수리통계학 × 통계학의 역사

10회 워크북

숨은 변수와 EM
손으로 푸는 34문항

강의노트 10회를 읽은 뒤 손으로 푼다. 정보 행렬의 블록 계산, E 단계와 M 단계의 한 바퀴, 잃어버린 정보의 비율은 눈으로 따라간 것과 빈 종이에 다시 쓰는 것이 다르다. 문제 아래의 줄 친 공간에 직접 쓴다. 반복을 여러 번 돌려야 하는 문항은 처음 한두 번만 손으로 하고 나머지는 노트북으로 확인한다.

부	하는 일	문항	예상 시간
A 계산 드릴	정보 행렬, Schur 보수, E·M 단계를 손으로 계산한다	14	2시간
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 50분
C 증명 문제	변형, 일반화, 반례 만들기	9	2시간 35분
D 퀄형	여러 절을 잇는 종합 문제. 제한 시간 안에	4	2시간 30분
	합계	34	약 10시간

노트 `10_em_notes.pdf` — 문항의 "정리 3.2" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T1] 지시함수, [T2] 반례, [T4] 부등식·완전제곱, [T5] 급수, [T6] 미분과 적분의 교환, [T8] Taylor 전개

검산 노트북 `10_em.ipynb`의 "워크북 검산" 절: `check("A3a", "3/8")`

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. 반복을 수렴할 때까지 돌려야 하는 값은 소수 넷째 자리까지 넣으면 된다.
5. **증명은 근거를 단다.** 등호와 부등호마다 어느 정의, 정리, 앞 회차의 결과를 썼는지 오른쪽에 적는다. 미분과 적분의 순서를 바꾸는 곳에는 조건을 적는다.

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

예: 빈칸과 풀이 공간

① $D(p\,\|\,q) \geq 0$이다. 왜? _____

A

계산 드릴

정의와 공식을 손으로 돌린다. 답이 수나 식으로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다. 계산기는 마지막 수치 계산에만 쓴다.

A1 정규의 정보 행렬 ★

정의 1.1, 정리 1.2 · 6분

$X \sim \mathcal{N}(\mu, \sigma^2)$를 모수 (μ, υ), $\upsilon = \sigma^2$로 둔다. 정리 1.2의 둘째 표현으로 정보 행렬을 구하라. (a) $\upsilon = 4$에서 $I_{\upsilon\upsilon}$를 구하라. (b) $n = 50$이면 $\hat{\upsilon}$의 점근분산 $(I^{-1})_{\upsilon\upsilon}/n$은 얼마인가?

A2 로짓 모수의 정보 ★

정리 1.2 · 6분

$X \sim \text{Bernoulli}(p)$를 로짓 $\eta = \log \frac{p}{1-p}$로 모수화한다. $I(\eta) = p(1-p)$임을 보이고, $p = 0.2$, $n = 100$일 때 $\hat{\eta}$의 점근분산 $1/(nI(\eta))$을 구하라.

A3 두 모수의 Schur 보수 ★

보조정리 1.5, 정리 1.6 · 8분

정보 행렬이 $I = \left(\begin{smallmatrix} 4 & 2 \\ 2 & 3 \end{smallmatrix}\right)$이고 첫 성분 ψ가 관심 모수, 둘째 성분 λ가 귀찮은 모수다. (a) $I^{\psi\psi}$, (b) $I_{\psi\psi}^{-1}$, (c) 둘의 비 $I^{\psi\psi} / I_{\psi\psi}^{-1}$를 구하라. (c)는 무엇을 뜻하는가?

A4 귀찮은 모수가 둘 ★★

보조정리 1.5 · 10분

정보 행렬이 다음과 같고 첫 성분이 관심 모수 ψ, 나머지 둘이 귀찮은 모수다.

$$I = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 4 \end{pmatrix}$$

(a) Schur 보수 $S = I_{\psi\psi} - I_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi}$를 구하라. (b) $I^{\psi\psi}$를 구하라. (c) 귀찮은 모수를 알 때보다 점근분산이 몇 배인가?

A5 정규 분위수의 점근분산 ★★

따름정리 1.4(b) · 10분

$X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, v)$, $v = \sigma^2$에서 분위수 $g = \mu + z\sigma$ (z는 상수)의 MLE $\hat{\mu} + z\hat{\sigma}$를 생각한다. 따름정리 1.4(b)로 $\sqrt{n}(\hat{g} - g)$의 점근분산을 v와 z의 식으로 구하라. (`check()`에는 기호 `v`와 `z`로 넣는다.)

A6 분산의 프로파일 가능도비 ★★

정의 1.7, 정리 1.8 · 10분

정규 표본 $n=20$에서 $\hat{v}=\frac{1}{n}\sum(x_i-\bar{x})^2=1.5$이다. μ를 귀찮은 모수로 두고 $H_0: v=1$의 프로파일 가능도비 $\Lambda=2[\ell(\hat{\theta})-\ell_p(1)]$를 구하라. $\chi^2(1)$의 5% 임계값 3.84와 비교하라.

A7 다른 유전 연관 자료 ★★

예제 2.3, 알고리즘 3.1 · 12분

노트 예제 2.3과 같은 모형(범주 확률 $\frac{1}{2}+\frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}$)에서 3,839개체가 $\mathbf{y}=(1997, 906, 904, 32)$로 나뉘었다. (a) $\theta_0=\frac{1}{20}$에서 EM을 한 번 돌린 θ_1을 분수로 구하라. (b) 가능도 방정식을 이차방정식으로 바꿔 MLE를 구하라.

A8 잃어버린 정보의 비율 ★★

정리 3.5, 3.6 · 12분

A7의 자료에서 $\hat{\theta}=0.03571$일 때 Louis 공식으로 I_m과 I_c를 구하고 EM의 수렴 비율 I_m/I_c를 구하라. 노트 예제 3.7의 비율 0.1328보다 큰 이유를 설명하라.

A9 빠진 관측이 있는 정규 평균 ★

알고리즘 3.1, 정리 3.6 · 8분

$X_1, \dots, X_{20} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$을 계획했는데 15개만 관측했고 그 평균이 3이다. 빠진 5개를 숨은 자료로 보고 EM을 돌린다. E 단계는 빠진 값을 지금의 μ_t로 채우고, M 단계는 채운 20개의 평균을 구한다. (a) $\mu_0 = 0$에서 μ_1, (b) μ_2를 구하라. (c) 수렴 비율은 얼마인가?

A10 책임 하나 ★

명제 4.1 · 6분

$0.4\,\mathcal{N}(0,1) + 0.6\,\mathcal{N}(3,1)$에서 관측 $y = 1.2$가 첫 성분에서 왔을 책임 r_1을 구하라.

A11 M 단계 한 번 ★

명제 4.1 · 8분

관측 $y = (0, 1, 3, 4)$의 첫 성분 책임이 $r_{i1} = (0.9,\ 0.8,\ 0.2,\ 0.1)$이다. M 단계의 (a) π_1, (b) μ_1, (c) σ_1^2을 구하라.

A12 k-평균 한 바퀴 ★

명제 8.1 · 8분

점 $1, 2, 6, 7, 8$과 두 중심 $2, 6$에서 k-평균을 한 바퀴 돌린다. (a) 배정 뒤의 $W = \sum_i (y_i - c_{c(i)})^2$ (중심은 아직 $2, 6$), (b) 새 첫째 중심, (c) 새 둘째 중심, (d) 새 중심에서의 W를 구하라.

A13 KL 발산의 비대칭 ★

명제 3.3 · 6분

$D\big(\mathcal{N}(\mu_1, s_1^2) \,\|\, \mathcal{N}(\mu_2, s_2^2)\big) = \log \frac{s_2}{s_1} + \frac{s_1^2 + (\mu_1 - \mu_2)^2}{2 s_2^2} - \frac{1}{2}$이다(5회 명제 6.2의 1차원판). (a) $D\big(\mathcal{N}(1, 4) \,\|\, \mathcal{N}(0, 1)\big)$, (b) $D\big(\mathcal{N}(0, 1) \,\|\, \mathcal{N}(1, 4)\big)$를 구하라.

A14 임의효과 모형의 MLE ★★

명제 7.2 · 10분

네 사람을 두 번씩 잰 값이 $(3, 5)$, $(6, 6)$, $(2, 4)$, $(9, 7)$이다. 명제 7.2의 식으로 (a) $\hat{\sigma}^2$, (b) $\hat{m}$, (c) $\hat{\tau}^2$을 구하라.

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 정리 번호, 앞 회차의 정리, 도구 번호를 쓴다.

B1 EM의 단조성, 근거 채우기 ★

① · 정리 3.2 · 15분

EM의 반복은 $\ell(\theta_{t+1}) \geq \ell(\theta_t)$를 만족한다. 각 단계의 근거를 써라.

① $\ell(\theta_{t+1}) - \ell(\theta_t) = \big[Q(\theta_{t+1} \mid \theta_t) - Q(\theta_t \mid \theta_t)\big] - \big[H(\theta_{t+1} \mid \theta_t) - H(\theta_t \mid \theta_t)\big]$이다.

왜? _____

② 첫 괄호는 0 이상이다.

왜? _____

③ 둘째 괄호는 $-D\big(k(\cdot \mid \mathbf{y}; \theta_t) \,\|\, k(\cdot \mid \mathbf{y}; \theta_{t+1})\big)$이다.

왜? _____

④ 둘째 괄호는 0 이하이다.

왜? _____

⑤ $\ell(\theta_{t+1}) - \ell(\theta_t) \geq 0$이다.

왜? _____

⑥ 등호는 Q가 늘지 않고 $k(\cdot \mid \mathbf{y}; \theta_{t+1}) = k(\cdot \mid \mathbf{y}; \theta_t)$일 때뿐이다.

왜? _____

B2 귀찮은 모수의 대가, 근거 채우기 ★★

① · 정리 1.6 · 20분

$I \succ 0$이면 $I^{\psi\psi} \geq I_{\psi\psi}^{-1}$이고 등호는 $I_{\psi\lambda} = \mathbf{0}$일 때뿐이다. 근거를 채워라. $S = I_{\psi\psi} - I_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi}$이다.

① 모든 $\mathbf{u}$에서 $\mathbf{u}' I_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi} \mathbf{u} \geq 0$이다. 왜? _____

② $S \preceq I_{\psi\psi}$이다. 왜? _____

③ $A \geq B \succ 0$이면 $A^{-1} \preceq B^{-1}$이다. 왜? _____

④ $I^{\psi\psi} = S^{-1}$이다. 왜? _____

⑤ $I^{\psi\psi} \geq I_{\psi\psi}^{-1}$이다. 왜? _____

⑥ 등호이면 ①의 값이 모든 $\mathbf{u}$에서 0이므로 $I_{\lambda\psi} = \mathbf{0}$이다. 왜? _____

B3 잃어버린 정보의 원리, 근거 채우기 ★★

① · 정리 3.5 · 20분

$J(\theta) = I_c(\theta \mid \mathbf{y}) - I_m(\theta \mid \mathbf{y})$이고 $I_m = \mathrm{Cov}(\nabla \log f(\mathbf{y}, \mathbf{Z}; \theta) \mid \mathbf{y})$이다. 근거를 채워라.

① 모든 $\mathbf{z}$에서 $\ell(\theta) = \log f(\mathbf{y}, \mathbf{z}; \theta) - \log k(\mathbf{z} \mid \mathbf{y}; \theta)$이다. 왜? _____

② $-\nabla^2 \ell = -\nabla^2 \log f + \nabla^2 \log k$이다. 왜? _____

③ $J(\theta) = I_c(\theta \mid \mathbf{y}) - I_m(\theta \mid \mathbf{y})$이다. 왜? _____

④ $-E[\nabla^2 \log k \mid \mathbf{y}] = \mathrm{Cov}(\nabla \log k \mid \mathbf{y})$이다. 왜? _____

⑤ $\nabla \log k = \nabla \log f - \nabla \ell(\theta)$이다. 왜? _____

⑥ $I_m = \mathrm{Cov}(\nabla \log f \mid \mathbf{y})$이다. 왜? _____

B4 EM의 단조성, 단계의 이름만 ★★

② · 정리 3.2 · 25분

정리 3.2를 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **가능도의 증가를 Q의 차와 H의 차로 나눈다**

② Q의 차의 부호를 정한다

③ H의 차를 Kullback–Leibler 발산으로 쓴다

④ H의 차의 부호를 정한다

⑤ 결론과 등호 조건

B5 귀찮은 모수의 대가, 단계의 이름만 ★★

② · 보조정리 1.5, 정리 1.6 · 25분

보조정리 1.5의 $I^{\psi\psi} = S^{-1}$을 받아들이고 정리 1.6을 증명하라.

① $I_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi}$가 양의 준정부호임을 보인다

② S와 $I_{\psi\psi}$의 순서

③ 역행렬을 취하면 순서가 뒤집힘을 보인다

④ **결론**

⑤ **등호 조건**

B6 EM의 단조성, 전략만 ★★

③ · 정리 3.2 · 30분

정리 3.2를 증명하라. 이번에는 명제 3.3(증거 하한)으로 증명하라.

전략 $q_t = k(\cdot \mid \mathbf{y}; \theta_t)$로 두면 하한 $\mathcal{F}(q_t, \cdot)$가 θ_t에서 ℓ에 닿는다. M 단계는 이 하한을 올리고, 하한은 어디서나 ℓ 아래에 있다.

B7 잃어버린 정보의 원리, 전략만 ★★★ ③ · 정리 3.5 · 35분

정리 3.5를 증명하라. 그다음 이 결과를 노트 예제 2.2(b)의 유전 연관 자료에 써서 $I_m = \mathrm{Var}(x_2 \mid \mathbf{y})/\theta^2$ 임을 보여라.

전략 $\ell = \log f - \log k$를 두 번 미분하고 숨은 자료의 조건부분포로 평균한다. 둘째 등식은 k가 $\mathbf{z}$의 밀도라는 것과 정리 1.2의 두 표현이다.

C

증명 문제

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

C1 감마의 직교 모수 ★★

[T6] · 정리 1.2, 1.6 · 20분

감마분포(모양 α, 척도 β)의 로그밀도는 $\log f = -\log\Gamma(\alpha) - \alpha\log\beta + (\alpha-1)\log x - x/\beta$ 이고 평균은 $m = \alpha\beta$이다.

(a) 모수 (α, β)에서 $I_{\alpha\beta} = 1/\beta \neq 0$임을 보여라.

(b) 모수를 (α, m)으로 바꾸면 $I_{\alpha m} = 0$, 곧 두 모수가 직교함을 보여라.

(c) $\psi'(2) = \pi^2/6 - 1$이다. $\alpha = 2$에서 $\sqrt{n}(\hat{\alpha} - \alpha)$의 점근분산을 구하라. m을 알 때와 모를 때 이 값은 어떻게 다른가?

C2 모수를 바꾸면 정보 행렬은 ★★

[T6] · 정리 1.2, 1.3 · 15분

$\theta = h(\eta)$가 미분가능한 일대일 변환이고 야코비 행렬 $A = \partial\theta/\partial\eta'$가 가역이다.

(a) η의 점수가 $A'\mathbf{s}_\theta$이고 정보 행렬이 $I_\eta = A' I_\theta A$임을 보여라.

(b) 정리 1.3이 주는 $\hat\eta$의 점근공분산 I_η^{-1}이, $\hat\theta$에 다변량 델타 방법을 써서 얻은 점근공분산과 같음을 보여라.

(c) 따름정리 1.4(a)의 이차형식 $n(\hat\theta-\theta_0)' I(\theta_0)(\hat\theta-\theta_0)$은 모수를 바꾸면 그대로인가? 일차 근사의 범위에서 답하라.

C3 EM은 MM 알고리즘이다 ★★

보조정리 2.4, 정리 3.2 · 20분

$G(\theta \mid \theta_t) = \ell(\theta_t) + Q(\theta \mid \theta_t) - Q(\theta_t \mid \theta_t)$로 둔다.

(a) 모든 θ에서 $G(\theta \mid \theta_t) \le \ell(\theta)$이고 $G(\theta_t \mid \theta_t) = \ell(\theta_t)$임을 보여라. ($G$를 ℓ의 **하한 함수**(minorizer)라 한다.)

(b) ℓ의 하한 함수 $G(\cdot \mid \theta_t)$가 무엇이든, $G(\theta_{t+1} \mid \theta_t) \ge G(\theta_t \mid \theta_t)$인 θ_{t+1}을 고르면 $\ell(\theta_{t+1}) \ge \ell(\theta_t)$임을 보여라.

(c) $\log x \le x - 1$을 써서 $\ell(\theta) = \log\theta - \theta^2$ $(\theta > 0)$의 하한 함수를 하나 만들고, 그 하한 함수의 최대점을 다음 반복점으로 하는 알고리즘을 적어라.

C4 완전 자료가 지수족일 때 ★★

명제 2.5, 따름정리 3.4 · 20분

완전 자료의 밀도가 $f(\mathbf{y},\mathbf{z};\boldsymbol{\eta}) = h(\mathbf{y},\mathbf{z})\exp\big(\boldsymbol{\eta}'T(\mathbf{y},\mathbf{z}) - A(\boldsymbol{\eta})\big)$이다. $\nabla A(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}}T$임은 받아들인다.

(a) $Q(\boldsymbol{\eta}\mid\boldsymbol{\eta}_t) = \boldsymbol{\eta}'E_{\boldsymbol{\eta}_t}[T\mid\mathbf{y}] - A(\boldsymbol{\eta})+$($\boldsymbol{\eta}$에 무관한 항)임을 보여라.

(b) M 단계가 방정식 $E_{\boldsymbol{\eta}}T = E_{\boldsymbol{\eta}_t}[T\mid\mathbf{y}]$를 푸는 것임을 보여라.

(c) 관측 점수가 $\nabla\ell(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}}[T\mid\mathbf{y}] - E_{\boldsymbol{\eta}}T$임을 보이고, EM의 고정점이 가능도 방정식의 근임을 결론지어라.

(d) 노트 예제 2.2(b)의 유전 연관 자료에서 T와 E 단계를 적어라.

C5 식별되지 않는 혼합 ★★

[T2] · 정리 1.3 · 15분

(a) 두 Bernoulli의 혼합 $\pi\,\mathrm{Ber}(p_1)+(1-\pi)\mathrm{Ber}(p_2)$는 $\mathrm{Ber}(\pi p_1+(1-\pi)p_2)$와 같음을 보이고, 서로 다른 모수 (π, p_1, p_2)가 같은 분포를 주는 예를 들어라.

(b) (a)의 모형에서 정보 행렬이 모든 점에서 특이함을 보여라. 정리 1.3의 어느 가정이 깨지는가?

(c) 이항 $\mathrm{b}(2, p)$ 두 개의 혼합은 모수가 셋이다. 관측 가능한 확률은 몇 개의 자유도를 갖는가? 이 모형도 식별되지 않음을 보여라.

C6 프로파일 점수의 치우침 ★★

정의 1.7 · 20분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, v)$에서 μ를 귀찮은 모수로 둔 v의 프로파일 로그가능도는 $\ell_p(v) = -\frac{n}{2}\log v - \frac{\mathrm{SS}}{2v}$이다($\mathrm{SS} = \sum(X_i - \bar{X})^2$, 상수 생략).

(a) 프로파일 점수 $\ell_p'(v)$의 기댓값을 v의 식으로 구하라(`check()`에는 기호 `v`로). 진짜 점수의 기댓값과 무엇이 다른가?

(b) ℓ_p의 n을 $n-1$로 바꾼 $\tilde{\ell}(v) = -\frac{n-1}{2}\log v - \frac{\mathrm{SS}}{2v}$의 점수는 기댓값이 0임을 보이고, 최대점이 S^2임을 보여라.

(c) n쌍의 관측 $X_{i1}, X_{i2} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_i, v)$에서 $\mu_1, \dots, \mu_n$을 모두 귀찮은 모수로 두면 프로파일 점수의 기댓값이 $-\frac{n}{2v}$임을 보이고, 프로파일 가능도의 최대점이 $v/2$로 감을 설명하라(노트 7절).

C7 여러 모수의 점수 검정 ★★

정리 1.2, 따름정리 1.4 · 15분

$H_0 : \theta = \theta_0$에서 점수 통계량 $R = \frac{1}{n}\nabla\ell(\theta_0)' I(\theta_0)^{-1}\nabla\ell(\theta_0)$을 생각한다.

(a) H_0 아래에서 $R \xrightarrow{D} \chi^2(p)$임을 보여라.

(b) R은 모수를 바꿔도(연습 C2의 $\theta = h(\eta)$) 정확히 같은 값임을 보여라.

(c) R을 계산하는 데 MLE가 필요하지 않다는 것이 EM으로 맞추는 모형에서 왜 편리한지 설명하라.

C8 비율만 모르는 혼합 ★★

[T4] · 정리 3.2, 3.6 · 20분

두 성분의 밀도 f_1, f_2를 알고 비율 π만 모르는 혼합 $g = \pi f_1 + (1-\pi) f_2$에서 관측 $y_1, \dots, y_n$이 주어졌다.

(a) $\ell(\pi) = \sum_i \log g(y_i)$가 $[0,1]$에서 오목임을 보여라.

(b) EM의 반복이 $\pi_{t+1} = \frac{1}{n}\sum_i \frac{\pi_t f_1(y_i)}{\pi_t f_1(y_i) + (1-\pi_t) f_2(y_i)}$ 임을 보여라.

(c) $\ell'(1) \ge 0$이면 MLE가 $\hat\pi = 1$임을 보이고, 그 조건을 $\sum_i f_2(y_i)/f_1(y_i) \le n$으로 바꿔 써라.

(d) $f_1 = \mathcal{N}(0,1)$, $f_2 = \mathcal{N}(2,1)$, $y = (0,\ 0.5,\ 2)$이면 $\hat\pi$는 경계에 있는가? EM을 노트북으로 돌려 $\hat\pi$를 소수 넷째 자리까지 구하라.

C9 Poisson 혼합의 가능도는 유계다 ★

[T2] · 명제 6.1 · 10분

(a) Poisson 혼합 $\sum_k \pi_k \,\mathrm{Poi}(\lambda_k)$의 로그가능도가 $\sum_i \log \max_\lambda p(y_i;\lambda)$ 이하임을 보이고, 정규 혼합(명제 6.1)과 다른 이유를 설명하라.

(b) $y = (0, 1, 3)$일 때 이 상한 $\sum_i \log \max_\lambda \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$을 구하라.

D

퀄형

여러 절의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 모형을 완전 자료와 관측 자료로 나눠 적는 것부터 답안의 일부다. 수렴할 때까지 반복해야 하는 값은 노트북으로 구한다.

D1 0이 기록되지 않는 Poisson ★★★

[T5], [T6] · 알고리즘 3.1, 정리 3.5, 3.6 · 제한 40분

사건이 한 번 이상 일어난 단위만 기록된다. 기록된 n개 단위의 사건 수 $y_i \geq 1$은 Poisson(λ)를 $y \geq 1$로 제한한 분포 $p(y) = \dfrac{e^{-\lambda}\lambda^{y}}{y!\,(1-e^{-\lambda})}$를 따른다. $\bar{y} = \dfrac{1}{n}\sum y_i$로 쓴다.

(a) 관측 로그가능도를 쓰고 가능도 방정식이 $\dfrac{\lambda}{1-e^{-\lambda}} = \bar{y}$임을 보여라. $\bar{y} > 1$이면 해가 하나뿐임을 보여라.

(b) 기록되지 않은 0의 개수 N_0를 숨은 자료로 두고, 관측 단위 하나마다 기록되지 않은 단위가 평균 $\dfrac{e^{-\lambda}}{1-e^{-\lambda}}$개라는 것을 받아들인다. 완전 자료(모든 $n+N_0$개 단위의 Poisson 표본)의 MLE는 $\sum y_i/(n+N_0)$이다. E 단계와 M 단계를 적고 EM 사상이 $\lambda \mapsto \bar{y}\,(1-e^{-\lambda})$임을 보여라.

(c) EM 사상의 고정점이 (a)의 해임을 보여라.

(d) $\bar{y} = 2$이면 $\hat{\lambda}$는 얼마인가? 소수 넷째 자리까지 구하라.

(e) 수렴 비율(EM 사상의 $\hat{\lambda}$에서의 미분)을 구하고, 그것이 $I_m/I_c = 1 - J/I_c$와 같음을 보여라.

D2 동전 두 개 ★★★

명제 4.1, 4.3, 예제 4.4 · 제한 35분

앞면 확률이 θ_A, θ_B 인 동전 A, B가 있다. 다섯 번에 걸쳐, 매번 같은 확률 $\frac{1}{2}$로 동전 하나를 골라 10번 던졌다. 앞면의 수는 $(5,\ 9,\ 8,\ 4,\ 7)$이고, 어느 동전을 골랐는지는 기록되지 않았다.

(a) 관측 로그가능도를 적어라.

(b) E 단계의 책임 $w_i = P(\text{동전 A} \mid h_i)$를 적어라.

(c) M 단계의 식을 유도하라.

(d) $(\theta_A, \theta_B) = (0.6,\ 0.5)$에서 한 번 돌린 θ_A를 구하라.

(e) 같은 반복의 θ_B를 구하라.

(f) 노트북으로 수렴시킨 θ_A를 소수 넷째 자리까지 구하라.

(g) 수렴시킨 θ_B를 구하라. (θ_B, θ_A)를 바꾼 점의 가능도는 어떤가? $(0.5,\ 0.5)$에서 출발하면 어떻게 되는가?

D3 비율의 정보와 잃어버린 정보 ★★★

[T6] · 정리 1.2, 3.5, 3.6, 3회 정리 2.6 · 제한 35분

알려진 두 밀도 $f_1 = \varphi(\cdot;0,1)$, $f_2 = \varphi(\cdot;\mu,1)$의 혼합 $g = \pi f_1 + (1-\pi) f_2$에서 π만 모른다. 한 관측의 정보 $I(\pi)$를 생각한다.

(a) 점수가 $s = \frac{f_1 - f_2}{g}$이고 $I(\pi) = \int \frac{(f_1-f_2)^2}{g}\,dy$임을 보여라.

(b) 책임 $r = \pi f_1/g$로 쓰면 $s = \frac{r-\pi}{\pi(1-\pi)}$임을 보이고, 이름표 $Z = 1\{\text{성분 1}\}$에 대해 $r = E[Z \mid Y]$임을 써서 $I(\pi) \le \frac{1}{\pi(1-\pi)}$을 보여라. 오른쪽은 무엇의 정보인가?

(c) $\mu \to 0$이면 $I(\pi) \to 0$임을 보여라.

(d) $\pi = \frac{1}{2}$, $\mu = 3$이면 $I(\pi)$는 얼마인가? (수치 적분은 노트북에서)

(e) 같은 경우 표본이 클 때 EM의 수렴 비율 $1 - I(\pi)\pi(1-\pi)$를 구하라.

D4 ABO 혈액형의 유전자 빈도 ★★★

명제 4.1, 정리 3.2, 2회 정리 8.3 · 제한 40분

대립유전자 A, B, O의 빈도를 $p, q, r (p+q+r=1)$이라 하고 Hardy–Weinberg 평형을 가정한다. 유전자형 AA, AO, BB, BO, AB, OO의 확률은 p^2, $2pr$, q^2, $2qr$, $2pq$, r^2이고, 표현형 A는 AA와 AO, B는 BB와 BO이다. 521명의 표현형이 A 186, B 38, AB 13, O 284였다.

(a) 관측 로그가능도를 적어라.

(b) 유전자형의 수 $(n_{AA}, n_{AO}, n_{BB}, n_{BO}, n_{AB}, n_{OO})$를 모두 안다면 MLE가 "유전자 세기" $p = \frac{2n_{AA}+n_{AO}+n_{AB}}{2n}$ 임을 보여라.

(c) E 단계를 적어라.

(d) $(p,q,r) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$에서 한 번 돌린 p를 분수로 구하라.

(e) 같은 반복의 q를 분수로 구하라.

(f) 노트북으로 수렴시킨 $\hat{p}$를 소수 넷째 자리까지 구하라.

(g) 수렴시킨 $\hat{q}$를 구하라.

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 $\log f = -\frac{1}{2}\log(2\pi v) - \frac{(x-\mu)^2}{2v}$를 두 번 미분한다. $E(X-\mu)=0$, $E(X-\mu)^2 = v$.

답 (a) $\frac{1}{32}$ (b) $\frac{16}{25} = 0.64$

A2 힌트 $\log f = x\eta - \log(1+e^{\eta})$이다. 점수는 $x - p$이다.

답 $\frac{1}{16}$

A3 힌트 $I^{\psi\psi} = (I_{\psi\psi} - I_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi})^{-1}$. 2×2 역행렬의 공식으로도 확인한다.

답 (a) $\frac{3}{8}$ (b) $\frac{1}{4}$ (c) $\frac{3}{2}$

A4 힌트 $I_{\lambda\lambda} = \mathrm{diag}(2,4)$이므로 역행렬이 쉽다. $I_{\psi\lambda} = (1,2)$.

답 (a) $\frac{7}{2}$ (b) $\frac{2}{7}$ (c) $\frac{10}{7}$

A5 힌트 A1의 $I^{-1} = \mathrm{diag}(v, 2v^2)$와 $\nabla g = \left(1,\ \frac{z}{2\sqrt{v}}\right)$.

답 $v\left(1 + \frac{z^2}{2}\right)$

A6 힌트 v를 고정해도 μ의 MLE는 $\bar{x}$이다. 그러면 $\ell_p(v) = -\frac{n}{2}\log(2\pi v) - \frac{n\hat{v}}{2v}$이다.

답 $10 - 20\log\frac{3}{2} \approx 1.891$

A7 힌트 (a) E 단계는 $x_2 = 1997\cdot\frac{\theta_0}{2+\theta_0}$, M 단계는 $\theta_1 = \frac{x_2+32}{x_2+32+906+904}$. (b) 점수에 $\theta(2+\theta)(1-\theta)$를 곱한다.

답 (a) $\frac{3309}{77519} \approx 0.04269$ (b) $\frac{-1655+\sqrt{3721809}}{7678} \approx 0.03571$

A8 힌트 $q = \hat{\theta}/(2+\hat{\theta})$로 두면 $x_2 \mid \mathbf{y} \sim \mathrm{b}(1997, q)$, $I_m = \mathrm{Var}(x_2 \mid \mathbf{y})/\hat{\theta}^2$, $I_c = \frac{E[x_2|\mathbf{y}]+32}{\hat{\theta}^2} + \frac{1810}{(1-\hat{\theta})^2}$.

답 0.4951

A9 힌트 $\mu_{t+1} = \frac{15\cdot 3 + 5\mu_t}{20}$. 이 사상의 고정점과 기울기를 본다.

답 (a) $\frac{9}{4}$ (b) $\frac{45}{16}$ (c) $\frac{1}{4}$

A10 힌트 분자와 분모를 첫 항으로 나누면 $r_1 = \left(1 + \frac{0.6}{0.4}e^{-[(y-3)^2 - y^2]/2}\right)^{-1}$.

답 $\left(1 + 1.5e^{-0.9}\right)^{-1} \approx 0.621$

A11 힌트 $n_1 = \sum r_{i1}$. σ_1^2에는 새 μ_1을 쓴다.

답 (a) $\frac{1}{2}$ (b) 0.9 (c) 1.29

A12 힌트 6은 둘째 중심에 정확히 놓여 있다. 배정은 $\{1,2\}$와 $\{6,7,8\}$이다.

답 (a) 6 (b) $\frac{3}{2}$ (c) 7 (d) $\frac{5}{2}$

A13 힌트 (a)는 $s_1 = 2, s_2 = 1$. (b)는 $s_1 = 1, s_2 = 2$.

답 (a) $2 - \log 2 \approx 1.307$ (b) $\log 2 - \frac{1}{4} \approx 0.443$

A14 힌트 $D = (-2, 0, -2, 2)$, $S = (8, 12, 6, 16)$. 분산 MLE는 n으로 나눈다.

답 (a) $\frac{3}{2}$ (b) $\frac{21}{4}$ (c) $\frac{47}{16}$

B1 힌트 근거는 짧게 쓴다. 예: "보조정리 2.4", "2회 정리 8.3". 둘째 단계는 M 단계가 무엇을 하는지 쓴다.

B2 힌트 ③은 행렬의 순서를 고유값의 순서로 바꾸는 단계다. ⑤에서 $A = I_{\psi\psi}$, $B = S$로 둔다.

B3 힌트 ④에서 θ마다 $k(\cdot \mid \mathbf{y}; \theta)$가 하나의 확률모형이라는 것을 쓴다.

B4 힌트 B1에서 채운 근거를 다시 보지 말고 쓴다. 셋째 단계에서 기댓값이 어느 분포에 대한 것인지 적는다.

B5 힌트 셋째 단계는 $B^{-1/2} A B^{-1/2}$의 고유값으로 한다.

B6 힌트 세 줄짜리 부등식 사슬 $\ell(\theta_{t+1}) \geq \mathcal{F}(q_t, \theta_{t+1}) \geq \mathcal{F}(q_t, \theta_t) = \ell(\theta_t)$의 각 부등호와 등호에 근거를 단다.

B7 힌트 유전 연관 자료의 완전 자료 점수는 $\frac{x_2 + 34}{\theta} - \frac{38}{1-\theta}$이고, $\mathbf{y}$가 주어지면 확률변수는 x_2 뿐이다.

C1 힌트 (b) $\beta = m/\alpha$를 넣고 $\partial_m \log f = \alpha(x - m)/m^2$을 α로 미분한 뒤 기댓값을 취한다. (c) 직교이면 $I^{\alpha\alpha} = 1/I_{\alpha\alpha}$이다. $I_{\alpha\alpha}$는 (α, m) 모수에서 구한다.

답 (c) $\frac{1}{\pi^2/6 - 3/2} \approx 6.90$

C2 힌트 (a) 연쇄법칙. (b) $\eta = h^{-1}(\theta)$의 야코비 행렬은 A^{-1}이다.

C3 힌트 (a) 보조정리 2.4로 $\ell(\theta) - G(\theta \mid \theta_t) = H(\theta_t \mid \theta_t) - H(\theta \mid \theta_t)$이다. (c) $\log\theta = \log\theta_t + \log(\theta/\theta_t)$.

C4 힌트 (c)는 명제 2.5: 관측 점수는 완전 자료 점수 $T - \nabla A$의 조건부 기댓값이다.

C5 힌트 (b) 밀도가 $\bar{p} = \pi p_1 + (1-\pi)p_2$ 하나를 통해서만 모수에 의존하면 점수 벡터는 $\nabla \bar{p}$에 평행하다.

C6 힌트 $E\,\mathrm{SS} = (n-1)\upsilon$(6회). (c)에서는 각 쌍 안의 제곱합 $\mathrm{SS}_i = \frac{1}{2}(X_{i1}-X_{i2})^2$의 기댓값이 υ다.

답 (a) $-\frac{1}{2\upsilon}$

C7 힌트 (a) 정리 1.3의 (2)와 5회 정리 3.17. (b) $\mathbf{s}_\eta = A'\mathbf{s}_\theta$, $I_\eta = A' I_\theta A$.

C8 힌트 (a) $\ell''(\pi) = -\sum_i \frac{(f_1-f_2)^2}{g^2} \le 0$. (d) $\ell'(0)$과 $\ell'(1)$의 부호를 먼저 본다.

답 (d) $\hat{\pi} = 0.6625$

C9 힌트 $\max_\lambda e^{-\lambda}\lambda^y/y!$의 최대점은 $\lambda = y$이다($y=0$이면 $\lambda \to 0$).

답 (b) $-4 + \log\frac{9}{2} \approx -2.496$

D1 힌트 (a) $h(\lambda) = \lambda/(1-e^{-\lambda})$는 $\lambda \to 0$에서 1로 가고 증가한다. (e) $J = n\big[\frac{\bar{y}}{\lambda^2} - \frac{e^{-\lambda}}{(1-e^{-\lambda})^2}\big]$, $I_c = \frac{n\bar{y}}{\lambda^2}$이고, $\hat{\lambda}$에서 $\bar{y} = \hat{\lambda}/(1-e^{-\hat{\lambda}})$를 쓴다.

답 (d) $\hat{\lambda} = 1.5936$ (e) 0.4064

D2 힌트 (b) $w_i = \frac{\theta_A^{h_i}(1-\theta_A)^{10-h_i}}{\theta_A^{h_i}(1-\theta_A)^{10-h_i} + \theta_B^{h_i}(1-\theta_B)^{10-h_i}}$. (c) 명제 4.1의 정규를 이항으로 바꾼다. (g) 명제 4.3과 예제 4.4.

답 (d) 0.7130 (e) 0.5813 (f) 0.7968 (g) 0.5196

D3 힌트 (b) 전분산 공식(3회 정리 2.6)으로 $\mathrm{Var}\,E[Z \mid Y] \le \mathrm{Var}\,Z = \pi(1-\pi)$. (c) 지배수렴 [T6].

답 (d) 3.211 (e) 0.197

D4 힌트 (b) 완전 자료의 로그가능도는 $(2n_{AA}+n_{AO}+n_{AB})\log p + (2n_{BB}+n_{BO}+n_{AB})\log q + (n_{AO}+n_{BO}+2n_{OO})\log r$+상수다. (c) A 186명 가운데 AA의 기대 수는 $186\frac{p^2}{p^2+2pr}$.

답 (d) $\frac{261}{1042} \approx 0.2505$ (e) $\frac{191}{3126} \approx 0.0611$ (f) 0.2136 (g) 0.0501