수리통계학 × 통계학의 역사

6회 워크북 — 풀이판

Student의 t 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	표본분산, χ^2, t, F, 구간을 손으로 계산한다	14	1시간 52분
B 빈칸 증명	노트의 단계형 증명을 책을 덮고 다시 쓴다	7	2시간 50분
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 `06_student_t_notes.pdf` — 문항의 "정리 2.2" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T2] 귀류·반례, [T4] 부등식, [T6] 극한의 교환, [T7] 변수변환, [T8] O 표기

검산 노트북 `06_student_t.ipynb`의 "워크북 검산" 절: `check("A3", "5*exp(-4)")`

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. 분위수표의 값으로 계산한 답은 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

표본분산, χ^2, t, F와 구간을 손으로 계산한다. 분위수가 필요한 문항은 문제에 값을 적어 두었다. 답이 수나 식으로 떨어지므로 다 풀면 뒤의 최종답이나 노트북의 `check()`로 맞춰 본다.

A1 표본분산을 세 가지로 ★

정리 1.1, 연습 1.1 · 6분

자료 3, 7, 5, 9, 6의 표본분산 s^2를 (i) 편차제곱합, (ii) $\sum x_i^2 - n\bar{x}^2$, (iii) $\frac{1}{2n}\sum_{i,j}(x_i - x_j)^2$의 세 가지로 계산해 같은 값이 나오는지 확인하라.

풀이 $\bar{x} = 6$이고 편차는 $-3, 1, -1, 3, 0$이므로 (i) 편차제곱합은 $9 + 1 + 1 + 9 + 0 = 20$이다. (ii) $\sum x_i^2 = 9 + 49 + 25 + 81 + 36 = 200$이고 $200 - 5 \cdot 36 = 20$이다. (iii) $i < j$인 쌍의 차는 $4, 2, 6, 3, 2, 2, 1, 4, 1, 3$이고 제곱의 합은 $16 + 4 + 36 + 9 + 4 + 4 + 1 + 16 + 1 + 9 = 100$이다. $\sum_{i,j} = 2 \cdot 100 = 200$이므로 $\frac{1}{10} \cdot 200 = 20$이다. 세 경우 모두 $s^2 = 20/4 = 5$이다.

최종답 $s^2 = 5$

A2 $n = 4$의 Helmert 회전 ★

보조정리 2.1, 정리 2.2 · 8분

$n = 4$의 Helmert 행렬 H의 네 번째 행을 쓰고, $\mathbf{z} = (1, 2, 3, 6)'$에 대해 $\mathbf{Y} = H\mathbf{z}$의 성분 $Y_1, \dots, Y_4$를 구하라. (a) Y_4는 얼마인가? (b) $Y_2^2 + Y_3^2 + Y_4^2$을 구하고 $\sum_i (z_i - \bar{z})^2$과 같음을 확인하라.

풀이 네 번째 행은 $\frac{1}{\sqrt{12}}(1, 1, 1, -3)$이다. $Y_1 = \frac{12}{2} = 6$, $Y_2 = \frac{1-2}{\sqrt{2}} = -\frac{1}{\sqrt{2}}$, $Y_3 = \frac{1+2-6}{\sqrt{6}} = -\frac{3}{\sqrt{6}}$, $Y_4 = \frac{1+2+3-18}{\sqrt{12}} = -\frac{12}{\sqrt{12}} = -2\sqrt{3} \approx -3.464$이다. 제곱의 합은 $\frac{1}{2} + \frac{3}{2} + 12 = 14$이고, $\bar{z} = 3$이므로 $\sum (z_i - \bar{z})^2 = 4 + 1 + 0 + 9 = 14$로 같다. $Y_1^2 = 36$을 더하면 $\|\mathbf{z}\|^2 = 50$이다.

최종답 (a) $-2\sqrt{3} \approx -3.464$ (b) 14

A3 표본분산이 두 배 넘게 나올 확률 ★

정리 2.2(c) · 8분

$X_1, \dots, X_5 \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 4)$이다. $P(S^2 > 8)$을 정확히 구하라. (힌트: $\chi^2(4)$는 모양모수 2인 감마분포다.)

풀이 정리 2.2(c)로 $\frac{(n-1)S^2}{\sigma^2} = \frac{4S^2}{4} = S^2 \sim \chi^2(4)$이다. $\chi^2(4)$의 밀도는 $\frac{xe^{-x/2}}{\Gamma(2)2^2} = \frac{1}{4}xe^{-x/2}$이고, 부분적분으로 $P(\chi^2(4) > c) = e^{-c/2}\left(1 + \frac{c}{2}\right)$이다. $c = 8$이면 $5e^{-4} \approx 0.0916$이다. 참분산의 두

배가 넘는 표본분산이 열에 한 번쯤 나온다.

최종답 $5e^{-4} \approx 0.0916$

A4 표본분산의 흔들림 ★

따름정리 2.3 · 6분

$X_1, \dots, X_9 \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 2)$이다. (a) $\operatorname{Var} S^2$를 구하라. (b) S^2의 상대 표준편차 $\sqrt{\operatorname{Var} S^2}/ES^2$를 구하라.

풀이 (a) $\frac{2 \cdot 2^2}{8} = 1$. (b) $ES^2 = 2$이므로 $\frac{1}{2}$이다. 일반적으로 $\sqrt{2/(n-1)}$이고 $n = 9$이면 $\frac{1}{2}$이다. 아홉 관측으로 잰 분산도 참값의 ±50%가 한 표준편차다.

최종답 (a) 1 (b) $\frac{1}{2}$

A5 세 관측의 표본표준편차 ★

연습 2.4 · 8분

$n = 3$인 정규 표본에서 $ES = c_4(3)\,\sigma$의 $c_4(3)$을 감마함수로 구하고 소수로 나타내라.

풀이 $c_4(3) = \sqrt{\frac{2}{2}}\,\frac{\Gamma(3/2)}{\Gamma(1)} = \frac{\sqrt{\pi}}{2} \approx 0.886$이다. 세 관측의 S는 σ를 평균 11% 작게 어림한다.

최종답 $\frac{\sqrt{\pi}}{2} \approx 0.886$

A6 $t(3)$의 봉우리 ★

정리 3.2 · 6분

$t(3)$의 밀도의 0에서의 값 $f_3(0)$을 정확히 구하고, 표준정규의 $\phi(0) = \frac{1}{\sqrt{2\pi}} \approx 0.399$와 비교하라.

풀이 $f_3(0) = \frac{\Gamma(2)}{\sqrt{3\pi}\,\Gamma(3/2)} = \frac{1}{\sqrt{3\pi}\cdot\sqrt{\pi}/2} = \frac{2}{\sqrt{3}\,\pi} \approx 0.368$이다. 정규보다 낮다. 밀도의 넓이가 같으므로 가운데가 낮은 만큼 꼬리가 두껍다.

최종답 $\frac{2}{\sqrt{3}\,\pi} \approx 0.368$

A7 자유도 1의 t ★

정리 3.3(c) · 8분

$T \sim t(1)$이다. (a) $P(T > \sqrt{3})$, (b) $P(|T| < 1)$을 정확히 구하라.

풀이 정리 3.3(c)로 밀도는 $\frac{1}{\pi(1+t^2)}$이고 분포함수는 $F(t) = \frac{1}{2} + \frac{1}{\pi}\arctan t$이다. (a) $1 - F(\sqrt{3}) = \frac{1}{2} - \frac{1}{\pi}\cdot\frac{\pi}{3} = \frac{1}{6}$. (b) $F(1) - F(-1) = \frac{2}{\pi}\cdot\frac{\pi}{4} = \frac{1}{2}$.

최종답 (a) $\frac{1}{6}$ (b) $\frac{1}{2}$

A8 t와 F의 적률 ★

정리 3.3(b), 4.2(b)(c) · 6분

(a) $T \sim t(6)$의 분산을 구하라. (b) $F \sim F(3,8)$의 평균을 구하라. (c) T^2의 분포를 쓰고 (a)와 ET^2이 맞는지 확인하라.

풀이 (a) $\frac{6}{6-2} = \frac{3}{2}$. (b) $\frac{8}{8-2} = \frac{4}{3}$. (c) $T^2 \sim F(1,6)$이고 $EF(1,6) = \frac{6}{4} = \frac{3}{2} = ET^2$로 (a)와 같다 ($ET = 0$이므로 $\operatorname{Var} T = ET^2$).

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

A9 다섯 관측의 t-구간 ★

정리 5.2 · 10분

정규 모집단에서 10.2, 9.6, 10.8, 10.1, 9.8을 얻었다. $t_{4,\,0.025} = 2.776$을 써서 μ의 95% 신뢰구간의 (a) 아래 끝과 (b) 위 끝을 구하라.

풀이 $\bar{x} = 10.1$, 편차는 $0.1, -0.5, 0.7, 0, -0.3$이고 편차제곱합은 0.84이므로 $s^2 = 0.21$, $s = 0.4583$이다. 반폭은 $2.776 \cdot 0.4583/\sqrt{5} = 0.569$이므로 구간은 $[9.531,\ 10.669]$이다.

최종답 (a) 9.531 (b) 10.669

A10 대응 t 통계량 ★

정리 3.4, 예제 5.3 · 10분

환자 6명의 처리 전후 차이가 2, 4, −1, 3, 5, 1이다. (a) 대응 t 통계량 $t = \bar{d}/(s_d/\sqrt{n})$을 정확한 수로 구하라. (b) 귀무가설 $\mu_D = 0$ 아래의 자유도는?

풀이 $\bar{d} = \frac{14}{6} = \frac{7}{3}$이고 편차제곱합은 $\sum d_i^2 - n\bar{d}^2 = 56 - 6 \cdot \frac{49}{9} = \frac{70}{3}$이므로 $s_d^2 = \frac{14}{3}$이다. $s_d^2/n = \frac{7}{9}$이므로 $t = \frac{7/3}{\sqrt{7/3}} = \sqrt{7} \approx 2.646$이다. (b) 자유도는 $n - 1 = 5$이다. $t_{5,0.025} = 2.571$보다 크므로 양쪽 5% 수준에서 차이가 있다.

최종답 (a) $\sqrt{7} \approx 2.646$ (b) 5

A11 합동분산의 두 표본 t ★

정리 5.4 · 10분

X: 5, 7, 6과 Y: 3, 4, 2, 3이 등분산 정규 모집단의 독립 표본이다. (a) $\mu_1 = \mu_2$ 아래의 통계량 T를 정확한 수로 구하라. (b) 자유도는?

풀이 $\bar{x} = 6$, $\bar{y} = 3$이고 편차제곱합은 각각 2와 2이므로 $S_p^2 = \frac{4}{5}$이다. $S_p^2(\frac{1}{3} + \frac{1}{4}) = \frac{4}{5} \cdot \frac{7}{12} = \frac{7}{15}$이므로 $T = \frac{3}{\sqrt{7/15}} = 3\sqrt{15/7} \approx 4.392$이다. (b) $m + n - 2 = 5$.

최종답 (a) $3\sqrt{15/7} \approx 4.392$ (b) 5

A12 분산비의 구간 ★

정리 4.3 · 10분

$m=7$, $n=9$인 독립 정규 표본에서 $s_X^2=6$, $s_Y^2=2$를 얻었다. $F_{6,8,\,0.025}=4.652$, $F_{8,6,\,0.025}=5.600$을 써서 σ_1^2/σ_2^2의 95% 신뢰구간의 (a) 아래 끝과 (b) 위 끝을 구하라.

풀이 노트 연습 4.3의 구간 $\left[\dfrac{s_X^2/s_Y^2}{F_{m-1,n-1,0.025}},\ \dfrac{s_X^2/s_Y^2}{F_{m-1,n-1,0.975}}\right]$에 비 3을 넣는다. $F_{6,8,0.975}=1/5.600$이므로 구간은 $[3/4.652,\ 3\cdot 5.600]=[0.645,\ 16.80]$이다. 1을 포함한다.

최종답 (a) 0.645 (b) 16.80

A13 다음 한 관측의 구간 ★

정리 8.1 · 8분

A9의 자료로 다음 관측 하나에 대한 95% 예측구간의 (a) 아래 끝과 (b) 위 끝을 구하라. A9의 신뢰구간보다 몇 배 넓은가?

풀이 반폭은 $2.776\cdot 0.4583\cdot\sqrt{1.2}=1.394$이므로 구간은 $[8.706,\ 11.494]$이다. 신뢰구간의 반폭과의 비는 $\sqrt{1.2}\cdot\sqrt{5}=\sqrt{6}\approx 2.45$배다.

최종답 (a) 8.706 (b) 11.494

A14 세 관측에 1.96을 쓰면 ★★

정리 3.4, 노트 연습 3.3 · 8분

$n=3$인 정규 표본에서 $\bar X \pm 1.96\,S/\sqrt{3}$의 실제 포함확률을 정확한 식으로 구하라. $t(2)$의 분포함수 $F_2(t)=\frac{1}{2}\left(1+\frac{t}{\sqrt{2+t^2}}\right)$를 써도 된다.

풀이 포함되는 사건은 $|T|\le 1.96$, $T\sim t(2)$이다. $P(|T|\le c)=F_2(c)-F_2(-c)=\dfrac{c}{\sqrt{2+c^2}}$이므로 $\dfrac{1.96}{\sqrt{5.8416}}\approx 0.811$이다. “95%” 구간이 다섯 번에 한 번 가까이 틀린다.

최종답 $\dfrac{1.96}{\sqrt{2+1.96^2}}\approx 0.811$

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 Student의 정리, 근거 채우기 ★★

① · 정리 2.2 · 15분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이면 $\bar{X} \perp S^2$, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$이다. 각 단계의 근거를 써라.

① $Z_i = (X_i - \mu)/\sigma$로 두면 $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, I)$이고 $\mathbf{Z}$에 대해 보이면 된다. — 표준화

② $\mathbf{Y} = H\mathbf{Z}$이면 $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, I)$이다. — 5회 정리 3.8(정리 A), $HH' = I$

③ $Y_1, \dots, Y_n$은 독립인 표준정규다. — 공분산행렬이 대각, 5회 따름정리 3.12

④ $Y_1 = \sqrt{n}\,\bar{Z}$이다. — H의 첫 행이 $\mathbf{1}/\sqrt{n}$

⑤ $\sum_{i \ge 2} Y_i^2 = \sum_i (Z_i - \bar{Z})^2$이다. — $\|H\mathbf{Z}\| = \|\mathbf{Z}\|$와 ④

⑥ $\bar{Z} \perp \sum_i (Z_i - \bar{Z})^2$이다. — 서로 다른 Y들의 함수, ③

⑦ $\bar{Z} \sim \mathcal{N}(0, \frac{1}{n})$이다. — $\bar{Z} = Y_1/\sqrt{n}$, ③

⑧ $\sum_i (Z_i - \bar{Z})^2 \sim \chi^2(n-1)$이다. — 표준정규 $n-1$개의 제곱합, 4회 정리 3.9

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 2.2의 단계형 증명과 같다. 다섯째 단계는 $\sum_i Y_i^2 = \|H\mathbf{Z}\|^2 = \|\mathbf{Z}\|^2$에서 $Y_1^2 = n\bar{Z}^2$을 빼고 $\sum_i Z_i^2 - n\bar{Z}^2 = \sum_i (Z_i - \bar{Z})^2$을 쓴다.

B2 t의 밀도, 근거 채우기 ★★

① · 정리 3.2 · [T7] · 20분

$Z \sim \mathcal{N}(0,1)$과 $V \sim \chi^2(r)$이 독립이면 $T = Z/\sqrt{V/r}$의 밀도는 다음과 같다. 근거를 채워라.

$$f_r(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)}\left(1+\frac{t^2}{r}\right)^{-\frac{r+1}{2}}$$

① (Z, V)의 결합밀도는 $\phi(z) \cdot \frac{v^{r/2-1}e^{-v/2}}{\Gamma(r/2)2^{r/2}}$이다. — 독립, 4회 정의 3.8

② 보조 변수 $W = V$를 두면 역변환은 $z = t\sqrt{w/r}$, $v = w$이다. — 일대일 변환 [T7]

③ 야코비안은 $\sqrt{w/r}$이다. — 위삼각 행렬의 행렬식

④ $f_{T,W}(t,w) = \phi\left(t\sqrt{w/r}\right) \cdot \frac{w^{r/2-1}e^{-w/2}}{\Gamma(r/2)2^{r/2}} \cdot \sqrt{w/r}$이다. — 3회 정리 5.1

⑤ $f_T(t) = \int_0^\infty f_{T,W}(t,w)\,dw$이다. — 주변밀도, 피적분함수 ≥ 0 (Tonelli) [T6]

⑥ $\int_0^\infty w^{\frac{r+1}{2}-1}e^{-w\frac{1+t^2/r}{2}}dw = \Gamma\left(\frac{r+1}{2}\right)\left(\frac{2}{1+t^2/r}\right)^{\frac{r+1}{2}}$이다. — 감마 적분, 4회 정의 3.4

풀이 빈칸의 빨간 글씨가 근거다. 마지막으로 상수를 모으면 $\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\Gamma(r/2)2^{r/2}} \cdot \frac{1}{\sqrt{r}} \cdot 2^{(r+1)/2}\Gamma\left(\frac{r+1}{2}\right) = \frac{\Gamma((r+1)/2)}{\sqrt{r\pi}\,\Gamma(r/2)}$이다.

B3 두 표본 t, 근거 채우기 ★★

① · 정리 5.4 · 15분

등분산 정규 표본 $X_1, \ldots, X_m$과 $Y_1, \ldots, Y_n$이 독립이면 $T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/m+1/n}} \sim t(m+n-2)$이다. 근거를 채워라.

① $Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{1/m+1/n}} \sim \mathcal{N}(0,1)$이다. — 정리 2.2(a), 5회 정리 1.8(iv)

② $\frac{(m-1)S_X^2}{\sigma^2}$와 $\frac{(n-1)S_Y^2}{\sigma^2}$는 독립인 $\chi^2(m-1)$, $\chi^2(n-1)$이다. — 정리 2.2(c), 두 표본의 독립

③ $V = \frac{(m+n-2)S_p^2}{\sigma^2} \sim \chi^2(m+n-2)$이다. — 독립인 χ^2의 합, 4회 정리 3.7

④ $Z \perp V$이다. — 정리 2.2(b)를 두 표본에, 함수의 독립

⑤ $T = Z/\sqrt{V/(m+n-2)} \sim t(m+n-2)$이다. — σ 소거, 정의 3.1

풀이 빈칸의 빨간 글씨가 근거다. 셋째 단계에서 두 χ^2를 같은 σ^2로 나눴기 때문에 합이 $\frac{(m+n-2)S_p^2}{\sigma^2}$이 된다. 등분산이 아니면 합동분산은 하나의 σ^2를 추정하지 않는다.

B4 Student의 정리, 단계의 이름만 ★★

② · 정리 2.2 · 30분

정리 2.2의 (a)(b)(c)를 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **표준화해 $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, I)$로 줄인다**

② **Helmert 행렬로 돌린다**

③ **돌린 좌표가 독립인 표준정규임을 보인다**

④ **첫 성분이 평균임을 보인다**

⑤ **나머지 성분의 제곱합이 편차제곱합임을 보인다**

⑥ **(b): 독립을 결론짓는다**

⑦ **(a)와 (c)를 결론짓는다**

풀이

① $Z_i = (X_i - \mu)/\sigma$, $\bar{X} = \mu + \sigma\bar{Z}$, $(n-1)S^2 = \sigma^2 \sum (Z_i - \bar{Z})^2$. — 표준화

② $\mathbf{Y} = H\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, HH') = \mathcal{N}_n(\mathbf{0}, I)$. — 5회 정리 3.8

③ 공분산행렬이 대각이므로 Y_i들은 독립인 $\mathcal{N}(0,1)$. — 5회 따름정리 3.12

④ $Y_1 = \frac{1}{\sqrt{n}} \sum Z_i = \sqrt{n}\bar{Z}$. — H의 첫 행

⑤ $\sum_{i \geq 2} Y_i^2 = \|\mathbf{Z}\|^2 - n\bar{Z}^2 = \sum (Z_i - \bar{Z})^2$. — $\|H\mathbf{Z}\| = \|\mathbf{Z}\|$, 정리 1.1

⑥ $\bar{Z}$는 Y_1의, 편차제곱합은 $Y_2, \ldots, Y_n$의 함수이므로 독립. — 3회 3.2절의 상자

⑦ $\bar{Z} \sim \mathcal{N}(0, \frac{1}{n})$이고 $\sum_{i \geq 2} Y_i^2 \sim \chi^2(n-1)$. — 4회 정리 3.9

B5 t의 밀도, 단계의 이름만 ★★

② · 정리 3.2 · [T7] · 25분

정리 3.2를 증명하라. 단계의 이름만 주어진다.

① **결합밀도를 쓴다**

② **보조 변수를 두고 역변환을 구한다**

③ **야코비안을 구한다**

④ **(T, W)의 결합밀도를 쓴다**

⑤ **보조 변수를 적분한다(감마 적분)**

⑥ **상수를 모은다**

풀이 B2의 단계에 다음 식을 채우면 된다. 결합밀도는

$$f_{T,W}(t,w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2 w}{2r}} \cdot \frac{w^{r/2-1} e^{-w/2}}{\Gamma(r/2) 2^{r/2}} \cdot \sqrt{\frac{w}{r}}$$

이다. w의 거듭제곱은 $w^{(r+1)/2-1}$이고 지수는 $e^{-w(1+t^2/r)/2}$이다. 감마 적분의 공식

$$\int_0^\infty w^{a-1}e^{-bw}\,dw = \Gamma(a)\,b^{-a}, \qquad a=\frac{r+1}{2},\quad b=\frac{1+t^2/r}{2}$$

을 쓰면

$$\int_0^\infty w^{\frac{r+1}{2}-1}e^{-w\frac{1+t^2/r}{2}}\,dw = \Gamma\!\left(\frac{r+1}{2}\right)2^{\frac{r+1}{2}}\left(1+\frac{t^2}{r}\right)^{-\frac{r+1}{2}}$$

이다. 상수는 $\dfrac{2^{(r+1)/2}}{\sqrt{2\pi}\,2^{r/2}\sqrt{r}\,\Gamma(r/2)} = \dfrac{1}{\sqrt{r\pi}\,\Gamma(r/2)}$이다.

B6 Student의 정리, 전략만 ★★

③ · 정리 2.2 · 35분

$X_1,\dots,X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma^2)$이면 $\bar X \sim \mathcal{N}(\mu,\sigma^2/n)$, $\bar X \perp S^2$, $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$임을 증명하라.

전략 표본을 표준화하고 직교행렬로 돌린다. 돌린 좌표에서 첫 성분은 평균이고, 나머지 성분의 제곱합은 편차제곱합이며, 좌표들은 서로 독립인 표준정규다.

풀이 B4의 풀이와 같다. 정규성은 두 곳에서 쓰인다. 직교변환한 벡터가 다시 결합정규라는 것(5회 정리 A)과, 결합정규에서는 공분산행렬이 대각이면 성분이 독립이라는 것(5회 정리 B)이다. 지수 표본에서는 둘째가 깨져서 $\bar X$와 S^2가 의존한다(노트 연습 2.3).

B7 두 표본 t, 전략만 ★★

③ · 정리 5.4 · 30분

정리 5.4를 증명하라.

전략 표준정규를, 그와 독립인 χ^2를 자유도로 나눈 것의 제곱근으로 나눈 꼴을 만든다. 두 표본에 Student의 정리를 따로 쓰고 독립을 합친다.

풀이 B3의 단계와 같다. 네 변수 $\bar X, S_X^2, \bar Y, S_Y^2$가 상호 독립이라는 것이 넷째 단계의 내용이다. 표본 안에서는 정리 2.2(b)가, 표본 사이에서는 두 표본의 독립이 독립을 준다. 그러면 $(\bar X,\bar Y)$의 함수인 Z와 (S_X^2,S_Y^2)의 함수인 V가 독립이다.

C

증명 문제

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

C1 Helmert가 아니어도 된다 ★★

보조정리 2.1, 정리 2.2 · 20분

H가 첫 행이 $\mathbf{1}'/\sqrt{n}$인 임의의 $n\times n$ 직교행렬이고 $\mathbf{Y}=H\mathbf{Z}$이다.

(a) H의 둘째 행부터 마지막 행까지는 모두 $\mathbf{1}$과 직교함을 보여라.

(b) $\sum_{i\ge 2} Y_i^2 = \sum_i (Z_i-\bar{Z})^2$ 임을 보여라.

(c) 정리 2.2의 증명이 이 H로도 그대로 됨을 밝혀라.

(d) $n=3$에서 Helmert 행렬과 다른 예를 하나 만들어라.

풀이 (a) $HH'=I$이므로 행들은 정규직교이고, 첫 행이 $\mathbf{1}/\sqrt{n}$이므로 나머지 행은 $\mathbf{1}$과 직교한다.

(b) $Y_1=\mathbf{1}'\mathbf{Z}/\sqrt{n}=\sqrt{n}\bar{Z}$이고 $H'H=I$이므로 $\sum_i Y_i^2=\|\mathbf{Z}\|^2$이다. 따라서 $\sum_{i\ge2}Y_i^2=\|\mathbf{Z}\|^2-n\bar{Z}^2=\sum_i(Z_i-\bar{Z})^2$이다.

(c) 증명에 쓴 것은 H가 직교라는 것(5회 정리 A, B)과 첫 행이 $\mathbf{1}/\sqrt{n}$이라는 것(④ ⑤)뿐이다. Helmert 행렬의 나머지 행의 모양은 쓰지 않았다.

(d) Helmert의 둘째, 셋째 행 $u=\frac{1}{\sqrt{2}}(1,-1,0)$, $v=\frac{1}{\sqrt{6}}(1,1,-2)$를 각도 θ만큼 돌린 $\cos\theta\,u+\sin\theta\,v$, $-\sin\theta\,u+\cos\theta\,v$로 바꾼다. $\theta=\frac{\pi}{2}$이면 행이 v, $-u$이고, 이것도 직교행렬이다. 일반적으로 $\mathbf{1}$에 수직인 평면의 정규직교기저는 무엇이든 된다.

C2 μ를 알 때의 잘못된 t ★★

[T2], [T4] · 정리 2.2, 3.4 · 20분

$X_1,\dots,X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma^2)$이고 μ를 안다. σ^2을 $\hat{\sigma}_0^2=\frac{1}{n}\sum_i(X_i-\mu)^2$로 어림해 $W=\frac{\sqrt{n}(\bar{X}-\mu)}{\hat{\sigma}_0}$를 만들었다.

(a) Cauchy–Schwarz로 $|W|\le\sqrt{n}$임을 보이고, W가 $t(n)$을 따를 수 없음을 보여라.

(b) 정리 3.4의 증명의 어느 단계가 W에서는 성립하지 않는가?

(c) $\frac{W^2}{n}\sim\text{Beta}\left(\frac{1}{2},\frac{n-1}{2}\right)$임을 보여라.

풀이 (a) Cauchy–Schwarz [T4]로 $\left(\sum_i(X_i-\mu)\right)^2\le n\sum_i(X_i-\mu)^2$, 곧 $n^2(\bar{X}-\mu)^2\le n\cdot n\hat{\sigma}_0^2$이다. 따라서 $|\bar{X}-\mu|\le\hat{\sigma}_0$이고 $|W|\le\sqrt{n}$이다. $t(n)$의 밀도는 모든 실수에서 양수이므로

$P(|T|>\sqrt{n})>0$인데 $P(|W|>\sqrt{n})=0$이다. 두 분포는 같을 수 없다 [T2].

(b) 분자와 분모의 독립(④)이 깨진다. $n\hat{\sigma}_0^2=\sum(X_i-\bar{X})^2+n(\bar{X}-\mu)^2$은 분자의 제곱을 그 안에 담고 있다.

(c) 정리 2.2의 증명처럼 표준화하고 돌리면 $\frac{n\hat{\sigma}_0^2}{\sigma^2}=Y_1^2+\sum_{i\ge 2}Y_i^2$이고 $W^2/n=\frac{Y_1^2}{Y_1^2+\sum_{i\ge 2}Y_i^2}$이다. $Y_1^2/2\sim\Gamma(\frac{1}{2},1)$과 $\sum_{i\ge 2}Y_i^2/2\sim\Gamma(\frac{n-1}{2},1)$이 독립이므로 4회 정리 4.2로 $\text{Beta}(\frac{1}{2},\frac{n-1}{2})$이다. μ를 알 때의 올바른 축량은 $\frac{\sqrt{n}(\bar{X}-\mu)}{S}\sim t(n-1)$ 그대로다.

C3 분산의 비를 알 때의 두 표본 t ★★

정리 5.4 · 15분

$X_1,\dots,X_m\overset{\text{iid}}{\sim}\mathcal{N}(\mu_1,\sigma^2)$, $Y_1,\dots,Y_n\overset{\text{iid}}{\sim}\mathcal{N}(\mu_2,\lambda\sigma^2)$가 독립이고 $\lambda>0$은 안다. $t(m+n-2)$를 따르는 $\mu_1-\mu_2$의 축량을 만들고, 95% 신뢰구간을 써라. $\lambda=1$이면 정리 5.4가 됨을 확인하라.

풀이 $\bar{X}-\bar{Y}\sim\mathcal{N}\big(\mu_1-\mu_2,\ \sigma^2(\frac{1}{m}+\frac{\lambda}{n})\big)$이므로 $Z=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{1/m+\lambda/n}}\sim\mathcal{N}(0,1)$이다. $V=\frac{(m-1)S_X^2}{\sigma^2}+\frac{(n-1)S_Y^2}{\lambda\sigma^2}\sim\chi^2(m+n-2)$(정리 2.2(c), 4회 정리 3.7)이고 정리 5.4의 증명과 같은 이유로 $Z\perp V$이다. $S_\lambda^2=\frac{(m-1)S_X^2+(n-1)S_Y^2/\lambda}{m+n-2}$로 두면

$$T=\frac{Z}{\sqrt{V/(m+n-2)}}=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\lambda\sqrt{1/m+\lambda/n}}\sim t(m+n-2)$$

이고 σ가 소거된다. 구간은 $\bar{X}-\bar{Y}\pm t_{m+n-2,0.025}\,S_\lambda\sqrt{1/m+\lambda/n}$이다. $\lambda=1$이면 $S_\lambda=S_p$이다.

C4 t-구간이 z-구간보다 짧을 확률 ★★

정리 2.2(c) · 20분

$n=4$인 정규 표본에서 t-구간 $\bar{X}\pm t_{3,0.025}S/2$와 ($\sigma$를 알 때의) z-구간 $\bar{X}\pm 1.96\,\sigma/2$를 비교한다. $t_{3,0.025}=3.182$이다.

(a) t-구간이 더 짧은 사건이 $\{\chi^2(3)<c\}$ 꼴임을 보이고 c를 구하라.

(b) $\chi^2(3)$의 분포함수가 $2\Phi(\sqrt{x})-1-\sqrt{2x/\pi}\,e^{-x/2}$임을 보이고, (a)의 확률을 구하라.

풀이 (a) $3.182\,S<1.96\,\sigma\iff\frac{3S^2}{\sigma^2}<3\big(\frac{1.96}{3.182}\big)^2=1.138$이고 $\frac{3S^2}{\sigma^2}\sim\chi^2(3)$(정리 2.2(c))이다. $c=1.138$.

(b) $\chi^2(3)$의 밀도는 $\frac{x^{1/2}e^{-x/2}}{\Gamma(3/2)2^{3/2}}=\frac{\sqrt{x}e^{-x/2}}{\sqrt{2\pi}}$이다. $G(x)=2\Phi(\sqrt{x})-1-\sqrt{2x/\pi}\,e^{-x/2}$를 미분하면 $\frac{\phi(\sqrt{x})}{\sqrt{x}}-\frac{e^{-x/2}}{\sqrt{2\pi x}}+\sqrt{\frac{x}{2\pi}}e^{-x/2}=\frac{\sqrt{x}e^{-x/2}}{\sqrt{2\pi}}$이고 $G(0)=0$이므로 G가 분포함수다. $\sqrt{c}=1.0668$이면 $2\Phi(1.0668)-1=0.7140$이고 $\sqrt{2c/\pi}\,e^{-c/2}=0.4819$이므로 확률은 0.232다. 네 번 가운데 한 번 가까이 t-구간이 σ를 알 때보다도 짧고, 그런 때 틀리기 쉽다.

최종답 (a) $c\approx 1.138$ (b) 0.232

C5 무상관이지만 독립이 아니다 ★★

[T2] · 명제 7.1 · 15분

$X_1, X_2 \overset{\text{iid}}{\sim} \mathrm{U}(-1,1)$이고 $\bar{X} = \frac{X_1+X_2}{2}$, $S^2 = \frac{(X_1-X_2)^2}{2}$이다.

(a) 명제 7.1로 $\mathrm{Cov}(\bar{X}, S^2) = 0$임을 보여라.

(b) $P(|\bar{X}| > \frac{1}{2})$와 $P(S^2 > \frac{1}{2})$를 구하라.

(c) $P(|\bar{X}| > \frac{1}{2},\ S^2 > \frac{1}{2}) = 0$임을 보이고, $\bar{X}$와 S^2가 독립이 아님을 결론지어라.

풀이 (a) 균등분포는 대칭이므로 $\mu_3 = 0$이고 명제 7.1로 공분산은 0이다.

(b) $|\bar{X}| > \frac{1}{2} \iff |X_1 + X_2| > 1$이다. 넓이 4인 정사각형에서 이 영역은 직각변이 1인 삼각형 두 개로 넓이 1이므로 확률은 $\frac{1}{4}$이다. $S^2 > \frac{1}{2} \iff |X_1 - X_2| > 1$도 같은 넓이이므로 $\frac{1}{4}$이다.

(c) $|X_1 + X_2| + |X_1 - X_2| = 2\max(|X_1|, |X_2|) \le 2$이므로 두 값이 동시에 1보다 클 수 없다. 따라서 결합확률은 $0 \neq \frac{1}{4} \cdot \frac{1}{4}$이고 독립이 아니다 [T2]. 정규 표본에서 무상관이 독립을 준 것은 결합정규라는 성질 때문이었다.

최종답 (b) 둘 다 $\frac{1}{4}$

C6 k개 집단의 합동분산 ★★★

정리 2.2, 5.4 · 25분

k개의 독립 집단에서 $Y_{j1}, \dots, Y_{jn_j} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_j, \sigma^2)$ $(j = 1, \dots, k)$이고 $N = \sum_j n_j$, $S_p^2 = \frac{1}{N-k}\sum_j (n_j - 1)S_j^2$이다.

(a) $\frac{(N-k)S_p^2}{\sigma^2} \sim \chi^2(N-k)$이고 S_p^2가 $(\bar{Y}_1, \dots, \bar{Y}_k)$와 독립임을 보여라.

(b) 상수 $c_1, \dots, c_k$에 대해 $\theta = \sum_j c_j \mu_j$, $\hat{\theta} = \sum_j c_j \bar{Y}_j$이면 $T = \frac{\hat{\theta} - \theta}{S_p \sqrt{\sum_j c_j^2 / n_j}} \sim t(N-k)$임을 보여라.

(c) $k = 2$, $c = (1, -1)$이면 정리 5.4가 됨을 확인하라.

풀이 (a) 정리 2.2를 집단마다 쓰면 $\bar{Y}_j \perp S_j^2$, $\frac{(n_j-1)S_j^2}{\sigma^2} \sim \chi^2(n_j - 1)$이다. 집단들이 독립이므로 $\bar{Y}_1, S_1^2, \dots, \bar{Y}_k, S_k^2$는 상호 독립이다. 독립인 χ^2의 합(4회 정리 3.7)으로 $\frac{(N-k)S_p^2}{\sigma^2} \sim \chi^2\big(\sum(n_j - 1)\big) = \chi^2(N-k)$이고, S_p^2는 S_j^2들의 함수이므로 $\bar{Y}_j$들과 독립이다.

(b) $\hat{\theta} \sim \mathcal{N}\big(\theta,\ \sigma^2 \sum_j c_j^2/n_j\big)$(5회 정리 1.8(iv))이므로 $Z = \frac{\hat{\theta} - \theta}{\sigma\sqrt{\sum c_j^2/n_j}} \sim \mathcal{N}(0,1)$이고 (a)로 $V = \frac{(N-k)S_p^2}{\sigma^2}$와 독립이다. $T = Z/\sqrt{V/(N-k)}$이므로 $t(N-k)$이다.

(c) $k = 2$이면 $N - k = m + n - 2$, S_p^2는 정리 5.4의 합동분산, $\sum c_j^2/n_j = \frac{1}{m} + \frac{1}{n}$이다.

C7 단위를 바꾸어도 같은 구간 ★

정리 5.2 · 10분

$Y_i = a + bX_i$ ($b \neq 0$)이다. $\bar{Y} = a + b\bar{X}$, $S_Y = |b|S_X$ 임을 보이고, μ_Y 의 t-구간이 μ_X 의 t-구간을 같은 아핀 변환으로 옮긴 것임을 보여라($b<0$이면 양 끝이 바뀐다). 섭씨를 화씨로 바꿔도 결론이 같다는 뜻이다.

풀이 $\bar{Y} = \frac{1}{n}\sum(a+bX_i) = a + b\bar{X}$이고 $Y_i - \bar{Y} = b(X_i - \bar{X})$이므로 $S_Y^2 = b^2 S_X^2$, $S_Y = |b|S_X$이다. μ_Y의 구간은 $a + b\bar{X} \pm c|b|S_X/\sqrt{n}$이다. $b>0$이면 이것은 μ_X의 구간 $[\bar{X} - cS_X/\sqrt{n},\ \bar{X} + cS_X/\sqrt{n}]$의 양 끝에 $y = a + bx$를 쓴 것이고, $b<0$이면 옮긴 양 끝이 자리를 바꾼다. $\mu_Y = a + b\mu_X$이므로 두 구간이 참값을 포함하는 사건도 같다. t 통계량은 $\frac{\bar{Y}-\mu_Y}{S_Y/\sqrt{n}} = \operatorname{sign}(b)\,\frac{\bar{X}-\mu_X}{S_X/\sqrt{n}}$이다.

C8 t의 분위수는 정규의 분위수로 간다 ★★★

[T6] · 정리 3.3(d) · 30분

(a) $r \geq 2$이면 $\left(1+\frac{t^2}{r}\right)^r \geq \left(1+\frac{t^2}{2}\right)^2$임을 이항전개의 처음 세 항으로 보여라. 이것으로 $r \geq 2$에서 $\left(1+\frac{t^2}{r}\right)^{-(r+1)/2} \leq \left(1+\frac{t^2}{2}\right)^{-1}$을 얻어라.

(b) 정리 3.3(d)와 지배수렴 정리로, 모든 x에서 $F_r(x) = \int_{-\infty}^{x} f_r \to \Phi(x)$임을 보여라. 상수 c_r이 유계인 것도 확인하라.

(c) (b)로 $t_{r,p} \to z_p$ ($r \to \infty$)를 보여라.

풀이 (a) $u = t^2 \geq 0$이면 이항전개의 항이 모두 음이 아니므로 $(1+u/r)^r \geq 1 + u + \frac{r(r-1)}{2}\frac{u^2}{r^2} = 1 + u + \frac{r-1}{2r}u^2 \geq 1 + u + \frac{u^2}{4} = (1+\frac{u}{2})^2$이다. 따라서 $(1+u/r)^{-(r+1)/2} \leq (1+u/r)^{-r/2} \leq (1+u/2)^{-1}$이다. 첫 부등호는 $1 + u/r \geq 1$에서 나온다.

(b) $c_r = \frac{\Gamma((r+1)/2)}{\sqrt{r\pi}\,\Gamma(r/2)}$은 노트 연습 3.1로 $\frac{1}{\sqrt{2\pi}}$로 가므로 유계이고, $C = \sup_{r\geq 2} c_r < \infty$이다. 그러면 $r \geq 2$에서 $f_r(t) \leq C(1 + t^2/2)^{-1} =: g(t)$이고 $\int g = C\sqrt{2}\,\pi < \infty$이다. 정리 3.3(d)로 $f_r \to \phi$이므로 지배수렴 정리 [T6]으로 $\int_{-\infty}^{x} f_r \to \int_{-\infty}^{x} \phi = \Phi(x)$이다.

(c) $\varepsilon > 0$이면 $\Phi(z_p + \varepsilon) > 1 - p > \Phi(z_p - \varepsilon)$이다. (b)로 충분히 큰 r에서 $F_r(z_p+\varepsilon) > 1-p > F_r(z_p-\varepsilon)$이고, F_r이 연속이고 증가하므로 $F_r(t_{r,p}) = 1-p$인 $t_{r,p}$는 $(z_p-\varepsilon,\ z_p+\varepsilon)$ 안에 있다. ε이 임의이므로 $t_{r,p} \to z_p$이다 [T5]. 노트 3.2절의 $t_{30,0.025} = 2.04$가 1.96에 가까운 이유다.

C9 지수분포의 축량 ★★

정의 5.1 · 15분

$X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ 지수분포(평균 θ)이다.

(a) $Q = \frac{2}{\theta}\sum_i X_i \sim \chi^2(2n)$임을 보이고, Q가 축량임을 밝혀라.

(b) θ의 95% 신뢰구간을 쓰고 포함확률이 정확히 0.95임을 보여라.

(c) $n = 5$, $\sum x_i = 10$이고 $\chi^2_{10,0.025} = 20.483$, $\chi^2_{10,0.975} = 3.247$이면 구간의 아래 끝은?

(d) 같은 자료에서 위 끝은?

풀이 (a) X_i/θ는 평균 1인 지수이고 $\frac{2X_i}{\theta} \sim \Gamma(1,2) = \chi^2(2)$이다. 독립인 $\chi^2(2)$ n개의 합이므로 $\chi^2(2n)$이다(4회 정리 3.7). 분포에 θ가 없으므로 축량이다.

(b) $P\left(\chi^2_{2n,0.975} \le Q \le \chi^2_{2n,0.025}\right) = 0.95$를 θ에 대해 풀면 $\left[\frac{2\sum X_i}{\chi^2_{2n,0.025}},\ \frac{2\sum X_i}{\chi^2_{2n,0.975}}\right]$이다. 사건이 같으므로 포함확률은 모든 θ에서 정확히 0.95이다.

(c)(d) $[20/20.483,\ 20/3.247] = [0.976,\ 6.160]$이다. 정규 모집단이 아니어도 축량이 있으면 정확한 구간을 만들 수 있다.

최종답 (c) 0.976 (d) 6.160

D

퀄형

여러 절과 앞 회차의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 분포를 유도할 때는 어느 정리를 쓰는지 밝히는 것까지 답안이다.

D1 σ**의 불편추정량과 평균제곱오차** ★★★ 정리 2.2(c), 연습 2.4 · 제한 35분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$이고 $c_4(n) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$로 $ES = c_4(n)\sigma$이다(노트 연습 2.4).

(a) $k_n S$가 σ의 불편추정량이 되는 k_n을 쓰고, $n=5$에서 정확한 값을 구하라.

(b) $n=5$에서 $E(k_5 S - \sigma)^2/\sigma^2$을 구하라.

(c) $n=5$에서 $E(S-\sigma)^2/\sigma^2$을 구하고 (b)와 비교하라.

(d) $E(aS-\sigma)^2$을 가장 작게 하는 a를 $c_4(n)$으로 쓰고, $n=5$의 값을 구하라.

(e) (d)의 최솟값 $E(a^*S-\sigma)^2/\sigma^2$을 $n=5$에서 구하라.

풀이 $c = c_4(n)$으로 쓴다. $ES^2 = \sigma^2$이므로 모든 a에서 $E(aS-\sigma)^2 = a^2\sigma^2 - 2a\sigma\cdot c\sigma + \sigma^2 = \sigma^2(a^2 - 2ac + 1)$이다.

(a) $k_n = 1/c_4(n)$이다. $c_4(5) = \frac{1}{\sqrt{2}}\cdot\frac{3\sqrt{\pi}}{4} = \frac{3\sqrt{\pi}}{4\sqrt{2}} \approx 0.940$이므로 $k_5 = \frac{4\sqrt{2}}{3\sqrt{\pi}} \approx 1.064$이다.

(b) $a = 1/c$를 넣으면 $\frac{1}{c^2} - 2 + 1 = \frac{1}{c^2} - 1 = \frac{32}{9\pi} - 1 \approx 0.1318$이다.

(c) $a=1$이면 $2 - 2c = 2 - \frac{3\sqrt{\pi}}{2\sqrt{2}} \approx 0.1200$이다. 불편인 $k_5 S$의 평균제곱오차가 치우친 S보다 크다.

(d) $a^2 - 2ac + 1$은 $a = c$에서 최소이므로 $a^* = c_4(n)$이고, $n=5$이면 $\frac{3\sqrt{\pi}}{4\sqrt{2}} \approx 0.940$이다.

(e) 최솟값은 $1 - c^2 = 1 - \frac{9\pi}{32} \approx 0.1164$이다. S를 키우는 불편화와 반대로, 평균제곱오차를 줄이려면 S를 오히려 줄인다. 노트 연습 1.4의 σ^2에 대한 결론($\frac{1}{n+1}$이 최적)과 같은 방향이다.

최종답 (a) $k_5 = \frac{4\sqrt{2}}{3\sqrt{\pi}} \approx 1.064$ (b) $\frac{32}{9\pi} - 1 \approx 0.1318$ (c) $2 - \frac{3\sqrt{\pi}}{2\sqrt{2}} \approx 0.1200$ (d) $a^* = c_4(5) \approx 0.940$ (e) $1 - \frac{9\pi}{32} \approx 0.1164$

D2 대응을 무시하면 ★★★

정리 5.4, 예제 5.3, 3회 정리 4.2 · 제한 40분

(X_i, Y_i), $i = 1, \dots, n$은 독립인 이변량 정규 쌍이고, $\mathrm{Var}\, X_i = \mathrm{Var}\, Y_i = \sigma^2$, $\mathrm{Corr}(X_i, Y_i) = \rho > 0$, $\Delta = EX_i - EY_i$이다.

(a) $D_i = X_i - Y_i \overset{\text{iid}}{\sim} \mathcal{N}\big(\Delta,\ 2\sigma^2(1-\rho)\big)$임을 보이고, 대응 t 통계량이 $\Delta = 0$에서 $t(n-1)$을 따름을 보여라.

(b) $\mathrm{Var}(\bar X - \bar Y) = \frac{2\sigma^2(1-\rho)}{n}$임을 보여라. 대응을 무시하고 두 표본을 독립으로 보면 이 분산을 $\frac{2\sigma^2}{n}$으로 잘못 어림한다.

(c) σ와 ρ를 안다고 하자. $n = 10$, $\rho = 0.8$, $\Delta = 0.5\sigma$일 때 올바른 z-검정 $|\bar D|/\sqrt{2\sigma^2(1-\rho)/n} > 1.96$의 검정력을 구하라.

(d) 같은 조건에서 대응을 무시한 검정 $|\bar X - \bar Y|/\sqrt{2\sigma^2/n} > 1.96$의 검정력을 구하라.

(e) (d)의 검정의 실제 유의수준($\Delta = 0$에서 기각할 확률)을 구하라.

풀이 (a) (X_i, Y_i)가 결합정규이므로 선형결합 D_i는 정규이고(5회 정리 A), $\mathrm{Var}\, D_i = \sigma^2 + \sigma^2 - 2\rho\sigma^2$이다(3회 정리 4.2). 쌍들이 독립이므로 D_i들은 iid이고, 정리 3.4를 D_i에 쓰면 대응 t는 $t(n-1)$이다.

(b) $\bar X - \bar Y = \bar D$이므로 $\mathrm{Var} = \frac{2\sigma^2(1-\rho)}{n}$이다. 독립으로 보면 공분산 항 $-2\rho\sigma^2/n$을 빠뜨린다.

(c) $\bar D/\sqrt{2\sigma^2(1-\rho)/n} \sim \mathcal{N}(\delta, 1)$, $\delta = \frac{0.5}{\sqrt{2\cdot 0.2/10}} = \frac{0.5}{0.2} = 2.5$이다. 검정력은 $P(|\mathcal{N}(2.5,1)| > 1.96) = \Phi(0.54) + \Phi(-4.46) \approx 0.705$이다.

(d) 통계량 $\frac{\bar D}{\sqrt{2\sigma^2/n}} = \sqrt{1-\rho}\cdot\frac{\bar D}{\sqrt{2\sigma^2(1-\rho)/n}}$은 $\mathcal{N}\big(\sqrt{0.2}\cdot 2.5,\ 0.2\big) = \mathcal{N}(1.118,\ 0.2)$이다. 표준편차 $\sqrt{0.2} = 0.447$로 표준화하면 검정력은 $1 - \Phi\big(\frac{1.96-1.118}{0.447}\big) + \Phi\big(\frac{-1.96-1.118}{0.447}\big) = 1 - \Phi(1.883) + \Phi(-6.88) \approx 0.030$이다.

(e) $\Delta = 0$이면 통계량은 $\mathcal{N}(0, 0.2)$이므로 $P(|Z| > 1.96/\sqrt{0.2}) = 2\big(1 - \Phi(4.383)\big) \approx 1.2 \times 10^{-5}$이다. 대응을 무시한 검정은 거의 기각하지 않고, 효과가 있어도 찾지 못한다. 노트 예제 5.3에서 t가 4.06에서 1.86으로 떨어진 것과 같은 현상이다.

최종답 (c) ≈ 0.705 (d) ≈ 0.030 (e) $\approx 1.2 \times 10^{-5}$

D3 균등분포의 로그와 χ^2, F ★★★

2회 정리 4.3, 4회 정리 3.7, 정의 4.1 · 제한 35분

$U_1, U_2, \dots$ 와 $V_1, V_2, \dots$ 가 모두 독립이고 $\mathrm{U}(0,1)$을 따른다.

(a) $-2\log U_1 \sim \chi^2(2)$임을 보여라.

(b) $-2\sum_{i=1}^{k}\log U_i \sim \chi^2(2k)$임을 보여라.

(c) $\dfrac{\frac{1}{m}\sum_{i=1}^{m}\log U_i}{\frac{1}{n}\sum_{j=1}^{n}\log V_j} \sim F(2m,2n)$임을 보여라.

(d) $F(2,2)$의 분포함수가 $\dfrac{x}{1+x}$ $(x>0)$임을 보이고, $P\big(\log U_1/\log V_1 > 3\big)$을 구하라.

(e) 독립인 두 검정의 유의확률이 $p_1 = 0.04$, $p_2 = 0.10$이다. 귀무가설 아래에서 유의확률은 $\mathrm{U}(0,1)$을 따른다. 결합 통계량 $X = -2(\log p_1 + \log p_2)$의 유의확률 $P(\chi^2(4) \ge X)$를 $q = p_1 p_2$로 나타내고 값을 구하라.

풀이 (a) $x>0$이면 $P(-2\log U_1 \le x) = P(U_1 \ge e^{-x/2}) = 1 - e^{-x/2}$이고, 이것은 평균 2인 지수분포 $\Gamma(1,2) = \chi^2(2)$의 분포함수다(2회 정리 4.3, 4회 정의 3.8).

(b) 독립인 $\chi^2(2)$ k개의 합이므로 4회 정리 3.7로 $\chi^2(2k)$이다.

(c) 분자와 분모에 -2를 곱해도 비는 그대로이므로, 비는 $\dfrac{A/(2m)}{B/(2n)}$이다. 여기서 $A = -2\sum\log U_i \sim \chi^2(2m)$, $B = -2\sum\log V_j \sim \chi^2(2n)$이고 둘은 독립이다. 정의 4.1로 $F(2m,2n)$이다.

(d) 정리 4.2(a)에 $r_1 = r_2 = 2$를 넣으면 $B(1,1) = 1$이므로 밀도는 $(1+x)^{-2}$이고 분포함수는 $1 - \dfrac{1}{1+x} = \dfrac{x}{1+x}$이다. $m = n = 1$이면 (c)의 비가 $\log U_1/\log V_1 \sim F(2,2)$이므로 $P(F>3) = 1 - \dfrac{3}{4} = \dfrac{1}{4}$이다.

(e) A3의 계산처럼 $P(\chi^2(4) \ge x) = e^{-x/2}(1 + \frac{x}{2})$이고 $e^{-X/2} = p_1 p_2 = q$, $\dfrac{X}{2} = -\log q$이므로 유의확률은 $q(1 - \log q)$이다. $q = 0.004$이면 $0.004\,(1 + 5.521) \approx 0.0261$이다. 따로는 5%를 넘는 증거(0.10)도 합치면 2.6%의 증거가 된다(Fisher 1932의 결합 검정).

최종답 (d) $\dfrac{1}{4}$ (e) $q(1-\log q) \approx 0.0261$

D4 관측 두 개의 t ★★★

정리 2.2, 3.3(c), 3회 정리 5.8 · 제한 30분

$X_1, X_2 \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma^2)$이고 $T = \dfrac{\bar X - \mu}{S/\sqrt{2}}$이다.

(a) $T = \dfrac{X_1 + X_2 - 2\mu}{|X_1 - X_2|}$임을 보여라.

(b) $Y_1 = \dfrac{Z_1+Z_2}{\sqrt{2}}$, $Y_2 = \dfrac{Z_1 - Z_2}{\sqrt{2}}$ $(Z_i = (X_i - \mu)/\sigma)$로 $T = Y_1/|Y_2|$임을 보여라. 정리 3.3(c)를 쓰지 말고 3회 정리 5.8로 T가 Cauchy를 따름을 보여라.

(c) $t_{1,0.025}$를 정확한 식으로 쓰고 값을 구하라.

(d) $\bar X \pm 1.96\,S/\sqrt{2}$의 실제 포함확률을 정확한 식으로 쓰고 값을 구하라.

풀이 (a) $n=2$이면 $S^2=\frac{(X_1-X_2)^2}{2}$이므로 $S/\sqrt{2}=\frac{|X_1-X_2|}{2}$이고, $\bar X-\mu=\frac{X_1+X_2-2\mu}{2}$이다. 나누면 결론이다.

(b) σ로 나누면 $T=\frac{Z_1+Z_2}{|Z_1-Z_2|}=\frac{\sqrt{2}Y_1}{\sqrt{2}|Y_2|}=\frac{Y_1}{|Y_2|}$이다. (Y_1,Y_2)는 $n=2$의 Helmert 회전이므로 독립인 표준정규다(정리 2.2의 증명). Y_1이 Y_2와 독립이고 대칭이므로 $\operatorname{sign}(Y_2)Y_1$은 Y_1과 같은 분포를 갖고 Y_2와 독립이다. 따라서 $Y_1/|Y_2|=\operatorname{sign}(Y_2)Y_1/Y_2$는 Y_1/Y_2와 같은 분포이고, 3회 정리 5.8로 Cauchy다.

(c) $\frac12+\frac1\pi\arctan t=0.975$에서 $t_{1,0.025}=\tan(0.475\pi)\approx 12.71$이다.

(d) $P(|T|\le 1.96)=\frac{2}{\pi}\arctan 1.96\approx 0.700$이다. 노트 1.1절의 $n=2$의 값과 같다. 관측 두 개로 95%를 말하려면 1.96 대신 12.71을 곱해야 한다.

최종답 (c) $\tan(0.475\pi)\approx 12.71$ (d) $\frac{2}{\pi}\arctan 1.96\approx 0.700$