[2강]

통계학과 확률(1)

정보통계학과 이긍희 교수

학습목표

1. 확률을 이해한다.
2. 확률변수를 이해한다.
3. 확률분포의 성질을 이해한다.
4. 기댓값을 계산한다.

[2강]

통계학과 확률(1)

1 확률의 정의

1.1 통계적 추론과 확률

- 우리가 알고자 하는 세계인 모집단은 불확실
  → 불확실성은 확률, 확률분포로 측정

1.1 통계적 추론과 확률

- 확률은 동전던지기와 같은 확률 실험으로 이해
  - 동전 던지기, 주사위던지기
  - 확률실험의 규칙성

1.2 확률의 정의

- 표본공간(sample space) : 이산형, 연속형

예 2.1

동전 던지기 실험과 자동차 수명의 표본공간은?

1.2 확률의 정의

- 사건(event) : 표본공간의 부분집합

예 2.2

주사위 던지기에서 표본공간과 홀수가 나타날 사건은?

1.3 확률의 고전적 정의

- 이산형 : $P(A) = \dfrac{n(A)}{n(S)}$
- 연속형 : $P(A) = \dfrac{m(A)}{m(S)}$

1.4 확률의 공리적 정의

- 다음을 만족하는 측도 P가 확률
  ① $0 \leq P(A) \leq 1$
  ② $P(S) = 1$
  ③ $A_1, \cdots, A_i, \cdots$ 서로 배반
  $\rightarrow P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

1.4 확률의 공리적 정의

예 2.3

주사위를 던져서 홀수가 나타날 확률은?

1.5 확률의 연산

- 여사건의 확률 : $P(A^c) = 1 - P(A)$
- 합집합의 확률 :
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $A \cap B = \phi$: $P(A \cup B) = P(A) + P(B)$

1.5 확률의 연산

- 표본공간 : $B_1, \ldots, B_m$로 분할$\left(S = \cup_{i=1}^{m} B_i\right.$,

1.5 확률의 연산

예 2.4

동전을 2번 던질 때 앞면이 적어도 한 번 나올 확률은?

2 조건부 확률과 독립

2.1 조건부 확률

- 사건 B 발생 조건 하 사건 A가 발생할 조건부 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2.1 조건부 확률

예 2.5

주사위 눈이 짝수라는 조건 하에서 한 번 던져서 3 이하의 숫자가 나올 확률은?

2.2 역확률

- 역확률(inverse probability) :

$$P(A|B) \rightarrow P(B|A)$$

- $P(B|A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(A|B)P(B)}{P(A)}$

2.2 역확률

- 표본공간을 B와 B^c으로 분할할 경우
  - $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$
  - $P(B|A) = \dfrac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)}$

2.3 베이즈(Bayes) 정리

- 표본공간을 $B_1, \dots, B_k$로 분할, A가 발생하였다는 정보가 주어졌을 때, B_i의 조건부확률

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + \cdots + P(A|B_k)P(B_k)}$$

2.3 베이즈(Bayes) 정리

- 역추정 기반을 제공 :

"원인$(B) \rightarrow$ 결과(A)" $\Rightarrow$ "결과$(A) \rightarrow$ 원인(B)"

2.4 독립성

- 사건 A와 사건 B는 서로 독립(independent)

$$P(B|A) = P(B) \rightarrow P(A \cap B) = P(A)P(B)$$

2.4 독립성

- n개 사건의 독립 :

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)$$

2.4 독립성

예 2.6

500원 동전과 100원 동전을 동시에 던질 때 사건 A와 B가 서로 독립인지 밝혀라.

A : 500원 앞면인 사건, B : 100원 앞면인 사건

3 확률변수와 확률분포

3.1 확률변수

- **확률변수(Random Variable)** : 사건을 숫자로 변환해 주는 함수
  - 표본공간 정의역, 실수 공역으로 하는 함수

3.1 확률변수

- 확률변수의 구분
  - 이산형 확률변수
  - 연속형 확률변수

3.2 확률분포

- **확률분포** : 확률변수 X로부터 유도되는 측도 P
  - 확률적 실험의 규칙성

- **누적확률분포함수** :
  - $F(x) = P(X \le x) = P(-\infty, x)$

3.2 확률분포

- 누적확률분포함수의 특성
  - $\lim_{x \to -\infty} F(x) = 0, \ \lim_{x \to \infty} F(x) = 1$
  - $F(x)$는 x의 비감소함수
  - $F(x)$는 오른쪽 방향으로 연속

3.2 확률분포

- 점확률로 표현되는 확률분포 :
  - 확률질량함수(이산형 확률변수)
  - 확률밀도함수(연속형 확률변수)

3.2 확률분포

- **확률질량함수** : 이산형 확률변수 X의 분포

$$f(x) = P(X = x), \qquad x = 0,1,2,\ldots$$

$$F(x) = P(X \le x) = \sum_{t=0}^{x} f(t)$$

3.2 확률분포

- 확률질량함수의 성질
  - ① $0 \le f(x) \le 1, \quad x = 0,1,2,\ldots$
  - ② $\sum_{x=0}^{\infty} f(x) = 1$
  - ③ $P(a \le X \le b) = \sum_{x=a}^{b} f(x)$

3.2 확률분포

- **확률밀도함수** : 연속형 확률변수 X의 분포

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad f(x) = \frac{d}{dx}F(x)$$

3.2 확률분포

- 확률밀도함수의 성질

  ① $f(x) \geq 0$

  ② $\int_{-\infty}^{\infty} f(x)dx = 1$

  ③ $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$

3.2 확률분포

예 2.7

2개의 불량품, 2개의 정상제품 상자에서 2개의 제품을 동시에 꺼낼 때 불량품 개수를 X라 할 때 X의 확률질량함수는?

3.2 확률분포

예 2.8

정확히 20분 간격으로 도착하는 버스,
버스 기다리는 시간 X의 확률밀도함수와
누적확률분포함수는?

3.2 확률분포

예 2.9

확률밀도함수가 다음과 같을 때 상수 C와 $P\left(\frac{1}{2} < X \leq 2\right)$는?

$$f(x) = \begin{cases} Ce^{-2x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

3.2 확률분포

예 2.9

확률밀도함수가 다음과 같을 때 상수 C와 $P\left(\frac{1}{2} < X \le 2\right)$는?

4 기댓값

4.1 기댓값의 정의

- **기댓값 $E(X)$** : 확률분포의 무게중심

$$E(X) = \begin{cases} \sum_{x=0}^{\infty} xf(x), & X: \text{이산형 확률변수} \\ \int_{-\infty}^{\infty} xf(x)dx, & X: \text{연속형 확률변수} \end{cases}$$

4.1 기댓값의 정의

예 2.10

동전 2번 던져서 나온 앞면의 수 X의 기댓값을 구하라.

4.1 기댓값의 정의

예 2.11

확률밀도함수 $f(x) = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & x < 0,\ x > 1 \end{cases}$

에서 X의 기댓값은?

4.2 확률변수 X의 함수 $g(X)$의 기댓값

$$E[g(X)] = \begin{cases} \displaystyle\sum_{x=0}^{\infty} g(x)f(x), & X: \text{이산형} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f(x)dx, & X: \text{연속형} \end{cases}$$

4.3 분산과 표준편차

- **분산** : 모집단이 중심(기댓값)으로부터 흩어진 정도를 측정

$$Var(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

4.3 분산과 표준편차

- **표준편차** : 확률변수와 단위 일치

$$Sd(X) = \sqrt{Var(X)}$$

4.3 분산과 표준편차

예 2.12

동전을 2번 던져서 나온 앞면의 수 X의 분산과 표준편차는?

4.4 기댓값의 성질

- a, b가 상수이고 X, Y가 확률변수
  ① $E(aX + b) = aE(X) + b$
  ② $\mathrm{E}(aX + bY) = aE(X) + bE(Y)$
  ③ $\mathrm{E}(X_1 + \cdots + X_n) = E(X_1) + \cdots E(X_n)$

4.4 기댓값의 성질

- a, b가 상수이고 X가 확률변수

① $\mathrm{Var}(\mathrm{aX} + b) = a^2 Var(X)$

② $Sd(aX + b) = |a| Sd(X)$

4.4 기댓값의 성질

예 2.13

a, b가 상수일 때 $E(aX+b) = aE(X) + b$를 증명하여라.

4.4 기댓값의 성질

예 2.14

a, b가 상수, $Var(aX + b) = a^2 Var(X)$를 증명하여라.

4.4 기댓값의 성질

예 2.15

$X \sim (\mu, \sigma^2)$, $Y = \frac{X-\mu}{\sigma}$의 기댓값과 분산은?

5 정리하기

정리하기

1. 확률은 어떤 사건(A)이 일어날 가능성을 0과 1 사이의 실수로 표현한 것이다.

2. B 발생 조건 하의 A가 발생 조건부 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

정리하기

3. 사건 A, B의 독립성은 다음과 같이 정의된다.

$$P(A \cap B) = P(A)P(B)$$

4. 확률변수는 확률적 실험에서 실험결과를 관심의 대상이 되는 수 값으로 나타낸 함수이다.

정리하기

5. 확률질량함수는 이산형 확률변수의 분포를 결정하는 함수이며, 확률밀도함수는 연속형 확률변수의 분포를 결정하는 함수이다.

정리하기

6. 기댓값과 분산은 다음과 같이 정의된다.

$$E(X) = \begin{cases} \displaystyle\sum_{x=0}^{\infty} xf(x), & X\text{: 이산형 확률변수} \\ \displaystyle\int_{-\infty}^{\infty} xf(x)dx, & X\text{: 연속형 확률변수} \end{cases}$$

$$Var(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

다음 시간 안내 ▼

3강. 통계학과 확률(2)

수고하셨습니다.