LGE - SNU AI Scientist 고급과정

확률통계 및 통계 방법론

과제2

2026년 01월

문제 1. 어떤 온라인 강의 플랫폼에서 한 사용자의 강의 시청 시간은 평균 30분, 표준편차 5분인 분포를 따른다고 하자. 서로 다른 사용자의 시청 시간은 서로 독립이라고 가정한다. 무작위로 $n = 100$명의 사용자를 선택하여 이들의 평균 시청 시간을 $\bar{X}$라 하자. $\bar{X}$가 29분 이상 31분 이하일 확률을 아래의 표를 참고하여 구하여라. (단, Z는 표준정규분포를 따르는 확률변수이다.)

z	$P(0 \le Z \le z)$
1.0	0.341
1.5	0.433
2.0	0.477
2.5	0.494
3.0	0.499

문제 2. 음악 스트리밍 서비스에서 임의로 한 명의 사용자를 선택했을 때, 그 사용자가 시스템이 추천한 곡을 끝까지 재생하면 1, 중간에 종료하면 0으로 기록한다고 하자. 이를 확률변수 X로 두면, 확률변수 X는 다음과 같은 분포를 가진다.

$$P(X=1)=0.7, \qquad P(X=0)=0.3$$

서로 다른 다섯 명의 사용자를 무작위로 선택하여 관측한 반응을 X_1, X_2, X_3, X_4, X_5라고 하자. 각 사용자의 반응은 서로 독립이며 동일한 확률분포를 따른다고 가정한다. 이때 표본평균을 $\bar{X} = \frac{X_1+X_2+X_3+X_4+X_5}{5}$라고 할 때, 다음의 물음에 답하여라.

(a) $\bar{X}$의 확률질량함수를 구하여라.

(b) $P(\bar{X} \geq 0.6)$을 구하여라.

(c) $E(\bar{X})$와 $\mathrm{Var}(\bar{X})$를 구하여라.

문제 3. 음악 스트리밍 서비스에서 임의로 한 명의 사용자를 선택했을 때, 그 사용자가 시스템이 추천한 곡을 끝까지 재생하면 1, 중간에 종료하면 0으로 기록한다고 하자. 이를 확률변수 X로 두고, 각 사용자의 반응은 서로 독립이며

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p$$

를 따른다고 하자. 서로 다른 $n = 100$명의 사용자를 무작위로 선택하여 얻은 결과를 $X_1, \ldots, X_{100}$이라 하고, 추천된 음악을 끝까지 재생한 사용자의 수를 $Y = \sum_{i=1}^{100} X_i$라 하자. 관측 결과 $Y = 70$으로 나타났다. 표본비율을 $\hat{p} = Y/100$이라 할 때, 모비율 p의 95% 신뢰구간을 근사적으로 구하여라. (단, $Z \sim N(0,1)$에 대하여 $P(Z \leq 1.64) = 0.95$, $P(Z \leq 1.96) = 0.975$이다.)

문제 4. 어떤 배달 주문 플랫폼에서는 하나의 주문 건에 대해 사용자가 남기는 리뷰 별점 유형을 세 가지로 분류한다. 각 반응은 다음과 같이 숫자 1, 2, 3으로 기록된다.

- 1: 낮은 수준의 만족도
- 2: 보통 수준의 만족도
- 3: 높은 수준의 만족도

한 사용자가 남기는 리뷰 별점을 확률변수 X로 나타낼 때, 이 반응은 다음과 같은 분포를 따른다고 알려져 있다.

$$P(X=1)=\theta, \quad P(X=2)=2\theta(1-\theta), \quad P(X=3)=(1-\theta)^2, \qquad (0<\theta<1).$$

이제 서로 독립인 사용자의 리뷰 별점 $X_1,\ldots,X_n$을 관측한 결과, 리뷰 별점이 1인 경우가 n_1번, 2인 경우가 n_2번, 3인 경우가 n_3번 관측되었을 때, ($n_1+n_2+n_3=n$이 성립함은 알려져있다) 모수 θ의 최대가능도추정량 $\hat{\theta}_{\text{MLE}}$을 구하여라.

Hint (로그 미분 공식)
양의 함수 $f(\theta)>0$에 대하여

$$\frac{d}{d\theta}\log f(\theta)=\frac{f'(\theta)}{f(\theta)}.$$

문제 5. 한 프로야구 타자가 실내 배팅 연습장에서 친 타구가 공중에 머무는 시간(단위: 초)을 X라 하자. 실내 환경에서 동일한 조건으로 타구가 발생하며, 각 타구는 서로 영향을 주지 않는다고 가정한다. 타구가 떨어질 때까지의 비행 시간 X는 평균이 μ인 지수분포

$$f(x;\mu) = \frac{1}{\mu}e^{-x/\mu}, \qquad x > 0$$

를 따른다고 하자. 이 타자가 친 n개의 독립적인 타구 비행 시간을 $X_1, \ldots, X_n$이라 할 때, 모수 μ의 최대가능도추정량 $\hat{\mu}_{\mathrm{MLE}}$을 구하여라.

Hint: 미분 공식

- (로그 미분 공식)
  양의 함수 $f(\mu) > 0$에 대하여
  $$\frac{d}{d\mu}\log f(\mu) = \frac{f'(\mu)}{f(\mu)}.$$

- (지수함수의 합성함수 미분)
  미분 가능한 함수 $g(\mu)$에 대하여
  $$\frac{d}{d\mu}e^{g(\mu)} = e^{g(\mu)}g'(\mu).$$
  특히,
  $$\frac{d}{d\mu}e^{-a/\mu} = e^{-a/\mu} \cdot \frac{a}{\mu^2}, \qquad (a > 0).$$

- (거듭제곱 함수의 미분)
  상수 k에 대하여
  $$\frac{d}{d\mu}\mu^{-k} = -k\mu^{-(k+1)}, \qquad \frac{d^2}{d\mu^2}\mu^{-k} = k(k+1)\mu^{-(k+2)}.$$