수리통계학 × 통계학의 역사

15회 워크북

가정이 틀리면
손으로 푸는 35문항

강의노트 15회를 읽은 뒤 손으로 푼다. 순위의 정확한 분포를 세고, M-추정량을 반복으로 풀고, 부트스트랩 분포를 모든 표본을 적어 계산한다. 노트의 증명은 빈 종이에 다시 쓴다. 문제 아래의 줄 친 공간에 직접 쓴다.

부	하는 일	문항	예상 시간
A 계산 드릴	순위 · 효율 · M-추정 · 잭나이프 · 부트스트랩을 손으로 계산한다	14	2시간 16분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 50분
C 증명 문제	변형, 일반화, 반례 만들기	10	2시간 47분
D 퀄형	여러 절을 잇는 종합 문제. 제한 시간 안에	4	2시간 20분
	합계	35	약 10시간

노트 `15_robust_bootstrap_notes.pdf` — 문항의 "정리 3.2" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T1] 원소 추적, [T3] 귀납, [T4] 부등식, [T5] 극한, [T6] 극한과 적분의 교환, [T7] 변수변환, [T8] Taylor

검산 노트북 `15_robust_bootstrap.ipynb`의 "워크북 검산" 절: `check("A2c", "11/35")`

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. 표를 써서 구한 소수 답은 표의 정밀도만큼 맞으면 맞다고 본다.
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.

예: 빈칸과 풀이 공간

① 순위 벡터는 $N!$개 순열 위에서 균등하다. 왜? _____

A

계산 드릴

순위, 정확한 귀무분포, 효율, M-추정, 잭나이프와 부트스트랩을 손으로 계산한다. 답이 수나식으로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다. 표준정규 분위수가 필요한 곳은 표를 쓴다.

A1 순위합과 Mann–Whitney ★

정의 1.4, 정리 1.5(a) · 6분

$X = (1.8,\ 3.1,\ 4.6)$, $Y = (2.2,\ 2.9,\ 5.0,\ 5.7)$이다. (a) 합친 표본에서의 순위로 Wilcoxon 순위합 W를 구하라. (b) Mann–Whitney 통계량 $U = \#\{(i,j) : X_i > Y_j\}$를 직접 세고 $W = U + m(m+1)/2$를 확인하라.

A2 작은 표본의 정확한 귀무분포 ★★

정리 1.5(b) · 12분

$m = 3$, $n = 4$이고 $F = G$가 연속이다. (a) EW, (b) $\operatorname{Var} W$를 정리의 공식으로 구하라. (c) $\{1, \dots, 7\}$에서 고른 세 수의 합이 10 이하인 경우를 모두 세어 $P(W \le 10)$을 구하라.

A3 부호순위의 정확한 꼬리 확률 ★★

정리 1.8 · 12분

짝 차이가 $d = (1.2,\ -0.4,\ 2.5,\ 0.9,\ -1.7,\ 3.1)$이다. (a) T^+를 구하라. (b) 귀무가설 아래 $\mathrm{Var}\, T^+$를 구하라. (c) 정리 1.8의 표현 $T^+ = \sum k I_k$를 써서 정확한 $P(T^+ \geq 16)$을 구하라.

A4 Hodges–Lehmann 추정 ★

정의 1.9 · 8분

짝 차이 $d = (-1,\ 2,\ 3,\ 7)$에서 Walsh 평균 $(d_i + d_j)/2$ $(i \leq j)$를 모두 적고 Hodges–Lehmann 추정값을 구하라. 양수인 Walsh 평균의 수가 T^+와 같음도 확인하라.

A5 두 표본 순열검정 ★

정리 1.2 · 8분

$X = (5, 7, 9)$, $Y = (1, 2, 4)$이다. $T = \bar{X} - \bar{Y}$로 한쪽 순열검정을 할 때 정확한 p-값을 구하라.

A6 두 분포에서의 효율 ★

정리 2.3 · 10분

$\mathrm{ARE}(W,t) = 12\sigma^2(\int f^2)^2$을 (a) 지수분포 $f(x)=e^{-x}$ $(x>0)$와 (b) 삼각분포 $f(x)=1-|x|$ $(|x|<1)$에서 구하라.

A7 부호검정의 효율 ★

명제 2.2, 노트 연습 2.4 · 8분

부호검정의 t 검정에 대한 ARE는 $4\sigma^2 f(0)^2$이다. (a) 표준 Laplace 밀도 $\frac{1}{2}e^{-|x|}$와 (b) 표준 로지스틱 밀도 $f(x)=e^{-x}/(1+e^{-x})^2$ $(\sigma^2=\pi^2/3)$에서 구하라.

A8 Huber 추정량을 손으로 ★★

정의 3.1, 알고리즘 6.2 · 12분

자료 $(1,\ 2,\ 3,\ 4,\ 20)$, $k=1.5$, 척도 $s=1$로 고정한다. (a) $\sum_i \psi_k(x_i-\theta)=0$의 해를 구하라. (b) 평균 $\theta^{(0)}=6$에서 시작해 무게 $w_i=\min(1,\,k/|x_i-\theta^{(0)}|)$로 가중평균 한 번 $(\theta^{(1)}=\sum w_i x_i/\sum w_i)$을 계산하라.

A9 영향함수의 값 ★

명제 3.5 · 8분

$F = \mathcal{N}(0,1)$에서 $x = 3$의 영향함수 값을 (a) Huber 추정량($k = 1.345$), (b) 평균, (c) 중앙값에 대해 구하라. $\Phi(1.345) = 0.9107$을 쓴다.

A10 잭나이프의 치우침 수정 ★★

정의 5.6 · 10분

자료 $(1, 3, 4, 8)$에서 $\hat\theta = \bar X^2$으로 μ^2을 추정한다. 관측 i를 뺀 값 $\hat\theta_{(i)}$와 그 평균 $\hat\theta_{(\cdot)}$으로 (a) 잭나이프 치우침 추정 $(n-1)(\hat\theta_{(\cdot)} - \hat\theta)$와 (b) 수정한 추정량 $\hat\theta - (n-1)(\hat\theta_{(\cdot)} - \hat\theta)$를 구하라. (b)가 $\bar X^2 - s^2/n$과 같은지 확인하라.

A11 정확한 부트스트랩 분포 ★★

정리 5.3(a) · 10분

자료 $(0, 3, 9)$에서 크기 3의 부트스트랩 표본의 평균 $\bar X^*$를 생각한다. (a) $\mathrm{Var}^*\,\bar X^*$를 정리 5.3(a)로 구하라. (b) $3^3 = 27$가지 순서 있는 표본을 세어 $P^*(\bar X^* > \bar X)$를 구하라.

A12 부트스트랩 최댓값이 둘째로 큰 값일 확률 ★★

정리 7.1 · 8분

서로 다른 값 $n = 5$개에서 부트스트랩 표본을 뽑는다. $X^*_{(n)} = X_{(n-1)}$, 곧 부트스트랩 표본의 최댓값이 원래 자료의 둘째로 큰 값일 확률을 구하라.

A13 BCa의 수정된 분위수 ★★

정의 5.4 · 12분

$B = 1000$개의 부트스트랩 값 가운데 560개가 $\hat\theta$보다 작았고, 잭나이프로 구한 가속은 $\hat a = 0.05$이다. (a) $\hat z_0$, (b) 95% 구간의 아래쪽 수준 α_1 ($z_{(1)} = -1.960$), (c) 위쪽 수준 α_2 ($z_{(2)} = 1.960$)를 구하라. $\Phi^{-1}(0.56) = 0.151$을 쓴다.

A14 Laplace에서의 Huber 점근분산 ★★

정리 3.2 · 12분

표준 Laplace 밀도 $\tfrac{1}{2}e^{-|x|}$에서 $k = 1$인 Huber 추정량의 $V(\psi_1, F) = E\psi_1^2/(E\psi_1')^2$을 구하라. 같은 분포에서 평균의 V는 2, 중앙값의 V는 1이다.

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 정리나 정의의 번호, 앞 회차의 정리, 도구 번호를 쓴다.

B1 순위합의 귀무분포, 근거 채우기 ★

① · 정리 1.5(a)(b) · 15분

$F = G$가 연속일 때 $W = U + \frac{m(m+1)}{2}$, $EW = \frac{m(N+1)}{2}$, $\operatorname{Var} W = \frac{mn(N+1)}{12}$이다. 각 단계의 근거를 써라.

① $X_{(k)}$의 합친 순위는 $(k-1) + \#\{j : Y_j < X_{(k)}\} + 1$이다. 왜? _____

② k에 대해 더하면 $W = \frac{m(m+1)}{2} + U$이다. 왜? _____

③ $(R(X_1), \dots, R(X_m))$은 $\{1, \dots, N\}$에서 m개를 비복원으로 뽑은 것과 같은 분포이다. 왜? _____

④ 모집단 $\{1, \dots, N\}$의 평균은 $\frac{N+1}{2}$, 분산은 $\frac{N^2-1}{12}$이다. 왜? _____

⑤ $\operatorname{Var} W = m \cdot \frac{N^2-1}{12} \cdot \frac{N-m}{N-1} = \frac{mn(N+1)}{12}$이다. 왜? _____

⑥ W와 $m(N+1) - W$는 같은 분포다. 왜? _____

B2 M-추정량의 점근정규성, 근거 채우기 ★★

① · 정리 3.2 · 20분

정리 3.2의 조건(대칭, 홀함수이고 단조증가하는 ψ, ψ'' 유계)에서 $\sqrt{n}\,\hat\theta \xrightarrow{D} \mathcal{N}(0, V)$이다. 근거를 채워라. $\Psi_n(t) = \frac{1}{n}\sum \psi(X_i - t)$, $\lambda(t) = E\psi(X - t)$.

① Ψ_n은 감소하고 $\hat\theta$는 그 영점이다. 왜? _____

② $\lambda(0) = 0$이고 $\lambda'(0) = -E\psi'(X) < 0$이다. 왜? _____

③ $\hat\theta \xrightarrow{P} 0$이다. 왜? _____

④ $0 = \Psi_n(0) - \hat\theta A_n + \frac{1}{2}\hat\theta^2 B_n$이다. 왜? _____

⑤ $\sqrt{n}\,\hat\theta = \dfrac{n^{-1/2}\sum \psi(X_i)}{A_n - \frac{1}{2}\hat\theta B_n}$이다. 왜? _____

⑥ 분자 $\xrightarrow{D} \mathcal{N}(0, E\psi^2)$이다. 왜? _____

⑦ 분모 $\xrightarrow{P} E\psi'(X)$이다. 왜? _____

⑧ $\sqrt{n}\,\hat\theta \xrightarrow{D} \mathcal{N}(0, E\psi^2/(E\psi')^2)$이다. 왜? _____

B3 하한 108/125, 근거 채우기 ★★

① · 정리 2.5 · 15분

분산이 유한한 모든 밀도에서 $12\sigma^2(\int f^2)^2 \geq \frac{108}{125}$이다. 근거를 채워라. $h(x) = c(5 - x^2)_+$, $c = \frac{3}{20\sqrt{5}}$, $a = 5c$, $b = c$.

① 평균 0, 분산 1로 놓아도 된다. 왜? _____

② $\int h = 1$, $\int x^2 h = 1$이다. 왜? _____

③ $\int f^2 \geq 2\int fh - \int h^2$이다. 왜? _____

④ $\int fh \geq \int f(a - bx^2) = a - b$이다. 왜? _____

⑤ $\int h^2 = a - b = \frac{3\sqrt{5}}{25}$이다. 왜? _____

⑥ $12(\int f^2)^2 \geq 12 \cdot \frac{45}{625} = \frac{108}{125}$이다. 왜? _____

B4 순위합의 귀무분포, 단계의 이름만 ★★

② · 정리 1.5 · 25분

정리 1.5의 (a)(b)(c)를 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **(a) $X_{(k)}$의 순위를 세 부분으로 쪼갠다**

② **k에 대해 더한다**

③ **(b) X 쪽 순위를 비복원 표본으로 본다**

④ **모집단 $\{1, \dots, N\}$의 평균과 분산**

⑤ **비복원 표본합의 평균과 분산**

⑥ **대칭: $r \mapsto N + 1 - r$**

⑦ **(c) U를 지시변수의 합으로 쓰고 기댓값을 취한다**

B5 M-추정량의 점근정규성, 단계의 이름만 ★★

② · 정리 3.2 · 30분

정리 3.2를 증명하라(ψ''가 유계라고 가정해도 된다).

① **영점 $\hat{\theta}$와 감소하는 Ψ_n**

② **$\lambda(t) = E\psi(X - t)$의 0 근처 모양**

③ **일치성 $\hat{\theta} \xrightarrow{P} 0$**

④ **영점의 Taylor 전개**

⑤ **비로 쓰기**

⑥ **분자: 중심극한정리**

⑦ **분모: 대수의 법칙과 나머지**

⑧ **Slutsky로 합치기**

B6 순위합의 귀무분포, 전략만 ★★

③ · 정리 1.5 · 30분

정리 1.5의 (a)(b)(c)를 증명하라.

전략 (a)는 X의 순위를 "작은 X의 수 + 작은 Y의 수 + 1"로 센다. (b)는 순위 벡터의 균등성으로 W를 비복원 표본합으로 바꾼다. (c)는 지시변수의 합이다.

B7 하한 108/125, 전략만 ★★★ ③ · 정리 2.5 · 35분

분산이 유한한 모든 밀도 f에서 $\mathrm{ARE}(W,t) = 12\sigma^2(\int f^2)^2 \geq \frac{108}{125}$임을 증명하고, 등호가 성립하는 밀도를 밝혀라.

전략 위치와 척도를 정규화하고, 포물선 밀도 h와의 차의 제곱을 적분해 $\int f^2$를 아래에서 누른다.

C

증명 문제

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

C1 순열 p-값의 정확한 분포 ★★

정리 1.2 · 15분

정리 1.2의 상황에서 $M = \binom{N}{m}$이라 하자. 합친 표본의 값이 주어졌을 때 M가지 나눔의 T 값이 모두 다르면, 귀무가설 아래 $k = 1, \dots, M$에서 $P(p \le k/M) = k/M$임을 보여라. p의 가장 작은 값은 무엇이고, 그것이 크기 $\alpha < 1/M$의 검정에 대해 무엇을 뜻하는가?

C2 지시변수로 구하는 $\operatorname{Var} U$ ★★

정리 1.5(b), 3회 정리 4.2(c) · 20분

$U = \sum_{i,j} I_{ij}$, $I_{ij} = \mathbf{1}\{X_i > Y_j\}$이다. 귀무가설($F = G$ 연속) 아래 다음을 보이고 $\operatorname{Var} U = \frac{mn(N+1)}{12}$을 순위를 쓰지 않고 다시 얻어라.

(a) $\operatorname{Var} I_{ij} = \frac{1}{4}$.

(b) $j \neq k$이면 $\operatorname{Cov}(I_{ij}, I_{ik}) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$. $i \neq l$이면 $\operatorname{Cov}(I_{ij}, I_{lj}) = \frac{1}{12}$.

(c) 첨자가 하나도 겹치지 않으면 공분산은 0이다.

C3 Hodges–Lehmann 추정량의 붕괴점 ★★★

정의 1.9, 3.6 · 25분

한 표본의 Hodges–Lehmann 추정량은 Walsh 평균 $M = n(n+1)/2$개의 중앙값이다.

(a) k개의 관측을 바꾸면 영향을 받는 Walsh 평균의 수가 $M - \frac{(n-k)(n-k+1)}{2}$ 임을 보여라.

(b) 영향을 받는 Walsh 평균이 $M/2$보다 많아야 추정량을 끝없이 보낼 수 있다. 붕괴점 k^*/n이 $n \to \infty$에서 $1 - 1/\sqrt{2} \approx 0.293$으로 감을 보여라.

(c) 평균(0), 중앙값($\frac{1}{2}$)과 비교하라.

C4 부호검정의 효율은 0에서 무한대까지 ★★

명제 2.2, 노트 연습 2.4 · 15분

부호검정의 t 검정에 대한 ARE $4\sigma^2 f(0)^2$에 대해

(a) ARE가 0인 대칭 밀도를 만들어라.

(b) t_ν 분포($\nu > 2$, $\sigma^2 = \nu/(\nu-2)$)에서 $\nu \downarrow 2$이면 ARE가 무한대로 감을 보여라.

(c) t_3에서 ARE를 구하라. t_3의 $f(0) = \dfrac{2}{\pi\sqrt{3}}$이다.

C5 M-추정량의 이동 등변성과 유일성 ★

정의 3.1 · 12분

척도 s를 고정한 위치 M-추정량에서 ψ가 연속이고 순증가하는 홀함수라 하자.

(a) 방정식 $\sum \psi((x_i - \theta)/s) = 0$의 해가 하나뿐임을 보여라.

(b) 모든 x_i에 c를 더하면 해도 c만큼 움직임을 보여라.

(c) 자료가 어떤 점 c에 대해 대칭이면(곧 $\{x_i - c\} = \{c - x_i\}$) 해는 c임을 보여라.

C6 분산 추정량의 잭나이프 ★★

정의 5.6 · 15분

$\hat\theta = \hat\sigma^2 = \frac{1}{n}\sum (X_i - \bar X)^2$ 이다. 잭나이프로 치우침을 고친 추정량 $n\hat\theta - (n-1)\hat\theta_{(\cdot)}$ 가 $s^2 = \frac{1}{n-1}\sum (X_i - \bar X)^2$ 과 정확히 같음을 보여라.

C7 부트스트랩은 왜도를 본다 ★★

정리 5.3 · 15분

(a) $E^*\big[(\bar X^* - \bar X)^3\big] = \hat\mu_3 / n^2$, $\hat\mu_3 = \frac{1}{n}\sum (X_i - \bar X)^3$ 임을 보여라. (b) 자료 $(0, 3, 9)$ 에서 이 값을 구하고, A11의 27가지 표본으로 확인하라. (c) 정규근사 $\mathcal{N}(\bar X, \hat\sigma^2/n)$ 은 이 3차 적률을 얼마로 주는가?

C8 IRLS는 목적함수를 줄인다 ★★★

명제 6.3, 명제 8.3 · 25분

위치 문제에서 Huber 목적함수 $L(\theta)=\sum_i \rho_k(x_i-\theta)$를 생각한다(척도 1).

(a) $g(t)=\rho_k(\sqrt{t})$ $(t\ge 0)$가 오목함수임을 보여라.

(b) 오목성으로 $\rho_k(u)\le \rho_k(u_0)+\frac{w_0}{2}(u^2-u_0^2)$, $w_0=\psi_k(u_0)/u_0$을 보여라.

(c) IRLS의 한 걸음 $\theta^{(1)}=\sum w_i x_i/\sum w_i$ $(w_i=\psi_k(r_i)/r_i,\ r_i=x_i-\theta^{(0)})$이 $L(\theta^{(1)})\le L(\theta^{(0)})$을 만족함을 보여라.

C9 독립이 아니어도 교환가능하면 ★★

정의 1.1, 정리 1.2 · 15분

$Z_i=\eta+\varepsilon_i$ $(i=1,\dots,N)$이고 $\eta\sim\mathcal{N}(0,\tau^2)$, $\varepsilon_i\overset{\text{iid}}{\sim}\mathcal{N}(0,1)$이 서로 독립이다.

(a) Z_i들은 독립이 아님을 보여라$(\mathrm{Cov}(Z_1,Z_2))$.

(b) $(Z_1,\dots,Z_N)$이 교환가능함을 보여라.

(c) 처음 m개를 X, 나머지를 Y로 둔 순열검정이 여전히 정확함을 보여라. 정리 1.2의 증명 가운데 어느 단계가 독립을 쓰지 않았는가?

C10 순위 통계량은 단조변환에 불변이다 ★

정의 1.4 · 10분

h가 순증가함수이면 X_i, Y_j를 $h(X_i)$, $h(Y_j)$로 바꿔도 W와 U가 그대로임을 보여라. 두 표본 t 통계량은 그렇지 않음을 $X=(1,2)$, $Y=(3,10)$과 $h(x)=\log x$로 보여라.

D

퀄형

여러 절과 앞 회차의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 표준정규 분위수와 분포함수 값은 표를 쓴다.

D1 정규 이동 모형에서 두 검정의 국소 검정력 ★★★

명제 2.2, 정리 2.3 · 제한 35분

$X_1, \dots, X_m \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta,1)$이고 $m = n = 50$이다. 크기 5%의 한쪽 검정으로 $H_0 : \theta = 0$ 대 $\theta > 0$을 검정한다.

(a) $p(\theta) = P(Y > X)$를 Φ로 쓰고 $\theta = 0.5$에서 값을 구하라.

(b) 명제 2.2와 정리 2.3의 증명에 나온 효능 e_W로 Wilcoxon 검정의 $\theta = 0.5$에서의 근사 검정력을 구하라($N = 100$, $\lambda = \frac{1}{2}$).

(c) 같은 방법으로 t 검정의 근사 검정력을 구하고, 두 검정력의 차이를 ARE $3/\pi$와 연결해 설명하라.

D2 대립가설 아래의 Mann–Whitney 분산 ★★★ 정리 1.5(c), 2회 정리 2.8 · 제한 40분

$X \sim \mathrm{Exp}(1)$, $Y \sim \mathrm{Exp}(2)$ (평균 $\frac{1}{2}$)이고 모든 관측이 독립이다. $p = P(X > Y)$, $q_1 = P(X > Y_1, X > Y_2)$, $q_2 = P(X_1 > Y, X_2 > Y)$로 둔다.

(a) p를 구하라.

(b) q_1을 구하라.

(c) q_2를 구하라.

(d) C2의 방법으로 일반적으로 $\mathrm{Var}\,U = mn\big[p(1-p) + (n-1)(q_1 - p^2) + (m-1)(q_2 - p^2)\big]$ 임을 보이고, $m = n = 10$에서 $\mathrm{Var}\big(U/(mn)\big)$을 구하라. 귀무가설 아래의 값 $\frac{N+1}{12mn}$과 비교하라.

D3 치우친 자료의 부트스트랩 구간 ★★★

정리 5.3, 알고리즘 5.2 · 제한 30분

자료가 $(0, 0, 0, 0, 10)$이다. K를 부트스트랩 표본에 10이 뽑힌 횟수라 하자.

(a) $\bar{X}^* = 2K$이고 $K \sim \mathrm{b}(5, \frac{1}{5})$임을 보이고 $P^*(\bar{X}^* = 0)$을 구하라.

(b) $\mathrm{Var}^* \bar{X}^*$를 두 가지로 구하라(이항분포와 정리 5.3(a)).

(c) 95% 백분위 구간의 위 끝(97.5% 분위수)을 구하라. 아래 끝은 무엇인가?

(d) t 구간 $\bar{X} \pm t_{0.975,4}\, s/\sqrt{5}$ ($t_{0.975,4} = 2.776$)와 비교하고, 어느 쪽이 이 자료에 맞는지 논하라.

D4 순열분포의 평균과 분산, 그리고 정규근사 ★★★

정리 1.2, 4회 정리 1.5 · 제한 35분

합친 값 $\mathbf{Z} = (Z_1, \dots, Z_N)$을 고정하고, 그 가운데 m개를 무작위로 골라 X로, 나머지 n개를 Y로 둔다. $\tau^2 = \frac{1}{N}\sum (Z_i - \bar{Z})^2$이다.

(a) 순열분포에서 $E(\bar{X} - \bar{Y}) = 0$임을 보여라.

(b) $\mathrm{Var}(\bar{X} - \bar{Y}) = \dfrac{N^2 \tau^2}{mn(N-1)}$임을 보이고, A5의 자료 $\{1, 2, 4, 5, 7, 9\}$, $m = n = 3$에서 값을 구하라.

(c) 정규근사 $P\big(\mathcal{N}(0, \mathrm{Var}) \geq 14/3\big)$을 구하고 A5의 정확한 p-값 0.05와 비교하라.

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 일곱 값을 크기 순서로 늘어놓고 X의 자리에 표시한다. U는 각 X_i보다 작은 Y의 수를 더한 것이다.

답 (a) $W = 10$ (b) $U = 4$

A2 힌트 (c) 가능한 부분집합은 $\binom{7}{3} = 35$개다. 가장 작은 합은 $1+2+3 = 6$이다. 합이 6, 7, 8, 9, 10인 부분집합을 차례로 적는다.

답 (a) $EW = 12$ (b) $\operatorname{Var} W = 8$ (c) $\frac{11}{35} \approx 0.314$

A3 힌트 (c) $\{1, \dots, 6\}$의 부분집합 $2^6 = 64$개 가운데 합이 16 이상인 것을 센다. 전체 합이 21이므로 "빠진 수의 합이 5 이하"인 부분집합을 세는 편이 쉽다.

답 (a) $T^+ = 16$ (b) $\frac{91}{4} = 22.75$ (c) $\frac{5}{32} \approx 0.156$

A4 힌트 $i \le j$이므로 Walsh 평균은 $4 \cdot 5/2 = 10$개이고, 그 가운데 4개는 d_i 자신이다. 짝수 개의 중앙값은 가운데 두 값의 평균이다.

답 2.75

A5 힌트 합친 값의 합이 28로 고정되므로 T는 X 쪽 세 값의 합 s의 증가함수다. $T \ge T_{\text{obs}}$는 $s \ge 21$과 같다.

답 $\frac{1}{20}$

A6 힌트 (a) $\sigma^2 = 1$, $\int_0^\infty e^{-2x}dx$. (b) 대칭이므로 $\sigma^2 = 2\int_0^1 x^2(1-x)\,dx$, $\int f^2 = 2\int_0^1 (1-x)^2 dx$.

답 (a) 3 (b) $\frac{8}{9} \approx 0.889$

A7 힌트 $f(0)$만 구하면 된다. 로지스틱은 $f(0) = 1/4$.

답 (a) 2 (b) $\frac{\pi^2}{12} \approx 0.822$

A8 힌트 (a) 답이 3 근처라고 짐작하고, 어느 잔차가 잘리는지 정한 뒤 θ에 대한 일차방정식을 풀고, 짐작이 맞는지 확인한다.

답 (a) $\hat\theta = 3$ (b) $\theta^{(1)} = \frac{2154}{569} \approx 3.786$

A9 힌트 명제 3.5의 세 식에 넣는다. Huber는 $\psi_k(3) = k$이다.

답 (a) 1.637 (b) 3 (c) $\sqrt{\pi/2} \approx 1.253$

A10 힌트 하나를 뺀 평균은 $(16 - x_i)/3$이다. $\hat\theta_{(i)}$는 그 제곱이다.

답 (a) $\frac{13}{6}$ (b) $\frac{83}{6} \approx 13.83$

A11 힌트 (b) $\bar X^* > 4$는 세 값의 합이 12보다 큰 것이다. 9가 몇 번 뽑혔는지로 나눠 센다.

답 (a) $\frac{14}{3}$ (b) $\frac{10}{27}$

A12 힌트 "최댓값이 $X_{(n-1)}$ 이하"에서 "최댓값이 $X_{(n-2)}$ 이하"를 뺀다.

답 $\frac{781}{3125} \approx 0.250$

A13 힌트 $\hat z_0 + z_{(1)} = -1.809$이다. 분모 $1 - \hat a(\hat z_0 + z)$를 먼저 계산한다.

답 (a) $\hat z_0 \approx 0.151$ (b) $\alpha_1 \approx 0.0658$ (c) $\alpha_2 \approx 0.9940$

A14 힌트 $E\psi_1' = P(|X| < 1)$이다. $E\psi_1^2 = \int_0^1 x^2 e^{-x} dx + P(|X| > 1)$이고 앞의 적분은 부분적분 두 번이다.

답 $\frac{2 - 4/e}{(1 - 1/e)^2} \approx 1.323$

B1 힌트 근거는 짧게 쓴다. 넷째 단계는 $\sum_{k=1}^{N} k^2 = N(N+1)(2N+1)/6$에서 평균의 제곱을 뺀다.

B2 힌트 대수의 법칙이 쓰이는 곳이 두 군데, 중심극한정리가 한 군데다. ③에서는 감소함수의 영점이 두 점 사이에 갇힌다는 것을 쓴다.

B3 힌트 넷째 단계는 받침 밖에서 $a - bx^2 < 0 \le h$이고 받침 안에서 등호라는 것이다.

B4 힌트 B1에서 채운 근거를 다시 보지 말고 쓴다. (c)는 귀무가설을 쓰지 않는다.

B5 힌트 9회 정리 3.1의 증명과 같은 뼈대다. 점수함수 $s(x;\theta)$의 자리에 $\psi(x - \theta)$가 들어간다.

B6 힌트 $\operatorname{Var} W$의 계산에서 $m \cdot \frac{N^2-1}{12} \cdot \frac{N-m}{N-1}$을 정리하는 단계까지 쓴다.

B7 힌트 h의 상수는 $\int h = 1$과 $\int x^2 h = 1$에서 정한다. 받침의 끝 $\sqrt{5}$는 두 조건을 함께 맞추는 값이다.

C1 힌트 조건부로 관측된 나눔은 M가지 위에서 균등하다. T 값이 모두 다르면 $p =$ (관측값 이상인 T의 수)$/M$은 $1/M, 2/M, \dots, 1$을 한 번씩 갖는다.

C2 힌트 (b) $P(X>Y_1, X>Y_2)$는 세 iid 연속변수 가운데 X가 가장 클 확률이다. 공분산 항의 개수를 센다: 같은 i를 공유하는 순서쌍은 $mn(n-1)$개.

C3 힌트 (a) 바뀌지 않은 $n-k$개끼리 만든 Walsh 평균만 영향을 받지 않는다. (b) $\frac{(n-k)^2}{2} \approx \frac{n^2}{4}$을 k/n에 대해 푼다 [T5].

답 (c) $1-\frac{1}{\sqrt{2}} \approx 0.293$

C4 힌트 (a) 0 근처에 질량이 없는 대칭 밀도. (b) $f(0)$는 $\nu \to 2$에서 양수로 수렴한다.

답 (c) $\frac{16}{\pi^2} \approx 1.62$

C5 힌트 (a) $g(\theta)=\sum \psi((x_i-\theta)/s)$는 순감소하고 연속이며, $\theta \to \mp\infty$에서 부호가 바뀐다 [T5].

C6 힌트 $\hat\theta_{(\cdot)}=\frac{1}{n}\sum_i \hat\theta_{(i)}$에서 $\sum_i \hat\theta_{(i)}$를 계산한다. 각 쌍 (j,l), $j<l$가 몇 번 나오는지 센다: $\hat\sigma^2=\frac{1}{n^2}\sum_{j<l}(X_j-X_l)^2$.

C7 힌트 $D_i=X_i^*-\bar X$는 조건부로 iid, 평균 0이다. $(\sum D_i)^3$을 전개하면 $E^*D_iD_jD_k$는 세 첨자가 모두 같을 때만 0이 아니다.

답 (b) $\frac{20}{9}$

C8 힌트 (a) $g'(t)$는 $t<k^2$에서 $\frac{1}{2}$, $t>k^2$에서 $\frac{k}{2\sqrt{t}}$이다. (b) 오목함수는 접선 아래에 있다 [T4]. (c) (b)의 오른쪽을 θ에 대해 더하면 가중 제곱합이고, 그 최소점이 $\theta^{(1)}$이다.

C9 힌트 (b) 조건부로 η가 주어지면 Z_i는 iid다. 결합분포를 η에 대해 적분한 꼴로 써서 순열에 불변임을 본다.

C10 힌트 h가 순증가이면 $X_i > Y_j \iff h(X_i) > h(Y_j)$이다.

D1 힌트 (a) $Y-X \sim \mathcal{N}(\theta, 2)$. (b) $e_W=\sqrt{12\lambda(1-\lambda)}\int\phi^2$, $\int\phi^2=\frac{1}{2\sqrt{\pi}}$, 검정력 $\approx 1-\Phi(z_{0.05}-he)$, $h=\theta\sqrt{N}$.

답 (a) $\Phi(0.354)\approx 0.638$ (b) ≈ 0.788 (c) ≈ 0.804

D2 힌트 (a) $P(X>Y)=E\,P(X>Y\mid Y)=Ee^{-Y}$. (b) $q_1=E\big[(1-e^{-2X})^2\big]$. (c) $\min(X_1,X_2)\sim$ Exp(2)이므로 $q_2=P(Y<\min)$은 같은 비율의 두 지수의 경주다.

답 (a) $p=\frac{2}{3}$ (b) $q_1=\frac{8}{15}$ (c) $q_2=\frac{1}{2}$ (d) $\frac{137}{9000}\approx 0.0152$

D3 힌트 $P(K=k)=\binom{5}{k}(0.2)^k(0.8)^{5-k}$: 0.328, 0.410, 0.205, 0.051, … . 누적확률이 처음으로 0.975 이상이 되는 k를 찾는다.

답 (a) $\frac{1024}{3125}\approx 0.328$ (b) $\frac{16}{5}$ (c) 6

D4 힌트 $\bar{X}-\bar{Y}=\frac{N}{n}\left(\frac{1}{m}\sum_{i\in X}Z_i-\bar{Z}\right)$이다. X 쪽 합은 크기 N의 모집단에서 비복원으로 뽑은 m개의 합이다(4회 정리 1.5, 유한 모집단 수정).

답 (b) $\frac{272}{45}\approx 6.04$ (c) ≈ 0.029