수리통계학 × 통계학의 역사

5회 워크북

정규분포와 Galton의 타원 손으로 푸는 34문항

강의노트 5회를 읽은 뒤 손으로 푼다. 다변량 정규의 계산은 블록 행렬, 완전제곱, mgf 세 가지로 거의 모두 끝난다. 이 워크북은 그 세 가지를 손에 익히고, 정리 C와 de Moivre–Laplace 정리의 증명을 빈 종이에 다시 쓰게 한다. 문제 아래의 줄 친 공간에 직접 쓴다.

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.

예: 빈칸과 풀이 공간

① $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$이면 $C\mathbf{X} \sim \mathcal{N}(C\boldsymbol{\mu}, C\Sigma C')$이다. 왜? _____

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

$S \sim \mathrm{Bin}(400, \frac{1}{2})$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 ($\Phi(1.05) = 0.85314$).

A2 mgf로 적률 구하기 ★

정리 1.8(ii) · 10분

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

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.

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?

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.

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.

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.

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?

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.

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?

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) $\operatorname{Var} X$와 (b) 첨도 $EX^4/(\operatorname{Var} X)^2$을 구하라. 정규의 첨도 3과 비교하라.

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 ($\Phi(0.894) = 0.8144$).

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.

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?

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.

왜? _____

② $\operatorname{Cov}(\mathbf{W}, \mathbf{X}_2) = \Sigma_{12} - B\Sigma_{22} = 0$ 이다. 왜? _____

③ $\mathbf{W} \perp \mathbf{X}_2$ 이다. 왜? _____

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

왜? _____

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.

왜? _____

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. 왜? _____

② $\sqrt{\frac{n}{2\pi k(n-k)}} = \frac{1}{\sqrt{2\pi npq}}\left(1+O(n^{-1/2})\right)$이다. 왜? _____

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

④ $\log(1+w) = w - \frac{w^2}{2} + R(w)$, $|R(w)| \le |w|^3$을 넣는다. 왜? _____

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. 왜? _____

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

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. 왜? _____

② $I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\,r\,dr\,d\varphi = 2\pi$이다. 왜? _____

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. 왜? _____

④ $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$이다. 왜? _____

⑤ $EZ^4 = 4! \cdot \frac{1}{2^2\,2!} = 3$이다. 왜? _____

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

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

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.

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.

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?

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.)

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) $\mathrm{Cov}(\bar X, X_i - \bar X) = 0$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?

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 $\mathrm{Corr}(X,Y) = 0.4$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.

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.

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?

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

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

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

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.

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?

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.

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.

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?

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?

힌트와 최종답

막히면 먼저 힌트를 보고, 다 푼 뒤에 최종답과 맞춰 본다. 그래도 어긋나면 풀이판을 본다. 최종답은 노트북의 `check()`로도 확인할 수 있다.

A1 힌트 $np = 200$, $\sqrt{npq} = 10$이다. (b)는 정수 $190, \dots, 210$이 실수축에서 $[189.5,\ 210.5]$를 차지한다는 것을 쓴다.

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

A2 힌트 (a)는 M_X를 세 번 미분해 $t = 0$을 넣는다. (b)는 $X - \mu$의 mgf $e^{\sigma^2 t^2/2}$를 거듭제곱급수로 펼쳐 t^6의 계수를 읽는 편이 빠르다.

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

A3 힌트 $\chi^2(2) = \Gamma(1, 2)$는 평균 2인 지수분포다. 꼬리는 $P(W > w) = e^{-w/2}$이다.

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

A4 힌트 $\mathbf{Y} = C\mathbf{X}$, $C = \left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{smallmatrix}\right)$로 쓰고 $C\Sigma C'$를 계산한다.

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

A5 힌트 두 선형결합은 함께 결합정규이므로(정리 3.8) 공분산이 0이면 독립이다(정리 3.11).

답 (a) $a = 1$ (b) $b = 2$

A6 힌트 $E[Y \mid X = x] = \rho\frac{\sigma_Y}{\sigma_X}x$, $\mathrm{Var}(Y \mid X) = \sigma_Y^2(1 - \rho^2)$. (c)는 X와 Y의 자리를 바꾼다.

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

A7 힌트 $\Sigma_{22} = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$의 역행렬은 $\frac{1}{3}\left(\begin{smallmatrix} 2 & -1 \\ -1 & 2 \end{smallmatrix}\right)$이다. $\Sigma^{-1} = \frac{1}{4}\left(\begin{smallmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{smallmatrix}\right)$.

답 (a) 1 (b) $\frac{4}{3}$

A8 힌트 (a)(b)는 정리 3.15로 Λ의 첫 행만 쓴다. (c)는 $\Sigma = \Lambda^{-1}$이 필요하다. $\det\Lambda = 4$.

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

A9 힌트 $n = 2$이면 $P(Q \le c^2) = 1 - e^{-c^2/2}$. 넓이는 $\pi c^2\sqrt{|\Sigma|}$.

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

A10 힌트 집단 지시변수 I에 대한 전분산 공식: 집단 안의 분산의 평균 + 집단 평균의 분산.

답 (a) 2.8 (b) 6.46

A11 힌트 $EX^2 = (1-\varepsilon) + \varepsilon\sigma_c^2$, $EX^4 = 3\big[(1-\varepsilon) + \varepsilon\sigma_c^4\big]$.

답 (a) 2.2 (b) $\dfrac{2415}{121} \approx 19.96$

A12 힌트 $\operatorname{Var} X = \operatorname{Var} Y = 125$, $\operatorname{Cov}(X,Y) = 100$, 상관 $\rho = 0.8$.

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

A13 힌트 노트 연습 3.4의 식 $\frac{1}{4} + \frac{\arcsin\rho}{2\pi}$. (b)는 $P(X_1 > 0) = \frac{1}{2}$에서 뺀다.

답 (a) $\frac{1}{3}$ (b) $\frac{1}{6}$

A14 힌트 명제 6.2에서 $n=1$: $\frac{1}{2}\big[\sigma_0^2/\sigma_1^2 + (\mu_1-\mu_0)^2/\sigma_1^2 - 1 + \log(\sigma_1^2/\sigma_0^2)\big]$.

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

B1 힌트 ③의 근거에는 "결합정규" 라는 가정이 들어간다. 무상관만으로는 독립이 나오지 않는다(노트 5.1절).

B2 힌트 ②에서 $(1+h)^{-1/2} = 1 + O(h)$를 쓴다. ④의 조건 $|u|, |v| \le \frac{1}{2}$은 n이 충분히 크면 성립한다.

B3 힌트 ①에서 $I^2 = \big(\int e^{-z_1^2/2}dz_1\big)\big(\int e^{-z_2^2/2}dz_2\big)$을 이중적분으로 바꾸는 근거가 필요하다. 이상적분이므로 순서를 바꾸는 정리를 쓴다.

B4 힌트 B1에서 채운 근거를 다시 보지 말고 쓴다. $\operatorname{Cov}\mathbf{W}$를 계산할 때 네 항 가운데 셋이 같은 행렬이다.

B5 힌트 ④에서 $-(np+\delta)\big(\frac{\delta}{np} - \frac{\delta^2}{2n^2p^2}\big)$을 전개하면 δ^2 항이 두 번 나온다. 둘을 합치면 $-\frac{\delta^2}{2np}$이다.

B6 힌트 증명 2에서는 $\tilde{\mathbf{x}} = \mathbf{x} - \boldsymbol{\mu}$로 두고 $\tilde{\mathbf{x}}_1 - B\tilde{\mathbf{x}}_2 = \mathbf{x}_1 - \boldsymbol{\mu}_{1\cdot 2}$임을 확인한다.

B7 힌트 (i)의 마지막에 $z = (x-\mu)/\sigma$로 치환해 $\int f = 1$까지 보인다. (ii)의 4차 적률은 $Z = (X-\mu)/\sigma$로 표준화해 $E(X-\mu)^4 = \sigma^4 EZ^4$로 쓴다.

C1 힌트 (a) $\int_{-R}^{R} g(z)\,z\varphi(z)\,dz = \big[-g\varphi\big]_{-R}^{R} + \int_{-R}^{R} g'\varphi$로 유한 구간에서 부분적분한 뒤 $R\to\infty$로 보낸다. 극한과 적분의 교환에는 $E|g'(Z)| < \infty$가 쓰인다.

C2 힌트 (R,Φ)는 독립이고 $\Phi \sim \mathrm{U}(0,2\pi)$, $P(R>r) = P(U_1 < e^{-r^2/2}) = e^{-r^2/2}$이다. 그다음

극좌표에서 직교좌표로의 변수변환(노트 연습 0.1의 야코비안 r)을 쓴다.

C3 힌트 (a) $(n+1)$차원 벡터가 $\mathbf{X}$의 선형변환 $C\mathbf{X}$다. (c) 독립인 두 확률벡터의 함수는 독립이다.

C4 힌트 성분 지시변수 I에 조건을 걸면 $S \mid I \sim \mathcal{N}(0, 2+2\rho_I)$이다. $ES^4 = E\big[3(2+2\rho_I)^2\big]$.

답 (c) $\frac{663}{196} \approx 3.383$

C5 힌트 정리 3.13을 $\mathbf{X}_1 = (X_1, X_2)$, $\mathbf{X}_2 = X_3$에 쓴다. $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$에서 $\Sigma_{22} = 1$이다.

C6 힌트 X_2에 대한 조건부분포 $(X_1, X_3) \mid X_2$도 결합정규이다(정리 3.13). 그 분포에 따름정리 3.14(c)를 다시 쓴다.

C7 힌트 노트 연습 1.11의 Mills 부등식으로 $1-\Phi(z) = \frac{\varphi(z)}{z}(1+O(z^{-2}))$이다. 분자와 분모에 쓰고 $\varphi(z+x/z)/\varphi(z)$를 계산한다.

C8 힌트 $P = \Sigma^{1/2} A \Sigma^{1/2}$이 대칭이고 $P^2 = P$임을 보인다. 대칭 멱등행렬은 $P = UU'$, $U'U = I_r$로 쓸 수 있다(스펙트럼 분해에서 고유값이 0 또는 1).

C9 힌트 (a)는 (θ, X)의 결합정규에서 $\mathrm{Var}(\theta \mid X)$가 $B\sigma^2$이고 $E[\theta \mid X] = BX$임을 쓴다. (b)는 $BX - \theta = B(X-\theta) - (1-B)\theta$로 쪼갠다.

D1 힌트 (a) $X - Y \sim \mathcal{N}(0, 2-2\rho)$(정리 3.8). (b) $M^2 + m^2 = X^2 + Y^2$ $(m = \min(X,Y))$이고, $(X,Y) \mapsto (-X,-Y)$는 M을 $-m$으로 보낸다.

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

D2 힌트 (a) $W_{2n} = 0$은 $2n$걸음 가운데 정확히 n걸음이 $+1$인 사건이다. $\mathrm{Bin}(2n, \frac{1}{2})$에서 $\sqrt{npq} = \sqrt{n/2}$, $x_k = 0$. (c) $\exp\!\big(\frac{1}{24n} - \frac{2}{12n}\big)$을 1차까지 전개한다. (d) $\sum 1/\sqrt{n} = \infty$와 비교 판정 [T5].

답 (a) $1/\sqrt{\pi n}$ (c) $c = -\frac{1}{8}$

D3 힌트 (a) $X_i = \theta + E_i$, $E_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ 독립. (c) $X_{n+1} = \theta + E_{n+1}$이고 E_{n+1}은 $(\theta, X_1, \dots, X_n)$과 독립이다.

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

D4 힌트 (a) $\int_c^\infty x\varphi(x)\,dx = \varphi(c)$. (b) 반복기댓값: $E[Y \mid X] = \rho X$이므로 $E[Y\mathbf{1}\{X > c\}] = \rho E[X\mathbf{1}\{X > c\}]$.

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