12.24 중심극한정리

학습지

문제

STEP III 2008, 12번 문제

더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:

step.maths.org/questions/08-s3-q12

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 밀도함수 $\frac{1}{2}e^{-|x|}$는 대칭이다. 그래프는 어떤 모양인가? 평균은 얼마인가?

1단계 적률생성함수

다루는 내용: 그래프 개형과 적률생성함수

생각해 보기 $\mathrm{E}(e^{\theta X})$의 적분을 0에서 나누어라. 왜 $|\theta| < 1$이어야 하는가?

생각해 보기 $\frac{1}{1-\theta^2}$을 급수(series)로 전개하여 평균(mean)과 분산(variance)을 읽어 내라.

2단계 개구리의 변위

다루는 내용: $Y/\sqrt{2n}$의 적률생성함수

잠깐, 직접 해 보기 각 점프는 독립(independent)이므로, 그 합 Y의 적률생성함수(moment generating function)는 무엇인가? Y를 $\sqrt{2n}$으로 나누면 θ에는 어떤 변화가 생기는가?

3단계 극한

다루는 내용: 극한

생각해 보기 로그를 취하고, 작은 u에 대하여 $\ln(1-u) = -u - \frac{1}{2}u^2 - \cdots$를 이용하라. 극한으로 얻은 적률생성함수는 어떤 분포의 것인가?

4단계 개구리의 확률

다루는 내용: 최소 점프 횟수

> ? **생각해 보기** n이 클 때 Y는 근사적으로 $\mathrm{N}(0, 2n)$을 따른다. 표준정규분포(standard normal distribution)를 따르는 Z에 대하여 $\mathrm{P}(|Z| \geq 1.96) = 0.05$이다. n에 대하여 풀어라.

숨은 수학

일반적인 정리

같은 논증은 평균이 0이고 분산 σ^2이 유한한 임의의 분포에 대해서도 성립한다(적률생성함수가 존재해야 하지만, 이 조건은 피할 수 있다). 이때 $\mathrm{M}(\theta) = 1 + \frac{1}{2}\sigma^2\theta^2 + o(\theta^2)$이고,

$$\left(\mathrm{M}\left(\frac{\theta}{\sigma\sqrt{n}}\right)\right)^n = \left(1 + \frac{\theta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{\theta^2/2}.$$

극한에서 살아남는 것은 분산뿐이고, 원래 분포의 모양은 잊힌다. 정규분포가 보편적인 이유가 바로 이것이다. 대학에서는 항상 존재하는 **특성함수**(*characteristic function*) $\mathrm{E}(e^{i\theta X})$를 이용하여 이 정리를 증명하고, 분포의 수렴은 레비(Lévy)의 연속성 정리로 엄밀하게 다룬다.

얼마나 빠른가?

중심극한정리는 극한에서 무슨 일이 일어나는지 말해 줄 뿐, 얼마나 빨리 일어나는지는 말해 주지 않는다. **베리-에센 정리**(*Berry–Esseen theorem*)는 오차를 상수 곱하기 $\frac{\mathrm{E}|X|^3}{\sigma^3\sqrt{n}}$으로 제한한다. 개구리의 분포처럼 대칭인 분포에서는 수십 번만 점프해도 근사가 좋다. 치우친(skewed) 분포이거나 꼬리의 극단적인 부분에서는 n이 커도 근사가 나쁠 수 있다. 정규근사는 매우 드문 사건이 일어날 확률을 과소평가하며, 이 사실은 보험과 금융에 중대한 영향을 미친다.

드무아브르에서 린데베리까지

드무아브르(de Moivre)는 1733년에 이항분포의 정규근사를 발견했고(Section 12.23), 라플라스(Laplace)는 1810년에 이를 더 일반적인 확률변수들의 합으로 확장하면서 측정 오차가 많은 작은 효과의 합이라는 생각을 처음 내놓았다. 엄밀한 증명은 훨씬 뒤에 나왔다. 1901년에 알렉산드르 랴푸노프(Aleksandr Lyapunov)가, 그리고 1922년에 야를 린데베리(Jarl Lindeberg)가 증명했는데, 린데베리는 정리가 성립하는 본질적으로 가장 약한 조건을 찾아냈다. "중심극한정리"라는 이름은 1920년에 조지 포여(George Pólya)가 붙인 것으로, 이 정리가 확률론의 중심에 있다는 뜻이다.

더 도전하기

1 **점프 두 번의 정확한 분포.** $Y = X_1 + X_2$의 밀도함수가 $\frac{1}{4}(1 + |y|)e^{-|y|}$임을 보이고, 분산이 4임을 확인하라.

2 **이항분포에서 정규분포로.** 평균이 0이 되도록 평행이동한(즉 $\pm\frac{1}{2}$의 값을 갖는) $X \sim \mathrm{B}(1, \frac{1}{2})$에 대하여 $\mathrm{M}_X(\theta) = \cosh\frac{\theta}{2}$임을 구하고, 이러한 확률변수 n개의 합을 표준화한 것의 적률생성함수가 $e^{\theta^2/2}$로 수렴함을 보여라.

3 **개구리, 정확하게.** 컴퓨터를 이용하여(또는 검산용으로 $n = 2$일 때의 정확한 밀도함수를 이용하여) $n = 82$일 때 정규근사가 얼마나 정확한지 알아보라.

4 **코시 분포는 왜 안 되는가?** 코시 분포(Cauchy distribution)(Section 12.22)는 적률생성함수가 없다. 논증의 어느 단계가 실패하는가? 그리고 코시 분포를 따르는 확률변수 n개의 평균은 어떻게 되는가?

더 읽을거리

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University Press), 생성함수와 극한정리를 다룬 장.
- **Hans Fischer, *A History of the Central Limit Theorem*** (Springer). 라플라스부터 린데베리까지를 다룬다.
- **David Spiegelhalter, *The Art of Statistics*** (Pelican), 정규분포가 어디에나 나타나는 이유를 다룬다.

안내에 따른 전체 풀이는 제 12.24절에 있다:

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

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