

사회통계

확률의 기초

김현우, PhD¹

¹충북대학교 사회학과 조교수



진행 순서

- 1 확률 개념의 이해
- 2 확률법칙
- 3 결합확률, 한계확률과 조건부확률
- 4 베이즈 정리

확률 개념의 이해

확률 개념의 이해

확률이론의 역사는 도박과 밀접한 연관을 갖고 있었다.

- 17세기 프랑스 도박사였던 Antoine Gombaud는 도박과 판돈의 분배 문제에 관해 고민하다가 결국 해법을 찾는데 실패하고, 당대의 천재 수학자였던 Blaise Pascal과 Pierre de Fermat에게 이에 관해 물어보았다.
- 이에 Pascal과 Fermat는 서신을 교환하면서 확률이론(probability theory)의 토대를 쌓기 시작했다.
- 감사하라. 이들 덕택에 여러분이 이 고생을 하는 거다.



Gombaud



Pascal



Fermat



확률 개념의 이해

왜 사회학에서 확률이론을 필요로 할까?

- 우리는 모집단(population)에서 표본(sample)을 골라낼 것이다.
- 그 뒤, (모집단을 전수조사하지 않고) 표본만을 관찰하여 어떤 성질을 찾을 것이다.
- 우리는 그 표본의 성질로부터 모집단의 성질을 유추한다(make inference).
- 표본으로부터 유추된 성질이 (모집단과는 상관없이) 그저 표본만에서 우연히 나타났을 확률을 알고 싶기 때문에 확률이론을 필요로 한다.



확률 개념의 이해

예제 1. 동전을 한 번 던져서 앞면이 나올 가능성은 $1/2$ 이다. 그렇다면 동전을 두 번 던져서 두 번 모두 앞면이 나올 가능성은 얼마인가?



확률 개념의 이해

- 일단 동전을 (한 번) 던졌을 때, 있을 수 있는 모든 결과(all possible outcomes)는
암만 생각해도 두 가지 뿐이다: 앞면(head)과 뒷면(tail).
- 앞면(H)이 나올 확률은 $1/2$, 뒷면(T)이 나올 확률도 $1/2$.
- 이제 두 번 던졌을 때, 있을 수 있는 모든 결과(outcomes)의 집합(set)은 다음과 같다

$$S = \{HH, HT, TH, TT\}.$$

- “답은 $1/4$ 이구나.”
- 이 아이디어를 (동전 던지기 뿐만 아니라) 모든 현상에도 일반화할 수 있다.



확률 개념의 이해

이 예제로부터 몇 가지 유용한 개념을 배울 수 있다.

- 확률(probability)이란 “어떤 사건(event)이 발생할 가능성을 나타낸 값”이다.
- 사건이란 “표본공간(sample space)의 특정한 부분집합”을 의미한다.
- 앞의 예제에서 사건은 “두 번 모두 앞면(HH)”으로 문제에서 주어졌다.
- 표본공간(sample space) S 란 “어떤 시행(trial) 또는 실험(experiment)에서 가능한 모든 결과(all possible outcomes)의 집합”을 의미한다.
- 앞의 예제에서 표본공간은 $S = \{HH, HT, TH, TT\}$ 이다.



확률 개념의 이해

- **결과(outcome)**는 시행 또는 실험에서 수집된 관측치를 뜻한다.
- 앞의 예제에서 “동전을 두 번 던진” 결과들은 HH, HT, TH, TT 로 모두 4개가 있다.
- 결과와는 달리 사건은 연구자의 관심에 따라 달라질 수 있다!
- 앞에서는 “두 번 모두 앞면”이라는 사건에 관심을 두었는데, 만일 “적어도 한 번은 앞면”이라는 사건으로 관심을 바꿀 수도 있다. 이 경우 사건은 HH, HT, TH 이다.



확률 개념의 이해

확률을 몇 가지 방식으로 정의할 수 있다.

- 고전적 확률(classical probability)은 “논리적으로 생각하여” 알 수 있는 가능성을 확률로 정의하는 입장이다.
- 경험적 관찰에 근거하지 않고 논리적 판단에 의해 정의된다. 이 때문에 분석적 확률이라고도 불리운다.
- 표본공간 $S = \{s_1, s_2, \dots, s_N\}$ 에서 각각의 결과(outcome)가 일어날 가능성이 동등하다면, m 개의 결과(outcomes)를 포함하는 사건(event)의 확률은 다음과 같다.

$$P(E) = \frac{m}{N}$$



확률 개념의 이해

- **경험적 확률(empirical probability)**은 “장기간에 걸쳐 사건의 발생을 관찰하여(in a long run of observations) 얻은 상대적인 빈도”로 확률을 정의하는 관점이다.
- 주사위를 단 몇 차례만 던졌다면, 6이 나올 확률은 $1/6$ 이 아닐 수도 있다.
- 그러나 주사위를 제법 여러 차례 던졌다면, 6이 나올 확률은 $1/6$ 에 제법 근접할 것이다.
- 주사위를 무한히 많이 던진다면, 6이 나올 확률은 $1/6$ 에 무한히 근접할 것이다.
- 고전적 접근은 확률을 분석적으로 **연역**하고 있지만, 경험적 접근은 확률을 자료로부터 **귀납**한다.



확률 개념의 이해

확률은 사건에 대응하여 정의된다.

- “오늘 비가 내릴” 가능성이나 “동전을 열 번 던져 모두 뒷면이 나올” 가능성과 같이 관심있는 사건(event)이 먼저 정의되어 있어야 한다.
- 그 뒤 개별적인 사건에 대응하여 각각의 발생 확률을 제시할 수 있어야 한다.

사건	확률
H	$1/2$
T	$1/2$



확률 개념의 이해

이때 사건과 확률은 반드시 다음의 두 가지 법칙을 따라야 한다.

- 포괄적이고 상호배타적 사건들의 확률의 합은 1이다.
- 첫째, (구분한 후에 얻어진) 자항의 총합이 모항과 같아야 한다(exhaustive). 예컨대 “10대”, “20대”, “30대 이상”은 자항의 총합이 모항과 같지 않은 구분이다.
- 둘째, (구분한 후에 얻어진) 자항들은 서로 배제해야 한다(mutually exclusive). 예컨대 “군필”, “미필”, “장교 전역”은 서로 배제하지 못하는 구분이다.
- 이에 따라 표본공간 S 에 대하여 $P(S) = 1$ 이다.

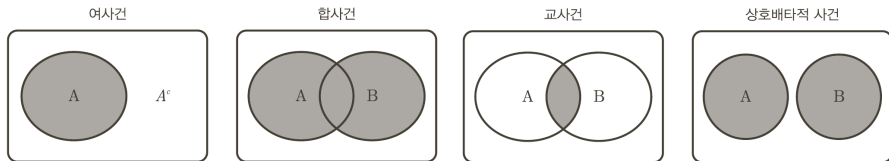


확률법칙

확률법칙

둘 이상의 사건들 사이 관계를 그림으로 표현할 수 있다.

- 흔히 **벤 다이어그램(Venn Diagram)**으로 두 사건의 확률이 어떤 관련을 갖고 있는가를 나타낸다.
- A 와 B 의 **합사건(union)**은 $A \cup B$ 로 표현한다
- A 와 B 의 **교사건(intersection)**은 $A \cap B$ 로 표현한다.
- A 의 **여사건(complement)**은 A^C 로 표현한다



둘 이상의 사건들 사이 관계에 관하여 두 가지 기본 법칙이 성립한다.

- 첫째, 두 사건 A 와 B 중 “최소한 한 사건이 발생”할 확률에 관해 다음의 **덧셈 법칙 (addition rule)**이 성립한다.

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

- 여기서 $P(A \cap B)$ 는 (곧 설명할) **결합확률(joint probability)**이다.
- 만일 사건 A 와 B 가 상호배타적(=배반사건)이라면 다음과 같이 단순화된다.

$$P(A \cup B) = P(A) + P(B)$$



- 둘째, 두 사건 A 와 B 가 “둘 다 발생할 확률”에 관해 다음의 곱셈 법칙 (multiplication rule)이 성립한다.

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

- 여기서 $P(A|B)$ 또는 $P(B|A)$ 는 (곧 설명할) 조건부확률(conditional probability)이다.



두 사건의 관계를 독립사건 혹은 종속사건으로 이해할 수 있다.

- 두 사건 A 와 B 가 있을 때, $P(A|B) = P(A)$ 이거나 $P(B|A) = P(B)$ 이거나 $P(A \cap B) = P(A)P(B)$ 이면 두 사건은 독립(independent) 사건이다(Why?).
- (두 사건은 독립 아니면 종속이므로) 위 식이 성립하지 않으면 두 사건은 종속(dependent) 사건이다.
- A 와 B 가 독립적인 사건들이라면 덧셈 법칙은 좀 더 단순해진다(Why?).

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

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



- 세 독립사건인 경우에는 다음이 성립한다(Why?).

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3) \\ &\quad - P(E_1 E_2) - P(E_1 E_3) - P(E_2 E_3) + P(E_1 E_2 E_3) \end{aligned}$$

- 이와 마찬가지로 세 독립사건인 경우에 다음이 성립한다(Why?).

$$P(E_1 \cap E_2 \cap E_3) = P(E_1) \cdot P(E_2) \cdot P(E_3)$$

- A와 B가 상호배타적(=배반사건)이라면, A와 B는 함께 일어날 수 없으므로 독립사건이 될 수조차 없다(Why?).



결합확률, 한계확률과 조건부확률

결합확률, 한계확률과 조건부확률

지금까지 배운 것을 토대로 자료 설명에 유용한 확률 개념을 도출한다.

- 어떤 자료가 주어졌을 때, “두 사건이 동시에 일어날 확률”을 **결합확률(joint probability)**이라고 부른다.

$$\begin{aligned}P(A \cap B) &= P(A|B) \cdot P(B) \\ &= P(B|A) \cdot P(A)\end{aligned}$$

- 결합확률은 곱셈법칙을 그대로 이용한다.
- 위 수식의 의미를 천천히 음미해보아야 한다.



결합확률, 한계확률과 조건부확률

- 조건부확률(conditional probability)이란 “다른 사건(B)이 이미 일어났다는 전제 아래, 한 사건(A)이 일어날 확률”을 의미한다.

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

- 조건부확률 또한 곱셈법칙을 그대로 이용한다(Why?).
- 결합확률과는 달리 조건부확률에서는 교환법칙(commutative property)이 성립하지 않는다.

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

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



결합확률, 한계확률과 조건부확률

- 한계확률(marginal probability)은 “연관된 결합확률의 합”이다.

$$P(A) = P(A \cap B) + P(A \cap B^C) \quad (1a)$$

$$= P(A|B) \cdot P(B) + P(A|B^C) \cdot P(B^C) \quad (1b)$$

- (1b)는 조건부확률 $P(A|B)$ 의 정의를 이용하여 (1a)에서 도출된 것이다.



결합확률, 한계확률과 조건부확률

우리는 사실 이 개념들을 이미 배웠고 심지어 엑셀에서 활용까지 했다.

- 자료부터 교차표를 만든 뒤, 행 합계(row total)로 표준화한다.
- 아래 교차표를 들여다보면 $P(\text{종교인}|\text{여자}) = 0.44$, $P(\text{비종교인}|\text{여자}) = 0.56$, $P(\text{종교인}|\text{남자}) = 0.23$, $P(\text{비종교인}|\text{남자}) = 0.77$ 임을 알 수 있다.
- 이것이 바로 **조건부확률(conditional probability)**이다.

	종교인	비종교인	행 합계
여자	97	124	221
남자	68	232	300
열 합계	165	356	521

	종교인	비종교인	행 합계
여자	0.44	0.56	1
남자	0.23	0.77	1
열 합계	0.32	0.68	1



결합확률, 한계확률과 조건부확률

- 만약 열 합계(column total)로 표준화한 경우 다른 해석을 해야 한다(Why?)
- 아래 교차표를 들여다보면 $P(\text{여자}|\text{종교인})$, $P(\text{남자}|\text{종교인})$, $P(\text{여자}|\text{비종교인})$, $P(\text{남자}|\text{비종교인})$ 을 구할 수 있다.
- 이것들도 물론 조건부확률(conditional probability)이다.

	종교인	비종교인	행 합계
여자	97	124	221
남자	68	232	300
열 합계	165	356	521

	종교인	비종교인	행 합계
여자	0.59	0.35	0.42
남자	0.41	0.65	0.58
열 합계	1	1	1



결합확률, 한계확률과 조건부확률

- 마지막으로 총 합계(grand total)로 표준화한 경우 다른 해석을 해야 한다.
- 아래 교차표를 들여다보면 $P(\text{종교인} \cap \text{여자}) = 0.19$, $P(\text{비종교인} \cap \text{여자}) = 0.24$, $P(\text{종교인} \cap \text{남자}) = 0.13$, $P(\text{비종교인} \cap \text{남자}) = 0.45$ 임을 알 수 있다.
- 이것이 바로 **결합확률(joint probability)**이다.

	종교인	비종교인	행 합계
여자	97	124	221
남자	68	232	300
열 합계	165	356	521

	종교인	비종교인	행 합계
여자	0.19	0.24	0.42
남자	0.13	0.45	0.58
열 합계	0.32	0.68	1



결합확률, 한계확률과 조건부확률

- 한계확률(marginal probability) 역시 교차표에서 계산할 수 있다.

$$\begin{aligned}P(\text{종교인}) &= P(\text{종교인} \cap \text{여자}) + P(\text{종교인} \cap \text{남자}) \\&= P(\text{종교인}|\text{여자}) \cdot P(\text{여자}) + P(\text{종교인}|\text{남자}) \cdot P(\text{남자})\end{aligned}$$

$$\begin{aligned}P(\text{비종교인}) &= P(\text{비종교인} \cap \text{여자}) + P(\text{비종교인} \cap \text{남자}) \\&= P(\text{비종교인}|\text{여자}) \cdot P(\text{여자}) + P(\text{비종교인}|\text{남자}) \cdot P(\text{남자})\end{aligned}$$

- $P(\text{여자})$ 와 $P(\text{남자})$ 도 연습삼아 계산해보자.



결합확률, 한계확률과 조건부확률

- 교차표가 2×2 가 아니라 3×3 라도 조금 더 길어지지만 본질은 똑같다.

	보수적 성역할 태도	중립적 성역할 태도	진보적 성역할 태도	합계
소득: 상	486	761	279	1526
소득: 중	1024	1080	967	3071
소득: 하	662	822	434	1918
합계	2172	2663	1680	6515

- 여기서도 표준화 방식에 따라 상이하게 조건부확률과 결합확률을 구할 수 있다.
- 한계확률도 마찬가지다. 다만 2×2 방식이 아니므로,

$$P(A) = P(A \cap B) + P(A \cap B^C)$$

대신 다음과 같이 표현한다.

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$



베이지 정리

베이즈 정리

이제 인공지능 분야에서 새롭게 피어난 확률이론의 꽃을 살펴보자.

- 지금까지 배운 개념들을 토대로 **베이즈 정리(Bayes' Theorem)**를 살펴보자

$$\begin{aligned}P(A|B) &= \frac{P(A \cap B)}{P(B)} && (\text{if } P(B) \neq 0) \\&= \frac{P(B|A) \cdot P(A)}{P(B)} \\&= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^C) \cdot P(A^C)}\end{aligned}$$

- **사전확률(prior probability)**인 $P(A)$ 와 $P(A^C)$ 그리고 **우도(likelihood)**인 $P(B|A)$ 를 알면 **사후확률(posterior probability)** $P(A|B)$ 를 알 수 있다.
- 좌변의 $P(A|B)$ 가 우변에서는 모두 $P(B|A)$ 와 같이 바뀌었음에 주목하자!



예제 2. 사지선다 문제를 맞추었을 때 알고 풀었을 확률을 구하시오(단, 답을 아는 사전확률과 모르는 사전확률은 같다고 전제한다).



베이즈 정리

- “답을 알(K)” 사전확률과 “답을 모르는(DK)” 사전확률이 같다면 이는 상호배타적 사건임을 의미한다.

$$P(A_K) = P(A_{DK}) = \frac{1}{2}$$

- (다른 문제도 그렇지만 특히) 베이즈 확률을 계산할 때는 문제가 묻는 것이 정확히 무엇인지 먼저 식별해야 한다.
- 여기서는 “답을 맞추었다(C)”는 조건 아래 답을 알(A_K) 확률, 즉 $P(A_K|C)$ 를 묻고 있다.

$$P(A_K|C) = \frac{P(C|A_K) \cdot P(A_K)}{P(C|A_K) \cdot P(A_K) + P(C|A_{DK}) \cdot P(A_{DK})}$$



베이즈 정리

- 곰곰히 생각하면 우도(likelihood)가 무엇인지 알 수 있다(Why?).

$$P(C|A_K) = 1$$
$$P(C|A_{DK}) = \frac{1}{4}$$

- 그러므로 문제가 찾는 답은 다음과 같다.

$$P(A_K|C) = \frac{1}{1 + \frac{1}{4}} = \frac{4}{5}$$

- 문제를 맞췄다고 하여 반드시 답을 알고 있었던 것은 아니다. 베이즈 정리는 그 확률마저 예측한다!



베이즈 정리

베이즈 정리에는 독특한 역사가 있다.

- 장로교 목사이자 아마추어 수학자였던 Thomas Bayes (1701?-1761)는 David Hume의 회의주의적 경험론에 반발하여 (아마도) 신의 존재를 수학으로 증명하는데 관심이 있었다.
- 그의 논문은 한때 완전히 잊혀졌다가 Pierre-Simon Laplace (1749-1827)에 의해 재발견되었다.
- 그러나 20세기 들어 통계학이 급성장하면서 Karl Pearson, Ronald Fisher, Jerzy Neyman 등이 베이즈의 관점과 상이한 통계추론의 논리를 제시하면서 다시 한 번 잊혀졌다.



Bayes



Laplace



Pearson



Fisher



Neyman

베이즈 정리

- 21세기 들어 인공지능(AI)과 기계학습(machine learning)의 시대가 도래하면서 다시 한 번 베이즈 통계학(Bayesian statistics)이 부상하게 되었다.
- 20세기 통계학의 관점을 빈도주의 통계학(frequentist statistics)이라고도 부른다.
- 우리 수업은 (아쉽게도 베이즈 통계학이 아니라) 빈도주의 통계학의 기초를 배우는 과정이다.
- 뇌의 작동 방식이 베이즈 정리에 의해 잘 설명된다는 주장이 있지만, 아쉽게도 베이즈 통계학은 계산집약적(computationally intensive)인 점이 흠이다.



베イズ 정리

- 확률의 기초는 베イズ 정리를 알기 전과 이것을 안 뒤로 나뉜다.
- 몇 가지 흥미로운 논리적 문제를 풀 수 있다: 몬티 홀 문제(the Monty Hall Problem), 다운증후군 임신 테스트 신뢰성 문제 등등...
- 그 밖에도 몇 가지 실생활에서 할 수 있는게 있다: 사람의 말과 글을 알아듣는 기계나 자율주행하는 자동차를 만든다던가 등등...
- (R이나 Python 등을 통해) 20-30줄 정도의 코드로도 꽤 쓸모있는 인공지능을 구현할 수 있다.
- 3학년이나 4학년 때 소셜데이터사이언스 수업에서 살짝 다룬다.

