컴퓨터 응용통계

4-1. 확률과 확률변수

최경미

확률(Probability)

정의 4.1 실험과 관련된 확률 용어

① 실험단위(Experimental unit)는 사람, 물건, 현상과 같이 우리가 자료(data)를 모으는 대상이다.

② 시행(Trial)은 실험 한 번 실시를 일컫는다.

③ 결과(Outcome)는 시행의 측정값이다.

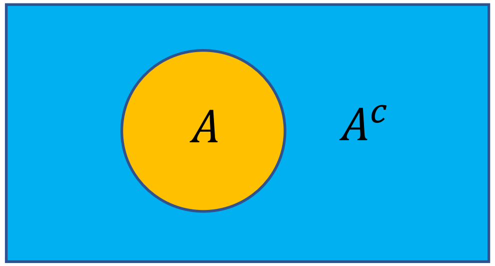
④ 표본공간(Sample space) S는 모든 가능한 시행결과의 집합이다.

⑤ 사건(Event) A는 표본공간 S의 부분집합이며, $A \subseteq S$로 나타낸다.

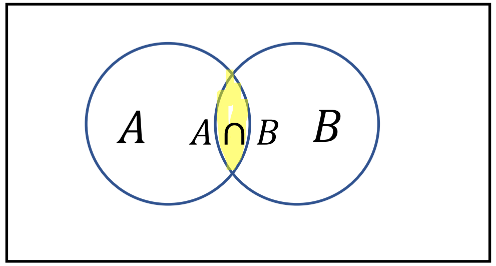
네 가지 사건

정의 4.2

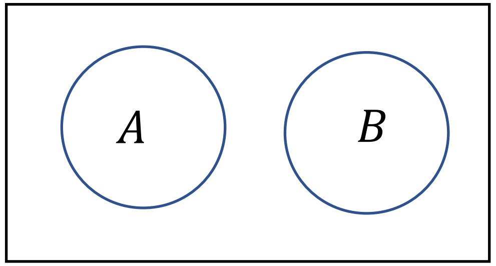
① 여사건 A^c는 A가 발생하지 않는 사건이다.



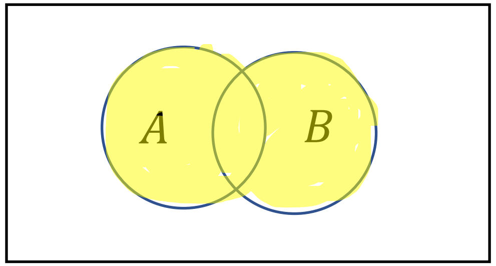
② 교사건, 곱사건 $A \cap B$는 두 사건 A와 B가 동시에 발생하는 사건이다.



③ $A \cap B = \emptyset$이면, A와 B는 배반사건(exclusive event)이다.



④ 합사건 $A \cup B$는 A 또는 B가 발생하는 사건이다.



예제 4.1 동전 한 개를 던지기

- 시행: 동전 한 개 던지기
- 결과: 동전의 앞면(H) 또는 뒷면(T)
- 표본공간 $S = \{H, T\}$
- 앞면이 나올 사건을 A, 뒷면이 나올 사건을 B라고 두자.

  $A = \{H\}$, $B = \{T\}$
- 앞면이 나오지 않는 사건 A^c
- 뒷면이 나오니 않는 사건 B^c
- 동전을 한 번 던져서 앞면과 뒷면이 동시에 나올 수 없기 때문에, $A \cap B = \emptyset$이므로, A와 B는 배반사건이다.
- 앞면 또는 뒷면이 나올 사건은 $A \cup B = S$이므로, 전체 표본공간이다

예제 4.2 공정한 주사위 한 개 던지기

- 시행: 주사위 한 개 던지기
- 결과: 주사위 눈의 수
- 표본공간 $S = \{1,2,3,4,5,6\}$
- 3의 배수인 사건을 A, 최소값이 나타나는 사건을 B라고 두자. $A = \{3,6\}$, $B = \{1\}$
- A가 일어나지 않는 사건 $A^c = \{1,2,4,5\}$
- B가 일어나지 않는 사건 $B^c = \{2,3,4,5,6\}$
- $A \cap B = \emptyset$이므로 A와 B는 동시에 발생할 수 없는 배반사건이다.
- $A \cup B = \{1,3,6\}$

확률의 경험적, 고전적 정의

- 셀 수 있는 유한 표본공간 S에서 사건 A가 발생하는 확률

$$P(A) = \frac{A\text{의 원소의 수}}{S\text{의 원소의 수}}$$

- 예제 주사위 한 개를 던질 때, 3의 배수인 사건을 A라 두자.

S={1,2,3,4,5,6}

A={3,6}

$$P(A) = \frac{A\text{의 원소의 수}}{\text{표본공간 } S\text{의 원소의 수}} = \frac{2}{6} = \frac{1}{3}$$

집합과 공리를 이용한 확률의 정의

확률의 정의 4.3 표본공간 S, 사건 A에 대하여, 다음의 세 가지 공리를 만족하는 P를 확률이라고 정의한다.

공리 1. $0 \le P(A) \le 1$

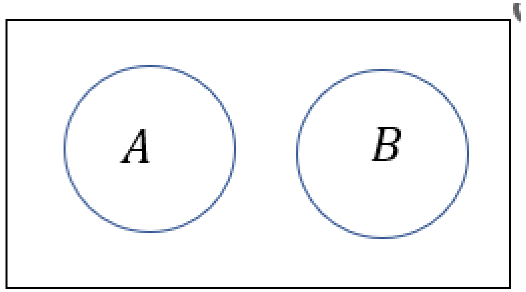
공리 2. $P(S) = 1$

공리3. 사건 $A_1, A_2, A_3 \ldots$에 대하여 $A_i \cap A_j = \emptyset$ $(i \neq j)$ 이면, 다음 식이 성립한다.

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

- 공리 3으로부터 A와 B가 배반사건이면, $A \cap B = \emptyset$이므로,

$P(A \cup B) = P(A) + P(B)$가 성립함을 알 수 있다.

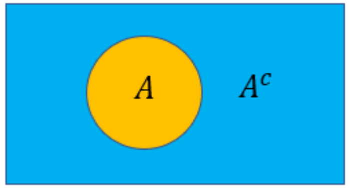


정리 4.1 확률의 성질

표본공간 S와 사건 A, B에 대하여 다음이 성립한다.

① 사건 A가 발생하지 않는 여사건 A^c에 대한 확률은 다음과 같다.

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



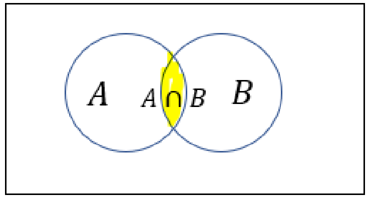
② $P(\emptyset) = 0$

③ 사건 A 또는 사건 B가 발생하는 확률은 다음과 같으며, 이를 합의 법칙이라고 부른다.

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

$P(A) + P(B)$에 두 번 포함된

$P(A \cap B)$를 한 번 뺀다.



$$\begin{aligned} P(S) &= P(A \cup A^c) \\ &= P(A) + P(A^c) \text{ ③} \\ &= 1 \text{ ②} \end{aligned}$$

$$\therefore P(A^c) = 1 - P(A)$$

$$P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$$

$$P(\emptyset) = 0$$

$$\begin{aligned} P(A \cup B) &= P\big(A \cup (B \cap A^c)\big) \\ &= P(A) + P(B \cap A^c) \text{ ③} \\ P(B) &= P(\,(A \cap B) \cup (B \cap A^c)) \\ &= P(A \cap B) + P(B \cap A^c) \text{ ③} \\ P(B \cap A^c) &= P(B) - P(A \cap B) \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

대입

예제 4.3 동전 두 개를 던질 때, 앞면이 나오지 않는 사건을 A, 앞면이 한번만 나오는 사건을 B, 앞면이 적어도 한번 나오는 사건을 C라고 두고, 세 사건의 확률을 구해보자.

표본공간

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

$$P(\{HH\}) = P(\{HT\}) = P(\{TH\}) = P(\{TT\}) = \frac{1}{4}$$

따라서,

$$P(\text{앞면이 나오지 않는 사건}) = P(A) = P(\{TT\}) = \frac{1}{4}$$

$$P(\text{앞면이 한번만 나오는 사건}) = P(B) = P(\{HT, TH\}) = \frac{1}{2}$$

$$P(\text{앞면이 적어도 한번 나오는 사건}) = P(C) = P(\{HH, HT, TH\}) = \frac{3}{4}$$

또는

$$P(\text{앞면이 적어도 한번 나타남}) = \ 1 - P(\text{앞면이 나오지 않음}) = \ 1 - P(\{TT\})$$

$$P(C) = 1 - P(C^c) = 1 - \frac{1}{4} = \frac{3}{4}$$

예제 4.4 주사위 한 개를 던질 때, 홀수가 나올 사건을 A, 짝수가 나올 사건을 B, 3의 배수가 나올 사건을 C라고 두고, 각각의 확률과 $B \cap C$와 $B \cup C$의 확률을 계산해보자.

S={1,2,3,4,5,6}

$$A = \{1,3,5\}, \quad B = \{2,4,6\}, \quad C = \{3,6\}$$

$$P(A) = \tfrac{1}{2}, P(B) = \tfrac{1}{2}, \ P(C) = \tfrac{1}{3}$$

$$B \cap C = \{6\}$$
$$B \cup C = \{2,3,4,6\}$$

$$P(\text{짝수이고}, 3\text{의 배수}) = P(B \cap C) = P(\{6\}) = \frac{1}{6}$$

$$P(\text{짝수이거나}, 3\text{의 배수}) = P(B \cup C) = P(\{2,3,4,6\}) = \frac{2}{3}$$

합의 법칙을 사용하여 다시 계산해보자.

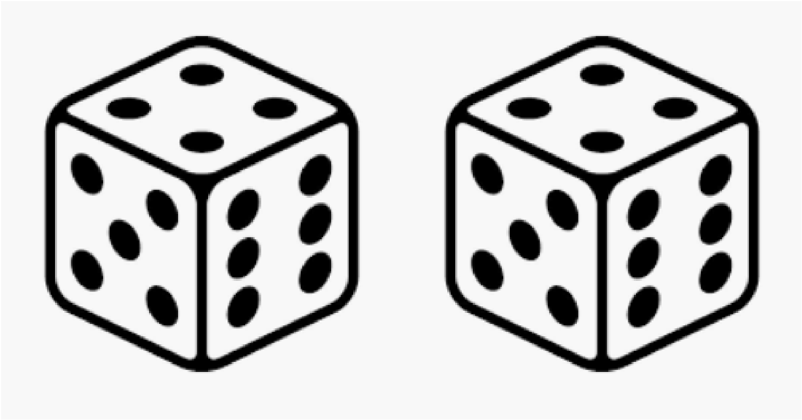
$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}$$

예제 4.5 두 개의 주사위를 동시에 던져서, 두 눈의 합이 10이상인 사건 A의 확률을 계산해보자.

표본공간 S와 사건 A가 다음과 같다.

$$S = \{(x, y); x, y = 1,2,3,4,5,6\}, \quad A = \{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}$$

S의 각 원소의 확률은 $\frac{1}{36}$이므로, $P(A) = \frac{6}{36} = \frac{1}{6}$이다



+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

경우의 수

정리 4.2

① 곱의 법칙. 사건 A가 발생할 경우의 수가 a이고 사건 B가 발생할 경우의 수가 b이면, 발생 가능한 모든 경우의 수는 $a \times b$이다.

② 순열의 법칙. 서로 다른 n개에서 r개를 뽑아서 순서 있게 나열하는 방법의 수는 다음과 같다.

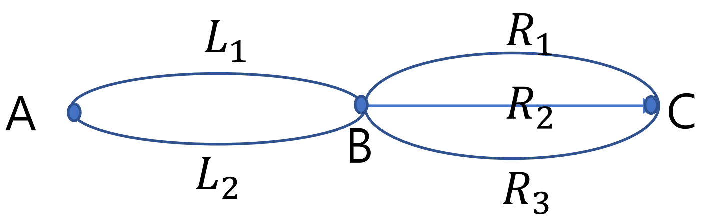
$$nPr = \frac{n!}{(n-r)!}$$

③ 조합의 법칙. 서로 다른 n개에서 순서 없이 r개를 뽑는 방법의 수는 다음과 같다.

$$nCr = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

- $n! = n(n-1)(n-2)\cdots 2 \cdot 1$

①



$2 \times 3 = 6$

$L_1R_1 \quad L_1R_2 \quad L_1R_3$

$L_2R_1 \quad L_2R_2 \quad L_2R_3$

② abc…z 26에서 2개 반복 x 순서 o 나열하기

$$\boxed{26} \times \boxed{25} = \frac{26!}{24!}$$

③ 네 사람 P_1, P_2, P_3, P_4 중에 반복 x 순서 x 둘 뽑기

$P_1\,P_2 = P_2\,P_1 \quad P_2\,P_3 = P_3\,P_2 \quad P_3\,P_4 = P_4\,P_3$

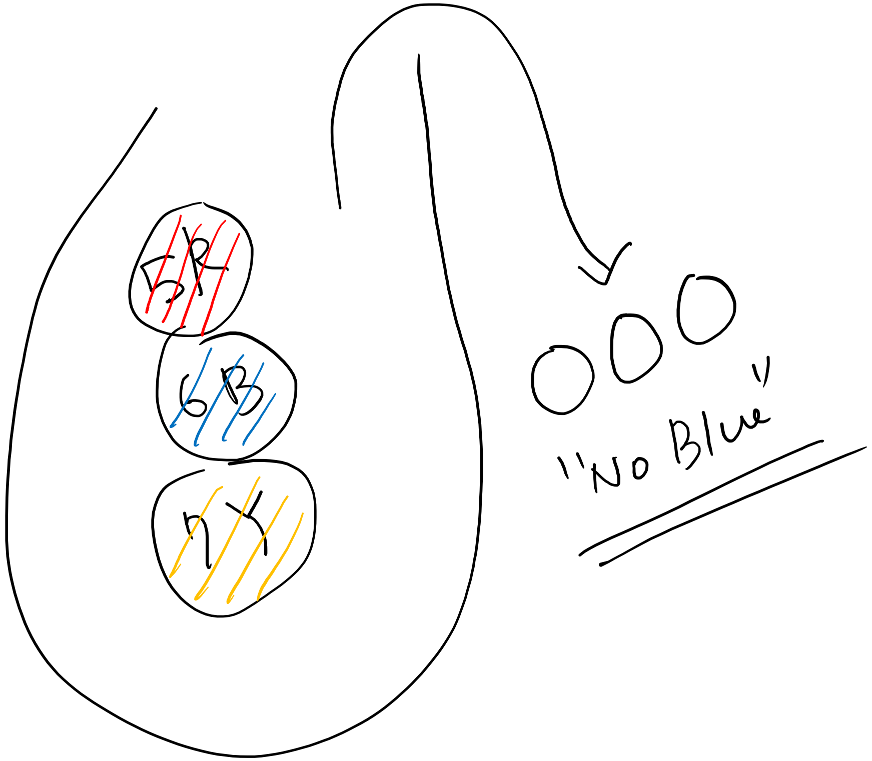
$P_1\,P_3 = P_3\,P_1 \quad P_2\,P_4 = P_4\,P_2$

$P_1\,P_4 = P_4\,P_1$

$$\binom{4}{2} = \frac{4!}{2!2!} = 6$$

조합의 예제

예제 4.6 주머니 안에 빨간 공 5개와 파란 공 6개, 노란 공 7가 들어있다고 가정하자. 주머니에서 3개의 공을 뽑을 때, 파란 공이 없을 확률을 구해보자. 단, 모든 공이 뽑힐 가능성은 동일하다고 가정하자.



풀이 주머니 안에 있는 18개의 공 중에서 3개를 뽑는 경우의 수= $\binom{18}{3}$

파란 공 중에서 하나도 안 뽑는 경우의 수= $\binom{6}{0}$

빨간 공과 노란 공 중에서 3개를 뽑는 경우의 수= $\binom{12}{3}$

곱의 법칙에 따라서 구하는 확률

$P(\text{뽑힌 세 공 중에 파란 공이 없음})$

$$= \frac{\text{빨강 노랑 중 3개 뽑기} \times \text{파랑 중 하나도 안 뽑기}}{\text{전체 18개 공 중 3개를 뽑기}}$$

$$= \frac{\binom{12}{3}\binom{6}{0}}{\binom{18}{3}}$$

$$= \frac{\frac{12!}{3!9!}\ \frac{6!}{0!6!}}{\frac{18!}{3!15!}}$$

$= 0.2696078$

4.2 조건부 확률과 독립사건

- 사건 B가 발생한 사실을 알 때와 모를 때, 사건 A의 발생 확률이 달라질까?
- 예를 들어, 눈을 가리고 주사위 한 개를 던지고, 주사위 눈을 맞춰보자.
- 아무 정보도 없는 상태에서 주사위 눈이 {1}일 확률은 $1/6$이다.
- 옆에서 누군가 주사위 눈이 홀수임을 알려준다면, 주사위 눈이 {1}일 확률은 어떻게 달라질까?
- 옆에서 누군가 주사위 눈이 3의 배수임을 알려준다면, 주사위 눈이 {1}일 확률은 어떻게 달라질까?

예제 4.7 주사위 한 개를 던지는 실험에서 사건 $A = \{1\}$와 사건 $B = \{1,3,5\}$에 대하여, B를 알 때 A의 조건부 확률을 계산해보자.

표본공간 $S = \{1,2,3,4,5,6\}$

목표 $A = \{1\}$, 홀수 $B = \{1,3,5\}$

$$P(A) = \frac{n(A)}{n(S)} = \frac{n(\{1\})}{n(\{1,2,3,4,5,6\})} = \frac{1}{6}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{n(\{1,3,5\})}{n(\{1,2,3,4,5,6\})} = \frac{1}{2}$$

주사위의 눈이 홀수 B 중 하나라는 사실을 알면, 주사위의 눈이 1일 확률은 셋 중 하나이므로 1/3이다.

$$P(A|B) = \frac{n(A \cap B)}{n(B)} = \frac{n(\{1\})}{n(\{1,3,5\})} = \frac{1}{3}$$

분자: {1}이 {1,3,5} 중에 있는지 살핌.

분모: 주사위의 눈이 홀수라는 정보가 주어졌으므로, 표본공간이 $S = \{1,2,3,4,5,6\}$에서 $B = \{1,3,5\}$로 바뀌기 때문에, $n(A \cap B)$을 $n(B)$로 나눈다.

분모와 분자를 $n(S)$로 나누어, 조건부 확률을 구할 수 있다.

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

① B가 발생했다는 조건 하에,

② B가 발생했을 때,

③ B가 발생했다는 사실을 안다면,

A가 발생할 조건부 확률

P(A|B), P of A given B

조건부확률 $P(A|B)$

정의 4.4 두 사건 $A, B \subseteq S$에 대하여, 사건 B가 발생했다는 가정 하에 사건 A가 발생하는 확률로 정의되는 조건부확률 $P(A|B)$는 두 사건 A, B가 동시에 발생하는 확률을 사건 B의 확률로 나누어서 얻을 수 있다. 즉,

$$P(B) > 0\text{일 때, } P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A) > 0\text{일 때, } P(B|A) = \frac{P(A \cap B)}{P(A)}$$

정리 4.3 곱의 규칙. $P(A), P(B) > 0$일 때,

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

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

가 성립한다.

민감도(Sensitivity)와 특이도(specificity)

	양성(+)	음성(−)	합
질병	9 (TP)	1 (FN)	10
정상	2 (FP)	188 (TN)	190
합	11 (P)	189 (N)	200

민감도 (Sensitivity) = 진양성률(TP; True Positive rate)

$$= P(\text{양성}|\text{질병}) = \frac{P(\text{질병} \cap \text{양성})}{P(\text{질병})} = \frac{9/200}{10/200} = 0.900$$

특이도(Specificity) = 진음성률(TN; True Negative rate)

$$= P(\text{음성}|\text{정상}) = \frac{P(\text{정상} \cap \text{음성})}{P(\text{정상})} = \frac{188}{190} = 0.989$$

위음성

$$P(-|\text{질병}) = \frac{1}{10}$$

위양성

$$P(+|\text{정상}) = \frac{2}{190}$$

한국경제TV 2022년 2월 7일
코로나 자가진단키트 KFDA 승인기준
민감도 90%
특이도 99%이상

중앙재난안전대책본부
광주, 전남, 경기도 평택과 안성 1/26~ 3/1
8만 4천건
자가진단키트 신속항원검사 + 687 (0.8%) P
PCR + 523 (76.1%) TP
PCR – 164 (23.9%) FP

독립사건

정의 4.4 두 사건 $A, B \subseteq S$에 대하여, $P(B) > 0$일 때,

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

정의 4.5 $P(A), P(B) > 0$일 때,

$$P(A|B) = P(A) \text{ 또는 } P(B|A) = P(B)$$

이면, 두 사건 A와 B는 독립이라고 정의된다.

사건 B의 발생이 사건 A의 방생에 영향을 미치지 않는다.

사건 A의 발생이 사건 B의 방생에 영향을 미치지 않는다.

정리 4.4 두 사건 A와 B는 독립사건(independent events)일 필요충분 조건은

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

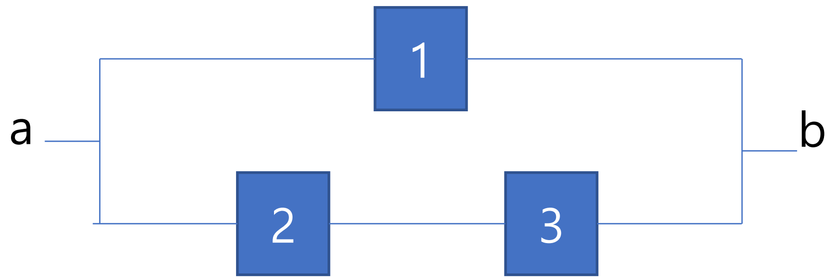
이다. 독립이 아닌 두 사건은 종속사건(dependent events)이다.

증명. 정의 4.4와 4.5로부터 다음의 결과를 얻을 수 있다.

$$P(A|B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)} = P(A) \ (\because \text{독립})$$

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

예제 4.9 어떤 통신시스템에 3개의 부품이 아래 그림4.1과 같이 연결되어 있다. 각 부품의 고장을 R_i 라고 두고, 각 부품의 고장확률은 $P(R_1) = 0.002, P(R_2) = P(R_3) = 0.001$이라고 가정하자. a에서 b로 신호를 보낼 때, 신호가 전달되지 않을 확률은 얼마인가? 부품의 고장은 독립이라고 가정하자.



$P($ 신호전달 안됨$)$

$= P($위로 전달안됨 $\cap$ 아래로 전달안됨$)$

$= P($위로 전달안됨$)P($아래로 전달안됨$)$ (독립)

$= P(R_1)P(R_2 \cup R_3)$

$= P(R_1)(P(R_2) + P(R_3) - P(R_2 \cap R_3))$ (합의 법칙)

$= P(R_1)(P(R_2) + P(R_3) - P(R_2)P(R_3))$ (독립)

$= (0.002)(0.001 + 0.001 - (0.001)^2)$

$= 0.000003998$

A와 B 독립 $\Leftrightarrow P(A \cap B) = P(A)P(B)$

혼자 풀기 (과제)

1. 월드오미터 (https://www.worldometers.info/coronavirus/)에 따라서, 2021년 10월11일 UK와 프랑스의 Covid-19 환자의 사망 또는 생존을 나타낸 자료이다. 두 나라의 Covid-19 자료만을 근거로, 다음 확률 중 바르게 계산된 것을 모두 고르시오. 2, 4

두 나라별 Covid-19사망/생존	Covid-19 사망	Covid-19 생존
프랑스	117082	4468303
UK	137763	4203973

1 $\mathrm{P}(\text{프랑스}\,|\mathrm{Covid}-19\text{ 사망}) = \frac{4468303}{4585385}$

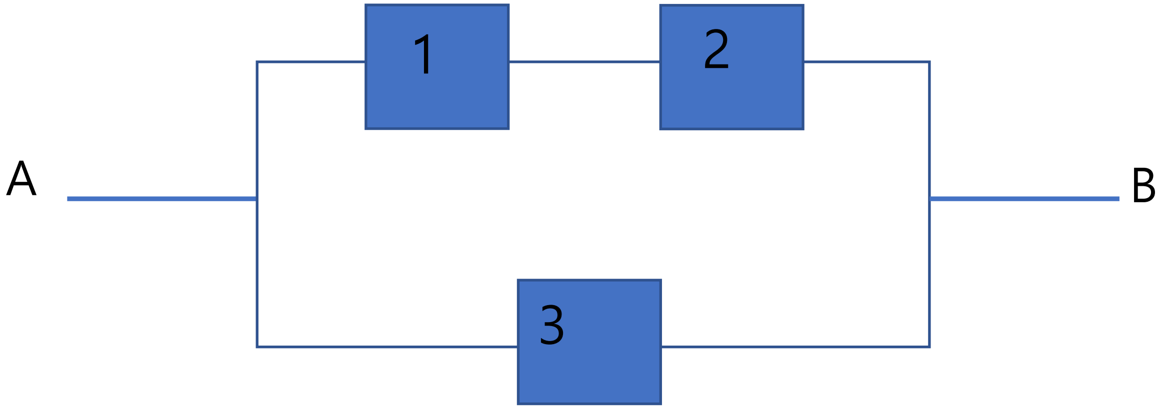
2 $\mathrm{P}(\mathrm{UK}|\mathrm{Covid}-19\text{ 생존}) = \frac{4203973}{8672276}$

3 $\mathrm{P}(\mathrm{Covid}-19\text{ 사망 }|\mathrm{UK}) = \frac{4203973}{8672276}$

4 $\mathrm{P}(\mathrm{Covid}-19\text{ 생존}|\text{프랑스}) = \frac{4468303}{4585385}$

2. 세 개의 중계기 1,2,3의 고장을 $R1, R2, R3$라 두면, 각 중계기의 고장률이 $P(R1) = P(R2) = 0.02, P(R3) = 0.01$이고, 각 중계기의 고장은 독립이다. A에서 B로 신호가 전달되지 않을 확률은 얼마인가? 1

10.000396 2 0.000639 3 0.000693 4 0.000963
5 위 보기 중 답 없음



4.3 베이즈 공식

- 사전확률 (prior probability): 이미 알려진 사건의 확률 $P(A)$또는 $P(B)$
- 사후확률 (posterior probability) : $P(A|B)$ 또는 $P(B|A)$ 등의 조건부 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(A \cap B) = P(B)P(A|B), P(B) > 0$$

A와 B는 독립 $\Leftrightarrow$ $P(A|B) = P(A)$ 또는 $P(B|A) = P(B)$

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

예제 4.10 이진신호채널 (1) 0 보내고 0 받기 (2) 0 보내고 1 받기 (3) 1 보내고 0 받기 (4) 1 보내고 1 받기. 기호로 S0은 0 보내기, S1은 1 보내기, R0은 0받기, R1은 1 받기라고 두자. 보낸 신호의 약 30%가 0이고, 약 70%가 1이라고 가정하자. 또한 0을 보낼 때 0을 받을 확률과 1을 보낼 때 1을 받을 확률이 99%라고 두면, 0을 보낼 때 1로 잘못 받을 확률과 1을 보낼 때 0으로 잘못 받을 확률이 1%이다. 그러면, 받은 신호가 0일 때, 이 신호가 원래 0이었을 확률은 얼마일까?

$P(S0) = 0.3$

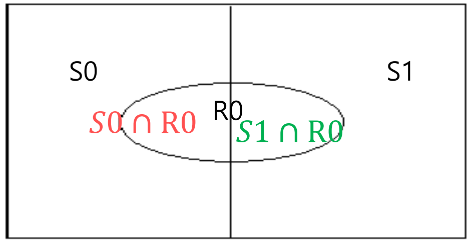
$P(S1) = 0.7$

$P(R0|S0) = P(R1|S1) = 0.99$

$P(R1|S0) = P(R0|S1) = 0.01$

$P(R0) = P(S0 \cap R0) + P(S1 \cap R0)$ (확률공리3)

$= P(S0)P(R0|S0) + P(S1)P(R0|S1)$ (배반사건)



$P(S0|R0)$

$= \dfrac{P(S0 \cap R0)}{P(R0)}$

$= \dfrac{P(S0 \cap R0)}{P(S0 \cap R0)+P(S1 \cap R0)}$ (베이즈 공식)

$= \dfrac{P(S0)P(R0|S0)}{P(S0)P(R0|S0)+ P(S1)P(R0|S1)}$

$= \dfrac{(0.3)(0.99)}{(0.3)(0.99) + (0.7)(0.01)} = 0.9769737$

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

A와 B는 독립 $\Leftrightarrow P(A \cap B) = P(A)P(B)$

혼자 풀기 (과제)

- 3. 어떤 채널을 통해서 이진신호를 전달할 때 다음의 4가지 사건이 발생할 수 있다: (1) 0 보내고 0 받기 (2) 0 보내고 1 받기 (3) 1 보내고 0 받기 (4) 1 보내고 1 받기. 기호로 $\mathrm{S0}$은 0 보내기, $\mathrm{S1}$은 1 보내기, $\mathrm{R0}$은 0받기, $\mathrm{R1}$은 1 받기라고 두자. 보낸 신호의 약 60%가 0이고, 약 40%가 1이라고 가정하자. 또한 0을 보낼 때 0을 받을 확률과 1을 보낼 때 1을 받을 확률이 98%라고 두면, 0을 보낼 때 1로 잘못 받을 확률과 1을 보낼 때 0으로 잘못 받을 확률이 2%이다. 그러면, 받은 신호가 0일 때, 이 신호가 원래 0이었을 확률은 얼마일까?

복습

확률의 정의 표본공간 S, 사건 A...

공리 1. $0 \le P(A) \le 1$

공리 2. $P(S) = 1$

공리3. $A_1 \cap A_2 = \emptyset \Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$

성질

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

$P(\emptyset) = 0$

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

$P(B) > 0$일 때, $P(A|B) = \frac{P(A \cap B)}{P(B)}$

$P(A) > 0$일 때, $P(B|A) = \frac{P(A \cap B)}{P(A)}$

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

정의

A와 B 독립 $\Leftrightarrow P(A|B) = P(A)$ 또는 $P(B|A) = P(B)$

정리

A와 B 독립 $\Leftrightarrow P(A \cap B) = P(A)P(B)$

베이즈공식

$$P(S0|R0) = \frac{P(S0)P(R0|S0)}{P(S0)P(R0|S0) + P(S1)P(R0|S1)}$$