풀이

STEP II 2001, 문제 14

확률과 통계 • 최소분산 추정량

> **STEP II 2001, 14번 문제**
>
> 더 읽어 나가기 전에 문제부터 읽어 보자. 이 문제는 STEP Support Programme의 문제 데이터베이스에 있다:
>
> step.maths.org/questions/01-s2-q14

핵심 아이디어

T는 서로 독립인 두 베르누이(Bernoulli) 확률변수의 가중 결합이므로, 그 평균과 분산은 기댓값의 선형성(linearity of expectation)과 독립인 X, Y에 대한 규칙 $\operatorname{Var}(aX+cY) = a^2\operatorname{Var}(X) + c^2\operatorname{Var}(Y)$에서 곧바로 나온다. 가중치는 λ가 무엇이든 평균이 p가 되도록 정해져 있다. 즉 T는 p의 **불편추정량**(*unbiased estimator*)이고, 이 문제는 그중 가장 좋은 것을 구하라는 것이다. 마지막 부분에서는 중심극한정리(central limit theorem)에 의해 $\overline{T}$가 근사적으로 정규분포를 따르며, 평균이 고정된 정규 확률변수는 분산이 가장 작을 때 가장 집중되어 있다.

풀이

T의 평균과 분산

여기서 X와 Y는 다음을 만족하는 베르누이 확률변수이다.

$$\mathrm{E}(X) = p, \quad \operatorname{Var}(X) = p(1-p), \qquad \mathrm{E}(Y) = 2p, \quad \operatorname{Var}(Y) = 2p(1-2p).$$

(확률 q로 값 1을 갖고 그렇지 않으면 0을 갖는 확률변수에 대하여 $\mathrm{E} = q$이고 $\mathrm{E}(\cdot^2) = q$이므로, 분산은 $q - q^2 = q(1-q)$이다.) 기댓값의 선형성에 의해

$$\mathrm{E}(T) = \lambda p + \tfrac{1}{2}(1-\lambda)\cdot 2p = \lambda p + (1-\lambda)p = p,$$

이므로 요구된 결과를 얻는다. 동전들이 서로 독립이므로 X와 Y는 독립이고, 따라서 공분산(covariance) 항이 없으며

$$\begin{aligned}\operatorname{Var}(T) &= \lambda^2\operatorname{Var}(X) + \tfrac{1}{4}(1-\lambda)^2\operatorname{Var}(Y)\\ &= \lambda^2 p(1-p) + \tfrac{1}{2}(1-\lambda)^2 p(1-2p).\end{aligned}$$

분산 최소화하기

$V(\lambda) = \mathrm{Var}(T)$라 하자. 이것은 최고차항의 계수가 양수 $p(1-p) + \frac{1}{2}p(1-2p)$인 λ에 대한 이차식이므로($0 < p < \frac{1}{2}$이므로 두 항 모두 양수이다), $V'(\lambda) = 0$인 곳에서 유일한 최솟값을 갖는다.

$$V'(\lambda) = 2\lambda p(1-p) - (1-\lambda)p(1-2p) = 0.$$

$p > 0$으로 나누고 λ가 있는 항을 모으면

$$\lambda[2(1-p) + (1-2p)] = 1 - 2p \quad \Longrightarrow \quad \lambda = \frac{1-2p}{3-4p},$$

이고, 이것이 문제에 주어진 값이다. (여기서 $3 - 4p > 1 > 0$이므로 0으로 나누는 일은 없다.)

최솟값 자체도 필요하다. $a = p(1-p)$, $c = \frac{1}{2}p(1-2p)$라 하면 $V = a\lambda^2 + c(1-\lambda)^2$이고, 이는 $\lambda = c/(a+c)$(위에서 구한 값)에서 최소가 되므로

$$V_{\min} = a\Big(\frac{c}{a+c}\Big)^2 + c\Big(\frac{a}{a+c}\Big)^2 = \frac{ac(a+c)}{(a+c)^2} = \frac{ac}{a+c}.$$

이제 $a + c = \frac{1}{2}p[2(1-p) + (1-2p)] = \frac{1}{2}p(3-4p)$이고 $ac = \frac{1}{2}p^2(1-p)(1-2p)$이므로

$$V_{\min} = \frac{p(1-p)(1-2p)}{3-4p}.$$

n번 던진 결과의 평균

$T_1, \ldots, T_n$을 n번 던졌을 때의 T의 값이라 하자. 이들은 서로 독립이고, 각각 평균이 p, 분산이 $V(\lambda)$이다. 그러면

$$\mathrm{E}(\overline{T}) = p, \qquad \mathrm{Var}(\overline{T}) = \frac{V(\lambda)}{n}.$$

? **생각해 보기** T는 네 가지 값만 갖는다. 왜 $\overline{T}$를 정규분포로 다루어도 될까?

$\overline{T}$는 분산이 유한하고 서로 독립이며 같은 분포를 따르는 n개의 확률변수의 평균이고 $n > 30$은 충분히 크므로, 중심극한정리에 의해 하나의 T의 분포 모양이 어떠하든 $\overline{T}$는 근사적으로 $\mathrm{N}(p, V(\lambda)/n)$를 따른다. $\sigma_\lambda = \sqrt{V(\lambda)/n}$, $Z = (\overline{T} - p)/\sigma_\lambda \approx \mathrm{N}(0,1)$이라 하면, 대칭성 $\Phi(-z) = 1 - \Phi(z)$를 이용하여

$$\mathrm{P}(|\overline{T} - p| < b) = \mathrm{P}\Big(|Z| < \frac{b}{\sigma_\lambda}\Big) \approx \Phi\Big(\frac{b}{\sigma_\lambda}\Big) - \Phi\Big(-\frac{b}{\sigma_\lambda}\Big) = 2\Phi\Big(\frac{b}{\sigma_\lambda}\Big) - 1,$$

을 얻는다.

$b > 0$은 고정되어 있고 Φ는 순증가하므로, 이 값은 b/σ_λ가 가장 클 때, 즉 σ_λ가 가장 작을 때 가장 크며, 이는 $V(\lambda)$가 최소일 때이다. 그것은 $\lambda = (1-2p)/(3-4p)$에서 일어나고, 이때

$$\sigma_\lambda^2 = \frac{V_{\min}}{n} = \frac{p(1-p)(1-2p)}{(3-4p)n} = s^2.$$

따라서 λ가 변할 때 $\mathrm{P}(|\overline{T} - p| < b)$의 최댓값은 근사적으로 $2\Phi(b/s) - 1$이고, 이것이 요구된 결과이다.

검산

- $\lambda = 1$이면 $T = X$이고 공식은 $\mathrm{Var}(T) = p(1-p)$를 준다. $\lambda = 0$이면 $T = \frac{1}{2}Y$이고 공식은 $\frac{1}{4} \cdot 2p(1-2p)$를 준다. 둘 다 옳다.
- $V_{\min}$은 이 두 값보다 작다. 이는 $p(1-p)$에 $\frac{1-2p}{3-4p} < 1$을 곱한 값이기 때문이다. 두 동전을 결합하는 것이 어느 하나만 쓰는 것보다 낫다.
- 수치로 확인해 보자. $p = 0.3$, $n = 100$, $b = 0.03$일 때 최적의 λ는 $\frac{2}{9}$이고, $s \approx 0.0216$, $2\Phi(b/s) - 1 \approx 0.835$이다. 100번 던지기를 40 000회 시행한 모의실험에서는 0.836이 나왔고, $\lambda = 1$(동전 A만 사용)일 때는 약 0.487에 그쳤다.

흔한 실수

- $\mathrm{Var}(\frac{1}{2}(1-\lambda)Y) = \frac{1}{2}(1-\lambda)\,\mathrm{Var}(Y)$라고 쓰는 것. 상수는 제곱되어 분산 밖으로 나온다.
- 분산을 최소화하면 확률이 최대가 되는 **이유**를 말하지 않는 것. 이를 위해서는 Φ가 증가함수라는 것과, 모든 λ에 대하여 $\overline{T}$의 평균이 p라는 것이 필요하다.

STEP 너머

이것은 추정 이론(estimation theory)의 첫 맛보기이다. p의 모든 불편추정량 $\lambda X + \frac{1}{2}(1-\lambda)Y$ 가운데 분산이 최소인 것을 찾은 것이다. 답은 각 관측값에 그 분산에 반비례하는 가중치를 주는데, 이는 통계학에서 정밀도가 서로 다른 측정값을 결합할 때 쓰는 바로 그 규칙(역분산 가중법, inverse-variance weighting)이며, 최솟값 $ac/(a+c)$는 "병렬 저항" 공식 $1/V_{\min} = 1/a + 1/c$이다. Grimmett와 Welsh의 *Probability: An Introduction*(확률론 입문)은 마지막 단계의 바탕이 되는 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.