수리통계학 × 통계학의 역사

11회 워크북 — 풀이판

충분성
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	조건부분포, UMVUE, 로그분배함수, 정보를 손으로 계산한다	14	2시간 4분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 40분
C 증명 문제	변형, 일반화, 반례 만들기	9	2시간 50분
D 퀄형	여러 절을 잇는 종합 문제. 제한 시간 안에	4	2시간 20분
	합계	34	약 10시간

노트 `11_sufficiency_notes.pdf` — 문항의 "정리 2.1" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T1] 원소 추적, [T2] 귀류·반례, [T4] 부등식, [T6] 극한의 교환, [T7] 변수변환, [T8] 델타 방법

검산 노트북 `11_sufficiency.ipynb`의 "워크북 검산" 절: `check("A3a", "2/15")`

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. 식에 λ가 들어가면 `lam`으로 쓴다(`lambda`는 파이썬의 예약어다).
5. **증명은 근거를 단다.** 등호와 부등호마다 어느 정의, 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

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

A1 조건부분포에서 모수가 사라진다 ★

정의 1.1, 예제 1.2 · 8분

(a) $X_1,\dots,X_5 \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, $T=\sum X_i$이다. $P(\mathbf{X}=(1,1,0,1,0)\mid T=3)$을 구하고 p가 사라짐을 확인하라. (b) $X_1,X_2,X_3 \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$, $T=\sum X_i$이다. $P(\mathbf{X}=(2,1,1)\mid T=4)$를 구하라.

풀이 (a) $\{\mathbf{X}=(1,1,0,1,0)\}\subset\{T=3\}$이므로 조건부확률은 $\dfrac{p^3(1-p)^2}{\binom{5}{3}p^3(1-p)^2}=\dfrac{1}{10}$이다. 성공이 세 번인 열 10개가 모두 같은 확률이다.

(b) 4회 정리 2.4로 $T\sim\text{Poi}(3\lambda)$이므로

$$\frac{\frac{e^{-\lambda}\lambda^2}{2!}\cdot e^{-\lambda}\lambda\cdot e^{-\lambda}\lambda}{\frac{e^{-3\lambda}(3\lambda)^4}{4!}}=\frac{4!}{2!\,1!\,1!}\left(\frac{1}{3}\right)^4=\frac{12}{81}=\frac{4}{27}.$$

다항분포 $\text{M}(4;\tfrac13,\tfrac13,\tfrac13)$의 확률이다(노트 연습 1.1).

최종답 (a) $\dfrac{1}{10}$ (b) $\dfrac{4}{27}$

A2 Poisson 표본의 UMVUE 값 ★

정리 2.1, 3.3 · 8분

$X_1,\dots,X_4 \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$에서 $T=\sum X_i=6$을 관측했다. (a) $e^{-\lambda}=P(X=0)$의 UMVUE의 값을 구하라. (b) $P(X=2)=\tfrac12\lambda^2e^{-\lambda}$의 UMVUE를 $\mathbb{1}\{X_1=2\}$에 T로 조건을 걸어 구하고, 그 값을 계산하라.

풀이 T는 완비충분통계량이고(정리 4.5), $T=t$가 주어지면 $X_1\sim\text{b}(t,\tfrac14)$이다. Lehmann–Scheffé(정리 3.3)로 불편추정량에 T로 조건을 건 것이 UMVUE다.

(a) $P(X_1=0\mid T)=(\tfrac34)^T$이고 $T=6$이면 $\dfrac{729}{4096}\approx0.178$이다.

(b) $E\big[\mathbb{1}\{X_1=2\}\mid T\big]=\binom{T}{2}\left(\tfrac14\right)^2\left(\tfrac34\right)^{T-2}$이다. $T=6$이면 $15\cdot\dfrac{1}{16}\cdot\dfrac{81}{256}=\dfrac{1215}{4096}\approx0.297$이다.

최종답 (a) $(\tfrac34)^6=\dfrac{729}{4096}\approx0.178$ (b) $\dfrac{1215}{4096}\approx0.297$

A3 Bernoulli의 UMVUE ★

예제 2.4, 정리 3.3 · 6분

$X_1, \dots, X_{10} \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$에서 $T = \sum X_i = 4$를 관측했다. (a) p^2의 UMVUE의 값, (b) $p(1-p)$의 UMVUE의 값을 구하라. 각각을 MLE $\hat{p}^2 = 0.16$, $\hat{p}(1-\hat{p}) = 0.24$와 비교하라.

풀이 T는 완비충분이다(예제 3.2(a)). (a) $E(X_1X_2 \mid T=t) = \frac{t(t-1)}{n(n-1)} = \frac{12}{90} = \frac{2}{15} \approx 0.133$. (b) $E(X_1(1-X_2) \mid T=t) = \frac{t(n-t)}{n(n-1)} = \frac{24}{90} = \frac{4}{15} \approx 0.267$. UMVUE는 p^2에서 MLE보다 작고, $p(1-p)$에서 MLE보다 크다. 둘 다 MLE의 치우침($E\hat{p}^2 = p^2 + \frac{p(1-p)}{n}$)을 되돌리는 방향이다.

최종답 (a) $\frac{2}{15}$ (b) $\frac{4}{15}$

A4 균등분포의 UMVUE ★★

예제 3.2(c), 정리 3.3 · 8분

$X_1, \dots, X_4 \overset{\text{iid}}{\sim} \text{U}(0, \theta)$에서 1.2, 3.1, 0.7, 2.5를 얻었다. (a) θ의 UMVUE의 값을 구하라. (b) $1/\theta$의 UMVUE를 $c/X_{(n)}$ 꼴에서 찾고, 그 값을 구하라.

풀이 $X_{(n)}$은 완비충분이다(예제 3.2(c)). (a) $EX_{(n)} = \frac{n}{n+1}\theta$이므로 $\frac{n+1}{n}X_{(n)} = \frac{5}{4} \cdot 3.1 = 3.875$이다.

(b) $EX_{(n)}^{-1} = \int_0^\theta t^{-1}\frac{nt^{n-1}}{\theta^n}dt = \frac{n}{n-1}\cdot\frac{1}{\theta}$ ($n \geq 2$)이므로 $\frac{n-1}{n}X_{(n)}^{-1}$이 불편이고, 완비충분통계량의 함수이므로 UMVUE다. 값은 $\frac{3}{4}\cdot\frac{1}{3.1} = \frac{15}{62} \approx 0.242$이다. (a)의 역수 $1/3.875 = 0.258$과 다르다. UMVUE는 일대일 변환에 대해 불변이 아니다.

최종답 (a) 3.875 (b) $\frac{15}{62} \approx 0.242$

A5 독일 전차 ★

예제 2.3, 3.4(d) · 8분

$\{1, \dots, N\}$에서 비복원으로 뽑은 일련번호 19, 40, 42, 60을 얻었다. (a) N의 UMVUE의 값과 (b) 적률법 추정값 $2\bar{X} - 1$을 구하라. (a)가 (b)를 Rao–Blackwell화한 것임을 한 줄로 말하라.

풀이 (a) $k=4$, $M=60$이므로 $\frac{5}{4}\cdot 60 - 1 = 74$. (b) $\bar{X} = 40.25$이므로 $2\bar{X} - 1 = 79.5$. 예제 2.3에서 $E(2\bar{X} - 1 \mid M=m) = \frac{k+1}{k}m - 1$이었으므로 (a)는 (b)에 충분통계량 M으로 조건을 건 것이고, M이 완비이므로 UMVUE다(연습 3.3).

최종답 (a) 74 (b) 79.5

A6 기하분포의 로그분배함수 ★★

정의 4.1, 정리 4.3 · 12분

$P(X=x)=p(1-p)^{x-1}$ $(x=1,2,\dots)$를 자연모수 $\eta=\log(1-p)$, $T(x)=x$, $h(x)=1$인 지수족으로 쓴다. (a) $A(\eta)$를 구하라. (b) $A'(\eta)$로 $p=\frac14$에서 EX를, (c) $A''(\eta)$로 $\operatorname{Var}X$를 구하라.

풀이 (a) $\sum_x e^{\eta x-A(\eta)}=1$이므로 $e^{A(\eta)}=\sum_{x\ge1}e^{\eta x}=\frac{e^\eta}{1-e^\eta}$ $(\eta<0)$, 곧 $A(\eta)=\eta-\log(1-e^\eta)$이다. 확인: $e^{\eta x-A(\eta)}=(1-p)^x\cdot\frac{p}{1-p}=p(1-p)^{x-1}$.

(b) $A'(\eta)=1+\frac{e^\eta}{1-e^\eta}=\frac{1}{1-e^\eta}=\frac1p$이므로 $p=\frac14$에서 $EX=4$.

(c) $A''(\eta)=\frac{e^\eta}{(1-e^\eta)^2}=\frac{1-p}{p^2}$이므로 $\operatorname{Var}X=\frac{3/4}{1/16}=12$.

최종답 (a) $A(\eta)=\eta-\log(1-e^\eta)$ (b) 4 (c) 12

A7 두 모수화의 정보 ★

따름정리 4.4 · 8분

Bernoulli(p), $p=0.2$에서 (a) 자연모수 $\eta=\log\frac{p}{1-p}$에 대한 한 관측의 정보 $I(\eta)$와 (b) p에 대한 정보 $I(p)$를 구하라. 두 값의 관계를 $\frac{d\eta}{dp}$로 설명하라.

풀이 (a) 따름정리 4.4로 $I(\eta)=A''(\eta)=p(1-p)=0.16$. (b) $I(p)=\frac{1}{p(1-p)}=6.25$. $\frac{d\eta}{dp}=\frac{1}{p(1-p)}$이므로 $I(\eta)\big(\frac{d\eta}{dp}\big)^2=p(1-p)\cdot\frac{1}{p^2(1-p)^2}=I(p)$이다(9회 연습 2.2의 변환 규칙). 정보는 모수의 단위에 따라 달라진다.

최종답 (a) 0.16 (b) 6.25

A8 지수 표본의 최댓값의 몫 ★★

정리 5.2 · 10분

$X_1,\dots,X_4\overset{\text{iid}}{\sim}\text{Exp}(\text{평균 }\theta)$이다. $EX_{(n)}=\theta\big(1+\frac12+\cdots+\frac1n\big)$을 써서 $E\big[X_{(4)}/\sum_i X_i\big]$를 구하라.

풀이 $S=\sum X_i$는 1-모수 지수족의 완비충분통계량이고(정리 4.5), $R=X_{(4)}/S$는 $X_i\mapsto cX_i$에 대해 변하지 않으므로 분포가 θ에 무관한 보조통계량이다. Basu(정리 5.2)로 $R\perp S$이므로 $EX_{(4)}=E[RS]=ER\cdot ES$이고

$$ER=\frac{\theta(1+\frac12+\frac13+\frac14)}{4\theta}=\frac{25/12}{4}=\frac{25}{48}\approx0.521.$$

최종답 $\frac{25}{48}\approx0.521$

A9 균등 표본의 순서통계량의 몫 ★

예제 5.5(b) · 8분

$X_1, \dots, X_5 \overset{\text{iid}}{\sim} \mathrm{U}(0,\theta)$이다. $EX_{(k)} = \frac{k}{n+1}\theta$를 써서 $E\big[X_{(2)}/X_{(5)}\big]$를 구하라.

풀이 $X_{(5)}$는 완비충분이고 $X_{(2)}/X_{(5)}$는 척도에 불변이므로 보조다. Basu로 독립이므로 $EX_{(2)} = E\frac{X_{(2)}}{X_{(5)}}\,EX_{(5)}$이고 $E\frac{X_{(2)}}{X_{(5)}} = \frac{2\theta/6}{5\theta/6} = \frac{2}{5}$이다.

최종답 $\frac{2}{5}$

A10 UMVUE의 분산과 Cramér–Rao 한계 ★★

정리 2.1, 9회 정리 2.4 · 10분

$X_1, \dots, X_{10} \overset{\text{iid}}{\sim} \mathrm{Poi}(\lambda)$, $\lambda = 1$에서 $e^{-\lambda}$의 UMVUE $(1-\frac{1}{n})^T$를 생각한다. 확률생성함수 $Es^T = e^{n\lambda(s-1)}$을 써서 (a) $\mathrm{Var}\big((1-\frac{1}{n})^T\big)$, (b) Cramér–Rao 한계 $\frac{\tau'(\lambda)^2}{nI(\lambda)}$, (c) 둘의 비를 구하라.

풀이 (a) $E(1-\frac{1}{n})^T = e^{-\lambda}$, $E(1-\frac{1}{n})^{2T} = e^{n\lambda((1-1/n)^2-1)} = e^{-2\lambda+\lambda/n}$이므로 분산은 $e^{-2\lambda}(e^{\lambda/n}-1) = e^{-2}(e^{0.1}-1) \approx 0.01423$이다. (b) $\tau' = -e^{-\lambda}$, $I = 1/\lambda$이므로 $\frac{\lambda e^{-2\lambda}}{n} = \frac{e^{-2}}{10} \approx 0.01353$이다. (c) 비는 $\frac{e^{\lambda/n}-1}{\lambda/n} = 10(e^{0.1}-1) \approx 1.0517$이다. UMVUE는 한계에 5% 못 미친다. 정리 4.6대로 $e^{-\lambda}$는 ET의 일차함수가 아니어서 한계에 닿을 수 없다.

최종답 (a) $e^{-2}(e^{0.1}-1) \approx 0.01423$ (b) $\frac{e^{-2}}{10} \approx 0.01353$ (c) $10(e^{0.1}-1) \approx 1.0517$

A11 모수가 두 개뿐이면 완비가 아니다 ★★

정의 3.1 · [T2] · 10분

$X_1, X_2 \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(p)$이고 모수공간이 $p \in \{\frac{1}{4}, \frac{3}{4}\}$ 두 점뿐이다. $T = X_1 + X_2$의 함수 g 가운데 두 p에서 모두 $E_p g(T) = 0$인 것을 $g(0) = 1$로 정규화해 찾아라. (a) $g(1)$과 (b) $g(2)$를 구하라.

풀이 $p = \frac{1}{4}$: $9g(0) + 6g(1) + g(2) = 0$. $p = \frac{3}{4}$: $g(0) + 6g(1) + 9g(2) = 0$ (양변에 16을 곱했다). 빼면 $8g(0) - 8g(2) = 0$이므로 $g(2) = g(0) = 1$이고, 첫 식에서 $10 + 6g(1) = 0$, 곧 $g(1) = -\frac{5}{3}$이다. 0이 아닌 $g(T)$의 평균이 모수공간 전체에서 0이므로 이 모수공간에서 T는 완비가 아니다. 모수공간이 $(0,1)$이면 예제 3.2(a)로 완비다.

최종답 (a) $g(1) = -\frac{5}{3}$ (b) $g(2) = 1$

A12 Rayleigh 분포의 MLE ★★

정리 4.3, 따름정리 4.4 · 8분

Rayleigh 밀도 $f(x;\theta) = \frac{x}{\theta}e^{-x^2/(2\theta)}$ $(x>0)$를 $T(x) = x^2$인 지수족으로 쓰고, $A'(\eta) = EX^2$를 구하라. 자료 1, 2, 2, 3에서 가능도 방정식 "충분통계량의 평균 = 모형의 평균"으로 $\hat\theta$를 구하라.

풀이 $\eta = -\frac{1}{2\theta}$이면 $f(x) = x\exp\big(\eta x^2 - A(\eta)\big)$, $A(\eta) = -\log(-2\eta)$이다. $A'(\eta) = -\frac{1}{\eta} = 2\theta$이므로 $EX^2 = 2\theta$다. 따름정리 4.4(c)로 $\frac{1}{n}\sum x_i^2 = 2\hat\theta$이고, $\sum x_i^2 = 1+4+4+9 = 18$이므로 $\hat\theta = \frac{18}{8} = \frac{9}{4}$이다.

최종답 $\hat\theta = \frac{9}{4}$

A13 오염 정규에서 평균편차의 효율 ★

명제 7.1 · 10분

$X \sim (1-\varepsilon)\mathcal{N}(0,1) + \varepsilon\mathcal{N}(0,9)$, $\varepsilon = 0.02$에서 μ_2, μ_4, ν_1을 구하고, 명제 7.1의 식으로 평균편차의 상대효율을 소수 셋째 자리까지 구하라.

풀이 $\mu_2 = 1.16$, $\mu_4 = 3\cdot 2.6 = 7.8$, $\nu_1 = \sqrt{2/\pi}\cdot 1.04 = 0.8298$. $\log s$의 점근분산은 $\frac{1}{4}(7.8/1.3456 - 1) = 1.1992$, $\log d$의 점근분산은 $1.16/0.68858 - 1 = 0.6846$이므로 효율은 $1.1992/0.6846 \approx 1.751$이다. 50개에 하나꼴의 큰 오차로 평균편차가 표준편차보다 75% 효율적이 된다.

최종답 ≈ 1.751

A14 "0인가 아닌가"가 담는 정보 ★★

명제 1.7 · 10분

$X \sim \mathrm{Poi}(\lambda)$ 하나에서 $T = 1\{X = 0\}$만 기록한다. (a) T의 정보 $I_T(\lambda)$를 구하라(답에는 λ를 `lam`으로 쓴다). (b) $\lambda = 2$에서 I_T/I_X를 구하라.

풀이 (a) $q = e^{-\lambda}$, $q' = -e^{-\lambda}$이므로 $I_T = \dfrac{e^{-2\lambda}}{e^{-\lambda}(1-e^{-\lambda})} = \dfrac{1}{e^{\lambda}-1}$이다. (b) $I_X = 1/\lambda$이므로 비는 $\frac{\lambda}{e^{\lambda}-1}$이고 $\lambda = 2$에서 $\frac{2}{e^2-1} \approx 0.313$이다. $\lambda \to 0$이면 비가 1로 간다. 사건이 드물면 "0인가 아닌가"가 거의 모든 정보를 담는다. 명제 1.7대로 비는 1보다 작다.

최종답 (a) $I_T(\lambda) = \frac{1}{e^{\lambda}-1}$ (b) $\frac{2}{e^2-1} \approx 0.313$

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, 앞 회차의 정리, 도구 번호를 쓴다.

B1 인수분해 정리, 근거 채우기 ★★

① · 정리 1.3 · 20분

이산 표본에서 T가 충분통계량일 필요충분조건은 $f(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)h(\mathbf{x})$인 것이다. 각 단계의 근거를 써라.

(⇒)

① $t = T(\mathbf{x})$이면 $\{\mathbf{X} = \mathbf{x}\} \subset \{T = t\}$이다. — T는 $\mathbf{X}$의 함수

② $P_\theta(T = t) > 0$이면 $f(\mathbf{x};\theta) = P_\theta(T = t)\,P_\theta(\mathbf{X} = \mathbf{x} \mid T = t)$이다. — ①, 1회 정리 5.3

③ 둘째 인수 $h(\mathbf{x}) = P(\mathbf{X} = \mathbf{x} \mid T = T(\mathbf{x}))$는 θ에 의존하지 않는다. — 충분성, 정의 1.1

④ $g(t;\theta) = P_\theta(T = t)$로 두면 $f(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)\,h(\mathbf{x})$이다. — ②③

(⇐)

① $T(\mathbf{x}) = t$, $P_\theta(T = t) > 0$이면 $P_\theta(T = t) = g(t;\theta)\sum_{\mathbf{y}:\,T(\mathbf{y})=t} h(\mathbf{y})$이다. — 가산가법성, 인수분해 가정

② $P_\theta(\mathbf{X} = \mathbf{x} \mid T = t) = \dfrac{g(t;\theta)h(\mathbf{x})}{g(t;\theta)\sum_{\mathbf{y}} h(\mathbf{y})} = \dfrac{h(\mathbf{x})}{\sum_{\mathbf{y}:\,T(\mathbf{y})=t} h(\mathbf{y})}$이다. — 1회 정의 5.1, ①

③ 오른쪽에 θ가 없으므로 T는 충분하다. — 정의 1.1

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 1.3의 단계형 증명과 같다. (⇐)의 둘째 단계에서 $g(t;\theta)$를 약분하려면 $g(t;\theta) > 0$이어야 하고, $P_\theta(T = t) > 0$과 첫째 단계가 그것을 준다. (⇒)에서 $P_\theta(T = t) = 0$이면 $f(\mathbf{x};\theta) = 0$이므로 $g(t;\theta) = 0$으로 두면 된다.

B2 Rao–Blackwell, 근거 채우기 ★★

① · 정리 2.1 · 15분

W가 $\tau(\theta)$의 불편추정량이고 $\mathrm{Var}_\theta\, W < \infty$, T가 충분통계량이면 $\varphi(T) = E(W \mid T)$는 통계량이고 불편이며 $\mathrm{Var}\,\varphi(T) \le \mathrm{Var}\, W$이다. 근거를 채워라.

① $T = t$가 주어질 때 $\mathbf{X}$의 조건부분포는 θ에 의존하지 않는다. — 정의 1.1

② $\varphi(t) = E(w(\mathbf{X}) \mid T = t)$는 θ 없이 계산된다. — ①, 조건부분포의 평균

③ $E_\theta \varphi(T) = E_\theta W = \tau(\theta)$이다. — 3회 정리 2.5

④ $\mathrm{Var}_\theta\, W = \mathrm{Var}_\theta\, \varphi(T) + E_\theta\, \mathrm{Var}(W \mid T)$이다. — 3회 정리 2.6, $\mathrm{Var}\, W < \infty$

⑤ $\mathrm{Var}_\theta\, \varphi(T) \le \mathrm{Var}_\theta\, W$이다. — $\mathrm{Var}(W \mid T) \ge 0$

⑥ 등호이면 $E_\theta(W - \varphi(T))^2 = 0$, 곧 $W = \varphi(T)$ (확률 1)이다. — 2회 따름정리 5.6

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 2.1의 단계형 증명과 같다. 충분성은 첫째 단계에서만 쓰인다. 충분하지 않은 T로 조건을 걸어도 셋째–다섯째 단계는 그대로 성립하지만, 둘째 단계가 깨져 φ가 θ를 포함할 수 있다.

B3 Basu의 정리, 근거 채우기 ★★

① · 정리 5.2 · 15분

T가 완비충분통계량이고 V가 보조통계량이면 $T \perp V$이다. 근거를 채워라.

① $p_B = P_\theta(V \in B)$는 θ에 의존하지 않는다. — V는 보조, 정의 5.1

② $\eta(T) = P(V \in B \mid T)$는 θ를 포함하지 않는 T의 함수다. — T는 충분, 정의 1.1

③ 모든 θ에서 $E_\theta[\eta(T) - p_B] = 0$이다. — 3회 정리 2.5, ①

④ 모든 θ에서 $\eta(T) = p_B$ (확률 1)이다. — T의 완비성, 정의 3.1

⑤ $P_\theta(T \in C, V \in B) = E_\theta[1\{T \in C\}\eta(T)] = p_B P_\theta(T \in C)$이다. — 반복기댓값, ④

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 5.2의 단계형 증명과 같다. 넷째 단계에서 완비성을 쓰는 함수는 $g(T) = \eta(T) - p_B$이고, 이것이 T만의 함수라는 것(둘째 단계, 충분성)과 그 평균이 θ에 무관한 0이라는 것(첫째와 셋째 단계, 보조성)이 함께 필요하다.

B4 인수분해 정리, 단계의 이름만 ★★

② · 정리 1.3 · 25분

이산 표본에서 정리 1.3의 두 방향을 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **(⇒) $\{\mathbf{X} = \mathbf{x}\}$와 $\{T = t\}$의 관계**

② **결합확률을 곱셈법칙으로 쪼갠다**

③ **충분성으로 h를 정하고 g를 정한다**

④ **(⇐) $P_\theta(T=t)$를 g와 h로 쓴다**

⑤ **조건부확률을 계산하고 g를 약분한다**

⑥ **$P_\theta(T=t)=0$인 경우를 처리한다**

풀이

① $t=T(\mathbf{x})$이면 $\{\mathbf{X}=\mathbf{x}\}\subset\{T=t\}$. — T는 $\mathbf{X}$의 함수

② $f(\mathbf{x};\theta)=P_\theta(T=t)\,P_\theta(\mathbf{X}=\mathbf{x}\mid T=t)$. — 1회 정리 5.3

③ $h(\mathbf{x})=P(\mathbf{X}=\mathbf{x}\mid T=T(\mathbf{x}))$는 θ에 무관하고, $g(t;\theta)=P_\theta(T=t)$. — 정의 1.1

④ $P_\theta(T=t)=\sum_{T(\mathbf{y})=t} g(t;\theta)h(\mathbf{y})=g(t;\theta)\sum_{T(\mathbf{y})=t}h(\mathbf{y})$. — 가산가법성

⑤ $P_\theta(\mathbf{X}=\mathbf{x}\mid T=t)=h(\mathbf{x})/\sum_{T(\mathbf{y})=t}h(\mathbf{y})$, θ가 없다. — 1회 정의 5.1

⑥ $P_\theta(T=t)=0$이면 ($\Rightarrow$)에서 $f(\mathbf{x};\theta)=0$이므로 $g(t;\theta)=0$으로 둔다. ($\Leftarrow$)에서는 조건부확률을 정의하지 않는다. — 1회 정의 5.1

B5 Rao–Blackwell, 단계의 이름만 ★★

② · 정리 2.1 · 25분

정리 2.1의 (a)–(c)를 증명하라.

① **(a) 충분성으로 φ가 통계량임을 보인다**

② **(b) 반복기댓값으로 불편성을 보인다**

③ **(c) 전분산 공식을 쓴다**

④ **조건부분산이 음이 아님을 쓴다**

⑤ **등호 조건을 분석한다**

풀이

① $\mathbf{X}\mid T=t$의 분포가 θ에 무관하므로 $\varphi(t)=E(W\mid T=t)$도 θ에 무관하다. — 정의 1.1

② $E_\theta\varphi(T)=E_\theta W=\tau(\theta)$. — 3회 정리 2.5

③ $\operatorname{Var}W=\operatorname{Var}E(W\mid T)+E\operatorname{Var}(W\mid T)$. — 3회 정리 2.6

④ $E\operatorname{Var}(W\mid T)\ge 0$이므로 $\operatorname{Var}\varphi(T)\le\operatorname{Var}W$. — 조건부분산 ≥ 0

⑤ $\operatorname{Var}(W\mid T)=E[(W-\varphi(T))^2\mid T]$이므로 등호이면 $E(W-\varphi(T))^2=0$이고 $W=\varphi(T)$ (확률 1). — 반복기댓값, 2회 따름정리 5.6

B6 인수분해 정리, 전략만 ★★

③ · 정리 1.3 · 30분

이산 표본에서 정리 1.3을 증명하라.

전략 (⇒)는 결합확률을 "T의 확률 × T가 주어진 조건부확률"로 쪼갠다. (⇐)는 조건부확률을 계산해 g를 약분한다.

풀이 B4의 풀이와 같다. 이 증명은 이산 표본에서만 성립한다. 연속 표본에서는 $\{T = t\}$의 확률이 0이어서 조건부분포를 정의하는 데 측도론이 필요하다(노트 1.1절의 상자). C6에서는 이 정리를 써서 충분통계량의 일대일 함수도 충분함을 보인다.

B7 Basu의 정리, 전략만 ★★

③ · 정리 5.2 · 30분

정리 5.2를 증명하라.

전략 사건 $\{V \in B\}$의 조건부확률 $\eta(T)$를 만든다. 그 평균이 모든 θ에서 같은 상수이므로 완비성으로 $\eta(T)$ 자체가 그 상수다.

풀이 B3의 풀이와 같다. 보조성은 p_B가 θ에 무관하다는 데, 충분성은 $\eta(T)$가 통계량이라는 데, 완비성은 $\eta(T) = p_B$를 끌어내는 데 쓰인다. 완비성이 빠지면 노트 5절의 가정 주의처럼 $\mathrm{U}(\theta, \theta+1)$에서 최소충분통계량과 범위가 독립이 아니다. 충분성이 빠지면 $\eta(T)$가 θ를 포함해 완비성을 쓸 대상이 없다.

C

증명 문제

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

C1 충분하지 않은 통계량에 조건을 걸면 ★

[T2] · 정리 2.1 · 15분

$X_1, X_2 \overset{\text{iid}}{\sim} \mathcal{N}(\theta, 1)$이고 $W = \bar{X}$이다.

(a) $T = X_1$이 θ의 충분통계량이 아님을 X_2의 조건부분포로 보여라.

(b) $E(W \mid X_1)$을 구하고, 그것이 추정량이 아닌 이유를 말하라.

(c) $\operatorname{Var} E(W \mid X_1) < \operatorname{Var} W$임을 확인하라. 정리 2.1의 어느 부분이 성립하고 어느 부분이 깨지는가?

풀이 (a) 독립이므로 $X_1 = x_1$이 주어질 때 (X_1, X_2)의 조건부분포는 $X_1 = x_1$에 점질량, $X_2 \sim \mathcal{N}(\theta, 1)$이다. 이것은 θ에 의존하므로 정의 1.1로 X_1은 충분하지 않다.
(b) $E(\bar{X} \mid X_1) = \frac{1}{2}\big(X_1 + E(X_2 \mid X_1)\big) = \frac{X_1 + \theta}{2}$이다. 계산하려면 θ를 알아야 하므로 통계량이 아니다.
(c) $\operatorname{Var}\frac{X_1+\theta}{2} = \frac{1}{4} < \frac{1}{2} = \operatorname{Var}\bar{X}$이다. 정리 2.1의 (b)(평균 θ)와 (c)(분산 감소)는 반복기댓값과 전분산 공식만 쓰므로 성립한다. 깨지는 것은 (a)뿐이고, 그것이 충분성을 쓰는 유일한 곳이다.

C2 굽은 정규족 ★★

[T2] · 정리 1.6, 4.5 · 20분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \theta^2)$, $\theta > 0$, $n \geq 2$이다.

(a) $\big(\sum X_i,\ \sum X_i^2\big)$가 최소충분통계량임을 정리 1.6으로 보여라. 모수는 하나인데 통계량은 2차원이다.

(b) $E\bar{X}^2$과 ES^2을 구하고, 모든 θ에서 평균이 0이지만 0이 아닌 함수 $g(\bar{X}, S^2)$을 찾아 이 통계량이 완비가 아님을 보여라.

(c) 정리 4.5가 여기에 쓰이지 않는 이유를 자연모수의 모양으로 설명하라.

풀이 (a) $\log\frac{f(\mathbf{x};\theta)}{f(\mathbf{y};\theta)} = -\frac{1}{2\theta^2}\big(\sum x_i^2 - \sum y_i^2\big) + \frac{1}{\theta}\big(\sum x_i - \sum y_i\big)$이다. 이것이 모든 $\theta > 0$에서 상수이려면 $u = 1/\theta$의 다항식 $-\frac{a}{2}u^2 + bu$가 상수여야 하므로 $a = b = 0$, 곧 두 합이 같아야 한다. 거꾸로 두 합이 같으면 비는 1이다. 정리 1.6으로 최소충분이다.

(b) $E\bar X^2 = \operatorname{Var}\bar X + (E\bar X)^2 = \frac{\theta^2}{n} + \theta^2 = \frac{n+1}{n}\theta^2$ 이고 $ES^2 = \theta^2$ 이다(6회 정리 1.1). 따라서 $g = \frac{n}{n+1}\bar X^2 - S^2$ 은 모든 θ에서 평균이 0이다. g는 연속분포를 갖는 확률변수이므로 확률 1로 0일 수 없다. 완비가 아니다.

(c) 자연모수 $(\eta_1, \eta_2) = (\frac{1}{\theta}, -\frac{1}{2\theta^2})$는 $\eta_2 = -\frac{1}{2}\eta_1^2$ 인 곡선 위만 움직인다. $\mathbb{R}^2$의 열린 집합을 포함하지 않으므로 정리 4.5의 가정이 성립하지 않는다. 이런 족을 "굽은 지수족"이라 한다.

C3 이동한 지수분포의 UMVUE ★★

[T6] · 정리 1.3, 3.3 · 20분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} f(x;\theta) = e^{-(x-\theta)}\ (x > \theta)$이다.

(a) $X_{(1)}$이 충분통계량임을 인수분해 정리로 보여라.

(b) $P(X_{(1)} - \theta > u) = e^{-nu}\ (u > 0)$임을 보여라.

(c) $X_{(1)}$이 완비임을 보여라.

(d) θ의 UMVUE를 구하고, $n = 5$, $X_{(1)} = 2.3$일 때 그 값을 구하라.

풀이 (a) $\prod f(x_i;\theta) = e^{n\theta}\mathbb{1}\{x_{(1)} > \theta\} \cdot e^{-\sum x_i}$ 이므로 정리 1.3으로 $X_{(1)}$이 충분하다.

(b) 독립으로 $P(X_{(1)} > \theta + u) = \prod_i P(X_i > \theta + u) = (e^{-u})^n$ 이다. 따라서 $X_{(1)}$의 밀도는 $ne^{-n(t-\theta)}\ (t > \theta)$이다.

(c) 모든 θ에서 $E_\theta g(X_{(1)}) = ne^{n\theta}\int_\theta^\infty g(t)e^{-nt}dt = 0$이면 $G(\theta) = \int_\theta^\infty g(t)e^{-nt}dt = 0$이 모든 θ에서 성립한다. G는 상수 0이므로 미적분의 기본정리로 거의 모든 θ에서 $G'(\theta) = -g(\theta)e^{-n\theta} = 0$, 곧 $g = 0$ (거의 어디서나)이고 $P_\theta(g(X_{(1)}) = 0) = 1$이다. (적분가능성은 $E_\theta|g(X_{(1)})| < \infty$에서 나온다 [T6].)

(d) (b)로 $EX_{(1)} = \theta + \frac{1}{n}$이므로 $X_{(1)} - \frac{1}{n}$은 완비충분통계량의 불편 함수, 곧 UMVUE다(정리 3.3). 값은 $2.3 - 0.2 = 2.1$이다.

최종답 (d) 2.1

C4 치우친 추정량의 Rao–Blackwell화 ★

정리 2.1 · 15분

W가 $\tau(\theta)$의 추정량이고(불편일 필요는 없다) $\operatorname{Var}_\theta W < \infty$, T가 충분통계량이다. $\varphi(T) = E(W \mid T)$이면 모든 θ에서

$$E_\theta\big(\varphi(T) - \tau(\theta)\big)^2 \le E_\theta\big(W - \tau(\theta)\big)^2$$

임을 보여라. 평균제곱오차를 분산과 치우침으로 나누어 설명하라.

풀이 반복기댓값으로 $E_\theta\varphi(T) = E_\theta W$이므로 두 추정량의 치우침 $b(\theta) = E_\theta W - \tau(\theta)$가 같다. 전분산 공식으로 $\operatorname{Var}_\theta\varphi(T) \le \operatorname{Var}_\theta W$이다. 평균제곱오차는 분산 + $b(\theta)^2$이므로 부등식이 나온다. 충분성은 φ가 통계량이라는 데만 쓰인다. 노트 연습 2.4(볼록 손실)는 이것을 볼록

손실로 넓힌 것이다.

C5 정규 표본의 범위 ★★

정리 5.2 · 20분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이고 $R = X_{(n)} - X_{(1)}$은 범위다.

(a) σ를 고정하고 $\bar{X} \perp R$임을 보여라.

(b) 두 모수가 모두 미지일 때 R/S가 $(\bar{X}, S)$와 독립임을 보이고, $E(R/S) = ER/ES$를 얻어라.

(c) R과 S는 독립이 아니다. $\operatorname{Cov}(R, S) = E(R/S)\operatorname{Var} S > 0$을 보여라.

풀이 (a) σ를 고정하면 $\bar{X}$는 μ의 완비충분통계량이다(정리 4.5, 1-모수 정규). $R = \max_i(X_i - \mu) - \min_i(X_i - \mu)$은 $\mathcal{N}(0, \sigma^2)$ 표본만의 함수이므로 분포가 μ에 무관한 보조통계량이다. Basu로 독립이고, 모든 σ에서 그렇다.

(b) $(\bar{X}, S)$는 (μ, σ)의 완비충분통계량이다(정리 4.5). $X_i \mapsto a + bX_i$ $(b > 0)$이면 R과 S가 모두 b배가 되므로 R/S는 변하지 않고, 분포가 (μ, σ)에 무관하다. Basu로 독립이고, $ER = E[(R/S)S] = E(R/S)\,ES$이다.

(c) (b)로 $E[RS] = E[(R/S)S^2] = E(R/S)\,ES^2$이고 $ER\,ES = E(R/S)(ES)^2$이다. 빼면 $\operatorname{Cov}(R, S) = E(R/S)\big(ES^2 - (ES)^2\big) = E(R/S)\operatorname{Var} S > 0$이다. 둘 다 척도 σ를 따라 움직이므로 양의 상관이다. σ를 고정해도 S는 보조가 아니므로 Basu를 쓸 수 없다.

C6 충분통계량의 함수 ★★

[T2] · 정리 1.3 · 15분

(a) T가 충분하고 어떤 통계량 U와 함수 r로 $T = r(U)$이면 U도 충분함을 보여라.

(b) 충분통계량의 일대일 함수는 충분함을 보여라.

(c) 반례: $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$에서 두 모수가 모두 미지이면 $\bar{X}$ 하나는 충분하지 않음을 $\operatorname{Var}(X_1 \mid \bar{X})$로 보여라.

풀이 (a) 정리 1.3으로 $f(\mathbf{x}; \theta) = g(T(\mathbf{x}); \theta)h(\mathbf{x}) = \tilde{g}(U(\mathbf{x}); \theta)h(\mathbf{x})$, $\tilde{g} = g(r(\cdot); \theta)$이다. 다시 정리 1.3으로 U는 충분하다. (b) $U = s(T)$이고 s가 일대일이면 $T = s^{-1}(U)$이므로 (a)에 $r = s^{-1}$을 쓴다. (c) $(X_1, \bar{X})$는 결합정규이고 $\operatorname{Cov}(X_1, \bar{X}) = \sigma^2/n$이므로 $\operatorname{Var}(X_1 \mid \bar{X}) = \sigma^2 - \dfrac{(\sigma^2/n)^2}{\sigma^2/n} = \sigma^2\dfrac{n-1}{n}$이다. $\bar{X}$가 주어진 $\mathbf{X}$의 조건부분포가 σ에 의존하므로 정의 1.1로 충분하지 않다. 충분통계량 $(\bar{X}, S^2)$의 함수 $\bar{X}$는 충분성을 잃는다.

C7 평균모수와 Cramér–Rao 한계 ★★

정리 4.3, 4.6 · 20분

1-모수 지수족 $f(x;\eta)=h(x)\exp\big(\eta T(x)-A(\eta)\big)$에서 $\operatorname{Var}_\eta T>0$이고, 평균모수를 $\mu=A'(\eta)$라 하자.

(a) μ가 η의 순증가함수임을 보여라.

(b) μ에 대한 한 관측의 정보가 $I(\mu)=1/\operatorname{Var}T$임을 보여라.

(c) $\bar T=\frac{1}{n}\sum T(X_i)$가 μ의 불편추정량이고 분산이 Cramér–Rao 한계와 같음을 보여라. 정리 4.6과 어떻게 맞는가?

풀이 (a) 정리 4.3으로 $\frac{d\mu}{d\eta}=A''(\eta)=\operatorname{Var}_\eta T>0$이므로 순증가다. 그래서 μ로 모수를 바꿔 쓸 수 있다. (b) 따름정리 4.4로 $I(\eta)=A''(\eta)$이고, 9회 연습 2.2의 변환 규칙으로 $I(\mu)=I(\eta)\big(\frac{d\eta}{d\mu}\big)^2=A''(\eta)\cdot\frac{1}{A''(\eta)^2}=\frac{1}{\operatorname{Var}T}$이다. (c) $E\bar T=A'(\eta)=\mu$이고 $\operatorname{Var}\bar T=\frac{\operatorname{Var}T}{n}=\frac{1}{nI(\mu)}$이다. $\tau(\mu)=\mu$에 대한 한계 $\frac{1}{nI(\mu)}$와 같다. 정리 4.6은 한계에 닿는 통계량이 지수족의 자연 충분통계량의 일차함수라고 말했고, $\bar T$가 바로 그것이다.

C8 UMVUE는 0의 불편추정량과 상관이 없다 ★★★

[T4] · 정리 3.3 · 20분

W는 $\tau(\theta)$의 불편추정량이고 분산이 유한하다. $\mathcal{U}$를 0의 불편추정량($E_\theta U=0$, 모든 θ) 가운데 분산이 유한한 것의 모임이라 하자.

(a) W가 UMVUE이면 모든 $U\in\mathcal{U}$와 θ에서 $\operatorname{Cov}_\theta(W,U)=0$임을 보여라.

(b) 거꾸로 모든 $U\in\mathcal{U}$와 θ에서 $\operatorname{Cov}_\theta(W,U)=0$이면 W가 UMVUE임을 보여라.

(c) 완비충분통계량 T가 있으면 $\mathcal{U}$에서 T의 함수는 0뿐이다. 이것과 (b)가 Lehmann–Scheffé 정리와 어떻게 이어지는지 말하라.

풀이 (a) $U\in\mathcal{U}$, $a\in\mathbb{R}$이면 $W_a=W+aU$는 불편이고 $\operatorname{Var}_\theta W_a=\operatorname{Var}_\theta W+2a\operatorname{Cov}_\theta(W,U)+a^2\operatorname{Var}_\theta U$이다. 어떤 θ_0에서 $c=\operatorname{Cov}_{\theta_0}(W,U)\neq 0$이라 하자($\operatorname{Var}_{\theta_0}U>0$이다). $a=-c/\operatorname{Var}_{\theta_0}U$로 두면 $\operatorname{Var}_{\theta_0}W_a=\operatorname{Var}_{\theta_0}W-c^2/\operatorname{Var}_{\theta_0}U<\operatorname{Var}_{\theta_0}W$이므로 W가 UMVUE라는 데 모순이다.

(b) W'이 분산이 유한한 불편추정량이면 $U=W'-W\in\mathcal{U}$이고 $\operatorname{Var}W'=\operatorname{Var}W+\operatorname{Var}U+2\operatorname{Cov}(W,U)=\operatorname{Var}W+\operatorname{Var}U\geq\operatorname{Var}W$이다. 분산이 무한인 W'에서는 자명하다.

(c) $W=\varphi(T)$이고 $U\in\mathcal{U}$이면 $\operatorname{Cov}(\varphi(T),U)=E[\varphi(T)U]=E[\varphi(T)E(U\mid T)]$이다. $E(U\mid T)$는 평균이 언제나 0인 T의 함수이므로 완비성으로 0이다. 따라서 (b)로 $\varphi(T)$는 UMVUE이고, 이것이 Lehmann–Scheffé 정리의 다른 증명이다.

C9 정보를 보존하면 충분하다 ★★★

[T6] · 명제 1.7 · 25분

이산 표본이고 9회의 정칙조건이 성립하며, 모수공간 Ω는 구간이다. 모든 θ에서 $I_T(\theta) = I_{\mathbf{X}}(\theta) < \infty$이면 T가 충분함을 보여라. (명제 1.7의 역)

풀이 노트 연습 1.6의 (a)(b)로 $I_{\mathbf{X}} = I_T + E_\theta \operatorname{Var}(s_{\mathbf{X}} \mid T)$이므로 등호이면 $E_\theta \operatorname{Var}(s_{\mathbf{X}} \mid T) = 0$이다. 따라서 확률이 양수인 모든 $\mathbf{x}$에서 $s_{\mathbf{X}}(\mathbf{x};\theta) = E[s_{\mathbf{X}} \mid T = T(\mathbf{x})] = s_T(T(\mathbf{x});\theta)$이다. $t = T(\mathbf{x})$이면

$$\frac{\partial}{\partial\theta}\log P_\theta(\mathbf{X}=\mathbf{x} \mid T=t) = \frac{\partial}{\partial\theta}\big(\log f(\mathbf{x};\theta) - \log P_\theta(T=t)\big) = s_{\mathbf{X}}(\mathbf{x};\theta) - s_T(t;\theta) = 0$$

이 모든 $\theta \in \Omega$에서 성립한다. 구간 위에서 도함수가 0인 함수는 상수이므로 조건부확률은 θ에 의존하지 않고, 정의 1.1로 T는 충분하다. Ω가 구간이 아니면(예: 두 점) 도함수가 0이어도 상수일 필요가 없다 [T2].

D

퀄형

여러 절의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. UMVUE를 구하는 문제는 "완비충분통계량 찾기 → 불편추정량 하나 → 조건부기댓값" 의 순서를 답안에 드러낸다.

D1 지수분포의 생존확률 ★★★

정리 3.3, 5.2, 4회 정리 4.2 · 제한 40분

$X_1, \dots, X_n \overset{\text{iid}}{\sim}$ Exp(평균 θ), $n \geq 2$이고 $c > 0$은 알려진 상수다. $\tau(\theta) = P_\theta(X_1 > c) = e^{-c/\theta}$ 의 UMVUE를 구한다. $S = \sum X_i$이다.

(a) S가 완비충분통계량임을 보여라.

(b) $X_1/S \sim$ 베타$(1, n-1)$이고 X_1/S가 S와 독립임을 보여라.

(c) $P(X_1 > c \mid S = s) = (1 - c/s)^{n-1}$ $(s > c)$, 0 $(s \leq c)$임을 보이고 UMVUE를 적어라.

(d) $n = 5$, $\sum x_i = 10$, $c = 1$이면 UMVUE의 값은 얼마인가? MLE $e^{-c/\bar{x}}$와 비교하라.

풀이 (a) 밀도 $\frac{1}{\theta}e^{-x/\theta}$는 자연모수 $\eta = -1/\theta \in (-\infty, 0)$, $T(x) = x$인 지수족이고 자연모수 공간이 열린 구간이므로 정리 4.5로 S는 완비충분이다.

(b) $X_1 \sim$ 감마$(1, \theta)$와 $X_2 + \cdots + X_n \sim$ 감마$(n-1, \theta)$는 독립이므로 4회 정리 4.2(베타와 감마)로 $X_1/S \sim$ 베타$(1, n-1)$이고 S와 독립이다. Basu로도 같다. X_1/S는 $X_i \mapsto bX_i$에 대해 변하지 않아 보조이고 S는 완비충분이다.

(c) $W = 1\{X_1 > c\}$는 τ의 불편추정량이다. (b)의 독립으로 $P(X_1 > c \mid S = s) = P(X_1/S > c/s \mid S = s) = P(B > c/s)$이고, $B \sim$ 베타$(1, n-1)$의 밀도 $(n-1)(1-u)^{n-2}$를 적분하면 $(1 - c/s)^{n-1}$ $(c < s)$이다. $s \leq c$이면 $X_1 \leq S \leq c$이므로 0이다. 정리 3.3으로 UMVUE는 $\varphi(S) = (1 - c/S)^{n-1} 1\{S > c\}$이다.

(d) $(1 - 0.1)^4 = 0.6561$. MLE는 $e^{-1/2} = 0.6065$이다. 표본이 커지면 $(1 - c/S)^{n-1} \approx e^{-c(n-1)/S} \to e^{-c/\theta}$로 두 추정량이 가까워진다.

최종답 (c) $(1 - c/S)^{n-1}$ $(S > c)$ (d) $0.9^4 = 0.6561$

D2 Poisson에서 "정확히 한 번"의 확률 ★★★

정리 2.1, 3.3, 4.5 · 제한 30분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$이고 $\tau(\lambda) = P(X = 1) = \lambda e^{-\lambda}$이다. $T = \sum X_i$이다.

(a) τ의 불편추정량을 하나 찾아라.

(b) (a)에 T로 조건을 걸어 UMVUE $\varphi(T)$를 구하라(답의 변수는 t와 n).

(c) φ가 UMVUE인 근거를 정리 번호로 적어라.

(d) $n = 10$, $T = 7$이면 UMVUE의 값은 얼마인가?

(e) 같은 자료에서 MLE $\tau(\bar{x})$의 값을 구하고 비교하라.

풀이 (a) $W = 1\{X_1 = 1\}$은 $E_\lambda W = \lambda e^{-\lambda}$이므로 불편이다. (b) $T = t$가 주어지면 $X_1 \sim \text{b}(t, \frac{1}{n})$이므로(노트 연습 1.1) $\varphi(t) = P(X_1 = 1 \mid T = t) = t \cdot \frac{1}{n}\big(1 - \frac{1}{n}\big)^{t-1}$이다. (c) T는 1-모수 지수족의 완비충분통계량이고(정리 4.5), $\varphi(T)$는 불편추정량에 조건을 건 것이므로 불편이다(정리 2.1(b)). 정리 3.3으로 유일한 UMVUE다. (d) $0.7 \cdot 0.9^6 = 0.3720$이다. (e) $\bar{x} = 0.7$이므로 $0.7e^{-0.7} = 0.3476$이다. 두 값의 차는 0.024이고, $\lambda = 0.7$에서 UMVUE의 표준편차 0.051보다 작다.

최종답 (b) $\frac{T}{n}(1 - \frac{1}{n})^{T-1}$ (d) $0.7 \cdot 0.9^6 \approx 0.3720$ (e) $0.7e^{-0.7} \approx 0.3476$

D3 정규분포의 σ와 μ^2 ★★★

정리 3.3, 4.5, 6회 정리 2.2 · 제한 35분

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

(a) $ES = \sigma\sqrt{\frac{2}{n-1}}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$임을 보여라.

(b) σ의 UMVUE가 $c_n S$, $c_n = \sqrt{\frac{n-1}{2}}\,\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}$임을 보여라.

(c) c_5를 닫힌 꼴과 소수로 구하라.

(d) μ^2의 UMVUE를 구하고, $n = 5$, $\bar{x} = 3$, $s^2 = 2$일 때 값을 구하라.

(e) (d)의 UMVUE가 음수가 될 수 있는가? 7절의 터무니없는 UMVUE와 비교하라.

풀이 (a) $V = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$이다(6회 정리 2.2). $k = n - 1$이면

$$E\sqrt{V} = \int_0^\infty v^{1/2}\frac{v^{k/2-1}e^{-v/2}}{\Gamma(k/2)2^{k/2}}dv = \frac{\Gamma((k+1)/2)2^{(k+1)/2}}{\Gamma(k/2)2^{k/2}} = \sqrt{2}\,\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$$

이고 $S = \sigma\sqrt{V/(n-1)}$이므로 주어진 식이 나온다. (b) (a)로 $c_n S$는 불편이고 $(\bar{X}, S^2)$는 완비충분이므로(정리 4.5) 정리 3.3으로 UMVUE다. (c) $c_5 = \sqrt{2}\,\frac{\Gamma(2)}{\Gamma(5/2)} = \sqrt{2} \cdot \frac{4}{3\sqrt{\pi}} = \frac{4\sqrt{2}}{3\sqrt{\pi}} \approx 1.0638$이다. S는 σ를 약 6% 작게 추정한다. (d) $E\bar{X}^2 = \mu^2 + \sigma^2/n$, $ES^2 = \sigma^2$이므로 $\bar{X}^2 - S^2/n$은 불편이고 UMVUE다. 값은 $9 - 0.4 = 8.6$이다. (e) $\bar{X}^2 < S^2/n$이면 음수다. $\mu^2 \geq 0$인데 음수로 추정할 수 있다. 노트 예제 7.2의 $(-1)^X$처럼 불편이라는 조건이 값의 범위를 지키지 못한다. $\max(\bar{X}^2 - S^2/n, 0)$은 평균제곱오차가 더 작지만 불편이 아니다.

최종답 (c) $c_5 = \frac{4\sqrt{2}}{3\sqrt{\pi}} \approx 1.0638$ (d) 8.6

D4 균등분포의 분포함수 값 ★★★

정리 2.1, 3.3 · 제한 35분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{U}(0,\theta)$이고 $c > 0$은 알려진 상수다. $\tau(\theta) = P_\theta(X_1 \le c) = \min(c/\theta, 1)$의 UMVUE를 구한다. $M = X_{(n)}$이다.

(a) $M = m$이 주어지면 확률 $\frac{1}{n}$로 $X_1 = m$이고, 나머지 확률로 $X_1 \sim \mathrm{U}(0, m)$임을 설명하라.

(b) $\varphi(m) = E\big[1\{X_1 \le c\} \mid M = m\big]$을 구하라.

(c) $\varphi(M)$이 UMVUE인 근거를 적어라.

(d) $n = 4$, $M = 5$, $c = 2$이면 UMVUE의 값은 얼마인가?

풀이 (a) 표본은 교환가능하므로 n개 가운데 X_1이 최댓값일 확률은 $\frac{1}{n}$이다. X_1이 최댓값이 아니면, $M = m$이 주어질 때 최댓값이 아닌 $n-1$개는 $(0, m)$에서 독립인 균등분포다(독일 전차의 연속판, 노트 연습 2.2와 같은 논리). (b) $m \le c$이면 $X_1 \le M \le c$이므로 $\varphi(m) = 1$이다. $m > c$이면 $\varphi(m) = \frac{1}{n}\cdot 0 + \frac{n-1}{n}\cdot\frac{c}{m} = \frac{(n-1)c}{nm}$이다. (c) $1\{X_1 \le c\}$는 불편이고 M은 완비충분이므로(예제 3.2(c)) 정리 2.1과 3.3으로 $\varphi(M)$이 UMVUE다. 직접 확인: $\theta > c$이면 $\int_0^c \frac{nm^{n-1}}{\theta^n} dm + \int_c^\theta \frac{(n-1)c}{nm}\frac{nm^{n-1}}{\theta^n} dm = \frac{c^n}{\theta^n} + \frac{c(\theta^{n-1} - c^{n-1})}{\theta^n} = \frac{c}{\theta}$이다. (d) $\frac{3\cdot 2}{4\cdot 5} = \frac{3}{10}$이다. MLE $c/M = 0.4$보다 작다.

최종답 (d) $\frac{3}{10}$