**4회 워크북 — 풀이판**

# 셈하는 분포와 섞인 분포 손으로 푸는 35문항

학생판과 같은 문제에 빈칸의 답과 완전한 풀이를 붙였다. 학생판으로 먼저 풀고, 힌트와 최종답으로도 막힐 때 이 판을 본다.

| 부 | 하는 일 | 문항 | 예상 시간 |
|---|---|---|---|
| **A 계산 드릴** | pmf · 밀도 · 적률 · 사후분포를 손으로 계산한다 | 14 | 2시간 10분 |
| **B 빈칸 증명** | 노트의 단계형 증명을 책을 덮고 다시 쓴다 | 7 | 2시간 35분 |
| **C 증명 문제** | 변형, 일반화, 반례 만들기 | 10 | 3시간 25분 |
| **D 퀄형** | 여러 회차를 잇는 종합 문제. 제한 시간 안에 | 4 | 2시간 35분 |
| | 합계 | 35 | 약 10시간 45분 |

**노트** `04_counting_notes.pdf` — 문항의 "정리 3.3" 같은 번호는 노트의 번호다. 앞 회차는 "3회 정리 5.1" 처럼 쓴다

**도구** `00_toolkit_notes.pdf` — [T1] 지시함수, [T2] 귀류, [T3] 귀납, [T4] 부등식, [T5] 극한, [T6] 합과 적분의 교환, [T7] 변수변환, [T8] 급수 전개

**검산** 노트북 `04_counting.ipynb`의 "워크북 검산" 절: `check("A10c", "11/16")`

## 이 워크북을 쓰는 법

1. **노트를 덮고 푼다.** 문제의 머리 오른쪽에 쓸 도구, 관련 정리 번호, 예상 시간이 있다. 예상 시간의 두 배가 지나도 진척이 없으면 그때 노트를 편다.
2. **막히면 3단으로 본다.** 뒤의 **힌트와 최종답**에서 힌트를 먼저 본다. 다 푼 뒤에는 최종답과 맞춰 본다. 그래도 어긋나거나 증명이 막히면 풀이판을 본다.
3. **B부는 순서대로 푼다.** 초기하분포의 분산, 무기억성의 역, 베타와 감마의 관계를 여러 번 만난다. 처음에는 근거만 빈칸이고(①), 다음에는 단계의 이름만 있고(②), 마지막에는 전략만 있다(③). 앞 단계의 답을 보면서 다음 단계를 쓰지 않는다.
4. **계산은 검산한다.** 최종답이 수나 식인 문항은 노트북의 `check()`에 넣으면 맞는지만 알려 준다. 정답은 보여 주지 않는다.
5. **증명은 근거를 단다.** 등호와 부등호마다 어느 정의, 정리, 도구를 썼는지 오른쪽에 적는다. 합과 적분의 순서를 바꿀 때는 조건을 확인한다.

**표시** ★ 정의를 바로 쓰는 문제, ★★ 두세 단계를 잇는 문제, ★★★ 퀄에 나오는 길이의 문제. [T1]–[T8]은 증명 도구 상자의 도구 번호다.

# A

# 계산 드릴

pmf, 밀도, 적률, 사후분포를 손으로 계산한다. 답이 수나 식으로 떨어지므로 다 풀면 뒤의 최종 답이나 노트북의 `check()`로 맞춰 본다. 계산기는 마지막 수치에만 쓴다.

### A1 세 번째 6까지 ★

정리 1.3 · 8분

공정한 주사위를 6이 세 번 나올 때까지 던진다. (a) 6이 아닌 눈이 정확히 5번 나올 확률을 구하라. (b) 던지는 횟수의 평균을 구하라.

**풀이** (a) 6이 아닌 눈의 수 $Y \sim \mathrm{NB}(3, \frac{1}{6})$이므로 정리 1.3으로

$$P(Y=5) = \binom{5+3-1}{5}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^5 = 21 \cdot \frac{5^5}{6^8} = \frac{21875}{559872} \approx 0.0391.$$

(b) 던지는 횟수는 $Y+3$이고 $EY = \frac{rq}{p} = \frac{3 \cdot 5/6}{1/6} = 15$이므로 평균은 18이다. 6 하나를 보는 데 평균 6번이 걸리고, 그것이 세 번이다.

**최종답** (a) $\binom{7}{5}\frac{5^5}{6^8} \approx 0.0391$ (b) 18

### A2 브리지 패의 하트 ★

정의 1.4, 정리 1.5 · 10분

잘 섞은 52장에서 13장을 받는다. 하트의 수를 $X$라 하자. (a) $EX$, (b) $\operatorname{Var} X$, (c) $P(X=0)$을 구하라.

**풀이** $p = \frac{13}{52} = \frac{1}{4}$이다. (a) $EX = 13 \cdot \frac{1}{4} = \frac{13}{4}$. (b) 정리 1.5로

$$\operatorname{Var} X = 13 \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{52-13}{52-1} = \frac{39}{16} \cdot \frac{39}{51} = \frac{507}{272} \approx 1.864.$$

되돌려 넣는 뽑기라면 $\frac{39}{16} \approx 2.44$이다. (c) 하트가 아닌 39장에서 13장을 모두 받아야 하므로 $P(X=0) = \binom{39}{13}/\binom{52}{13} \approx 0.0128$이다.

**최종답** (a) $\frac{13}{4}$ (b) $\frac{507}{272} \approx 1.864$ (c) $\binom{39}{13}\Big/\binom{52}{13} \approx 0.0128$

### A3 열 명의 혈액형 ★★

정의 1.7, 정리 1.8 · 10분

혈액형 A, B, O, AB의 비율이 0.34, 0.27, 0.28, 0.11인 집단에서 열 명을 독립적으로 뽑는다. (a) A, B, O가 3명씩이고 AB가 1명일 확률을 구하라. (b) A형의 수와 B형의 수의 공분산을 구하라. (c) 두 수의 상관계수를 구하라.

**풀이** (a) $\frac{10!}{3!\,3!\,3!\,1!} = 16800$이므로 확률은 $16800 \cdot 0.34^3 \cdot 0.27^3 \cdot 0.28^3 \cdot 0.11 \approx 0.0314$이다. (b) 정리 1.8로 $\operatorname{Cov}(X_A, X_B) = -10 \cdot 0.34 \cdot 0.27 = -0.918$이다. (c) $\operatorname{Var} X_A = 10 \cdot 0.34 \cdot 0.66$, $\operatorname{Var} X_B = 10 \cdot 0.27 \cdot 0.73$이므로

$$\rho = \frac{-0.918}{\sqrt{2.244 \cdot 1.971}} = -\sqrt{\frac{0.34 \cdot 0.27}{0.66 \cdot 0.73}} \approx -0.437.$$

$n$이 상관계수에서 사라진다. 상관은 몇 명을 뽑든 범주의 비율로만 정해진다.

**최종답** (a) $\approx 0.0314$ (b) $-0.918$ (c) $\approx -0.437$

### A4 불량품 셋 이상 ★

정리 2.3 · 10분

불량률이 0.002인 공정에서 1000개를 만든다(각각 독립). 불량품이 3개 이상일 확률을 (a) Poisson 근사로, (b) 이항분포로 정확히 구하라. (c) 2.1절의 Le Cam 한계 $np^2$은 얼마인가? (a)와 (b)의 차이와 비교하라.

**풀이** (a) $\lambda = 2$이므로 $1 - e^{-2}(1 + 2 + 2) = 1 - 5e^{-2} \approx 0.323324$이다. (b) $0.998^{1000} \approx 0.135065$, $1000 \cdot 0.002 \cdot 0.998^{999} \approx 0.270671$, $\binom{1000}{2}(0.002)^2(0.998)^{998} \approx 0.270942$이므로 $1 - 0.676677 = 0.323323$이다. (c) $np^2 = 0.004$이다. 실제 차이는 $10^{-6}$ 정도로 한계보다 훨씬 작다. Le Cam의 한계는 모든 사건에 대한 최대 차이의 상한이다.

**최종답** (a) $1 - 5e^{-2} \approx 0.3233$ (b) $\approx 0.323323$ (c) $0.004$

### A5 콜센터 ★

정의 3.1, 정리 3.2 · 10분

전화가 시간당 평균 6통씩 Poisson 과정으로 걸려 온다. 시간은 분 단위로 쓴다. (a) 10분 동안 전화가 한 통도 없을 확률을 구하라. (b) 세 번째 전화가 올 때까지의 시간의 평균과 (c) 분산을 구하라. (d) 첫 전화가 10분과 20분 사이에 올 확률을 구하라.

**풀이** 분당 비율 $\lambda = 0.1$이다. (a) $N(10) \sim \mathrm{Poi}(1)$이므로 $e^{-1} \approx 0.368$이다. (b)(c) 정리 3.2로 $T_3 \sim \Gamma(3, 10)$이므로 평균 $3 \cdot 10 = 30$분, 분산 $3 \cdot 10^2 = 300$분$^2$이다. (d) $T_1 \sim \mathrm{Exp}(0.1)$이므로 $P(10 < T_1 \le 20) = e^{-1} - e^{-2} \approx 0.233$이다.

**최종답** (a) $e^{-1} \approx 0.368$ (b) 30분 (c) 300분$^2$ (d) $e^{-1} - e^{-2} \approx 0.233$

### A6 방문자와 구매자 ★★

정리 2.5 · 10분

하루 방문자 수가 $\mathrm{Poi}(10)$이고, 방문자는 각자 독립적으로 확률 0.3으로 물건을 산다. (a) 구매자가 한 명도 없을 확률을 구하라. (b) 사지 않은 사람이 7명이었다는 것을 알 때 구매자 수의 조건부분산을 구하라. (c) 방문자가 10명이었다는 것을 알 때 구매자 수의 조건부분산을 구하라.

**풀이** (a) 정리 2.5로 구매자 수 $N_1 \sim \mathrm{Poi}(3)$이므로 $e^{-3} \approx 0.0498$이다. (b) $N_1$과 비구매자 수 $N_2$는 독립이므로 $N_2 = 7$을 알아도 $N_1$의 분포는 그대로 $\mathrm{Poi}(3)$이고 분산은 3이다. (c) $N = 10$이 주어지면 $N_1 \sim \mathrm{b}(10, 0.3)$이므로 분산은 $10 \cdot 0.3 \cdot 0.7 = 2.1$이다. 전체를 알면 구매자 수에 대한 정보가 생기고, 비구매자 수만 알면 정보가 없다.

**최종답** (a) $e^{-3} \approx 0.0498$ (b) 3 (c) 2.1

### A7 감마함수로 적분하기 ★

정의 3.4, 정리 3.5 · 8분

다음을 감마함수로 계산하라. (a) $\Gamma(\frac{7}{2})$ (b) $\int_0^\infty x^4 e^{-2x}\,dx$ (c) $\int_0^\infty \sqrt{x}\,e^{-x}\,dx$

**풀이** (a) $\Gamma(\frac{7}{2}) = \frac{5}{2}\cdot\frac{3}{2}\cdot\frac{1}{2}\Gamma(\frac{1}{2}) = \frac{15\sqrt{\pi}}{8}$. (b) $y = 2x$이면 $\int_0^\infty \frac{y^4}{16}e^{-y}\frac{dy}{2} = \frac{\Gamma(5)}{32} = \frac{24}{32} = \frac{3}{4}$. (c) $\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$.

**최종답** (a) $\frac{15\sqrt{\pi}}{8}$ (b) $\frac{3}{4}$ (c) $\frac{\sqrt{\pi}}{2}$

### A8 감마분포의 적률과 확률 ★★

정리 3.2, 3.7 · 10분

$X \sim \Gamma(3, 2)$이다. (a) $EX^2$, (b) $E(1/X)$, (c) $P(X \le 2)$를 구하라.

**풀이** (a) $EX^2 = 2^2\frac{\Gamma(5)}{\Gamma(3)} = 4 \cdot 12 = 48$이다. 분산 $\alpha\beta^2 = 12$와 평균의 제곱 36을 더한 값과 같다. (b) $E(1/X) = 2^{-1}\frac{\Gamma(2)}{\Gamma(3)} = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$이다. $1/EX = \frac{1}{6}$보다 크다. (c) $\Gamma(3,2)$는 비율 $\lambda = \frac{1}{2}$인 Poisson 과정의 $T_3$이므로 정리 3.2로 $P(X \le 2) = P(N(2) \ge 3)$, $N(2) \sim \mathrm{Poi}(1)$이다. $1 - e^{-1}(1 + 1 + \frac{1}{2}) = 1 - \frac{5}{2}e^{-1} \approx 0.0803$이다.

**최종답** (a) 48 (b) $\frac{1}{4}$ (c) $1 - \frac{5}{2}e^{-1} \approx 0.0803$

### A9 표준정규의 제곱합 ★

정의 3.8, 정리 3.9 · 8분

$Z_1, Z_2, Z_3 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$이다. (a) $P(Z_1^2 + Z_2^2 > 4)$, (b) $\mathrm{Var}(Z_1^2 + Z_2^2 + Z_3^2)$, (c) $Ee^{(Z_1^2+Z_2^2+Z_3^2)/4}$를 구하라.

**풀이** (a) 정리 3.9로 $Z_1^2 + Z_2^2 \sim \chi^2(2) = \Gamma(1,2)$, 평균 2인 지수분포이므로 $P(>4) = e^{-2} \approx 0.135$. (b) $\chi^2(3)$의 분산 $2 \cdot 3 = 6$. (c) mgf $(1-2t)^{-3/2}$에 $t = \frac{1}{4}$를 넣어 $(\frac{1}{2})^{-3/2} = 2\sqrt{2}$.

**최종답** (a) $e^{-2} \approx 0.135$ (b) 6 (c) $2\sqrt{2}$

### A10 베타분포 ★

정의 4.1, 따름정리 4.3 · 8분

$U \sim \mathrm{Beta}(3,2)$이다. (a) $EU$, (b) $\mathrm{Var}\,U$, (c) $P(U > \frac{1}{2})$를 구하라.

**풀이** (a) $\frac{3}{3+2} = \frac{3}{5}$. (b) $\frac{3\cdot 2}{5^2 \cdot 6} = \frac{1}{25}$. (c) $\int_{1/2}^{1} 12u^2(1-u)\,du = 12\big[\frac{u^3}{3} - \frac{u^4}{4}\big]_{1/2}^{1} = 12\big(\frac{1}{12} - \frac{5}{192}\big) = \frac{11}{16}$.

**최종답** (a) $\frac{3}{5}$ (b) $\frac{1}{25}$ (c) $\frac{11}{16}$

### A11 사후분포와 다음 시행 ★★

정리 4.5, 따름정리 4.3 · 10분

성공 확률 $p$의 사전분포가 $\mathrm{Beta}(2,2)$이고 10번 가운데 7번 성공했다. (a) 다음 시행이 성공할 확률, (b) 사후분산, (c) 다음 두 시행이 모두 성공할 확률을 구하라.

**풀이** 정리 4.5로 사후분포는 $\mathrm{Beta}(2+7, 2+3) = \mathrm{Beta}(9,5)$이다. (a) 반복기댓값으로 $E(p \mid \text{자료}) = \frac{9}{14}$. (b) $\frac{9\cdot 5}{14^2 \cdot 15} = \frac{3}{196}$. (c) $p$가 주어지면 두 시행이 모두 성공할 확률은 $p^2$이므로 $E(p^2 \mid \text{자료}) = \frac{9\cdot 10}{14\cdot 15} = \frac{3}{7} \approx 0.429$이다. $(\frac{9}{14})^2 \approx 0.413$보다 크다. 차이가 사후분산 $\frac{3}{196}$이다.

**최종답** (a) $\frac{9}{14}$ (b) $\frac{3}{196}$ (c) $\frac{3}{7}$

### A12 두 Poisson의 혼합 ★★

정의 5.1, 정리 5.2 · 10분

확률 0.3으로 Poi(1), 확률 0.7로 Poi(4)에서 뽑은 $X$가 있다. (a) $EX$, (b) $\mathrm{Var}\,X$를 구하라. (c) $X = 0$이었을 때 Poi(1)에서 뽑았을 확률을 구하라.

**풀이** (a) $EX = 0.3 \cdot 1 + 0.7 \cdot 4 = 3.1$. (b) 집단 안: $0.3 \cdot 1 + 0.7 \cdot 4 = 3.1$. 집단 사이: $0.3 \cdot 1^2 + 0.7 \cdot 4^2 - 3.1^2 = 11.5 - 9.61 = 1.89$. 합은 4.99로 평균보다 크다. (c) 베이즈 정리로

$$\frac{0.3e^{-1}}{0.3e^{-1} + 0.7e^{-4}} \approx \frac{0.1104}{0.1232} \approx 0.896.$$

0이 나온 것은 위험이 낮은 집단이었을 증거가 된다. 사전확률 0.3이 사후확률 0.896이 된다.

**최종답** (a) 3.1 (b) 4.99 (c) $\approx 0.896$

### A13 다섯 점의 중앙값 ★★

예제 4.4, 따름정리 4.3 · 10분

$U_1, \dots, U_5 \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$의 중앙값을 $M$이라 하자. (a) $P(M \le \frac{1}{4})$를 구하라. (b) $\operatorname{Var} M$을 구하라.

**풀이** (a) $\frac{1}{4}$ 이하인 점의 수는 $\mathrm{b}(5, \frac{1}{4})$이므로

$$P(M \le \tfrac{1}{4}) = \sum_{j=3}^{5} \binom{5}{j} \left(\tfrac{1}{4}\right)^{j} \left(\tfrac{3}{4}\right)^{5-j} = \frac{10 \cdot 9 + 5 \cdot 3 + 1}{1024} = \frac{106}{1024} = \frac{53}{512} \approx 0.104.$$

(b) 예제 4.4로 $M = U_{(3)} \sim \mathrm{Beta}(3,3)$이고 $\operatorname{Var} M = \frac{9}{36 \cdot 7} = \frac{1}{28}$이다. 표준편차 0.19는 한 점의 표준편차 0.29보다 작다.

**최종답** (a) $\frac{53}{512} \approx 0.104$ (b) $\frac{1}{28}$

### A14 닳는 전구와 닳지 않는 전구 ★

정리 3.3 · 8분

이미 500시간 켜진 전구가 앞으로 1000시간 이상 버틸 확률을 구하라. 수명이 (a) 평균 1000시간인 지수분포일 때, (b) $\Gamma(2, 500)$일 때. 두 분포의 평균은 같다.

**풀이** (a) 무기억성(정리 3.3)으로 $P(X > 1000) = e^{-1} \approx 0.368$이다. (b) 정리 3.2로 $P(X > x) = P(N(x) \le 1) = e^{-x/500}(1 + \frac{x}{500})$이므로

$$\frac{P(X > 1500)}{P(X > 500)} = \frac{4e^{-3}}{2e^{-1}} = 2e^{-2} \approx 0.271.$$

같은 평균이라도 $\alpha = 2 > 1$인 감마는 이미 쓴 시간만큼 닳아서 남은 수명이 짧아진다.

**최종답** (a) $e^{-1} \approx 0.368$ (b) $2e^{-2} \approx 0.271$

# B

# 빈칸 증명

노트의 단계형 증명을 책을 덮고 다시 쓴다. 같은 정리를 여러 번 만나며 도움이 줄어든다. ① 단계는 주어지고 근거와 몇 개의 식만 빈칸이다. ② 단계의 이름만 주어진다. ③ 전략만 주어진다. 근거에는 정의나 정리의 번호, 앞 회차의 번호, 도구 번호를 쓴다.

### B1 초기하분포의 분산, 근거 채우기 ★

① · 정리 1.5 · 15분

$N$개 가운데 $K$개가 성공인 모집단에서 $n$개를 되돌려 넣지 않고 뽑는다. $p = K/N$이다. 근거와 빈칸을 채워라.

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

② $P(\mathbf{1}_i = 1) = p$이므로 $EX = np$이다. 뽑는 순서의 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 $\mathrm{Cov}(\mathbf{1}_i, \mathbf{1}_j) = -\frac{p(1-p)}{N-1}$이다. ②③, $\mathrm{Cov} = E[\mathbf{1}_i\mathbf{1}_j] - p^2$

⑤ 분산을 공분산의 합으로 쓴다.

$$\mathrm{Var}\, X = \sum_i \mathrm{Var}\, \mathbf{1}_i + \sum_{i \neq j} \mathrm{Cov}(\mathbf{1}_i, \mathbf{1}_j) \qquad \text{3회 정리 4.2(c)}$$

$$= np(1-p) - n(n-1) \cdot \frac{p(1-p)}{N-1} = np(1-p)\frac{N-n}{N-1} \qquad \text{④, 순서쌍 } i \neq j \text{의 수}$$

**풀이** 빈칸의 빨간 글씨가 답이다. 노트 정리 1.5의 증명과 같다. 둘째 단계의 대칭은 "$n$개를 뽑는 모든 순서가 같은 확률이므로, $i$번째 자리에 오는 것은 $N$개 가운데 고르게 하나"라는 뜻이다. 공분산이 음수인 것은 성공 하나가 뽑히면 남은 모집단에서 성공의 비율이 줄기 때문이다.

### B2 무기억성의 역, 근거 채우기 ★★

① · 정리 3.3 · 15분

양수인 연속 확률변수 $X$가 모든 $s,t>0$에서 $P(X>s+t\mid X>s)=P(X>t)$를 만족하면 지수분포임을 보인다. $G(x)=P(X>x)$이다. 근거를 채워라.

① 모든 $s>0$에서 $G(s)>0$이고 $G(s+t)=G(s)G(t)$이다. — 조건부확률의 정의, $\{X>s+t\}\subset\{X>s\}$

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

③ $0<c<1$이다. — $c=1$이면 $G(n)=1$인데 $G(n)\to 0$, 귀류 [T2]

④ 양의 유리수 $q=\frac{n}{m}$에서 $G(q)=c^q$이다. — $c=G(\frac{1}{m})^m$, 그리고 ②와 같은 귀납

⑤ 실수 $x>0$에서 $G(x)=c^x$이다. — 유리수열 $q_k\downarrow x$, $G$의 오른쪽 연속

⑥ $\lambda=-\log c$로 두면 $X\sim\mathrm{Exp}(\lambda)$이다. — ③으로 $\lambda>0$

**풀이** 빈칸의 빨간 글씨가 근거다. 노트 정리 3.3의 증명과 같다. 셋째 단계에서 $G(n)\to 0$은 1회 정리 6.3(b)의 $F(n)\to 1$이다. 다섯째 단계는 $G$가 유리수에서만 정해져 있으면 실수에서의 값을 정하는 데 연속성 같은 성질이 하나 더 필요하다는 것을 보여 준다. 그런 성질이 없으면 $G(s+t)=G(s)G(t)$의 기이한 해들이 존재한다.

### B3 베타와 감마, 근거와 식 채우기 ★★

① · 정리 4.2 · 15분

$X\sim\Gamma(a,1)$과 $Y\sim\Gamma(b,1)$이 독립이고 $U=\frac{X}{X+Y}$, $V=X+Y$이다. 빈칸을 채워라.

① 역변환은 $x=uv$, $y=(1-u)v$이고 $(0,1)\times(0,\infty)$ 위의 일대일 대응이다. — 대입해 확인

② 역변환의 야코비안은 $v$이다. — $2\times 2$ 행렬식 [T7]

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

④ $f_{U,V}(u,v)=\frac{u^{a-1}(1-u)^{b-1}}{\Gamma(a)\Gamma(b)/\Gamma(a+b)}\cdot\frac{v^{a+b-1}e^{-v}}{\Gamma(a+b)}$이다. — 3회 정리 5.1

⑤ $U$와 $V$는 독립이고 $V\sim\Gamma(a+b,1)$이다. — 인수분해, 3회 정리 3.2

⑥ $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$이고 $U\sim\mathrm{Beta}(a,b)$이다. — $U$의 밀도의 적분은 1

**풀이** 빈칸의 빨간 글씨가 답이다. 노트 정리 4.2의 증명과 같다. 야코비안은 $\det\left(\begin{smallmatrix} v & u \\ -v & 1-u\end{smallmatrix}\right)=v(1-u)+uv=v$이다. 마지막 단계는 "밀도의 적분은 1"이라는 사실 하나로 베타함수와 감마함수의 관계를 얻는다. 적분을 직접 계산하지 않는다.

### B4 초기하분포의 분산, 단계의 이름만 ★★

② · 정리 1.5 · 25분

정리 1.5를 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① $X$를 지시변수의 합으로 쓴다

② 한 지시변수의 기댓값과 $EX$를 구한다

③ 두 지시변수가 함께 1일 확률을 구한다

④ 공분산을 계산한다

⑤ 분산을 분산과 공분산의 합으로 쓰고 정리한다

**풀이**

① $X = \sum_{i=1}^{n} \mathbf{1}_i$. [T1]

② 대칭으로 $E\mathbf{1}_i = \frac{K}{N} = p$, $EX = np$. 선형성

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

④ $\operatorname{Cov}(\mathbf{1}_i, \mathbf{1}_j) = \frac{K}{N}\left(\frac{K-1}{N-1} - \frac{K}{N}\right) = \frac{K}{N} \cdot \frac{K-N}{N(N-1)} = -\frac{p(1-p)}{N-1}$. 계산

⑤ $\operatorname{Var} X = np(1-p) - n(n-1)\frac{p(1-p)}{N-1} = np(1-p)\frac{N-n}{N-1}$. 3회 정리 4.2(c)

### B5 무기억성의 역, 단계의 이름만 ★★

② · 정리 3.3 · 25분

정리 3.3의 역, 곧 무기억성을 갖는 양수인 연속 확률변수는 지수분포라는 것을 증명하라.

① 성질을 생존함수의 함수방정식으로 바꾼다

② 자연수에서 푼다

③ $c = G(1)$이 0과 1 사이에 있음을 보인다

④ 양의 유리수로 넓힌다

⑤ 실수로 넓힌다

⑥ 지수분포임을 결론짓는다

**풀이** 노트 정리 3.3의 역방향 증명과 같다. 단계별 근거는 B2의 풀이를 본다. 셋째 단계에서 $c > 0$은 첫 단계의 $G > 0$에서, $c < 1$은 귀류로 나온다. $c = 1$이면 $G(n) = 1$이 모든 $n$에서 성립해 $P(X > n) = 1$인데, $X$가 실숫값이면 $P(X > n) \to 0$이어야 한다.

### B6 베타와 감마, 전략만 ★★

③ · 정리 4.2 · 30분

정리 4.2를 증명하라. 결론에는 $U \perp V$, $U \sim \mathrm{Beta}(a,b)$, $V \sim \Gamma(a+b,1)$, $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$가 모두 들어가야 한다.

**전략** $(X,Y) \mapsto (U,V)$의 변수변환으로 결합밀도를 구하고, 그것이 $u$만의 함수와 $v$만의 함수의 곱으로 인수분해됨을 보인다.

**풀이** B3의 풀이와 노트 정리 4.2의 증명과 같다. 이 증명은 같은 틀로 세 개 이상의 감마로 넓어진다(노트 연습 4.5의 Dirichlet). 또 $R = X/Y = U/(1-U)$의 밀도를 구하면 노트 연습 4.4의 $F$ 분포의 씨앗이 된다.

### B7 초기하분포의 분산, 전략만 ★★

③ · 정리 1.5 · 30분

정리 1.5를 증명하라. 그리고 $n = N$이면 분산이 0이 되는 이유를 식과 말로 설명하라.

**전략** $X$를 지시변수의 합으로 쪼갠다. 지시변수들은 독립이 아니므로 분산에는 공분산 항이 남는다.

**풀이** B4의 풀이와 같다. $n = N$이면 $\frac{N-n}{N-1} = 0$이므로 분산이 0이다. 모집단 전체를 뽑으면 $X = K$로 확정되기 때문이다. 식으로는 $\mathrm{Var}\sum_i \mathbf{1}_i = Np(1-p) - N(N-1)\frac{p(1-p)}{N-1} = 0$이다. 분산 항의 합과 공분산 항의 합이 정확히 상쇄된다.

# C

# 증명 문제

노트의 정리를 변형하고, 일반화하고, 반례로 경계를 확인한다. 증명은 노트의 단계형 모양으로 쓴다. 한 줄에 한 단계, 줄마다 근거.

### C1 빈 상자의 수 ★★

[T1] · 정리 1.5와 같은 방법 · 25분

공 $n$개를 상자 $m$개에 하나씩 독립적으로, 고르게 넣는다. 빈 상자의 수를 $E$라 하자.

(a) $EE = m\left(1-\frac{1}{m}\right)^n$ 임을 보여라.

(b) $\operatorname{Var} E = m\left(1-\frac{1}{m}\right)^n + m(m-1)\left(1-\frac{2}{m}\right)^n - m^2\left(1-\frac{1}{m}\right)^{2n}$ 임을 보여라.

(c) $m=n=10$이면 $EE$는 얼마인가? (d) 그때 $\operatorname{Var} E$는 얼마인가?

**풀이** (a) $\mathbf{1}_j$를 상자 $j$가 비었다는 지시변수라 하면 $E=\sum_{j=1}^{m}\mathbf{1}_j$이다. 공 하나가 상자 $j$를 피할 확률은 $1-\frac{1}{m}$이고 공들은 독립이므로 $E\mathbf{1}_j=(1-\frac{1}{m})^n$이다. 선형성으로 $EE=m(1-\frac{1}{m})^n$이다.

(b) $j\neq k$이면 $E[\mathbf{1}_j\mathbf{1}_k]=(1-\frac{2}{m})^n$이다. $\operatorname{Var} E = E[E^2]-(EE)^2$이고

$$E[E^2]=\sum_j E\mathbf{1}_j+\sum_{j\neq k}E[\mathbf{1}_j\mathbf{1}_k]=m\left(1-\frac{1}{m}\right)^n+m(m-1)\left(1-\frac{2}{m}\right)^n$$

이므로 주어진 식이 나온다. 정리 1.5와 같이 지시변수들이 독립이 아니어서 교차항이 남는다.

(c) $10\cdot 0.9^{10}\approx 3.487$이다. 열 개를 열 상자에 넣으면 평균 세 개 반이 빈다. (d) $3.487+90\cdot 0.8^{10}-100\cdot 0.9^{20}\approx 3.487+9.664-12.158\approx 0.993$이다. 독립인 지시변수의 합이라면 분산은 $10\cdot 0.9^{10}(1-0.9^{10})\approx 2.27$이다. 음의 상관이 분산을 절반 아래로 줄인다.

**최종답** (c) $10(0.9)^{10}\approx 3.487$ (d) $\approx 0.993$

### C2 음이항의 Poisson 극한 ★★

[T5], [T8] · 정리 1.3, 2.3 · 20분

$\lambda>0$을 고정하고 $Y_r\sim \mathrm{NB}\left(r,\ p_r=\frac{r}{r+\lambda}\right)$라 하자.

(a) $EY_r=\lambda$이고 $\operatorname{Var} Y_r=\lambda+\lambda^2/r$임을 보여라.

(b) 고정된 $k$에서 $P(Y_r=k)\to e^{-\lambda}\lambda^k/k!$ $(r\to\infty)$임을 보여라.

(c) 노트 5.2절의 Poisson–감마 혼합으로 이 극한을 해석하라.

**풀이** (a) $q_r = \frac{\lambda}{r+\lambda}$이므로 $EY_r = \frac{rq_r}{p_r} = \lambda$, $\operatorname{Var} Y_r = \frac{EY_r}{p_r} = \lambda\frac{r+\lambda}{r} = \lambda + \frac{\lambda^2}{r}$이다.
(b) 정리 1.3의 pmf를 세 인수로 나눈다.

$$P(Y_r = k) = \frac{r(r+1)\cdots(r+k-1)}{(r+\lambda)^k} \cdot \frac{\lambda^k}{k!} \cdot \left(1+\frac{\lambda}{r}\right)^{-r}.$$

첫 인수는 $k$개의 인수 $\frac{r+j}{r+\lambda} \to 1$의 곱이므로 1로 간다($k$ 고정 [T5]). 셋째 인수는 $r\log(1+\lambda/r) \to \lambda$ [T8]이므로 $e^{-\lambda}$로 간다. 극한의 곱으로 결론이 나온다.
(c) 정리 5.3으로 $\mathrm{NB}(r, \frac{1}{1+\beta})$는 $\Lambda \sim \Gamma(r,\beta)$의 Poisson 혼합이다. 여기서 $\beta = \lambda/r$이므로 $E\Lambda = \lambda$, $\operatorname{Var}\Lambda = \lambda^2/r \to 0$이다. 위험률의 흩어짐이 사라지면 혼합은 하나의 $\mathrm{Poi}(\lambda)$가 되고, 과산포 항 $\lambda^2/r$도 0으로 간다.

### C3 여러 범주로 나누기 ★★

정리 2.5, 정의 1.7 · 20분

$N \sim \mathrm{Poi}(\lambda)$이고, $N = n$이 주어지면 각 사건이 서로 독립적으로 범주 $1,\dots,m$에 확률 $p_1,\dots,p_m$으로 나뉜다. 범주별 수를 $N_1,\dots,N_m$이라 하자.

(a) $N_1,\dots,N_m$이 상호 독립이고 $N_j \sim \mathrm{Poi}(\lambda p_j)$임을 보여라.

(b) $N$이 $\mathrm{Poi}(\lambda)$가 아니라 정해진 $n$이면 (a)는 어떻게 바뀌는가?

**풀이** (a) $s = a_1 + \cdots + a_m$이라 하면 곱셈법칙과 정의 1.7로

$$P(N_1 = a_1,\dots,N_m = a_m) = e^{-\lambda}\frac{\lambda^s}{s!} \cdot \frac{s!}{a_1!\cdots a_m!} p_1^{a_1}\cdots p_m^{a_m} = \prod_{j=1}^{m} e^{-\lambda p_j}\frac{(\lambda p_j)^{a_j}}{a_j!}.$$

$\sum_j p_j = 1$로 $e^{-\lambda} = \prod_j e^{-\lambda p_j}$를 썼다. 결합 pmf가 각 $a_j$만의 함수의 곱이고 각 인수가 $\mathrm{Poi}(\lambda p_j)$의 pmf이므로, 주변분포가 그것이고 상호 독립이다(3회 정리 3.2를 $m$개로). $m=2$가 정리 2.5다.
(b) 전체가 $n$으로 고정되면 $(N_1,\dots,N_m)$은 다항분포이고, 정리 1.8로 $\operatorname{Cov}(N_i,N_j) = -np_ip_j < 0$이다. 독립이 깨진다. Poisson의 "전체 수도 무작위"라는 성질이 범주들 사이의 독립을 만든다.

### C4 이항을 나누면 독립이 아니다 ★★

[T2] · 정리 2.5, 1.8 · 20분

정리 2.5에서 Poisson 가정이 필요함을 반례로 보인다. $N \sim \mathrm{b}(n,\theta)$이고 $N$이 주어지면 각 사건이 독립적으로 확률 $p$로 표시된다. 표시된 수를 $N_1$, 나머지를 $N_2 = N - N_1$이라 하자 ($q = 1-p$).

(a) $(N_1, N_2, n-N)$이 $\mathrm{Mult}(n;\theta p, \theta q, 1-\theta)$임을 보여라.

(b) $\operatorname{Cov}(N_1,N_2) = -n\theta^2 pq$임을 보여라.

(c) $n=10$, $\theta = p = \frac{1}{2}$이면 $\operatorname{Cov}(N_1,N_2)$는 얼마인가?

**풀이** (a) $a+b \le n$이면

$$P(N_1=a, N_2=b) = \binom{n}{a+b}\theta^{a+b}(1-\theta)^{n-a-b}\binom{a+b}{a}p^a q^b$$
$$= \frac{n!}{a!\,b!\,(n-a-b)!}(\theta p)^a(\theta q)^b(1-\theta)^{n-a-b}$$

이고, 이것은 $\mathrm{Mult}(n;\theta p, \theta q, 1-\theta)$의 pmf다.

(b) 정리 1.8로 $\mathrm{Cov}(N_1,N_2) = -n(\theta p)(\theta q) = -n\theta^2 pq$이다. $0<\theta, p<1$이면 0이 아니므로 $N_1$과 $N_2$는 독립이 아니다.

(c) $-10\cdot\frac{1}{4}\cdot\frac{1}{4} = -\frac{5}{8}$이다. $\theta\to 0$, $n\theta=\lambda$로 두면 Cov $\to 0$이고 이항이 Poisson이 되면서 독립이 회복된다(정리 2.3).

**최종답** (c) $-\frac{5}{8}$

### C5 경쟁하는 지수분포 여럿 ★★

정리 3.3 · 3회 연습 3.4 · 20분

$X_1,\dots,X_k$가 독립이고 $X_i \sim \mathrm{Exp}(\lambda_i)$이다. $L=\min_i X_i$, $J$를 최솟값을 갖는 첨자라 하고 $\Lambda=\sum_i \lambda_i$로 둔다.

(a) $L\sim\mathrm{Exp}(\Lambda)$임을 보여라.

(b) $P(L>t,\ J=i)=\frac{\lambda_i}{\Lambda}e^{-\Lambda t}$임을 보이고, $L$과 $J$가 독립임을 결론지어라.

(c) $\lambda=(1,2,3)$이면 $P(J=2)$는? (d) 그때 $EL$은?

**풀이** (a) $P(L>t)=\prod_i P(X_i>t)=e^{-\Lambda t}$이다(독립).

(b) $X_i$의 값으로 나눠 적분하면

$$P(L>t, J=i) = \int_t^\infty \lambda_i e^{-\lambda_i s}\prod_{j\neq i} e^{-\lambda_j s}\,ds = \lambda_i\int_t^\infty e^{-\Lambda s}ds = \frac{\lambda_i}{\Lambda}e^{-\Lambda t}.$$

$t=0$이면 $P(J=i)=\lambda_i/\Lambda$이고, 일반적인 $t$에서 $P(L>t,J=i)=P(L>t)P(J=i)$이므로 독립이다.

(c) $\frac{2}{1+2+3}=\frac{1}{3}$. (d) $L\sim\mathrm{Exp}(6)$이므로 $EL=\frac{1}{6}$. 누가 먼저 끝나는지는 언제 끝나는지에 대해 아무것도 알려 주지 않는다.

**최종답** (c) $\frac{1}{3}$ (d) $\frac{1}{6}$

### C6 Poisson을 가리는 항등식 ★★★

[T3] · 정의 2.1 · 25분

(a) $X \sim \mathrm{Poi}(\lambda)$이면 모든 유계함수 $g$에 대해 $E[\lambda g(X+1)] = E[Xg(X)]$임을 보여라.

(b) (a)가 $g(x) = x^2$에서도 성립한다고 받아들이고, $EX^3 = \lambda^3 + 3\lambda^2 + \lambda$임을 보여라.

(c) 거꾸로, 음이 아닌 정수값 확률변수 $X$가 모든 유계함수 $g$에 대해 (a)의 항등식을 만족하면 $X \sim \mathrm{Poi}(\lambda)$임을 보여라.

**풀이** (a) $E[Xg(X)] = \sum_{k\ge 1} k\, g(k) e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k\ge 1} g(k) e^{-\lambda}\frac{\lambda^k}{(k-1)!} = \lambda \sum_{j\ge 0} g(j+1) e^{-\lambda}\frac{\lambda^j}{j!} = E[\lambda g(X+1)]$. $g$가 유계이면 두 급수가 절대수렴한다.

(b) $EX^3 = E[X \cdot X^2] = \lambda E(X+1)^2 = \lambda(EX^2 + 2EX + 1) = \lambda(\lambda^2 + \lambda + 2\lambda + 1) = \lambda^3 + 3\lambda^2 + \lambda$이다. $EX^2 = \lambda + \lambda^2$은 정리 2.2다.

(c) $g = \mathbf{1}_{\{k+1\}}$이면 왼쪽은 $\lambda P(X=k)$, 오른쪽은 $(k+1)P(X=k+1)$이다. 따라서 $P(X=k+1) = \frac{\lambda}{k+1} P(X=k)$이고, $k$에 대한 귀납으로 $P(X=k) = \frac{\lambda^k}{k!} P(X=0)$이다 [T3]. 확률의 합이 1이므로 $P(X=0)\sum_k \frac{\lambda^k}{k!} = 1$, 곧 $P(X=0) = e^{-\lambda}$이다. 이 항등식이 Poisson 근사의 오차를 재는 Stein–Chen 방법의 출발점이다.

**최종답** (b) $\lambda^3 + 3\lambda^2 + \lambda$

### C7 섞으면 0이 늘어난다 ★★

[T4] · 정리 5.2 · 20분

$\Lambda \ge 0$이 $E\Lambda = \mu < \infty$인 확률변수이고, $\Lambda = \lambda$가 주어지면 $X \sim \mathrm{Poi}(\lambda)$이다.

(a) $\mathrm{Var}\, X \ge EX$이고, 등호는 $\mathrm{Var}\,\Lambda = 0$일 때뿐임을 보여라.

(b) $P(X=0) \ge e^{-\mu}$임을 보여라. 같은 평균의 Poisson보다 0이 많다.

(c) 노트 예제 6.3의 *may* 자료에서 0의 상대도수 $\frac{156}{262}$와 $e^{-0.656}$을 비교하고, (b)로 해석하라.

**풀이** (a) 정리 5.2로 $\mathrm{Var}\, X = E\Lambda + \mathrm{Var}\,\Lambda \ge E\Lambda = EX$이고, 등호는 $\mathrm{Var}\,\Lambda = 0$, 곧 $\Lambda$가 상수일 때뿐이다.

(b) 반복기댓값으로 $P(X=0) = E[P(X=0 \mid \Lambda)] = Ee^{-\Lambda}$이다. $e^{-x}$는 볼록이므로 Jensen 부등식으로 $Ee^{-\Lambda} \ge e^{-E\Lambda} = e^{-\mu}$이다. 등호는 역시 $\Lambda$가 상수일 때뿐이다.

(c) $\frac{156}{262} \approx 0.595$이고 $e^{-0.656} \approx 0.519$이다. 관측된 0의 비율이 같은 평균의 Poisson보다 크다. (b)에 따르면 이것은 조각마다 위험률이 다른 혼합이라는 모형과 맞는다. 노트 연습 6.3에서 0의 비율로 추정한 $\lambda$가 평균보다 작았던 것도 같은 현상이다.

### C8 음이항의 꼬리와 이항 ★

정리 1.3 · 15분

$Y \sim \mathrm{NB}(r, p)$이고 $B \sim \mathrm{b}(k+r, p)$이다.

(a) $P(Y \le k) = P(B \ge r)$임을 같은 사건을 두 가지로 읽어 보여라.

(b) $r = 3$, $p = \frac{1}{2}$, $k = 2$이면 $P(Y \le 2)$는 얼마인가?

**풀이** (a) 같은 독립 시행열에서 $\{Y \le k\}$는 $r$번째 성공이 $(k+r)$번째 시행까지 나온다는 사건이다. 그것은 처음 $k+r$번 안에 성공이 $r$번 이상이라는 사건과 같다. 처음 $k+r$번의 성공 수는 $\mathrm{b}(k+r, p)$이다.

(b) $P(\mathrm{b}(5, \frac{1}{2}) \ge 3) = \frac{10+5+1}{32} = \frac{1}{2}$이다. 대칭으로 $P(\mathrm{b}(5, \frac{1}{2}) \ge 3) = P(\mathrm{b}(5, \frac{1}{2}) \le 2)$이므로 $\frac{1}{2}$이다. 정리 3.2가 연속 시간에서 감마와 Poisson을 잇는 것처럼, 이 등식은 이산 시간에서 음이항과 이항을 잇는다.

**최종답** (b) $\frac{1}{2}$

### C9 사건 수를 알 때의 위치 ★★

정의 3.1 · 20분

비율 $\lambda$의 Poisson 과정에서 $0 < s < t$이다.

(a) $P(T_1 \le s \mid N(t) = 1) = \frac{s}{t}$임을 보여라. 사건이 하나 있었다는 것을 알면 그 위치는 $(0, t)$ 위의 균등분포다.

(b) $N(t) = n$이 주어지면 $N(s) \sim \mathrm{b}(n, s/t)$임을 보여라. $\lambda$가 사라진다.

(c) $t = 1$, $s = 0.3$, $n = 5$이면 $P(N(s) = 2 \mid N(t) = 5)$는?

**풀이** (a) $\{T_1 \le s,\ N(t) = 1\} = \{N(s) = 1,\ N(t) - N(s) = 0\}$이므로 정의 3.1로

$$P(T_1 \le s \mid N(t) = 1) = \frac{\lambda s e^{-\lambda s} \cdot e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t}.$$

(b) 같은 방법으로

$$P(N(s) = k \mid N(t) = n) = \frac{e^{-\lambda s}\frac{(\lambda s)^k}{k!} e^{-\lambda(t-s)}\frac{(\lambda(t-s))^{n-k}}{(n-k)!}}{e^{-\lambda t}\frac{(\lambda t)^n}{n!}} = \binom{n}{k}\left(\frac{s}{t}\right)^k\left(1 - \frac{s}{t}\right)^{n-k}.$$

노트 연습 2.4와 같은 계산이다. (c) $\binom{5}{2}(0.3)^2(0.7)^3 = 0.3087$.

**최종답** (c) $\binom{5}{2}(0.3)^2(0.7)^3 = 0.3087$

### C10 $\Gamma(\frac{1}{2})$를 베타로 ★★

[T7] · 정리 4.2 · 20분

가우스 적분을 쓰지 않고 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$를 증명한다.

(a) $x = \sin^2\theta$로 치환해 $B(\frac{1}{2},\frac{1}{2}) = \int_0^1 x^{-1/2}(1-x)^{-1/2}dx$를 계산하라.

(b) 정리 4.2로 $\Gamma(\frac{1}{2})^2 = B(\frac{1}{2},\frac{1}{2})\Gamma(1)$을 보이고 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$를 얻어라. 양의 제곱근을 고르는 근거는?

(c) 거꾸로 $\int_{-\infty}^{\infty} e^{-z^2/2}dz = \sqrt{2\pi}$를 (b)에서 얻어라.

**풀이** (a) $\theta \in (0, \frac{\pi}{2})$에서 $x = \sin^2\theta$는 증가하는 일대일 대응이고 $dx = 2\sin\theta\cos\theta\, d\theta$이므로 $B(\frac{1}{2},\frac{1}{2}) = \int_0^{\pi/2} \frac{2\sin\theta\cos\theta}{\sin\theta\cos\theta}d\theta = \pi$이다.

(b) 정리 4.2로 $B(\frac{1}{2},\frac{1}{2}) = \frac{\Gamma(1/2)^2}{\Gamma(1)}$이고 $\Gamma(1) = 1$이므로 $\Gamma(\frac{1}{2})^2 = \pi$이다. $\Gamma(\frac{1}{2})$는 양의 함수의 적분이므로 양수이고 $\sqrt{\pi}$이다. 정리 4.2의 증명은 3회 정리 5.1(변수변환)과 감마 밀도의 정규화만 썼고 가우스 적분은 쓰지 않았다.

(c) $x = z^2/2$로 치환하면 $\Gamma(\frac{1}{2}) = \sqrt{2}\int_0^{\infty} e^{-z^2/2}dz$이므로 $\int_0^{\infty} e^{-z^2/2}dz = \sqrt{\pi/2}$이고, 대칭으로 전체 적분은 $\sqrt{2\pi}$이다.

**최종답** (a) $\pi$

# D

# 퀄형

여러 절과 앞 회차의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 모형을 확률변수와 분포로 적는 것부터 답안의 일부다.

### D1 두 Poisson 과정의 경주 ★★★

정리 2.4, 3.2, 3.3 · 1회 정리 1.10 · 제한 40분

종류 A의 사건이 비율 $\lambda$, 종류 B의 사건이 비율 $\mu$인 독립인 Poisson 과정으로 일어난다.

(a) 두 과정을 합친 사건의 수가 비율 $\lambda+\mu$의 Poisson 과정임을 보여라.

(b) 첫 사건이 A일 확률이 $\frac{\lambda}{\lambda+\mu}$임을 보여라. 무기억성으로, 차례로 일어나는 사건들의 종류가 서로 독립이고 각각 확률 $\frac{\lambda}{\lambda+\mu}$로 A임을 설명하라.

(c) A가 $n$번 일어나는 것이 B가 $m$번 일어나는 것보다 먼저일 확률이 $P\big(\mathrm{b}(n+m-1, \frac{\lambda}{\lambda+\mu}) \geq n\big)$임을 보여라. 1회의 어떤 문제와 같은가?

(d) $\lambda=2$, $\mu=1$, $n=m=2$이면 이 확률은 얼마인가?

**풀이** (a) 겹치지 않는 구간들에서 합친 수는 각 과정의 수들의 함수이고, 두 과정이 독립이며 각 과정 안에서 구간들이 독립이므로 구간별 합친 수는 독립이다. 길이 $t$인 구간의 합친 수는 독립인 $\mathrm{Poi}(\lambda t)$와 $\mathrm{Poi}(\mu t)$의 합이므로 정리 2.4로 $\mathrm{Poi}((\lambda+\mu)t)$이다.

(b) 첫 A의 시각 $T^A \sim \mathrm{Exp}(\lambda)$와 첫 B의 시각 $T^B \sim \mathrm{Exp}(\mu)$는 독립이고(정리 3.2), $P(T^A < T^B) = \frac{\lambda}{\lambda+\mu}$이다(3회 연습 3.4). 첫 사건이 일어난 시각 이후로는 두 과정이 다시 처음처럼 시작한다. 지수분포의 무기억성(정리 3.3)으로 아직 오지 않은 쪽의 남은 대기시간도 같은 지수분포이기 때문이다. 그래서 다음 사건의 종류도 같은 확률로, 앞의 종류와 독립으로 정해진다.

(c) 사건들의 종류는 성공 확률 $\pi = \frac{\lambda}{\lambda+\mu}$인 독립 시행이다. "A가 $n$번 먼저"는 처음 $n+m-1$개의 사건 가운데 A가 $n$개 이상이라는 것과 같다. 그렇지 않으면 B가 $m$개 이상이기 때문이다. 이것은 1회 정리 1.10(점수 분배 문제의 Fermat 공식)과 같은 구조다. A가 $n$판, B가 $m$판을 더 이겨야 하는 상태의 승부다.

(d) $\pi=\frac{2}{3}$, $n+m-1=3$이므로 $P(\mathrm{b}(3,\frac{2}{3}) \geq 2) = 3\cdot\frac{4}{9}\cdot\frac{1}{3} + \frac{8}{27} = \frac{20}{27} \approx 0.741$이다.

**최종답** (d) $\frac{20}{27}$

### D2 사고를 잘 당하는 사람 ★★★

정리 4.5, 5.2, 5.3 · 3회 정리 2.5 · 제한 40분

Greenwood와 Yule(1920)의 모형이다. 한 노동자의 사고 위험률 $\Lambda$가 $\Gamma(r,\beta)$를 따르고, $\Lambda=\lambda$가 주어지면 첫해와 둘째 해의 사고 수 $X_1, X_2$는 독립인 $\mathrm{Poi}(\lambda)$이다.

(a) $\mathrm{Cov}(X_1,X_2)=r\beta^2$임을 보여라. $X_1$과 $X_2$는 독립인가?

(b) $X_1=k$가 주어지면 $\Lambda\sim\Gamma\big(r+k,\frac{\beta}{1+\beta}\big)$임을 보여라.

(c) $E(X_2\mid X_1=k)=\frac{(r+k)\beta}{1+\beta}$임을 보여라.

(d) $r=2,\beta=\frac12$이면 $E(X_2\mid X_1=0)$은? (e) $E(X_2\mid X_1=3)$은?

**풀이** (a) $\Lambda$가 주어지면 $X_1,X_2$는 독립이므로 $E(X_1X_2\mid\Lambda)=\Lambda^2$이고, 반복기댓값(3회 정리 2.5)으로 $E[X_1X_2]=E\Lambda^2$이다. $EX_i=E\Lambda$이므로 $\mathrm{Cov}(X_1,X_2)=E\Lambda^2-(E\Lambda)^2=\mathrm{Var}\,\Lambda=r\beta^2>0$이다. 독립이 아니다. 같은 사람의 두 해는 같은 $\Lambda$를 공유한다.

(b) $\lambda$의 함수로

$$g(\lambda)P(X_1=k\mid\lambda)\propto\lambda^{r-1}e^{-\lambda/\beta}\cdot e^{-\lambda}\lambda^k=\lambda^{r+k-1}e^{-\lambda(1+\beta)/\beta}$$

이고 이것은 $\Gamma(r+k,\frac{\beta}{1+\beta})$의 밀도에 비례한다. 정리 4.5와 같은 논리로 그것이 조건부 밀도다.

(c) $E(X_2\mid X_1=k)=E\big[E(X_2\mid\Lambda,X_1)\mid X_1=k\big]=E(\Lambda\mid X_1=k)=\frac{(r+k)\beta}{1+\beta}$이다. 첫째 등호는 반복기댓값이고, $\Lambda$가 주어지면 $X_2$가 $X_1$과 독립이므로 $E(X_2\mid\Lambda,X_1)=\Lambda$이다.

(d) $\frac{2\cdot\frac12}{\frac32}=\frac23$. (e) $\frac{5\cdot\frac12}{\frac32}=\frac53$. 전체 평균 $r\beta=1$을 사이에 두고, 사고가 없던 사람은 다음 해의 기대 사고가 $\frac23$으로, 세 번 사고를 낸 사람은 $\frac53$으로 예측된다. 위험률이 사람마다 다르면 과거가 미래를 예측한다. 이것이 "사고를 잘 당하는 사람"의 통계적 의미다.

**최종답** (d) $\frac23$ (e) $\frac53$

### D3 지수분포 표본의 평균 ★★★

정리 3.2, 3.7, 정의 3.8 · 제한 35분

$X_1,\dots,X_n$이 독립이고 평균 $\theta$인 지수분포를 따른다. $\bar X$는 표본평균이다.

(a) $\sum_i X_i\sim\Gamma(n,\theta)$임을 보여라.

(b) $\frac{2n\bar X}{\theta}\sim\chi^2(2n)$임을 보여라. 이 분포는 $\theta$에 의존하지 않는다.

(c) 정수 $n$과 $c>0$에서 $P\big(\chi^2(2n)\le 2c\big)=P\big(\mathrm{Poi}(c)\ge n\big)$임을 보여라.

(d) $n=5$이면 $P\big(\frac{2n\bar X}{\theta}\le 10\big)$은 얼마인가?

**풀이** (a) 평균 $\theta$인 지수분포는 $\Gamma(1,\theta)$이고, 같은 척도의 독립인 감마의 합은 정리 3.7로 $\Gamma(n,\theta)$이다.

(b) $W=\frac{2}{\theta}\sum X_i$의 mgf는 $Ee^{t\frac{2}{\theta}\sum X_i}=(1-\theta\cdot\frac{2t}{\theta})^{-n}=(1-2t)^{-n}$이고, 이것은 $\Gamma(n,2)=\chi^2(2n)$의 mgf다(정의 3.8). 유일성으로 $W\sim\chi^2(2n)$이다. $\theta$가 사라지는 것이 8회 이후에 $\theta$에 대한

구간을 만드는 열쇠가 된다.

(c) $\Gamma(n,2)$는 비율 $\frac{1}{2}$인 Poisson 과정의 $n$번째 사건의 시각이므로 정리 3.2로 $P(T_n \le 2c) = P(N(2c) \ge n)$이고 $N(2c) \sim \mathrm{Poi}(c)$이다.

(d) $2c = 10$, $c = 5$이므로 $P(\mathrm{Poi}(5) \ge 5) = 1 - e^{-5}\left(1 + 5 + \frac{25}{2} + \frac{125}{6} + \frac{625}{24}\right) = 1 - \frac{523}{8}e^{-5} \approx 0.5595$이다. 감마의 적분을 Poisson의 유한 합으로 계산했다.

**최종답** (d) $1 - \frac{523}{8}e^{-5} \approx 0.5595$

### D4 기병대의 사후분포 ★★★

정리 4.5, 5.3 · 예제 2.6 · 제한 40분

위험률 $\Lambda$의 사전분포가 $\Gamma(\alpha, \beta)$이고, $\Lambda = \lambda$가 주어지면 $X_1, \dots, X_n$이 독립인 $\mathrm{Poi}(\lambda)$이다. $s = \sum_i x_i$로 쓴다.

(a) 사후분포가 $\Gamma\left(\alpha + s,\ \frac{\beta}{1+n\beta}\right)$임을 보여라.

(b) 사후평균이 사전평균 $\alpha\beta$와 표본평균 $\bar{x}$의 가중평균임을 보이고 가중치를 구하라.

(c) 다음 관측 $X_{n+1}$의 예측분포가 음이항분포임을 정리 5.3으로 보여라.

(d) 예제 2.6의 기병대 자료($n = 200$, $s = 122$)에 사전분포 $\Gamma(1,1)$을 쓰면 사후평균은 얼마인가?

(e) 그때 다음 군단–해에 사망이 없을 예측확률은 얼마인가? $e^{-0.61}$과 비교하라.

**풀이** (a) $\lambda$의 함수로 $\lambda^{\alpha-1}e^{-\lambda/\beta} \cdot e^{-n\lambda}\lambda^{s} = \lambda^{\alpha+s-1}e^{-\lambda(1/\beta+n)}$이고, $\frac{1}{\beta} + n = \frac{1+n\beta}{\beta}$이므로 $\Gamma\left(\alpha+s, \frac{\beta}{1+n\beta}\right)$의 밀도에 비례한다.

(b) 사후평균은 $\frac{(\alpha+s)\beta}{1+n\beta} = \frac{1}{1+n\beta}\cdot\alpha\beta + \frac{n\beta}{1+n\beta}\cdot\bar{x}$이다. 표본평균의 가중치 $\frac{n\beta}{1+n\beta}$는 $n$이 커질수록 1로 간다.

(c) $\Lambda$의 사후분포가 주어지면 $X_{n+1} \mid \Lambda \sim \mathrm{Poi}(\Lambda)$이므로 예측분포는 $\Gamma(\alpha', \beta')$로 섞은 Poisson이고, 정리 5.3으로 $\mathrm{NB}(\alpha', \frac{1}{1+\beta'})$이다($\alpha' = \alpha + s$, $\beta' = \frac{\beta}{1+n\beta}$).

(d) $\alpha' = 123$, $\beta' = \frac{1}{201}$이므로 사후평균은 $\frac{123}{201} \approx 0.612$이다. 표본평균 0.61과 거의 같다.

(e) $P(X_{201} = 0) = \left(\frac{1}{1+\beta'}\right)^{\alpha'} = \left(\frac{201}{202}\right)^{123} \approx 0.5431$이다. $e^{-0.61} \approx 0.5434$와 거의 같다. 자료가 200개이면 $\lambda$의 불확실성이 작아서 음이항 예측이 Poisson 예측과 거의 같다. 자료가 적으면 예측분포의 0의 확률이 Poisson보다 커진다(C7).

**최종답** (d) $\frac{123}{201} \approx 0.612$ (e) $\left(\frac{201}{202}\right)^{123} \approx 0.5431$