종합 모의고사

수리통계 자격시험 모의고사 A · B · C

1–16회 전체를 범위로 하는 대학원 수리통계 자격시험 형식의 모의고사 세 세트다. 세트 하나는 180분 동안 푸는 6문항, 100점이다. 강의노트, 워크북, 증명 도구 상자를 덮고 푼다.

세트	문항	배점	시간
모의고사 A	A1–A6	100점	180분
모의고사 B	B1–B6	100점	180분
모의고사 C	C1–C6	100점	180분

수치 정규분포·t·χ^2·F의 분위수처럼 필요한 수치는 문제에 준다. 계산기는 쓰지 않는다

답안 증명은 등호와 부등호마다 근거를 단다. 근거가 없는 단계는 채점 기준에서 감점한다

채점 풀이판의 채점 기준으로 채점한다. 최종답이 수나 식인 소문항은 노트북 `17_mock_qual.ipynb`의 `check("A3b", …)`로도 확인한다

응시 요령

1. **세트 하나를 한 번에 푼다.** 180분을 재고, 중간에 노트를 펴지 않는다. 문항 머리의 배점과 권장 시간을 보고 시간을 나눈다.
2. **모든 문항을 푼다.** 선택 문항은 없다. 막힌 소문항은 앞 소문항의 결과를 가정하고 다음으로 넘어가도 된다. 그때는 무엇을 가정했는지 적는다.
3. **가정을 확인한다.** 극한과 적분의 교환, 정칙 조건, 표본 크기에 대한 조건처럼 정리를 쓰기 위한 조건을 답안에 적는다. 채점 기준에 그 배점이 있다.
4. **채점은 다음 날 한다.** 풀이판의 채점 기준으로 소문항마다 점수를 매기고, 틀린 곳의 **관련** 회차로 돌아가 해당 정리의 단계형 증명과 워크북 B부를 다시 푼다.

A

모의고사 A

시간은 180분이고 6문항, 100점이다. 강의노트, 워크북, 증명 도구 상자를 덮고 푼다. 분위수처럼 계산에 필요한 수치는 문제에 주어져 있다. 증명에는 단계마다 근거를 단다.

A1 기록값의 수

15점 · 25분

$X_1, X_2, \dots$ 가 연속분포를 따르는 iid 확률변수이다. X_1 은 기록값이고, $k \ge 2$ 이면 $X_k > \max(X_1, \dots, X_{k-1})$ 일 때 X_k 를 기록값이라 한다. I_k 를 X_k 가 기록값이라는 사건의 지시함수, $R_n = I_1 + \cdots + I_n$ 을 처음 n 개 가운데 기록값의 수라 하자. $(X_1, \dots, X_k)$ 의 $k!$ 가지 크기 순서가 모두 같은 확률이라는 사실은 증명 없이 써도 된다.

(a) (4점) $P(I_k = 1) = 1/k$ 임을 보여라. $j < k$ 이면 $P(I_j = 1, I_k = 1) = 1/(jk)$ 임을 보이고, I_j 와 I_k 가 독립임을 결론지어라.

(b) (4점) $ER_n = \sum_{k=1}^{n} \frac{1}{k}$, $\operatorname{Var} R_n = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k^2}\right)$ 임을 보여라.

(c) (3점) $n = 10$ 이다. $\sum_{k=1}^{10} \frac{1}{k} \approx 2.929$, $\sum_{k=1}^{10} \frac{1}{k^2} \approx 1.550$ 을 써서 $P(R_{10} \ge 6)$ 의 Markov 상한과 Chebyshev 상한을 구하라.

(d) (4점) $0 \le \sum_{k=1}^{n} \frac{1}{k} - \log n \le 1$ 을 써서 $R_n / \log n$ 이 1로 확률수렴함을 보여라.

A2 지수분포의 비에 대한 조건부 분포 15점 · 25분

X와 Y가 독립이고 둘 다 밀도 e^{-x} $(x>0)$를 갖는다. $U = X/Y$, $V = Y$로 둔다.

(a) (4점) (U, V)의 결합밀도를 구하고, U의 밀도가 $f_U(u) = (1+u)^{-2}$ $(u>0)$임을 보여라. $EU = \infty$임을 보여라.

(b) (4점) $U = u$일 때 V의 조건부 분포를 구하고, $E(X \mid U = u)$를 구하라.

(c) (3점) $W = U/(1+U)$가 $\mathrm{U}(0,1)$을 따름을 보여라.

(d) (4점) $\mathrm{Var}(X \mid U)$를 W로 쓰고, 전분산 공식으로 $\mathrm{Var}\,X = 1$을 다시 얻어라. $E[\mathrm{Var}(X \mid U)]$는 얼마인가?

A3 생존확률의 두 추정량 15점 · 25분

$X_1, \dots, X_n$이 밀도 $\theta^{-1}e^{-x/\theta}$ $(x>0)$인 iid 표본이다. $t>0$을 고정하고 $\psi = P(X_1 > t) = e^{-t/\theta}$를 추정한다. 두 추정량

$$\hat{\psi}_1 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i > t\}, \qquad \hat{\psi}_2 = \exp(-t/\bar{X}_n)$$

을 비교한다.

(a) (3점) $\sqrt{n}(\hat{\psi}_1 - \psi)$의 극한분포를 구하라.

(b) (4점) $\sqrt{n}(\hat{\psi}_2 - \psi)$의 극한분포를 구하라. 근사분산 $v_2(t,\theta)$를 적어라.

(c) (4점) $x = t/\theta$로 두고 두 근사분산의 비 v_2/v_1을 x의 식으로 쓴 뒤, 모든 $x>0$에서 1보다 작음을 보여라. $t=\theta$일 때 그 비는 얼마인가? ($e \approx 2.718$)

(d) (4점) 실제 자료가 $\mathrm{U}(0,2)$에서 나왔고 $t=1$이다. 두 추정량은 각각 무엇으로 확률수렴하는가? 이 결과와 (c)에서 추정량을 고르는 기준을 한 문장으로 적어라.

A4 Pareto 분포: 추정과 검정 20점 · 40분

$X_1, \dots, X_n$ 이 밀도

$$f(x;\theta) = \theta\, x^{-(\theta+1)}, \qquad x > 1,\ \theta > 0$$

인 iid 표본이다. $T = \sum_{i=1}^{n} \log X_i$ 로 둔다.

(a) (4점) θ 의 MLE를 구하고, 그것이 가능도의 유일한 최댓점임을 보여라.

(b) (4점) Fisher 정보 $I(\theta)$를 구하고 MLE의 극한분포를 적어라. 정칙 조건 가운데 이 분포족에서 확인해야 할 것을 하나 들어 확인하라.

(c) (5점) $Y_i = \log X_i$ 의 분포를 구하라. T가 완비충분통계량임을 보이고 θ의 UMVUE를 구하라. 그 분산을 Cramér–Rao 한계와 비교하라.

(d) (4점) $H_0 : \theta \le \theta_0$ 대 $H_1 : \theta > \theta_0$ 의 크기 α UMP 검정을 구하고, 기각역의 경계를 χ^2 분위수로 적어라.

(e) (3점) $n = 10$, $\theta_0 = 2$, $\alpha = 0.05$ 이다. $\chi^2(20)$의 하위 5% 분위수는 10.851이다. 기각역의 경계를 수로 구하라. 관측값이 $T = 2.5$이면 결론은? ($P(\chi^2(20) \le 10) \approx 0.032$)

A5 직교 다항식 회귀 20점 · 40분

$x = -2, -1, 0, 1, 2$에서 반응을 하나씩 관측해 $y = (5,\ 2,\ 1,\ 3,\ 9)$를 얻었다. 모형은 $z_i = x_i^2 - 2$로 두고

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

이다. $t_{2,0.975} = 4.303$, $F_{1,2;0.95} = 18.51$, $\sqrt{0.02} \approx 0.1414$를 써도 된다.

(a) (4점) 설계행렬 X를 쓰고 $X'X$가 대각행렬임을 보여라. 최소제곱추정값 $\hat{\beta}$를 구하라.

(b) (4점) 잔차제곱합과 s^2을 구하라. $\hat{\beta}$와 s^2의 분포, 그리고 둘의 관계를 적어라.

(c) (4점) $\tilde{\beta}_1 = (y_5 - y_1)/4$가 β_1의 선형불편추정량임을 보이고, $\operatorname{Var}\hat{\beta}_1 / \operatorname{Var}\tilde{\beta}_1$을 구하라. 이 비가 1보다 작다는 것은 어느 정리의 한 예인가?

(d) (5점) 이차항이 필요한지 $H_0 : \beta_2 = 0$을 수준 0.05에서 F 검정으로 검정하라. F 통계량의 값을 적어라.

(e) (3점) β_1의 95% 신뢰구간을 구하라.

A6 척도 손실에서의 Bayes 추정과 minimax

15점 · 25분

$X_1, \dots, X_n$ ($n \ge 3$)이 밀도 $\lambda e^{-\lambda x}$ ($x > 0$)인 iid 표본이고 $T = \sum X_i$이다. 사전분포는 $\lambda \sim \Gamma(\alpha,\ \text{율}\ \beta)$, 곧 밀도가 $\lambda^{\alpha-1} e^{-\beta\lambda}$에 비례한다. 손실은 상대오차의 제곱 $L(\lambda, d) = (d/\lambda - 1)^2$이다. 답의 식에서 α, β는 `alpha`, `beta`로 쓴다.

(a) (3점) λ의 사후분포를 구하라.

(b) (4점) 이 손실에서 Bayes 추정량을 구하고, 그때의 사후 기대손실이 $1/(\alpha + n - 1)$임을 보여라.

(c) (5점) $c > 0$에 대해 $d_c = c/T$의 위험 $R(\lambda, d_c)$를 구하라. 위험이 λ에 의존하지 않음을 보이고, MLE의 위험을 구하라. MLE는 허용적인가?

(d) (3점) $d_{n-2} = (n-2)/T$가 minimax임을 보이고, 그 위험을 구하라.

B

모의고사 B

시간 180분, 6문항, 100점. 강의노트, 워크북, 증명 도구 상자를 덮고 푼다. 필요한 수치(분위수, 로그값 등)는 문제에 있고 계산기는 쓰지 않는다. 노트북의 `check()`에 식을 넣을 때는 λ를 `lam`, μ를 `mu`, ρ를 `rho`로 쓴다.

B1 두 Poisson 과정과 Chernoff 한계

15점 · 25분

두 종류의 사건 A와 B가 서로 독립인 Poisson 과정으로 일어난다. A의 비율은 λ, B의 비율은 μ다.

(a) (4점) A가 처음 일어나기 전에 B가 일어난 횟수를 N이라 하자. $P(N = k)$ $(k = 0, 1, \dots)$를 구하고, N이 어떤 분포인지 밝혀라.

(b) (4점) $X \sim \mathrm{Poi}(\lambda)$의 적률생성함수를 구하고, $a > \lambda$이면 $P(X \geq a) \leq e^{-\lambda}(e\lambda/a)^a$ 임을 보여라.

(c) (4점) $\lambda = 2$, $a = 6$에서 (b)의 한계, Markov 부등식의 한계, Chebyshev 부등식으로 얻는 한계를 구해 비교하라. $e^4 \approx 54.6$이다.

(d) (3점) (b)로 $P(X \geq 2\lambda) \leq (e/4)^{\lambda}$ 임을 보이고, $\lambda \to \infty$에서 Chebyshev 부등식이 주는 한계와 줄어드는 빠르기를 비교하라.

B2 무게가 다른 정규 관측의 t 구간 15점 · 25분

$X_1, \dots, X_n$이 독립이고 $X_i \sim \mathcal{N}(\mu, \sigma^2/w_i)$이다. 무게 $w_i > 0$는 알려져 있고 μ와 σ^2은 모른다. $W = \sum_i w_i$, $\hat{\mu} = \sum_i w_i X_i / W$, $S_w^2 = \frac{1}{n-1}\sum_i w_i (X_i - \hat{\mu})^2$으로 둔다.

(a) (4점) $Z_i = \sqrt{w_i}(X_i - \mu)/\sigma$, $\mathbf{a} = (\sqrt{w_1}, \dots, \sqrt{w_n})'/\sqrt{W}$로 두면 $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, I)$, $\mathbf{a}'\mathbf{Z} = \sqrt{W}(\hat{\mu} - \mu)/\sigma$, $(n-1)S_w^2/\sigma^2 = \mathbf{Z}'(I - \mathbf{a}\mathbf{a}')\mathbf{Z}$임을 보여라.

(b) (4점) $\hat{\mu}$와 S_w^2이 독립이고 $(n-1)S_w^2/\sigma^2 \sim \chi^2(n-1)$임을 보여라.

(c) (4점) $T = \sqrt{W}(\hat{\mu} - \mu)/S_w \sim t(n-1)$임을 보이고 μ의 $1-\alpha$ 신뢰구간을 써라.

(d) (3점) $n = 4$, 무게 $(1, 1, 2, 4)$, 관측값 $(6, 12, 11, 14)$에서 95% 구간을 $\hat{\mu} \pm h$ 꼴로 구하라. $t(3)$의 0.975 분위수는 3.182이다.

B3 분포에 무관한 분위수 구간과 허용구간 15점 · 25분

$X_1, \dots, X_n$이 iid이고 분포함수 F는 연속이라는 것 말고는 모른다. $0 < p < 1$에서 $F(\xi_p) = p$인 점 ξ_p가 하나뿐이라고 하자.

(a) (4점) $r < s$이면 $P\big(X_{(r)} \le \xi_p < X_{(s)}\big) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$이고, 따라서 이 확률이 F에 의존하지 않음을 보여라.

(b) (3점) $n = 10$, $p = 0.75$에서 구간 $[X_{(5)},\ X_{(10)})$이 상사분위수 $\xi_{0.75}$를 포함할 확률을 구하라. $B \sim \mathrm{b}(10, 0.75)$의 분포함수는 $P(B \le 4) = 0.0197$, $P(B \le 9) = 0.9437$이다.

(c) (5점) $R = F(X_{(n)}) - F(X_{(1)})$은 구간 $[X_{(1)}, X_{(n)}]$이 덮는 모집단의 비율이다. R의 분포가 F에 의존하지 않음을 보이고, $0 < \gamma < 1$에서 $P(R \ge \gamma) = 1 - n\gamma^{n-1} + (n-1)\gamma^n$임을 보여라. $n = 20$, $\gamma = 0.9$에서 이 확률을 구하라. $0.9^{19} \approx 0.1351$, $0.9^{20} \approx 0.1216$이다.

(d) (3점) $H_0 : \xi_{0.5} = m_0$ 대 $H_1 : \xi_{0.5} > m_0$를 $S = \#\{i : X_i > m_0\}$가 크면 기각하는 부호검정으로 한다. p-값 $p(s) = P_{H_0}(S \ge s)$에 대해 H_0 아래 모든 $u \in (0,1)$에서 $P\{p(S) \le u\} \le u$임을 보이고, $n = 10$에서 $P_{H_0}\{p(S) \le 0.05\}$를 구하라.

B4 분산이 같은 이변량 정규의 상관계수

20점 · 40분

(X_i, Y_i), $i = 1, \dots, n$이 iid이고, 평균이 $(0,0)$, 두 분산이 모두 σ^2, 상관계수가 ρ인 이변량 정규분포를 따른다. $\sigma^2 > 0$과 $|\rho| < 1$을 모두 모른다. $S = \sum_i (X_i^2 + Y_i^2)$, $P = \sum_i X_i Y_i$로 둔다.

(a) (4점) 로그가능도를 S와 P로 쓰고, MLE가 $\hat{\sigma}^2 = S/(2n)$, $\hat{\rho} = 2P/S$임을 보여라.

(b) (5점) 한 쌍의 Fisher 정보 행렬을 (σ^2, ρ)에 대해 구하라. 비대각 원소가 0이 아님을 확인하라.

(c) (5점) $\sqrt{n}(\hat{\rho} - \rho)$의 점근분산을 구하라. σ^2을 알 때 ρ의 MLE의 점근분산과 비교해, σ^2을 모르는 대가를 두 분산의 비로 나타내라.

(d) (4점) $n = 50$, $\hat{\rho} = 0.3$에서 $H_0 : \rho = 0$의 가능도비 통계량 $-2\log\Lambda$를 구하고 5% 수준에서 판정하라. $\log 0.91 \approx -0.0943$이고 $\chi^2(1)$의 0.95 분위수는 3.841이다.

(e) (2점) 같은 자료에서 Wald 통계량 $n\hat{\rho}^2/(1-\hat{\rho}^2)^2$을 구하고 (d)와 비교하라. $0.91^2 = 0.8281$이다.

B5 두 지수 표본의 비교와 축차 검정 20점 · 40분

$X_1, \dots, X_m$은 평균 θ_1인 지수분포에서, $Y_1, \dots, Y_n$은 평균 θ_2인 지수분포에서 온 서로 독립인 표본이다. $H_0 : \theta_1 = \theta_2$ 대 $H_1 : \theta_1 \neq \theta_2$를 검정한다. $B = \sum_i X_i \big/ \left(\sum_i X_i + \sum_j Y_j\right)$로 둔다.

(a) (5점) 가능도비가 $\Lambda = \dfrac{(m+n)^{m+n}}{m^m n^n} B^m (1-B)^n$ 임을 보여라.

(b) (5점) H_0 아래 $B \sim \mathrm{Beta}(m, n)$이고 $F = \dfrac{\sum_i X_i / m}{\sum_j Y_j / n} \sim F(2m, 2n)$임을 보여라. 가능도비 검정의 기각역이 $\{B \le c_1\} \cup \{B \ge c_2\}$ 꼴임을 보이고, c_1과 c_2가 만족하는 두 식을 써라.

(c) (5점) $m = n = 5$, $\sum x_i = 30$, $\sum y_j = 10$이다. 크기 0.05의 정확한 가능도비 검정을 F로 수행하라($F(10,10)$의 0.975 분위수는 3.717). 또 $-2\log\Lambda$를 계산해 Wilks 근사(3.841)의 판정과 비교하라. $\log 2 \approx 0.6931$, $\log 3 \approx 1.0986$이다.

(d) (5점) 이번에는 지수분포의 관측 $X_1, X_2, \dots$를 하나씩 보며 $H_0 : \theta = 1$ 대 $H_1 : \theta = 2$를 $\alpha = \beta = 0.05$의 SPRT로 검정한다. Wald의 경계를 쓰고, H_1 아래 평균 표본 크기의 Wald 근사를 구하라. $\log 19 \approx 2.944$이다.

B6 Kendall의 불일치 쌍의 수

15점 · 25분

연속분포에서 온 쌍 (X_i, Y_i), $i = 1, \dots, n$이 iid다. $i < j$인 쌍 가운데 $(X_i - X_j)(Y_i - Y_j) < 0$인 것(불일치 쌍)의 수를 K라 하고, H_0: X와 Y가 독립이다를 검정한다.

(a) (4점) X의 순서로 늘어놓은 Y의 순위를 $\pi = (\pi_1, \dots, \pi_n)$이라 하자. $K = \#\{i < j : \pi_i > \pi_j\}$임을 보이고, H_0 아래 $V_j = \#\{i < j : \pi_i > \pi_j\}$ $(j = 2, \dots, n)$가 상호 독립이며 V_j가 $\{0, 1, \dots, j-1\}$ 위의 균등분포임을 보여라.

(b) (4점) H_0 아래 $EK = n(n-1)/4$, $\operatorname{Var} K = n(n-1)(2n+5)/72$임을 보여라.

(c) (4점) $n = 4$, 자료 (1.2, 0.8), (2.5, 1.9), (3.1, 1.7), (4.0, 3.5)에서 K를 구하라. 양의 연관을 대립가설로 하는 정확한 p-값 $P_{H_0}(K \le k_{\text{obs}})$를 구하라.

(d) (3점) $n = 10$에서 $K = 12$가 관측되었다. 정규근사로 한쪽 p-값을 구하고, 정규근사를 쓸 수 있는 근거를 한 줄로 밝혀라. $\sqrt{31.25} \approx 5.590$, $\Phi(-1.88) \approx 0.030$이다.

C

모의고사 C

시간 180분, 6문항, 100점이다. 강의노트, 워크북, 증명 도구 상자를 덮고 푼다. 분위수와 거듭제곱처럼 필요한 수치는 문제에 주어져 있다. 각 소문항의 배점은 문항 안에 적혀 있다.

C1 독립인 사건열과 두 가지 수렴

15점 · 25분

사건열 $A_1, A_2, \dots$ 가 상호 독립이고 $P(A_n) = p_n$ 이며 $X_n = \mathbb{1}_{A_n}$ 이다. 확률변수열 Y_n 이 0으로 **거의 확실하게** 수렴한다는 것은 사건 $\{\omega : Y_n(\omega) \to 0\}$ 의 확률이 1이라는 뜻으로 쓴다. $\limsup_n A_n = \bigcap_{n \ge 1} \bigcup_{k \ge n} A_k$ 는 무한히 많은 A_k 가 일어나는 사건이다.

(a) (3점) X_n 이 0으로 확률수렴할 필요충분조건은 $p_n \to 0$ 임을 보여라.

(b) (5점) $n \le m$ 이면 $P\left(\bigcap_{k=n}^{m} A_k^c\right) = \prod_{k=n}^{m} (1 - p_k) \le \exp\left(-\sum_{k=n}^{m} p_k\right)$ 임을 보이고, $\sum_{n} p_n = \infty$ 이면 $P(\limsup_n A_n) = 1$ 임을 보여라.

(c) (4점) $p_n = 1/n$ $(n \ge 2)$ 이면 $P\left(\bigcap_{k=10}^{100} A_k^c\right)$ 은 얼마인가? 이때 X_n 은 0으로 확률수렴하지만 거의 확실하게 수렴하지는 않음을 보여라.

(d) (3점) $p_n = 1/n^2$ $(n \ge 2)$ 이면 X_n 이 0으로 거의 확실하게 수렴함을 보이고, 어떤 A_n $(n \ge 2)$ 도 일어나지 않을 확률을 구하라.

C2 Poisson 과정의 도착 시각 15점 · 25분

도착 비율이 λ인 Poisson 과정의 사건 사이 시간 E_1, E_2, E_3가 iid $\mathrm{Exp}(\lambda)$이고 $T_k = E_1 + \cdots + E_k$는 k번째 도착 시각이다.

(a) (5점) (T_1, T_2, T_3)의 결합밀도가 $\lambda^3 e^{-\lambda t_3}$ $(0 < t_1 < t_2 < t_3)$임을 변수변환으로 보여라. 그리고 $U_1 = T_1/T_3$, $U_2 = T_2/T_3$로 두면 (U_1, U_2)는 T_3와 독립이고 삼각형 $\{0 < u_1 < u_2 < 1\}$ 위의 밀도 2인 균등분포를 따름을 보여라.

(b) (4점) $E\left[T_1 T_2 / T_3^2\right]$를 구하라.

(c) (3점) $E(T_1 \mid T_3 = s)$를 구하라.

(d) (3점) 첫 도착이 셋째 도착 시각의 절반보다 늦을 확률 $P(T_1 > T_3/2)$를 구하라.

C3 Pareto 척도 모수: 비정칙 모형의 추정과 검정 20점 · 40분

$X_1, \dots, X_n$이 iid이고 밀도가 $f(x;\theta) = a\theta^a x^{-(a+1)}$ $(x \ge \theta)$이다. 모양 $a > 0$은 알고 척도 $\theta > 0$은 모른다. 필요하면 $0.05^{1/15} = 0.8190$, $1.1^{15} = 4.177$을 쓴다.

(a) (4점) θ의 MLE $\hat\theta$를 구하고 $P_\theta(\hat\theta > t)$를 구하라.

(b) (3점) $n(\hat\theta - \theta)$의 극한분포를 구하고 그 평균을 적어라. 정칙 모형의 $\sqrt{n}$과 비교하라.

(c) (5점) $X_{(1)}$이 완비충분통계량임을 보이고 θ의 UMVUE와 그 분산을 구하라($na > 2$).

(d) (3점) 점수의 평균 $E_\theta\big[\frac{\partial}{\partial\theta}\log f(X;\theta)\big]$를 구해 Cramér–Rao 부등식의 조건이 깨짐을 보여라. $n = 5$, $a = 3$에서 UMVUE의 분산과 형식적인 한계 $\theta^2/(na^2)$의 비 $\mathrm{Var}(\text{UMVUE})\big/\big(\theta^2/(na^2)\big)$를 구하라.

(e) (3점) $n = 5$, $a = 3$에서 $H_0 : \theta \le 1$ 대 $H_1 : \theta > 1$의 크기 0.05 UMP 검정을 구하라. 기각 경계 c는 얼마인가?

(f) (2점) (e)의 검정의 $\theta = 1.1$에서의 검정력을 구하라.

C4 평균을 아는 정규 표본의 척도: Basu와 Rao–Blackwell

15점 · 25분

$X_1, \dots, X_n$이 iid $\mathcal{N}(0, \sigma^2)$이고 $T = \sum_{i=1}^{n} X_i^2$, $W = X_1^2 / T$이다. $EZ^4 = 3$ ($Z \sim \mathcal{N}(0,1)$)을 써도 된다. $F(1, 9)$의 위쪽 5% 분위수는 5.117이다.

(a) (4점) T가 σ^2의 완비충분통계량이고 W가 보조통계량임을 보이고, W와 T가 독립임을 보여라.

(b) (4점) $EX_1^4 = E[W^2 T^2]$에 (a)를 써서 EW^2를 구하라. $n = 10$이면 얼마인가?

(c) (4점) $E(X_1^4 \mid T)$를 구하고 σ^4의 UMVUE를 구하라. $n = 10$, $T = 20$이면 그 값은?

(d) (3점) $(n-1)W/(1-W) \sim F(1, n-1)$임을 보이고, $n = 10$에서 $P(W > w) = 0.05$인 w를 구하라.

C5 세 처리의 완전무작위설계 20점 · 40분

12개 실험 단위를 세 처리 A, B, C에 4개씩 완전히 무작위로 배정했다. 반응은 다음과 같다.

처리	반응			
A	8	10	9	13
B	12	14	11	15
C	16	17	19	20

모형은 $y_{ij} = \mu_i + \varepsilon_{ij}$, ε_{ij} iid $\mathcal{N}(0, \sigma^2)$이다. $F_{0.05}(2, 9) = 4.256$, $\sqrt{17/9} = 1.374$, $\sqrt{2 \times 4.256} = 2.918$, t_9의 위쪽 0.05/6 분위수는 2.933이다.

(a) (6점) 분산분석표(SS, 자유도, MS)를 만들고 F 통계량을 구해 $H_0 : \mu_1 = \mu_2 = \mu_3$을 수준 0.05에서 검정하라.

(b) (4점) 반응이 처리와 무관하다는 날카로운 귀무가설 아래의 무작위화 검정을 생각한다. 가능한 배정은 몇 가지인가? F 통계량을 쓰는 무작위화 검정의 p-값이 가질 수 있는 가장 작은 값은 얼마인가? 이 검정의 크기가 정확한 이유를 한 문단으로 쓰라.

(c) (5점) $\mu_3 - \mu_1$의 95% Scheffé 구간을 구하라. 세 쌍의 비교에 Bonferroni를 쓴 구간의 반폭과 비교하라.

(d) (5점) 참 평균이 $(10, 12, 14)$이고 $\sigma = 2$이면 F 검정의 비중심도 λ는 얼마인가? 이때 검정력은 $P\big(F(2, 9, \lambda) > 4.256\big) = 0.560$이다. 단위 수를 처리마다 8개로 늘리면 λ는 얼마가 되고 검정력은 왜 커지는가?

C6 사고 수의 경험적 Bayes

15점 · 25분

운전자 i의 한 해 사고 수가 $X_i \mid \lambda_i \sim \mathrm{Poi}(\lambda_i)$이고 사고 성향 λ_i는 iid $\Gamma(\alpha, \beta)$(모양 α, 비율 β, 평균 α/β)이다. 노트북의 `check`에서는 α, β를 `alpha`, `beta`로 쓴다. $0.75^{10} = 0.0563$이다.

(a) (4점) λ의 사후분포와 사후평균 $E(\lambda \mid X = x)$를 구하고, 사후평균이 x와 사전평균의 가중평균임을 보여라.

(b) (4점) 제곱오차 손실에서 X와 사후평균의 Bayes 위험을 각각 구하라.

(c) (4점) 운전자 200명의 사고 수의 표본평균이 2.0, 표본분산이 3.0이었다. X의 주변 평균과 분산을 써서 α, β를 적률법으로 추정하고, 사고를 6번 낸 운전자의 경험적 Bayes 추정값을 구하라.

(d) (3점) (c)의 추정값을 참값으로 보고, 사고를 6번 낸 운전자가 다음 해에 사고를 한 번도 내지 않을 확률 $P(Y = 0 \mid X = 6)$을 구하라. $Y \mid \lambda \sim \mathrm{Poi}(\lambda)$는 λ가 주어지면 X와 독립이다.