풀이

STEP II 2003, 문제 13

확률과 통계 • 절단 푸아송 분포

> **STEP II 2003, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/03-s2-q13

핵심 아이디어

이 분포는 포아송 분포(Poisson distribution)에서 값 0을 제거하고, 남은 확률들의 합이 여전히 1이 되도록 비례적으로 키운 것이다. 따라서 우리에게 필요한 모든 합은 $k=0$ 항이 빠진 포아송 합이며, 평균(mean)과 분산(variance)에서는 어차피 $k=0$ 항이 아무것도 기여하지 않는다. 그러므로 익숙한 포아송 분포의 사실 $\sum k\,p_k = \lambda$와 $\sum k^2 p_k = \lambda + \lambda^2$를 빌려와서 A를 곱하기만 하면 된다.

풀이

상수 A

일반적인 $\mathrm{Po}(\lambda)$의 확률을 $p_k = \dfrac{\lambda^k \mathrm{e}^{-\lambda}}{k!}$로 쓰자. $\sum_{k\geq 0} \lambda^k/k! = \mathrm{e}^{\lambda}$이므로

$$\sum_{k=1}^{\infty} p_k = \mathrm{e}^{-\lambda}(\mathrm{e}^{\lambda} - 1) = 1 - \mathrm{e}^{-\lambda}.$$

확률 $\mathrm{P}(X=k) = Ap_k$의 합은 1이어야 하므로 $A(1-\mathrm{e}^{-\lambda}) = 1$, 즉

$$A = (1-\mathrm{e}^{-\lambda})^{-1},$$

이고, 이것이 요구된 결과이다. $\lambda > 0$일 때 $0 < 1-\mathrm{e}^{-\lambda} < 1$이므로 $A > 1$임에 주목하자. 즉 남은 확률 p_0이 $k \geq 1$인 값들에 나누어 분배된 것이다.

평균

$\sum k\,p_k$의 $k=0$ 항은 0이므로

$$\sum_{k=1}^{\infty} k\,p_k = \sum_{k=0}^{\infty} k\,p_k = \lambda,$$

이고, 이것은 $\mathrm{Po}(\lambda)$의 평균이다. (직접 계산하면 $\sum_{k\geq 1} k\lambda^k \mathrm{e}^{-\lambda}/k! = \lambda \sum_{k\geq 1} \lambda^{k-1}\mathrm{e}^{-\lambda}/(k-1)! = \lambda$이다.) 따라서

$$\mu = \mathrm{E}(X) = A\lambda = \frac{\lambda}{1-\mathrm{e}^{-\lambda}}.$$

분산

마찬가지로 $\sum k^2 p_k$의 $k=0$ 항도 0이며, $\mathrm{Po}(\lambda)$에 대해서는 $\mathrm{E}(Y^2) = \mathrm{Var}(Y) + (\mathrm{E}\,Y)^2 = \lambda + \lambda^2$이다. 이 사실을 그냥 인용하고 싶지 않다면 $k^2 = k(k-1)+k$로 쓰면 된다.

$$\sum_{k\geq 2} k(k-1)\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} = \lambda^2 \sum_{k\geq 2} \frac{\lambda^{k-2}\mathrm{e}^{-\lambda}}{(k-2)!} = \lambda^2, \qquad \text{이므로} \qquad \sum_{k\geq 1} k^2 p_k = \lambda^2 + \lambda.$$

따라서

$$\begin{aligned}\mathrm{E}(X^2) &= A\lambda(1+\lambda) = \mu(1+\lambda),\\ \mathrm{Var}(X) &= \mathrm{E}(X^2) - \mu^2 = \mu(1+\lambda) - \mu^2 = \mu(1-\mu+\lambda),\end{aligned}$$

이고, 이것이 요구된 결과이다.

부등식

> **생각해 보기** 엄격한 부등식(strict inequality) $\mu < 1+\lambda$는 어디에서 나올 수 있을까?

분산에서 나온다. X는 모든 값 $1, 2, 3, \ldots$을 양의 확률로 취하므로 상수가 아니며, 상수가 아닌 확률변수(random variable)의 분산은 0보다 엄격하게 크다($\mathrm{Var}(X) = \mathrm{E}((X-\mu)^2)$이고 $(X-\mu)^2 > 0$이 양의 확률로 성립하기 때문이다). 따라서

$$\mu(1-\mu+\lambda) > 0.$$

$\mu > 0$이므로(양의 확률변수의 평균이므로) μ로 나누어 $1-\mu+\lambda > 0$, 즉 $\mu < 1+\lambda$를 얻는다.

하한에 대해서는, $\mu = A\lambda$이고 $A > 1, \lambda > 0$이므로 $\mu > \lambda$이다. 이를 종합하면

$$\lambda < \mu < 1+\lambda.$$

$\lambda = 100$일 때의 정규근사

이제 e^{-100}은 천문학적으로 작으므로(약 4×10^{-44}), 어떤 합리적인 정확도로 보아도

$$\mu = \frac{100}{1 - e^{-100}} \approx 100, \qquad \mathrm{Var}(X) = \mu(101 - \mu) \approx 100 \times 1 = 100.$$

따라서 X를 $Y \sim N(100, 10^2)$로 근사한다. 이를 정규근사(normal approximation)라 한다. X는 정수값을 가지므로 연속성 수정(continuity correction)을 사용한다. 즉 사건 $X = 100$은 $99.5 < Y < 100.5$에 대응한다. $Z = (Y - 100)/10$으로 표준화하면

$$\begin{aligned} P(X = 100) &\approx P(-0.05 < Z < 0.05) \\ &= 2\Phi(0.05) - 1 \\ &\approx 2(0.51994) - 1 = 0.0399, \end{aligned}$$

이므로 소수점 아래 둘째 자리까지 $P(X = 100) \approx 0.04$이다.

검산

- $\lambda \to 0$일 때 $\mu = \lambda/(1 - e^{-\lambda}) \to 1$이고 분산은 $\to 0$이다. 즉 거의 모든 확률이 $k = 1$에 몰리며, 이는 당연한 결과이다. 이것은 부등식 $\lambda < \mu < 1 + \lambda$가 μ를 1 쪽으로 조여 가는 것과도 일치한다.
- 여러 임의의 λ에 대해 급수를 수치적으로 합하면 A, μ, $\mu(1 - \mu + \lambda)$가 기계 정밀도까지 재현된다.
- $\lambda = 100$일 때 정확한 확률은 $\dfrac{100^{100} e^{-100}}{100!\,(1 - e^{-100})} \approx 0.03986$이고, 정규근사값은 0.03988이다. 둘 다 반올림하면 0.04이다.

흔한 실수

- 연속성 수정을 잊고 $P(Y = 100) = 0$을 계산하거나, 정당화 없이 확률밀도 값을 사용하는 것.
- $\mu \le 1 + \lambda$만 증명하는 것. 엄격한 부등식을 얻으려면 X가 상수가 아니라는 관찰이 필요하다.
- $\mu > 0$임을 언급하지 않고 μ로 나누는 것.

STEP 너머

X는 **영절단 포아송**(zero-truncated Poisson) 분포이다. 통계학자들은 개수가 0인 경우를 관측할 수 없을 때 이 분포를 사용한다(예를 들어 검문소를 지나는 차량의 탑승자 수가 그렇다. 탑승자가 0명인 차는 없다). 정규근사가 잘 맞는 이유는 Po(100) 확률변수가 서로 독립인 100개의 Po(1) 확률변수의 합이어서 중심극한정리(central limit theorem)가 적용되기 때문이다. 이 규모에서는 절단의 효과가 보이지 않는다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 이 문제의 두 부분 모두의 바탕이 되는 생성함수(generating function)와 중심극한정리를 다룬다.

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.