수리통계학 × 통계학의 역사

9회 워크북 — 풀이판

최대가능도 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	MLE, 정보, 한계, 구간, 세 검정을 손으로 계산한다	14	1시간 56분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 45분
C 증명 문제	변형, 일반화, 반례 만들기	9	2시간 50분
D 퀄형	여러 절을 잇는 종합 문제. 제한 시간 안에	4	2시간 30분
	합계	34	약 10시간

노트 `09_mle_notes.pdf` — 문항의 "정리 3.1" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T1] 원소 추적, [T2] 귀류·반례, [T5] 극한, [T6] 적분 기호 아래의 미분, [T8] Taylor 전개와 o_P

검산 노트북 `09_mle.ipynb`의 "워크북 검산" 절: `check("A1a", "1/3")`

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. 그리스 문자는 `theta`, `sigma`, `lam`(`λ`)으로 쓴다.
5. **증명은 근거를 단다.** 등호와 부등호마다 어느 정칙조건, 정리, 앞 회차의 결과를 썼는지 오른쪽에 적는다. 노트의 단계형 증명과 같은 모양이다.

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

A

계산 드릴

가능도를 세우고, 최대화하고, 정보와 구간을 손으로 계산한다. 답이 수나 식으로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다. 정규분위수는 $z_{0.025}=1.96$을 쓴다.

A1 기하분포의 MLE ★

정의 1.1, 정리 1.3 · 6분

첫 성공까지의 시행 횟수 X가 $P(X=x)=p(1-p)^{x-1}$ $(x=1,2,\dots)$를 따른다. 관측값이 $3,1,4,2,5$이다. (a) p의 MLE를 구하라. (b) $P(X>2)=(1-p)^2$의 MLE를 구하라.

풀이 (a) $\ell(p)=n\log p+(\sum x_i-n)\log(1-p)$이고 $\ell'(p)=\frac{n}{p}-\frac{\sum x_i-n}{1-p}=0$에서 $\hat p=n/\sum x_i=1/\bar x$이다. $n=5$, $\sum x_i=15$이므로 $\hat p=\frac{1}{3}$이다. $\ell''<0$이므로 최대점이다. (b) 불변성(정리 1.3)으로 $(1-\hat p)^2=\frac{4}{9}$이다.

최종답 (a) $\frac{1}{3}$ (b) $\frac{4}{9}$

A2 Rayleigh 분포 ★

정의 1.1, 2.1 · 8분

밀도 $f(x;\theta)=\frac{x}{\theta}e^{-x^2/(2\theta)}$ $(x>0,\ \theta>0)$에서 관측값이 $1,2,2,3$이다. (a) θ의 MLE를 구하라. (b) 한 관측의 Fisher 정보 $I(\theta)$를 구하라. $EX^2=2\theta$를 써도 된다.

풀이 (a) $\ell(\theta)=\sum\log x_i-n\log\theta-\frac{1}{2\theta}\sum x_i^2$, $\ell'(\theta)=-\frac{n}{\theta}+\frac{\sum x_i^2}{2\theta^2}=0$에서 $\hat\theta=\frac{\sum x_i^2}{2n}=\frac{18}{8}=\frac{9}{4}$이다. (b) $\frac{\partial^2}{\partial\theta^2}\log f=\frac{1}{\theta^2}-\frac{x^2}{\theta^3}$이고 $EX^2=2\theta$이므로 $I(\theta)=-\frac{1}{\theta^2}+\frac{2\theta}{\theta^3}=\frac{1}{\theta^2}$이다.

최종답 (a) $\frac{9}{4}$ (b) $1/\theta^2$

A3 베타$(\theta,1)$ ★

정의 1.1, 정리 1.3 · 8분

밀도 $f(x;\theta)=\theta x^{\theta-1}$ $(0<x<1)$에서 관측값이 $\frac{1}{2},\frac{1}{4},\frac{1}{2},\frac{1}{8}$이다. (a) θ의 MLE를 구하라($\log 2$를 남겨 둔다). (b) 평균 $EX=\frac{\theta}{\theta+1}$의 MLE를 구하라.

풀이 (a) $\ell'(\theta)=\frac{n}{\theta}+\sum\log x_i=0$에서 $\hat\theta=-n/\sum\log x_i$이다. $\sum\log x_i=-(1+2+1+3)\log 2=-7\log 2$이므로 $\hat\theta=\frac{4}{7\log 2}\approx 0.824$이다. (b) 불변성으로 $\frac{\hat\theta}{\hat\theta+1}=\frac{4}{4+7\log 2}\approx 0.452$이다.

최종답 (a) $\frac{4}{7\log 2}\approx 0.824$ (b) $\frac{4}{4+7\log 2}\approx 0.452$

A4 $e^{-\lambda}$의 Cramér–Rao 한계 ★

정리 2.4 · 6분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$에서 $P(X=0) = e^{-\lambda}$의 불편추정량의 분산의 Cramér–Rao 하한을 λ와 n의 식으로 구하라. (노트북의 `check()`에는 λ를 `lam`으로 쓴다.)

풀이 $k'(\lambda) = -e^{-\lambda}$, $nI(\lambda) = n/\lambda$이므로 정리 2.4로 하한은 $\dfrac{e^{-2\lambda}}{n/\lambda} = \dfrac{\lambda e^{-2\lambda}}{n}$이다.

최종답 $\lambda e^{-2\lambda}/n$

A5 표준편차의 정보 ★

정의 2.1, 정리 2.4 · 8분

$X \sim \mathcal{N}(0, \sigma^2)$이고 모수는 표준편차 σ다. (a) 한 관측의 Fisher 정보 $I(\sigma)$를 구하라. (b) σ의 불편추정량의 분산의 Cramér–Rao 하한을 구하라(표본 n개).

풀이 (a) $\dfrac{\partial^2}{\partial\sigma^2}\log f = \dfrac{1}{\sigma^2} - \dfrac{3x^2}{\sigma^4}$이고 $EX^2 = \sigma^2$이므로 $I(\sigma) = -\dfrac{1}{\sigma^2} + \dfrac{3}{\sigma^2} = \dfrac{2}{\sigma^2}$이다. (b) $\dfrac{1}{nI(\sigma)} = \dfrac{\sigma^2}{2n}$이다. $\eta = \sigma^2$이면 $I^*(\eta) = I(\sigma)/(d\eta/d\sigma)^2 = \dfrac{2/\sigma^2}{4\sigma^2} = \dfrac{1}{2\sigma^4}$로 예제 2.3과 맞는다.

최종답 (a) $2/\sigma^2$ (b) $\sigma^2/(2n)$

A6 비율의 Wald 구간 ★

따름정리 3.2 · 6분

400명 가운데 248명이 찬성했다. (a) $\hat{p}$의 표준오차 $\sqrt{\hat{p}(1-\hat{p})/n}$을, (b) 95% Wald 구간의 아래 끝을, (c) 위 끝을 구하라.

풀이 $\hat{p} = 0.62$이고 $nI(\hat{p}) = \dfrac{400}{0.62\cdot 0.38}$이므로 표준오차는 $\sqrt{0.62\cdot 0.38/400} = 0.02427$이다. 구간은 $0.62 \pm 1.96\cdot 0.02427 = 0.62 \pm 0.0476$, 곧 $[0.5724,\ 0.6676]$이다.

최종답 (a) 0.0243 (b) 0.5724 (c) 0.6676

A7 지수분포의 생존확률 ★★

따름정리 3.2(b) · [T8] · 10분

수명이 평균 θ인 지수분포를 따르고, 50개의 수명의 평균이 $\bar{x} = 4$이다. (a) $P(X>5) = e^{-5/\theta}$의 MLE를 구하라. (b) 델타 방법으로 그 표준오차를 구하라. 평균 θ의 정보는 $1/\theta^2$이다.

풀이 (a) $\hat{\theta} = \bar{x} = 4$이므로 불변성으로 $e^{-5/4} = 0.2865$이다. (b) $\text{Var}\,\hat{\theta} \approx \dfrac{1}{nI(\theta)} = \dfrac{\theta^2}{n}$이므로 따름정리 3.2(b)로 표준오차는 $g'(\hat{\theta})\dfrac{\hat{\theta}}{\sqrt{n}} = \dfrac{5}{16}e^{-5/4}\cdot\dfrac{4}{\sqrt{50}} = \dfrac{5}{4\sqrt{50}}e^{-5/4} \approx 0.0506$이다.

최종답 (a) $e^{-5/4} \approx 0.2865$ (b) $\dfrac{5}{4\sqrt{50}}e^{-5/4} \approx 0.0506$

A8 Cauchy 위치의 한 걸음 ★★

알고리즘 4.1 · 10분

표준 Cauchy 위치모형 $f(x;\theta)=\frac{1}{\pi(1+(x-\theta)^2)}$에서 관측값이 $-1, 0, 3$이다. 중앙값 $\theta^{(0)}=0$에서 시작해 (a) Newton–Raphson의 한 걸음 $\theta^{(1)}$을, (b) Fisher 채점의 한 걸음을 구하라. 한 관측의 정보는 $\frac{1}{2}$이다.

풀이 $\theta=0$에서 $\ell'(0)=\frac{-2}{2}+0+\frac{6}{10}=-\frac{2}{5}$, $\ell''(0)=-\big(0+2+\frac{2\cdot(-8)}{100}\big)=-\frac{46}{25}$이다. (a) $\theta^{(1)}=0-\frac{-2/5}{-46/25}=-\frac{5}{23}\approx -0.217$. (b) $nI=\frac{3}{2}$이므로 $\theta^{(1)}=0+\frac{-2/5}{3/2}=-\frac{4}{15}\approx -0.267$. 관측된 곡률 $\frac{46}{25}$이 기대 곡률 $\frac{3}{2}$보다 커서 Newton의 걸음이 더 짧다.

최종답 (a) $-\frac{5}{23}$ (b) $-\frac{4}{15}$

A9 지수분포의 세 검정 ★★

정의 5.1 · 12분

$X_i \overset{\text{iid}}{\sim}$ Exp (비율 λ), $n=10$, $\sum x_i=15$이다. $H_0:\lambda=1$에 대해 (a) 가능도비 통계량 Λ_{LR}, (b) Wald 통계량 W, (c) 점수 통계량 R을 구하라. 한 관측의 정보는 $1/\lambda^2$이다. $W=R$이 우연인가?

풀이 (a) $\Lambda=2[\ell(\frac{2}{3})-\ell(1)]=2\big[10\log\frac{2}{3}-10+15\big]=20\log\frac{2}{3}+10\approx 1.891$. (b) $W=nI(\hat\lambda)(\hat\lambda-1)^2=\frac{10}{(2/3)^2}\cdot\frac{1}{9}=\frac{5}{2}$. (c) $\ell'(1)=n-\sum x_i=-5$, $nI(1)=10$이므로 $R=\frac{25}{10}=\frac{5}{2}$. 우연이 아니다. $W=\frac{n(\hat\lambda-\lambda_0)^2}{\hat\lambda^2}=n(1-\lambda_0\bar x)^2$이고 $R=\frac{n^2(1/\lambda_0-\bar x)^2}{n/\lambda_0^2}=n(1-\lambda_0\bar x)^2$이므로 언제나 같다. 가능도비만 다르다.

최종답 (a) $20\log\frac{2}{3}+10\approx 1.891$ (b) $\frac{5}{2}$ (c) $\frac{5}{2}$

A10 두 KL 발산 ★

보조정리 1.4 · 8분

(a) $\mathcal{N}(m_1,s^2)$에서 $\mathcal{N}(m_2,s^2)$로의 KL 발산 $D=E_1\log\frac{f_1(X)}{f_2(X)}$를 구하라. (b) Bernoulli($\frac{1}{2}$)에서 Bernoulli($\frac{1}{4}$)로의 KL 발산을 구하라. (`check()`에는 m_1, m_2, s를 `m1, m2, s`로 쓴다.)

풀이 (a) $\log\frac{f_1(x)}{f_2(x)}=\frac{(x-m_2)^2-(x-m_1)^2}{2s^2}=\frac{(m_1-m_2)(2x-m_1-m_2)}{2s^2}$이고 $E_1X=m_1$이므로 $D=\frac{(m_1-m_2)^2}{2s^2}$이다. (b) $\frac{1}{2}\log\frac{1/2}{1/4}+\frac{1}{2}\log\frac{1/2}{3/4}=\frac{1}{2}\log\frac{4}{3}\approx 0.144$이다.

최종답 (a) $\frac{(m_1-m_2)^2}{2s^2}$ (b) $\frac{1}{2}\log\frac{4}{3}\approx 0.144$

A11 평균과 분산이 같은 정규 ★★

정리 2.2 · 10분

$X \sim \mathcal{N}(\theta, \theta)$ ($\theta > 0$, 평균과 분산이 모두 θ)의 한 관측의 Fisher 정보 $I(\theta)$를 구하라. θ가 평균에만 들어 있을 때의 정보 $1/\theta$보다 큰 이유를 말하라.

풀이 $\frac{\partial}{\partial\theta}\log f = -\frac{1}{2\theta} + \frac{x-\theta}{\theta} + \frac{(x-\theta)^2}{2\theta^2}$, $\frac{\partial^2}{\partial\theta^2}\log f = \frac{1}{2\theta^2} - \frac{1}{\theta} - \frac{x-\theta}{\theta^2} - \frac{x-\theta}{\theta^2} - \frac{(x-\theta)^2}{\theta^3}$이다. 기댓값을 취하면 $\frac{1}{2\theta^2} - \frac{1}{\theta} - \frac{1}{\theta^2}$이므로 $I(\theta) = \frac{1}{\theta} + \frac{1}{2\theta^2}$이다. 첫 항은 평균에서, 둘째 항은 분산에서 오는 정보다. 자료의 흩어짐도 θ에 대해 알려 준다.

최종답 $\frac{1}{\theta} + \frac{1}{2\theta^2}$

A12 균등분포의 가능도비 ★★

정리 5.2 · 10분

$X_i \overset{\text{iid}}{\sim} \mathrm{U}(0,\theta)$, $n = 5$, $x_{(5)} = 0.8$이고 $H_0 : \theta = 1$이다. (a) $\Lambda = 2[\ell(\hat\theta) - \ell(1)]$을 구하라. (b) H_0 아래에서 $\Lambda = -2n\log X_{(n)}$이 정확히 $\chi^2(2)$, 곧 평균 2인 지수분포를 따름을 이용해 $P_{H_0}(\Lambda > 3.84)$를 구하라. Wilks의 정리와 다른 이유는?

풀이 (a) $\ell(\hat\theta) - \ell(1) = -5\log 0.8$이므로 $\Lambda = -10\log 0.8 \approx 2.231$이다. (b) $P(\Lambda > c) = P(X_{(n)}^n < e^{-c/2}) = e^{-c/2}$이므로 $e^{-1.92} \approx 0.147$이다. 공칭 0.05의 세 배다. 받침이 θ에 의존해 정칙조건 (R1)이 깨지고, 최대점이 가능도 방정식의 근이 아니다. 정리 5.2의 첫 단계인 $\ell'(\hat\theta) = 0$이 성립하지 않는다.

최종답 (a) $-10\log 0.8 \approx 2.231$ (b) $e^{-1.92} \approx 0.147$

A13 Laplace 위치모형의 효율 ★★

정의 2.5 · 8분

밀도 $f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}$에서 (a) 한 관측의 Fisher 정보 $E\,s^2$, (b) $n\,\mathrm{Var}\,\bar X$, (c) 점근 효율 $\frac{1/(nI)}{\mathrm{Var}\,\bar X}$를 구하라.

풀이 (a) $s = \mathrm{sign}(x-\theta)$이므로 $s^2 = 1$ (확률 1)이고 $I = 1$이다. (b) $\mathrm{Var}\,X = \int x^2 \frac{1}{2}e^{-|x|}dx = 2$이므로 $n\,\mathrm{Var}\,\bar X = 2$. (c) $\frac{1/n}{2/n} = \frac{1}{2}$. 표본평균은 정보의 절반을 버린다. MLE인 표본중앙값은 점근분산 $\frac{1}{n}$으로 한계에 닿는다.

최종답 (a) 1 (b) 2 (c) $\frac{1}{2}$

A14 정보로 표본 크기 정하기 ★

따름정리 3.2 · 6분

찬성 비율이 대략 0.3으로 예상된다. 95% Wald 구간의 반폭이 0.02 이하가 되게 하는 가장 작은 n을 구하라.

풀이 $1.96\sqrt{0.21/n} \le 0.02 \iff n \ge \frac{1.96^2 \cdot 0.21}{0.0004} = 2016.84$이므로 $n = 2017$이다. 정보가 n에 비례하므로 반폭을 절반으로 줄이려면 표본이 네 배 필요하다.

최종답 2017

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 정칙조건의 번호, 정리 번호, 앞 회차의 정리, 도구 번호를 쓴다.

B1 MLE의 점근정규성, 근거 채우기 ★★

① · 정리 3.1 · 15분

정칙조건 아래 $\hat{\theta}_n$이 가능도 방정식의 일치하는 근이면 $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, 1/I(\theta_0))$이다. 각 단계의 근거를 써라.

① $0 = \ell'(\hat{\theta}_n) = \ell'(\theta_0) + (\hat{\theta}_n - \theta_0)\ell''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2\ell'''(\theta_n^*)$이다. — Taylor, (R3) [T8]

② $\sqrt{n}(\hat{\theta}_n - \theta_0) = \dfrac{n^{-1/2}\ell'(\theta_0)}{-n^{-1}\ell''(\theta_0) - \frac{1}{2}(\hat{\theta}_n - \theta_0)n^{-1}\ell'''(\theta_n^*)}$이다. — ①을 정리

③ $n^{-1/2}\ell'(\theta_0) \xrightarrow{D} \mathcal{N}(0, I(\theta_0))$이다. — 정리 2.2, 7회 정리 3.1

④ $-n^{-1}\ell''(\theta_0) \xrightarrow{P} I(\theta_0)$이다. — 정리 2.2, 7회 정리 1.6

⑤ $\frac{1}{2}(\hat{\theta}_n - \theta_0)\, n^{-1}\ell'''(\theta_n^*) \xrightarrow{P} 0$이다. — (R5), 정리 1.6, 7회 정리 2.5

⑥ $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, 1/I(\theta_0))$이다. — Slutsky, 7회 정리 4.2

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 3.1의 단계형 증명과 같다. 다섯째 단계에서 θ_n^*가 (R5)의 근방 안에 들어갈 확률이 1로 가는 것은 $|\theta_n^* - \theta_0| \le |\hat{\theta}_n - \theta_0|$와 일치성(정리 1.6) 때문이다.

B2 Cramér–Rao 부등식, 근거 채우기 ★★

① · 정리 2.4 · 15분

정칙조건 아래 $E_\theta T = k(\theta)$이고 $\mathrm{Var}_\theta\, T < \infty$이면 $\mathrm{Var}_\theta\, T \ge k'(\theta)^2/(nI(\theta))$이다. 근거를 채워라.

① $k(\theta) = \int T(\mathbf{x})L(\theta;\mathbf{x})\,d\mathbf{x}$이다. — 기댓값의 정의

② $k'(\theta) = \int T\,\dfrac{\partial L}{\partial\theta}\,d\mathbf{x} = E_\theta[T\,\ell'(\theta)]$이다. — (R4) 미분과 적분의 교환 [T6], $\partial_\theta L = \ell' L$

③ $E_\theta[T\ell'(\theta)] = \mathrm{Cov}(T - k(\theta), \ell'(\theta))$이다. — $E_\theta\ell'(\theta) = 0$, 정리 2.2

④ $k'(\theta)^2 \le \mathrm{Var}\,T \cdot \mathrm{Var}\,\ell'(\theta)$이다. — Cauchy–Schwarz, 3회 정리 4.4

⑤ $\mathrm{Var}\,\ell'(\theta) = nI(\theta)$이다. — 정리 2.2(독립합의 분산)

풀이 빈칸의 빨간 글씨가 근거다. 둘째 단계에서 엄밀하게는 $T \cdot L$에 대해 순서를 바꾸므로 $|T\,\partial_\theta L|$이 적분가능한 함수로 눌려야 한다(노트의 조건 확인). 넷째 단계의 등호는 $T - k(\theta)$가 $\ell'(\theta)$의 상수배일 때뿐이다.

B3 기대 로그가능도의 최대, 근거 채우기 ★★

① · 보조정리 1.4 · 15분

식별가능하고 기댓값이 존재하면 $\theta \neq \theta_0$에서 $E_{\theta_0} \log f(X;\theta) < E_{\theta_0} \log f(X;\theta_0)$이다. 근거를 채워라. $A = \{x : f(x;\theta_0) > 0\}$, $Y = f(X;\theta)/f(X;\theta_0)$이다.

① Y는 확률 1로 정의되고 $Y \geq 0$이다. — $P_{\theta_0}(X \in A) = 1$, 받침의 정의

② $E_{\theta_0} \log Y \leq \log E_{\theta_0} Y$이다. — log는 오목, Jensen(2회 정리 5.9)

③ $E_{\theta_0} Y = \int_A f(x;\theta)\,dx \leq 1$이다. — 기댓값의 정의, 밀도의 적분은 1

④ $E_{\theta_0} \log f(X;\theta) \leq E_{\theta_0} \log f(X;\theta_0)$이다. — ②③, log의 단조성

⑤ 등호이면 $Y \equiv 1$이고 두 밀도가 같다. — log의 순오목, Jensen의 등호 조건

⑥ $\theta \neq \theta_0$이면 부등식은 엄격하다. — 식별가능성 [T2]

풀이 빈칸의 빨간 글씨가 근거다. 다섯째 단계에서 Jensen의 등호로 Y가 상수 c이고, 셋째 단계의 등호와 $\log c = 0$에서 $c = 1$이다. 그러면 A 위에서 두 밀도가 같고, $\int_A f(x;\theta)dx = 1$이므로 A 밖에서는 $f(x;\theta) = 0$이다.

B4 MLE의 점근정규성, 단계의 이름만 ★★

② · 정리 3.1 · 25분

정리 3.1을 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **가능도 방정식을 θ_0 둘레에서 Taylor 전개한다**

② **$\sqrt{n}(\hat\theta_n - \theta_0)$를 분자와 분모의 비로 쓴다**

③ **분자의 극한분포**

④ **분모의 첫 항의 확률극한**

⑤ **분모의 둘째 항이 0으로 감**

⑥ **Slutsky로 합친다**

풀이

① $0 = \ell'(\hat\theta_n) = \ell'(\theta_0) + (\hat\theta_n - \theta_0)\ell''(\theta_0) + \frac{1}{2}(\hat\theta_n - \theta_0)^2 \ell'''(\theta_n^*)$. — (R3) [T8]

② $\sqrt{n}(\hat\theta_n - \theta_0) = \dfrac{n^{-1/2}\ell'(\theta_0)}{-n^{-1}\ell''(\theta_0) - \frac{1}{2}(\hat\theta_n - \theta_0)n^{-1}\ell'''(\theta_n^*)}$. — ①

③ $\ell'(\theta_0) = \sum s(X_i;\theta_0)$, 평균 0, 분산 I: $n^{-1/2}\ell'(\theta_0) \xrightarrow{D} \mathcal{N}(0, I)$. — 정리 2.2, CLT

④ $-n^{-1}\ell''(\theta_0) \xrightarrow{P} I(\theta_0)$. — 정리 2.2, Khinchin

⑤ $|n^{-1}\ell'''(\theta_n^*)| \le n^{-1}\sum M(X_i)$가 확률적으로 유계이고 $\hat\theta_n - \theta_0 \xrightarrow{P} 0$. (R5), 정리 1.6

⑥ 분모 $\xrightarrow{P} I$이므로 극한은 $\mathcal{N}(0,I)/I = \mathcal{N}(0,1/I)$. Slutsky

B5 Cramér–Rao 부등식, 단계의 이름만 ★★

② · 정리 2.2, 2.4 · 25분

정리 2.4를 증명하라. 증명에 쓰이는 정리 2.2의 두 사실($E\ell' = 0$, $\operatorname{Var}\ell' = nI$)도 $\int f\,dx = 1$에서 출발해 함께 보여라.

① $\int f\,dx = 1$**을 미분해 점수의 평균이 0임을 보인다**

② $k(\theta)$**를 적분으로 쓰고 미분한다**

③ $k'(\theta)$**를** T**와 점수의 공분산으로 바꾼다**

④ **Cauchy–Schwarz를 쓴다**

⑤ **점수의 분산을 정보로 바꾸고 정리한다**

풀이

① $0 = \frac{d}{d\theta}\int f\,dx = \int s f\,dx = E\,s$, 표본이면 $E\ell' = 0$. (R4) [T6]

② $k'(\theta) = \int T\,\partial_\theta L\,d\mathbf{x} = E[T\ell'(\theta)]$. (R4)를 TL에

③ $E[T\ell'] = \operatorname{Cov}(T-k, \ell')$. ①

④ $k'(\theta)^2 \le \operatorname{Var}T \cdot \operatorname{Var}\ell'$. 3회 정리 4.4

⑤ $\operatorname{Var}\ell' = \sum \operatorname{Var}s(X_i) = nI(\theta)$; 나누면 끝. 독립합의 분산

B6 MLE의 점근정규성, 전략만 ★★

③ · 정리 3.1 · 35분

정리 3.1을 증명하라. 쓰는 정칙조건을 단계마다 밝혀라.

전략 가능도 방정식을 θ_0 둘레에서 Taylor 전개해 분자와 분모의 비로 쓴다. 분자는 중심극한정리로, 분모는 대수의 법칙으로 다루고 Slutsky로 합친다.

풀이 B4의 풀이와 같다. (R3)은 Taylor 전개에, (R4)는 정리 2.2를 거쳐 분자의 평균과 분산, 분모의 극한에, (R5)와 일치성은 나머지항을 누르는 데 쓰인다. $0 < I(\theta_0)$은 마지막에 분모로 나누는 데, $I < \infty$는 중심극한정리에 쓰인다.

B7 기대 로그가능도의 최대, 전략만 ★★

③ · 보조정리 1.4 · 35분

보조정리 1.4를 증명하라. 등호의 경우까지 다룬다.

전략 두 밀도의 비의 로그에 Jensen 부등식을 써서 기댓값과 로그의 순서를 바꾸고, 밀도의 적분이 1을 넘지 않음을 쓴다. 등호의 경우를 따로 보아 식별가능성으로 엄격한 부등식을 얻는다.

풀이 노트 보조정리 1.4의 단계형 증명과 같다(B3의 풀이). $\int_A f(x;\theta)dx < 1$인 경우는 $f(\cdot;\theta)$가 A 밖에도 질량을 둘 때다. 그때는 등호가 될 수 없으므로 부등식이 바로 엄격하다. 이산이면 적분을 합으로 바꾸면 된다.

C

증명 문제

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

C1 MLE가 하나가 아닐 때 ★★

[T1], [T5] · 정의 1.1 · 15분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{U}(\theta, \theta+1)$이다.

(a) $L(\theta) = 1\{x_{(n)} - 1 \le \theta \le x_{(1)}\}$임을 보여라.

(b) 구간 $[x_{(n)} - 1,\ x_{(1)}]$의 모든 점이 MLE임을 보여라.

(c) 어떤 MLE $\hat\theta_n$을 고르든 $\hat\theta_n \xrightarrow{P} \theta$임을 보여라.

풀이 (a) $L(\theta) = \prod_i 1\{\theta \le x_i \le \theta+1\}$이고, 모든 i에서 $\theta \le x_i$인 것은 $\theta \le x_{(1)}$과, 모든 i에서 $x_i \le \theta+1$인 것은 $\theta \ge x_{(n)}-1$과 같다 [T1]. (b) L은 그 구간에서 1, 밖에서 0이므로 구간의 모든 점에서 최댓값 1을 갖는다. 관측값은 모두 $(\theta, \theta+1)$ 안에 있으므로 구간은 비어 있지 않다. (c) 참값 θ도 구간 안에 있으므로 $|\hat\theta_n - \theta| \le 1 - (X_{(n)} - X_{(1)})$이다. 구간의 길이가 ε보다 크면 $X_{(1)} > \theta + \frac{\varepsilon}{2}$이거나 $X_{(n)} < \theta + 1 - \frac{\varepsilon}{2}$이다. Boole 부등식으로 $P(|\hat\theta_n - \theta| > \varepsilon) \le 2(1-\frac{\varepsilon}{2})^n \to 0$이다. 수렴 속도는 $1/n$이다(정칙조건 (R1)이 깨지는 모형).

C2 가능도가 끝없이 커질 때 ★★

[T5] · 정의 1.1, 정리 1.6 · 15분

밀도 $f(x;\mu,\sigma) = \frac{1}{2}\phi(x) + \frac{1}{2\sigma}\phi\big(\frac{x-\mu}{\sigma}\big)$ (ϕ는 표준정규 밀도)에서 서로 다른 관측값 $x_1, \dots, x_n$을 얻었다.

(a) $\mu = x_1$이면 $L(x_1, \sigma) \ge \frac{\phi(0)}{2\sigma} \prod_{i \ge 2} \frac{1}{2}\phi(x_i)$임을 보여라.

(b) $\sigma \to 0$이면 $L(x_1, \sigma) \to \infty$, 곧 (μ, σ)의 MLE가 없음을 보여라.

(c) 이것이 가능도 방정식의 일치하는 근이 있다는 정리 1.6과 어긋나지 않는 이유를 설명하라.

풀이 (a) 두 항이 모두 0 이상이므로 $f(x_1; x_1, \sigma) \ge \frac{1}{2\sigma}\phi(0)$이고 $i \ge 2$에서 $f(x_i; x_1, \sigma) \ge \frac{1}{2}\phi(x_i)$이다. 곱한다. (b) (a)의 오른쪽은 $\sigma \to 0$에서 $\frac{C}{\sigma} \to \infty$ ($C > 0$은 σ에 무관)이다. 가능도에 최댓값이 없으므로 MLE가 없다. (c) 정리 1.6은 참값 근처의 **극대점**이 확률 1로 가는 사건 위에서 있다는 주장이다. 전역 최대점의 존재를 말하지 않는다. 무한대로 가는 것은 $\sigma \to 0$인 경계 쪽의 가능도이고, 정리 1.6의 증명은 θ_0의 근방만 본다.

C3 Cramér–Rao 한계에 닿는 추정량 ★★★

[T2] · 정리 2.4, 3회 정리 4.4 · 30분

(a) 정리 2.4의 등호가 성립할 필요충분조건이 "확률 1로 $T - k(\theta) = a(\theta)\,\ell'(\theta)$인 상수 $a(\theta)$가 있다" 임을 보여라.

(b) $X_i \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$이면 $\ell'(\lambda) = n(\bar{x}/\lambda - 1)$이다. 모든 λ에서 한계에 닿는 불편추정량을 갖는 $k(\lambda)$는 λ의 일차함수뿐임을 보여라. 따라서 $e^{-\lambda}$의 어떤 불편추정량도 한계에 닿지 않는다.

(c) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\theta, 1)$에서 θ^2의 불편추정량 $\bar{X}^2 - \frac{1}{n}$의 분산과 Cramér–Rao 하한을 비교하라.

풀이 (a) 정리 2.4의 증명에서 부등식은 Cauchy–Schwarz 한 번뿐이다. 그 등호는 $T - k$와 ℓ'가 확률 1로 일차 관계일 때이고, 둘 다 평균이 0이므로 상수항 없이 $T - k = a\ell'$이다(3회 정리 4.4의 등호 조건). (b) (a)로 $T = k(\lambda) + a(\lambda)n(\bar{x}/\lambda - 1) = \big[k(\lambda) - n a(\lambda)\big] + \frac{n a(\lambda)}{\lambda}\bar{x}$이다. 왼쪽은 λ에 의존하지 않고 $\bar{x}$는 둘 이상의 값을 양의 확률로 가지므로, 기울기와 절편이 λ에 무관한 상수 B, A다. 곧 $T = A + B\bar{X}$이고 $k(\lambda) = ET = A + B\lambda$다. $e^{-\lambda}$는 일차함수가 아니므로 한계에 닿는 불편추정량이 없다 [T2]. (c) $\text{Var}(\bar{X}^2 - \frac{1}{n}) = \frac{4\theta^2}{n} + \frac{2}{n^2}$이고, $k(\theta) = \theta^2$, $I = 1$이므로 하한은 $\frac{(2\theta)^2}{n} = \frac{4\theta^2}{n}$이다. 차이 $\frac{2}{n^2}$는 n보다 빨리 줄어 점근적으로는 한계에 닿는다. $\theta = 0$에서는 하한이 0인데 분산은 $\frac{2}{n^2} > 0$이다.

C4 정보는 더해진다 ★★

정의 2.1, 정리 2.2 · 20분

(a) $X \sim f(x;\theta)$와 $Y \sim g(y;\theta)$가 독립이면 (X,Y)의 정보가 $I_X(\theta)+I_Y(\theta)$임을 보여라.

(b) 비율 λ인 Poisson 과정을 겹치지 않는 길이 t_1, t_2의 두 구간에서 세면 정보는 $\frac{t_1+t_2}{\lambda}$임을 보여라.

(c) 수명 $X \sim \text{Exp}$ (비율 λ)를 시간 c까지만 지켜보아 $\mathbb{1}\{X \le c\}$만 안다. 이 관측의 정보가 $\frac{c^2 e^{-\lambda c}}{1-e^{-\lambda c}}$임을 보여라.

(d) $u=\lambda c=1$이면 (c)의 정보는 완전한 관측의 정보 $1/\lambda^2$의 몇 배인가?

풀이 (a) $\log[f(x;\theta)g(y;\theta)]$의 점수는 s_X+s_Y이고 두 점수는 독립이며 평균이 0이다. $\text{Var}(s_X+s_Y)=I_X+I_Y$이다. (b) 두 구간의 개수는 독립인 $\text{Poi}(\lambda t_j)$이다. $\text{Poi}(\lambda t)$의 λ에 대한 점수는 $\frac{N}{\lambda}-t$이고 분산은 $\frac{\lambda t}{\lambda^2}=\frac{t}{\lambda}$이다. (a)로 더한다. (c) $q(\lambda)=1-e^{-\lambda c}$이고 Bernoulli$(q)$의 q에 대한 정보는 $\frac{1}{q(1-q)}$이다. 모수를 바꾸면(노트 연습 2.2) λ에 대한 정보는 $\frac{q'(\lambda)^2}{q(1-q)}=\frac{c^2e^{-2\lambda c}}{(1-e^{-\lambda c})e^{-\lambda c}}=\frac{c^2e^{-\lambda c}}{1-e^{-\lambda c}}$이다. (d) 비는 $\frac{u^2e^{-u}}{1-e^{-u}}$이고 $u=1$이면 $\frac{1}{e-1}\approx 0.582$이다. 끝났는지만 보면 정보의 42%를 잃는다.

최종답 (c) $\frac{c^2e^{-\lambda c}}{1-e^{-\lambda c}}$ (d) $\frac{1}{e-1}\approx 0.582$

C5 치우친 추정량의 한계 ★★

정리 2.4 · 15분

(a) $E_\theta T=\theta+b(\theta)$이면 $E_\theta(T-\theta)^2 \ge b(\theta)^2+\frac{(1+b'(\theta))^2}{nI(\theta)}$임을 보여라.

(b) $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\theta,1)$, $T=c\bar X$ $(0<c<1)$의 평균제곱오차를 구하고 (a)의 하한과 같음을 보여라.

(c) $\theta=0$ 근처에서 $c\bar X$의 평균제곱오차는 $\frac{1}{n}$보다 작다. 정리 2.4와 모순이 아닌 이유를 말하라.

풀이 (a) $E(T-\theta)^2=\text{Var}\,T+b(\theta)^2$이고 정리 2.4로 $\text{Var}\,T \ge \frac{k'(\theta)^2}{nI}=\frac{(1+b'(\theta))^2}{nI}$이다. (b) $E(c\bar X-\theta)^2=\frac{c^2}{n}+(1-c)^2\theta^2$이다. $b(\theta)=(c-1)\theta$, $b'=c-1$, $I=1$이므로 하한은 $(c-1)^2\theta^2+\frac{c^2}{n}$으로 같다. $c\bar X-c\theta=\frac{c}{n}\ell'(\theta)$가 점수의 상수배이므로 C3(a)의 등호 조건이 성립한다. (c) 정리 2.4의 $\frac{1}{nI}$은 불편추정량의 한계다. $c\bar X$는 치우쳐 있어 (a)의 하한을 따르고, θ가 0 근처이면 치우침의 제곱이 작아 $\frac{1}{n}$ 아래로 내려간다. 대가는 $|\theta|$가 클 때 $(1-c)^2\theta^2$로 치른다.

C6 KL 발산은 거리가 아니다 ★

[T2] · 보조정리 1.4 · 10분

Poisson 사이의 KL 발산은 $D(\lambda\|\lambda') = \lambda\log\frac{\lambda}{\lambda'} - \lambda + \lambda'$ 이다. (a) $D(1\|2)$와 (b) $D(2\|1)$을 구해 대칭이 아님을 보여라. (c) $D(1\|4) - \big[D(1\|2) + D(2\|4)\big]$을 구해 삼각부등식이 깨짐을 보여라.

풀이 (a) $D(1\|2) = \log\frac{1}{2} - 1 + 2 = 1 - \log 2 \approx 0.307$. (b) $D(2\|1) = 2\log 2 - 2 + 1 = 2\log 2 - 1 \approx 0.386$. 서로 다르다. (c) $D(1\|4) = 3 - 2\log 2$, $D(2\|4) = 2\log\frac{1}{2} - 2 + 4 = 2 - 2\log 2$이므로 차는 $(3 - 2\log 2) - (3 - 3\log 2) = \log 2 > 0$이다. 거리라면 이 차는 0 이하여야 한다 [T2]. KL 발산은 "거리"가 아니라 정보량의 차이다. 그래도 가까운 두 분포에서는 $\frac{1}{2}I\delta^2$로 대칭인 이차식이 된다(노트 8.2절).

최종답 (a) $1 - \log 2 \approx 0.307$ (b) $2\log 2 - 1 \approx 0.386$ (c) $\log 2$

C7 지수족의 가능도는 오목하다 ★★

정리 2.2 · 25분

$f(x;\theta) = h(x)\exp\big(\theta T(x) - A(\theta)\big)$, θ는 열린 구간에 있고 $A''(\theta) > 0$이며, $\int f\,dx = 1$을 적분 기호 안에서 두 번 미분할 수 있다.

(a) $E_\theta T(X) = A'(\theta)$, $\mathrm{Var}_\theta\, T(X) = A''(\theta)$임을 보여라.

(b) $\ell''(\theta) = -nA''(\theta) < 0$을 보이고, 가능도 방정식 $\bar{T} = A'(\theta)$의 근은 있으면 하나뿐이며 전역 최대점임을 보여라.

(c) $I(\theta) = A''(\theta)$이고 관측 정보 $-\ell''(\theta)$가 자료에 의존하지 않음을 보여라. Newton–Raphson과 Fisher 채점이 같은 이유다.

(d) Poisson을 이 꼴로 쓰면 $\theta = \log\lambda$, $T(x) = x$, $A(\theta) = e^\theta$이다. θ의 MLE와 정보 $I(\theta)$를 구하라.

풀이 (a) 점수는 $s = T(x) - A'(\theta)$이므로 정리 2.2의 $Es = 0$에서 $ET = A'(\theta)$이고, $\frac{\partial^2}{\partial\theta^2}\log f = -A''(\theta)$이므로 $\mathrm{Var}\,T = \mathrm{Var}\,s = I(\theta) = A''(\theta)$이다. (b) $\ell(\theta) = \theta\sum T(x_i) - nA(\theta) +$ 상수이므로 $\ell'' = -nA'' < 0$이다. ℓ'가 순감소하므로 영점은 많아야 하나이고, 영점 $\hat\theta$의 왼쪽에서 $\ell' > 0$, 오른쪽에서 $\ell' < 0$이므로 ℓ은 $\hat\theta$에서 전역 최대다. (c) (a)와 같이 $I = A''(\theta)$이고 $-\ell''(\theta) = nA''(\theta) = nI(\theta)$는 자료를 포함하지 않는다. 알고리즘 4.1의 두 분모가 같다. (d) 가능도 방정식 $\bar{x} = e^\theta$에서 $\hat\theta = \log\bar{x}$ (불변성과도 맞는다), $I(\theta) = A''(\theta) = e^\theta = \lambda$이다. 노트 연습 2.2의 공식 $I(\lambda)\cdot(d\lambda/d\theta)^2 = \frac{1}{\lambda}\lambda^2 = \lambda$와 같다.

최종답 (d) $I(\theta) = e^\theta$

C8 정규 평균의 가능도비는 정확히 χ^2이다 ★★

정리 5.2, 6회 정리 3.4 · 25분

$X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, $H_0 : \mu = \mu_0$이다.

(a) σ^2을 알면 $\Lambda_{\text{LR}} = W = R = \frac{n(\bar{X}-\mu_0)^2}{\sigma^2}$이고, 이것이 모든 n에서 정확히 $\chi^2(1)$임을 보여라.

(b) σ^2을 모르면 두 가설 아래에서 σ^2도 최대화한 가능도비가 $\Lambda = n\log\!\left(1+\frac{t^2}{n-1}\right)$임을 보여라. t는 일표본 t 통계량이다.

(c) (b)의 Λ는 $|t|$의 증가함수이므로 가능도비 검정은 t 검정과 같음을 설명하고, $n \to \infty$에서 $\Lambda - t^2 \xrightarrow{P} 0$임을 보여라.

풀이 (a) $\ell(\mu) = -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 +$상수는 μ의 이차식이고 $\hat{\mu} = \bar{x}$, $\ell(\bar{x}) - \ell(\mu_0) = \frac{n(\bar{x}-\mu_0)^2}{2\sigma^2}$이다. $I = 1/\sigma^2$로 상수이므로 W도 같고, $\ell'(\mu_0) = \frac{n(\bar{x}-\mu_0)}{\sigma^2}$이므로 R도 같다. $\sqrt{n}(\bar{X}-\mu_0)/\sigma \sim \mathcal{N}(0,1)$이므로 정확히 $\chi^2(1)$이다. 로그가능도가 이차식이면 Taylor 전개에 나머지항이 없다. (b) 최대화한 로그가능도는 $-\frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2} +$상수이므로 $\Lambda = n\log\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}$이다. $\sum(x_i-\mu_0)^2 = \sum(x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2$이므로 $\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = 1 + \frac{n(\bar{x}-\mu_0)^2}{\sum(x_i-\bar{x})^2} = 1 + \frac{t^2}{n-1}$이다($t^2 = \frac{n(\bar{x}-\mu_0)^2}{S^2}$). (c) $\log(1+u)$는 증가하므로 $\Lambda > c \iff |t| > c'$이다. $\log(1+u) = u + O(u^2)$이고 $t^2 = O_P(1)$이므로 $\Lambda = \frac{n}{n-1}t^2 + O_P(\frac{1}{n})$, 곧 $\Lambda - t^2 \xrightarrow{P} 0$이다 [T8].

C9 위치모형의 MLE는 등변이다 ★★

[T1] · 정의 1.1, 8회 정리 2.2 · 15분

위치모형 $f(x;\theta) = f_0(x-\theta)$에서 MLE가 자료마다 하나로 정해진다고 하자.

(a) 자료를 모두 c만큼 옮기면 $\hat{\theta}(\mathbf{x}+c) = \hat{\theta}(\mathbf{x}) + c$임을 보여라.

(b) $\hat{\theta} - \theta$의 분포가 θ에 의존하지 않음을 보여라.

(c) 그러므로 $\hat{\theta}-\theta$의 분위수 $q_{\alpha/2}, q_{1-\alpha/2}$로 만든 구간 $[\hat{\theta}-q_{1-\alpha/2},\ \hat{\theta}-q_{\alpha/2}]$의 포함확률은 모든 θ에서 정확히 $1-\alpha$임을 보여라.

풀이 (a) $L(\theta;\mathbf{x}+c) = \prod f_0(x_i+c-\theta) = L(\theta-c;\mathbf{x})$이므로 θ가 $L(\cdot;\mathbf{x}+c)$를 최대로 하는 것과 $\theta-c$가 $L(\cdot;\mathbf{x})$를 최대로 하는 것은 같다. 최대점이 하나이므로 $\hat{\theta}(\mathbf{x}+c)-c = \hat{\theta}(\mathbf{x})$이다. (b) $X_i = \theta + \varepsilon_i$이면 (a)로 $\hat{\theta}(\mathbf{X}) = \theta + \hat{\theta}(\boldsymbol{\varepsilon})$이다. $\hat{\theta}-\theta = \hat{\theta}(\boldsymbol{\varepsilon})$의 분포는 f_0와 n에만 의존한다. (c) $P(q_{\alpha/2} \le \hat{\theta}-\theta \le q_{1-\alpha/2}) = 1-\alpha$이고, 이 사건은 $\theta \in [\hat{\theta}-q_{1-\alpha/2}, \hat{\theta}-q_{\alpha/2}]$와 같다. $\hat{\theta}-\theta$가 축량이므로 8회 정리 2.2의 축량 방법이다. 분위수는 $\theta=0$에서 시뮬레이션으로 구할 수 있다.

D

퀄형

여러 절의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 가능도를 정확히 적는 것부터 답안의 일부다. 정규분위수는 $z_{0.025} = 1.96$, $\chi^2_{1,0.05} = 3.84$를 쓴다.

D1 중도절단된 수명 ★★★

정리 1.6, 2.2, 3.1 · 제한 45분

전구 n개의 수명 $X_i \overset{\text{iid}}{\sim} \text{Exp}$ (비율 λ)를 시각 c까지만 지켜본다. 관측은 $Y_i = \min(X_i, c)$와 $\delta_i = 1\{X_i \le c\}$이다. $d = \sum \delta_i$는 꺼진 전구의 수, $T = \sum Y_i$는 총 관측 시간이다.

(a) 한 관측 (y, δ)의 가능도가 $\lambda^{\delta} e^{-\lambda y}$임을 보이고, $d \ge 1$이면 MLE가 $\hat\lambda = d/T$임을 보여라.

(b) 한 관측의 Fisher 정보가 $I(\lambda) = \frac{1 - e^{-\lambda c}}{\lambda^2}$임을 보여라. (`check()`에는 λ를 `lam`으로 쓴다.)

(c) $n = 50$, $d = 30$, $T = 120$이면 $\hat\lambda$는 얼마인가?

(d) (c)에서 관측 정보 $-\ell''(\hat\lambda)$로 만든 표준오차를 구하고, 95% 구간을 써라.

풀이 (a) $\delta = 1$이면 $X = y \le c$의 밀도 $\lambda e^{-\lambda y}$, $\delta = 0$이면 $P(X > c) = e^{-\lambda c} = e^{-\lambda y}$ $(y = c)$이다. 둘을 합치면 $\lambda^{\delta} e^{-\lambda y}$이다. $\ell(\lambda) = d \log \lambda - \lambda T$, $\ell'(\lambda) = \frac{d}{\lambda} - T = 0$에서 $\hat\lambda = d/T$이고 $\ell'' = -d/\lambda^2 < 0$이므로 최대점이다. $d = 0$이면 ℓ은 증가하지 않고 최대점이 $\lambda = 0$ 쪽 경계에 있다. (b) 한 관측의 $\frac{\partial^2}{\partial \lambda^2} \log f = -\frac{\delta}{\lambda^2}$이므로 $I(\lambda) = \frac{E\delta}{\lambda^2} = \frac{1 - e^{-\lambda c}}{\lambda^2}$이다 (정리 2.2). $c \to \infty$이면 완전한 관측의 $1/\lambda^2$이다. 지켜보지 못한 부분만큼 정보가 준다. (c) $\hat\lambda = 30/120 = 0.25$. (d) $-\ell''(\hat\lambda) = d/\hat\lambda^2$이므로 표준오차는 $\hat\lambda/\sqrt{d} = 0.25/\sqrt{30} \approx 0.0456$이고, 95% 구간은 $0.25 \pm 0.0894 = [0.161,\ 0.339]$이다(따름정리 3.2(a)).

최종답 (b) $\frac{1 - e^{-\lambda c}}{\lambda^2}$ (c) 0.25 (d) $\frac{0.25}{\sqrt{30}} \approx 0.0456$

D2 Neyman–Scott, k번 측정 ★★★

정리 3.1, 7회 정리 1.6 · 제한 30분

n명에게서 k번씩 측정한다: $X_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \dots, n$, $j = 1, \dots, k$ ($k \ge 2$ 고정).

(a) $(\mu_1, \dots, \mu_n, \sigma^2)$의 MLE를 구하라.

(b) $n \to \infty$에서 $\hat\sigma^2$의 확률극한을 k와 σ^2의 식으로 구하라. (`check()`에는 σ를 `sigma`로 쓴다.)

(c) $\hat\sigma^2$에 곱해 일치추정량을 만드는 상수를 구하라. 정리 3.1의 어느 가정이 깨졌는가?

풀이 (a) σ^2을 고정하면 μ_i는 사람 i의 제곱합만 최소화하므로 $\hat\mu_i = \bar X_i$이다. 이것을 넣고 σ^2로 미분하면 $\hat\sigma^2 = \frac{1}{nk}\sum_i\sum_j (X_{ij}-\bar X_i)^2$이다. (b) $W_i = \sum_j (X_{ij}-\bar X_i)^2$는 iid이고 $EW_i = (k-1)\sigma^2$이다(6회 정리 2.2, $\chi^2(k-1)$의 평균). 대수의 법칙으로 $\hat\sigma^2 = \frac{1}{k}\cdot\frac{1}{n}\sum W_i \xrightarrow{P} \frac{k-1}{k}\sigma^2$이다. (c) $\frac{k}{k-1}$을 곱하면 $\frac{1}{n(k-1)}\sum W_i \xrightarrow{P} \sigma^2$이다. 정리 3.1은 모수의 수가 고정되어 있다고 가정했는데, 여기서는 모수가 $n+1$개로 자료와 함께 는다. $k=2$이면 노트 예제 7.2의 $\frac12$이다.

최종답 (b) $\frac{k-1}{k}\sigma^2$ (c) $\frac{k}{k-1}$

D3 Hardy–Weinberg 평형 ★★★

정리 2.2, 5.2, 8회 정리 5.1 · 제한 40분

유전자형 AA, Aa, aa의 확률이 Hardy–Weinberg 모형에서 θ^2, $2\theta(1-\theta)$, $(1-\theta)^2$이다. 100명을 조사해 $(n_1, n_2, n_3) = (50, 30, 20)$을 얻었다.

(a) θ의 MLE가 $\frac{2n_1+n_2}{2n}$임을 보이고 값을 구하라.

(b) 한 사람의 Fisher 정보 $I(\theta)$를 구하라.

(c) 모형이 맞는지를 일반 다항 모형에 대해 가능도비로 검정하라. $G = 2\sum O\log\frac{O}{E}$를 구하라. 자유도는?

(d) Pearson의 $X^2 = \sum \frac{(O-E)^2}{E}$도 구하고 결론을 내려라.

풀이 (a) $\ell'(\theta) = \frac{2n_1+n_2}{\theta} - \frac{n_2+2n_3}{1-\theta} = 0$에서 $\hat\theta = \frac{2n_1+n_2}{2n} = \frac{130}{200} = 0.65$이다. (b) 한 사람의 로그가능도는 A의 수 $Y\sim\mathrm{Bin}(2,\theta)$에 대해 $Y\log\theta + (2-Y)\log(1-\theta)$이므로 $I(\theta) = \frac{2}{\theta(1-\theta)}$이다. (c) 기대도수는 $100\cdot 0.4225 = 42.25$, $100\cdot 0.455 = 45.5$, $100\cdot 0.1225 = 12.25$이다. $G = 2\big[50\log\frac{50}{42.25} + 30\log\frac{30}{45.5} + 20\log\frac{20}{12.25}\big] \approx 11.46$이다. 일반 다항 모형의 자유로운 모수는 2개, Hardy–Weinberg는 1개이므로 자유도는 1이다(노트 5.1절의 상자). (d) $X^2 = \frac{7.75^2}{42.25} + \frac{15.5^2}{45.5} + \frac{7.75^2}{12.25} \approx 11.60$이다. 둘 다 3.84를 크게 넘으므로($p\approx 0.0007$) Hardy–Weinberg 평형을 기각한다. 이형접합 Aa가 기대보다 적다.

최종답 (a) 0.65 (b) $\frac{2}{\theta(1-\theta)}$ (c) $G\approx 11.46$ (d) $X^2\approx 11.60$

D4 변동계수가 1인 정규 ★★★

정리 1.6, 2.2, 3.1 · 제한 35분

$X_i \overset{\text{iid}}{\sim} \mathcal{N}(\theta,\theta^2)$, $\theta>0$이다. $\bar x = \frac1n\sum x_i$, $m_2 = \frac1n\sum x_i^2$으로 쓴다.

(a) 가능도 방정식이 $\theta^2+\bar x\theta - m_2 = 0$임을 보이고 MLE를 $\bar x$와 m_2의 식으로 구하라. (`check()`에는 `xbar`, `m2`로 쓴다.)

(b) 한 관측의 Fisher 정보 $I(\theta)$를 구하라.

(c) $\bar X$도 θ의 불편추정량이다. $\bar X$의 점근 효율 $\frac{1/(nI)}{\mathrm{Var}\,\bar X}$를 구하라.

(d) $\bar x = 2$, $m_2 = 6$이면 $\hat\theta$는 얼마인가?

풀이 (a) $\ell'(\theta) = -\frac{n}{\theta} + \frac{\sum(x_i-\theta)^2}{\theta^3} + \frac{\sum(x_i-\theta)}{\theta^2}$ 이다. θ^3/n을 곱하면 $-\theta^2 + (m_2 - 2\theta\bar{x} + \theta^2) + \theta(\bar{x}-\theta) = m_2 - \theta\bar{x} - \theta^2 = 0$이다. 양의 근은 $\hat{\theta} = \frac{-\bar{x}+\sqrt{\bar{x}^2+4m_2}}{2}$ 이다. 두 근의 곱이 $-m_2 < 0$ 이므로 양의 근은 하나다. (b) $\log f = -\log\theta - \frac{(x-\theta)^2}{2\theta^2}$ + 상수에서 $\frac{\partial^2}{\partial\theta^2}\log f = \frac{1}{\theta^2} - \frac{3}{\theta^4}(x-\theta)^2 - \frac{4}{\theta^3}(x-\theta) - \frac{1}{\theta^2}$ 이다. 기댓값은 $\frac{1}{\theta^2} - \frac{3}{\theta^2} - 0 - \frac{1}{\theta^2} = -\frac{3}{\theta^2}$ 이므로 $I(\theta) = \frac{3}{\theta^2}$ 이다. (c) $\mathrm{Var}\,\bar{X} = \frac{\theta^2}{n}$ 이고 한계는 $\frac{\theta^2}{3n}$ 이므로 효율은 $\frac{1}{3}$ 이다. 표본평균은 분산에 들어 있는 정보를 버린다. (d) $\hat{\theta} = \frac{-2+\sqrt{4+24}}{2} = \sqrt{7}-1 \approx 1.646$ 이다.

최종답 (a) $\frac{-\bar{x}+\sqrt{\bar{x}^2+4m_2}}{2}$ (b) $3/\theta^2$ (c) $\frac{1}{3}$ (d) $\sqrt{7}-1 \approx 1.646$