풀이

STEP II 2016, 문제 13

확률과 통계 • 정규근사

STEP II 2016, 13번 문제

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

step.maths.org/questions/16-s2-q13

이 문제는 *Beyond STEP*의 12.23 절 **확률로 얻는 스털링 공식**에서 단계별 안내와 함께 자세히 다룬다: https://beyondstep.keithmcnulty.org/ko/ch12/stirling-probability.html

핵심 아이디어

이산 분포(discrete distribution)의 확률 하나 $\mathrm{P}(X = k)$는 k를 중심으로 하는 폭 1인 막대의 넓이이다. 분포가 정규분포(normal distribution)에 가깝다면, 그 넓이는 대략 k에서의 정규분포 밀도(density)에 폭 1을 곱한 값이다. 이 사실을 평균(mean)에서 적용하면, 그곳에서 표준정규분포(standard normal distribution)의 밀도는 $\frac{1}{\sqrt{2\pi}}$이므로, 정확히 알고 있는 확률이 하나의 근사식으로 바뀐다. 분포를 영리하게 고르면 그 정확한 확률 안에 우리가 원하는 계승(factorial)이 들어 있게 된다.

풀이

풀이 전체에서 쓰는 근사

W가 정숫값을 갖고 평균이 μ, 분산(variance)이 σ^2이며, $V \sim \mathrm{N}(\mu, \sigma^2)$로 잘 근사된다고 하자. V의 밀도는

$$f(v) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2}.$$

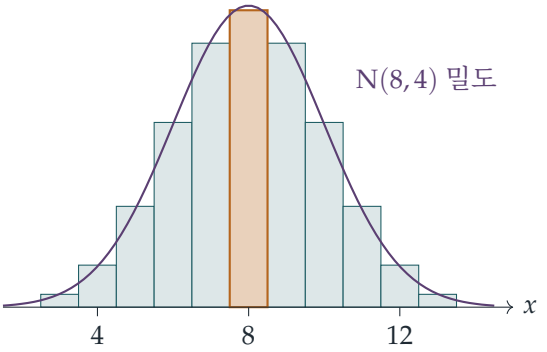
이다. 연속성 수정(continuity correction)을 하면

$$\mathrm{P}(W = k) \approx \mathrm{P}\left(k - \tfrac{1}{2} < V < k + \tfrac{1}{2}\right) = \int_{k-1/2}^{k+1/2} f(v)\,\mathrm{d}v \approx 1 \times f(k),$$

이다. 마지막 단계는 구간의 폭이 1이고, σ가 작지 않을 때 그 구간에서 f가 거의 변하지 않기 때문이다. $k = \mu$에서 지수 부분은 $e^0 = 1$이므로

$$P(W = \mu) \approx \frac{1}{\sigma\sqrt{2\pi}}. \qquad (*)$$

이다.



소문항 (i)

$X \sim B(16, \frac{1}{2})$에 대하여

$$E(X) = np = 8, \qquad Var(X) = np(1-p) = 16 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 4,$$

이므로 $\sigma = 2$이다. $p = \frac{1}{2}$이므로 분포는 8에 대하여 대칭(symmetric)이고 n도 꽤 크므로, $N(8, 4)$는 좋은 근사이다(그림은 막대들을 정규분포 곡선과 함께 보여 준다). $(*)$에 의해 색칠한 막대의 넓이는

$$P(X = 8) \approx \frac{1}{2\sqrt{2\pi}},$$

이고, 이것이 구하는 결과이다.

소문항 (ii)

$Y \sim B(2n, \frac{1}{2})$라 하자. 그러면 $E(Y) = n$이고 $Var(Y) = 2n \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{n}{2}$이므로 $\sigma = \sqrt{n/2}$이다. 정확한 값은

$$P(Y = n) = \binom{2n}{n} \left(\tfrac{1}{2}\right)^n \left(\tfrac{1}{2}\right)^n = \frac{(2n)!}{(n!)^2\, 2^{2n}}.$$

이다. n이 크면 정규근사를 적용할 수 있고, $(*)$에 의해

$$P(Y = n) \approx \frac{1}{\sqrt{n/2}\,\sqrt{2\pi}} = \frac{1}{\sqrt{n\pi}}.$$

이다. 두 식을 같다고 놓고 양변에 $2^{2n}(n!)^2$을 곱하면

$$(2n)! \approx \frac{2^{2n}(n!)^2}{\sqrt{n\pi}},$$

이고, 이것이 구하는 결과이다. (소문항 (i)은 $n = 8$인 경우이다: $\sqrt{8\pi} = 2\sqrt{2\pi}$.)

소문항 (iii)

$Z \sim \mathrm{Po}(n)$이라 하면 $\mathrm{E}(Z) = \mathrm{Var}(Z) = n$이고 $\sigma = \sqrt{n}$이다. 정확한 값은

$$\mathrm{P}(Z = n) = \frac{\mathrm{e}^{-n} n^n}{n!}.$$

이다.

> **생각해 보기** 여기서 왜 포아송 분포(Poisson distribution)를 정규분포로 근사해도 되는가?

$\mathrm{Po}(n)$ 확률변수는 서로 독립(independent)인 $\mathrm{Po}(1)$ 확률변수 n개의 합과 같은 분포를 가지므로, 중심극한정리(central limit theorem)에 의해 n이 클 때 근사적으로 $\mathrm{N}(n, n)$을 따른다. 이것은 λ가 클 때 $\mathrm{Po}(\lambda) \approx \mathrm{N}(\lambda, \lambda)$라는 표준적인 규칙이다.

따라서 $(*)$에 의해

$$\frac{\mathrm{e}^{-n} n^n}{n!} \approx \frac{1}{\sqrt{n}\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi n}},$$

이고, 정리하면

$$n! \approx \sqrt{2\pi n}\, \mathrm{e}^{-n}\, n^n,$$

이 되어 구하는 결과를 얻는다.

검산

- 소문항 (i): 정확한 값은 $\mathrm{P}(X = 8) = \binom{16}{8}/2^{16} = 12870/65536 \approx 0.1964$이고, $\frac{1}{2\sqrt{2\pi}} \approx 0.1995$이다. 오차는 약 1.6% 이내이다.
- $n = 10$일 때의 소문항 (ii): $\binom{20}{10}/2^{20} \approx 0.1762$이고 $1/\sqrt{10\pi} \approx 0.1784$이다. 계승 근사식의 양변의 비는 n이 커질수록 1로 수렴한다($n = 100$에서 약 0.9988).
- $n = 10$일 때의 소문항 (iii): $10! = 3\,628\,800$이고 $\sqrt{20\pi}\,\mathrm{e}^{-10}\,10^{10} \approx 3\,598\,696$이므로, 1% 미만만큼 작다.
- 일관성: (iii)의 결과를 (ii)의 $(2n)!$과 $n!$에 모두 대입하면 양변이 모두 $\sqrt{4\pi n}\,\mathrm{e}^{-2n}(2n)^{2n}$이 되므로, 두 근사식은 서로 맞아떨어진다.

흔한 실수

- 표준정규분포의 밀도 $\frac{1}{\sqrt{2\pi}}$를 σ로 나누지 않고 그대로 쓰는 것. $\mathrm{N}(\mu, \sigma^2)$의 평균에서의 밀도는 $\frac{1}{\sigma\sqrt{2\pi}}$이다.

- 소문항 (ii)에서 표준편차(standard deviation)로 $\sqrt{n/2}$ 대신 $\frac{n}{2}$를 쓰거나, $\mathrm{B}(2n, \frac{1}{2})$의 분산을 잘못 구하는 것.
- 소문항 (iii)에서 포아송 분포를 정규분포로 근사해도 되는 이유를 밝히지 않는 것.

STEP 너머

소문항 (iii)의 결과는 스털링 공식(Stirling's formula)이고, 이 논증은 중심극한정리를 한 점에서 적용한 발견적(heuristic) 형태, 곧 "국소 극한 정리(local limit theorem)"이다. 이를 엄밀하게 만들려면 주의가 필요하다. 분포함수(distribution function)가 정규분포의 분포함수에 가깝다는 사실만으로는 개별 확률이 밀도에 가깝다고 말할 수 없기 때문이다. 스털링 공식은 직접 증명할 수도 있는데, 예를 들어 $\ln n! = \sum \ln k$를 $\int \ln x \, \mathrm{d}x$와 비교하거나, 감마함수(gamma function)에서 출발하여 라플라스(Laplace)의 방법을 쓰면 된다. Blitzstein과 Hwang의 *Introduction to Probability*(확률론 입문)는 중심극한정리와 포아송 분포와 정규분포의 관계를 다루며, Grimmett과 Welsh의 *Probability: An Introduction*(확률: 입문)은 함께 읽기 좋은 책이다.

Beyond STEP 풀이: https://beyondstep.keithmcnulty.org/ko/. 풀이 글은 CC0 1.0이다. 문제 자체의 저작권은 © Cambridge University Press & Assessment에 있으며 여기에 싣지 않는다. 영어 원문을 기계 번역하였으며 아직 검토되지 않았다.