W_1, W_2, \ldots$ (위)는 독립인 지수분포이고, 정해진 길이의 구간 안의 사건 수(아래)는 Poisson이다. r번째 사건의 시각 $T_r = W_1 + \cdots + W_r$은 감마분포다.

정리 3.2 (기다리는 시간) 비율 λ의 Poisson 과정에서 r번째 사건의 시각을 T_r이라 하면

$$P(T_r \le t) = P\big(N(t) \ge r\big) = 1 - \sum_{k=0}^{r-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \qquad f_{T_r}(t) = \frac{\lambda^r t^{r-1}}{(r-1)!}e^{-\lambda t} \quad (t>0).$$

특히 첫 사건까지의 시간 T_1은 $\mathrm{Exp}(\lambda)$이다.

증명. **전략** "r번째 사건이 t까지 일어났다"를 "t까지의 사건 수가 r 이상"으로 바꿔 Poisson 확률로 쓴다. 밀도는 그 분포함수를 미분해 얻는다.

① $\{T_r \le t\} = \{N(t) \ge r\}$이다. r번째 사건이 t까지 일어났다는 것과 t까지 사건이 r개 이상이라는 것은 같은 결과들의 모임이다. 원소 추적 [T1]

② $P(N(t)\ge r) = 1 - \sum_{k=0}^{r-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!}$이다. 정의 3.1(ii), 여사건

③ ②를 t로 미분하면 각 항의 도함수 $\lambda e^{-\lambda t}\big[\frac{(\lambda t)^{k-1}}{(k-1)!} - \frac{(\lambda t)^k}{k!}\big]$에서 이웃 항이 소거되어 $f_{T_r}(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{r-1}}{(r-1)!}$만 남는다. $k=0$인 항에는 첫째 항이 없다. 망원 합 [T4], 연습 3.1

④ $r=1$이면 $P(T_1 \le t) = 1 - e^{-\lambda t}$이므로 $T_1 \sim \mathrm{Exp}(\lambda)$이다. ②

■

3회 예제 5.7에서 독립인 지수분포 n개의 합이 같은 밀도를 가졌다. 이것은 우연이 아니다. 사건 사이의 시간들이 독립인 $\mathrm{Exp}(\lambda)$이고(연습 3.2), 그 합이 T_r이다. 세는 것(Poisson)과 기다리는 것(감마)은 같은 과정의 두 얼굴이다.

정리 3.3 (지수분포의 무기억성) $X \sim \mathrm{Exp}(\lambda)$이면 모든 $s, t > 0$에 대해

$$P(X > s+t \mid X > s) = P(X > t).$$

거꾸로, 연속이고 양수인 확률변수가 이 성질을 가지면 지수분포다.

증명. **전략** 정방향은 생존함수 $G(x) = P(X>x)$를 넣어 계산한다. 역방향은 $G(s+t) = G(s)G(t)$를 자연수, 유리수, 실수의 순서로 풀고, 실수로 넓힐 때 오른쪽 연속성을 쓴다.

정방향

① $G(x) = e^{-\lambda x}$이므로 $P(X>s+t \mid X>s) = \frac{G(s+t)}{G(s)} = e^{-\lambda t} = G(t)$이다. 조건부확률의 정의, $\{X>s+t\} \subset \{X>s\}$

역방향

① 조건부확률이 정의되려면 모든 $s>0$에서 $G(s)>0$이고, 성질은 $G(s+t)=G(s)G(t)$ $(s,t>0)$와 같다. 조건부확률의 정의

② $c=G(1)$로 두면 자연수 n에서 $G(n)=c^n$이다. ①, 귀납 [T3]

③ $0<c<1$이다. $c>0$은 ①이다. $c=1$이면 ②로 모든 n에서 $G(n)=1$인데, X는 실숫값이므로 $G(n)\to 0$이어야 한다. 1회 정리 6.3(b), 귀류 [T2]

④ 자연수 m에서 $c=G(1)=G(\tfrac{1}{m})^m$이므로 $G(\tfrac{1}{m})=c^{1/m}$이고, ②와 같은 귀납으로 $G(\tfrac{n}{m})=c^{n/m}$이다. 곧 양의 유리수 q에서 $G(q)=c^q$이다. ①, [T3]

⑤ 실수 $x>0$이면 $q_k\downarrow x$인 유리수열을 잡는다. $G=1-F$는 오른쪽 연속이므로 $G(x)=\lim_k c^{q_k}=c^x$이다. 1회 정리 6.3(c), [T5]

⑥ $\lambda=-\log c$로 두면 ③으로 $\lambda>0$이고, $P(X>x)=e^{-\lambda x}$이므로 $X\sim\mathrm{Exp}(\lambda)$이다. ③⑤

■

이미 s시간을 기다렸어도 앞으로 기다릴 시간의 분포는 처음과 같다. 사건이 "밀린다"는 생각이 틀리는 이유다. 버스가 Poisson 과정으로 온다면, 정류장에 늦게 왔다고 버스가 곧 온다는 보장은 없다. 이산판은 기하분포다(연습 3.3).

3.2 감마함수와 감마분포

정리 3.2의 밀도에서 $(r-1)!$을 실수로 넓히면 감마분포가 된다.

정의 3.4 (감마함수) $\alpha>0$에 대해

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,dx.$$

정리 3.5 (감마함수의 성질) (a) $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$. (b) 자연수 n에 대해 $\Gamma(n)=(n-1)!$. (c) $\Gamma(\tfrac{1}{2})=\sqrt{\pi}$.

증명. **전략** (a)는 부분적분이고 경계항이 사라지는 것을 확인한다. (b)는 (a)의 귀납 [T3], (c)는 치환으로 가우스 적분을 만든다 [T7]. 단계별 증명은 연습 3.7(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 43쪽에 있다.

정의 3.6 (감마분포) $\alpha, \beta > 0$에 대해 밀도

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \qquad x > 0$$

를 갖는 분포를 **감마분포** $\Gamma(\alpha, \beta)$라 한다. α는 모양, β는 척도 모수다.

$x = \beta y$로 치환하면 적분이 $\Gamma(\alpha)$의 정의가 되어 1이다. $\Gamma(1, \beta)$는 평균 β의 지수분포이고, 정리 3.2의 T_r은 $\Gamma(r, 1/\lambda)$이다. 그림 3.2처럼 $\alpha < 1$이면 0에서 밀도가 무한대이고, α가 커지면 종 모양이 된다.

정리 3.7 (감마분포의 mgf와 적률) $X \sim \Gamma(\alpha, \beta)$이면 $t < 1/\beta$에서 $M(t) = (1 - \beta t)^{-\alpha}$이고, $EX = \alpha\beta$, $\operatorname{Var} X = \alpha\beta^2$. 독립인 $X_i \sim \Gamma(\alpha_i, \beta)$(같은 β)의 합은 $\Gamma(\sum \alpha_i, \beta)$이다.

증명. **전략** mgf의 적분을 치환으로 감마함수의 정의로 바꾼다. 적률은 누율생성함수의 미분, 합은 mgf의 곱과 유일성이다. 단계별 증명은 연습 3.8(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 43쪽에 있다.

3.3 χ^2 분포

정의 3.8 (χ^2 분포) $r > 0$에 대해 $\Gamma(\frac{r}{2}, 2)$를 자유도 r인 **카이제곱 분포** $\chi^2(r)$이라 한다. 평균은 r, 분산은 $2r$, mgf는 $(1 - 2t)^{-r/2}$이다.

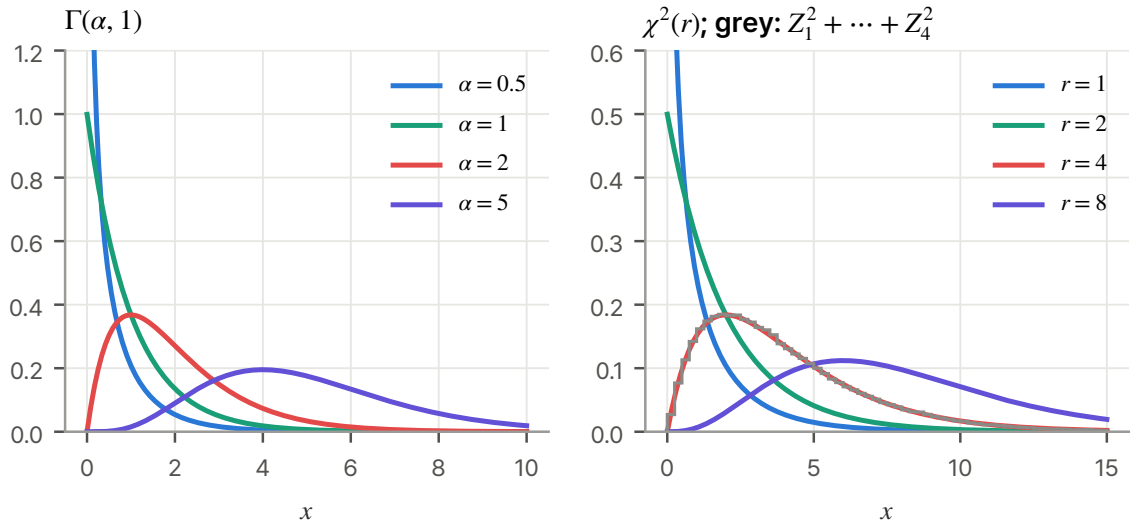


그림 3.2 왼쪽: 감마 밀도 $\Gamma(\alpha, 1)$. $\alpha < 1$이면 0에서 무한대, $\alpha = 1$이면 지수, α가 커지면 종 모양이 된다. 오른쪽: $\chi^2(r)$ 밀도.

정리 3.9 (표준정규의 제곱합) $Z \sim \mathcal{N}(0,1)$이면 $Z^2 \sim \chi^2(1)$이다. 따라서 $Z_1, \dots, Z_r \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$이면 $\sum_{i=1}^{r} Z_i^2 \sim \chi^2(r)$.

증명. **전략** Z^2의 분포함수를 Φ로 쓰고 미분해 $\Gamma(\frac{1}{2}, 2)$의 밀도를 알아본다. 합은 감마의 합이다. 단계별 증명은 연습 3.9(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 43쪽에 있다.

3회 연습 5.6에서 Box–Muller의 $R^2 = Z_1^2 + Z_2^2$이 평균 2인 지수분포였다. 그것이 $\chi^2(2) = \Gamma(1, 2)$다. χ^2 분포는 6회(Student의 t)와 8회(적합도 검정)의 중심이다. 자유도 r은 "독립인 표준정규를 몇 개 제곱해 더했는가"로 읽는다.

연습문제

연습 3.1 (망원급수 ★) 정리 3.2의 증명에서 $\frac{d}{dt}\sum_{k=0}^{r-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!} = -\lambda e^{-\lambda t}\frac{(\lambda t)^{r-1}}{(r-1)!}$ 임을 보여라. 풀이 42쪽

연습 3.2 (사건 사이의 시간 ★★) $W_1 = T_1$, $W_2 = T_2 - T_1$이다. 정의 3.1의 (i)–(ii)에서 $P(W_1 > s,\ W_2 > t) = e^{-\lambda s}e^{-\lambda t}$ 임을 보여라. (힌트: W_1이 s 근처일 때 $(s, s+t]$에 사건이 없을 확률. 엄밀한 증명은 조건부 밀도를 쓴다.) 풀이 42쪽

연습 3.3 (기하분포의 무기억성 ★) 실패 수 G가 기하분포이면 $P(G \ge s+t \mid G \ge s) = P(G \ge t)$ (정수 $s, t \ge 0$)임을 보여라. 풀이 42쪽

연습 3.4 (감마의 적률 ★) $X \sim \Gamma(\alpha, \beta)$이면 $EX^k = \beta^k \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}$ 임을 적분으로 보이고 $k = 1, 2$로 평균과 분산을 확인하라. $k > -\alpha$인 실수 k에서도 성립한다. 풀이 42쪽

연습 3.5 (기다림과 셈 ★) 서버에 요청이 분당 평균 3개씩 Poisson 과정으로 들어온다. 5번째 요청이 2분 안에 올 확률을 (a) 감마분포로, (b) Poisson 분포로 구하라. 같은 값인가? 풀이 42쪽

연습 3.6 ($\Gamma(\frac{3}{2})$와 반정수 ★) $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$, $\Gamma(\frac{5}{2}) = \frac{3\sqrt{\pi}}{4}$를 보이고, $\chi^2(3)$의 밀도를 쓰라. 이것은 3차원 표준정규 벡터의 길이의 제곱의 분포이며, Maxwell(1860)이 기체 분자의 속력 분포를 유도한 것과 같은 계산이다. 풀이 43쪽

연습 3.7 (감마함수의 성질: 증명 ★★ 유도형) 정리 3.5를 다음 순서로 증명하라.

(a) $\alpha > 0$이면 $\Gamma(\alpha)$의 적분이 수렴함을 보여라. 0 근처에서는 $x^{\alpha-1}e^{-x} \le x^{\alpha-1}$, 무한대 근처에서는 $x^{\alpha-1}e^{-x} \le e^{-x/2}$ (충분히 큰 x)를 쓴다.

(b) $0 < \varepsilon < M$에서 $\int_\varepsilon^M x^\alpha e^{-x}dx$를 부분적분하고, $\varepsilon \to 0$, $M \to \infty$에서 경계항 $-x^\alpha e^{-x}$이 사라짐을 보여 정리의 (a) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$를 얻어라.

(c) $\Gamma(1) = 1$을 계산하고 n에 대한 귀납법으로 $\Gamma(n) = (n-1)!$을 보여라.

(d) $x = z^2/2$로 치환해 $\Gamma(\frac{1}{2}) = \sqrt{2}\int_0^\infty e^{-z^2/2}dz$를 보이고, 3회 정리 5.3으로 $\sqrt{\pi}$를 얻어라.

풀이 43쪽

연습 3.8 (감마분포의 mgf와 적률: 증명 ★ 유도형) $X \sim \Gamma(\alpha, \beta)$이다. 정리 3.7를 다음 순서로 증명하라.

(a) $t < 1/\beta$이면 $M(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-x(1/\beta - t)}dx$에서 $y = x(1/\beta - t)$로 치환해 $M(t) = (1-\beta t)^{-\alpha}$를 얻어라. $t \geq 1/\beta$이면 무엇이 문제인가?

(b) 누율생성함수 $\psi(t) = -\alpha\log(1-\beta t)$를 두 번 미분해 평균과 분산을 얻어라.

(c) 같은 β의 독립인 감마의 합에 대해 mgf의 곱과 유일성으로 결론을 얻어라. β가 다르면 무엇이 깨지는가?

풀이 43쪽

연습 3.9 (표준정규의 제곱합: 증명 ★★ 유도형) 정리 3.9을 다음 순서로 증명하라.

(a) $y > 0$이면 $P(Z^2 \leq y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$임을 보이고, 미분해 $f_{Z^2}(y) = \frac{\phi(\sqrt{y}) + \phi(-\sqrt{y})}{2\sqrt{y}}$를 얻어라.

(b) 정리 3.5(c)를 써서 (a)가 $\Gamma(\frac{1}{2}, 2)$의 밀도임을 확인하라.

(c) 독립인 Z_i^2들이 같은 척도 $\beta = 2$의 감마이므로 정리 3.7로 합의 분포를 얻어라. Z_i^2들이 독립이라는 것은 어디서 나오는가?

풀이 43쪽

4

베타분포 — Bayes의 당구대

지금까지 성공 확률 p는 알려진 수였다. 모른다면? 18세기의 Bayes와 Laplace는 p 자체에 확률분포를 주고, 관측한 성공 수로 그 분포를 고쳤다. 그 분포가 베타분포이고, 베타분포는 두 감마분포의 비에서 나온다.

4.1 베타함수와 베타분포

정의 4.1 (베타분포) $a, b > 0$에 대해 밀도

$$f(x) = \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1, \qquad B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

를 갖는 분포를 **베타분포** $\mathrm{Beta}(a,b)$라 하고 B를 **베타함수**라 한다.

$\mathrm{Beta}(1,1)$은 $\mathrm{U}(0,1)$이다. a, b가 크면 $\frac{a}{a+b}$ 근처에 몰리고, $a = b < 1$이면 양 끝에 몰리는 U자 모양이 된다(그림 4.2).

정리 4.2 (베타와 감마) $X \sim \Gamma(a,1)$과 $Y \sim \Gamma(b,1)$이 독립이면

$$U = \frac{X}{X+Y} \sim \mathrm{Beta}(a,b), \qquad V = X + Y \sim \Gamma(a+b,1),$$

이고 U와 V는 **독립**이다. 따라서 $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

증명. **전략** $(X,Y) \mapsto (U,V)$의 변수변환으로 결합밀도를 구하고 [T7], 그것이 u만의 함수와 v만의 함수의 곱으로 인수분해됨을 보인다. 인수분해에서 독립과 두 주변분포, 그리고 정규화 상수가 함께 나온다.

① 변환 $(x,y) \mapsto (u,v) = \left(\frac{x}{x+y},\, x+y\right)$는 $(0,\infty)^2$에서 $(0,1)\times(0,\infty)$로 가는 일대일 대응이고, 역변환은 $x = uv$, $y = (1-u)v$이다. 대입해 확인

② 역변환의 야코비안은 $\det\begin{pmatrix} v & u \\ -v & 1-u \end{pmatrix} = v(1-u) + uv = v$이다. [T7]

③ X와 Y가 독립이므로 $f_{X,Y}(x,y) = \frac{x^{a-1}e^{-x}}{\Gamma(a)} \cdot \frac{y^{b-1}e^{-y}}{\Gamma(b)}$ $(x,y>0)$이다. 3회 정리 3.2

④ 변수변환 공식에 ①–③을 넣고 u와 v로 나눠 묶는다($0<u<1, v>0$).

$$f_{U,V}(u,v) = f_{X,Y}\big(uv,(1-u)v\big)\cdot v \qquad \text{3회 정리 5.1}$$
$$= \frac{u^{a-1}(1-u)^{b-1}}{\Gamma(a)\Gamma(b)/\Gamma(a+b)}\cdot\frac{v^{a+b-1}e^{-v}}{\Gamma(a+b)}. \qquad e^{-uv}e^{-(1-u)v}=e^{-v}$$

⑤ 결합밀도가 u만의 함수와 v만의 함수의 곱이므로 U와 V는 독립이다. 둘째 인수는 $\Gamma(a+b,1)$의 밀도다. 3회 정리 3.2, 정의 3.6

⑥ v로 적분하면 둘째 인수의 적분은 1이므로, 첫째 인수가 U의 밀도이고 그 적분도 1이다. 따라서 $\int_0^1 u^{a-1}(1-u)^{b-1}du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, 곧 $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$이고 $U\sim\text{Beta}(a,b)$이다. 밀도의 적분은 1

■

따름정리 4.3 (베타분포의 적률) $U\sim\text{Beta}(a,b)$이면

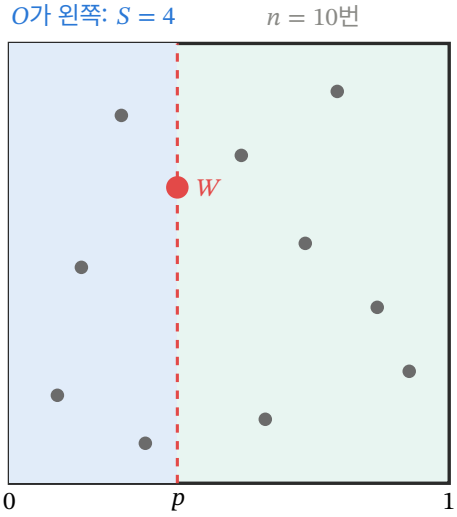
$$EU=\frac{a}{a+b},\qquad \text{Var}\,U=\frac{ab}{(a+b)^2(a+b+1)}.$$

증명. $EU^k=\frac{B(a+k,b)}{B(a,b)}=\frac{\Gamma(a+k)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+k)}$이고 $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$를 쓴다. $k=1$이면 $\frac{a}{a+b}$, $k=2$이면 $\frac{a(a+1)}{(a+b)(a+b+1)}$. ■

예제 4.4 (균등분포의 순서통계량) $U_1,\dots,U_n \overset{\text{iid}}{\sim} \text{U}(0,1)$을 크기순으로 늘어놓은 k번째 값 $U_{(k)}$는 $\text{Beta}(k,n-k+1)$이다. $U_{(k)}\le x$일 확률은 n개 가운데 k개 이상이 x 이하일 확률 $\sum_{j\ge k}\binom{n}{j}x^j(1-x)^{n-j}$이고, 미분하면 망원급수로 $\frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k}$이 남는다(연습 4.1). $k=n$이 3회 예제 3.5의 최댓값이다. 평균은 $\frac{k}{n+1}$이다. n개의 점이 구간을 평균적으로 $n+1$등분한다.

4.2 Bayes의 당구대 (1763)

역사 Thomas Bayes의 유고 *An Essay towards Solving a Problem in the Doctrine of Chances*(1763, Richard Price가 정리)는 이렇게 묻는다. 정사각형 당구대에 공 W를 굴려 멈춘 위치의 가로 좌표를 p라 하자(우리는 p를 보지 못한다). 이어서 공 O를 n번 굴려, O가 W보다 왼쪽에 멈춘 횟수 S만 듣는다. $S=s$일 때 p가 a와 b 사이에 있을 확률은? Bayes는 기하학적 논증으로 답했다. 오늘의 말로 그것은 균등 사전분포와 이항 가능도에서 베타 사후분포를 얻는 계산이다.



W를 굴려 멈춘 가로 좌표 $p \sim \mathrm{U}(0,1)$은 보이지 않는다.

O를 n번 굴려 W의 왼쪽에 멈춘 횟수 S만 듣는다. p가 주어지면 $S \sim \mathrm{b}(n,p)$.

Bayes의 답: $p \mid S=s \sim \mathrm{Beta}(s+1, n-s+1)$ (정리 4.5).

다음 O가 왼쪽일 확률은 $\frac{s+1}{n+2}$ (Laplace의 계승규칙).

그림 4.1 Bayes의 당구대. 공 W의 위치 p는 숨겨져 있고, 공 O를 n번 굴려 W의 왼쪽에 멈춘 횟수 S만 안다.

정리 4.5 (Bayes의 사후분포) $p \sim \mathrm{Beta}(a,b)$이고 p가 주어지면 $S \sim \mathrm{b}(n,p)$일 때, $S=s$가 주어진 p의 조건부분포는

$$p \mid S=s \sim \mathrm{Beta}(a+s,\ b+n-s).$$

특히 균등 사전분포($a=b=1$, Bayes의 당구대)이면 $\mathrm{Beta}(s+1, n-s+1)$이다.

증명. **전략** 결합 "밀도–질량" $f(p)P(S=s \mid p)$를 s를 고정한 p의 함수로 보고 정규화한다. p에 대해 비례하는 꼴만 맞으면 정규화 상수를 따로 계산하지 않아도 된다. 단계별 증명은 연습 4.6(유도형)에서 직접 쓴다. 완성된 증명은 부록 A의 풀이 44쪽에 있다.

사전분포와 사후분포가 같은 가족이다. 베타는 이항의 **켤레 사전분포**다. 성공 하나는 a에, 실패 하나는 b에 1을 더한다. a와 b를 "사전에 본 것과 같은 가상의 성공·실패 수"로 읽을 수 있다.

따름정리 4.6 (Laplace의 계승규칙 (1774)) 균등 사전분포에서 n번 가운데 s번 성공했다면 다음 시행이 성공할 확률은

$$P(\text{다음 성공} \mid S=s) = E(p \mid S=s) = \frac{s+1}{n+2}.$$

증명. p가 주어지면 다음 시행은 확률 p로 성공하므로, 반복기댓값(3회 정리 2.5)으로 조건부 확률은 $E(p \mid S=s)$이고, 따름정리 4.3로 $\frac{s+1}{(s+1)+(n-s+1)}$. ∎

Laplace(1814)는 이것으로 "해가 내일도 뜰 확률"을 계산했다. 역사가 5,000년, 곧 1,826,213일 동안 매일 해가 떴다면, 내일도 뜰 확률은 $\frac{1{,}826{,}214}{1{,}826{,}215}$, "1,826,214 대 1"이다. 그는 곧이어, 자연의

법칙을 아는 사람에게는 이 확률이 훨씬 크다고 덧붙였다. 사전분포가 모든 것을 정한다는 비판은 이때부터 있었다(16회).

연습문제

연습 4.1 (순서통계량의 밀도 ★★) 예제 4.4의 미분을 실행해 $U_{(k)}$의 밀도를 유도하라. 또 정리 4.2를 쓰지 않고 $B(k, n-k+1) = \frac{(k-1)!(n-k)!}{n!}$임을 이 결과에서 얻어라. 풀이 43쪽

연습 4.2 (균등 사전분포의 예측 ★★) Bayes의 당구대에서 $p \sim \mathrm{U}(0,1)$, $S \mid p \sim \mathrm{b}(n,p)$이면 $P(S=s) = \frac{1}{n+1}$ $(s = 0, \dots, n)$임을 보여라. 공 O를 굴리기 전에는 모든 성공 수가 똑같이 그럴듯하다. 풀이 44쪽

연습 4.3 (사후분포의 평균과 분산 ★) $\mathrm{Beta}(1,1)$ 사전분포에서 20번 가운데 6번 성공했다. 사후평균과 사후 표준편차를 구하고, $\hat{p} = 0.3$과 그 표준오차 $\sqrt{\hat{p}(1-\hat{p})/n}$과 비교하라. 풀이 44쪽

연습 4.4 (F 분포의 씨앗 ★★) 정리 4.2의 X, Y로 $R = X/Y$의 밀도를 구하라. $\frac{x^{a-1}}{B(a,b)(1+x)^{a+b}}$ $(x>0)$이다. 6회에서 χ^2의 비로 F 분포를 만들 때 같은 계산을 쓴다. 풀이 44쪽

연습 4.5 (Dirichlet로 가는 길 ★★★) X_1, X_2, X_3이 독립이고 $X_j \sim \Gamma(a_j, 1)$이면 $\left(\frac{X_1}{S}, \frac{X_2}{S}\right)$, $S = \sum X_j$의 결합밀도가 $\frac{\Gamma(a_1+a_2+a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} u_1^{a_1-1} u_2^{a_2-1} (1-u_1-u_2)^{a_3-1}$임을 보여라. 이것이 다항분포의 켤레 사전분포인 Dirichlet 분포다(16회, 주제 모형). 풀이 44쪽

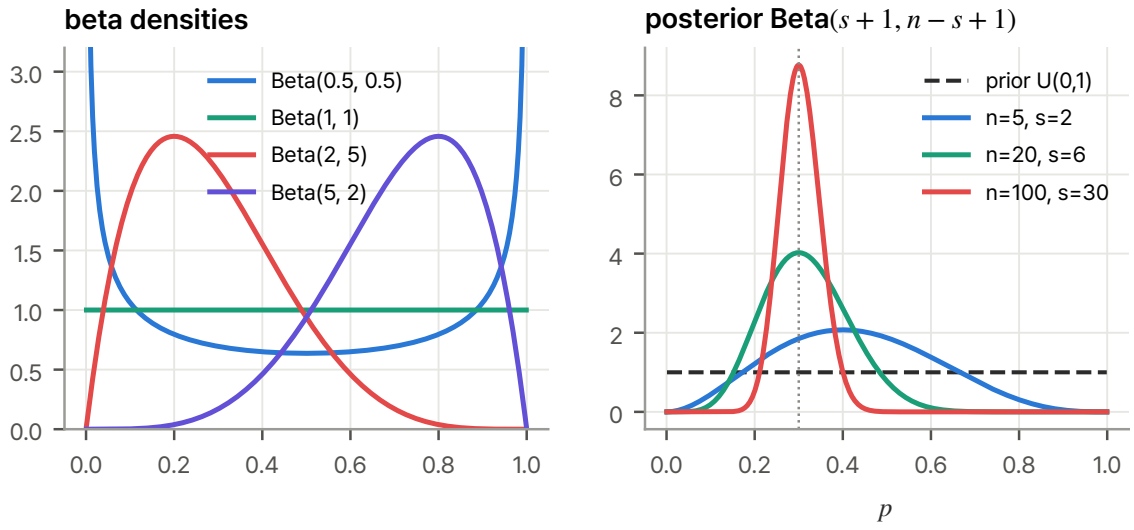


그림 4.2 왼쪽: 베타 밀도의 여러 모양. 오른쪽: Bayes의 당구대. 균등 사전분포(점선)에서 시작해 $n = 5, 20, 100$번 가운데 성공 비율 0.3을 관측했을 때의 사후분포 $\mathrm{Beta}(s+1, n-s+1)$. 관측이 늘수록 참값 근처로 좁아진다.

연습 4.6 (Bayes의 사후분포: 증명 ★★ 유도형) $p \sim \mathrm{Beta}(a,b)$이고 p가 주어지면 $S \sim \mathrm{b}(n,p)$이다. 정리 4.5를 다음 순서로 증명하라.

(a) 결합 "밀도–질량" $h(p,s) = f(p)P(S=s \mid p)$를 쓰라.

(b) p로 적분해 $P(S=s) = \binom{n}{s}\frac{B(a+s,\,b+n-s)}{B(a,b)}$를 얻어라.

(c) 조건부 밀도 $h(p,s)/P(S=s)$가 $\mathrm{Beta}(a+s,\,b+n-s)$의 밀도와 같음을 확인하라.

(d) (b)의 계산 없이 "p에 대해 $p^{a+s-1}(1-p)^{b+n-s-1}$에 비례한다"는 것만으로 같은 결론이 나오는 이유를 설명하라.

풀이 44쪽

5

혼합분포 — 섞인 집단

모든 관측이 같은 분포에서 나오지 않을 수 있다. 군단마다 위험이 다르고, 저자마다 단어 습관이 다르고, 게의 무리마다 체형이 다르다. 관측값이 "먼저 집단을 무작위로 고르고, 그 집단의 분포에서 뽑는다"는 두 단계로 만들어지면 그 분포를 혼합분포라 한다. 3회의 전분산 공식이 혼합의 평균과 분산을 한 줄로 준다.

5.1 정의와 적률

정의 5.1 (혼합분포) 잠재변수 Θ가 분포 G를 따르고, $\Theta=\theta$일 때 X가 밀도(또는 pmf) $f(x\mid\theta)$를 따르면, X의 분포

$$f_X(x)=\sum_{k}\pi_k f(x\mid\theta_k)\quad\text{(유한 혼합)},\qquad f_X(x)=\int f(x\mid\theta)\,g(\theta)\,d\theta\quad\text{(연속 혼합)}$$

를 **혼합분포**라 한다. $\pi_k=P(\Theta=\theta_k)$를 혼합 비율이라 한다.

식은 X와 Θ의 결합분포에서 Θ를 합하거나 적분해 없앤 주변분포(3회 정리 1.6)다.

정리 5.2 (혼합의 평균과 분산) $\mu(\theta)=E(X\mid\Theta=\theta)$, $\sigma^2(\theta)=\mathrm{Var}(X\mid\Theta=\theta)$이면

$$EX=E\,\mu(\Theta),\qquad \mathrm{Var}\,X=\underbrace{E\,\sigma^2(\Theta)}_{\text{집단 안}}+\underbrace{\mathrm{Var}\,\mu(\Theta)}_{\text{집단 사이}}.$$

증명. 3회 정리 2.5와 2.6을 X와 Θ에 쓴다. ∎

혼합의 분산은 언제나 집단 안의 평균 분산보다 크다. 집단 사이의 차이가 더해지기 때문이다. 5회 5.3절에서 Galton의 아들과 딸의 키를 섞었을 때 분산의 절반 가까이가 "집단 사이" 항이었던 것이 이 정리다.

5.2 Poisson–감마 혼합은 음이항

정리 5.3 (Poisson–감마 혼합) $\Lambda\sim\Gamma(r,\beta)$이고 $\Lambda=\lambda$일 때 $X\sim\mathrm{Poi}(\lambda)$이면

$$X\sim\mathrm{NB}\Big(r,\ p=\frac{1}{1+\beta}\Big),\qquad EX=r\beta,\qquad \mathrm{Var}\,X=r\beta+r\beta^2.$$

r은 정수가 아니어도 된다.

증명. **전략** 주변 pmf를 혼합의 정의대로 적분으로 쓰고, 그 적분을 감마함수로 알아본다. 적률은 혼합의 평균과 분산 공식으로 구한다.

① $k=0,1,\dots$에 대해 혼합의 정의로 쓰고 λ의 거듭제곱과 지수를 모은다.

$$P(X=k)=\int_0^\infty e^{-\lambda}\frac{\lambda^k}{k!}\cdot\frac{\lambda^{r-1}e^{-\lambda/\beta}}{\Gamma(r)\beta^r}\,d\lambda \qquad \text{정의 5.1}$$
$$=\frac{1}{k!\,\Gamma(r)\beta^r}\int_0^\infty \lambda^{k+r-1}e^{-\lambda(1+1/\beta)}\,d\lambda.$$

② $y=\lambda(1+1/\beta)$로 치환하면 적분은 $\Gamma(k+r)\big(\frac{\beta}{1+\beta}\big)^{k+r}$이다. 정의 3.4, 치환 [T7]

③ ①②에서 $P(X=k)=\frac{\Gamma(k+r)}{k!\,\Gamma(r)}\big(\frac{1}{1+\beta}\big)^r\big(\frac{\beta}{1+\beta}\big)^k$이다. 정리

④ 정수 r이면 $\frac{\Gamma(k+r)}{k!\,\Gamma(r)}=\binom{k+r-1}{k}$이므로, 이것은 정리 1.3의 pmf에 $p=\frac{1}{1+\beta}$, $q=\frac{\beta}{1+\beta}$를 넣은 것이다. 정수가 아닌 r에서는 1.2절처럼 이 식이 $\mathrm{NB}(r,p)$의 정의다. 정리 3.5(a)(b)

⑤ 적률은 $EX=E\Lambda=r\beta$, $\mathrm{Var}\,X=E\Lambda+\mathrm{Var}\,\Lambda=r\beta+r\beta^2$이다. 정리 5.2, 2.2, 3.7

조건 확인 혼합의 pmf가 합이 1인 것은 음이 아닌 항의 합과 적분의 순서를 바꿔 $\sum_k\int=\int\sum_k=\int g=1$로 확인한다(Tonelli [T6]). ②의 적분은 $k+r>0$이므로 수렴한다. ∎

이것이 음이항분포가 셈 자료의 표준 모형이 된 이유다. 개체마다 위험률 λ가 다르고(감마로 흩어져 있고), 주어진 위험률에서는 Poisson이라는 그림이다. 분산 $r\beta + r\beta^2 = EX + \frac{(EX)^2}{r}$에서 둘째 항이 위험률의 흩어짐이 만든 **과산포**다. $r \to \infty$, $r\beta = \lambda$로 고정하면 감마가 한 점에 몰리고 음이항은 Poisson으로 돌아간다. Greenwood와 Yule(1920)은 이 모형으로 공장 노동자의 사고 횟수가 Poisson보다 더 흩어져 있다는 것, 곧 "사고를 잘 당하는 사람"이 있다는 것을 보였다.

5.3 정규의 혼합 — Pearson의 게 (1894)

> **역사** 동물학자 Weldon은 나폴리 만의 게 1,000마리의 이마 폭을 쟀다. 분포가 약간 비대칭이었고, 그는 두 아종이 섞여 있는 것이 아닌지 Karl Pearson에게 물었다. Pearson(1894)은 두 정규분포의 혼합 $\pi\mathcal{N}(\mu_1, \sigma_1^2) + (1-\pi)\mathcal{N}(\mu_2, \sigma_2^2)$의 모수 다섯 개를, 자료의 처음 다섯 적률을 이 모형의 적률과 같게 놓아 풀었다. 그 결과는 9차 방정식이었고, Pearson은 그것을 손으로 풀었다. 이것이 **적률법**의 시작이다. 9회에서 최대가능도와 비교하고, 10회의 EM 알고리즘이 같은 문제를 다른 방법으로 푼다.

유한 정규 혼합의 평균과 분산은 정리 5.2로

$$EX = \sum_k \pi_k \mu_k, \qquad \operatorname{Var} X = \sum_k \pi_k \sigma_k^2 + \sum_k \pi_k (\mu_k - EX)^2.$$

평균이 같고 분산만 다른 정규를 섞으면($\mu_k = \mu$) 한 봉우리이지만 꼬리가 정규보다 무거워진다(그림 5.1 오른쪽). 5회 5.2절의 "오염된 정규"가 이것이다.

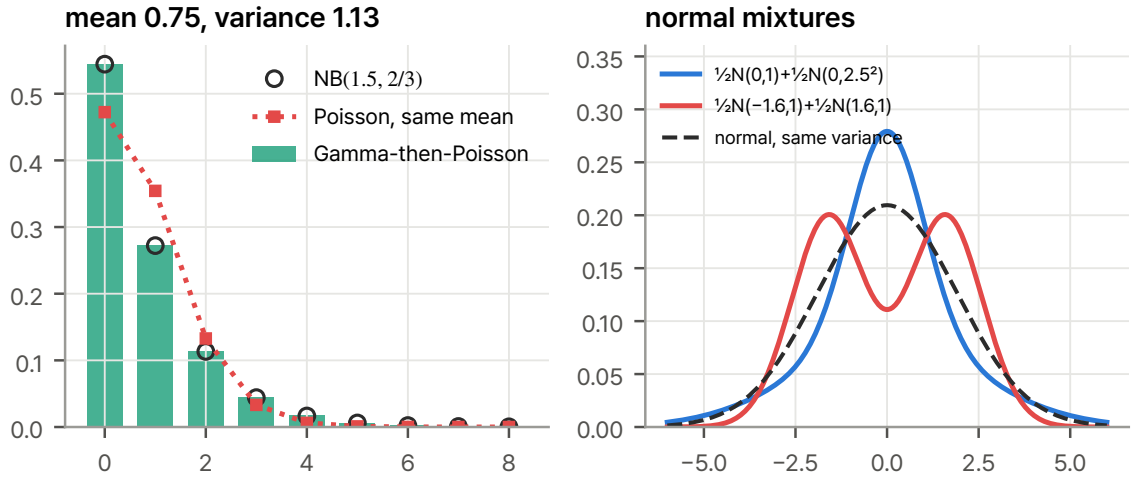


그림 5.1 왼쪽: 위험률 $\Lambda \sim \Gamma(r = 1.5, \beta = 0.5)$를 먼저 뽑고 Poisson을 뽑은 시뮬레이션(막대)과 NB$(1.5, \frac{2}{3})$(원), 같은 평균의 Poisson(점선). 오른쪽: 정규분포 두 개의 혼합. 한 봉우리이면서도 정규보다 꼬리가 무겁거나, 두 봉우리일 수 있다.

예제 5.4 (Laplace의 오차곡선은 정규의 혼합) Laplace(1774)는 관측 오차의 분포로 이중지수 밀도 $f(x)=\frac{1}{2b}e^{-|x|/b}$를 제안했다(Laplace 분포). 분산이 $V\sim$ Exp(평균 $2b^2$)로 흩어져 있고 $V=v$일 때 $X\sim\mathcal{N}(0,v)$이면 X가 바로 이 분포다(연습 5.3). 오차의 크기가 관측마다 다른 정밀도로 흩어져 있다는 그림이다. 13회의 최소절대편차 회귀가 이 분포의 최대가능도이고, 15회의 로버스트 방법들도 같은 생각에서 나온다.

연습문제

연습 5.1 (두 집단 ★) 남자의 키가 $\mathcal{N}(175,7^2)$, 여자가 $\mathcal{N}(162,6^2)$이고 비율이 반반이다. 섞은 키의 평균과 분산을 구하고, 분산에서 "집단 사이" 항이 차지하는 비율을 구하라. 풀이 44쪽

연습 5.2 (음이항의 과산포 ★) 정리 5.3에서 분산/평균 비가 $1+\beta$임을 보여라. 평균이 0.66이고 분산/평균 비가 1.54인 자료라면 β와 r은 얼마인가? 풀이 44쪽

연습 5.3 (Laplace는 정규의 혼합 ★★★) $V\sim$ Exp(평균 $2b^2$)이고 $V=v$일 때 $X\sim\mathcal{N}(0,v)$이면 X의 mgf가 $\frac{1}{1-b^2t^2}$ $(|t|<1/b)$임을 보여라. (힌트: $E(e^{tX}\mid V)=e^{t^2V/2}$이고 반복기댓값.) 이것이 Laplace 분포의 mgf임을 확인하고 mgf의 유일성으로 결론지어라. 풀이 45쪽

연습 5.4 (베타–이항 ★★) $p\sim\text{Beta}(a,b)$이고 p가 주어지면 $S\sim\text{b}(n,p)$일 때 $ES=\frac{na}{a+b}$, $\operatorname{Var}S=n\bar{p}(1-\bar{p})\frac{a+b+n}{a+b+1}$, $\bar{p}=\frac{a}{a+b}$임을 정리 5.2로 보여라. 이항보다 $\frac{a+b+n}{a+b+1}$배 흩어진다. 풀이 45쪽

연습 5.5 (혼합의 mgf ★) 유한 혼합 $f=\sum\pi_k f_k$의 mgf가 $\sum\pi_k M_k(t)$임을 보이고, 독립합의 mgf(곱)와 혼동하지 말아야 할 이유를 설명하라. 풀이 45쪽

6

데이터로 — 말발굽, 효모, 그리고 may

모형이 맞는지는 자료가 말한다. 이 절은 세 원자료에 Poisson 모형을 맞추고, 평균과 분산의 비교라는 가장 간단한 진단으로 모형을 시험한다. 정식 검정(카이제곱 적합도 검정)은 8회에서 한다.

6.1 적률을 맞추기

Poisson 모형에는 모수가 λ 하나이고 $EX = \lambda$이므로, 자연스러운 추정은 표본평균 $\hat{\lambda} = \bar{x}$이다. 음이항에는 모수가 둘이므로 평균과 분산 두 적률을 맞춘다. 정리 1.3에서 $EX = \frac{rq}{p}$, $\operatorname{Var} X = \frac{EX}{p}$ 이므로

$$\hat{p} = \frac{\bar{x}}{s^2}, \qquad \hat{r} = \frac{\bar{x}\,\hat{p}}{1-\hat{p}}$$

이다(s^2는 표본분산). 모형의 적률과 표본의 적률을 같게 놓는 이 방법이 Pearson(1894)의 **적률법**이다. 9회에서 이것을 최대가능도와 비교한다.

정의 6.1 (분산/평균 비) 음이 아닌 정수값 자료의 $D = s^2/\bar{x}$를 **분산 지수**(index of dispersion)라 한다. Poisson이면 1 근처, 과산포면 1보다 크고, 과소산포(규칙적인 배치)면 1보다 작다.

D는 2회 6.6절의 Lexis 분산비와 같은 생각이다. 표본 크기가 n이면 Poisson 모형 아래 $(n-1)D$는 근사적으로 $\chi^2(n-1)$을 따른다(8회). 그러면 D의 표준오차는 대략 $\sqrt{2/(n-1)}$이고, $n = 200$이면 0.1이다.

6.2 세 자료

자료	n	$\bar{x}$	s^2	D	읽기
기병대, Fisher의 10개 군단	200	0.610	0.611	1.00	Poisson과 거의 완벽하게 맞는다
기병대, 14개 군단 전체	280	0.700	0.763	1.09	군단 사이의 차이가 조금 더해진다
효모, 농도 A–D	400씩	0.68–4.68	—	0.95–1.19	고르게 섞인 액체
*The Federalist*의 *may*	262	0.656	1.008	1.54	과산포. 음이항이 낫다

예제 6.2 (14개 군단 — 혼합의 흔적) Bortkiewicz의 14개 군단 전체(280개 군단–해)에서는 $D = 1.09$이다. 군단마다 20년의 사망 합계를 보면 VIII 군단은 7명, XI 군단은 25명이다. 모든 군단의 위험률이 같다면 군단 합계는 같은 Poisson을 따라야 한다. 그 산포를 재는 $\sum_c (T_c - \bar{T})^2/\bar{T}$는 27.3이고, 자유도 13의 χ^2 아래에서 이만큼 클 확률은 약 0.011이다(8회). 군단마다 위험률이 다르다는 증거, 곧 혼합(정리 5.2)의 흔적이다. Bortkiewicz가 조직이 다른 네 군단(근위대, I, VI, XI)을 따로 다룬 이유가 여기 있다. 남은 10개 군단에서는 $D = 1.00$이다.

예제 6.3 (Madison의 may — 과산포) Mosteller와 Wallace(1963)는 저자가 논란이던 *The Federalist* 12편이 Hamilton의 것인지 Madison의 것인지 가리려고, 두 사람의 확실한 글에서 기능어의 빈도를 셌다. Madison의 글을 일정한 길이의 조각 262개로 나눠 *may*의 수를 세면 0이 156번, 1이 63번, 2가 29번, … 6이 1번이다(`vcd::Federalist`). 평균은 0.656인데 분산은 1.008이다. Poisson의 기대 빈도는 0의 개수를 136으로 적게, 4 이상을 1.2로 적게 예측한다. 적률로 맞춘 음이항 NB($\hat{r} = 1.23, \hat{p} = 0.651$)은 0을 155, 4 이상을 5.4로 예측해 관측(156, 6)에 훨씬 가깝다. 같은 저자라도 주제에 따라 *may*를 쓰는 비율이 다르다. 조각마다 위험률이 다른 Poisson, 곧 정리 5.3의 그림이다. Mosteller와 Wallace는 바로 이 음이항 모형으로 12편 모두가 Madison의 글이라는 결론을 냈다.

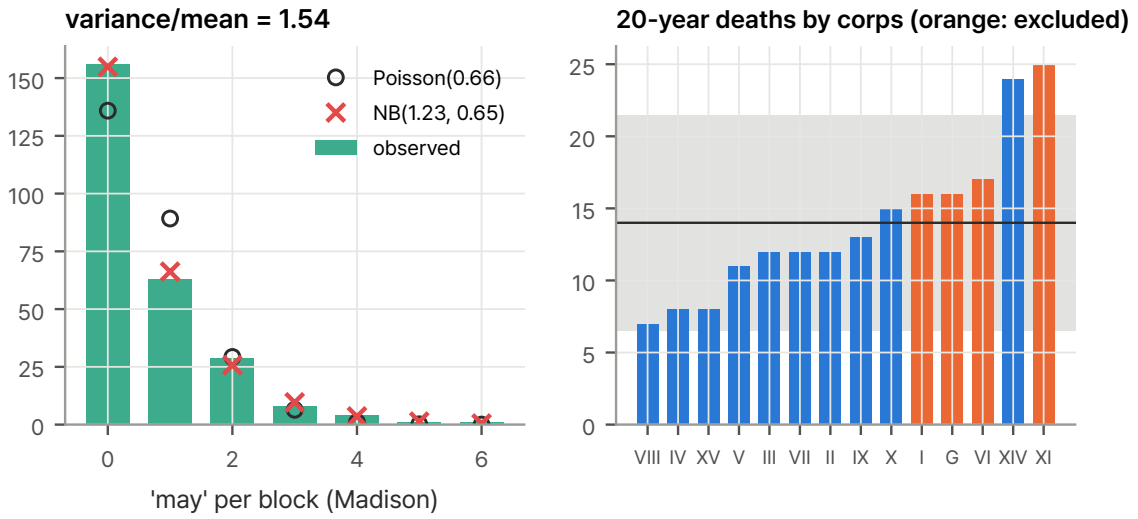


그림 6.1 왼쪽: Madison의 *may* 수(막대), Poisson(원)과 음이항(×)의 기대 빈도. 오른쪽: 14개 군단의 20년 사망 합계. 회색 띠는 모든 군단이 같은 위험률이라는 모형 아래 ±2 표준편차다.

연습문제

연습 6.1 (효모의 농도 ★) 효모 표본 D의 칸당 평균은 4.68이다. Poisson 모형에서 400칸의 평균의 표준오차는? 농도를 ±1% 안으로 알려면 몇 칸을 세야 하나(2배 표준오차 기준)? 풀이 45쪽

연습 6.2 (적률법 ★) *may* 자료의 $\bar{x} = 0.656$, $s^2 = 1.008$로 음이항의 $\hat{p}$, $\hat{r}$을 계산하고, 정리 5.3의 감마 혼합의 모수 $\hat{\beta}$를 구하라. 풀이 45쪽

연습 6.3 (0의 개수 ★★) Poisson 모형에서 $P(X=0) = e^{-\lambda}$이다. *may* 자료의 0의 비율 $\frac{156}{262}$로 λ를 추정하면 얼마인가? 평균으로 추정한 0.656과 왜 다른가? 두 추정이 다르다는 것 자체가 무엇의 증거인가? 풀이 45쪽

7

부쉬보기 — 가족의 경계

7.1 Poisson인 척하는 음이항

과산포가 작으면 음이항과 Poisson은 눈으로 구별되지 않는다. 그런데 꼬리에서는 크게 다르다. 평균이 같은 Poi(2)와 NB($r=4,\ p=\frac{2}{3}$) (분산 3)에서 10 이상이 나올 확률은 각각 4.6×10^{-5}와 1.6×10^{-3}, 약 35배 차이다. 보험의 큰 손해, 서버의 순간 폭주, 유전자 발현량의 이상치처럼 **드문 큰 값**이 중요한 문제에서, 평균만 보고 Poisson을 쓰면 위험을 크게 얕잡는다. 분산 지수 D를 먼저 보는 습관이 필요한 이유다.

7.2 Poisson 근사가 틀릴 때

정리 2.3은 $p\to0$을 요구한다. b(20, 0.4)를 Poi(8)로 어림하면 두 분포의 차이(전변동 거리)가 0.12로 쓸모가 없다. 이항의 분산 $npq=4.8$은 Poisson의 분산 8보다 훨씬 작다. 이항은 "기회의 수 n이 상한"이라 과소산포이고, p가 크면 그 차이가 드러난다. Poisson은 "기회는 사실상 무한하고 각 기회는 아주 드물다"는 모형이다.

7.3 무기억성은 지수와 기하뿐

정리 3.3에 따르면 연속분포 가운데 무기억성을 갖는 것은 지수분포뿐이다. 감마분포 $\Gamma(\alpha,\beta)$에서 $\alpha>1$이면 이미 기다린 시간이 길수록 남은 시간이 짧아지고("닳는다"), $\alpha<1$이면 길어진다("살아남은 것은 더 오래 산다"). 기계의 수명, 사람의 수명, 실직 기간을 지수분포로 모형화하면 이 차이를 잃는다. 생존분석의 위험함수 $h(t)=f(t)/(1-F(t))$가 상수인 것이 지수분포다(연습 7.2).

7.4 혼합은 이름표를 잃는다

정규 두 개의 혼합 $\pi\mathcal{N}(\mu_1,\sigma^2)+(1-\pi)\mathcal{N}(\mu_2,\sigma^2)$에서 (π,μ_1,μ_2)와 $(1-\pi,\mu_2,\mu_1)$은 똑같은 분포를 준다. 자료만으로는 어느 성분이 "첫째"인지 알 수 없다(이름표 바꾸기). 더 나쁘게는, 두 평균이 표준편차의 두 배보다 가까우면 혼합의 밀도는 봉우리가 하나뿐이어서(연습 7.3), 한 정규인지 두

정규의 혼합인지 눈으로는 구별할 수 없다. Pearson의 게가 정말 두 아종이었는지는 끝내 결론이 나지 않았다. 적률법이든 최대가능도든, 혼합 모형의 추정은 이 근본적인 어려움을 안고 있다(10회).

연습문제

연습 7.1 (꼬리의 차이 ★) $\mathrm{Poi}(2)$와 $\mathrm{NB}(4, \frac{2}{3})$의 평균과 분산을 확인하고, 노트북으로 $P(X \geq 10)$을 각각 구하라. 풀이 46쪽

연습 7.2 (위험함수 ★★) $h(t) = f(t)/(1-F(t))$를 위험함수라 한다. 지수분포에서 h가 상수임을 보이고, $\Gamma(2,1)$에서는 $h(t) = \frac{t}{1+t}$로 증가함을 보여라. 풀이 46쪽

연습 7.3 (두 봉우리의 조건 ★★★) $\frac{1}{2}\mathcal{N}(-\delta,1) + \frac{1}{2}\mathcal{N}(\delta,1)$의 밀도가 $x=0$에서 극소가 될 필요충분조건이 $\delta > 1$임을 보여라. (힌트: $x=0$에서의 2차 미분의 부호.) 두 평균의 거리가 표준편차의 두 배를 넘어야 봉우리가 둘이 된다. 풀이 46쪽

8

Beautiful insights — 같은 가족, 오늘의 문제

8.1 계승규칙과 add-one smoothing

언어 모형이 단어 다음에 올 단어의 확률을 상대도수 $\frac{c_w}{N}$으로 추정하면, 훈련 자료에 한 번도 없던 단어의 확률은 0이 된다. 그러면 그 단어가 들어간 문장 전체의 확률이 0이 되어 모형이 무너진다. Laplace의 처방은 모든 단어에 가상의 관측 1을 더하는 것이다. 어휘가 V개이면

$$\hat{p}_w = \frac{c_w + 1}{N + V}.$$

이것은 따름정리 4.6을 V개의 범주로 넓힌 것이다. 균등한 Dirichlet 사전분포(연습 4.5)의 사후평균이다. 더하는 수를 α로 바꾸면(Lidstone) 사전분포의 강도를 조절하는 것이 된다. 분류 신경망의 **label smoothing**도 같은 생각이다. 정답 범주의 확률 1을 $1-\varepsilon$으로, 나머지에 $\varepsilon/(V-1)$씩 나눠 주어 모형이 지나친 확신을 갖지 않게 한다.

8.2 베타–이항과 Thompson 샘플링

광고 두 개 가운데 어느 쪽의 클릭률이 높은지 실험하며 동시에 많이 클릭되게 하고 싶다(다중 슬롯머신 문제). Thompson(1933)의 방법은 각 광고의 클릭률에 Beta(1, 1) 사전분포를 주고, 매번 (1) 각 광고의 사후분포 $\mathrm{Beta}(1+s_i, 1+f_i)$에서 클릭률을 하나씩 뽑아, (2) 뽑힌 값이 가장 큰 광고를 보여 주고, (3) 결과로 사후분포를 고친다(정리 4.5). 확신이 없는 광고는 사후분포가 넓어 가끔 큰 값이 뽑히므로 탐색되고, 좋은 광고는 점점 자주 선택된다. 켤레 사전분포 덕분에 갱신이 덧셈 두 번이다. 노트북에서 구현한다.

8.3 셈 자료의 회귀 — Poisson과 음이항

셈 자료를 설명변수로 설명하는 Poisson 회귀는 $\log EY = \beta_0 + \beta_1 x$로 놓는다(일반화 선형모형, 13회). 실제 자료는 거의 언제나 과산포여서, 유전체학의 RNA-seq(유전자마다 읽힌 조각의 수)에서는 음이항 회귀가 표준이다(DESeq2, edgeR). 정리 5.3의 그림, 곧 "표본마다 생물학적 변동이 감마처럼 흩어져 있고, 주어진 발현량에서는 Poisson 측정 잡음" 이 그 근거다.

8.4 혼합 모형 — GMM과 전문가 혼합

정규 혼합(5.3절)은 군집 분석의 기본 모형이다(가우스 혼합 모형, GMM). 각 관측이 어느 성분에서 왔는지(잠재변수 Θ)를 모르는 것이 어려움이고, 10회의 EM 알고리즘이 이것을 "Θ를 추측하고(E), 모수를 고치기(M)"의 반복으로 푼다. 신경망의 **전문가 혼합**(mixture of experts)은 혼합 비율 π_k를 입력 x의 함수(게이트)로 만든 것이다. 큰 언어 모형이 매 토큰마다 일부 전문가만 쓰는 구조가 이것이다.

연습문제

연습 8.1 (스무딩의 효과 ★) 어휘 $V = 10{,}000$, 훈련 단어 $N = 10^6$개에서 단어 w가 0번, 다른 단어 u가 1,000번 나왔다. add-one 추정 $\hat{p}_w$, $\hat{p}_u$를 상대도수와 비교하라. 흔한 단어의 확률은 얼마나 줄어드나? 풀이 46쪽

연습 8.2 (Thompson의 첫걸음 ★★) 광고 A는 10번 보여 3번, B는 2번 보여 1번 클릭되었다. 사후분포가 Beta(4, 8)과 Beta(2, 2)일 때, 다음에 B가 선택될 확률 $P(p_B > p_A)$를 노트북의 시뮬레이션으로 구하라. B의 사후 평균이 A보다 크지만 확률은 1에 가깝지 않은 이유를 말하라. 풀이 46쪽

9

정리와 역사 노트

9.1 한 장 요약

	pmf / pdf, 평균, 분산, mgf	어디서 나오나
이항 b(n,p)	$\binom{n}{k}p^kq^{n-k},\ np,\ npq,\ (q+pe^t)^n$	n번 중 성공 수
음이항 NB(r,p)	$\binom{k+r-1}{k}p^rq^k,\ \frac{rq}{p},\ \frac{rq}{p^2},\ \left(\frac{p}{1-qe^t}\right)^r$	r번째 성공까지의 실패 · Poisson–감마
초기하	$\binom{K}{x}\binom{N-K}{n-x}/\binom{N}{n},\ np,\ npq\frac{N-n}{N-1}$	되돌려 넣지 않는 뽑기
다항	$\frac{n!}{\prod x_j!}\prod p_j^{x_j},\ \mathrm{Cov}(X_i,X_j)=-np_ip_j$	범주가 여럿
Poisson Poi(λ)	$e^{-\lambda}\frac{\lambda^k}{k!},\ \lambda,\ \lambda,\ e^{\lambda(e^t-1)}$	이항의 극한 · 과정의 셈
감마 $\Gamma(\alpha,\beta)$	$\frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha},\ \alpha\beta,\ \alpha\beta^2,\ (1-\beta t)^{-\alpha}$	과정의 기다림 · 지수의 합
$\chi^2(r)$	$\Gamma(\frac{r}{2},2),\ r,\ 2r,\ (1-2t)^{-r/2}$	표준정규 r개의 제곱합
베타 Beta(a,b)	$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)},\ \frac{a}{a+b},\ \frac{ab}{(a+b)^2(a+b+1)}$	감마의 비 · 사후분포 · 순서통계량
혼합	$\sum\pi_kf_k,\ E\mu(\Theta),\ E\sigma^2(\Theta)+\mathrm{Var}\,\mu(\Theta)$	섞인 집단

증명의 핵심은 세 가지 도구의 되풀이였다. **mgf의 곱**(독립합: 이항, Poisson, 감마, χ^2), **야코비안**(감마의 비 → 베타), **전분산 공식**(혼합의 분산, 과산포).

9.2 Stigler의 명명 법칙

> **Stigler's law of eponymy** "어떤 과학 법칙도 처음 발견한 사람의 이름을 달고 있지 않다."
>
> - **Poisson 분포**: de Moivre(1711)가 126년 먼저 같은 어림을 썼다. 그리고 Poisson은 이것을 하나의 분포로 다루지 않았다. 분포로서의 자리를 준 것은 Bortkiewicz(1898)였다.
> - **베타분포와 감마분포**: Euler가 적분 $B(a,b)$와 $\Gamma(\alpha)$를 연구했고(1730년대), 이것들을 확률분포로 이름 붙인 것은 Pearson(1895)의 분포 체계(Type I, Type III)였다. Bayes(1763)는 이미 베타 사후분포를 계산하고 있었다.

- χ^2 **분포**: Pearson(1900)의 이름이 붙지만, Abbe(1863)와 Helmert(1875)가 정규 제곱합의 분포로 먼저 유도했다(6회).

9.3 7 Pillars 지도

이번 회차는 기둥을 새로 세우기보다 기둥들이 쓸 **재료**를 만들었다. 셋째 기둥 **Likelihood**(9회)는 모수의 가족이 있어야 시작할 수 있고, 6절의 적률법은 그 직전 단계다. 넷째 기둥 **Intercomparison**(6회 이후)은 χ^2와 t 같은 분포로 자료 안의 변동을 서로 비교한다. 그리고 6절의 분산 지수는 일곱째 기둥 **Residual**의 씨앗이다. 모형이 설명하지 못하고 남는 것(과산포)을 보고 모형을 고친다.

9.4 다음 회 — 정규분포의 탄생과 Galton의 타원 (1733 → 1901)

이 회차에서 이항(정리 2.3의 다른 극한), 감마(큰 α), 베타(큰 a, b)가 모두 종 모양으로 가는 것을 그림에서 보았다. 5회는 그 종, 곧 정규분포가 어떻게 태어났는지(de Moivre 1733, Gauss 1809, Laplace 1810)를 따라간다. 이 노트의 χ^2 분포(정리 3.9)와 감마함수가 5회의 다변량 정규분포와 이차형식에서 쓰인다.

참고문헌

- Stigler, S. M. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard, 1986. (Bayes, Laplace, Pearson의 장)
- von Bortkiewicz, L. *Das Gesetz der kleinen Zahlen*. Teubner, 1898.
- Student. On the error of counting with a haemacytometer. *Biometrika* 5 (1907).
- Bayes, T. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions* 53 (1763).
- Laplace, P.-S. Mémoire sur la probabilité des causes par les événements (1774). · *Essai philosophique sur les probabilités* (1814).
- Pearson, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions A* 185 (1894).
- Greenwood, M., Yule, G. U. An inquiry into the nature of frequency distributions representative of multiple happenings. *JRSS* 83 (1920).
- Mosteller, F., Wallace, D. L. Inference in an authorship problem. *JASA* 58 (1963).
- Le Cam, L. An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* 10 (1960).
- Thompson, W. R. On the likelihood that one unknown probability exceeds another. *Biometrika* 25 (1933).

- 자료: R 패키지 `vcd`의 `HorseKicks`, `VonBort`, `Federalist`; `HistData`의 `Yeast`.

A

연습문제 풀이

각 풀이의 머리에 있는 문제 ○○쪽을 누르면 문제로 돌아간다. 계산 문제의 수치는 노트북에서 같은 계산으로 확인했다.

1절

연습 1.1 문제 7쪽

$P(0) = 0.9 \cdot 0.1 = 0.09$, $P(1) = 0.1 \cdot 0.1 + 0.9 \cdot 0.9 = 0.82$, $P(2) = 0.09$. 평균은 1로 $\mathrm{b}(2, 0.5)$와 같지만 분산은 $0.1 \cdot 0.9 + 0.9 \cdot 0.1 = 0.18$로 0.5보다 훨씬 작다. 확률이 서로 다른 시행의 합은 같은 평균의 이항보다 덜 흩어진다(과소산포).

연습 1.2 문제 7쪽

갑 A를 "성공", B를 "실패"로 보자. A가 비었다는 것을 처음 알게 되는 순간은 A가 $N+1$번째로 선택된 때이고, 그때 B에 k개가 남았다면 B는 $N-k$번 선택되었다. 그 전까지의 실패 수 $N-k$가 $\mathrm{NB}(N+1, \frac{1}{2})$의 값이므로 확률은 $\binom{(N-k)+N}{N-k} 2^{-(2N-k+1)}$. B가 먼저 비는 경우가 대칭으로 같으므로 2배 하면 $\binom{2N-k}{N} 2^{-(2N-k)}$.

연습 1.3 문제 8쪽

$\psi(t) = r \log p - r \log(1 - qe^t)$. $\psi'(t) = \frac{rqe^t}{1-qe^t}$에서 $\psi'(0) = \frac{rq}{p}$. $\psi''(t) = \frac{rqe^t}{(1-qe^t)^2}$에서 $\psi''(0) = \frac{rq}{p^2}$.

연습 1.4 문제 8쪽

$N = 45$, $K = 6$, $n = 6$, $p = \frac{6}{45}$. $EX = 0.8$, $\operatorname{Var} X = 6 \cdot \frac{6}{45} \cdot \frac{39}{45} \cdot \frac{39}{44} \approx 0.615$. $P(X = 3) = \frac{\binom{6}{3}\binom{39}{3}}{\binom{45}{6}} = \frac{20 \cdot 9139}{8{,}145{,}060} \approx 0.0224$.

연습 1.5 문제 8쪽

$$\frac{L(N)}{L(N-1)} = \frac{(N-K)(N-n)}{N(N-K-n+x)}$$

이고(이항계수를 전개해 소거), 이것이 1 이상이라는 부등식을 전개하면 $N^2-(K+n)N+Kn \geq N^2-(K+n)N+xN$, 곧 $N \leq \frac{nK}{x}$이다. 따라서 L은 $N \leq 1250$에서 증가, 그 뒤 감소하고 최대는 $N=1250$이다. 비가 정확히 1이면 1249와 1250이 같은 값이다.

연습 1.6 문제 8쪽

$$P(X_2=y \mid X_1=x)=\frac{\frac{n!}{x!\,y!\,(n-x-y)!}p_1^x p_2^y p_3^{n-x-y}}{\binom{n}{x}p_1^x(1-p_1)^{n-x}}=\binom{n-x}{y}\Big(\frac{p_2}{p_2+p_3}\Big)^y\Big(\frac{p_3}{p_2+p_3}\Big)^{n-x-y},$$

$1-p_1=p_2+p_3$을 썼다. 범주 1이 아닌 $n-x$번의 시행이 범주 2와 3으로 나뉘는 것이다.

연습 1.7 문제 8쪽

(a) 2회 정리 2.4로 $\mathrm{b}(n_i,p)$의 mgf는 $(q+pe^t)^{n_i}$이고, 독립합의 mgf는 곱이다(3회 정리 3.4(d)). 곱은 $(q+pe^t)^{n_1+n_2}$, 곧 $\mathrm{b}(n_1+n_2,p)$의 mgf다. (b) 두 mgf는 모든 실수 t에서 정의되고 같으므로 0을 포함하는 어떤 열린구간에서도 같다. 2회 3.3절의 상자(mgf의 유일성)로 분포가 같다. (c) m개의 공을 남자 n_1명과 여자 n_2명의 모임에서 고르는 방법 $\binom{n_1+n_2}{m}$을 고른 남자의 수 j로 나누면 $\sum_j \binom{n_1}{j}\binom{n_2}{m-j}$이다. 그러면

$$\begin{aligned}P(K_1+K_2=m)&=\sum_j \binom{n_1}{j}p^j q^{n_1-j}\binom{n_2}{m-j}p^{m-j}q^{n_2-m+j}\\&=p^m q^{n_1+n_2-m}\sum_j\binom{n_1}{j}\binom{n_2}{m-j}=\binom{n_1+n_2}{m}p^m q^{n_1+n_2-m}.\end{aligned}$$

(d) (a)에서는 두 mgf의 밑 $q+pe^t$가 같아야 곱이 한 거듭제곱이 된다. (c)에서는 $p^j p^{m-j}=p^m$으로 묶을 때 쓴다. p가 다르면 두 곳 모두 묶이지 않는다.

연습 1.8 문제 8쪽

(a) 각 시행에서 "범주 j"의 확률은 p_j이고 시행들은 독립이므로, 범주 j가 나온 횟수는 성공확률 p_j인 n번의 독립 시행의 성공 수다. 곧 $X_j \sim \mathrm{b}(n,p_j)$이다. (b) "범주 i 또는 j"를 성공으로 보면 확률은 p_i+p_j이다(서로소인 사건). 같은 논리로 $X_i+X_j \sim \mathrm{b}(n,p_i+p_j)$이다. (c) $\operatorname{Var}(X_i+X_j)=\operatorname{Var}X_i+\operatorname{Var}X_j+2\operatorname{Cov}(X_i,X_j)$에서

$$2\operatorname{Cov}(X_i,X_j)=n(p_i+p_j)(1-p_i-p_j)-np_i(1-p_i)-np_j(1-p_j)=-2np_ip_j.$$

(d) 한 시행은 한 범주에만 속하므로 $\mathbf{1}_{t,i}\mathbf{1}_{t,j}=0$이고 $\operatorname{Cov}(\mathbf{1}_{t,i},\mathbf{1}_{t,j})=0-p_ip_j$이다. 다른 시행 $t\neq s$의 지시변수는 독립이므로 공분산이 0이다. 공분산의 이중선형성(3회 정리 4.2)으로 $\operatorname{Cov}(X_i,X_j)=\sum_t \operatorname{Cov}(\mathbf{1}_{t,i},\mathbf{1}_{t,j})=-np_ip_j$이다.

2절

연습 2.1 문제 13쪽

$\frac{P(X=k)}{P(X=k-1)} = \frac{e^{-\lambda}\lambda^k/k!}{e^{-\lambda}\lambda^{k-1}/(k-1)!} = \frac{\lambda}{k}$. $k < \lambda$이면 1보다 커서 증가하고 $k > \lambda$이면 감소하므로 최빈값은 $\lfloor\lambda\rfloor$. λ가 정수이면 $k = \lambda$에서 비가 1이어서 $\lambda - 1$과 λ가 같은 확률이다.

연습 2.2 문제 13쪽

$x_n = \lambda(e^t - 1)$로 두면 $(1 + x_n/n)^n \to e^{x_n}$이다($n\log(1 + x/n) \to x$). 극한이 $\mathrm{Poi}(\lambda)$의 mgf다.

연습 2.3 문제 13쪽

오탈자가 없는 쪽의 수는 $\mathrm{b}(300, e^{-0.3})$이고 평균 $300e^{-0.3} \approx 222$쪽. 책 전체의 오탈자 수는 정리 2.4으로 $\mathrm{Poi}(90)$.

연습 2.4 문제 13쪽

$$P(X_1 = k \mid X_1 + X_2 = n) = \frac{e^{-\lambda_1}\frac{\lambda_1^k}{k!}\, e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}} = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}.$$

분모는 정리 2.4이다.

연습 2.5 문제 13쪽

$P(X \geq 3) = 1 - e^{-0.61}(1 + 0.61 + 0.186) \approx 0.024$, $P(X \geq 4) \approx 0.0035$. 200개 군단-해의 기대 횟수는 각각 4.8과 0.69이고, 관측은 $4(= 3 + 1)$와 1이다.

연습 2.6 문제 13쪽

(a) $M(t) = e^{-\lambda}\sum_k \frac{(\lambda e^t)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = \exp(\lambda(e^t - 1))$이다. 지수함수의 급수는 모든 실수에서 수렴하므로 모든 t에서 정의된다. (b) $\psi(t) = \lambda(e^t - 1)$이므로 $\psi'(t) = \psi''(t) = \lambda e^t$이고, $t = 0$에서 $EX = \psi'(0) = \lambda$, $\mathrm{Var}\,X = \psi''(0) = \lambda$이다. (c) $k = 0$인 항은 0이므로 $EX = \sum_{k\geq 1} e^{-\lambda}\frac{\lambda^k}{(k-1)!} = \lambda\sum_{j\geq 0} e^{-\lambda}\frac{\lambda^j}{j!} = \lambda$이다. 같은 방법으로 $E[X(X-1)] = \sum_{k\geq 2} e^{-\lambda}\frac{\lambda^k}{(k-2)!} = \lambda^2$이므로 $\mathrm{Var}\,X = \lambda^2 + \lambda - \lambda^2 = \lambda$이다.

연습 2.7 문제 13쪽

(a) 독립으로

$$P(X_1+X_2=n)=\sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda_1^k\lambda_2^{n-k}=e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}.$$

(b) m개의 합이 $\mathrm{Poi}(\lambda_1+\cdots+\lambda_m)$이라 하자. $X_1+\cdots+X_m$은 $(X_1,\dots,X_m)$의 함수이고 X_{m+1}은 그것들과 독립이므로 둘은 독립이다(3회의 "독립인 변수의 함수의 독립"). (a)를 두 변수에 쓰면 $m+1$개의 합이 $\mathrm{Poi}(\lambda_1+\cdots+\lambda_{m+1})$이다. (c) mgf의 곱은 $\prod_i \exp(\lambda_i(e^t-1))=\exp\big(\sum_i\lambda_i(e^t-1)\big)$이고, 유일성으로 분포가 정해진다. 합성곱으로 pmf를 직접 계산하는 일과 귀납을 mgf의 곱셈 한 번과 유일성 정리가 대신한다.

3절

연습 3.1 문제 18쪽

$\frac{d}{dt}\Big[e^{-\lambda t}\frac{(\lambda t)^k}{k!}\Big]=\lambda e^{-\lambda t}\Big[\frac{(\lambda t)^{k-1}}{(k-1)!}-\frac{(\lambda t)^k}{k!}\Big]$ ($k=0$이면 첫 항이 없다). $k=0,\dots,r-1$로 더하면 이웃 항이 소거되어 $-\lambda e^{-\lambda t}\frac{(\lambda t)^{r-1}}{(r-1)!}$만 남는다.

연습 3.2 문제 18쪽

첫 사건이 $(u,u+du]$에 있고($u>s$) 그 뒤 $(u,u+t]$에 사건이 없을 확률은 정의 (i)–(ii)로 $\lambda e^{-\lambda u}du\cdot e^{-\lambda t}$이다. $u>s$로 적분하면 $e^{-\lambda s}e^{-\lambda t}$이고, 결합 생존함수가 곱으로 인수분해되므로 W_1,W_2는 독립인 $\mathrm{Exp}(\lambda)$이다.

연습 3.3 문제 18쪽

$P(G\ge m)=\sum_{k\ge m}q^kp=q^m$이므로 $P(G\ge s+t\mid G\ge s)=q^{s+t}/q^s=q^t=P(G\ge t)$.

연습 3.4 문제 18쪽

$EX^k=\frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha+k-1}e^{-x/\beta}dx=\frac{\Gamma(\alpha+k)\beta^{\alpha+k}}{\Gamma(\alpha)\beta^\alpha}$. $k=1$: $\alpha\beta$. $k=2$: $\alpha(\alpha+1)\beta^2$이고 분산 $\alpha\beta^2$. 적분은 $\alpha+k>0$이면 유한하다.

연습 3.5 문제 18쪽

$T_5\sim\Gamma(5,\frac{1}{3})$이고 $P(T_5\le 2)=P(N(2)\ge 5)$, $N(2)\sim\mathrm{Poi}(6)$. 두 계산 모두 0.7149로 같다(정리 3.2).

연습 3.6 문제 18쪽

$\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$, $\Gamma(\frac{5}{2}) = \frac{3}{2}\Gamma(\frac{3}{2}) = \frac{3\sqrt{\pi}}{4}$. $\chi^2(3)$의 밀도는 $\frac{x^{1/2}e^{-x/2}}{\Gamma(3/2)2^{3/2}} = \frac{\sqrt{x}\,e^{-x/2}}{\sqrt{2\pi}}$ $(x>0)$.

연습 3.7 문제 18쪽

(a) $\int_0^1 x^{\alpha-1}e^{-x}dx \le \int_0^1 x^{\alpha-1}dx = \frac{1}{\alpha} < \infty$이다. $x^{\alpha-1}e^{-x/2} \to 0$ $(x\to\infty)$이므로 어떤 x_0 이후로 $x^{\alpha-1}e^{-x} \le e^{-x/2}$이고, $\int_{x_0}^\infty e^{-x/2}dx < \infty$이다. 사이의 구간 $[1,x_0]$에서는 연속함수의 적분이다. (b) $\int_\varepsilon^M x^\alpha e^{-x}dx = \left[-x^\alpha e^{-x}\right]_\varepsilon^M + \alpha\int_\varepsilon^M x^{\alpha-1}e^{-x}dx$이다. $\varepsilon^\alpha e^{-\varepsilon} \to 0$ $(\alpha>0)$이고 $M^\alpha e^{-M}\to 0$이므로 극한에서 $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$이다. (c) $\Gamma(1) = \int_0^\infty e^{-x}dx = 1 = 0!$이다. $\Gamma(n) = (n-1)!$이면 (b)로 $\Gamma(n+1) = n\Gamma(n) = n!$이다. (d) $x = z^2/2$이면 $dx = z\,dz$, $x^{-1/2} = \sqrt{2}/z$이므로 $\Gamma(\frac{1}{2}) = \int_0^\infty \frac{\sqrt{2}}{z}e^{-z^2/2}z\,dz = \sqrt{2}\int_0^\infty e^{-z^2/2}dz$이다. 3회 정리 5.3의 절반 $\int_0^\infty e^{-z^2/2}dz = \frac{\sqrt{2\pi}}{2}$을 넣으면 $\sqrt{\pi}$이다.

연습 3.8 문제 19쪽

(a) $t<1/\beta$이면 $c = 1/\beta - t > 0$이고, $y = cx$로 치환하면 $\int_0^\infty x^{\alpha-1}e^{-cx}dx = c^{-\alpha}\Gamma(\alpha)$이다. 따라서 $M(t) = \frac{c^{-\alpha}}{\beta^\alpha} = (\beta c)^{-\alpha} = (1-\beta t)^{-\alpha}$이다. $t \ge 1/\beta$이면 $c\le 0$이고 피적분함수가 무한대에서 0으로 가지 않아 적분이 발산한다. (b) $\psi'(t) = \frac{\alpha\beta}{1-\beta t}$, $\psi''(t) = \frac{\alpha\beta^2}{(1-\beta t)^2}$이므로 $EX = \alpha\beta$, $\operatorname{Var}X = \alpha\beta^2$이다. (c) mgf의 곱은 $\prod_i (1-\beta t)^{-\alpha_i} = (1-\beta t)^{-\sum\alpha_i}$이고 $t<1/\beta$인 열린구간에서 $\Gamma(\sum\alpha_i,\beta)$의 mgf와 같다. β가 다르면 밑 $1-\beta_i t$가 달라 한 거듭제곱으로 묶이지 않고, 합은 감마분포가 아니다.

연습 3.9 문제 19쪽

(a) $\{Z^2 \le y\} = \{-\sqrt{y} \le Z \le \sqrt{y}\}$이므로 $P(Z^2\le y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$이다. 연쇄법칙으로 미분하면 $\phi(\sqrt{y})\frac{1}{2\sqrt{y}} + \phi(-\sqrt{y})\frac{1}{2\sqrt{y}}$이다. (b) ϕ는 짝함수이므로 (a)는 $\frac{\phi(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2}$이다. $\sqrt{2\pi} = \Gamma(\frac{1}{2})2^{1/2}$이므로 이것은 $\frac{1}{\Gamma(1/2)2^{1/2}}y^{\frac{1}{2}-1}e^{-y/2}$, 곧 $\Gamma(\frac{1}{2},2)$의 밀도다. (c) Z_i^2은 Z_i의 함수이므로 Z_i들이 독립이면 Z_i^2들도 독립이다(3회의 "독립인 변수의 함수의 독립"). 각각 $\Gamma(\frac{1}{2},2)$이므로 정리 3.7로 합은 $\Gamma(\frac{r}{2},2) = \chi^2(r)$이다.

4절

연습 4.1 문제 23쪽

$\frac{d}{dx}\sum_{j\ge k}\binom{n}{j}x^j(1-x)^{n-j}$의 각 항은 $n\binom{n-1}{j-1}x^{j-1}(1-x)^{n-j} - n\binom{n-1}{j}x^j(1-x)^{n-j-1}$이고, 이웃 항이 소거되어 $n\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k} = \frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k}$이 남는다. 밀도의 적분이 1이므로 $\int_0^1 x^{k-1}(1-x)^{n-k}dx = \frac{(k-1)!(n-k)!}{n!}$.

연습 4.2 문제 23쪽

$P(S=s)=\int_0^1 \binom{n}{s}p^s(1-p)^{n-s}dp=\binom{n}{s}B(s+1,n-s+1)=\binom{n}{s}\frac{s!(n-s)!}{(n+1)!}=\frac{1}{n+1}$.

연습 4.3 문제 23쪽

사후분포는 Beta(7, 15). 평균 $\frac{7}{22}\approx 0.318$, 분산 $\frac{7\cdot 15}{22^2\cdot 23}$, 표준편차 ≈ 0.097. 표본 비율 0.3과 표준오차 0.102에 가깝다. 사후 평균은 0.3을 사전 평균 0.5 쪽으로 조금 끌어당긴 값이다.

연습 4.4 문제 23쪽

$R=U/(1-U)$, $U\sim\text{Beta}(a,b)$ (정리 4.2). $u=\frac{r}{1+r}$, $\frac{du}{dr}=\frac{1}{(1+r)^2}$ 이므로 2회 정리 4.3으로

$$f_R(r)=\frac{1}{B(a,b)}\Big(\frac{r}{1+r}\Big)^{a-1}\Big(\frac{1}{1+r}\Big)^{b-1}\frac{1}{(1+r)^2}=\frac{r^{a-1}}{B(a,b)(1+r)^{a+b}}.$$

연습 4.5 문제 23쪽

$(x_1,x_2,x_3)\mapsto(u_1,u_2,s)$, $x_1=u_1s$, $x_2=u_2s$, $x_3=(1-u_1-u_2)s$의 야코비안은 s^2이다. 결합밀도는 $\prod_j \frac{x_j^{a_j-1}e^{-x_j}}{\Gamma(a_j)}\cdot s^2=\frac{u_1^{a_1-1}u_2^{a_2-1}(1-u_1-u_2)^{a_3-1}}{\prod\Gamma(a_j)}s^{\sum a_j-1}e^{-s}$ 이고, s를 적분하면 $\Gamma(\sum a_j)$가 나온다.

연습 4.6 문제 24쪽

(a) $h(p,s)=\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\binom{n}{s}p^s(1-p)^{n-s}$ $(0<p<1,\ s=0,\dots,n)$. (b) $P(S=s)=\int_0^1 h(p,s)\,dp=\frac{\binom{n}{s}}{B(a,b)}\int_0^1 p^{a+s-1}(1-p)^{b+n-s-1}dp=\binom{n}{s}\frac{B(a+s,b+n-s)}{B(a,b)}$ 이다. 마지막 적분은 베타함수의 정의다. (c) $\frac{h(p,s)}{P(S=s)}=\frac{p^{a+s-1}(1-p)^{b+n-s-1}}{B(a+s,b+n-s)}$ 이고, 이것이 $\text{Beta}(a+s,b+n-s)$의 밀도다. (d) 조건부 밀도는 p의 함수로서 $h(p,s)$에 p와 무관한 상수 $1/P(S=s)$를 곱한 것이고 적분이 1이다. $p^{a+s-1}(1-p)^{b+n-s-1}$에 상수를 곱해 적분이 1이 되는 함수는 하나뿐이므로, 그 상수는 $1/B(a+s,b+n-s)$일 수밖에 없다. 비례하는 꼴이 정해지면 정규화 상수는 저절로 정해진다.

5절

연습 5.1 문제 28쪽

평균 168.5. 분산 $=\frac{1}{2}(49+36)+\frac{1}{2}(6.5^2+6.5^2)=42.5+42.25=84.75$. 집단 사이 항이 약 50%이다.

연습 5.2 문제 28쪽

$\text{Var}\,X/EX=\frac{r\beta+r\beta^2}{r\beta}=1+\beta$. $D=1.54$이면 $\beta=0.54$, $r=0.66/0.54\approx 1.22$.

연습 5.3 문제 28쪽

$E(e^{tX} \mid V=v) = e^{t^2 v/2}$ 이다($\mathcal{N}(0,v)$의 mgf. 지수 $tx - \frac{x^2}{2v}$를 완전제곱 $-\frac{(x-tv)^2}{2v} + \frac{t^2 v}{2}$로 바꾸고 3회 정리 5.3을 쓴다). 그리고 V의 mgf는 $\frac{1}{1-2b^2 s}$이므로 반복기댓값으로 $M_X(t) = \frac{1}{1-2b^2\cdot t^2/2} = \frac{1}{1-b^2t^2}$. Laplace 분포의 mgf는 $\int \frac{1}{2b} e^{-|x|/b} e^{tx}\,dx = \frac{1}{2}\left(\frac{1}{1-bt} + \frac{1}{1+bt}\right) = \frac{1}{1-b^2t^2}$. 유일성으로 같은 분포다.

연습 5.4 문제 28쪽

$E(S \mid p) = np$, $\operatorname{Var}(S \mid p) = np(1-p)$. $Ep = \bar{p}$, $\operatorname{Var} p = \frac{\bar{p}(1-\bar{p})}{a+b+1}$(따름정리 4.3). 정리 5.2로 $ES = n\bar{p}$이고

$$\operatorname{Var} S = n\left(\bar{p} - \operatorname{Var} p - \bar{p}^2\right) + n^2 \operatorname{Var} p = n\bar{p}(1-\bar{p}) + n(n-1)\operatorname{Var} p = n\bar{p}(1-\bar{p})\frac{a+b+n}{a+b+1}.$$

연습 5.5 문제 28쪽

$E\,e^{tX} = \sum_k \pi_k E(e^{tX} \mid \Theta = \theta_k) = \sum_k \pi_k M_k(t)$. 혼합은 하나의 관측이 한 성분에서 나오는 것이므로 mgf의 **가중 평균**이고, 독립합은 관측마다 모든 성분이 더해지는 것이므로 mgf의 **곱**이다.

6절

연습 6.1 문제 31쪽

표준오차 $\sqrt{4.68/400} \approx 0.108$. $2\sqrt{4.68/n} \le 0.0468$에서 $n \ge \frac{4\cdot 4.68}{0.0468^2} \approx 8{,}550$칸.

연습 6.2 문제 31쪽

$\hat{p} = \frac{0.656}{1.008} \approx 0.651$, $\hat{r} = \frac{0.656\cdot 0.651}{0.349} \approx 1.22$. $p = \frac{1}{1+\beta}$이므로 $\hat{\beta} = \frac{1}{\hat{p}} - 1 \approx 0.54$.

연습 6.3 문제 31쪽

$\hat{\lambda}_0 = -\log\frac{156}{262} \approx 0.52$. 평균으로 추정한 0.656보다 작다. Poisson 모형이 맞다면 두 추정은 같은 λ를 향해야 한다. 0이 "평균이 말하는 것보다 많다"는 것은 과산포의 다른 얼굴이고, 음이항(같은 평균에서 0의 확률이 Poisson보다 크다)이 그것을 설명한다.

7절

연습 7.1 문제 33쪽

NB(4, $\frac{2}{3}$)의 평균은 $\frac{4\cdot 1/3}{2/3}=2$, 분산은 $\frac{4/3}{4/9}=3$. $P(X\ge 10)$은 Poisson 4.6×10^{-5}, 음이항 1.6×10^{-3}(노트북).

연습 7.2 문제 33쪽

지수: $h=\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}=\lambda$. $\Gamma(2,1)$: $f=te^{-t}$, $1-F=(1+t)e^{-t}$(부분적분)이므로 $h=\frac{t}{1+t}$, 증가하며 1로 간다.

연습 7.3 문제 33쪽

$f(x)\propto e^{-(x-\delta)^2/2}+e^{-(x+\delta)^2/2}$. 대칭이므로 $f'(0)=0$이고, $\frac{d^2}{dx^2}e^{-(x-\delta)^2/2}=\big((x-\delta)^2-1\big)e^{-(x-\delta)^2/2}$이므로 $f''(0)\propto 2(\delta^2-1)e^{-\delta^2/2}$. $\delta>1$이면 $f''(0)>0$, 곧 0이 극소이고 봉우리가 둘이다.

8절

연습 8.1 문제 35쪽

$\hat p_w=\frac{1}{1{,}010{,}000}\approx 9.9\times10^{-7}$ (상대도수 0). $\hat p_u=\frac{1001}{1{,}010{,}000}\approx 9.91\times10^{-4}$ (상대도수 10^{-3}), 약 0.9% 줄었다. 흔한 단어에서 조금씩 떼어 처음 보는 단어 1만 개에 나눠 주는 셈이다.

연습 8.2 문제 35쪽

$P(p_B>p_A)=\int f_B(x)F_A(x)\,dx\approx 0.725$. B의 사후 평균 0.5는 A의 $\frac{1}{3}$보다 크지만, B는 두 번밖에 보지 않아 사후분포가 넓다. Thompson 샘플링은 약 27%의 확률로 여전히 A를 보여 준다.