수리통계학 × 통계학의 역사

7회 워크북 — 풀이판

극한의 언어
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	정규근사, 델타 방법, Monte Carlo의 분산을 손으로 계산한다	14	1시간 50분
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 25분
	합계	34	약 10시간

노트 `07_limits_notes.pdf` — 문항의 "정리 3.1" 같은 번호는 노트의 번호다

도구 `00_toolkit_notes.pdf` — [T1] 원소 추적, [T2] 반례, [T4] 부등식, [T5] 극한과 급수, [T8] Taylor 전개

검산 노트북 `07_limits.ipynb`의 "워크북 검산" 절: `check("A5", "1/4")`

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. 정규분포표의 값으로 계산한 답은 넷째 자리까지 맞으면 맞다고 본다.
5. **극한 정리는 가정을 적는다.** 대수의 법칙, 중심극한정리, Slutsky, 델타 방법을 쓸 때마다 그 정리의 가정(평균의 존재, 분산의 존재, 상수로의 확률수렴, 미분가능성)을 확인한 줄을 남긴다.

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. 필요한 정규분포의 값은 문제에 적어 두었다.

A1 Chebyshev와 중심극한정리 ★

정리 3.1, 2회 정리 5.3 · 8분

성공 확률 0.3인 독립 시행 n번의 성공 비율을 $\hat{p}$라 하자. $P(|\hat{p}-0.3|\ge 0.02)\le 0.05$가 되는 가장 작은 n을 (a) Chebyshev 부등식으로, (b) 중심극한정리의 근사로($z_{0.025}=1.96$) 구하라.

풀이 $pq=0.21$이다. (a) Chebyshev로 $P(|\hat{p}-p|\ge 0.02)\le \frac{0.21}{n\cdot 0.0004}$이고, 이것이 0.05 이하이려면 $n\ge \frac{0.21}{0.0004\cdot 0.05}=10{,}500$이다. (b) $P(|\hat{p}-p|\ge 0.02)\approx P\big(|Z|\ge 0.02\sqrt{n}/\sqrt{0.21}\big)$이므로 $0.02\sqrt{n}/\sqrt{0.21}\ge 1.96$, 곧 $n\ge 1.96^2\cdot 0.21/0.0004=2016.84$이고 가장 작은 정수는 2017이다. Chebyshev는 모든 분포에 성립하는 보장이어서 약 다섯 배 보수적이다.

최종답 (a) 10,500 (b) 2,017

A2 이항의 정규근사 ★

따름정리 3.2(a) · 8분

$K\sim \mathrm{b}(400,\frac{1}{2})$에서 $P(K\le 190)$을 (a) 연속성 보정을 하여, (b) 보정 없이 정규근사로 구하라. $\Phi(0.95)=0.8289$, $\Phi(1)=0.8413$이다. 정확한 값은 0.1711이다.

풀이 $EK=200$, $\operatorname{Var}K=400\cdot\frac{1}{4}=100$이다. (a) $P(K\le 190)=P(K\le 190.5)\approx \Phi\big(\frac{190.5-200}{10}\big)=\Phi(-0.95)=1-0.8289=0.1711$. (b) $\Phi\big(\frac{190-200}{10}\big)=\Phi(-1)=0.1587$. 대칭인 이항에서는 보정한 값이 정확한 값과 넷째 자리까지 맞는다. 보정하지 않으면 190이라는 정수 위의 확률 절반을 빠뜨린다.

최종답 (a) $\Phi(-0.95)=0.1711$ (b) $\Phi(-1)=0.1587$

A3 Poisson의 정규근사 ★

따름정리 3.2(b) · 6분

$X\sim \mathrm{Poi}(100)$에서 $P(X\ge 115)$를 연속성 보정을 한 정규근사로 구하라. $\Phi(1.45)=0.9265$이다. 정확한 값 0.0759와 비교하라.

풀이 $P(X\ge 115)=P(X\ge 114.5)\approx 1-\Phi\big(\frac{114.5-100}{10}\big)=1-\Phi(1.45)=0.0735$이다. 정확한 값 0.0759보다 작다. Poisson은 오른쪽으로 치우쳐 있어 정규근사가 오른쪽 꼬리를 얕잡는다

(노트 예제 3.3).

최종답 $1-\Phi(1.45)=0.0735$

A4 최댓값의 극한 ★

예제 2.2(b) · 6분

$U_1,\dots,U_{10}\overset{\text{iid}}{\sim}\mathrm{U}(0,1)$의 최댓값을 M_{10}이라 하자. (a) $P\big(10(1-M_{10})>2\big)$을 정확히 구하라. (b) 극한분포 $\mathrm{Exp}(1)$로 근사한 값을 구하라.

풀이 (a) $P(M_{10}<0.8)=0.8^{10}=0.1074$이다. 독립이므로 열 개의 확률을 곱했다. (b) $P(\mathrm{Exp}(1)>2)=e^{-2}=0.1353$이다. $(1-\frac{x}{n})^n<e^{-x}$이므로 유한한 n의 값은 극한보다 작다.

최종답 (a) $0.8^{10}=0.1074$ (b) $e^{-2}=0.1353$

A5 Poisson 평균의 제곱근 ★

정리 4.4 · 6분

$X_1,\dots,X_n\overset{\text{iid}}{\sim}\mathrm{Poi}(\lambda)$이다. $\sqrt{n}\big(\sqrt{\bar X_n}-\sqrt{\lambda}\big)$의 극한분포의 분산을 구하라.

풀이 $\sqrt{n}(\bar X_n-\lambda)\xrightarrow{D}\mathcal{N}(0,\lambda)$이고 $g(x)=\sqrt{x}$, $g'(\lambda)=\frac{1}{2\sqrt{\lambda}}$이다. 델타 방법으로 극한분포의 분산은 $g'(\lambda)^2\lambda=\frac{1}{4\lambda}\cdot\lambda=\frac{1}{4}$이다. λ와 상관없으므로 분산 안정화 변환이다(노트 예제 4.7).

최종답 $\frac{1}{4}$

A6 지수분포의 비율 모수 ★

정리 4.4 · 8분

$X_1,\dots,X_n$이 iid이고 밀도가 $\lambda e^{-\lambda x}$ $(x>0)$이다. $\hat\lambda=1/\bar X_n$에 대해 $\sqrt{n}(\hat\lambda-\lambda)$의 극한분포의 분산을 λ의 식으로 구하라. (노트북의 `check()`에는 λ를 `lam`으로 쓴다.)

풀이 $\sqrt{n}(\bar X_n-\frac{1}{\lambda})\xrightarrow{D}\mathcal{N}(0,\frac{1}{\lambda^2})$이고 $g(x)=\frac{1}{x}$, $g'(\frac{1}{\lambda})=-\lambda^2$이다. 델타 방법으로 분산은 $\lambda^4\cdot\frac{1}{\lambda^2}=\lambda^2$이다.

최종답 λ^2

A7 Monte Carlo의 표본 크기 ★

정리 6.1 · 10분

$\theta=\int_0^1\sqrt{1-x^2}\,dx=\frac{\pi}{4}$를 $U_i\overset{\text{iid}}{\sim}\mathrm{U}(0,1)$과 $h(u)=\sqrt{1-u^2}$의 평균으로 추정한다. (a) $\sigma_h^2=\mathrm{Var}\,h(U)$를 정확히 구하라. (b) 표준오차 $\sigma_h/\sqrt{n}$이 0.001 이하가 되는 가장 작은 n을 구하라.

풀이 (a) $Eh(U)^2=\int_0^1(1-u^2)\,du=\frac{2}{3}$이므로 $\sigma_h^2=\frac{2}{3}-\frac{\pi^2}{16}=0.04982$이다. (b) $\sigma_h/\sqrt{n}\le 0.001$이면 $n\ge\sigma_h^2/10^{-6}=49{,}816.4$이므로 $n=49{,}817$이다. 노트 연습 6.1의 $4\cdot 1\{U_1^2+U_2^2\le 1\}$은

같은 $\frac{\pi}{4}\cdot 4$를 분산 2.70으로 추정한다. h를 쓰면 $4h$의 분산이 $16\cdot 0.0498 = 0.80$으로 세 배 넘게 작다.

최종답 (a) $\frac{2}{3} - \frac{\pi^2}{16} \approx 0.04982$ (b) 49,817

A8 베타 난수의 기각 표집 ★

알고리즘 6.3, 정리 6.4 · 8분

$f(x) = 6x(1-x)$ $(0<x<1)$를 제안분포 $\mathrm{U}(0,1)$로 기각 표집한다. (a) 쓸 수 있는 가장 작은 M은? (b) 한 번 시도에서 받아들일 확률은? (c) 정확히 세 번째 시도에서 처음 받아들일 확률은?

풀이 (a) $g=1$이므로 $M = \max f = f(\frac{1}{2}) = \frac{3}{2}$이다. (b) 정리 6.4로 $1/M = \frac{2}{3}$이다. (c) 처음 두 번 거절하고 세 번째에 받아들이므로 $(\frac{1}{3})^2\cdot\frac{2}{3} = \frac{2}{27}$이다.

최종답 (a) $\frac{3}{2}$ (b) $\frac{2}{3}$ (c) $\frac{2}{27}$

A9 평균의 제곱 ★

정리 4.4 · 6분

$X_i \overset{\text{iid}}{\sim} \mathrm{Exp}(1)$이면 $\sqrt{n}(\bar X_n^2 - 1)$의 극한분포의 분산은 얼마인가?

풀이 $\sqrt{n}(\bar X_n - 1) \xrightarrow{D} \mathcal{N}(0,1)$, $g(x)=x^2$, $g'(1)=2$이므로 분산은 $2^2\cdot 1 = 4$이다. $\mu=0$이었다면 $g'(0)=0$이어서 $\sqrt{n}$의 속도로는 퇴화하고 $n\bar X_n^2$을 봐야 한다(노트 연습 4.2).

최종답 4

A10 로그 오즈의 분산 ★

예제 4.5 · 6분

$p=0.2$, $n=100$일 때 $\log\frac{\hat p}{1-\hat p}$의 근사 분산을 델타 방법으로 구하라.

풀이 $g(p)=\log\frac{p}{1-p}$, $g'(p)=\frac{1}{pq}$이므로 근사 분산은 $\frac{pq}{n}\cdot\frac{1}{p^2q^2} = \frac{1}{npq} = \frac{1}{100\cdot 0.16} = \frac{1}{16} = 0.0625$이다. 표준오차는 0.25다.

최종답 $\frac{1}{16} = 0.0625$

A11 세 범주의 비율 ★★

예제 5.3 · 10분

세 범주의 확률이 $\mathbf{p}=(0.5,0.3,0.2)$이고 n번의 독립 시행에서 비율 $\hat p_1,\hat p_2,\hat p_3$을 얻는다. (a) $\sqrt{n}(\hat p_1 - \hat p_2 - 0.2)$의 극한분포의 분산을 구하라. (b) $\sqrt{n}\hat p_1$과 $\sqrt{n}\hat p_2$의 극한 공분산을 구하라.

풀이 $\Sigma_{\mathbf{p}}$의 성분은 $\Sigma_{11}=0.5\cdot 0.5=0.25$, $\Sigma_{22}=0.3\cdot 0.7=0.21$, $\Sigma_{12}=-0.5\cdot 0.3=-0.15$이다. (a) $\mathbf{a}=(1,-1,0)$이면 $\mathbf{a}'\Sigma\mathbf{a} = 0.25+0.21-2(-0.15) = 0.76$이다. (b) -0.15이다. 한

범주의 비율이 크면 나머지는 작아질 수밖에 없어 공분산이 음수다.

최종답 (a) 0.76 (b) −0.15

A12 지수 꼬리의 중요도 표집 ★★

정리 6.6 · 12분

$X \sim \mathrm{Exp}(1)$에서 $\theta = P(X > 3) = e^{-3}$을 추정한다. 제안분포는 평균 3인 지수분포 $g(y) = \frac{1}{3}e^{-y/3}$이고 $h = 1\{y > 3\}$이다. (a) 가중치 $w = f/g$를 쓰고 $E_g[h^2 w^2]$을 구하라. (b) 한 표본의 분산의 비 $\frac{\theta(1-\theta)}{E_g[h^2w^2]-\theta^2}$를 구하라(소수 둘째 자리까지).

풀이 (a) $w(y) = \frac{e^{-y}}{\frac{1}{3}e^{-y/3}} = 3e^{-2y/3}$이고

$$E_g[h^2w^2] = \int_3^\infty 9e^{-4y/3}\cdot\tfrac{1}{3}e^{-y/3}\,dy = 3\int_3^\infty e^{-5y/3}dy = 3\cdot\tfrac{3}{5}e^{-5} = \tfrac{9}{5}e^{-5} = 0.01213.$$

(b) 중요도 표집 한 표본의 분산은 $0.01213 - e^{-6} = 0.00965$이고, 보통 Monte Carlo는 $e^{-3}(1-e^{-3}) = 0.04731$이므로 비는 4.90이다. 같은 정확도에 표본이 다섯 배쯤 적게 든다.

최종답 (a) $\frac{9}{5}e^{-5} \approx 0.01213$ (b) ≈ 4.90

A13 Berry–Esseen 한계 ★

3.3절의 상수 · 6분

공정한 동전(Bernoulli($\frac{1}{2}$)) $n = 100$번의 표준화한 합에 대해 Berry–Esseen 상한 $C\rho/(\sigma^3\sqrt{n})$을 $C = 0.4748$로 구하라.

풀이 $|X - \frac{1}{2}| = \frac{1}{2}$이므로 $\rho = \frac{1}{8}$, $\sigma = \frac{1}{2}$, $\rho/\sigma^3 = 1$이다. 상한은 $0.4748/\sqrt{100} = 0.04748$이다. 실제 최대 오차는 이항의 점프 때문에 약 0.04로 상한에 가깝다. 이산분포는 분포함수가 뛰므로 연속성 보정 없이는 오차가 $1/\sqrt{n}$보다 빨리 줄지 않는다.

최종답 0.04748

A14 반대 변량의 이득 ★★

예제 6.8 · 10분

$h(u) = u^2$, $U \sim \mathrm{U}(0,1)$이다. (a) $\mathrm{Var}\,h(U)$, (b) $\mathrm{Cov}(h(U), h(1-U))$, (c) $\mathrm{Var}\,\frac{h(U)+h(1-U)}{2}$를 구하라. 독립인 난수 두 개의 평균과 비교하면 분산이 몇 배 줄었나?

풀이 (a) $\frac{1}{5} - \frac{1}{9} = \frac{4}{45}$. (b) $1-U$도 $\mathrm{U}(0,1)$이므로 $Eh(1-U) = \frac{1}{3}$이고 $\mathrm{Cov} = \frac{1}{30} - \frac{1}{9} = -\frac{7}{90}$. (c) $\frac{1}{2}\left(\frac{4}{45} - \frac{7}{90}\right) = \frac{1}{2}\cdot\frac{1}{90} = \frac{1}{180}$. 독립인 두 난수의 평균은 $\frac{4}{45}\cdot\frac{1}{2} = \frac{8}{180}$이므로 8배 줄었다. h가 $[0,1]$에서 증가함수이므로 공분산이 음수다(노트 연습 6.5).

최종답 (a) $\frac{4}{45}$ (b) $-\frac{7}{90}$ (c) $\frac{1}{180}$

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, 이 회차의 정리 번호, 도구 번호를 쓴다.

B1 중심극한정리, 근거 채우기 ★★

① · 정리 3.1 · 20분

iid이고 mgf가 $|t| < h$에서 존재하며 평균 μ, 분산 $\sigma^2 > 0$이면 $Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{D} \mathcal{N}(0,1)$이다. 각 단계의 근거를 써라.

① $Y_i = (X_i - \mu)/\sigma$는 iid, 평균 0, 분산 1이고 mgf가 $|s| < \sigma h$에서 존재한다.
2회 정리 3.6 (아핀변환의 mgf)

② $\psi = \log M_Y$에 대해 $\psi(0) = 0$, $\psi'(0) = 0$, $\psi''(0) = 1$이다. 2회 정리 3.3, 따름정리 3.4

③ $\log M_{Z_n}(t) = n\,\psi(t/\sqrt{n})$이다. 2회 정리 3.6, 독립(3회 정리 3.4)

④ $\psi(s) = \frac{1}{2}\psi''(\xi_n)s^2$인 ξ_n이 0과 $s = t/\sqrt{n}$ 사이에 있다. Taylor 정리 [T8], ②

⑤ $n\psi(t/\sqrt{n}) = \frac{t^2}{2}\psi''(\xi_n) \to \frac{t^2}{2}$이다. $\xi_n \to 0$, ψ''는 0에서 연속

⑥ $M_{Z_n}(t) \to e^{t^2/2}$이므로 $Z_n \xrightarrow{D} \mathcal{N}(0,1)$이다. 2.3절의 상자, 5회 정리 1.8(ii)

풀이 빈칸의 빨간 글씨가 근거다. 노트 정리 3.1의 증명과 같다. ④에서 $s = t/\sqrt{n}$이 ψ의 정의역 $(-\sigma h, \sigma h)$ 안에 있어야 하는데, $|t| < \sigma h$이면 모든 $n \geq 1$에서 그렇다.

B2 확률수렴은 분포수렴, 근거 채우기 ★★

① · 정리 2.3(a) · 15분

$X_n \xrightarrow{P} X$이면 $X_n \xrightarrow{D} X$이다. 근거를 써라.

① $\{X_n \leq x\} \subset \{X \leq x + \varepsilon\} \cup \{|X_n - X| \geq \varepsilon\}$이다.
원소 추적 [T1]: $X_n \leq x$, $|X_n - X| < \varepsilon$이면 $X < x + \varepsilon$

② $\{X \leq x - \varepsilon\} \subset \{X_n \leq x\} \cup \{|X_n - X| \geq \varepsilon\}$이다. 같은 원소 추적

③ $F(x-\varepsilon) - P(|X_n - X| \geq \varepsilon) \leq F_n(x) \leq F(x+\varepsilon) + P(|X_n - X| \geq \varepsilon)$이다. 1회 정리 4.3(d)(g)

④ $F(x-\varepsilon) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x+\varepsilon)$이다. 확률수렴, [T5]

⑤ x가 F의 연속점이면 $\lim_n F_n(x) = F(x)$이다. $\varepsilon \downarrow 0$, 연속점의 정의

풀이 빈칸의 빨간 글씨가 근거다. ⑤에서 연속점이 필요한 것은 $F(x-\varepsilon) \to F(x-)$이기 때문이다. 오른쪽 끝 $F(x+\varepsilon) \to F(x)$는 오른쪽 연속성으로 언제나 성립한다.

B3 델타 방법, 근거 채우기 ★★

① · 정리 4.4 · 15분

$\sqrt{n}(Y_n-\theta) \xrightarrow{D} \mathcal{N}(0,\sigma^2)$이고 g가 θ에서 미분가능하면 $\sqrt{n}(g(Y_n)-g(\theta)) \xrightarrow{D} \mathcal{N}(0, g'(\theta)^2\sigma^2)$이다. 근거를 써라.

① $h(y)=\frac{g(y)-g(\theta)}{y-\theta}$ $(y\neq\theta)$, $h(\theta)=g'(\theta)$는 θ에서 연속이다. — 미분의 정의

② $Y_n-\theta=\frac{1}{\sqrt{n}}\cdot\sqrt{n}(Y_n-\theta) \xrightarrow{P} 0$이다. — 정리 2.5 (확률적 유계 × 0으로 가는 수)

③ $h(Y_n) \xrightarrow{P} g'(\theta)$이다. — 연속사상 정리 1.3(a)

④ $\sqrt{n}(g(Y_n)-g(\theta)) = h(Y_n)\cdot\sqrt{n}(Y_n-\theta)$이다. — h의 정의, $y=\theta$에서도 양변 0

⑤ 극한은 $g'(\theta)\cdot\mathcal{N}(0,\sigma^2)=\mathcal{N}(0,g'(\theta)^2\sigma^2)$이다. — Slutsky 정리 4.2, 5회 정리 3.8

풀이 빈칸의 빨간 글씨가 근거다. 미분가능성만 쓰고 g'의 연속성은 쓰지 않았다. ③은 h가 θ 한 점에서 연속이면 충분하다.

B4 중심극한정리, 단계의 이름만 ★★

② · 정리 3.1 · 30분

정리 3.1을 증명하라. 단계의 이름만 주어진다. 각 단계 아래에 식과 근거를 써라.

① **표준화한다**

② **로그 mgf의 0에서의 값과 두 도함수를 적는다**

③ **독립합의 mgf로 $\log M_{Z_n}$을 쓴다**

④ **나머지항이 있는 Taylor 전개를 한다**

⑤ **n을 곱한 것의 극한을 구한다**

⑥ **mgf의 극한에서 분포수렴으로 간다**

풀이 B1의 여섯 단계와 같다. 핵심 식은 $n\psi(t/\sqrt{n})=\frac{t^2}{2}\psi''(\xi_n)$이다. 넷째 단계에서 상수항은 $\psi(0)=\log 1=0$, 1차 항은 $\psi'(0)=EY=0$으로 사라진다. 평균을 빼지 않으면 1차 항 $\sqrt{n}\,t\,\mu/\sigma$가 발산한다. 분산 1로 나누지 않으면 극한은 $e^{\sigma^2t^2/2}$, 곧 $\mathcal{N}(0,\sigma^2)$이다.

B5 두 수렴의 관계, 단계의 이름만 ★★

② · 정리 2.3 · 25분

정리 2.3의 (a)와 (b)를 증명하라.

① **(a) 두 포함관계를 쓴다**

② **(a) 확률을 취해 $F_n(x)$를 끼운다**

③ **(a) 상극한·하극한을 취하고 $\varepsilon \downarrow 0$**

④ **(b) 상수의 분포함수와 그 연속점을 적는다**

⑤ **(b) $\{|X_n - c| \ge \varepsilon\}$을 분포함수로 덮는다**

⑥ **(b) 극한을 취한다**

풀이 (a)는 B2의 풀이와 같다. (b): 상수 c의 분포함수는 c에서만 뛰므로 c가 아닌 모든 점이 연속점이다. $\{|X_n - c| \ge \varepsilon\} \subset \{X_n \le c - \varepsilon\} \cup \{X_n > c + \frac{\varepsilon}{2}\}$이므로 $P(|X_n - c| \ge \varepsilon) \le F_n(c - \varepsilon) + 1 - F_n(c + \frac{\varepsilon}{2}) \to 0 + 1 - 1 = 0$이다.

B6 중심극한정리, 전략만 ★★★

③ · 정리 3.1 · 35분

정리 3.1을 증명하라. 정의역에 관한 조건 확인도 함께 쓴다.

전략 표준화한 합의 mgf의 로그를 $n\psi(t/\sqrt{n})$으로 쓰고, ψ를 0 근처에서 2차까지 전개한다. 2차 항만 n을 곱해도 남는다.

풀이 노트 정리 3.1의 단계형 증명과 조건 확인 단락이 답이다. 근거는 B1의 풀이에 있다.

B7 델타 방법, 전략만 ★★

③ · 정리 4.4 · 30분

정리 4.4를 증명하라. 그리고 $g(y) = |y|$, $\theta = 0$이면 증명의 어느 단계가 무너지는지 밝혀라.

전략 차분몫 $h(y) = \frac{g(y) - g(\theta)}{y - \theta}$를 확률변수 계수로 보고, 그 계수가 $g'(\theta)$로 확률수렴함을 보인 뒤 Slutsky를 쓴다.

풀이 증명은 B3의 다섯 단계와 같다. $g(y) = |y|$, $\theta = 0$이면 $h(y) = \text{sign}(y)$이고 $h(0)$을 어떻게 정해도 0에서 연속이 아니다. 따라서 ①과 ③이 무너진다. 실제로 $\sqrt{n}|Y_n| \xrightarrow{D} \sigma|Z|$이고(연속사상), 이것은 정규가 아니라 반정규분포다(C5).

C

증명 문제

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

C1 극한이 확률변수인 사칙연산 ★★

[T1], [T4] · 정리 1.3, 2.5 · 15분

$X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$이다(X, Y는 확률변수). (a) $X_n + Y_n \xrightarrow{P} X + Y$임을 보여라. (b) $X_n Y_n \xrightarrow{P} XY$임을 보여라. (힌트: $X_n Y_n - XY = (X_n - X)Y_n + X(Y_n - Y)$.)

풀이 (a) 삼각부등식으로 $|X_n + Y_n - (X+Y)| \le |X_n - X| + |Y_n - Y|$이므로 $\{|X_n + Y_n - X - Y| \ge \varepsilon\} \subset \{|X_n - X| \ge \frac{\varepsilon}{2}\} \cup \{|Y_n - Y| \ge \frac{\varepsilon}{2}\}$이다. Boole 부등식으로 확률이 0으로 간다.
(b) $Y_n \xrightarrow{P} Y$이면 정리 2.3(a)로 $Y_n \xrightarrow{D} Y$이고, 정리 2.5로 $\{Y_n\}$은 확률적으로 유계다. $X_n - X \xrightarrow{P} 0$이므로 정리 2.5의 둘째 주장으로 $(X_n - X)Y_n \xrightarrow{P} 0$이다. 같은 이유로 상수 수열 $X, X, \ldots$는 확률적으로 유계이므로($P(|X| \le B) \to 1$, 1회 정리 6.3(b)) $X(Y_n - Y) \xrightarrow{P} 0$이다. (a)로 합도 0으로 확률수렴한다.

C2 평균제곱수렴과 확률수렴 ★★

[T2] · 2회 따름정리 5.2 · 15분

(a) $E(Y_n - c)^2 \to 0$이면 $Y_n \xrightarrow{P} c$임을 보여라. (b) $P(Y_n = \sqrt{n}) = \frac{1}{n}$, $P(Y_n = 0) = 1 - \frac{1}{n}$이면 $Y_n \xrightarrow{P} 0$이고 $EY_n \to 0$이지만 EY_n^2은 0으로 가지 않음을 보여라. 따라서 (a)의 역은 거짓이다.

풀이 (a) 2회 따름정리 5.2로 $P(|Y_n - c| \ge \varepsilon) \le \frac{E(Y_n - c)^2}{\varepsilon^2} \to 0$이다.
(b) $0 < \varepsilon$이면 $P(|Y_n| \ge \varepsilon) \le P(Y_n = \sqrt{n}) = \frac{1}{n} \to 0$이므로 $Y_n \xrightarrow{P} 0$이다. $EY_n = \sqrt{n} \cdot \frac{1}{n} = \frac{1}{\sqrt{n}} \to 0$이지만 $EY_n^2 = n \cdot \frac{1}{n} = 1$이다. 드문 큰 값이 평균에는 사라지지만 제곱의 평균에는 끝까지 남는다.

C3 유계이면 기댓값도 따라온다 ★★

[T4] · 정의 1.2 · 15분

$Y_n \xrightarrow{P} c$이고 모든 n에서 $|Y_n| \le K$이면 $EY_n \to c$임을 보여라. 유계라는 가정이 없으면 왜 성립하지 않는지 노트 연습 1.3의 예로 설명하라.

풀이 $A_n = \{|Y_n - c| \ge \varepsilon\}$이라 하면

$$|EY_n - c| \le E|Y_n - c| = E\big(|Y_n - c|; A_n^c\big) + E\big(|Y_n - c|; A_n\big) \le \varepsilon + (K + |c|)P(A_n).$$

첫 부등식은 $|EZ| \le E|Z|$(2회 정리 5.9의 Jensen이나 단조성)이다. $n \to \infty$에서 상극한은 ε 이하이고 ε은 임의이므로 $EY_n \to c$이다. 노트 연습 1.3의 Y_n(n 또는 0)은 $Y_n \xrightarrow{P} 0$이지만 $EY_n = 1$이다. 둘째 항 $(K+|c|)P(A_n)$에서 K 자리에 n이 들어가 $n \cdot \frac{1}{n} = 1$이 남는다.

C4 독립인 두 수열의 합 ★★

2회 정리 3.6 · 2.3절의 상자 · 20분

각 n에서 X_n과 Y_n이 독립이고, mgf가 0 근방에서 $M_{X_n} \to M_X$, $M_{Y_n} \to M_Y$로 수렴한다(M_X, M_Y는 X, Y의 mgf). (a) $X_n + Y_n \xrightarrow{D} X' + Y'$임을 보여라. 여기서 X', Y'은 X, Y와 각각 같은 분포이고 서로 독립이다. (b) 노트 연습 2.5의 $Y_n = -X_n$이 (a)의 반례가 아닌 이유는?

풀이 (a) 독립이므로 $M_{X_n+Y_n}(t) = M_{X_n}(t)M_{Y_n}(t) \to M_X(t)M_Y(t)$이다(2회 정리 3.6). 오른쪽은 독립인 X', Y'의 합의 mgf다. 2.3절의 상자로 $X_n + Y_n \xrightarrow{D} X' + Y'$이다. (b) $Y_n = -X_n$은 X_n과 독립이 아니다(X_n이 상수가 아닌 한). 극한분포만으로는 결합분포가 정해지지 않고, (a)는 독립이라는 결합분포의 정보를 더해 합의 극한을 정한다.

C5 미분가능하지 않으면 ★★

[T2] · 정리 4.4 · 20분

$\sqrt{n}Y_n \xrightarrow{D} \mathcal{N}(0, \sigma^2)$이고 $g(y) = |y|$이다($\theta = 0$). (a) $P(\sqrt{n}|Y_n| \le y) = F_n(y) - P(\sqrt{n}Y_n < -y)$로 $\sqrt{n}|Y_n| \xrightarrow{D} \sigma|Z|$임을 보여라($F_n$은 $\sqrt{n}Y_n$의 분포함수). (b) 극한의 평균 $E|Z|$을 구하라. (c) 극한이 정규분포가 아님을 보이고, 정리 4.4의 어느 가정이 깨졌는지 밝혀라.

풀이 (a) $y > 0$이면 $\{\sqrt{n}|Y_n| \le y\} = \{-y \le \sqrt{n}Y_n \le y\}$이므로 식이 성립한다. 극한 $\mathcal{N}(0, \sigma^2)$은 연속이므로 모든 점이 연속점이고, $F_n(-y-\delta) \le P(\sqrt{n}Y_n < -y) \le F_n(-y)$에서 $\delta \downarrow 0$으로 $P(\sqrt{n}Y_n < -y) \to \Phi(-y/\sigma)$이다. 따라서 $P(\sqrt{n}|Y_n| \le y) \to \Phi(y/\sigma) - \Phi(-y/\sigma) = P(\sigma|Z| \le y)$이다. (b) $E|Z| = 2\int_0^\infty z\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = \frac{2}{\sqrt{2\pi}} = \sqrt{2/\pi} \approx 0.798$이다. (c) $\sigma|Z| \ge 0$이므로 확률 1로 음이 아니다. 평균이 0이 아닌 양수 $\sigma\sqrt{2/\pi}$이고 음수 쪽 확률이 0인 정규분포는 없다. 깨진 가정은 g가 $\theta = 0$에서 미분가능하다는 것이다. 0에서 왼쪽 미분계수는 -1, 오른쪽은 1이다.

최종답 (b) $\sqrt{2/\pi} \approx 0.798$

C6 Wald 구간의 포함확률 ★★★

정리 4.2, 예제 2.2(c) · 25분

성공 확률 p인 독립 시행 n번의 비율 $\hat{p}$로 만든 구간 $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$을 Wald 구간이라 한다. (a) $p \in (0,1)$이 고정이면 포함확률이 $1-\alpha$로 감을 보여라. (b) $p = p_n = 1/n$이면 $P(\hat{p} = 0) \to e^{-1}$임을 보이고, 포함확률의 상극한이 $1 - e^{-1}$ 이하임을 보여라.

풀이 (a) 중심극한정리로 $\frac{\hat p - p}{\sqrt{pq/n}} \xrightarrow{D} \mathcal{N}(0,1)$이다. 대수의 법칙으로 $\hat p \xrightarrow{P} p$이고, $x \mapsto \sqrt{pq/(x(1-x))}$는 $p \in (0,1)$에서 연속이므로 정리 1.3으로 첫 인수는 1로 확률수렴한다. Slutsky로 스튜던트화한 비가 $\mathcal{N}(0,1)$로 분포수렴하고, $\pm z_{\alpha/2}$는 연속점이므로 포함확률 $P(|\cdot| \le z_{\alpha/2}) \to 1-\alpha$이다.

(b) $n\hat p \sim \mathrm{b}(n, 1/n)$이므로 $P(\hat p = 0) = (1-\frac{1}{n})^n \to e^{-1}$이다(예제 2.2(c)의 Poisson 극한과 같다). $\hat p = 0$이면 구간의 폭이 0이어서 구간은 $\{0\}$이고 $p_n = \frac{1}{n} > 0$을 포함하지 않는다. 따라서 포함확률은 $1 - P(\hat p = 0)$ 이하이고 상극한은 $1 - e^{-1} = 0.632$ 이하다. (a)의 극한은 p를 고정하고 $n \to \infty$로 보낸 것이어서, p가 n과 함께 0에 다가가는 경우는 보장하지 않는다.

최종답 (b) $1 - e^{-1} \approx 0.632$

C7 정규화 상수를 모를 때의 중요도 표집 ★★

정리 1.3, 1.6, 6.6 · 20분

목표 밀도가 $f = \tilde f / c$이고 $\tilde f$는 알지만 상수 c는 모른다. $Y_i \overset{\text{iid}}{\sim} g$, $\tilde w = \tilde f / g$에 대해

$$\hat\theta_n = \frac{\sum_{i=1}^n h(Y_i)\tilde w(Y_i)}{\sum_{i=1}^n \tilde w(Y_i)}$$

로 둔다($E_f|h| < \infty$, $\tilde f > 0$인 곳에서 $g > 0$). (a) $\hat\theta_n \xrightarrow{P} E_f h(X)$임을 보여라. (b) $n = 1$이면 $\hat\theta_1$의 기댓값은 무엇인가? 이 추정량은 일반적으로 불편인가?

풀이 (a) $E_g[\tilde w(Y)] = \int \tilde f = c$이고 $E_g[h\tilde w] = \int h\tilde f = c\,E_f h$이며, 둘 다 절댓값의 기댓값이 유한하다. Khinchin(정리 1.6)으로 $\frac{1}{n}\sum h\tilde w \xrightarrow{P} cE_f h$, $\frac{1}{n}\sum \tilde w \xrightarrow{P} c > 0$이다. 비는 정리 1.3(따름정리 1.4)으로 $E_f h$로 확률수렴한다. (b) $\hat\theta_1 = h(Y_1)$이므로 $E\hat\theta_1 = E_g h(Y)$이고, 이것은 $g \neq f$이면 일반적으로 $E_f h$와 다르다. 분모가 확률변수이므로 비의 기댓값은 기댓값의 비가 아니다. 이 추정량은 불편이 아니지만 일치추정량이다.

C8 독립이 아니어도 대수의 법칙 ★★

[T5] · 2회 정리 5.3 · 20분

$X_1, X_2, \ldots$의 평균이 모두 μ이고 분산이 모두 $v < \infty$이다. (a) 쌍마다 상관이 0이면 $\bar X_n \xrightarrow{P} \mu$임을 보여라. (b) $\mathrm{Cov}(X_i, X_j) = v\rho^{|i-j|}$ ($|\rho| < 1$)이면 $\mathrm{Var}\,\bar X_n \le \frac{v}{n}\cdot\frac{1+|\rho|}{1-|\rho|}$임을 보이고 $\bar X_n \xrightarrow{P} \mu$를 결론지어라. (c) 모든 i에서 $X_i = Z$이면 무엇이 깨지는가?

풀이 (a) $\mathrm{Var}\,\bar X_n = \frac{1}{n^2}\sum_i \mathrm{Var}\,X_i = \frac{v}{n}$이므로 Chebyshev로 $P(|\bar X_n - \mu| \ge \varepsilon) \le \frac{v}{n\varepsilon^2} \to 0$이다. 독립은 쓰지 않았고 공분산이 0이라는 것만 썼다. (b) $\mathrm{Var}\,\bar X_n = \frac{v}{n^2}\sum_{i,j}\rho^{|i-j|} \le \frac{v}{n^2}\sum_i\sum_j|\rho|^{|i-j|} \le \frac{v}{n^2}\cdot n\left(1 + \frac{2|\rho|}{1-|\rho|}\right) = \frac{v}{n}\cdot\frac{1+|\rho|}{1-|\rho|}$이다. 기하급수를 썼다 [T5]. Chebyshev로 결론이 나온다. (c) $\bar X_n = Z$이므로 Z가 상수가 아니면 $\bar X_n$은 μ로 가지 않는다. $\mathrm{Cov}(X_i, X_j) = v$로 상관이 1이어서 $\mathrm{Var}\,\bar X_n = v$가 줄지 않는다. 같은 정보를 n번 본 것은 정보가 하나다.

C9 기각 표집의 상수와 꼬리 ★★

정리 6.4 · 20분

(a) 알고리즘 6.3이 정당하려면 $M \geq \sup_x f(x)/g(x)$이어야 하고, 받아들일 확률은 $M = \sup_x f/g$에서 가장 큼을 보여라. (b) 목표 $f = \mathcal{N}(0,1)$, 제안 $g =$ 표준 Cauchy이면 가장 작은 M이 $\sqrt{2\pi/e}$임을 보여라. (c) 목표가 Cauchy, 제안이 $\mathcal{N}(0,1)$이면 쓸 수 있는 M이 없음을 보여라.

풀이 (a) $f(y) > Mg(y)$인 y의 집합의 확률이 양수이면 받아들이는 확률 $\min(1, f/(Mg))$이 그 위에서 1로 잘리고, 정리 6.4의 증명 ②의 등호가 깨진다. 받아들일 확률은 $1/M$이므로 쓸 수 있는 가장 작은 M에서 가장 크다. (b) $\frac{f}{g} = \frac{e^{-x^2/2}/\sqrt{2\pi}}{1/(\pi(1+x^2))} = \sqrt{\frac{\pi}{2}}(1+x^2)e^{-x^2/2}$이다. $(1+x^2)e^{-x^2/2}$의 도함수는 $x(1-x^2)e^{-x^2/2}$이므로 $x = \pm 1$에서 최대 $2e^{-1/2}$이다. 따라서 $M = \sqrt{\pi/2}\cdot 2e^{-1/2} = \sqrt{2\pi/e} \approx 1.520$이고 받아들일 확률은 0.658이다. (c) $\frac{f}{g} = \frac{\sqrt{2\pi}e^{x^2/2}}{\pi(1+x^2)} \to \infty$ $(|x| \to \infty)$이므로 상한이 없다. 제안분포의 꼬리가 목표보다 얇으면 기각 표집을 쓸 수 없다. 노트 연습 6.4의 중요도 표집과 같은 조건이다.

최종답 (b) $\sqrt{2\pi/e} \approx 1.520$

D

퀄형

여러 절과 앞 회차의 도구를 잇는 종합 문제다. 머리의 제한 시간을 재고, 노트를 덮고 푼다. 어느 극한 정리를 어떤 가정 아래 쓰는지 밝히는 것까지 답안의 일부다.

D1 Poisson에서 0의 확률 ★★★

정리 1.6, 3.1, 4.4 · 4회 · 제한 35분

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$에서 $\theta = P(X=0) = e^{-\lambda}$를 추정한다. 두 추정량

$$T_1 = \frac{1}{n}\#\{i : X_i = 0\}, \qquad T_2 = e^{-\bar{X}_n}$$

을 비교한다. (노트북의 `check()`에는 λ를 `lam`으로 쓴다.)

(a) T_1과 T_2가 모두 θ로 확률수렴함을 보여라.

(b) $\sqrt{n}(T_1 - \theta)$의 극한분포의 분산을 구하라.

(c) $\sqrt{n}(T_2 - \theta)$의 극한분포의 분산을 구하라.

(d) 점근분산의 비 $\frac{\text{(c)}}{\text{(b)}}$가 모든 $\lambda > 0$에서 1보다 작음을 보이고, $\lambda = 1$에서의 값을 구하라.

풀이 (a) $1\{X_i = 0\}$은 iid Bernoulli(θ)이므로 대수의 법칙으로 $T_1 \xrightarrow{P} \theta$이다. $\bar{X}_n \xrightarrow{P} \lambda$이고 e^{-x}는 연속이므로 정리 1.3으로 $T_2 \xrightarrow{P} e^{-\lambda}$이다. (b) 중심극한정리로 $\sqrt{n}(T_1 - \theta) \xrightarrow{D} \mathcal{N}(0, \theta(1-\theta))$이므로 분산은 $e^{-\lambda}(1-e^{-\lambda})$이다. (c) $\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{D} \mathcal{N}(0, \lambda)$(4회의 Poisson 분산), $g(x) = e^{-x}$, $g'(\lambda) = -e^{-\lambda}$이므로 분산은 $\lambda e^{-2\lambda}$이다. (d) $\frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1-e^{-\lambda})} = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} = \frac{\lambda}{e^{\lambda}-1}$이다. $e^{\lambda} > 1 + \lambda$ ($\lambda > 0$)이므로 1보다 작다. $\lambda = 1$이면 $\frac{1}{e-1} = 0.582$이다. T_1은 0의 개수만 쓰고 나머지 관측값을 버리지만, T_2는 모든 관측을 $\bar{X}_n$으로 쓴다. 대가는 T_2가 Poisson 모형을 가정한다는 것이다. 모형이 틀리면 T_2는 $e^{-\lambda}$가 아닌 값으로 간다.

최종답 (b) $e^{-\lambda}(1-e^{-\lambda})$ (c) $\lambda e^{-2\lambda}$ (d) $\frac{1}{e-1} \approx 0.582$

D2 균등분포의 끝점 ★★★

예제 2.2(b), 정리 3.1 · 3회 · 제한 40분

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

(a) $x \ge 0$에서 $P\big(n(\theta - M_n) > x\big) \to e^{-x/\theta}$ 임을 보여라.

(b) $2\bar X_n$의 분산을 구하라.

(c) $T = \frac{n+1}{n} M_n$이 불편이고 분산이 $\frac{\theta^2}{n(n+2)}$ 임을 보여라.

(d) $n = 10$에서 $\mathrm{Var}(2\bar X_n)/\mathrm{Var}\,T$를 구하라. 두 추정량의 오차는 n에 대해 각각 어떤 속도로 줄어드는가?

풀이 (a) $x < n\theta$이면 $P(n(\theta - M_n) > x) = P(M_n < \theta - \frac{x}{n}) = \big(1 - \frac{x}{n\theta}\big)^n \to e^{-x/\theta}$이다(독립이므로 n개의 확률을 곱한다). 극한은 평균 θ인 지수분포다. (b) $\mathrm{Var}\,X = \frac{\theta^2}{12}$이므로 $\mathrm{Var}(2\bar X_n) = \frac{4}{n}\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n}$이다. (c) $U = M_n/\theta$의 분포함수는 u^n, 밀도는 nu^{n-1}이다. $EU = \frac{n}{n+1}$, $EU^2 = \frac{n}{n+2}$이므로 $\mathrm{Var}\,U = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n}{(n+1)^2(n+2)}$이다. 따라서 $ET = \frac{n+1}{n}\theta \cdot \frac{n}{n+1} = \theta$이고 $\mathrm{Var}\,T = \frac{(n+1)^2}{n^2}\theta^2 \cdot \frac{n}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}$이다. (d) $n = 10$이면 $\frac{\theta^2/30}{\theta^2/120} = 4$이다. $2\bar X_n$의 표준편차는 $1/\sqrt{n}$, T의 표준편차는 $1/n$의 속도로 준다. (a)의 $n(\theta - M_n)$이 수렴한다는 것이 곧 M_n의 오차가 $1/n$의 크기라는 뜻이다. 표본의 받침이 모수에 의존하면 중심극한정리의 $\sqrt{n}$보다 빠를 수 있다.

최종답 (b) $\frac{\theta^2}{3n}$ (c) $\frac{\theta^2}{n(n+2)}$ (d) 4

D3 기하분포의 성공 확률 ★★★

정리 3.1, 4.2, 4.4 · 4회 · 제한 30분

$X_1, \dots, X_n$이 iid이고 $P(X = k) = (1-p)^{k-1}p$ $(k = 1, 2, \dots)$이다. $EX = \frac{1}{p}$, $\mathrm{Var}\,X = \frac{1-p}{p^2}$이다.

(a) $\hat p = 1/\bar X_n$에 대해 $\sqrt{n}(\hat p - p)$의 극한분포의 분산을 구하라.

(b) 그 분산의 p 자리에 $\hat p$를 넣은 근사 95% 구간이 점근적으로 옳음을 Slutsky로 보여라.

(c) $n = 50$, $\bar X = 4$이면 (b)에서 쓰는 표준오차 $\hat p\sqrt{(1-\hat p)/n}$은 얼마인가?

(d) 같은 자료로 (b)의 구간을 구하라($z_{0.025} = 1.96$).

풀이 (a) $\sqrt{n}(\bar X_n - \frac{1}{p}) \xrightarrow{D} \mathcal{N}(0, \frac{1-p}{p^2})$이고 $g'(\frac{1}{p}) = -p^2$이므로 분산은 $p^4 \cdot \frac{1-p}{p^2} = p^2(1-p)$이다. (b) $\frac{\sqrt{n}(\hat p - p)}{\hat p\sqrt{1-\hat p}} = \frac{p\sqrt{1-p}}{\hat p\sqrt{1-\hat p}} \cdot \frac{\sqrt{n}(\hat p - p)}{p\sqrt{1-p}}$이다. 둘째 인수는 (a)로 $\mathcal{N}(0,1)$로 분포수렴하고, $\hat p \xrightarrow{P} p$이므로 첫 인수는 1로 확률수렴한다(정리 1.3). Slutsky로 $\mathcal{N}(0,1)$이므로 $\hat p \pm 1.96\,\hat p\sqrt{(1-\hat p)/n}$의 포함확률은 0.95로 간다. (c) $\hat p = 1/4 = 0.25$이고 표준오차는 $0.25\sqrt{0.75/50} = 0.0306$이다. (d) $0.25 \pm 1.96 \times 0.0306 = 0.25 \pm 0.0600 = [0.190,\ 0.310]$이다.

최종답 (a) $p^2(1-p)$ (c) 0.0306 (d) $[0.190,\ 0.310]$

D4 정규 꼬리의 중요도 표집 ★★★

정리 6.6 · 5회 · 제한 40분

$\theta = P(Z > a) = 1-\Phi(a)$를 제안분포 $g = \mathcal{N}(\mu, 1)$의 중요도 표집으로 추정한다. $h = 1\{y > a\}$, $w(y) = \phi(y)/\phi(y-\mu)$이다.

(a) $w(y) = e^{-\mu y + \mu^2/2}$이고 $E_g[h^2 w^2] = e^{\mu^2}\big(1 - \Phi(a+\mu)\big)$임을 보여라.

(b) $a = \mu = 3$일 때 한 표본의 분산의 비 $\dfrac{\theta(1-\theta)}{E_g[h^2w^2]-\theta^2}$을 구하라. $1-\Phi(3) = 1.3499\times10^{-3}$, $1-\Phi(6) = 9.866\times10^{-10}$, $e^9 = 8103.08$이다.

(c) $E_g[h^2w^2]$을 최소로 하는 μ가 $2\mu\big(1-\Phi(a+\mu)\big) = \phi(a+\mu)$를 만족함을 보이고, $1-\Phi(x)\approx\phi(x)/x$를 써서 $\mu \approx a$임을 보여라.

풀이 (a) $w(y) = \exp\!\big(-\frac{y^2}{2} + \frac{(y-\mu)^2}{2}\big) = e^{-\mu y+\mu^2/2}$이다. 그러면

$$E_g[h^2w^2] = \int_a^\infty e^{-2\mu y+\mu^2}\phi(y-\mu)\,dy = e^{\mu^2}\int_a^\infty \phi(y+\mu)\,dy = e^{\mu^2}\big(1-\Phi(a+\mu)\big).$$

가운데 등호는 $-\frac{(y-\mu)^2}{2} - 2\mu y + \mu^2 = -\frac{(y+\mu)^2}{2} + \mu^2$에서 나온다. (b) $E_g[h^2w^2] = 8103.08\times 9.866\times10^{-10} = 7.995\times10^{-6}$이고 $\theta^2 = 1.822\times10^{-6}$이므로 분산은 6.17×10^{-6}이다. 보통 Monte Carlo의 $\theta(1-\theta) = 1.348\times10^{-3}$과의 비는 약 218이다. (c) $\log E_g[h^2w^2] = \mu^2 + \log(1-\Phi(a+\mu))$를 μ로 미분해 0으로 두면 $2\mu - \frac{\phi(a+\mu)}{1-\Phi(a+\mu)} = 0$이다. $1-\Phi(x)\approx\phi(x)/x$를 넣으면 $2\mu \approx a+\mu$, 곧 $\mu\approx a$이다. 제안분포의 중심을 드문 사건의 경계로 옮기라는 뜻이다. 수치로 최소점을 구하면 $a=3$에서 $\mu = 3.15$로 a보다 조금 크다.

최종답 (b) ≈ 218