일반통계학
제 3장
확률과 확률분포

통계학 2016.1학기 정혜영

1 확률의 정의

1.1 Why?

모집단에서 일부만 관측하고 이를 바탕으로 모집단 전체에 대한 결론을 이끌어 내는데 논리적 근거가 된다.

1.2 용어정리

(1) 표본공간(sample space)
- 통계적 조사에서 얻을 수 있는 모든 가능한 결과들의 집합

예) 주사위 눈금의 수

(2) 사건(event)
- 표본공간의 부분집합 (관심이 있는 실험 결과의 집합)

예) 주사위의 1과 2의 눈금이 나오는 경우

전구의 수명시간이 10이상인 경우

(3) 근원사건(elementary event), 단순사건(simple event)
-한 개의 원소로 이루어진 사건

1.3 사건의 연산

사건에 대하여

- 합사건 :
- 곱사건 :
- 여사건 :
- 이면 A와B 는 서로 배반사건

※ 공집합은 모든 집합과 배반임.

1.4 분할

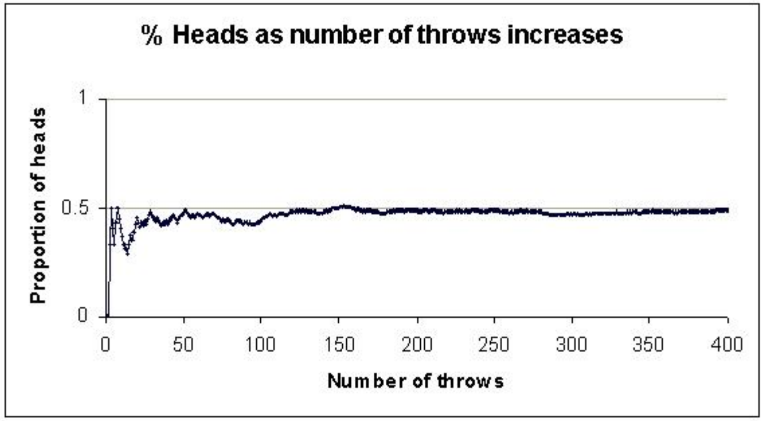
1.5 확률의 정의

고전적 확률

• 각각의 근원사상이 일어날 가능성이 같다는 가정
•표본공간 S가 n개의 근원사건으로 이루어져 있고, 각 근원사건이 일어날 가능성이 같다면, 확률은 (근원사건의 개수)/n
•예:
•단점 :

빈도적 확률

•무수히 많이 시행하였을 때 그 사건이 일어난 비율(relative frequency)이 수렴해 가는 값
•동전을 던져서 나오는 앞면이 나올 확률 = 0.5



1 확률의 정의

•단점 :

주관적 확률

•각자 생각하고 있는 어떤 사건이 일어날 가능성에 대한 믿음의 정도(degree of belief)

•예)

•단점 :

공리적 확률 - Kolmogorov(1903-1987)

• 현대수학자들이 보통 생각하는 확률개념

•큰 수의 법칙으로 빈도적 확률로 이어지고 고전적 확률을 포함.

(공리 1)

(공리 2)

(공리 3)

1.6 확률의 성질

(1) $P(\emptyset)=0$

(2) $P(A \cup B)=P(A)+P(B)-P(A \cap B)$

(3) $P(A^c)=1-P(A)$

(4) $A \subset B$ 이면 $P(A) \leq P(B)$

2.1 조건부 확률

- 사건 A가 주어졌을 때 사건B가 일어날 확률 (단, P(A)>0)

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

- 예) 두개의 주사위를 던지는 실험에서 첫 번째 던진 주사위 눈이 두 번째 던진 주사위 의 눈보다 클 때 두 주사위 눈의 합이 10일 확률은?

과제) 조건부 확률 P(· |A)가 표본공간 S에서의 확률임을 증명하여라.

- 곱셈법칙 : $P(A) > 0, P(B) > 0$ 이면 $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

예) 불량품 20개와 양호품 80개로 구성된 로트에서 2개의 제품을 단순랜덤추출 할 때, 2개 모두 불량품일 확률을 구하여라.

2.2 전확률 공식

사건 $A_1, A_2, \cdots, A_n$에 대하여 $A_i \cap A_j = \emptyset (i \neq j), A_1 \cup A_2 \cup \cdots \cup A_n = S$, $P(A_i)>0$ 이면 $P(B) = P(B|A_1)P(A_1) + \cdots P(B|A_n)P(A_n) = \sum_{k=1}^{n} P(B|A_k)P(A_k)$

예) 통계학과 : 1학년(30%), 2학년(25%), 3학년(25%), 4학년(20%)
수학과목 수강 : 1학년의 50%, 2학년의 30%, 3학년의 10%, 4학년의 2%
통계학과 학생 중 한 학생을 단순랜덤추출하였을 때 그 학생이 수학 과목의 수강생일 확률을 구하여라?

2.3 베이즈 정리

표본공간 S의 분할 $A_1, A_2, \cdots, A_n$ 과 $P(A_i)$>0 , $P(B)$>0 에 대하여

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

예) 어떤 지역의 결핵환자의 비율이 0.001로 알려져 있다. 결핵에 결려있는지를 알아보는 검사에서 결핵에 걸렸을 때 양성 반응이 나타날 확률은 0.95이고 그렇지 않을 때 양성 반응이 나타날 확률은 0.011이라고 한다. 양성 반응이 나타났을 때 결핵에 걸렸을 확률을 구하여라.

2.4 서로 독립 (mutually independent)

- $P(B|A) = \mathrm{P}(B)$ 이면 사건 A와 사건 B는 서로 독립이라 한다.
- 두 사건 A 와 B가 독립이면 $P(A \cap B) = P(A)P(B)$, $P(A|B) = \mathrm{P}(A)$, $P(B|A) = \mathrm{P}(B)$

예) "불량품 20개와 양호품 80개로 구성된 로트에서 2개의 제품을 단순랜덤추출할 때, 2개 모두 불량품일 확률을 구하여라. " 이 문제에 제시된 두 사건은 독립인가? 만약 단순랜덤복원추출을 한다면 독립인가?

- A_1, A_2,⋯, A_n에 대하여 다음이 성립할 때, A_1, A_2,⋯, A_n은 서로 독립이라고 한다.

P($A_i \cap A_j$)=P(A_i)P(A_j)

P($A_i \cap A_j \cap A_k$)=P(A_i)P(A_j)P(A_k) $(1 \le i < j < k \le n)$

$\cdots$

P($A_1 \cap A_2 \cap \cdots \cap A_n$)=P($A_1$)P($A_2$) ⋯P($A_n$)

3.1 동전 1개를 던져 나오는 결과를 관측하는 실험

- 표본공간 :S ,S={H,T}, H ='표면',T='이면',
- 확률 : P ({H})=0.5 P({T})=0.5
- 위와 같이 표본공간의 원소, 즉 가능한 결과가 숫자가 아닐 경우 표본공간과 확률을 나타내는 것이 복잡해진다. 따라서, 다음과 같이 표본공간 S 에서 정의되어 실수값을 갖는 함수 $X : S \rightarrow \boldsymbol{R}$ 를 정의하고 X를 확률변수라고 부른다.

예) 함수 X를 X(H)=1, X(T)=0로 정의하면 새로운 표본공간 S_X 와 확률 P_X 를 얻을 수 있다.

(S_X, P_X) : 표본공간 S_X = {0,1}
확률 P_X : P_X ({0})=P({T})=0.5 , P_X ({1})=P({H})=0.5
우리는 확률변수 X를 이용해서 P({H})=P(X=1)이라고 표기한다.

3.2 확률변수

- 표본공간 위에 정의된 실수값 함수 ($X : S \longrightarrow \mathbf{R}$)

(1) 이산형 확률변수

- 확률변수 X 가 취할 수 있는 모든 값이 셀 수 있을 경우

예) 동전을 2번 던지는 실험을 한다. X=앞면의 수

(2) 연속형 확률변수

-확률변수 X 가 어떤 구간 내의 모든 값을 취할 수 있는 경우

예) 표본공간 $S=\{x : x \in [0,1]\}$ 일 때, 확률변수X 를 바늘이 가리키는 눈금이라고 하면, 확률변수X 가 취할 수 있는 값은 $\{x: x \in [0,1]\}$ 이다.

3.3 확률분포

-확률변수의 값에 따라 확률이 어떻게 흩어져 있는지를 합이 1인 양수로서 나타낸 것

예)

x	0	1	합계
P(X=x)	1/2	1/2	1

- 확률밀도함수(probability density function)

1. 이산형 확률변수 X의 확률밀도함수 (일반적으로 확률질량함수라고 부름 probability mass function , pmf)

확률변수 X가 취할 수 있는 값이 $x_1, x_2, x_3, \ldots$일 때, 확률질량함수 p(x)를 다음과 같이 정의한다.

$$p(x) = \begin{cases} P(X = x_i) & x = x_i, i = 1,2,\ldots \\ 0 & o.w \end{cases}$$

이산형 확률질량함수의 성질

$(1) 0 \le p(x) \le 1,\ \sum_{\text{모든 } x} p(x) = 1$

$(2) P(a < X \le b) = \sum_{a<x\le b} p(x)$

예) 동전을 2번 던지는 실험에서 앞면의 수 X에 대한 확률분포

X	0	1	2
P(x=X)	1/4	1/2	1/4

2. 연속형 확률변수의 확률밀도함수(probability density function, pdf)

연속확률변수 X에 대하여 함수 $p(x)$가

$(1) p(x) > 0, \int_{-\infty}^{\infty} p(x)dx = 1$

$(2) P(a \le X \le b) = \int_{a}^{b} p(x)dx \quad$ (단, $-\infty \le a < b \le \infty$)

를 만족시킬 때, $p(x)$를 X의 확률밀도함수라고 한다.

예) 바늘이 구간 [a,b] (a,b ∈[0,1]) 사이의 눈금에 멈출 확률

X=바늘이 저절로 멈추면서 가리키는 눈금

$P(a \le X \le b)$=(b-a)/(1-0)=b-a (단, $0 \le a < b \le 1$)

$$p(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{o.w} \end{cases}$$

- 연속확률변수의 성질

(1)임의의 상수 c에 대하여 P(c)=0

(2) $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$

4.1 확률변수의 기대값(expected value) 또는 평균(mean)

$p(x)$ 를 X의 확률밀도함수라고 할 때, 확률변수 X의 기대값

$$E(X) = \begin{cases} \sum_{\text{모든 } x} xp(x) & X: \text{이산확률변수} \\ \int_{-\infty}^{\infty} xp(x)dx & X: \text{연속확률변수} \end{cases}$$

예) 동전을 2회 던질 때 표면의 개수 X 의 확률분포와 기대값

$$0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$$

예) 한 사람이 오후 12시부터 1시 사이에 우연히 정거장에 오는 시간 X의 확률분포와 기대값

4.1 확률변수의 기대값(expected value) 또는 평균(mean)

- 함수 g (**R**→**R**)에 대하여 확률변수g(X)의 기대값

$$E(g(X)) = \begin{cases} \sum_{\text{모든 } x} g(x)p(x) & X\text{: 이산확률변수} \\ \int_{-\infty}^{\infty} g(x)p(x)dx & X\text{: 연속확률변수} \end{cases}$$

예) $Y=(X-1)^4$의 확률분포와 기대값 (X는 앞의 동전 2회 던지는 실험의 확률변수)

- **기대값의 성질**

(1) 임의의 상수 a, b에 대하여 $E(aX+b) = aE(X) + b$
(2) 함수 g_1, g_2와 임의의 상수 a, b에 대하여,

$$E(ag_1(X) + bg_2(X)) = aE(g_1(X)) + b\mathrm{E}(g_2(X))$$

(3) $X \geq 0$이면 $E(X) \geq 0$

4.2 분산 (variance)과 표준편차 (standard deviation)
X의 평균이 μ이고 X의 확률밀도함수가 $p(x)$일 때,

(1) $\mathrm{Var}(X) = E[(X-\mu)^2] = \begin{cases} \sum_{\text{모든 } x} (x-\mu)^2 p(x) & X\text{: 이산확률변수} \\ \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx & X\text{: 연속확률변수} \end{cases}$

(2) sd(X)=$\sqrt{\mathrm{Var}(X)}$

(3) 분산의 간편계산 : $\mathrm{Var}(X) = \mathrm{E}(X^2) - [E(X)]^2$
(4) 분산의 성질 : $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$ $(a, b$는 상수)