수리통계학 × 통계학의 역사

10회 워크북 — 풀이판

숨은 변수와 EM
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	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.

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?

풀이 이차 미분은 $\partial_\mu^2 = -\frac{1}{\upsilon}$, $\partial_\mu\partial_\upsilon = -\frac{x-\mu}{\upsilon^2}$, $\partial_\upsilon^2 = \frac{1}{2\upsilon^2} - \frac{(x-\mu)^2}{\upsilon^3}$이다. 기댓값에 음수를 붙이면 $I = \operatorname{diag}\left(\frac{1}{\upsilon}, \frac{1}{2\upsilon^2}\right)$이다. (a) $\upsilon = 4$이면 $I_{\upsilon\upsilon} = \frac{1}{32}$. (b) I가 대각이므로 $(I^{-1})_{\upsilon\upsilon} = 2\upsilon^2 = 32$이고 $32/50 = \frac{16}{25} = 0.64$이다. μ와 υ는 직교하므로 μ를 알든 모르든 이 값은 같다.

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

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.

풀이 $p = \frac{e^\eta}{1+e^\eta}$이므로 $\log f(x;\eta) = x\log p + (1-x)\log(1-p) = x\eta - \log(1+e^\eta)$이다. 점수는 $x - \frac{e^\eta}{1+e^\eta} = x - p$이고 $I(\eta) = \operatorname{Var} X = p(1-p)$이다. 둘째 표현으로도 $-\partial_\eta^2 \log f = \frac{e^\eta}{(1+e^\eta)^2} = p(1-p)$이다. $p = 0.2$이면 $I = 0.16$이고 점근분산은 $\frac{1}{100\cdot 0.16} = \frac{1}{16}$이다.

최종답 $\frac{1}{16}$

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?

풀이 (a) Schur 보수는 $4 - 2\cdot\frac{1}{3}\cdot 2 = \frac{8}{3}$이므로 $I^{\psi\psi} = \frac{3}{8}$이다. 2×2 역행렬로도 $\frac{3}{4\cdot 3 - 2^2} = \frac{3}{8}$이다. (b) $\frac{1}{4}$. (c) $\frac{3/8}{1/4} = \frac{3}{2}$이다. λ를 모르면 $\hat{\psi}$의 점근분산이 λ를 알 때의 1.5배다(정리 1.6).

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

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?

풀이 (a) $I_{\lambda\lambda}^{-1} = \mathrm{diag}(\tfrac{1}{2}, \tfrac{1}{4})$이므로 $I_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi} = \tfrac{1^2}{2} + \tfrac{2^2}{4} = \tfrac{3}{2}$이고 $S = 5 - \tfrac{3}{2} = \tfrac{7}{2}$이다. (b) 보조정리 1.5로 $I^{\psi\psi} = S^{-1} = \tfrac{2}{7}$이다. 3×3 역행렬의 (1,1) 여인수 $\frac{2\cdot 4 - 0}{\det I} = \frac{8}{28}$로도 같다 ($\det I = 5\cdot 8 - 1\cdot 4 + 2\cdot(0-4) = 28$). (c) $I_{\psi\psi}^{-1} = \tfrac{1}{5}$이므로 $\frac{2/7}{1/5} = \frac{10}{7}$배다.

최종답 (a) $\tfrac{7}{2}$ (b) $\tfrac{2}{7}$ (c) $\tfrac{10}{7}$

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.)

풀이 $g = \mu + z\sqrt{v}$이므로 $\nabla g = \left(1, \frac{z}{2\sqrt{v}}\right)$이고, A1에서 $I^{-1} = \mathrm{diag}(v, 2v^2)$이다. 점근분산은 $\nabla g' I^{-1} \nabla g = v + \frac{z^2}{4v} \cdot 2v^2 = v\left(1 + \frac{z^2}{2}\right)$이다. $z = 0$이면 평균의 분산 v이고, 분위수가 꼬리로 갈수록 $\hat{\sigma}$의 불확실성이 더해진다.

최종답 $v\left(1 + \frac{z^2}{2}\right)$

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.

풀이 v를 고정하면 $\sum (x_i - \mu)^2$을 최소화하는 μ는 $\bar{x}$이므로 $\ell_p(v) = -\frac{n}{2}\log(2\pi v) - \frac{n\hat{v}}{2v}$이다. ℓ_p의 최대점은 $\hat{v}$이고

$$\Lambda = 2[\ell_p(\hat{v}) - \ell_p(1)] = n\left[\frac{\hat{v}}{1} - 1 - \log\frac{\hat{v}}{1}\right] = 20\left(0.5 - \log 1.5\right) = 10 - 20\log\tfrac{3}{2} \approx 1.891.$$

3.84보다 작으므로 5% 수준에서 H_0를 기각하지 않는다.

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

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.

풀이 (a) $x_2=1997\cdot\frac{1/20}{41/20}=\frac{1997}{41}\approx 48.71$이고 $\theta_1=\frac{1997/41+32}{1997/41+1842}=\frac{1997+1312}{1997+75522}=\frac{3309}{77519}\approx 0.04269$이다.

(b) 점수 $\frac{1997}{2+\theta}-\frac{1810}{1-\theta}+\frac{32}{\theta}=0$에 $\theta(2+\theta)(1-\theta)$를 곱하면 $1997\theta(1-\theta)-1810\theta(2+\theta)+32(2+\theta)(1-\theta)=0$, 곧 $3839\theta^2+1655\theta-64=0$이다. 양의 근은 $\hat\theta=\frac{-1655+\sqrt{3721809}}{7678}\approx 0.03571$이다.

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

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.

풀이 $q=\frac{0.03571}{2.03571}=0.01754$이므로 $E[x_2\mid\mathbf{y}]=1997q=35.03$, $\mathrm{Var}(x_2\mid\mathbf{y})=1997q(1-q)=34.42$이다. 완전 자료 로그가능도 $(x_2+32)\log\theta+1810\log(1-\theta)$의 점수는 $\frac{x_2+32}{\theta}-\frac{1810}{1-\theta}$이므로 Louis 공식으로 $I_m=\mathrm{Var}(x_2\mid\mathbf{y})/\hat\theta^2=26987$이고 $I_c=\frac{35.03+32}{\hat\theta^2}+\frac{1810}{(1-\hat\theta)^2}=54506$이다. 비율은 0.4951이고 관측 정보는 $J=I_c-I_m=27519$, 표준오차는 $1/\sqrt{J}=0.0060$이다. θ가 작으면 θ를 직접 알려 주는 칸이 "$\frac{\theta}{4}$"의 작은 칸들뿐이고, 그 가운데 큰 몫(x_2)이 첫 범주 안에 숨어 있다. 그래서 정보의 절반 가까이를 잃는다.

최종답 0.4951

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?

풀이 $\mu_{t+1}=\frac{45+5\mu_t}{20}$이다. (a) $\mu_1=\frac{45}{20}=\frac{9}{4}$. (b) $\mu_2=\frac{45+5\cdot 9/4}{20}=\frac{45}{16}$. (c) 사상의 기울기는 $\frac{5}{20}=\frac{1}{4}$이고 고정점은 관측 평균 3이다. 빠진 비율이 곧 잃어버린 정보의 비율이다($I_c=20$, $J=15$).

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

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.

풀이 $r_1=\dfrac{0.4\varphi(1.2)}{0.4\varphi(1.2)+0.6\varphi(1.2-3)}=\big(1+1.5\,e^{-[(1.8)^2-(1.2)^2]/2}\big)^{-1}=\big(1+1.5e^{-0.9}\big)^{-1}\approx 0.621$이다. $y=1.2$는 첫 평균에 더 가깝지만 둘째 성분의 비율이 커서 책임은 0.62에 그친다.

최종답 $\big(1+1.5e^{-0.9}\big)^{-1}\approx 0.621$

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.

풀이 $n_1=2.0$이므로 (a) $\pi_1=\dfrac{2}{4}=\dfrac{1}{2}$. (b) $\mu_1=\dfrac{0+0.8+0.6+0.4}{2}=0.9$. (c) $\sigma_1^2=\dfrac{0.9\cdot 0.81+0.8\cdot 0.01+0.2\cdot 4.41+0.1\cdot 9.61}{2}=\dfrac{2.58}{2}=1.29$.

최종답 (a) $\dfrac{1}{2}$ (b) 0.9 (c) 1.29

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.

풀이 배정은 $\{1,2\}\to 2$, $\{6,7,8\}\to 6$이다. (a) $W=1+0+0+1+4=6$. (b) $\dfrac{3}{2}$. (c) 7. (d) $W=\dfrac{1}{4}+\dfrac{1}{4}+1+0+1=\dfrac{5}{2}$. 다시 배정해도 바뀌지 않으므로 여기서 멈춘다. 두 단계가 모두 W를 늘리지 않는다(노트 연습 8.2).

최종답 (a) 6 (b) $\dfrac{3}{2}$ (c) 7 (d) $\dfrac{5}{2}$

A13 KL 발산의 비대칭 ★

명제 3.3 · 6분

$D\big(\mathcal{N}(\mu_1,s_1^2)\,\|\,\mathcal{N}(\mu_2,s_2^2)\big)=\log\dfrac{s_2}{s_1}+\dfrac{s_1^2+(\mu_1-\mu_2)^2}{2s_2^2}-\dfrac{1}{2}$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.

풀이 (a) $\log\dfrac{1}{2}+\dfrac{4+1}{2}-\dfrac{1}{2}=2-\log 2\approx 1.307$. (b) $\log 2+\dfrac{1+1}{8}-\dfrac{1}{2}=\log 2-\dfrac{1}{4}\approx 0.443$. 두 값이 다르다. 명제 3.3의 하한은 $D(q\|k)$, 곧 근사 q를 앞에 둔 방향의 발산을 쓴다.

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

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.

풀이 $D=(-2,0,-2,2)$이므로 (a) $\hat{\sigma}^2=\frac{4+0+4+4}{8}=\frac{3}{2}$. $S=(8,12,6,16)$, $\bar{S}=10.5$이므로 (b) $\hat{m}=\frac{21}{4}$. $\frac{1}{n}\sum(S_i-\bar{S})^2=\frac{6.25+2.25+20.25+30.25}{4}=\frac{59}{4}$이므로 (c) $\hat{\tau}^2=\frac{1}{4}\left(\frac{59}{4}-3\right)=\frac{47}{16}$이다. 양수이므로 경계 문제는 없다.

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

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. — 보조정리 2.4를 $\theta' = \theta_t$로 두 번

② 첫 괄호는 0 이상이다. — M 단계: θ_{t+1}은 $Q(\cdot \mid \theta_t)$의 최대점

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

④ 둘째 괄호는 0 이하이다. — Gibbs 부등식, 2회 정리 8.3

⑤ $\ell(\theta_{t+1}) - \ell(\theta_t) \geq 0$이다. — ①②④

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

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 3.2의 단계형 증명과 같다. 증명에서 M 단계가 쓰인 곳은 둘째 단계뿐이므로, Q를 늘리기만 하는 일반화 EM에서도 결론은 그대로다(노트 연습 3.4).

B2 귀찮은 모수의 대가, 근거 채우기 ★★

① · 정리 1.6 · 20분

$I \succ 0$이면 $I^{\psi\psi} \succeq I_{\psi\psi}^{-1}$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. — $= (I_{\lambda\psi}\mathbf{u})' I_{\lambda\lambda}^{-1} (I_{\lambda\psi}\mathbf{u})$, $I_{\lambda\lambda}^{-1} \succ 0$

② $S \preceq I_{\psi\psi}$이다. — ①과 S의 정의

③ $A \succeq B \succ 0$이면 $A^{-1} \preceq B^{-1}$이다. — $B^{-1/2} A B^{-1/2} \succeq \mathbf{1}$의 고유값

④ $I^{\psi\psi} = S^{-1}$이다. — 보조정리 1.5

⑤ $I^{\psi\psi} \succeq I_{\psi\psi}^{-1}$이다. — ②③④, $S \succ 0$

⑥ 등호이면 ①의 값이 모든 $\mathbf{u}$에서 0이므로 $I_{\lambda\psi} = \mathbf{0}$이다. — $I_{\lambda\lambda}^{-1} \succ 0$

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 1.6의 단계형 증명과 같다. 셋째 단계의 자세한 논증: $C = B^{-1/2}AB^{-1/2}$는 대칭이고 $C \succeq \mathbf{1}$이므로 고유값이 모두 1 이상이다. $C^{-1} = B^{1/2}A^{-1}B^{1/2}$의 고유값은 그 역수이므로 1 이하이고, $C^{-1} \preceq \mathbf{1}$이다. 양쪽에 $B^{-1/2}$를 곱하면 $A^{-1} \preceq B^{-1}$이다.

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. — 보조정리 2.4

② $-\nabla^2 \ell = -\nabla^2 \log f + \nabla^2 \log k$이다. — ①을 두 번 미분

③ $J(\theta) = I_c(\theta \mid \mathbf{y}) - I_m(\theta \mid \mathbf{y})$이다. — 왼쪽은 $\mathbf{z}$에 무관, k로 평균

④ $-E[\nabla^2 \log k \mid \mathbf{y}] = \mathrm{Cov}(\nabla \log k \mid \mathbf{y})$이다. — 정리 1.2를 $\mathbf{z}$의 밀도 k에

⑤ $\nabla \log k = \nabla \log f - \nabla \ell(\theta)$이다. — ①을 한 번 미분

⑥ $I_m = \mathrm{Cov}(\nabla \log f \mid \mathbf{y})$이다. — $\nabla \ell$은 $\mathbf{z}$에 대해 상수

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 3.5의 단계형 증명과 같다. 넷째 단계에는 $\int k \, d\mathbf{z} = 1$을 적분 기호 아래에서 두 번 미분할 수 있다는 조건이 필요하다 [T6].

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**

풀이 B1의 여섯 단계를 이름에 맞춰 쓰면 된다. 첫 단계는 $\ell(\theta) = Q(\theta \mid \theta_t) - H(\theta \mid \theta_t)$를 θ_{t+1}과 θ_t에서 빼는 것이다. 셋째 단계의 기댓값은 $\mathbf{Z} \sim k(\cdot \mid \mathbf{y}; \theta_t)$에 대한 것이고, 그래서 발산은 $D\big(k(\cdot \mid \mathbf{y}; \theta_t) \,\|\, k(\cdot \mid \mathbf{y}; \theta_{t+1})\big)$의 순서다.

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

풀이 B2의 풀이와 같다. 넷째 단계에서 $S \succ 0$이 필요하다. 이것은 보조정리 1.5의 증명에서 $\mathbf{u} \neq \mathbf{0}$이면 $\mathbf{u}'S\mathbf{u}$가 영이 아닌 벡터에서의 I의 이차형식이라는 것으로 나온다.

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.

풀이

① $q_t = k(\cdot \mid \mathbf{y}; \theta_t)$로 두면 $\mathcal{F}(q_t, \theta_t) = \ell(\theta_t)$이다. — 명제 3.3의 등호 조건

② $\mathcal{F}(q_t, \theta) = Q(\theta \mid \theta_t) - E_{q_t} \log q_t$이고 둘째 항은 θ에 무관하다. 따라서 θ_{t+1}은 $\mathcal{F}(q_t, \cdot)$의 최대점이고 $\mathcal{F}(q_t, \theta_{t+1}) \geq \mathcal{F}(q_t, \theta_t)$이다. — M 단계

③ $\ell(\theta_{t+1}) \geq \mathcal{F}(q_t, \theta_{t+1})$이다. — 명제 3.3

④ ①–③을 이으면 $\ell(\theta_{t+1}) \geq \ell(\theta_t)$이다.

첫 증명의 H의 차는 둘째 증명에서 $D(q_t \| k(\cdot \mid \mathbf{y}; \theta_{t+1})) = \ell(\theta_{t+1}) - \mathcal{F}(q_t, \theta_{t+1})$로 나타난다. 두 증명은 같은 등식을 다르게 읽은 것이다.

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.

풀이 정리 3.5의 증명은 B3의 여섯 단계다. 유전 연관 자료의 완전 자료 점수는 $\frac{x_2+34}{\theta} - \frac{38}{1-\theta}$이고, $\mathbf{y}$가 주어지면 확률변수는 계수가 $\frac{1}{\theta}$인 x_2뿐이므로 $I_m = \mathrm{Var}(x_2 \mid \mathbf{y})/\theta^2$이다. 노트 연습 2.2의 $x_2 \mid \mathbf{y} \sim \mathrm{b}\big(125, \frac{\theta}{2+\theta}\big)$를 넣으면 $\hat\theta$에서 $I_m = 57.8$이다.

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?

풀이 (a) $\partial_\alpha\partial_\beta \log f = -\frac{1}{\beta}$이므로 $I_{\alpha\beta} = \frac{1}{\beta}$.

(b) $\beta = m/\alpha$를 넣으면 $\log f = -\log\Gamma(\alpha) - \alpha\log m + \alpha\log\alpha + (\alpha-1)\log x - \alpha x/m$이다. $\partial_m \log f = -\frac{\alpha}{m} + \frac{\alpha x}{m^2} = \frac{\alpha(x-m)}{m^2}$이고, 이것을 α로 미분하면 $\frac{x-m}{m^2}$이다. $EX = m$이므로 $I_{\alpha m} = -E\frac{X-m}{m^2} = 0$이다.

(c) (α, m) 모수에서 $\partial_\alpha^2 \log f = -\psi'(\alpha) + \frac{1}{\alpha}$이므로 $I_{\alpha\alpha} = \psi'(\alpha) - \frac{1}{\alpha}$이다. 직교하므로 $I^{\alpha\alpha} = 1/I_{\alpha\alpha}$이고, $\alpha = 2$에서 $\frac{1}{\pi^2/6 - 3/2} \approx 6.90$이다. 직교 모수에서는 m을 알든 모르든 이 값이 같다(정리 1.6의 등호). 노트 1.2절의 "β를 알면 1.55, 모르면 6.90"에서 6.90과 같은 값이다. β를 아는 것과 m을 아는 것은 다른 정보다.

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

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.

풀이 (a) 연쇄법칙으로 $\partial \log f/\partial \eta_j = \sum_k (\partial \log f/\partial \theta_k)(\partial \theta_k/\partial \eta_j)$, 곧 $\mathbf{s}_\eta = A'\mathbf{s}_\theta$이다. $I_\eta = \mathrm{Cov}(A'\mathbf{s}_\theta) = A' I_\theta A$이다(3회의 선형변환 규칙).

(b) 불변성(9회 정리 1.3)으로 $\hat\eta = h^{-1}(\hat\theta)$이고 h^{-1}의 야코비 행렬은 A^{-1}이다. 다변량 델타 방법(7회 정리 5.4)의 점근공분산은 $A^{-1} I_\theta^{-1} (A^{-1})' = (A' I_\theta A)^{-1} = I_\eta^{-1}$이다.

(c) $\mathbf{e} = \hat\theta - \theta_0$로 쓰면 $\hat\eta - \eta_0 \approx A^{-1}\mathbf{e}$이므로 $(\hat\eta-\eta_0)' I_\eta (\hat\eta-\eta_0) \approx (A^{-1}\mathbf{e})' A' I_\theta A (A^{-1}\mathbf{e}) = \mathbf{e}' I_\theta \mathbf{e}$이다. 일차 근사의 범위에서 같고, 유한 표본에서는 다르다. Wald 신뢰영역의 모양이 모수화에 따라 달라지는 이유가 이 근사의 오차다. 가능도비 통계량은 모수를 바꿔도 정확히 같다.

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.

풀이 (a) 보조정리 2.4로 $\ell(\theta) = Q(\theta \mid \theta_t) - H(\theta \mid \theta_t)$이고 $\ell(\theta_t) = Q(\theta_t \mid \theta_t) - H(\theta_t \mid \theta_t)$이므로 $\ell(\theta) - G(\theta \mid \theta_t) = H(\theta_t \mid \theta_t) - H(\theta \mid \theta_t) = D\big(k(\cdot \mid \mathbf{y};\theta_t) \| k(\cdot \mid \mathbf{y};\theta)\big) \ge 0$이다. $\theta = \theta_t$이면 0이다.

(b) $\ell(\theta_{t+1}) \ge G(\theta_{t+1} \mid \theta_t) \ge G(\theta_t \mid \theta_t) = \ell(\theta_t)$. 차례로 하한 함수의 정의, θ_{t+1}의 선택, 닿는 점의 조건을 썼다.

(c) $\log x \le x-1$을 $x = \theta_t/\theta$에 쓰면 $\log\frac{\theta_t}{\theta} \le \frac{\theta_t}{\theta} - 1$, 곧 $\log\theta \ge \log\theta_t + 1 - \frac{\theta_t}{\theta}$이다. 따라서 $G(\theta \mid \theta_t) = \log\theta_t + 1 - \frac{\theta_t}{\theta} - \theta^2 \le \ell(\theta)$이고 $\theta = \theta_t$에서 같다. G를 최대화하면 $\frac{\theta_t}{\theta^2} = 2\theta$, 곧 $\theta_{t+1} = (\theta_t/2)^{1/3}$이다. 고정점은 $\theta^3 = \theta/2$, $\theta = 1/\sqrt{2}$이고 $\ell'(\theta) = \frac{1}{\theta} - 2\theta = 0$의 근과 같다.

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. $\nabla A(\boldsymbol\eta) = E_\eta T$임은 받아들인다.

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

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

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

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

풀이 (a) $\log f = \log h + \boldsymbol{\eta}'T - A(\boldsymbol{\eta})$를 $k(\cdot \mid \mathbf{y}; \boldsymbol{\eta}_t)$로 평균하면 $E[\log h \mid \mathbf{y}] + \boldsymbol{\eta}' E_{\boldsymbol{\eta}_t}[T \mid \mathbf{y}] - A(\boldsymbol{\eta})$이다. 첫 항은 $\boldsymbol{\eta}$에 무관하다.

(b) (a)를 $\boldsymbol{\eta}$로 미분해 0으로 두면 $\nabla A(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}_t}[T \mid \mathbf{y}]$, 곧 $E_{\boldsymbol{\eta}} T = E_{\boldsymbol{\eta}_t}[T \mid \mathbf{y}]$이다. A는 볼록이므로($\nabla^2 A = \operatorname{Cov} T \succeq 0$) 이 1계 조건이 최대점을 준다. 완전 자료의 MLE 방정식 $E_{\boldsymbol{\eta}} T = T$에서 T를 조건부 기댓값으로 바꾼 것이다.

(c) 완전 자료 점수는 $T - \nabla A(\boldsymbol{\eta})$이므로 명제 2.5로 $\nabla \ell(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}}[T \mid \mathbf{y}] - \nabla A(\boldsymbol{\eta})$이다. 고정점 $\boldsymbol{\eta}^*$에서 (b)는 $\nabla A(\boldsymbol{\eta}^*) = E_{\boldsymbol{\eta}^*}[T \mid \mathbf{y}]$이고, 이것이 $\nabla \ell(\boldsymbol{\eta}^*) = \mathbf{0}$이다.

(d) 완전 자료는 이항 모형 "$x_2 + y_4$번 성공, $y_2 + y_3$번 실패"이므로 자연모수 $\eta = \log \frac{\theta}{1-\theta}$에 대해 $T = x_2 + y_4$이다. E 단계는 $E[T \mid \mathbf{y}] = 125\frac{\theta_t}{2+\theta_t} + 34$이고, M 단계는 $(x_2 + y_4 + y_2 + y_3)\theta = E[T \mid \mathbf{y}]$를 푼다. x_2가 T와 시행 수에 함께 들어가므로 노트의 M 단계 $\theta_{t+1} = \frac{x_2^{(t)} + 34}{x_2^{(t)} + 72}$가 된다.

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.

풀이 (a) $P(X=1) = \pi p_1 + (1-\pi)p_2$이므로 혼합은 $\mathrm{Ber}(\bar{p})$다. $(\frac{1}{2}, 0.2, 0.6)$, $(\frac{1}{2}, 0.6, 0.2)$, $(1, 0.4, p_2)$(p_2는 아무 값)는 모두 $\mathrm{Ber}(0.4)$를 준다.

(b) $\log f = x \log \bar{p} + (1-x)\log(1-\bar{p})$이므로 점수는 $\frac{x-\bar{p}}{\bar{p}(1-\bar{p})}\nabla\bar{p}$이다. 정보 행렬은 $\frac{1}{\bar{p}(1-\bar{p})}\nabla\bar{p}\nabla\bar{p}'$로 계수가 1이고, 3×3 행렬로서 특이하다. 정리 1.3의 $I(\theta_0) \succ 0$이 깨진다. $\nabla\bar{p}$에 수직인 방향으로 모수를 움직여도 분포가 변하지 않는다.

(c) 관측값은 $\{0,1,2\}$이고 확률의 합이 1이므로 자유도는 2다. 모수 셋을 자유도 둘로 보내는 매끄러운 사상은 일대일일 수 없다. 구체적으로, 성공확률 P가 p_1(확률 π)이나 p_2인 확률변수라 하면 $P(X=0) = E(1-P)^2$, $P(X=1) = 2E[P(1-P)]$, $P(X=2) = EP^2$이므로 관측 분포는 EP와 EP^2에만 의존한다. $(\pi, p_1, p_2) = (\frac{1}{2}, 0.2, 0.6)$과 $(0.2,\ 0,\ 0.5)$는 모두 $EP = 0.4$, $EP^2 = 0.2$이므로 같은 분포를 준다.

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).

풀이 (a) $\ell_p'(v) = -\frac{n}{2v} + \frac{\mathrm{SS}}{2v^2}$이고 $E\,\mathrm{SS} = (n-1)v$이므로 $E\ell_p'(v) = -\frac{n}{2v} + \frac{n-1}{2v} = -\frac{1}{2v}$이다. 진짜 점수의 기댓값은 0이다(정리 1.2). 귀찮은 모수 하나를 추정해 넣은 대가가 $O(1)$의 치우침으로 나타난다.

(b) $\tilde\ell'(v) = -\frac{n-1}{2v} + \frac{\mathrm{SS}}{2v^2}$의 기댓값은 0이고, 0으로 두면 $v = \mathrm{SS}/(n-1) = S^2$이다. 이것이 제한 최대가능도(REML)의 가장 간단한 예다.

(c) 각 쌍에서 μ_i의 프로파일 추정은 쌍의 평균이고 남는 제곱합은 $\mathrm{SS}_i = \frac{1}{2}(X_{i1}-X_{i2})^2$이다. $E(X_{i1}-X_{i2})^2 = 2v$이므로 $E\,\mathrm{SS}_i = v$이다. 관측이 $2n$개이므로 $\ell_p(v) = -n\log v - \frac{\sum_i \mathrm{SS}_i}{2v}$이고

$$E\ell_p'(v) = -\frac{n}{v} + \frac{nv}{2v^2} = -\frac{n}{2v}.$$

쌍마다 (a)의 $-\frac{1}{2v}$가 하나씩 쌓인다. $\ell_p'(v)=0$의 근은 $\hat v = \frac{1}{2n}\sum_i \mathrm{SS}_i$이고, 대수의 법칙으로 $\frac{1}{n}\sum_i \mathrm{SS}_i \xrightarrow{P} v$이므로 $\hat v \xrightarrow{P} v/2$이다.

최종답 (a) $-\frac{1}{2v}$

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.

풀이 (a) $\mathbf{U} = n^{-1/2}\nabla\ell(\theta_0) \xrightarrow{D} \mathcal{N}_p(\mathbf{0},I)$(다변량 중심극한정리, 정리 1.3의 ②)이고 $R = \mathbf{U}'I^{-1}\mathbf{U}$는 $\mathbf{U}$의 연속함수이므로 연속사상정리로 $\mathbf{W}'I^{-1}\mathbf{W}$, $\mathbf{W}\sim\mathcal{N}_p(\mathbf{0},I)$로 분포수렴한다. $I^{-1/2}\mathbf{W}\sim\mathcal{N}_p(\mathbf{0},\mathbf{1})$이므로 $\chi^2(p)$이다(5회 정리 3.17).

(b) $\nabla_\eta \ell = A' \nabla_\theta \ell$ 이고 $I_\eta^{-1} = A^{-1} I_\theta^{-1} (A')^{-1}$ 이므로

$$\nabla_\eta \ell' I_\eta^{-1} \nabla_\eta \ell = \nabla_\theta \ell' A A^{-1} I_\theta^{-1} (A')^{-1} A' \nabla_\theta \ell = \nabla_\theta \ell' I_\theta^{-1} \nabla_\theta \ell .$$

근사 없이 같다.
(c) R은 θ_0에서의 점수와 정보만 쓴다. 숨은 자료가 있으면 점수는 명제 2.5로 E 단계 하나에서 얻고, 수렴이 느린 EM을 끝까지 돌리지 않아도 된다.

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.

풀이 (a) $\ell'(\pi) = \sum_i \frac{f_1(y_i) - f_2(y_i)}{g(y_i)}$, $\ell''(\pi) = -\sum_i \frac{(f_1 - f_2)^2}{g^2} \le 0$이다.
(b) 이름표를 숨은 자료로 두면 명제 4.1의 비율 부분과 같다. 책임 $r_i = \frac{\pi_t f_1}{\pi_t f_1 + (1-\pi_t) f_2}$의 평균이 새 비율이다.
(c) 오목함수이므로 $\ell'(1) \ge 0$이면 $[0,1]$에서 ℓ은 증가하고 최대점은 1이다. $\ell'(1) = \sum_i \frac{f_1 - f_2}{f_1} = n - \sum_i f_2(y_i)/f_1(y_i)$이다.
(d) $\ell'(0) = \sum (f_1 - f_2)/f_2 = 7.24 > 0$, $\ell'(1) = n - \sum f_2/f_1 = -4.89 < 0$이므로 최대점은 내부에 있고 오목성으로 하나뿐이다. EM은 정리 3.2로 ℓ을 늘리고 정류점이 하나뿐이므로 그 점으로 간다. $\hat\pi = 0.6625$이다.

최종답 (d) $\hat\pi = 0.6625$

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.

풀이 (a) $\pi_k p(y_i;\lambda_k) \le \pi_k \max_\lambda p(y_i;\lambda)$를 k에 대해 더하면 $g(y_i) \le \max_\lambda p(y_i;\lambda) \le 1$이다. 정규 혼합의 밀도는 $1/(\sqrt{2\pi}\sigma)$로 끝없이 커질 수 있었다. 확률은 유계이고 밀도는

유계가 아니다. (b) 최댓값은 $y=0$에서 1, $y=1$에서 e^{-1}, $y=3$에서 $\frac{27}{6}e^{-3}$이므로 합은 $-4+\log\frac{9}{2}\approx -2.496$이다. 두 성분으로 세 관측을 모두 완벽하게 맞출 수는 없으므로 이 상한에는 닿지 않는다.

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

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 $y_i \ge 1$은 Poisson(λ)를 $y \ge 1$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.

풀이 (a) $\ell(\lambda) = n\bar{y}\log\lambda - n\lambda - n\log(1-e^{-\lambda}) +$ 상수이고 $\ell'(\lambda) = \dfrac{n\bar{y}}{\lambda} - n - \dfrac{ne^{-\lambda}}{1-e^{-\lambda}} = \dfrac{n\bar{y}}{\lambda} - \dfrac{n}{1-e^{-\lambda}}$이다. 0으로 두면 $\dfrac{\lambda}{1-e^{-\lambda}} = \bar{y}$이다. $h(\lambda) = \dfrac{\lambda}{1-e^{-\lambda}}$는 $\lambda \to 0$에서 1로 가고($1 - e^{-\lambda} \sim \lambda$ [T8]), $h'(\lambda) = \dfrac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{(1-e^{-\lambda})^2} > 0$이다($e^{\lambda} > 1 + \lambda$). 따라서 $\bar{y} > 1$이면 해가 하나뿐이다.

(b) E 단계: $E[N_0 \mid \mathbf{y}] = n\dfrac{e^{-\lambda_t}}{1-e^{-\lambda_t}}$. M 단계: $\lambda_{t+1} = \dfrac{n\bar{y}}{n + E[N_0 \mid \mathbf{y}]} = \dfrac{\bar{y}}{1 + e^{-\lambda_t}/(1-e^{-\lambda_t})} = \bar{y}\,(1 - e^{-\lambda_t})$.

(c) 고정점 $\lambda = \bar{y}(1 - e^{-\lambda})$는 (a)의 방정식과 같다.

(d) $\lambda_0 = 2$에서 $1.7293,\ 1.6452,\ 1.6140, \ldots$ 로 가고 $\hat{\lambda} = 1.5936$이다.

(e) 사상의 미분은 $\bar{y}e^{-\lambda}$이고 $\hat{\lambda}$에서 $2e^{-1.5936} = 0.4064$이다. 완전 자료의 0들은 로그가능도에 $-\lambda$만 더하므로 $I_c = E[-\ell_c'' \mid \mathbf{y}] = n\bar{y}/\lambda^2$이다. $1 - \dfrac{J}{I_c} = \dfrac{\lambda^2 e^{-\lambda}}{\bar{y}(1-e^{-\lambda})^2}$이고 $\hat{\lambda}$에서 $\bar{y} = \hat{\lambda}/(1-e^{-\hat{\lambda}})$를 넣으면 $\dfrac{\hat{\lambda}e^{-\hat{\lambda}}}{1-e^{-\hat{\lambda}}} = \bar{y}e^{-\hat{\lambda}}$로 같다(정리 3.6).

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

D2 동전 두 개 ★★★

명제 4.1, 4.3, 예제 4.4 · 제한 35분

앞면 확률이 θ_A, θ_B인 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?

풀이 (a) $\ell(\theta_A,\theta_B) = \sum_{i=1}^{5} \log\!\left[\tfrac{1}{2}\theta_A^{h_i}(1-\theta_A)^{10-h_i} + \tfrac{1}{2}\theta_B^{h_i}(1-\theta_B)^{10-h_i}\right]$ + 상수.

(b) 베이즈 정리로 힌트의 식이다. 비율 $\frac{1}{2}$은 약분된다.

(c) $Q = \sum_i \big[w_i(h_i\log\theta_A + (10-h_i)\log(1-\theta_A)) + (1-w_i)(h_i\log\theta_B + (10-h_i)\log(1-\theta_B))\big]$를 미분하면 $\theta_A = \dfrac{\sum w_i h_i}{10\sum w_i}$, $\theta_B = \dfrac{\sum (1-w_i) h_i}{10\sum (1-w_i)}$이다. 각 동전에 책임만큼 나눠 준 앞면의 비율이다.

(d) $w = (0.449,\ 0.805,\ 0.734,\ 0.352,\ 0.647)$이고 A에 나눠 준 앞면은 21.30, 던진 수는 29.87이므로 $\theta_A = 0.7130$이다.

(e) B에는 앞면 11.70, 던진 수 20.13이므로 $\theta_B = 0.5813$이다.

(f) 0.7968.

(g) 0.5196. 이름표를 바꾼 $(0.5196,\ 0.7968)$은 가능도가 정확히 같다(명제 4.3). $(0.5, 0.5)$에서 출발하면 모든 책임이 $\frac{1}{2}$이고 두 동전 모두 전체 비율 $\frac{33}{50} = 0.66$으로 가서 갈라지지 않는다(예제 4.4).

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

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

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

풀이 (a) $\log g$를 π로 미분하면 $s = \frac{f_1 - f_2}{g}$이다. $Es = \int (f_1 - f_2)\,dy = 0$이므로 $I(\pi) = Es^2 = \int \frac{(f_1-f_2)^2}{g^2} g\,dy$이다.

(b) $\frac{f_1}{g} = \frac{r}{\pi}$, $\frac{f_2}{g} = \frac{1-r}{1-\pi}$이므로 $s = \frac{r}{\pi} - \frac{1-r}{1-\pi} = \frac{r-\pi}{\pi(1-\pi)}$이다. $r = P(Z=1 \mid Y) = E[Z \mid Y]$이고 $Er = EZ = \pi$이므로 $I(\pi) = \frac{\operatorname{Var} r}{\pi^2(1-\pi)^2}$이다. 전분산 공식으로 $\operatorname{Var} E[Z \mid Y] \le \operatorname{Var} Z = \pi(1-\pi)$이므로 $I(\pi) \le \frac{1}{\pi(1-\pi)}$이다. 오른쪽은 이름표를 볼 때, 곧 Bernoulli(π) 하나의 정보 I_c다. 등호는 $r = Z$, 곧 관측이 이름표를 완전히 정할 때뿐이다.

(c) $\mu \to 0$이면 $f_2 \to f_1$이고 $\frac{(f_1-f_2)^2}{g} \le \frac{(f_1+f_2)^2}{\min(\pi,1-\pi)(f_1+f_2)} = \frac{f_1+f_2}{\min(\pi,1-\pi)}$이다. μ가 유계인 범위에서 $f_1 + f_2$는 적분가능한 함수로 누를 수 있으므로 지배수렴으로 $I(\pi) \to 0$이다.

(d) 수치 적분으로 $I(\tfrac{1}{2}) = 3.211$이다. 상한 4의 80%다.

(e) $1 - 3.211/4 = 0.197$이다. EM 사상은 한 번에 오차를 약 5분의 1로 줄인다(정리 3.6에서 표본 평균을 기댓값으로 바꾼 것).

최종답 (d) 3.211 (e) 0.197

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.

풀이 (a) $\ell = 186\log(p^2+2pr) + 38\log(q^2+2qr) + 13\log(2pq) + 284\log r^2$.

(b) 각 유전자형 확률에 로그를 취해 더하면 힌트의 식이다. A, B, O 대립유전자가 각각 몇 번 나왔는지가 계수이고 합은 $2n$이다. $p+q+r=1$ 아래에서 Gibbs 부등식(2회 정리 8.3)으로 최대점은 각 계수를 $2n$으로 나눈 것이다.

(c) $E[n_{AA} \mid \mathbf{y}] = 186\frac{p^2}{p^2+2pr}$, $E[n_{AO} \mid \mathbf{y}] = 186\frac{2pr}{p^2+2pr}$, $E[n_{BB} \mid \mathbf{y}] = 38\frac{q^2}{q^2+2qr}$, $E[n_{BO} \mid \mathbf{y}] = 38\frac{2qr}{q^2+2qr}$이고 AB와 OO는 관측 그대로다.

(d) $p=q=r=\frac{1}{3}$이면 $\frac{p^2}{p^2+2pr}=\frac{1}{3}$이므로 $n_{AA}=62$, $n_{AO}=124$이다. $p=\frac{2\cdot 62+124+13}{1042}=\frac{261}{1042}\approx 0.2505$.
(e) $n_{BB}=\frac{38}{3}$, $n_{BO}=\frac{76}{3}$이므로 $q=\frac{76/3+76/3+13}{1042}=\frac{191}{3126}\approx 0.0611$.
(f) $\hat{p}=0.2136$. (g) $\hat{q}=0.0501$이고 $\hat{r}=0.7363$이다. 정리 3.2로 반복마다 (a)의 가능도가 늘고, 직접 최대화한 값과 같다.

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