제4장 확률

1. 서론
2. 사건의 확률
3. 확률을 배정하는 방법
4. 사상관계와 2개의 확률법칙
5. 조건부 확률과 독립
6. 베이즈 정리

1. 서론

- 통계적 조사에서는 모집단의 일부만 관측하고 이를 바탕으로 모집단 전체에 대한 결론을 귀납적으로 이끌어 낸다.

- 일반화의 근거 이해 ⇒ 확률의 이론 필요

- 확률은 가능성을 수치로 나타내는 수단

- 사건의 확률 :그 사건이 일어날 가능성을 수치로 나타낸 것.
  - 불확실성을 수치로 나타내는 것은 매우 유용하며 이 수치는 통계적 추론에서 중요한 역할을 한다.

2. 사건의 확률

- 실험(experiment)
  - 결과마다 변이를 가지는 현상을 관측하는 과정

- 시행(trial)
  - 실험을 실행하는 것
    - Ex. 주사위 던지기, 품질조사, 의견조사 등

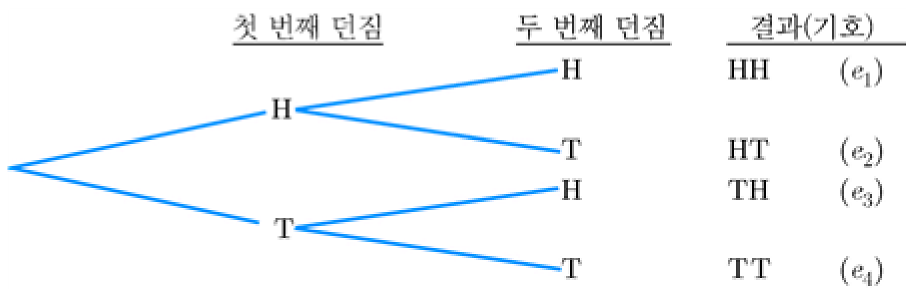
- 표본공간(sample space : S)
  - 실험의 서로 다른 모든 가능한 결과들의 모임

- 기본결과(elementary outcome), 원소(element)
  - 표본공간에 속하는 각각의 결과
  - $S = \{e_1, e_2, e_3, e_4 \cdots\}$

- 단순사건(simple event)또는 표본사건(event)
  - 특정한 성질을 지닌 기본 결과들의 모임
  - $A,\ B,\ C, \cdots$
  - 실험에서 사건A에 속하는 기본 결과 중 어느 하나라도 발생하면 사건 A가 발생했다고 한다.

2. 사건의 확률

- **예제1** 동전 던지기에 대한 나뭇가지 그림과 사건
  - 동전을 두 번 던져 앞면(H)이 나왔는지 뒷면(T)이 나왔는지를 기록한다. 앞면이 정확히 한 번 나올 사건을 A라 하고, 앞면이 한 번도 나오지 않을 사건을 B라 하자. 표본공간과 사건 A와 B의 구성을 나열하라.



- **예제2** 표본공간과 셀 수 있는 사건
  - 토요일 오후에 계산하는 135명의 고객을 관찰하여, 카드 (신용카드 또는 직불카드)로 계산한 고객의 수를 기록한다.
    - (a) 표본공간
    - (b) 구매자의 50% 이상이 카드로 계산할 사건을 표시

2. 사건의 확률

- 유한 표본공간 (finite sample space),
  - 예제1, 예제2

- 가산무한표본공간(countably infinite sample space),
  이산표본공간(discrete sample space)
  - 나열이 가능한 무한 표본공간.
    - 예) 로또에 당첨될 때 까지 로또를 구입하는 횟수

- 연속 표본공간(continuous sample space)
  - 표본 공간의 원소를 하나씩 셀 수 없는 형태로 일정한 범주의 모든 값을 취할 수 있는 형태의 표본공간.
    - 예) 휘발유를 가득 채운 자동차가 기름이 없어질 때까지 주행한 거리 기록
    - $S = \{d | d \geq 0\}$ - 0보다 크거나 같은 실수의 모임.

2. 사건의 확률

- 확률(probability)이란?
  - “창업한 벤처기업이 성공할 것인가?”
  - “내일 비가 올 것인가?”
  - “새로 산 복권이 당첨될 것인가?”
  - “나는 80세까지 살 수 있는가?”등
  - 불확실한 질문에 가능성의 정도를 확실한 수치로 측정하여 표현하는 이론
  - 어떤 사건의 발생가능성의 정도(degree of possibility)또는 믿음의 정도(degree of belief)를 0과 1사이의 수치로 측정하는 것.

2. 사건의 확률

- 사건의 확률(probability of an event)
  - 동일한 조건에서 실험을 여러 번 반복할 때 사건이 발생할 것이라고 기대되는 횟수의 비율을 나타내는 수치.
  - 그 사건에 포함된 모든 기본 결과에 할당된 확률의 합
  - P(A) : 사상 A의 확률.

- 확률의 성질

1. 모든 사건 A에 대하여 $0 \le P(A) \le 1$

2. $P(A) = \sum_{A\text{의 모든 원소 } e} P(e)$

3. $P(S) = \sum_{S\text{의 모든 원소 } e} P(e) = 1$

3. 확률을 배정하는 방법

3.1 일어날 가능성이 동일한 기본결과들-균일확률모형

- 균일확률모형(uniform probability model)
  - 모든 기본 결과가 나올 가능성이 같도록 모형화 하는 것.
  - K개의 상호배타적인(mutually exclusive)원소로 표본공간 $S = \{e_1, e_2, e_3, \cdots, e_k\}$가 구성되어 있을 때

m 개의 기본결과로 구성되어 있는 사건 A 의 확률은

$$P(A) = \frac{m}{k} = \frac{A\text{에 속하는 기본결과의 개수}}{S\text{에 속하는 기본결과의 개수}}$$

3.2 오랜 시행에서 상대도수인 확률

- 실험결과가 동일한 가능성을 가지면 균일 확률 모형 이용.
- 많은 경우 균일 확률 모형을 구성하는 것이 가능하지 않다.
- 사상의 확률을 구하는 방법 : 실험을 여러 번 반복하여 그 사상이 일어나는 비율을 관측하는 것.

① $r_N(A)$: N번 시행에서 사상 A의 상대도수

⇨ $r_N(A) = \dfrac{N \text{ 시행에서 } A\text{가 일어난 회수}}{N}$

② $P(A)$: N이 증가할 때 $r_N(A)$가 안정되는 값

⇨ $P(A) = \lim\limits_{N \to \infty} r_N(A)$

-> 오랜 시행에서의 상대도수의 안정성

오랜 시행에서의 상대도수인 확률
사건 A의 확률인 $P(A)$는 시행횟수가 증가됨에 따라 상대도수가 안정화 되는 값으로 정의한다. 비록 정확히 $P(A)$를 알 수 없다고 하더라도 실험을 많이 반복하여 정확하게 추정할 수 있다.

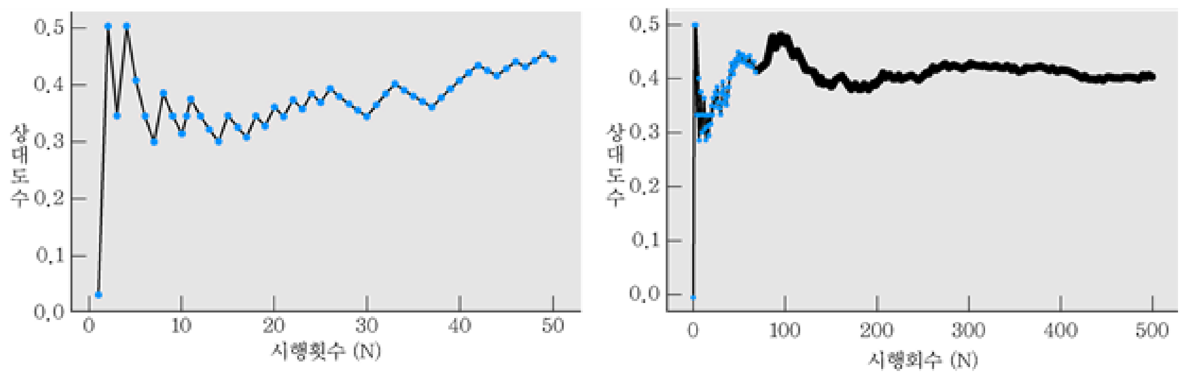
Ex. 보험회사의 보험 해약율, 불량품의 비율 등

3.2 오랜 시행에서 상대도수인 확률

예) N번 시행에 대한 사건A의 상대도수

그림 1

상대도수의 안정성

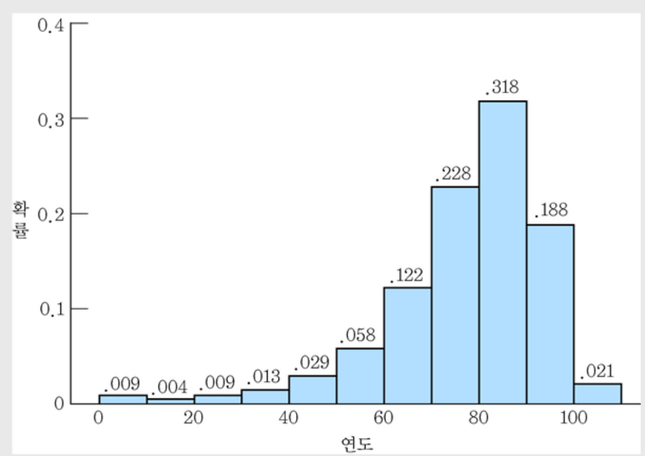


(a) 시행횟수와 상대도수, 처음 50회 (b) 시행횟수와 상대도수, 처음 500회

예) 아기가 태어날 요일의 확률을 정하는 문제

예) 아기는 얼마나 오래사는가?

아기는 얼마나 오래 사는가?



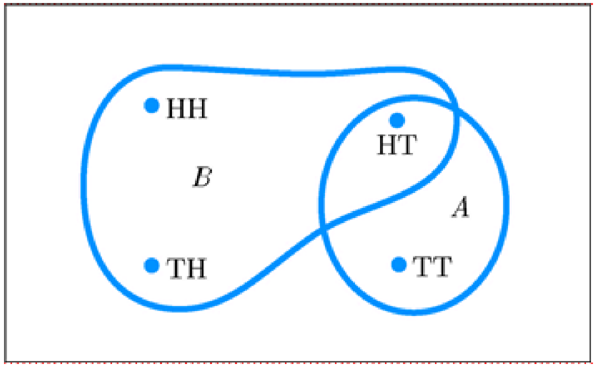
미국에서 태어난 아기의 수명 확률

(출처: National Vital Statistics Reports 54 [2006])

4. 사건 관계와 확률의 법칙

- 사건의 관계
  - 여사건(complement)
  - 합사건(union)
  - 교사건(intersection)
    - 이용한 몇 가지 확률 법칙 유도
  - 벤다이어그램 - 이해용이

- **예제5** 동전 던지기의 벤다이어그램
  - 동전을 두 번 던지는 실험에 대한 벤다이어그램을 그리고 다음 사건들을 표시하라.
    - A : 두 번째 시행에서 뒷면
      B : 적어도 한 개가 앞면



4. 사건 관계와 확률의 법칙

- 사건의 기본 연산

1. 여사건
   - 사건 A에 속하지 않는 모든 기본 결과들의 집합
   - $\bar{A}$: 사건 A의 여사건

2. 합사건
   - A또는 B에 속하거나, A와 B에 모두 속하는 모든 기본 결과들의 집합.
   - $A \cup B$: A와 B의 합사건

3. 교사건
   - A와 B에 동시에 속하는 모든 기본 결과들의 집합
   - AB : A와 B의 교사건

4. 배반사건
   - 두 사건 A와 B의 교사건 AB가 공집합 일 때, 두 사건 A와 B를 상호배반 또는 배반이라 한다.

4. 사건 관계와 확률의 법칙

- 확률법칙
  - 사건 A와 B가 표본 공간 S내에 정의된 임의의 두 사상.

1. 여사건의 법칙
   - $P(A) = 1 - P(\bar{A})$

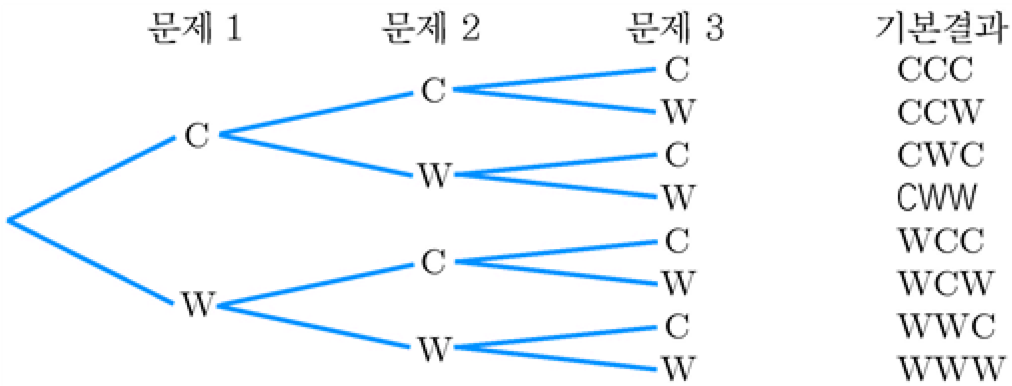
2. 덧셈법칙
   - $P(A \cup B) = P(A) + P(B) - P(AB)$

3. 배반사건 덧셈법칙
   - $P(A \cup B) = P(A) + P(B)$

4. 사건 관계와 확률의 법칙

- **예제 8** 여사건의 법칙을 이용한 확률계산
  - 어린이가 세 문제로 된 단어 맞추기 시험을 친다고 하자.
  - 각 문제마다 두 가지 답 – 맞는 답(C), 틀린 답(W)
  - 어린이가 단어의 뜻을 이해하지 못해 추측으로 답한다면,

    A = [적어도 한 문제를 맞힐 사건] 확률은 얼마인가?



5. 조건부 확률과 독립

- 사상 A의 확률은 관련된 사상 B가 일어났는지의 여부에 따라 정보가 얻어진 후에 조정되어야 하는 경우가 흔히 있다.

- B가 주어졌을 때 A의 조정된 확률
  - P(A|B) : B가 주어졌을 때 A의 조건부 확률

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

5. 조건부 확률과 독립

- **예제12** 체중이 주어졌을 경우 대체의학 사용의 조건부 확률
  - CAM : 보완대체의학
  - A = [보완 대체의학의 사용]
  - B = [과체중]

	저체중	정상체중	과체중	비만	합계
CAM 사용	0.01	0.13	0.12	0.12	0.38
CAM미사용	0.02	0.19	0.21	0.20	0.62
합계	0.03	0.32	0.33	0.32	1.00

a. 임의로 선택된 사람이 과거 보완 대체의학을 사용했을 확률
b. 임의로 뽑힌 사람이 과체중이었다. 이 과거 보완 대체의학을 사용했을 확률

5. 조건부 확률과 독립

1) 확률의 곱셈법칙

$$P(A|B) = P(AB)/P(B)$$
$$\Leftrightarrow P(AB) = P(A|B)P(B)$$

$$P(B|A) = P(AB)/P(A)$$
$$\Leftrightarrow P(AB) = P(B|A)P(A)$$

2) 독립

$$P(A|B) = P(A)$$

$\longrightarrow$ A 와 B는 서로 독립이다. 이와 동치조건은

$$P(B|A) = P(B)$$

또는

$$P(AB) = P(A)P(B)$$ 이다.

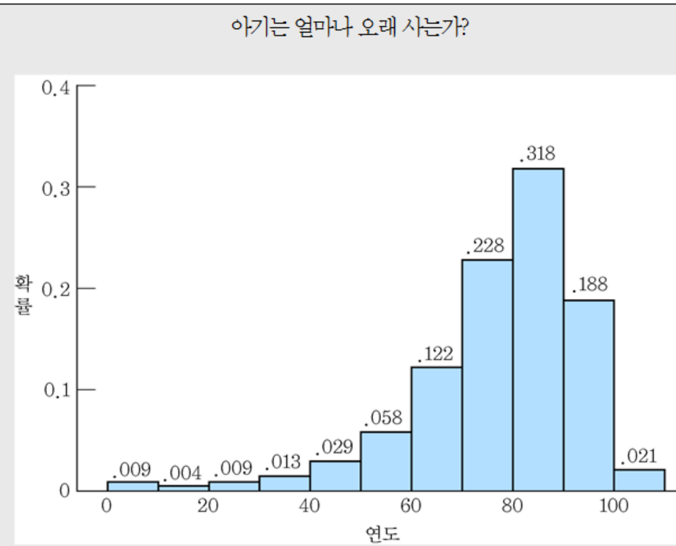
이유)

5. 조건부 확률과 독립

- **예제13** 생존의 조건부 확률(P113)
  - 4.3절 "아기는 얼마나 오래 사는가?"는 10살 단위로 사망확률
    - A = [새로 태어난 아기가 90세 이상 생존]
    - B = [80세 이상 생존]

a. 새로 태어나는 아기가 90세 이상 생존할 확률은 얼마인가?

b. 80세가 된 사람이 90세 이상 생존할 확률은 얼마인가?



미국에서 태어난 아기의 수명 확률
(출처: National Vital Statistics Reports 54 [2006])

5. 조건부 확률과 독립

- **예제16** 확률 배정을 위한 독립이용
  - 기술자들은 어떤 기계가 고장 나지 않을 확률을 다른 용어로 "신뢰도" 라고 한다. 어떤 기계가 독립적으로 작동하는 두 가지 부품으로 구성되어 있다고 하자. 오랜 기간의 검사로부터 부품 1의 신뢰도는 0.98이고 부품 2의 신뢰도는 0.95임이 알려져 있다고 하자. 만일 두 가지 부품이 모두 작동할 때만 시스템이 작동된다고 하면, 이 시스템의 신뢰도는 얼마인가?
  - A_1 : 부품1이 작동
  - A_2 : 부품2가 작동
  - S : 시스템이 작동

6. 베이즈 정리

- **총 확률의 법칙**

$$P(A) = P(AB) + P(A\bar{B})$$
$$= P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

- **베이즈의 정리**

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

$\overline{B}$의 사후확률은 $P(\overline{B}|A) = 1 - P(B|A)$이다.

6. 베이즈 정리

- **예제 19** 총 확률의 법칙과 의료검사의 불확실성
  - 사건 A를 심각한 바이러스 검사에서 어떤 사람이 양성반응을 나타낼 사건이라 하고, 사건 B를 실제로 그 사람이 바이러스에 감염될 사건이라 하자.
    - A=[어떤 사람이 양성반응을 나타낼 사건]
    - B=[실제로 그 사람이 바이러스에 감염될 사건]
  - $P(B) = 0.014$
  - $P(A|B) = 0.995$
  - $P(A|\bar{B}) = 0.01$
  - 어떤 사람이 검사에서 양성반응을 보일 확률 $P(A)$를 구하라.

6. 베이즈 정리

- **예제 20** 베이즈 정리와 의료검사의 불확실성
  - 사건 A를 심각한 바이러스 검사에서 어떤 사람이 양성반응을 나타낼 사건이라 하고, 사건 B를 실제로 그 사람이 바이러스에 감염될 사건이라 하자.
    - A=[어떤 사람이 양성반응을 나타낼 사건]
    - B=[실제로 그 사람이 바이러스에 감염될 사건]
  - 어떤 사람이 양성반응이 나타났다고 하자. 베이즈 정리를 이용하여 그 사람이 바이러스를 가지고 있을 확률을 갱신하라. 즉, 사후확률 $P(B|A)$를 구하라.
  -

7. 유한 모집단에서 확률표본 추출

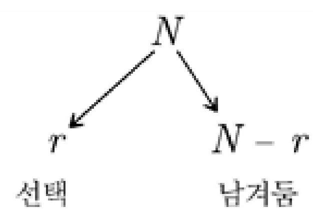
- **조합의 법칙**

기호: N개의 서로 다른 원소로 구성된 그룹에서 r개의 원소를 선택하는 방법의 수는 기호 $\binom{N}{r}$로 나타내며 "N 선택 r"로 읽는다.

공식:

$$\binom{N}{r} = \frac{N \times (N-1) \times \cdots \times (N-r+1)}{r \times (r-1) \times \cdots \times 2 \times 1}$$

- $\binom{N}{r} = \binom{N}{N-r}$



$$\therefore \binom{N}{N} = 1, \binom{N}{0} = 1$$

- 확률표본 추출의 개념

만일 크기 N인 모집단에서 크기 n인 표본을 각각 선택될 확률이 모두 $1 / \binom{N}{n}$이 되도록 선택할 때, 그 표본을 확률표본(random sample)이라 한다.