2016/01/06

[2016-동계세미나-확률 기초]

Chapter 2 확률(probability)

이 부 형(boohyung@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

Contents

- 표본공간
- 사상
- 경우의 수
- 사상의 확률
- 가법정리
- 조건부 확률, 독립사상, 승법정리
- 베이즈 정리

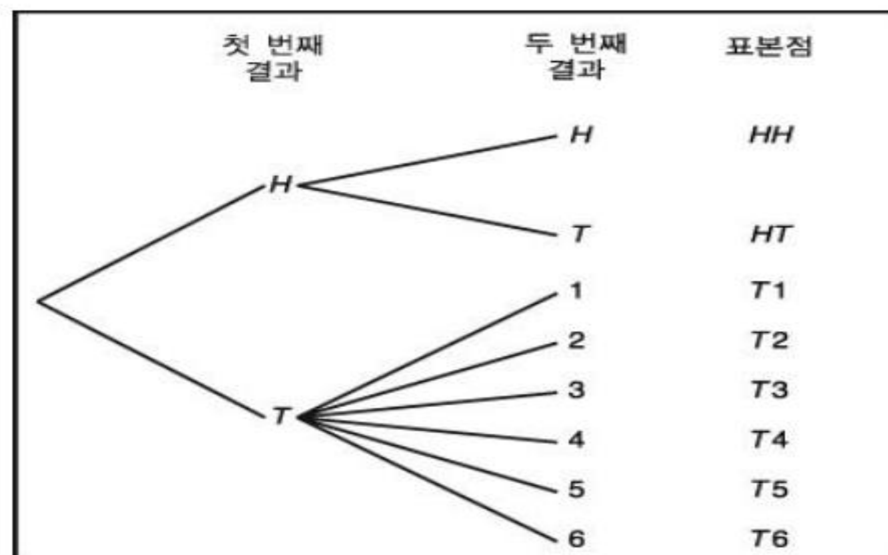
표본공간

- 표본공간(sample space): 통계적 실험에서 발생가능한 모든 결과들의 집합; S로 표시
  - 각 결과: 원소 or 표본점
  - 표현공간의 표현: 수형도, 서술식, 수식 등
    - 서술식이나 수식은 표본점의 수가 아주 크거나 무한인 경우 사용
  - 예)

$$S = \{\text{앞}, \text{뒤}\}$$
$$S = \{(1,1),(1,2),\ldots,(6,5),(6,6)\},$$
$$S = \{x \mid x\text{는 인구 백만이 넘는 도시}\},$$
$$S = \{(x,y) | x^2 + y^2 \leq 4\}$$

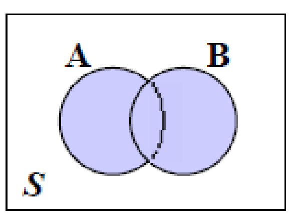
표본공간

- 표본공간(sample space): 통계적 실험에서 발생가능한 모든 결과들의 집합; S로 표시
  - 수형도
    - 예) 한 개의 동전을 던져서 앞면이 나오면 다시 동전을 던지고, 뒷면이 나오면 한 개의 주사위를 던지는 실험을 생각해 보자.
    - 표본공간의 원소들을 수형도로 나타내면

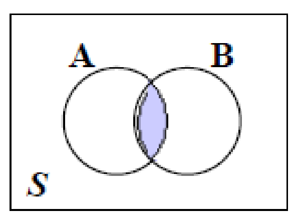


사상

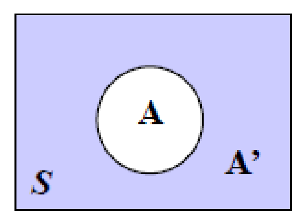
- 사상 or 사건(event): 표본공간의 부분집합
  - 표본공간 S에 대한 하나의 사상 A의 여집합: A의 원소가 아닌 S의 모든 원소들; A'로 표시
  - 두 사상 A와 B의 교집합: A와 B에 공통으로 속하는 모든 원소들을 포함하는 사상; $A \cap B$로 표시
    - $A \cap B = \emptyset$이라면 두 사상 A와 B는 상호 배반(mutually exclusive or disjoint)
  - 두 사상 A와 B의 합집합: A혹은 B에 속하는 모든 원소들을 포함하는 사상; $A \cup B$로 표시
  - 사상과 표본공간과의 표현: 벤 다이어그램



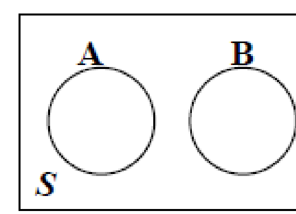
합집합(A∪B)



교집합(A∩B)



여집합 A^C



공집합(배반사건)

사상

- 사상 or 사건(event): 표본공간의 부분집합
  - 예) 한 개의 주사위를 던지는 시행해서 홀수의 눈이 나오는 사건을 A, 3 이상의 눈이 나오는 사건을 B라 하자. 이 때, 다음을 구하여라.
    (1) A와 B의 합집합(합사건)   (2) A와 B의 교집합(곱사건)
    (3) A의 여집합(여사건)

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법
  - 곱의 법칙: 만약 첫 번째 시행이 n_1가지 방법, 그리고 두 번째 시행이 n_2가지 방법이 있으면, 두 가지 시행이 이루어지는 것은 $n_1 n_2$가지 방법이 있음
    - 만약 k가지 시행이 이루어진다면 이 시행에 관련된 경우의 수는 $n_1 n_2 \ldots n_k$

    - 예) 샘이 컴퓨터를 조립하려고 할 때, CPU는 2개 회사, 하드디스크는 4개 회사, 메모리는 3개 회사로부터 주문할 수 있다면 샘이 부품을 주문할 수 있는 총 가짓 수는?

      풀이) $n_1 = 2,\ n_2 = 4,\ n_3 = 3$이므로 주문가능한 가짓 수는

      $$\boldsymbol{n_1\ X\ n_2\ X\ n_3 = 24}$$

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법 (순열과 조합은 상호 배반이 기본이 되어야 성립)
  - 순열(permutation): 어떤 대상물집합의 전체 또는 일부의 **순서적** 배열
    - n개의 서로 다른 대상물에 대한 순열의 수는 $n!$

$$n! = n(n-1)(n-2)\ldots 2 \, X \, 1$$

    - n개의 서로 다른 대상물 중 r개를 취하여 만든 순열의 수

$${}_{n}P_{r} = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

$$\Rightarrow \quad {}_{n}P_{n} = n!$$

    - n개의 서로 다른 대상물을 원형으로 배열하는 순열의 수는

$$(n-1)!$$

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법
  - 순열(permutation): 어떤 대상물집합의 전체 또는 일부의 **순서적** 배열
    - 예) 할아버지, 할머니, 아버지, 어머니, 두 딸로 구성된 6명의 가족이 박물관에 입장하려고 한다. 한 줄로 입장한다고 할 때, 두 딸이 연이어 입장하지 않는 방법의 수를 구하여라.

      풀이) 다른 사람들을 먼저 세우고 양 끝과 사이사이 중 2자리를 택하여 딸 둘을 세우면 된다.
      딸 둘을 제외한 나머지 4명의 가족을 일렬로 세우는 방법의 수는 4!, 딸 둘을 일렬로 세우는 방법의 수는 ${}_5P_2$

$$4! \text{ X } {}_5P_2 = 24 \text{ X } 20$$

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법
  - 순열(permutation): 어떤 대상물집합의 전체 또는 일부의 **순서적** 배열
    - n_1개는 첫 번째 종류, n_2개는 두 번째 종류, … , n_k개는 k번째 종류로 된 n개의 대상물의 순열의 수는

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

    - 분할(partition): n개의 서로 다른 대상물을 r개의 부분집합으로 나누는 것
      - 모든 부분집합의 합집합이 원래의 집합이 되는 경우
      - 경우의 수는 위의 식으로 구할 수 있음

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법
  - 순열(permutation): 어떤 대상물집합의 전체 또는 일부의 **순서적** 배열
    - 예) 7명의 대학원생을 3인용 객실 하나와 2인용 객실 둘에 투숙시키는 방법은 몇 가지인가?

      풀이) $\dfrac{7!}{3!2!2!} = 210$

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법
  - 순열(permutation): 어떤 대상물집합의 전체 또는 일부의 **순서적** 배열
    - 컴퓨터 공학 측면에서의 활용 예
      - 암호학: DES, 3-DES에서 각 라운드를 진행할 때 비트열의 순서를 바꾸는 과정이 존재함(비트열이 변화하는 경우의 수를 구할 때 순열을 사용)
      - 네트워크: 총 n개의 패킷 중 k개를 보낼 때 도착지에 k개의 패킷이 도착하는 순서의 수(이를 통해 패킷의 드랍율 등을 구할 때 사용)

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법
  - 조합(combination): 어떤 대상물집합의 전체 또는 일부의 **비순서적** 배열
    - 서로 다른 n개의 대상물 중에서 r개를 택하는 조합의 수는

$$\binom{n}{r} = \frac{{}_nP_r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

경우의 수

- 실험이 수행되었을 경우 총 시행횟수를 구하는 방법
  - 조합(combination): 어떤 대상물집합의 전체 또는 일부의 **비순서적** 배열
    - 예) 12명의 축구 선수 중에서 6명을 뽑아 선발 선수로 출전시킬 때, 특정한 두 선수를 포함하여 뽑는 경우의 수는?

      풀이) 두 선수가 이미 정해졌다고 생각하면, 10명 중 4명을 뽑는 경우의 수를 구한다.

$$ {}_{10}C_4 = \frac{10!}{4!6!} $$

    - 컴퓨터 공학 측면에서의 활용 예: 순열을 사용하는 경우 중 순서적인 배열에 의미를 두지 않을 경우 사용

사상의 확률

- 확률(probability) or 가중치(weight)의 정의: 사상이 발생할 가능성을 나타내는 0과 1 사이의 수
  - 표본 공간 S의 확률은 1

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

  - 상호배반(mutually exclusive)인 사상의 합의 확률은 각 사상의 확률을 합한 것과 동일

$$P(E_1 + E_2 + E_3) = P(E_1) + P(E_2) + P(E_3)$$

  - 어떤 실험에서 N개의 서로 다른 결과가 동일한 확률로 발생할 수 있고, 이 중 정확히 n개의 원소를 가지는 사상 A가 있으면, 사상 A의 확률은 $P(A) = \frac{n}{N}$

  예) 표본공간 S를 N개의 부분집합으로 분할한 경우

가법정리

- 가법정리(덧셈정리): 확률의 계산을 간단하게 해주는 법칙 중 하나; 합집합에 응용됨

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

- 만약 $P(A \cap B) = \emptyset$라면(상호배반) $P(A \cup B) = P(A) + P(B)$

- 위의 정리를 3가지 사상 A, B, C에 대하여 확장하면
  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$
  $\underline{P(A \cup B \cup C) = P(A \cup B) \cup P(C)}$<u>라고 생각하면</u>

$n(A \cup B \cup C) = n(A \cup B) + n(C) - n((A \cup B) \cap C)$

$= \{n(A) + n(B) - n(A \cap B)\} + n(C) - n((A \cap C) \cup (B \cap C))$

$= n(A) + n(B) - n(A \cap B) + n(C) - \{n(A \cap C) + n(B \cap C) - n((A \cap C) \cap (B \cap C))$

$= n(A) + n(B) + n(C) - n(A \cap B) - n(B \cap C) - n(C \cap A) + n(A \cap B \cap C)$

가법정리

- 가법정리(덧셈정리): 확률의 계산을 간단하게 해주는 법칙 중 하나; 합집합에 응용됨
  - 여사상의 확률 $P(A)=1-P(A^c)$
  - 여사상의 확률을 이용한 정리

    두 사상 A와 B에 대해

    i) $P(B)=P(A\cap B)+P(A^c\cap B)$

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

가법정리

- 가법정리(덧셈정리): 확률의 계산을 간단하게 해주는 법칙 중 하나; 합집합에 응용됨
  - 예) 1부터 20까지의 자연수가 각각 하나씩 적힌 20개의 구슬 중에서 한 개를 뽑을 때, 3 또는 4의 배수가 적힌 구슬이 나올 확률을 구하여라.

    풀이) 구슬에 적힌 수가 3의 배수인 사건을 A, 구슬에 적힌 수가 4의 배수인 사건을 B라 하면

$$P(A) = \frac{6}{20}$$

$$P(B) = \frac{5}{20}$$

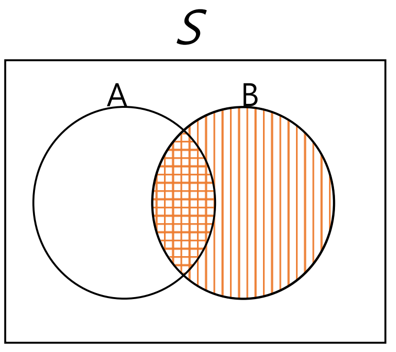
$$P(A \cap B) = \frac{1}{20}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2}$$

조건부 확률, 독립사상, 승법정리

- 조건부 확률(conditional probability): 사상 A가 일어났다고 알려진 조건 하에서 사상 B가 나타날 확률; $P(B\,|\,A)$로 표시
  (두 사상이 서로 상호 배반이 아닐 경우 성립)

  - $P(A\,|\,B) = \dfrac{P(A \cap B)}{P(B)}, \qquad P\left(B\right) > 0$



조건부 확률, 독립사상, 승법정리

- 조건부 확률(conditional probability): 사상 A가 일어났다고 알려진 조건 하에서 사상 B가 나타날 확률; $P(B \mid A)$로 표시
  - 예1) 한 개의 주사위를 던져서 짝수의 눈이 나왔을 때, 그 것이 소수일 확률을 구하여라.

    풀이) 짝수의 눈이 나오는 사건을 A, 소수의 눈이 나오는 사건을 B라고 하면

$$P(A)=\frac{1}{2}, P(B)=\frac{4}{6}, P(A\cap B)=\frac{1}{6}$$

$$P(B \mid A)=\frac{P(A\cap B)}{P(A)}=\frac{6/1}{1/2}=\frac{1}{3}$$

조건부 확률, 독립사상, 승법정리

- 조건부 확률(conditional probability): 사상 A가 일어났다고 알려진 조건 하에서 사상 B가 나타날 확률; $P(B \mid A)$로 표시
  - 예2) 두 사건 A, B에 대하여

$$P(A) = \frac{1}{2}, P(B) = \frac{2}{5}, P(A^c \cap B^c) = \frac{3}{10}$$

일 때, P(B|A)를 구하여라.

조건부 확률, 독립사상, 승법정리

- 독립 or 독립사상(independence): 두 사상 A, B에 대하여 사상 A가 일어나거나 일어나지 않거나 사상 B가 일어날 확률에 영향을 미치지 않는 경우
  이 경우 아래의 두 식이 성립

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

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

- 그렇지 않으면 두 사상 A, B는 종속
  - 두 사상 A, B가 독립이라면 $P(A \cap B) = P(A)\,P(B)$
  - 세 사상 A, B, C가 독립이라면 $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$
  - **독립인지 종속인지 판단하는데 사용!(필요충분조건)**

조건부 확률, 독립사상, 승법정리

- 독립 or 독립사상(independence): $P(A) = P(A \mid B)$ *or* $P(B) = P(B \mid A)$를 만족하는 두 사상 *A, B*
  - 대표적인 예: 동전이나 주사위를 던질 때, 복원추출할 때
  - 독립시행의 확률
    - 독립시행: 매회 시행에서 사건 A가 일어날 확률이 p로 일정하고 각 사건이 서로 영향을 주지 않을 때
    - n회 독립시행
    - 사건 A가 r번 일어날 확률? $P(X=r) = {}_nC_r p^r q^{n-r} \quad (q = 1-p)$
  - 예) 한 개의 동전을 5번 던질 때, 앞면이 적어도 2번 나올 확률은?

조건부 확률, 독립사상, 승법정리

- 승법정리: 조건부 확률 공식 $P(A|B)=\dfrac{P(A\cap B)}{P(B)}, \qquad P(B)>0$ 의 양변에 $P(B)$를 곱하면 두 사상이 함께 일어날 확률을 계산할 수 있음

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

  - 일반적인 승법정리는 종속관계일 때 성립

  - 예1) 양품 5개, 불량품 3개가 있는 상자에서 2개를 고를 때, 둘 다 불량일 확률은?

    풀이) 첫 번째가 불량품일 사상 A의 확률 $P(A)=\dfrac{3}{8}$

    두 번째가 불량품일 사상 B|A의 확률 $P(B|A)=\dfrac{2}{7}$

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

조건부 확률, 독립사상, 승법정리

- 승법정리: 조건부 확률 공식 $P(A|B)=\frac{P(A\cap B)}{P(B)}, \qquad P(B)>0$ 의 양변에 $P(B)$를 곱하면 두 사상이 함께 일어날 확률을 계산할 수 있음

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

- 예2) 앞의 예1에서 3개를 고를 때, 셋 다 불량일 확률은?

풀이) 첫 번째가 불량품일 사상 A의 확률 $P(A)$
두 번째가 불량품일 사상 $B|A$의 확률 $P(B|A)$
세 번째가 불량품일 사상 $C|A\cap B$의 확률 $P(C|A\cap B)$

세 가지 사상이 동시에 일어나야 셋 다 불량이기 때문에

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

조건부 확률, 독립사상, 승법정리

- 배반과 독립, 종속(두 사상 A와 B가 존재할 때)
  - 상호배반: 하나의 사상이 발생하면 다른 사상이 발생할 수 없는 경우
    - $P(A \cap B) = \emptyset$
    - 상호배반인 두 사건은 서로 종속: 상호배반이라면 사건 A가 일어나는 경우 사건 B가 일어날 확률은 0으로 변경됨
  - 상호독립: 하나의 사상이 발생하든 발생하지 않든 다른 사상이 일어날 확률이 변하지 않는 경우
    - $P(A \cap B) = P(A)P(B)$
  - 종속: 하나의 사상이 발생할 때 다른 사상이 일어날 확률에 영향을 미치는 경우
    - $P(A \cap B) = P(B|A)P(B) = P(A|B)P(A)$

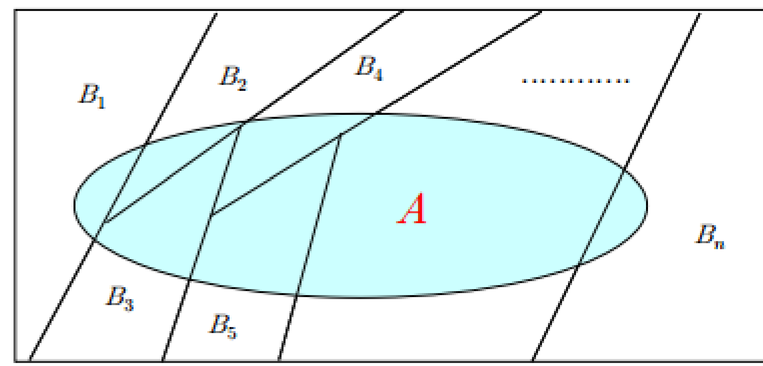
베이즈 정리

- 전확률의 법칙

B와 B^c 는 S 의 분할

$\Rightarrow \quad A \cap B$와 $A \cap B^c$ 는 A의 분할

$$\Rightarrow \quad P(A) = P(A \cap B) + P(A \cap B^c)$$
$$= P(B) \cdot P(A|B) + P(B^c) \cdot P(A|B^c)$$



$E_1, E_2, \cdots, E_k$ 가 표본공간 S 의 분할이면

$$P(A) = \sum_{i=1}^{k} P(E_i) P(A \mid E_i)$$

- 상황에 따라 S를 적절히 $E_1, \dots, E_k$로 분할하여 $P(A|E_i)$를 구함으로써 $P(A)$를 쉽게 구할 수 있음

베이즈 정리

- 전확률의 법칙
  - 예) 3대의 기계 B_1, B_2, B_3가 각각 전체 생산량의 30%, 45%, 25%를 생산하는 어느 조립 공장에서 과거의 경험으로부터 각 기계의 불량품 제조율이 2%, 3%, 2%임이 알려져 있다. 이제 완제품 중에서 임의로 하나를 선택했을 때, 그것이 불량품일 확률은 얼마인가?

    풀이) 사상 A: 그 제품이 불량품일 사상
    사상 B_1: 그 제품이 기계 B_1에서 제조되었을 사상
    사상 B_2: 그 제품이 기계 B_2에서 제조되었을 사상
    사상 B_3: 그 제품이 기계 B_3에서 제조되었을 사상

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

베이즈 정리

- 베이즈 정리
  - 전 페이지의 예에서 선택된 제품이 불량품이라는 사실이 알려졌을 때, 그 제품이 기계 B_1에서 제조되었을 확률은 얼마인가? -> **조건부 확률 $P(B_1|A)$**?

    - 사전확률: 사상 A or B_1의 확률
      $$P(A), P(B_1)$$
    - 사후확률: 사상 B_1가 일어난 조건에서의 사상 A의 확률
      $$P(A|B_1)$$
    - 우도확률: 사상 A가 일어난 조건에서의 사상 B_1의 확률
      $$P(B_1|A)$$

    - <u>사전확률과 사후확률을 이용하여 우도확률을 구할 수 있음</u>

베이즈 정리

- 베이즈 정리
  - 정의

$E_1, E_2, \cdots, E_k$ 가 표본공간 S 의 분할이되, $P\left(E_i\right) > 0$ 이면

$$P(E_j \mid A) = \frac{P(E_j)P(A \mid E_j)}{\sum_{i=1}^{k} P(E_i)P(A \mid E_i)}, \qquad A \subset S$$

(증) $P(E_j \mid A) = \dfrac{P(A \cap E_j)}{P(A)} = \dfrac{P(E_j)P(A \mid E_j)}{\sum_{i=1}^{k} P(E_i)P(A \mid E_i)}$

베이즈 정리

- 베이즈 정리
  - 예1) 몬티 홀 문제(Monty Hall Problem)
    스포츠카는 어디에? 문이 있는 방이 3개 있고 3개의 방 중 하나에는 스포츠카가, 나머지 2 곳에는 염소가 들어있다. 스포츠카가 있을 확률은 1/3, 퀴즈 참가자가 1번 문을 선택했다. 이 때, ‘스포츠카가 어디 있는지 알고 있는’ 사회자 홀(Hall)이 3번 문을 열었고 염소가 모습을 드러냈다. 선택을 바꿀 수 있는 기회가 1번 있다고 할 때, 바꾸는 것이 유리할까? 바꾸지 않는 것이 유리할까?

베이즈 정리

- 베이즈 정리
  - 풀이)
    P(문 1에 자동차) = P(A) = 1/3
    P(사회자가 문 3을 열 확률) = P(B) = 1/2
    P(문 2에 자동차) = P(B) = 1/3

    1) 선택을 바꾸지 않았을 경우 당첨될 확률
    (문 1에 자동차가 있을 확률)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{(1/2)(1/3)}{1/2} = 1/3$$

    2) 선택을 바꿨을 경우 당첨될 확률
    (문 2에 자동차가 있을 확률)

$$P(C|B) = \frac{P(B|C)P(C)}{P(B)} = \frac{(1)(1/3)}{1/2} = 2/3$$

베이즈 정리

- 베이즈 정리
  - 예2) 스팸메일 필터
    수신 메일의 내용에 ‘A’ 라는 문자가 있을 때 이 메일이 스팸 메일일 확률은?

    풀이)
    P(A): 수신 메일의 내용에 ‘A’ 가 포함되어 있는 경우의 확률 = 1/5
    P(S): 수신한 메일이 스팸 메일일 확률 = 1/100
    P(A|S): 수신한 메일이 스팸 메일일 때, 그 메일 안에 ‘A’ 라는 문자가 포함된 경우의 확률 = 1/4

$$P(S \mid A) = \frac{(1/4)(1/100)}{1/5} = \frac{1}{80}$$