풀이

STEP III 1996, 문제 13

확률과 통계 • 체비쇼프 부등식

> **STEP III 1996, 13번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/96-s3-q13

핵심 아이디어

분산(variance)은 각 값마다 하나씩 있는 음이 아닌 항 $p_i(x_i-\mu)^2$의 합이다. 이 합을 평균(mean)에서 먼 값(적어도 λ만큼 떨어진 값)과 가까운 값으로 나눈다. 먼 값의 확률에 대한 **상계**(upper bound)를 얻으려면 가까운 항을 버리고 먼 항 각각을 아래에서 누른다. **하계**(lower bound)를 얻으려면 반대로 한다. 즉 모든 항을 위에서 누르는데, 가까운 항은 λ^2로, 먼 항은 k^2로 누른다.

풀이

정의와 $aX+b$

$p_i=\mathrm{P}(X=x_i)$라 하면 $p_i\geq 0$이고 $\sum_{i=1}^{n}p_i=1$이다. 그러면

$$\mathrm{E}(X)=\sum_{i=1}^{n}x_ip_i,\qquad \mathrm{var}(X)=\mathrm{E}((X-\mu)^2)=\sum_{i=1}^{n}(x_i-\mu)^2p_i,$$

이고, 여기서 $\mu=\mathrm{E}(X)$이다. (동치로 $\mathrm{var}(X)=\mathrm{E}(X^2)-\mu^2$이지만, 앞으로 사용할 것은 첫 번째 형태이다.)

확률변수(random variable) $aX+b$는 확률 p_i로 값 ax_i+b를 갖는다. ($a=0$이면 이 값들이 모두 같아지지만, 같은 값의 확률을 더해도 아래의 합은 변하지 않는다.) 따라서 $\sum_i p_i=1$을 이용하면

$$\mathrm{E}(aX+b)=\sum_i(ax_i+b)p_i=a\sum_i x_ip_i+b\sum_i p_i=a\mathrm{E}(X)+b,$$

이다. 그러므로 $aX + b - \mathrm{E}(aX + b) = a(X - \mu)$이고,

$$\mathrm{var}(aX + b) = \sum_i (a(x_i - \mu))^2 p_i = a^2 \sum_i (x_i - \mu)^2 p_i = a^2 \,\mathrm{var}(X).$$

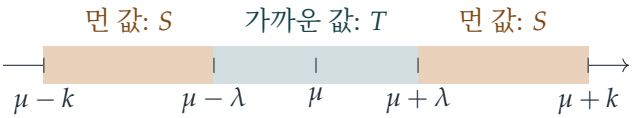
이다. b를 더하면 모든 값과 평균이 함께 평행이동하므로 퍼진 정도에는 아무 영향이 없다.

상계

$\mu = \mathrm{E}(X)$라 하고, $|x_i - \mu| \geq \lambda$인 첨자 i의 집합("먼" 값)을 S, 나머지, 즉 $|x_i - \mu| < \lambda$인 첨자의 집합을 T라 하자. 그러면

$$\mathrm{P}(|X - \mu| \geq \lambda) = \sum_{i \in S} p_i.$$

이다. 이 확률을 P라 하자.



분산의 모든 항은 음이 아니므로 T에 속하는 항을 버리면 합은 작아질 수밖에 없다. 또 $i \in S$이면 $(x_i - \mu)^2 \geq \lambda^2$이다. 따라서

$$\begin{aligned}\mathrm{var}(X) &= \sum_{i \in S} (x_i - \mu)^2 p_i + \sum_{i \in T} (x_i - \mu)^2 p_i \\ &\geq \sum_{i \in S} (x_i - \mu)^2 p_i \\ &\geq \lambda^2 \sum_{i \in S} p_i = \lambda^2 P.\end{aligned}$$

이다. $\lambda > 0$이므로 λ^2으로 나눌 수 있고,

$$\mathrm{P}(|X - \mu| \geq \lambda) \leq \frac{\mathrm{var}(X)}{\lambda^2},$$

를 얻는다. 이것이 증명할 내용이다. (S가 공집합이면 S에 대한 두 합이 모두 0이므로 부등식은 자명하게 성립한다.)

하계

이제 모든 i에 대하여 $|x_i - \mu| \leq k$라고 추가로 가정하자. 이번에는 분산의 모든 항을 위에서 누른다.

- $i \in S$이면 새 가정에 의해 $(x_i - \mu)^2 \leq k^2$이다.
- $i \in T$이면 T의 정의에 의해 $(x_i - \mu)^2 < \lambda^2$이다.

$\sum_{i\in T} p_i = 1 - P$이므로 다음을 얻는다.

$$\mathrm{var}(X) \le k^2 \sum_{i\in S} p_i + \lambda^2 \sum_{i\in T} p_i = k^2 P + \lambda^2(1-P) = \lambda^2 + (k^2-\lambda^2)P.$$

정리하면

$$(k^2-\lambda^2)\,P \ge \mathrm{var}(X) - \lambda^2.$$

이다.

? **생각해 보기** $k^2-\lambda^2$으로 나눌 수 있는가?

$k \ge \lambda > 0$이 주어졌으므로 $k^2-\lambda^2 \ge 0$이다. $k > \lambda$이면 이 값은 양수이고, 나누어도 부등호의 방향은 그대로이다.

$$\mathrm{P}\big(|X-\mu| \ge \lambda\big) \ge \frac{\mathrm{var}(X)-\lambda^2}{k^2-\lambda^2},$$

이것이 구하는 결과이다. 경계인 경우 $k = \lambda$에서는 우변의 분모가 0이 되어 이 명제는 의미가 없다. 그 바로 앞의 부등식은 여전히 성립하며 $\mathrm{var}(X) \le \lambda^2$을 뜻하는데, 모든 $(x_i-\mu)^2 \le k^2 = \lambda^2$이므로 이는 분명하다.

$\mathrm{var}(X) \le \lambda^2$일 때는 하계가 0 이하이므로 새로운 정보를 주지 않는다. 하계가 의미를 갖는 것은 분산이 λ^2에 비해 클 때뿐이며, 이때 확률의 일부가 꼬리 쪽으로 밀려날 수밖에 없다.

검산

- $X = \pm 1$을 각각 확률 0.4로, ± 4를 각각 확률 0.1로 갖는다고 하면 $\mu = 0$이고 $\mathrm{var}(X) = 0.8\times 1 + 0.2\times 16 = 4$이다. $\lambda = 1, k = 4$로 두면 $P = 1$이고, 상계는 4, 하계는 $\frac{3}{15} = \frac{1}{5}$이며, 실제로 $\frac{1}{5} \le 1 \le 4$이다.
- $X = 0$ 또는 $\pm c$이고 $\mathrm{P}(X = \pm c) = q/2$씩이라 하자. 그러면 $\mathrm{var}(X) = qc^2$이고 $\lambda = c$로 두면 상계는 정확히 $q = P$이다. 즉 체비쇼프 부등식(Chebyshev's inequality)은 일반적으로 더 개선할 수 없다.
- 값이 최대 일곱 개인 무작위 분포 20 000개에 대하여 정확한 유리수 연산으로 두 부등식과 $\mathrm{E}(aX+b)$, $\mathrm{var}(aX+b)$의 공식을 검사하였고, 실패한 경우는 없었다.

흔한 실수

- $\mathrm{var}(aX+b) = a\,\mathrm{var}(X)$ 또는 $a^2\mathrm{var}(X) + b$라고 쓰는 것. 분산은 편차에 대하여 이차이고, 평행이동은 어떤 편차도 바꾸지 않는다.
- 하계에서 가까운 항을 λ^2 대신 0으로 누르는 것. 그러면 $\mathrm{var}(X) \le k^2 P$를 얻는데, 이는 참이지만 묻는 내용이 아니다.

- $k^2 - \lambda^2$이 양수라는 점(그리고 $k = \lambda$를 제외해야 한다는 점)을 언급하지 않고 나누는 것.

STEP 너머

첫 번째 부등식은 체비쇼프 부등식이며, 항을 버리는 기법은 $Y \geq 0$에 대한 마르코프 부등식(Markov's inequality) $\mathrm{P}(Y \geq c) \leq \mathrm{E}(Y)/c$의 바탕에 있는 것과 같은 기법이다($Y = (X - \mu)^2$에 적용하면 된다). 대학 1학년 확률론 강의에서 체비쇼프 부등식을 쓰면 대수의 약법칙(weak law of large numbers)을 두 줄로 증명할 수 있다. X의 독립(independent)인 복사본 n개의 평균은 분산이 $\mathrm{var}(X)/n$이므로 μ에서 멀리 벗어날 가능성이 작다. 두 번째 부등식은 유계성(boundedness)이 필요한 "역" 형태로, Paley–Zygmund 부등식과 같은 정신을 따른다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문서)이 이 내용을 모두 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.