수리통계학 × 통계학의 역사

5회 워크북 — 풀이판

정규분포와 Galton의 타원
손으로 푸는 34문항

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	정의와 공식을 손으로 돌린다. 답이 수나 식으로 나온다	14	2시간 2분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	3시간 20분
C 증명 문제	변형, 일반화, 반례 만들기	9	3시간 20분
D 퀄형	여러 절을 잇는 종합 문제. 제한 시간 안에	4	2시간 20분
	합계	34	약 11시간

노트 `05_normal_galton_notes.pdf` — 문항의 "정리 3.13" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T4] 부등식, [T5] 극한, [T6] 적분 순서의 교환, [T7] 변수변환, [T8] Taylor와 O

검산 노트북 `05_normal_galton.ipynb`의 "워크북 검산" 절: `check("A6a", "3/2")`

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.** 같은 정리를 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. Φ가 들어간 답은 $\Phi(z) = \frac{1}{2}\big(1 + \operatorname{erf}(z/\sqrt{2})\big)$로 쓰거나 표에서 읽은 소수로 넣는다.
5. **증명은 근거를 단다.** 등호와 부등호마다 어느 정의, 정리, 도구를 썼는지 오른쪽에 적는다. 행렬 계산은 블록의 크기를 먼저 적는다.

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

A

계산 드릴

정의와 공식을 손으로 돌린다. 답이 수나 식으로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다. 표준정규표가 필요한 곳에는 값을 문제에 적어 두었다.

A1 이항분포의 정규 근사 ★

정리 1.2, 따름정리 1.3 · 10분

$S \sim \mathrm{Bin}(400, \tfrac{1}{2})$이다. (a) 정리 1.2의 국소 근사로 $P(S=210)$의 근삿값을 φ로 쓰고 수치를 구하라 ($\varphi(1)=0.24197$). (b) 연속성 보정을 한 정규 근사로 $P(190 \le S \le 210)$을 Φ로 쓰고 수치를 구하라 ($\Phi(1.05)=0.85314$).

풀이 (a) $x_k = (210-200)/10 = 1$이므로 $P(S=210) \approx \varphi(1)/10 = e^{-1/2}/(10\sqrt{2\pi}) \approx 0.02420$이다. 정확한 값은 0.024207로 상대오차가 0.04%다.
(b) $P(190 \le S \le 210) \approx \Phi\big(\frac{210.5-200}{10}\big) - \Phi\big(\frac{189.5-200}{10}\big) = 2\Phi(1.05)-1 \approx 0.7063$이다. 정확한 값은 0.70629이고, 보정 없이 $2\Phi(1)-1=0.6827$을 쓰면 3%p 넘게 어긋난다.

최종답 (a) $\varphi(1)/10 \approx 0.02420$ (b) $2\Phi(1.05)-1 \approx 0.7063$

A2 mgf로 적률 구하기 ★

정리 1.8(ii) · 10분

$X \sim \mathcal{N}(\mu, \sigma^2)$이다. mgf $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$만 써서 (a) EX^3을 μ, σ로 나타내고, (b) $E(X-\mu)^6$을 구하라.

풀이 (a) $M' = (\mu+\sigma^2 t)M$, $M'' = \big[\sigma^2 + (\mu+\sigma^2 t)^2\big]M$, $M''' = \big[2\sigma^2(\mu+\sigma^2 t) + (\mu+\sigma^2 t)\{\sigma^2 + (\mu+\sigma^2 t)^2\}\big]M$이다. $t=0$에서 $EX^3 = 2\sigma^2\mu + \mu(\sigma^2+\mu^2) = \mu^3 + 3\mu\sigma^2$이다.
(b) $e^{\sigma^2 t^2/2} = \sum_k \frac{\sigma^{2k} t^{2k}}{2^k k!}$이고 t^6의 계수는 $\frac{\sigma^6}{8\cdot 6}$이다. 이것이 $E(X-\mu)^6/6!$이므로 $E(X-\mu)^6 = \frac{720}{48}\sigma^6 = 15\sigma^6$이다. $15 = 5\cdot 3\cdot 1$로 노트 연습 1.9의 $(2k-1)!!$와 같다.

최종답 (a) $\mu^3 + 3\mu\sigma^2$ (b) $15\sigma^6$

A3 두 정규의 제곱합 ★

따름정리 1.9 · 8분

$Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$이다. (a) $P(Z_1^2 + Z_2^2 > 4)$를 구하라. (b) $P(Z_1^2 + Z_2^2 \le c) = 0.9$인 c를 구하라.

풀이 따름정리 1.9로 $W = Z_1^2 + Z_2^2 \sim \chi^2(2)$이고, 밀도 $\frac{1}{2}e^{-w/2}$에서 $P(W > w) = e^{-w/2}$이다. (a) $e^{-2} \approx 0.1353$. (b) $e^{-c/2} = 0.1$에서 $c = 2\log 10 \approx 4.605$이다.

최종답 (a) $e^{-2} \approx 0.1353$ (b) $2\log 10 \approx 4.605$

A4 두 선형결합의 결합분포 ★

정리 3.8 · 8분

$\mathbf{X} \sim \mathcal{N}_3(\boldsymbol{\mu}, \Sigma)$이고 $Y_1 = X_1 + X_2$, $Y_2 = X_2 - X_3$이다.

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

(a) $\operatorname{Var} Y_1$, (b) $\operatorname{Var} Y_2$, (c) $\operatorname{Cov}(Y_1, Y_2)$, (d) $\operatorname{Corr}(Y_1, Y_2)$를 구하라. (Y_1, Y_2)의 분포는 무엇인가?

풀이 $C = \left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{smallmatrix}\right)$이면 $C\boldsymbol{\mu} = (3,2)'$이고

$$C\Sigma = \begin{pmatrix} 3 & 4 & 1 \\ 1 & 2 & 0 \end{pmatrix}, \qquad C\Sigma C' = \begin{pmatrix} 7 & 3 \\ 3 & 2 \end{pmatrix}.$$

(a) 7, (b) 2, (c) 3, (d) $3/\sqrt{14} \approx 0.802$. 정리 3.8로 $\mathbf{Y} \sim \mathcal{N}_2\big((3,2)',\ C\Sigma C'\big)$이다.

최종답 (a) 7 (b) 2 (c) 3 (d) $3/\sqrt{14} \approx 0.802$

A5 독립이 되는 계수 ★

정리 3.11 · 8분

(X_1, X_2)는 결합정규이고 $\operatorname{Var} X_1 = 4$, $\operatorname{Var} X_2 = 1$, $\operatorname{Cov}(X_1, X_2) = 1$이다. (a) $X_1 - aX_2$가 X_2와 독립이 되는 a를 구하라. (b) $X_1 + bX_2$가 $X_1 - 2X_2$와 독립이 되는 b를 구하라.

풀이 두 선형결합 (U, V)는 $\mathbf{X}$의 선형변환이므로 결합정규이고, 정리 3.11로 독립 $\iff$ $\operatorname{Cov}(U, V) = 0$이다. (a) $\operatorname{Cov}(X_1 - aX_2, X_2) = 1 - a = 0$에서 $a = 1$이다. 이것은 정리 C의 $B = \Sigma_{12}\Sigma_{22}^{-1}$이다. (b) $\operatorname{Cov}(X_1 + bX_2,\ X_1 - 2X_2) = 4 - 2 \cdot 1 + b \cdot 1 - 2b \cdot 1 = 2 - b = 0$에서 $b = 2$이다.

최종답 (a) $a = 1$ (b) $b = 2$

A6 이변량 조건부분포 ★

따름정리 3.19 · 8분

(X,Y)는 이변량 정규이고 $\mu_X=\mu_Y=0$, $\sigma_X=2$, $\sigma_Y=3$, $\rho=0.5$이다. (a) $E[Y\mid X=2]$, (b) $\mathrm{Var}(Y\mid X=2)$, (c) $E[X\mid Y=3]$을 구하라.

풀이 따름정리 3.19로 (a) $0.5\cdot\frac{3}{2}\cdot 2=\frac{3}{2}$, (b) $9(1-0.25)=\frac{27}{4}$, (c) $0.5\cdot\frac{2}{3}\cdot 3=1$이다. 표준화하면 $X=2$는 $z=1$, 조건부 평균 $\frac{3}{2}$는 Y의 단위로 0.5, 곧 ρz다. $Y=3$에서 X의 기대값 1도 X의 단위로 $0.5=\rho\cdot 1$이다. 양쪽 방향 모두 평균 쪽으로 당겨진다.

최종답 (a) $\frac{3}{2}$ (b) $\frac{27}{4}$ (c) 1

A7 등상관 세 변수 ★★

정리 3.13, 3.15 · 10분

$\mathbf{X}\sim\mathcal{N}_3(\mathbf{0},\Sigma)$, $\Sigma=\left(\begin{smallmatrix}2&1&1\\1&2&1\\1&1&2\end{smallmatrix}\right)$이다. $X_1\mid(X_2,X_3)=(1,2)$의 (a) 평균과 (b) 분산을 구하라. 정밀도행렬로 검산하라.

풀이 $\Sigma_{12}=(1,1)$이므로 $B=\Sigma_{12}\Sigma_{22}^{-1}=\frac{1}{3}(1,1)$이다. (a) $\mu_{1\cdot 2}=\frac{1}{3}(x_2+x_3)=1$. (b) $\Sigma_{11\cdot 2}=2-\frac{1}{3}(1+1)=\frac{4}{3}$.
검산: $\Lambda=\Sigma^{-1}=\frac{1}{4}\left(\begin{smallmatrix}3&-1&-1\\-1&3&-1\\-1&-1&3\end{smallmatrix}\right)$에서 정리 3.15로 $1/\Lambda_{11}=\frac{4}{3}$, $-\Lambda_{12}/\Lambda_{11}=-\Lambda_{13}/\Lambda_{11}=\frac{1}{3}$로 일치한다.

최종답 (a) 1 (b) $\frac{4}{3}$

A8 정밀도행렬에서 읽기 ★★

정리 3.15 · 10분

$\mathbf{X}\sim\mathcal{N}_3(\mathbf{0},\Sigma)$이고 정밀도행렬이 $\Lambda=\Sigma^{-1}=\left(\begin{smallmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{smallmatrix}\right)$이다. (a) $\mathrm{Var}(X_1\mid X_2,X_3)$, (b) $E[X_1\mid X_2=1,X_3=5]$, (c) $\mathrm{Cov}(X_1,X_3)$을 구하라. X_1과 X_3은 독립인가? X_2가 주어지면 어떤가?

풀이 정리 3.15로 (a) $1/\Lambda_{11}=\frac{1}{2}$, (b) $-\frac{\Lambda_{12}}{\Lambda_{11}}x_2-\frac{\Lambda_{13}}{\Lambda_{11}}x_3=\frac{1}{2}\cdot 1+0\cdot 5=\frac{1}{2}$이다. X_3의 값은 영향을 주지 않는다.
(c) $\Sigma=\Lambda^{-1}=\frac{1}{4}\left(\begin{smallmatrix}3&2&1\\2&4&2\\1&2&3\end{smallmatrix}\right)$이므로 $\mathrm{Cov}(X_1,X_3)=\frac{1}{4}\neq 0$이다. 결합정규에서 무상관이 아니므로 X_1과 X_3은 독립이 아니다. 그러나 $\Lambda_{13}=0$이므로 X_2가 주어지면 조건부 독립이다(노트 연습 3.11). 예제 3.16과 반대 구조다.

최종답 (a) $\frac{1}{2}$ (b) $\frac{1}{2}$ (c) $\frac{1}{4}$

A9 90% 확률 타원 ★

정리 3.17, 따름정리 3.18 · 8분

$\mathbf{X} \sim \mathcal{N}_2(\mathbf{0}, \Sigma)$, $\Sigma = \left(\begin{smallmatrix} 4 & 2 \\ 2 & 2 \end{smallmatrix}\right)$이다. 확률 0.9를 담는 타원 $\{\mathbf{x} : \mathbf{x}'\Sigma^{-1}\mathbf{x} \le c^2\}$의 (a) c^2과 (b) 넓이를 구하라.

풀이 (a) $1 - e^{-c^2/2} = 0.9$에서 $c^2 = 2\log 10 \approx 4.605$. (b) $|\Sigma| = 8 - 4 = 4$이므로 넓이는 $\pi \cdot 2\log 10 \cdot 2 = 4\pi \log 10 \approx 28.94$이다.

최종답 (a) $2\log 10 \approx 4.605$ (b) $4\pi \log 10 \approx 28.94$

A10 두 정규의 혼합 ★

명제 5.4 · 8분

확률 0.3으로 $\mathcal{N}(0,1)$, 확률 0.7로 $\mathcal{N}(4,4)$에서 뽑은 X의 (a) 평균과 (b) 분산을 구하라. 분산 가운데 집단 평균의 차이에서 오는 몫은 얼마인가?

풀이 (a) $0.3 \cdot 0 + 0.7 \cdot 4 = 2.8$. (b) $E[\mathrm{Var}(X \mid I)] = 0.3 \cdot 1 + 0.7 \cdot 4 = 3.1$, $\mathrm{Var}(E[X \mid I]) = 0.3 \cdot 0.7 \cdot 16 = 3.36$이므로 6.46이다. 평균 차이에서 오는 몫은 $3.36/6.46 = 52\%$다.

최종답 (a) 2.8 (b) 6.46

A11 오염된 정규의 첨도 ★

명제 5.3 · 10분

$\varepsilon = 0.05$, $\sigma_c = 5$인 오염된 정규 $f(x) = (1-\varepsilon)\varphi(x) + \frac{\varepsilon}{\sigma_c}\varphi(x/\sigma_c)$에서 (a) $\mathrm{Var}\,X$와 (b) 첨도 $EX^4/(\mathrm{Var}\,X)^2$을 구하라. 정규의 첨도 3과 비교하라.

풀이 (a) $\mathrm{Var}\,X = 0.95 + 0.05 \cdot 25 = 2.2$. (b) $EX^4 = 3(0.95 + 0.05 \cdot 625) = 3 \cdot 32.2 = 96.6$이므로 첨도는 $96.6/4.84 = \frac{2415}{121} \approx 19.96$이다. 자료의 5%만 표준편차가 5배인 집단에서 와도 첨도는 정규의 여섯 배를 넘는다.

최종답 (a) 2.2 (b) $\frac{2415}{121} \approx 19.96$

A12 시험과 재시험 ★★

따름정리 3.19 · 10분

실력 $T \sim \mathcal{N}(100, 10^2)$, 성적 $X = T + E_1$, $Y = T + E_2$, $E_1, E_2 \overset{\text{iid}}{\sim} \mathcal{N}(0, 5^2)$이 T와 독립이다. 첫 성적이 $X = 130$인 사람에 대해 (a) $E[Y \mid X = 130]$, (b) $\mathrm{Var}(Y \mid X = 130)$, (c) $P(Y > 130 \mid X = 130)$을 구하라 ($\Phi(0.894) = 0.8144$).

풀이 (X, Y)는 독립 정규 (T, E_1, E_2)의 선형변환이므로 결합정규이고 $\mathrm{Var}\,X = \mathrm{Var}\,Y = 125$, $\mathrm{Cov} = 100$, $\rho = 0.8$이다. (a) $100 + 0.8 \cdot 30 = 124$. (b) $125(1 - 0.64) = 45$. (c) $P\left(Z > \frac{130-124}{\sqrt{45}}\right) = 1 - \Phi(0.894) \approx 0.186$이다. 첫 시험에서 130점을 받은 사람의 81%는 재시험에서 그보다 낮은 점수를 받는다. 실력이 변하지 않아도 그렇다.

최종답 (a) 124 (b) 45 (c) $1 - \Phi(6/\sqrt{45}) \approx 0.186$

A13 사분면 확률 ★

노트 연습 3.4 · 6분

(X_1, X_2)가 표준화 이변량 정규이고 $\rho = 0.5$이다. (a) $P(X_1 > 0, X_2 > 0)$, (b) $P(X_1 > 0, X_2 < 0)$을 구하라.

풀이 (a) $\frac{1}{4} + \frac{\pi/6}{2\pi} = \frac{1}{4} + \frac{1}{12} = \frac{1}{3}$. (b) $P(X_1 > 0) - P(X_1 > 0, X_2 > 0) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$이다. $\rho = 0$이면 두 값이 모두 $\frac{1}{4}$이다.

최종답 (a) $\frac{1}{3}$ (b) $\frac{1}{6}$

A14 두 정규 사이의 KL ★

명제 6.2 · 8분

$\mathrm{KL}\big(\mathcal{N}(0,1)\,\|\,\mathcal{N}(1,4)\big)$를 구하라. 순서를 바꾼 $\mathrm{KL}\big(\mathcal{N}(1,4)\,\|\,\mathcal{N}(0,1)\big)$과 같은가?

풀이 $\frac{1}{2}\big[\frac{1}{4} + \frac{1}{4} - 1 + \log 4\big] = \log 2 - \frac{1}{4} \approx 0.443$이다. 순서를 바꾸면 $\frac{1}{2}[4 + 1 - 1 - \log 4] = 2 - \log 2 \approx 1.307$로 다르다. KL은 대칭이 아니다.

최종답 $\log 2 - \frac{1}{4} \approx 0.443$

B

빈칸 증명

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

B1 정리 C, 근거 채우기 ★★

① · 정리 3.13 · 20분

$\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')' \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, $\Sigma_{22} \succ 0$, $B = \Sigma_{12}\Sigma_{22}^{-1}$, $\mathbf{W} = \mathbf{X}_1 - B\mathbf{X}_2$이다. 각 단계의 근거를 써라.

① $(\mathbf{W}, \mathbf{X}_2)' = T\mathbf{X}$, $T = \left(\begin{smallmatrix} I & -B \\ 0 & I \end{smallmatrix}\right)$이므로 $(\mathbf{W}, \mathbf{X}_2)$는 결합정규다. — 정리 3.8 (정리 A)

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

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

④ $E\mathbf{W} = \boldsymbol{\mu}_1 - B\boldsymbol{\mu}_2$이고 $\mathrm{Cov}\,\mathbf{W} = \Sigma_{11} - B\Sigma_{21} - \Sigma_{12}B' + B\Sigma_{22}B' = \Sigma_{11\cdot 2}$이다. — 보조정리 3.1

⑤ $\mathbf{X}_1 = \mathbf{W} + B\mathbf{X}_2$이므로 $\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_1 + B(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11\cdot 2})$이다.

③: 조건을 걸어도 $\mathbf{W}$의 분포는 그대로

풀이 빈칸의 빨간 글씨가 근거다. ④의 마지막 등호는 $B\Sigma_{21} = \Sigma_{12}B' = B\Sigma_{22}B' = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$이어서 세 항 가운데 둘이 지워지기 때문이다. ⑤는 $\mathbf{X}_2 = \mathbf{x}_2$에서 $\mathbf{X}_1 = \mathbf{W} + B\mathbf{x}_2$이고, ③의 독립 때문에 $\mathbf{W}$의 조건부분포가 무조건부 분포 $\mathcal{N}(\boldsymbol{\mu}_1 - B\boldsymbol{\mu}_2, \Sigma_{11\cdot 2})$와 같다는 것이다.

B2 국소 de Moivre–Laplace, 근거 채우기 ★★

① · 정리 1.2 · 25분

$S_n \sim \mathrm{Bin}(n, p)$, $\delta = k - np = x_k\sqrt{npq}$, $|x_k| \le A$이다. 각 단계의 근거를 써라.

① $P(S_n = k) = \sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}\left(1 + O(n^{-1})\right)$이다. — 보조정리 1.1 (Stirling)

② $\sqrt{\frac{n}{2\pi k(n-k)}} = \frac{1}{\sqrt{2\pi npq}}\left(1 + O(n^{-1/2})\right)$이다. — $k(n-k) = n^2pq(1 + O(n^{-1/2}))$, $\delta = O(\sqrt{n})$

③ $u = \frac{\delta}{np}$, $v = \frac{\delta}{nq}$이면 $\log(\text{나}) = -k\log(1+u) - (n-k)\log(1-v)$이다. — 로그의 성질

④ $\log(1+w) = w - \frac{w^2}{2} + R(w)$, $|R(w)| \le |w|^3$을 넣는다. — Taylor 전개, $|w| \le \frac{1}{2}$ [T8]

⑤ $\log(\text{나}) = -\frac{\delta^2}{2n}\left(\frac{1}{p} + \frac{1}{q}\right) + O(n^{-1/2}) = -\frac{\delta^2}{2npq} + O(n^{-1/2})$이다. — $\pm\delta$ 소거, $\frac{1}{p} + \frac{1}{q} = \frac{1}{pq}$

⑥ $(\text{나}) = e^{-x_k^2/2}\left(1 + O(n^{-1/2})\right)$이고, 오차는 $|x_k| \le A$에서 균등하다.

$e^h = 1 + O(h)$, 상수가 A와 p에만 의존

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 1.2의 단계형 증명과 같다. ⑤의 나머지항은 $k|R(u)| \le k|u|^3 = O(n \cdot n^{-3/2})$와 $\delta^3/n^2 = O(n^{-1/2})$로 모두 $O(n^{-1/2})$이다.

B3 가우스 적분과 정규 mgf, 근거 채우기 ★★

① · 정리 1.8(i)(ii) · 20분

정리 1.8의 (i)과 (ii)의 증명이다. 각 단계의 근거를 써라.

① $I = \int e^{-z^2/2}dz$이면 $I^2 = \int_{\mathbb{R}^2} e^{-(z_1^2+z_2^2)/2}dz_1dz_2$이다. — Tonelli, 피적분함수 ≥ 0 [T6]

② $I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\, r\, dr\, d\varphi = 2\pi$이다. — 극좌표, 야코비안 r [T7]

③ $tx - \frac{(x-\mu)^2}{2\sigma^2} = -\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2} + \mu t + \frac{\sigma^2 t^2}{2}$이다. — 완전제곱

④ $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$이다. — ③, 평균 $\mu + \sigma^2 t$인 정규밀도의 적분, (i)

⑤ $EZ^4 = 4! \cdot \frac{1}{2^2\, 2!} = 3$이다. — $e^{t^2/2}$와 $\sum EZ^m t^m/m!$의 계수 비교

풀이 빈칸의 빨간 글씨가 근거다. ①의 Tonelli는 음이 아닌 함수의 반복적분은 순서와 상관없이 이중적분과 같다는 것이다. 피적분함수가 양수이므로 조건을 따로 확인할 필요가 없다. ⑤의 계수 비교가 정당한 것은 M_Z가 모든 t에서 유한해 거듭제곱급수로 전개되고, 그 계수가 $EZ^m/m!$이기 때문이다(보조정리 0.1(b)).

B4 정리 C, 단계의 이름만 ★★

② · 정리 3.13 · 30분

정리 3.13을 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **잔차 $\mathbf{W}$를 정의하고 $(\mathbf{W}, \mathbf{X}_2)$가 결합정규임을 보인다**

② **$\mathbf{W}$와 $\mathbf{X}_2$가 무상관임을 계산한다**

③ **결합정규에서 무상관은 독립임을 쓴다**

④ **$\mathbf{W}$의 평균과 공분산을 구한다**

⑤ **$\mathbf{X}_1 = \mathbf{W} + B\mathbf{X}_2$에서 조건부분포를 읽는다**

풀이 B1의 단계와 같다. ① $\mathbf{W} = \mathbf{X}_1 - B\mathbf{X}_2$, $B = \Sigma_{12}\Sigma_{22}^{-1}$이고 $(\mathbf{W}, \mathbf{X}_2) = T\mathbf{X}$이므로 정리 3.8로 결합정규. ② $\mathrm{Cov}(\mathbf{W}, \mathbf{X}_2) = \Sigma_{12} - B\Sigma_{22} = 0$. ③ 정리 3.11로 $\mathbf{W} \perp \mathbf{X}_2$. ④ $E\mathbf{W} = \boldsymbol{\mu}_1 - B\boldsymbol{\mu}_2$, $\mathrm{Cov}\,\mathbf{W} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ (보조정리 3.1). ⑤ $\mathbf{X}_2 = \mathbf{x}_2$이면 $\mathbf{X}_1 = \mathbf{W} + B\mathbf{x}_2$이고 $\mathbf{W}$의 분포는 조건에 영향받지 않으므로 $\mathcal{N}(\boldsymbol{\mu}_1 + B(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11\cdot 2})$.

B5 국소 de Moivre–Laplace, 단계의 이름만 ★★★

② · 정리 1.2 · 35분

정리 1.2를 증명하라. $\delta = k - np$로 두고 시작한다.

① **Stirling 공식으로 앞인수와 거듭제곱 부분으로 나눈다**

② **앞인수를 $\frac{1}{\sqrt{2\pi npq}}$로 근사한다**

③ **거듭제곱 부분의 로그를 u, v로 쓴다**

④ **$\log(1+w)$를 2차까지 전개하고 두 줄을 더한다**

⑤ **나머지항의 크기를 센다**

⑥ **지수를 취하고 균등성을 확인한다**

풀이 노트 정리 1.2의 단계형 증명과 같다. ④의 두 줄은

$$-\delta - \frac{\delta^2}{2np} + O\!\left(\frac{\delta^3}{n^2}\right), \qquad +\delta - \frac{\delta^2}{2nq} + O\!\left(\frac{\delta^3}{n^2}\right)$$

이고, 더하면 $-\frac{\delta^2}{2npq} + O(n^{-1/2})$이다. 균등성은 모든 $O(\cdot)$의 상수가 A와 p에만 의존한다는 것이다. 이 균등성이 따름정리 1.3(적분형)에서 합의 오차를 하나로 묶게 한다(노트 연습 1.13).

B6 정리 C, 전략만 ★★

③ · 정리 3.13 · 35분

정리 3.13을 두 가지로 증명하라.

전략 (증명 1) 잔차 $\mathbf{W} = \mathbf{X}_1 - B\mathbf{X}_2$를 $\mathbf{X}_2$와 무상관이 되게 만들고, 결합정규이므로 독립임을 쓴다. (증명 2, $\Sigma \succ 0$) 보조정리 3.3(e)의 이차형식 분해와 (b)의 행렬식 분해로 결합밀도를 "조건부밀도 × 주변밀도"로 쪼갠다.

풀이 증명 1은 B4의 풀이와 같다. 증명 2: $\tilde{\mathbf{x}}'\Sigma^{-1}\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2)'\Sigma_{11\cdot 2}^{-1}(\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2) + \tilde{\mathbf{x}}_2'\Sigma_{22}^{-1}\tilde{\mathbf{x}}_2$(보조정리 3.3(e))이고 $|\Sigma| = |\Sigma_{22}|\,|\Sigma_{11\cdot 2}|$(보조정리 3.3(b))이다. 정리 3.6의 밀도에 넣으면 $f(\mathbf{x})$가 $\mathcal{N}(\boldsymbol{\mu}_{1\cdot 2}, \Sigma_{11\cdot 2})$의 밀도와 $\mathcal{N}(\boldsymbol{\mu}_2, \Sigma_{22})$의 밀도의 곱이 된다. 둘째 인수는 따름정리 3.9의 주변밀도이므로 첫째 인수가 $f(\mathbf{x})/f_2(\mathbf{x}_2)$, 곧 조건부밀도다.

B7 가우스 적분과 정규 mgf, 전략만 ★★

③ · 정리 1.8(i)(ii) · 35분

정리 1.8의 (i)과 (ii)를 증명하라. (ii)에서는 $E(X-\mu)^4 = 3\sigma^4$까지 보인다.

전략 (i)은 적분을 제곱해 극좌표로 바꾼다. (ii)는 지수를 완전제곱해 (i)로 돌아가고, 적률은 mgf의 거듭제곱급수 계수로 읽는다.

풀이 B3의 풀이와 노트 정리 1.8의 증명과 같다. 단계는 (i) Tonelli [T6] → 극좌표 [T7] → $I^2 = 2\pi$ → 치환, (ii) 완전제곱 → (i)로 적분 → M_X → $M_Z(t) = e^{t^2/2} = \sum t^{2k}/(2^k k!)$의 t^4

> 계수 $\frac{1}{8} = EZ^4/4!$ 에서 $EZ^4 = 3$, 따라서 $E(X-\mu)^4 = 3\sigma^4$ 이다.

C

증명 문제

노트의 정리를 변형하고, 일반화하고, 반례로 경계를 확인한다. 증명은 노트의 단계형 모양으로 쓴다. 한 줄에 한 단계, 줄마다 근거. 적분과 극한의 순서를 바꾸는 곳에는 조건을 확인한다.

C1 Stein 항등식 ★★

[T6] · 정리 1.8 · 20분

$Z \sim \mathcal{N}(0,1)$이고 g는 미분가능하며 $E|g'(Z)| < \infty$, $\lim_{|z|\to\infty} g(z)\varphi(z) = 0$이다.

(a) $\varphi'(z) = -z\varphi(z)$와 부분적분으로 $E[Zg(Z)] = E[g'(Z)]$를 보여라.

(b) $g(z) = z^3$을 넣어 $EZ^4 = 3$을 다시 얻어라.

(c) $X \sim \mathcal{N}(\mu,\sigma^2)$이면 $E[(X-\mu)g(X)] = \sigma^2 E[g'(X)]$임을 보여라. $g(x) = e^{tx}$를 넣으면 무엇이 나오는가?

풀이 (a) 유한 구간 $[-R,R]$에서 $z\varphi(z) = -\varphi'(z)$이므로 부분적분으로

$$\int_{-R}^{R} g(z)\,z\varphi(z)\,dz = \Big[-g(z)\varphi(z)\Big]_{-R}^{R} + \int_{-R}^{R} g'(z)\varphi(z)\,dz.$$

가정으로 경계항은 $R\to\infty$에서 0이고, $E|g'(Z)| < \infty$이므로 오른쪽 적분은 $E\,g'(Z)$로 수렴한다(지배수렴 [T6], 지배함수 $|g'|\varphi$). 따라서 왼쪽도 수렴하고 $E[Zg(Z)] = E[g'(Z)]$이다.
(b) $g(z) = z^3$이면 $g\varphi \to 0$이고 $E|3Z^2| < \infty$이므로 $EZ^4 = E[3Z^2] = 3$이다.
(c) $X = \mu + \sigma Z$로 쓰고 $h(z) = g(\mu+\sigma z)$에 (a)를 쓰면 $E[(X-\mu)g(X)] = \sigma E[Zh(Z)] = \sigma E[h'(Z)] = \sigma^2 E[g'(X)]$이다. $g(x) = e^{tx}$이면 $E[(X-\mu)e^{tX}] = \sigma^2 t M_X(t)$, 곧 $M_X'(t) - \mu M_X(t) = \sigma^2 t M_X(t)$라는 미분방정식이고, $M_X(0) = 1$로 풀면 $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$이다.

C2 Box–Muller 변환 ★★

[T7] · 보조정리 0.2 · 25분

$U_1, U_2 \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$이고

$$Z_1 = \sqrt{-2\log U_1}\cos(2\pi U_2), \qquad Z_2 = \sqrt{-2\log U_1}\sin(2\pi U_2)$$

이면 $Z_1, Z_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$임을 보여라. (힌트를 보기 전에 극좌표 $R = \sqrt{-2\log U_1}$, $\Phi = 2\pi U_2$의 분포부터 구해 본다.)

풀이 $R = \sqrt{-2\log U_1}$이면 $r > 0$에서 $P(R > r) = P(-2\log U_1 > r^2) = P(U_1 < e^{-r^2/2}) = e^{-r^2/2}$이므로 $f_R(r) = re^{-r^2/2}$이다. $\Phi = 2\pi U_2 \sim \mathrm{U}(0,2\pi)$이고 $R \perp \Phi$이므로 $f_{R,\Phi}(r,\varphi) =$

$\frac{1}{2\pi} r e^{-r^2/2}$ 이다. $(z_1, z_2) = (r\cos\varphi, r\sin\varphi)$는 $(0,\infty)\times(0,2\pi)$에서 $\mathbb{R}^2$ (한 반직선을 뺀 곳)으로 가는 일대일 변환이고, 역변환의 야코비안은 $1/r$이다(노트 연습 0.1의 역수). 보조정리 0.2로

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{2\pi} r\, e^{-r^2/2} \cdot \frac{1}{r} = \frac{1}{2\pi} e^{-(z_1^2+z_2^2)/2} = \varphi(z_1)\varphi(z_2).$$

결합밀도가 곱으로 쪼개지므로 Z_1, Z_2는 독립 표준정규다. 빠진 반직선은 넓이가 0이므로 밀도에 영향이 없다.

C3 정규표본의 평균과 잔차 ★★

정리 3.8, 3.11 · 20분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이다.

(a) $(\bar X, X_1 - \bar X, \dots, X_n - \bar X)$가 결합정규임을 보여라.

(b) $\operatorname{Cov}(\bar X, X_i - \bar X) = 0$을 계산해 $\bar X$가 잔차벡터 $(X_1 - \bar X, \dots, X_n - \bar X)$와 독립임을 보여라.

(c) $\bar X \perp S^2 = \frac{1}{n-1}\sum (X_i - \bar X)^2$을 결론지어라. 어느 단계에서 정규 가정을 썼는가?

풀이 (a) $\mathbf{X} \sim \mathcal{N}_n(\mu\mathbf{1}, \sigma^2 I)$이고 $(\bar X, X_1 - \bar X, \dots)' = C\mathbf{X}$, C의 첫 행은 $\frac{1}{n}\mathbf{1}'$, 나머지 행은 $\mathbf{e}_i' - \frac{1}{n}\mathbf{1}'$이다. 정리 3.8로 결합정규다.

(b) $\operatorname{Cov}(\bar X, X_i) = \frac{1}{n}\sigma^2$, $\operatorname{Var}\bar X = \frac{\sigma^2}{n}$이므로 $\operatorname{Cov}(\bar X, X_i - \bar X) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$이다. (a)의 결합정규에서 첫 블록 $\bar X$와 둘째 블록(잔차벡터)의 교차 공분산이 0이므로 정리 3.11로 독립이다.

(c) S^2은 잔차벡터의 함수이므로 $\bar X \perp S^2$이다. 정규 가정은 (a)(결합정규)에서만 썼다. (b)의 공분산 계산은 분산이 같은 독립 표본이면 어떤 분포에서나 성립하지만, 무상관에서 독립으로 가는 데 결합정규가 필요하다.

C4 주변은 정규, 상관은 ρ, 결합정규는 아니다 ★★

[T2] · 정리 3.10 · 25분

(X, Y)를 확률 $\frac{1}{2}$로 상관 0.9인 표준화 이변량 정규에서, 확률 $\frac{1}{2}$로 상관 -0.1인 표준화 이변량 정규에서 뽑는다.

(a) X와 Y가 각각 $\mathcal{N}(0,1)$이고 $\operatorname{Corr}(X,Y) = 0.4$임을 보여라.

(b) $S = X + Y$의 분산과 4차 적률을 구하라.

(c) S의 첨도 $ES^4/(ES^2)^2$를 계산해 S가 정규가 아님을 보이고, (X,Y)가 결합정규가 아님을 결론지어라.

풀이 (a) 성분마다 X는 $\mathcal{N}(0,1)$이므로 혼합도 $\mathcal{N}(0,1)$이다($P(X\le x) = \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(x)$). Y도 같다. $E[XY] = \frac{1}{2}(0.9) + \frac{1}{2}(-0.1) = 0.4$이다.

(b) $S \mid I \sim \mathcal{N}(0, 2 + 2\rho_I)$, 곧 분산 3.8 또는 1.8이다. $ES^2 = \frac{1}{2}(3.8 + 1.8) = 2.8$, $ES^4 = \frac{1}{2}\cdot 3(3.8^2 + 1.8^2) = 26.52$이다.

(c) 첨도는 $26.52/2.8^2 = \frac{663}{196} \approx 3.383 \neq 3$이므로 S는 정규가 아니다. 결합정규라면 정리 3.10으로 $X+Y$가 정규여야 하므로 (X,Y)는 결합정규가 아니다. 주변분포와 상관계수만으로는 결합정규를 판정할 수 없다.

최종답 (c) $\frac{663}{196} \approx 3.383$

C5 편상관계수 ★★

정리 3.13 · 20분

(X_1, X_2, X_3)이 결합정규이고 분산이 모두 1, 상관이 $\rho_{12}, \rho_{13}, \rho_{23}$ ($|\rho_{13}|, |\rho_{23}| < 1$)이다. $X_3 = x_3$이 주어질 때 X_1과 X_2의 조건부 상관이

$$\rho_{12\cdot 3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}$$

임을 보이고, 이것이 x_3에 의존하지 않음을 확인하라. $\rho_{12} > 0$인데 $\rho_{12\cdot 3} < 0$인 예를 하나 들어라.

풀이 $\Sigma_{11} = \left(\begin{smallmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{smallmatrix}\right)$, $\Sigma_{12} = (\rho_{13}, \rho_{23})'$, $\Sigma_{22} = 1$이므로

$$\Sigma_{11\cdot 2} = \begin{pmatrix} 1-\rho_{13}^2 & \rho_{12}-\rho_{13}\rho_{23} \\ \rho_{12}-\rho_{13}\rho_{23} & 1-\rho_{23}^2 \end{pmatrix}.$$

조건부 상관은 이 행렬의 상관이므로 주어진 식이고, $\Sigma_{11\cdot 2}$가 x_3에 의존하지 않으므로(따름정리 3.14(b)) 조건부 상관도 그렇다. 예: $\rho_{12} = 0.2$, $\rho_{13} = \rho_{23} = 0.6$이면 $\rho_{12\cdot 3} = \frac{0.2 - 0.36}{0.64} = -0.25$이다. 공통 원인 X_3이 만든 양의 상관을 걷어 내면 음의 관계가 남는다.

C6 조건을 더 걸면 분산은 줄어든다 ★★

[T4] · 따름정리 3.14 · 15분

(X_1, X_2, X_3)이 결합정규이고 $\Sigma \succ 0$이다. $\operatorname{Var}(X_1 \mid X_2, X_3) \le \operatorname{Var}(X_1 \mid X_2) \le \operatorname{Var} X_1$임을 보여라. 첫 부등식이 등호일 필요충분조건은 무엇인가?

풀이 따름정리 3.14(c)를 (X_1, X_2)에 쓰면 $\operatorname{Var}(X_1 \mid X_2) \le \operatorname{Var} X_1$이다. 정리 3.13으로 $(X_1, X_3) \mid X_2 = x_2$는 공분산이 $S = \Sigma_{\{13\}\cdot 2}$인 결합정규이고 S는 x_2에 의존하지 않는다. 조건부밀도의 정의로 $f(x_1 \mid x_2, x_3) = f(x_1, x_3 \mid x_2)/f(x_3 \mid x_2)$이므로, $X_1 \mid (X_2, X_3)$은 분포 $(X_1, X_3) \mid X_2 = x_2$에서 다시 X_3에 조건을 건 것이다. 이 분포에 정리 3.13을 쓰면 $\operatorname{Var}(X_1 \mid X_2, X_3) = S_{11} - S_{13}^2/S_{33} \le S_{11} = \operatorname{Var}(X_1 \mid X_2)$이다. 등호는 $S_{13} = 0$, 곧 X_2가 주어질 때 X_1과 X_3이 무상관(결합정규이므로 조건부 독립)일 때다.

C7 정규 꼬리를 넘어선 양은 지수분포에 가깝다 ★★

[T5] · 노트 연습 1.11 · 25분

$Z \sim \mathcal{N}(0,1)$이다. 모든 $x > 0$에 대해

$$\lim_{z\to\infty} P\Big(Z > z + \frac{x}{z} \,\Big|\, Z > z\Big) = e^{-x}$$

임을 보여라. 곧 $z(Z-z)$는 $Z > z$일 때 근사적으로 평균 1인 지수분포다.

풀이 Mills 부등식 $\big(\frac{1}{z} - \frac{1}{z^3}\big)\varphi(z) \le 1-\Phi(z) \le \frac{\varphi(z)}{z}$에서 $1-\Phi(z) = \frac{\varphi(z)}{z}(1+O(z^{-2}))$이다. $w = z + x/z$로 두면

$$\frac{1-\Phi(w)}{1-\Phi(z)} = \frac{\varphi(w)/w}{\varphi(z)/z}\,\frac{1+O(w^{-2})}{1+O(z^{-2})}, \qquad \frac{\varphi(w)}{\varphi(z)} = e^{-(w^2-z^2)/2} = e^{-x-x^2/(2z^2)}.$$

$z/w \to 1$, $e^{-x^2/(2z^2)} \to 1$이므로 비는 e^{-x}로 간다 [T5]. 정규분포의 꼬리는 깊이 들어갈수록 "넘어선 양"의 척도가 $1/z$로 줄어드는 지수분포처럼 행동한다.

C8 이차형식이 χ^2이 되는 조건 ★★★

정리 3.17 · 30분

$\mathbf{X} \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$, $\Sigma \succ 0$, A는 대칭행렬이다. $A\Sigma A = A$이고 $\operatorname{rank} A = r$이면 $\mathbf{X}'A\mathbf{X} \sim \chi^2(r)$임을 보여라. $A = \Sigma^{-1}$이면 정리 3.17이 됨을 확인하라.

풀이 $\mathbf{Z} = \Sigma^{-1/2}\mathbf{X} \sim \mathcal{N}_n(\mathbf{0}, I)$이고 $\mathbf{X}'A\mathbf{X} = \mathbf{Z}'P\mathbf{Z}$, $P = \Sigma^{1/2}A\Sigma^{1/2}$이다. P는 대칭이고 $P^2 = \Sigma^{1/2}A\Sigma A\Sigma^{1/2} = \Sigma^{1/2}A\Sigma^{1/2} = P$이다. 대칭 멱등행렬의 고유값은 $\lambda^2 = \lambda$에서 0 또는 1이므로 스펙트럼 분해(보조정리 3.2)로 $P = UU'$, $U'U = I_r$이고 $r = \operatorname{rank} P = \operatorname{rank} A$이다($\Sigma^{1/2}$는 가역). $\mathbf{W} = U'\mathbf{Z} \sim \mathcal{N}_r(\mathbf{0}, U'U) = \mathcal{N}_r(\mathbf{0}, I_r)$(정리 3.8)이므로 $\mathbf{Z}'P\mathbf{Z} = \|\mathbf{W}\|^2 \sim \chi^2(r)$이다(따름정리 3.12, 1.9). $A = \Sigma^{-1}$이면 $A\Sigma A = \Sigma^{-1} = A$, $r = n$이다.

C9 축소의 이득과 손해 ★★

따름정리 2.2 · 20분

$X \mid \theta \sim \mathcal{N}(\theta, \sigma^2)$, $\theta \sim \mathcal{N}(0, \tau^2)$, $B = \tau^2/(\tau^2+\sigma^2)$이다.

(a) θ를 무작위로 보고 평균한 오차제곱 $E(X-\theta)^2$와 $E(BX-\theta)^2$을 구해 축소가 이득임을 보여라.

(b) θ를 고정하면 $E_\theta(BX-\theta)^2 = B^2\sigma^2 + (1-B)^2\theta^2$임을 보여라. 이것이 σ^2보다 커지는 $|\theta|$의 범위는?

풀이 (a) $E(X-\theta)^2 = E[E((X-\theta)^2 \mid \theta)] = \sigma^2$이다. 따름정리 2.2(같은 계산)로 $E[\theta \mid X] = BX$이고 $\operatorname{Var}(\theta \mid X) = \frac{\tau^2\sigma^2}{\tau^2+\sigma^2} = B\sigma^2$이므로 $E(BX-\theta)^2 = E[\operatorname{Var}(\theta \mid X)] = B\sigma^2 < \sigma^2$이다.
(b) $BX - \theta = B(X-\theta) - (1-B)\theta$이고 첫 항의 평균은 0이므로 $E_\theta(BX-\theta)^2 = B^2\sigma^2 + (1-B)^2\theta^2$이다. 이것이 σ^2보다 크다는 것은 $(1-B)^2\theta^2 > (1-B^2)\sigma^2$, 곧 $\theta^2 > \frac{1+B}{1-B}\sigma^2$이다. 사전분포가 말하는 범위 밖의 θ에서는 축소가 손해다. 평균한 위험은 줄지만 모든 θ에서 줄지는 않는다.

D

퀄형

여러 절의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 어떤 분포가 결합정규인지, 그 근거가 무엇인지 적는 것부터 답안의 일부다.

D1 이변량 정규의 최댓값 ★★★

정리 1.8, 3.8 · 제한 30분

(X, Y)는 표준화 이변량 정규이고 상관이 ρ ($|\rho| < 1$)이다. $M = \max(X, Y)$라 하자.

(a) $\max(x, y) = \frac{x+y}{2} + \frac{|x-y|}{2}$와 $E|N| = s\sqrt{2/\pi}$ ($N \sim \mathcal{N}(0, s^2)$)를 써서 $EM = \sqrt{(1-\rho)/\pi}$ 임을 보여라.

(b) (X, Y)와 $(-X, -Y)$가 같은 분포임을 써서 $EM^2 = 1$을 보이고 $\operatorname{Var} M$을 구하라.

(c) $\rho \to 1$과 $\rho = -1$ 근방에서 (a)(b)를 해석하라.

풀이 (a) $E\frac{X+Y}{2} = 0$이고 $X - Y = \mathbf{a}'(X, Y)'$, $\mathbf{a} = (1, -1)$이므로 정리 3.8로 $X - Y \sim \mathcal{N}(0, 2-2\rho)$이다. $N \sim \mathcal{N}(0, s^2)$이면 $E|N| = 2\int_0^\infty x \frac{1}{s\sqrt{2\pi}} e^{-x^2/(2s^2)} dx = s\sqrt{2/\pi}$이다. 따라서 $EM = \frac{1}{2}\sqrt{2-2\rho}\sqrt{2/\pi} = \sqrt{(1-\rho)/\pi}$이다.

(b) $m = \min(X, Y)$이면 $M^2 + m^2 = X^2 + Y^2$이므로 $EM^2 + Em^2 = 2$이다. $(-X, -Y)$는 (X, Y)와 같은 분포(평균 0, 공분산 같음)이고 $\max(-X, -Y) = -m$이므로 M과 $-m$은 같은 분포, 특히 $EM^2 = Em^2 = 1$이다. 따라서 $\operatorname{Var} M = 1 - \frac{1-\rho}{\pi}$이다.

(c) $\rho \to 1$이면 $X \approx Y$이므로 $M \approx X$, $EM \to 0$, $\operatorname{Var} M \to 1$이다. $\rho = -1$이면 $Y = -X$, $M = |X|$이고 $EM = \sqrt{2/\pi}$, $\operatorname{Var} M = 1 - 2/\pi$로 (a)(b)의 극한과 일치한다. $\rho = 0$이면 $EM = 1/\sqrt{\pi} \approx 0.564$이다.

최종답 (a) $\sqrt{(1-\rho)/\pi}$ (b) $1 - (1-\rho)/\pi$

D2 가운데 이항계수와 되돌아옴 ★★★

[T5], [T8] · 보조정리 1.1, 정리 1.2 · 제한 40분

공정한 동전으로 ± 1 걸음을 걷는 무작위 걸음 $W_m = \xi_1 + \cdots + \xi_m$을 생각한다.

(a) $P(W_{2n} = 0) = \binom{2n}{n} 4^{-n}$ 임을 보이고, 정리 1.2로 $P(W_{2n} = 0) \sim 1/\sqrt{\pi n}$을 보여라.

(b) Stirling 공식만으로 같은 결론을 다시 얻어라.

(c) 보조정리 1.1의 부등식에서 $m! = \sqrt{2\pi m}(m/e)^m e^{1/(12m) + O(m^{-2})}$이다. 이것으로 $P(W_{2n} = 0)\sqrt{\pi n} = 1 + c/n + O(n^{-2})$의 상수 c를 구하라.

(d) 원점으로 되돌아오는 횟수의 기댓값 $\sum_{n \ge 1} P(W_{2n} = 0)$이 발산함을 보여라.

풀이 (a) $W_{2n}=0$은 $+1$이 정확히 n번인 사건이므로 확률은 $\binom{2n}{n}2^{-2n}$이다. 정리 1.2를 $\mathrm{Bin}(2n,\tfrac12)$, $k=n$에 쓰면 $x_k=0$, $npq=\tfrac{n}{2}$이므로 $P\approx\frac{1}{\sqrt{2\pi\cdot n/2}}=\frac{1}{\sqrt{\pi n}}$이다.

(b) $\binom{2n}{n}=\frac{(2n)!}{(n!)^2}=\frac{\sqrt{4\pi n}(2n/e)^{2n}}{2\pi n(n/e)^{2n}}(1+o(1))=\frac{4^n}{\sqrt{\pi n}}(1+o(1))$이다.

(c) 지수 부분만 남기면 $\binom{2n}{n}4^{-n}\sqrt{\pi n}=\exp\!\left(\frac{1}{24n}-2\cdot\frac{1}{12n}+O(n^{-2})\right)=\exp\!\left(-\frac{1}{8n}+O(n^{-2})\right)=1-\frac{1}{8n}+O(n^{-2})$이다 [T8]. 따라서 $c=-\frac18$이다. $n=50$이면 $1-\frac{1}{400}=0.9975$이고, 실제 비는 $0.079589/0.079788=0.99750$이다.

(d) (a)로 충분히 큰 n에서 $P(W_{2n}=0)\ge\frac{1}{2\sqrt{\pi n}}$이고 $\sum 1/\sqrt{n}=\infty$이므로 비교 판정으로 발산한다. 되돌아온 횟수 $V=\sum_n \mathbf{1}\{W_{2n}=0\}$의 기댓값은 이 급수(음이 아닌 항의 합과 기댓값의 교환, [T6])이므로 무한대다. 1차원 대칭 무작위 걸음은 원점으로 무한히 자주 돌아올 수 있다.

최종답 (a) $1/\sqrt{\pi n}$ (c) $c=-\frac18$

D3 정규 계층모형의 예측분포 ★★★

정리 3.8, 3.13 · 제한 40분

$\theta\sim\mathcal{N}(\mu_0,\tau^2)$이고, θ가 주어지면 $X_1,\dots,X_n \overset{\text{iid}}{\sim}\mathcal{N}(\theta,\sigma^2)$이다.

(a) $(\theta,X_1,\dots,X_n)$이 결합정규임을 보이고 공분산행렬을 적어라.

(b) 정리 3.13을 $\mathbf{X}_1=\theta$, $\mathbf{X}_2=(X_1,\dots,X_n)$에 써라. $\Sigma_{22}=\sigma^2 I+\tau^2\mathbf{1}\mathbf{1}'$의 역행렬은 $\frac{1}{\sigma^2}\big(I-\frac{\tau^2}{\sigma^2+n\tau^2}\mathbf{1}\mathbf{1}'\big)$이다. 조건부분포가 표본에 $\bar x$를 통해서만 의존함을 보여라.

(c) 다음 관측 X_{n+1}의 예측분포 $X_{n+1}\mid X_1,\dots,X_n$을 구하라.

(d) $\mu_0=0$, $\tau^2=4$, $\sigma^2=1$, $n=4$, $\bar x=2.5$이면 사후평균 $E[\theta\mid\mathbf{x}]$는 얼마인가?

(e) 같은 수치에서 사후분산 $\mathrm{Var}(\theta\mid\mathbf{x})$는?

(f) 같은 수치에서 예측분산 $\mathrm{Var}(X_{n+1}\mid\mathbf{x})$는?

풀이 (a) $(\theta,E_1,\dots,E_n)$은 독립 정규이고 $(\theta,X_1,\dots,X_n)$은 그 선형변환이므로 결합정규다(정리 3.8). $\mathrm{Var}\,\theta=\tau^2$, $\mathrm{Cov}(\theta,X_i)=\tau^2$, $\mathrm{Cov}(X_i,X_j)=\tau^2+\sigma^2\mathbf{1}\{i=j\}$이다.

(b) $\Sigma_{12}=\tau^2\mathbf{1}'$이므로 $B=\Sigma_{12}\Sigma_{22}^{-1}=\frac{\tau^2}{\sigma^2}\big(1-\frac{n\tau^2}{\sigma^2+n\tau^2}\big)\mathbf{1}'=\frac{\tau^2}{\sigma^2+n\tau^2}\mathbf{1}'$이다. 따라서

$$E[\theta\mid\mathbf{x}]=\mu_0+\frac{n\tau^2}{\sigma^2+n\tau^2}(\bar x-\mu_0),\qquad \mathrm{Var}(\theta\mid\mathbf{x})=\tau^2-\frac{n\tau^4}{\sigma^2+n\tau^2}=\frac{\tau^2\sigma^2}{\sigma^2+n\tau^2}$$

이고 표본은 $B(\mathbf{x}-\mu_0\mathbf{1})=\frac{n\tau^2}{\sigma^2+n\tau^2}(\bar x-\mu_0)$을 통해서만 들어온다. 역행렬 공식은 Σ_{22}에 곱해 I가 되는지 확인한다.

(c) $X_{n+1}=\theta+E_{n+1}$이고 E_{n+1}은 $(\theta,\mathbf{X})$와 독립이므로, $\mathbf{X}=\mathbf{x}$가 주어지면 $X_{n+1}\sim\mathcal{N}\big(E[\theta\mid\mathbf{x}],\ \mathrm{Var}(\theta\mid\mathbf{x})+\sigma^2\big)$이다.

(d) $n\tau^2=16$, $\sigma^2+n\tau^2=17$이므로 사후평균은 $\frac{16}{17}\cdot 2.5=\frac{40}{17}\approx 2.353$이다. (e) 사후분산은 $\frac{4\cdot1}{17}=\frac{4}{17}\approx0.235$이다. (f) 예측분산은 $1+\frac{4}{17}=\frac{21}{17}\approx1.235$이다. 예측의 불확실성은 관측

잡음 σ^2보다 언제나 크다.

최종답 (d) $\frac{40}{17} \approx 2.353$ (e) $\frac{4}{17} \approx 0.235$ (f) $\frac{21}{17} \approx 1.235$

D4 선택된 집단의 회귀 ★★★

따름정리 3.19 · 제한 30분

(X, Y)는 표준화 이변량 정규이고 상관이 $\rho \in (0,1)$이다. $\lambda(c) = \varphi(c)/(1-\Phi(c))$라 하자.

(a) $E[X \mid X > c] = \lambda(c)$임을 보여라.

(b) $E[Y \mid X > c] = \rho\lambda(c)$임을 보여라.

(c) 모든 c에서 $E[Y - X \mid X > c] < 0$이고 $E[X - Y \mid Y > c] < 0$임을 보여라. 여는 질문의 두 문장과 연결하라.

(d) $\rho = 0.6$, $c = 1$이면 $E[Y \mid X > 1]$은 얼마인가? ($\lambda(1) = 1.5251$)

(e) 같은 수치에서 $E[X - Y \mid X > 1]$은?

풀이 (a) $\varphi'(x) = -x\varphi(x)$이므로 $E[X\mathbf{1}\{X > c\}] = \int_c^\infty x\varphi(x)dx = \varphi(c)$이고, $P(X > c) = 1 - \Phi(c)$로 나눈다.

(b) 따름정리 3.19로 $E[Y \mid X] = \rho X$이므로 반복기댓값으로 $E[Y\mathbf{1}\{X > c\}] = E[\mathbf{1}\{X > c\}E(Y \mid X)] = \rho E[X\mathbf{1}\{X > c\}]$이고, $P(X > c)$로 나누면 $\rho\lambda(c)$다.

(c) $E[Y - X \mid X > c] = (\rho - 1)\lambda(c) < 0$이다($\lambda(c) > 0$, $\rho < 1$). 대칭이므로 $E[X - Y \mid Y > c] = (\rho - 1)\lambda(c) < 0$이다. 키 큰 부모($X > c$)의 자녀는 부모보다 평균적으로 작고, 키 큰 자녀($Y > c$)의 부모도 자녀보다 평균적으로 작다. 두 문장은 같은 대칭 타원의 성질이다.

(d) $E[Y \mid X > 1] = 0.6 \cdot 1.5251 = 0.915$이다. (e) $E[X - Y \mid X > 1] = (1 - 0.6) \cdot 1.5251 = 0.610$이다. 선택된 집단은 다음 측정에서 평균 0.61 표준편차만큼 내려온다.

최종답 (d) $0.6\,\lambda(1) \approx 0.915$ (e) $0.4\,\lambda(1) \approx 0.610$