

LGE - SNU AI Scientist 고급과정

확률통계 및 통계 방법론

과제1

2026년 01월

문제 1. 한 프로야구 경기에 대해, 다음과 같은 두 사건을 정의하자.

- 사건 A : 해당 경기에서 홈팀이 승리한 사건
- 사건 B : 해당 경기에서 홈팀 선발 투수가 5이닝 이상을 소화한 사건

이 때, 사건에 대한 확률이 다음과 같이 주어졌다고 가정하자.

$$P(A) = 0.55, \quad P(B) = 0.60, \quad P(A \cup B) = 0.75$$

다음의 물음에 답하여라.

- (a) 해당 경기에서 홈팀이 승리하고, 홈팀 선발 투수는 5이닝 이상을 소화한 확률을 구하여라.
- (b) 해당 경기에서 홈 팀이 승리한 사건과, 홈팀 선발 투수가 5이닝 이상을 소화한 사건은 서로 독립인지 판단하여라.

문제 2. LLM(대규모 언어 모델, large language model)을 이용한 어떤 AI 에이전트는 LLM이 답변을 생성한 뒤, 출력 검증 단계에서 출력 형식과 정책 위반 여부를 검사한다. 한 요청에 대한 결과를 최종적으로 실패로 판단하는 경우는 다음 세 가지 원인 중 하나에 의해 발생한다고 하자.

- H : 환각(hallucination)으로 인해 사실 오류가 포함된 경우
- T : 도구 호출 과정에서 오류가 발생한 경우
- F : 출력 형식 규칙을 어긴 경우

운영 로그에 따르면, 임의의 요청에 대해 각 원인이 발생할 확률은 다음과 같다.

$$P(H) = 0.10, \quad P(T) = 0.05, \quad P(F) = 0.15$$

세 원인은 서로 배타적이라고 가정하자. 즉, 한 요청에서 실패 원인은 정확히 하나만 발생한다고 하자. 또한 원인별로 출력 검증 단계에서 해당 요청을 실패로 판정할 조건부확률은 다음과 같다.

$$P(S | H) = 0.90, \quad P(S | T) = 0.80, \quad P(S | F) = 0.70$$

여기서 S 는 최종적으로 실패로 기록된 사건을 의미한다. 이 때, 다음 물음에 답하여라.

- 전확률 공식을 이용하여 $P(S)$ 를 구하여라.
- 실패 S 가 발생했다고 할 때, 그 원인이 환각 H 일 확률 $P(H | S)$ 를 구하여라.
- 실패 S 가 발생했다고 할 때, 원인이 형식 오류 F 일 확률 $P(F | S)$ 를 구하여라.

문제 3. 한 공장에서 생산되는 금속 막대 길이의 편차(mm)를 확률변수 X 라 하자. 이 막대는 규격상 기준 길이로부터 최대 2mm 이내의 편차만 허용되며, 정밀 가공 공정 특성상 편차가 작을수록 더 자주 발생하고 큰 편차는 드물다. 확률변수 X 의 확률밀도함수는 다음과 같이 주어진다.

$$f(x) = \begin{cases} k(2-x), & 0 \leq x \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

다음의 물음에 답하여라.

(a) 상수 k 의 값을 구하여라.

(b) 제품의 길이 편차가 0.5mm 이상 1.5mm 이하일 확률 $P(0.5 \leq X \leq 1.5)$ 을 구하여라.

문제 4. 한 공장에서 생산되는 제품은 출하 전에 자동 검사 시스템을 거친다. 각 제품은 검사 결과에 따라 경고 신호가 발생하거나 발생하지 않으며, 다음과 같은 특징을 가진다.

- 각 제품에서 경고 신호가 발생할 확률은 0.1이다.
- 서로 다른 제품에서 경고 신호가 발생하는 사건들은 서로 독립이다.

임의로 선택한 제품 4개를 검사할 때, 다음 물음에 답하여라.

- (a) 경고 신호가 발생한 제품의 수가 3개 이상일 확률을 구하여라.
- (b) 4개의 제품 중 경고 신호가 발생한 제품 수의 기댓값과 분산을 구하여라.

문제 5. 온라인 시험 시스템에서는 시험 도중 응시자의 화면에서 시험 페이지 외의 창이 활성화된 경우가 감지되면 규칙 위반 알림이 로그로 기록된다. 이 로그는 다음과 같은 특징을 가진다.

- 각 이탈 행위는 개별적으로 발생하며 서로 독립적이다.
- 이탈 행위는 드물게 발생한다.
- 로그 분석 결과, 평균적으로 30분 동안 1.8회의 규칙 위반 알림이 기록된다.

임의로 선택한 10분 구간에서 발생하는 규칙 위반 알림의 개수를 확률변수 X 라 하자. 다음의 물음에 답하여라.

- (a) 확률변수 X 의 확률분포를 명시하여라.
- (b) 해당 10분 구간에서 규칙 위반 알림이 2회 이상 발생할 확률 $P(X \geq 2)$ 를 구하여라.