12.25 여럿 가운데 가장 큰 값

학습지

문제

STEP III 2025, 11번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 OCR에서 내려받을 수 있는 2025년 STEP 시험지 묶음(zip 파일) 가운데 STEP III 시험지에 있다:

www.ocr.org.uk/Images/750486-2025-past-papers.zip

힌트

먼저 혼자 힘으로 문제를 풀어 보자. 막히면 힌트를 **하나씩** 읽고, 힌트를 하나 읽을 때마다 다시 문제로 돌아가자. **잠깐, 직접 해 보기**라고 표시된 힌트가 핵심 단계이다. 각 힌트 뒤에는 풀이를 적을 공간이 있다.

? **생각해 보기** $X_1, \ldots, X_n$ 가운데 가장 큰 값이 y 이하인 것은 정확히 어떤 일이 일어날 때인가?

1단계 최댓값의 분포

다루는 내용: 소문항 (i)

? **생각해 보기** X_i 중 가장 큰 값이 y 이하인 것은 모든 X_i가 y 이하일 때, 그리고 그때에만 성립한다. 독립성(independence)을 이용한다.

2단계 분위수

다루는 내용: 소문항 (ii)

생각해 보기 $\mathrm{G}(L) = \alpha$를 L에 대하여 풀고, $\mathrm{G}(U) = 1 - \alpha$를 U에 대하여 푼다.

3단계 큰 n

다루는 내용: 소문항 (iii)

잠깐, 직접 해 보기 $\alpha^{1/n} = e^{(\ln \alpha)/n}$으로 쓴다. n이 크면 지수(exponent)가 작으므로 $e^t \approx 1 + t$를 이용한다.

4단계 증가와 퍼짐

다루는 내용: 소문항 (iv)

생각해 보기 중앙값(median)은 $L(\frac{1}{2})$이다. $n \to \infty$일 때 이 값은 어떻게 되는가? U와 L에 대한 근삿값을 서로 빼 본다. 어떤 항들이 소거되는가?

5단계 90% **구간**

다루는 내용: 소문항 (v)

? **생각해 보기** $\alpha = 0.05$로 둔다. $\ln 20$과 $\ln \frac{1}{0.95}$의 근삿값을 구한다.

숨은 수학

검벨 분포

$\mathrm{P}(\lambda Y - \ln n \le x) = (1 - \frac{e^{-x}}{n})^n \to \exp(-e^{-x})$이므로, 평행이동한 최댓값 $\lambda Y - \ln n$은 극한에서 분포함수(distribution function)가 $\exp(-e^{-x})$인 **검벨 분포**를 따른다. 그 평균은 오일러(Euler) 상수 $\gamma \approx 0.577$이다. 이 분포는 이미 만난 적이 있다. 쿠폰 n종을 모두 모으는 데 걸리는 시간(Section 9.11)은 (거의 독립인) 대기 시간 n개의 최댓값이며, 그래서 그 값이 약 $n \ln n + \gamma n$이고 n 정도 크기의 검벨 요동을 보이는 것이다.

극값 이론

피셔(Fisher), Tippett, 그네덴코(Gnedenko)의 놀라운 정리(1928–1943)에 따르면, 독립이고 동일한 분포를 따르는 많은 확률변수의 최댓값은 적절히 평행이동하고 축척을 조정하면 극한에서 가능한 모양이 세 가지뿐이다. 검벨 분포(지수분포나 정규분포처럼 꼬리가 가벼운 경우), 프레셰(Fréchet) 분포(코시(Cauchy) 분포처럼 꼬리가 무거운 경우), 바이불(Weibull) 분포(엄격한 상한이 있는 확률변수의 경우)이다. 이는 중심극한정리(central limit theorem)(Section 12.24)의 극값판이라 할 수 있다. 공학자들은 이를 이용해 "백 년 만의 파도"에 대비한 홍수 방어 시설과 해상 구조물을 설계하고, 보험사는 재해 보험료를 책정하며, 기후과학자들은 폭염이 어떻게 변하고 있는지를 추정한다.

검벨과 홍수

에밀 율리우스 검벨(Emil Julius Gumbel)은 독일의 통계학자로, 나치에 공공연히 반대하다가 1932년 대학 교수직에서 쫓겨났고 결국 미국으로 망명했다. 그는 1930년대와 1940년대에 극값 이론을 강의 범람에 적용했다. 그의 책 *Statistics of Extremes*(극값의 통계학, 1958)는 이 분야를 실용적인 도구로 만들었다. 1953년 북해 대홍수 이후 네덜란드를 보호하기 위해 건설된 델타 공사(Delta Works)는 약 10 000년에 한 번 일어날 것으로 예상되는 폭풍 해일을 견디도록 극값 통계를 이용해 설계되었다.

더 도전하기

1 **최댓값의 기댓값.** $\mathrm{E}(Y) = \frac{1}{\lambda}(1 + \frac{1}{2} + \cdots + \frac{1}{n})$임을 보여라. (힌트: 무기억성(memorylessness)에 의해, n개의 수명 가운데 첫 번째 것이 끝난 뒤 남은 $n-1$개는 새로 시작하는 것과 같다.)

2 **최솟값.** $X_1, \ldots, X_n$ 가운데 가장 작은 값이 모수 $n\lambda$인 지수분포를 따름을 보여라.

3 **정확한 중앙값.** Y의 정확한 중앙값을 구하고, $n = 10$일 때 이를 근삿값 $\frac{1}{\lambda}(\ln n - \ln\ln 2)$와 비교하여라.

4 **균등분포의 최댓값.** $X_1, \ldots, X_n$이 $[0,1]$에서 균등분포(uniform distribution)를 따를 때, 그 최댓값 M의 분포를 구하고, n이 클 때 $n(1-M)$이 근사적으로 지수분포를 따름을 보여라. 이는 세 가지 극값 모양 가운데 어느 것인가?

더 읽을거리

- **Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values*** (Springer). 홍수와 해수면에 관한 실제 자료를 담은 표준 입문서이다.
- **Emil Julius Gumbel, *Statistics of Extremes*** (Dover). 고전이다.
- **David Spiegelhalter, *The Art of Uncertainty*** (Pelican). 드문 사건과 그것을 어떻게 생각해야 하는지에 관한 책이다.

안내에 따른 전체 풀이는 제 12.25절에 있다:

https://beyondstep.keithmcnulty.org/ko/ch12/extremes.html

매주 STEP 문제 하나를 끝까지 풀어 본다. The STEP Up은 이메일로 무료로 받아 볼 수 있다: keithmcnulty.substack.com