수리통계학 × 통계학의 역사

8회 워크북

추론의 기초
손으로 푸는 35문항

강의노트 8회를 읽은 뒤 손으로 푼다. 순서통계량의 분포, 신뢰구간과 검정의 계산, 쌍대성, p-값의 균등성, Pearson의 χ^2를 빈 종이에 다시 쓴다. 문제 아래의 줄 친 공간에 직접 쓴다.

부	하는 일	문항	예상 시간
A 계산 드릴	순서통계량, 구간, 검정력, p-값, χ^2를 손으로 계산한다	15	1시간 56분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 50분
C 증명 문제	변형, 일반화, 반례 만들기	9	2시간 40분
D 퀄형	여러 절과 앞 회차를 잇는 종합 문제. 제한 시간 안에	4	2시간 25분
	합계	35	약 10시간

노트 `08_inference_notes.pdf` — 문항의 "정리 5.1" 같은 번호는 노트의 번호다
도구 `00_toolkit_notes.pdf` — [T1] 원소 추적, [T5] 극한과 하한, [T8] o와 O
검산 노트북 `08_inference.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.** Pearson의 χ^2는 세 번, p-값의 분포와 순서통계량의 결합밀도는 두 번 만난다. 처음에는 근거만 빈칸이고(①), 다음에는 단계의 이름만 있고(②), 마지막에는 전략만 있다(③).
4. **계산은 검산한다.** 최종답이 수나 식인 문항은 노트북의 `check()`에 넣으면 맞는지만 알려 준다. 분위수를 표에서 읽은 답은 소수 넷째 자리까지 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.

A

계산 드릴

정의와 공식을 손으로 돌린다. 답이 수나 식으로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다. 분위수는 표나 계산기에서 읽는다. 표준정규의 위쪽 분위수는 $z_{0.05}=1.645$, $z_{0.025}=1.960$, $z_{0.10}=1.282$를 쓴다.

A1 균등분포의 둘째 순서통계량 ★

따름정리 1.3 · 8분

$X_1,\dots,X_5 \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$이다. (a) $EX_{(2)}$를 구하라. (b) $P(X_{(2)} \le \frac{1}{2})$를 구하라.

A2 지수 표본의 최솟값과 최댓값 ★

3회 예제 3.5 · 8분

$X_1,\dots,X_4$가 독립이고 평균 2인 지수분포를 따른다. (a) $EX_{(1)}$, (b) $P(X_{(4)} \le 2)$를 구하라.

A3 노획한 일련번호 ★

정리 1.6 · 6분

일련번호 $1,\dots,N$인 전차 가운데 4대를 노획했더니 번호가 19, 40, 42, 60이었다. (a) 정리 1.6의 불편추정량 $\hat{N}$의 값, (b) 적률법 추정량 $2\bar{X}-1$의 값을 구하라.

A4 최댓값의 분포 ★★

정리 1.6 · 10분

$\{1, \dots, 100\}$에서 4개를 비복원으로 뽑은 최댓값을 M이라 하자. (a) EM을 구하라. (b) $P(M \le 50)$을 구하라.

A5 지수 수명의 정확한 구간 ★★

정리 2.2 · 10분

수명이 평균 θ인 지수분포를 따르는 부품 5개의 수명의 합이 400시간이었다. 노트 예제 2.3의 축량으로 θ의 95% 신뢰구간을 구하라. ($\chi^2_{10,0.025} = 20.483$, $\chi^2_{10,0.975} = 3.247$)

A6 균등분포의 상한 ★

정리 2.2 · 8분

$X_1, \dots, X_4 \overset{\text{iid}}{\sim} \mathrm{U}(0, \theta)$에서 $M = X_{(4)} = 3.0$이었다. 축량 M/θ로 만든 90% 구간 $[M,\ M\alpha^{-1/n}]$의 위쪽 끝을 구하라.

A7 비율의 Wald 구간 ★

정리 2.4 · 6분

400명 가운데 88명이 “예” 라고 답했다. 비율의 95% Wald 구간의 (a) 아래쪽 끝, (b) 위쪽 끝을 구하라($z = 1.96$).

A8 분포를 가정하지 않는 중앙값의 구간 ★

정리 2.6 · 8분

연속분포의 표본 $n = 15$에서 구간 $[X_{(4)}, X_{(12)}]$가 중앙값을 덮을 확률을 정확히 구하라.

A9 이항 검정의 크기와 검정력 ★★

정의 3.2 · 10분

$X \sim \mathrm{b}(20, p)$로 $H_0 : p = \frac{1}{2}$ 대 $H_1 : p > \frac{1}{2}$를 검정하며 $X \geq 15$이면 기각한다. (a) 크기, (b) $p = 0.7$에서의 검정력을 구하라.

A10 표본 크기 ★

정리 3.4 · 8분

σ를 아는 정규 평균의 한쪽 z 검정을 크기 0.05로 할 때, $\delta = 0.25\sigma$의 차이를 검정력 0.9로 잡으려면 최소 몇 개의 관측이 필요한가?

A11 주어진 평균에서의 검정력 ★

정리 3.4 · 8분

$X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 10^2)$, $n = 25$로 $H_0 : \mu \le 100$을 크기 0.05의 한쪽 z 검정으로 한다. $\mu = 105$에서의 검정력을 구하라.

A12 p-값 ★

정의 4.1 · 6분

z 검정통계량의 관측값이 2.3이었다. (a) 양쪽 p-값, (b) 한쪽($H_1 : \mu > \mu_0$) p-값을 구하라. ($\Phi(2.3) = 0.98928$)

A13 주사위의 적합도 ★

정리 5.1 · 8분

주사위를 120번 던져 눈 1–6이 각각 18, 23, 16, 21, 18, 24번 나왔다. (a) 공평한 주사위에 대한 Pearson X^2, (b) 그 근사 p-값을 구하라. ($P(\chi^2(5) \ge 2.5) = 0.776$)

A14 2×2 **표의 독립성** ★ 따름정리 5.2 · 6분

치료(가, 나)와 결과(좋음, 나쁨)의 표가 가: $(30, 20)$, 나: $(20, 30)$이다. (a) 독립성 검정의 X^2, (b) 근사 p-값을 구하라. ($P(\chi^2(1)\geq 4) = 0.0455$)

A15 **동시 신뢰구간의 배수** ★ 명제 8.3 · 6분

평균 20개에 대해 동시에 95% 이상으로 덮는 Bonferroni 구간 $\bar{x}_j \pm c\cdot \mathrm{se}_j$를 만든다(양쪽, 정규근사). c를 구하라. ($\Phi(3.023) = 0.99875$)

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 앞 회차의 정리 번호, 정의, 도구 번호를 쓴다.

B1 Pearson의 χ^2, 근거 채우기 ★★

① · 정리 5.1 · 20분

$H_0: \mathbf{p} = \mathbf{p}^0$이 참이면 $X^2 = \sum_j \frac{(N_j - np_j^0)^2}{np_j^0} \xrightarrow{D} \chi^2(k-1)$이다. 각 단계의 근거를 써라.

① $D = \mathrm{diag}(\mathbf{p}^0)$, $\mathbf{V}_n = D^{-1/2}\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}^0)$로 두면 $X^2 = \|\mathbf{V}_n\|^2$이다.

왜? _____

② $\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}^0) \xrightarrow{D} \mathcal{N}_k(\mathbf{0},\ D - \mathbf{p}^0\mathbf{p}^{0\prime})$이다. 왜? _____

③ $\mathbf{u} = (\sqrt{p_j^0})_j$이면 $\mathbf{V}_n \xrightarrow{D} \mathcal{N}_k(\mathbf{0},\ I - \mathbf{u}\mathbf{u}')$이다. 왜? _____

④ $P = I - \mathbf{u}\mathbf{u}'$는 사영행렬이고 극한은 $P\mathbf{Z}$, $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, I)$의 분포다.

왜? _____

⑤ 첫 열이 $\mathbf{u}$인 직교행렬 Q로 $\mathbf{W} = Q'\mathbf{Z}$를 만들면 $\|P\mathbf{Z}\|^2 = \sum_{i \geq 2} W_i^2 \sim \chi^2(k-1)$이다.

왜? _____

⑥ $X^2 = \|\mathbf{V}_n\|^2 \xrightarrow{D} \|P\mathbf{Z}\|^2$이다. 왜? _____

B2 p-값의 분포, 근거 채우기 ★★

① · 정리 4.2 · 15분

$G(t) = P_{\theta_0}(T \geq t)$, $p(T) = G(T)$이다. 근거를 채워라.

(a) T의 분포함수 F가 연속일 때

① $G(t) = 1 - F(t-) = 1 - F(t)$이다. 왜? _____

② $F(T) \sim \mathrm{U}(0,1)$이다. 왜? _____

③ $p(T) = 1 - F(T) \sim \mathrm{U}(0,1)$이다. 왜? _____

(b) 일반적인 T

① $S_\alpha = \{t : G(t) \leq \alpha\}$는 $t_\alpha = \inf S_\alpha$에 대해 $[t_\alpha, \infty)$ 또는 (t_α, ∞)이다.
왜? _____

② $S_\alpha = [t_\alpha, \infty)$이면 $P(G(T) \leq \alpha) = G(t_\alpha) \leq \alpha$이다. 왜? _____

③ $S_\alpha = (t_\alpha, \infty)$이면 $P(T > t_\alpha) = \lim_n G(t_\alpha + \frac{1}{n}) \leq \alpha$이다. 왜? _____

B3 순서통계량의 결합밀도, 근거 채우기 ★★

① · 정리 1.2 · 15분

밀도 f를 갖는 iid의 순서통계량의 결합밀도가 $n! \prod f(y_i)$ $(y_1 < \cdots < y_n)$임의 증명이다. $A \subset \{y_1 < \cdots < y_n\}$이다. 근거를 채워라.

① $i \neq j$이면 $P(X_i = X_j) = 0$이다. 왜? _____

② $E_\pi = \{X_{\pi(1)} < \cdots < X_{\pi(n)}\}$들은 쌍마다 서로소이고 합집합의 확률이 1이며, E_π 위에서 $X_{(i)} = X_{\pi(i)}$이다. 왜? _____

③ $P\big((X_{(1)}, \ldots, X_{(n)}) \in A\big) = \sum_\pi P\big((X_{\pi(1)}, \ldots, X_{\pi(n)}) \in A\big)$이다.
왜? _____

④ 각 항은 $\int_A \prod_i f(y_i)\, d\mathbf{y}$이다. 왜? _____

⑤ 결합밀도는 $n! \prod_i f(y_i)$이다. 왜? _____

B4 Pearson의 χ^2, 단계의 이름만 ★★

② · 정리 5.1 · 35분

정리 5.1을 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① X^2를 벡터의 길이의 제곱으로 쓴다

② 비율 벡터의 다변량 중심극한정리

③ 표준화한 벡터의 극한과 그 공분산 $I - \mathbf{u}\mathbf{u}'$

④ 공분산이 사영행렬임을 확인하고 극한을 $P\mathbf{Z}$로 쓴다

⑤ 회전해 $\chi^2(k-1)$을 읽는다

⑥ 연속사상으로 마무리한다

B5 순서통계량의 결합밀도, 단계의 이름만 ★★

② · 정리 1.2 · 25분

정리 1.2를 증명하라. 각 단계 아래에 식과 근거를 써라.

① 같은 값이 나올 확률이 0임을 보인다

② **어느 X_i가 몇 번째인가로 표본공간을 나눈다**

③ **순서통계량의 사건을 조각들의 합으로 쓴다**

④ **조각마다 확률이 같음을 보인다**

⑤ **밀도를 읽는다**

B6 Pearson의 χ^2, 전략만 ★★★

③ · 정리 5.1 · 35분

정리 5.1을 증명하라.

전략 X^2를 한 벡터의 길이의 제곱으로 쓴다. 그 벡터의 정규 극한이 $(k-1)$차원 공간으로의 사영임을 보이고, 회전해 $\chi^2(k-1)$을 읽는다.

B7 p**-값의 분포, 전략만** ★★ ③ · 정리 4.2 · 25분

T가 어떤 분포를 따르든 모든 $\alpha \in (0,1)$에서 $P_{\theta_0}(p(T) \le \alpha) \le \alpha$임을 보이고, T의 분포함수가 연속이면 등호임을 보여라.

전략 $p(T) = G(T)$, $G(t) = P(T \ge t)$로 쓴다. $\{G(T) \le \alpha\}$는 T의 위쪽 꼬리 사건이고, 그 확률은 꼬리가 시작하는 점에서의 G 값으로 위에서 막힌다. 연속이면 확률적분변환을 쓴다.

C

증명 문제

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

C1 최솟값과 최댓값의 결합밀도 ★★

[T1] · 정리 1.2 · 20분

$X_1, \dots, X_n$ $(n \ge 2)$이 분포함수 F, 밀도 f를 갖는 iid이다.

(a) $u < v$이면 $P(u < X_{(1)},\ X_{(n)} \le v) = \big(F(v) - F(u)\big)^n$ 임을 보여라.

(b) $P(X_{(1)} \le u,\ X_{(n)} \le v) = F(v)^n - \big(F(v) - F(u)\big)^n$ $(u < v)$을 보여라.

(c) (b)를 u와 v로 미분해 $(X_{(1)}, X_{(n)})$의 결합밀도 $n(n-1)\big(F(v)-F(u)\big)^{n-2} f(u) f(v)$ $(u < v)$를 얻어라.

(d) $\mathrm{U}(0,1)$이면 범위의 기댓값 $E(X_{(n)} - X_{(1)})$은 n의 어떤 식인가?

C2 최댓값의 일치성과 $1/n$의 속도 ★★

[T5], [T8] · 정의 1.1 · 20분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{U}(0,\theta)$, $M_n = X_{(n)}$이다.

(a) 모든 $\varepsilon \in (0,\theta)$에서 $P(\theta - M_n > \varepsilon) = (1-\varepsilon/\theta)^n$을 보이고, M_n이 θ의 일치추정량임을 보여라.

(b) $t > 0$이면 $P\big(n(\theta - M_n) > t\big) \to e^{-t/\theta}$임을 보여라. 곧 $n(\theta - M_n)$은 평균 θ인 지수분포로 분포수렴한다.

(c) 표본평균에 바탕한 $2\bar X$는 $\sqrt{n}(2\bar X - \theta)$가 정규로 수렴한다. 두 추정량의 오차의 크기를 n의 거듭제곱으로 비교하라.

(d) $\theta = 1$, $n = 50$이면 $P(n(\theta - M_n) > 1)$의 정확한 값은 얼마인가?

C3 위치–척도 모형의 축량 ★★

정리 2.2 · 15분

$Z_1, \dots, Z_n$이 알려진 밀도 g를 갖는 iid이고 $X_i = \mu + \sigma Z_i$ $(\sigma > 0)$이다. $\bar X$, S는 표본평균과 표본표준편차다.

(a) $\bar X = \mu + \sigma \bar Z$, $S = \sigma S_Z$임을 보여라.

(b) $T = \dfrac{\bar X - \mu}{S}$의 분포가 μ와 σ에 의존하지 않음을 보여라. 곧 T는 축량이다.

(c) T의 분위수 $a < b$로 μ의 정확한 $1-\alpha$ 구간을 만들어라. g가 정규이면 무엇이 되는가?

(d) S/σ도 σ의 축량임을 보여라.

C4 두 신뢰구간을 함께 ★★

[T1] · 명제 8.3 · 15분

C_1은 θ_1의 $1-\alpha_1$ 신뢰구간, C_2는 θ_2의 $1-\alpha_2$ 신뢰구간이다(같은 자료로 만들어 독립이 아닐 수 있다).

(a) $P(\theta_1 \in C_1,\ \theta_2 \in C_2) \geq 1-\alpha_1-\alpha_2$를 보여라.

(b) 두 구간이 독립이면 동시 포함확률은 얼마인가? (a)의 하한보다 큰지 확인하라.

(c) (a)의 등호가 성립하는 예를 하나 만들어라.

(d) m개의 구간을 각각 $1-\alpha/m$으로 만들면 동시 포함확률이 $1-\alpha$ 이상임을 결론지어라.

C5 양쪽 z 검정은 불편이다 ★★

정리 3.4 · 20분

σ를 아는 정규 평균에서 $H_0 : \mu = \mu_0$를 $|\bar{X} - \mu_0| \geq z_{\alpha/2}\sigma/\sqrt{n}$일 때 기각한다. $d = \sqrt{n}(\mu - \mu_0)/\sigma$, $z = z_{\alpha/2}$로 둔다.

(a) 검정력 함수가 $\gamma(d) = \Phi(d - z) + \Phi(-d - z)$임을 보여라.

(b) γ가 d에 대해 짝함수이고 $\gamma(0) = \alpha$임을 보여라.

(c) $d > 0$이면 $\gamma'(d) = \varphi(d - z) - \varphi(d + z) > 0$임을 보이고, 모든 μ에서 $\gamma \geq \alpha$(불편성)를 결론지어라.

(d) $n = 25$, $\sigma = 10$, $\mu - \mu_0 = 5$, $\alpha = 0.05$이면 검정력은?

C6 복합 귀무가설의 p-값 ★★

정의 4.1, 정리 4.2 · 15분

정의 4.1의 $p(t) = \sup_{\theta \in \Omega_0} P_\theta(T \geq t)$에 대해 다음을 보여라.

(a) 모든 $\theta \in \Omega_0$에서 $p(t) \geq G_\theta(t) = P_\theta(T \geq t)$이다.

(b) $\{p(T) \leq \alpha\} \subset \{G_\theta(T) \leq \alpha\}$이고, 따라서 모든 $\theta \in \Omega_0$에서 $P_\theta(p(T) \leq \alpha) \leq \alpha$이다.

(c) 한쪽 z 검정($H_0 : \mu \leq \mu_0$)에서 상한이 $\mu = \mu_0$에서 이루어짐을 보이고, $\mu < \mu_0$이면 $P_\mu(p \leq \alpha) < \alpha$임을 보여라.

C7 X^2의 정확한 평균 ★

정리 5.1 · 15분

H_0가 참이고 n이 유한할 때 다음을 보여라.

(a) $X^2 = \sum_j \frac{N_j^2}{np_j^0} - n$.

(b) $E(N_j - np_j^0)^2 = np_j^0(1-p_j^0)$이고, 따라서 $EX^2 = k-1$이다.

(c) $\chi^2(k-1)$의 평균과 비교하라. (b)는 근사가 아니라 정확한 등식이다.

C8 Wilson 구간은 $[0,1]$을 벗어나지 않는다 ★★

정리 3.5 · 15분

노트 예제 3.6의 Wilson 구간은 $W = \{p_0 : (\hat{p}-p_0)^2 \le \frac{z^2}{n}p_0(1-p_0)\}$이다.

(a) $W \subset [0,1]$임을 보여라.

(b) Wald 구간은 $[0,1]$을 벗어날 수 있다. $n=20$, $\hat{p}=0.05$에서 Wald 구간의 아래쪽 끝을 구하라($z=1.96$).

(c) Wilson 구간의 중심 $\frac{\hat{p}+z^2/(2n)}{1+z^2/n}$이 $\hat{p}$와 $\frac{1}{2}$ 사이에 있음을 보여라.

C9 Holm의 절차 ★★★

[T1] · 명제 8.3 · 25분

p-값 m개를 정렬해 $p_{(1)} \le \cdots \le p_{(m)}$이라 하자. Holm(1979)의 절차는 $p_{(i)} \le \dfrac{\alpha}{m-i+1}$이 $i = 1, 2, \ldots$에서 성립하는 동안 차례로 기각하고, 처음 성립하지 않는 곳에서 멈춘다. 참인 귀무가설의 첨자 집합을 I_0, $m_0 = |I_0| \ge 1$이라 하자.

(a) 참인 귀무가설 가운데 p-값이 가장 작은 것이 정렬에서 j번째이면 $j \le m - m_0 + 1$임을 보여라.

(b) 참인 귀무가설이 하나라도 기각되면 $\min_{i \in I_0} p_i \le \dfrac{\alpha}{m_0}$임을 보여라.

(c) (b)와 명제 8.3의 논리로 가족별 오류율이 어떤 의존 구조에서도 α 이하임을 보여라.

(d) p-값 0.004, 0.011, 0.02, 0.03, 0.4에 $\alpha = 0.05$의 Holm을 쓰면 몇 개를 기각하는가?

(e) 같은 p-값에 $\alpha = 0.05$의 Bonferroni를 쓰면 몇 개를 기각하는가?

D

퀄형

여러 절과 앞 회차의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 모형과 가정을 적는 것부터 답안의 일부다.

D1 균등분포 상한의 가장 짧은 구간 ★★★

정리 2.2 · 제한 35분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{U}(0, \theta)$, $M = X_{(n)}$, $0 < \alpha < 1$이다.

(a) M/θ가 축량임을 보이고 그 분포함수를 구하라.

(b) $0 \le a < b \le 1$이고 $P(a \le M/\theta \le b) = 1 - \alpha$이면 $[M/b,\ M/a]$가 정확한 $1-\alpha$ 구간임을 보여라. 그 길이를 a, b로 써라.

(c) 제약 $b^n - a^n = 1 - \alpha$ 아래에서 길이를 가장 작게 하는 (a, b)가 $b = 1$, $a = \alpha^{1/n}$ 임을 보여라.

(d) $n = 5$, $M = 4.2$, $\alpha = 0.05$이면 가장 짧은 구간의 위쪽 끝은?

D2 지수 수명의 검정과 구간 ★★★

정리 3.4, 3.5, 4.2 · 4회 정리 3.7 · 제한 40분

수명이 평균 θ인 지수분포를 따르는 부품 n개를 검사한다. $H_0: \theta = \theta_0$ 대 $H_1: \theta < \theta_0$을 $T = 2\sum X_i/\theta_0$로 검정한다.

(a) H_0 아래에서 $T \sim \chi^2(2n)$임을 보이고, $T \le \chi^2_{2n,1-\alpha}$일 때 기각하는 검정의 크기가 α임을 보여라.

(b) 검정력 함수가 $\gamma(\theta) = P\left(\chi^2(2n) \le \frac{\theta_0}{\theta}\chi^2_{2n,1-\alpha}\right)$임을 보여라. θ에 대해 어떻게 변하는가?

(c) 이 검정을 모든 θ_0에 대해 뒤집어 θ의 $1-\alpha$ 위쪽 신뢰한계를 만들어라.

(d) $n = 10$, $\theta_0 = 1000$, $\sum x_i = 6000$, $\alpha = 0.05$이다($\chi^2_{20,0.95} = 10.851$). 기각하는가? p-값($P(\chi^2(20) \le 12) = 0.0839$), $\theta = 500$에서의 검정력($P(\chi^2(20) \le 21.70) = 0.643$), 위쪽 신뢰한계를 구하라.

D3 지수분포의 중앙값 ★★★

정리 1.4, 2.6 · 제한 35분

$X_1, \dots, X_n$이 평균 θ인 지수분포를 따른다. 모집단의 중앙값을 m이라 하자.

(a) $m = \theta \log 2$임을 보여라.

(b) 정리 1.4로 표본중앙값의 점근분산을 구하라. 답을 $c\,\theta^2/n$으로 쓸 때 c는?

(c) 다른 추정량 $(\log 2)\bar{X}$의 분산을 구하고, 표본중앙값의 점근 상대효율 $\dfrac{\mathrm{Var}((\log 2)\bar{X})}{\text{표본중앙값의 점근분산}}$을 구하라.

(d) 분포를 가정하지 않는 대칭 구간 $[X_{(i)}, X_{(101-i)}]$로 $n = 100$에서 신뢰계수 0.95 이상을 얻는 가장 큰 i는? (정리 2.6, 정규근사로 어림한 뒤 확인한다.)

(e) (d)의 구간의 정확한 신뢰계수는? ($K \sim \mathrm{b}(100, \tfrac{1}{2})$에서 $P(K \le 39) = 0.0176$)

D4 Poisson 적합도와 자유도 ★★★

정리 5.1, 5.2절의 상자 · 제한 35분

한 시간 단위 200개에서 사건 수를 세었다. 0회 62, 1회 60, 2회 42, 3회 20, 4회 이상 16이고, 4회 이상인 16개는 모두 정확히 4회였다.

(a) 원자료의 평균으로 λ를 추정하라.

(b) 범주 0, 1, 2, 3, 4 이상의 기대도수를 구하고 Pearson X^2를 계산하라. (기대도수 52.37, 70.17, 47.02, 21.00, 9.44)

(c) 노트 5.2절의 규칙대로 자유도를 $5-1-1=3$으로 읽으면 p-값은? 모수를 추정하지 않은 것처럼 4로 읽으면? ($P(\chi^2(3)\ge 8.39)=0.0386$, $P(\chi^2(4)\ge 8.39)=0.0783$)

(d) λ를 범주로 묶기 전의 원자료로 추정했다. 이때 X^2의 참 극한분포는 어디에 있는가? 5% 수준의 결론에 대해 무엇을 말할 수 있는가?

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 (a) $X_{(2)}$의 밀도는 베타 밀도다. (b)는 밀도를 적분하지 말고 "$\frac{1}{2}$ 이하인 관측이 2개 이상"의 확률로 센다.

답 (a) $\frac{1}{3}$ (b) $\frac{13}{16}$

A2 힌트 $P(X_{(1)} > x) = P(X_1 > x)^4$이고 $P(X_{(4)} \le x) = P(X_1 \le x)^4$이다.

답 (a) $\frac{1}{2}$ (b) $(1-e^{-1})^4 \approx 0.160$

A3 힌트 $\hat{N} = \frac{k+1}{k}M - 1$이다.

답 (a) 74 (b) 79.5

A4 힌트 (b) $M \le 50$은 네 개가 모두 $\{1,\dots,50\}$에서 뽑히는 것이다.

답 (a) 80.8 (b) $\frac{188}{3201} \approx 0.0587$

A5 힌트 $\frac{2}{\theta}\sum X_i \sim \chi^2(2n)$이다. 구간은 $\left[\frac{2\sum X_i}{\chi^2_{2n,\alpha/2}},\ \frac{2\sum X_i}{\chi^2_{2n,1-\alpha/2}}\right]$이다.

답 (a) 39.06 (b) 246.4

A6 힌트 $P(M/\theta \le x) = x^n$이다. $\alpha = 0.1$, $n = 4$.

답 $3 \cdot 10^{1/4} \approx 5.335$

A7 힌트 $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$.

답 (a) 0.1794 (b) 0.2606

A8 힌트 $K \sim \mathrm{b}(15, \frac{1}{2})$에 대해 $P(4 \le K \le 11)$이다. 대칭이므로 양쪽 꼬리 $P(K \le 3)$을 두 번 뺀다.

답 $\frac{247}{256} \approx 0.965$

A9 힌트 (a) $\sum_{k=15}^{20}\binom{20}{k} = 15504+4845+1140+190+20+1$이다. (b)는 같은 합에 $0.7^k 0.3^{20-k}$을 곱한다.

답 (a) $\frac{5425}{262144} \approx 0.0207$ (b) 0.4164

A10 힌트 $n \geq \big((z_\alpha + z_\beta)\sigma/\delta\big)^2$, $\beta = 0.1$이다.

답 138

A11 힌트 $\gamma(\mu) = \Phi\big(\sqrt{n}(\mu - \mu_0)/\sigma - z_\alpha\big)$.

답 0.8038

A12 힌트 양쪽이면 $P(|Z| \geq 2.3)$이다.

답 (a) 0.0214 (b) 0.0107

A13 힌트 기대도수는 모두 20이다. 자유도는 $6 - 1$.

답 (a) $X^2 = 2.5$ (b) 0.776

A14 힌트 기대도수는 모두 25다. 노트 연습 5.3의 공식으로도 확인한다.

답 (a) $X^2 = 4$ (b) 0.0455

A15 힌트 구간마다 놓칠 확률을 0.05/20로 한다. 양쪽 구간이므로 한쪽 꼬리는 그 절반이다.

답 $z_{0.00125} \approx 3.023$

B1 힌트 근거는 짧게 쓴다. 예: "7회 예제 5.3", "연속사상". 넷째 단계는 $\mathcal{N}_k(\mathbf{0}, P)$와 $P\mathbf{Z}$의 공분산을 비교한다.

B2 힌트 (b)의 첫 단계는 "G가 비증가이면 $G(t) \leq \alpha$인 t들의 집합은 위로 열린 반직선"이라는 것이다.

B3 힌트 셋째 단계에서 A의 점은 성분이 증가하므로 $\{(X_{\pi(1)}, \dots, X_{\pi(n)}) \in A\} \subset E_\pi$이다.

B4 힌트 B1의 근거를 다시 보지 말고 쓴다. 셋째 단계에서 $D^{-1/2}\mathbf{p}^0 = \mathbf{u}$를 쓰면 공분산 계산이 한 줄이다.

B5 힌트 조각은 순열 π마다 하나씩, $n!$개다.

B6 힌트 단계는 여섯이다. 증명을 다 쓴 뒤, 자유도가 k가 아니라 $k - 1$인 이유가 증명의 어느 단계에 있는지 표시한다.

B7 힌트 꼬리 사건이 $\{T \geq t_\alpha\}$인 경우와 $\{T > t_\alpha\}$인 경우를 나눈다. 둘째 경우에는 1회의 연속성 정리가 필요하다.

C1 힌트 (a) 모든 관측이 $(u, v]$에 있다는 사건이다. (b) $\{X_{(n)} \leq v\}$에서 (a)의 사건을 뺀다. (d)

는 결합밀도를 쓰지 않고 $EX_{(n)}$과 $EX_{(1)}$을 따로 구해도 된다.

답 (d) $\frac{n-1}{n+1}$

C2 힌트 (b) $(1-\frac{x}{n})^n \to e^{-x}$ [T5]. (c) 오차의 크기가 $1/n$인가 $1/\sqrt{n}$인가.

답 (d) $(1-\frac{1}{50})^{50} \approx 0.3642$

C3 힌트 (b) $T = \bar{Z}/S_Z$는 Z들만의 함수다. (c)는 정리 2.2에 단조성을 확인한다.

C4 힌트 (a) 동시에 덮지 못하는 사건은 두 "놓침" 사건의 합집합이다. (c) 두 놓침 사건이 서로소이면 Boole 부등식이 등호다.

C5 힌트 (c) $d > 0$이면 $|d - z| < d + z$이고 φ는 $|x|$에 대해 감소한다.

답 (d) 0.7054

C6 힌트 (b)의 둘째 부분은 정리 4.2(b)를 P_θ에 쓴다. 그 증명은 T의 분포에 아무 조건도 쓰지 않았다.

C7 힌트 (b) N_j의 주변분포는 이항분포 $\mathrm{b}(n, p_j^0)$이다(4회).

C8 힌트 (a) $p_0 < 0$이나 $p_0 > 1$이면 오른쪽이 음수다. (c) 중심은 $\hat{p}$와 $\frac{1}{2}$의 가중평균이다.

답 (b) -0.0455

C9 힌트 (b) 참인 귀무가설이 처음 기각되는 단계는 j번째 단계다. 그때 $p_{(j)} \le \alpha/(m-j+1)$이고 (a)를 쓴다.

답 (d) 2 (e) 1

D1 힌트 (c) a를 b의 함수 $a(b) = (b^n - (1-\alpha))^{1/n}$로 보고 $L(b) = \frac{1}{a(b)} - \frac{1}{b}$를 미분한다. $a'(b) = b^{n-1}/a^{n-1}$이다.

답 (d) $4.2 \cdot 0.05^{-1/5} \approx 7.646$

D2 힌트 (a) X_i/θ는 $\Gamma(1,1)$이고 4회 정리 3.7로 합은 $\Gamma(n,1)$이다. (c) 수락역 $\{2\sum x_i/\theta_0 > c\}$를 θ_0에 대해 푼다. p-값은 작은 T가 H_0에 불리하다는 것에 맞춰 $P(T \le t)$이다.

답 (a) 0.0839 (b) 0.643 (c) 1106

D3 힌트 (b) $f(m) = \frac{1}{\theta}e^{-m/\theta} = \frac{1}{2\theta}$이다. (d) $i \approx 50 - 1.96 \cdot 5$. (e) $P(40 \le K \le 60) = 1 - 2P(K \le 39)$.

답 (b) θ^2/n (c) $(\log 2)^2 \approx 0.480$ (d) $i = 40$ (e) 0.9648

D4 힌트 (a) 사건의 총수는 $0\cdot 62+1\cdot 60+2\cdot 42+3\cdot 20+4\cdot 16$이다. (d) 노트 5.2절 상자의 마지막 문장(Chernoff–Lehmann).

답 (a) $\hat{\lambda}=1.34$ (b) $X^2=8.39$ (c) 0.0386 (d) 0.0783