6.5 결합확률과 조건부확률

이 절에서는 결합확률과 조건부확률의 개념을 공부한다. 조건부확률은 데이터로부터 어떤 결론을 얻을수 있는가를 나타내는 숫자이므로, 데이터 분석에 있어 가장 중요한 개념이다.

확률론 = 논리학,
가장 중요한 두 개념 : 조건부 확률 & 베이즈정리

조건부확률
- $A \rightarrow B$ 80% $P(B|A)$
- $B \rightarrow C$ 90% $P(C|B)$

베이즈정리 $\longrightarrow A \rightarrow C$

$\boxed{\text{범인찾기문제}}$

범인 표본공간 Ω: $\boxed{(w_1)\ (w_2)\ \cdots\ (w_{20})}$

- 남자 - 12명
- 여자 - 8명

주장.명제 = 사건 = subset
$A = \{②, ④, ⑤\}$
"②, ④, ⑤ 중에 1명이 범인이다"
($P(A) =$ _ % $\longrightarrow$ 주장의 신뢰도)

추리의 관점 : '범인 = 남자'라는 사건의 신뢰도 $= P(A)$
- 아무런 정보 없음 = 모든 표본이 범인이 될 확률은 동일하다고 가정
- 따라서 $P(A)$는 전체 용의자 수를 남자 용의자 수로 나눈 값.

(남자) $P(A) = \dfrac{|A|}{|\Omega|} = \dfrac{12}{12+8} = \dfrac{12}{20} = .6$

(여자) $P(A^C) = \dfrac{|A^C|}{|\Omega|} = .4$

④ 새로운 정보 (주장)

"범인은 머리가 긴 사람" = 사건 B

표본공간에서 긴머리 10명, 짧은머리 10명
이에 대한 추가정보가 없다면,

$$P(B) = \frac{|B|}{|\Omega|} = \frac{10}{10+10} = .5$$

$$P(B^c) = 1 - .5 = .5$$

Summary

- 살인사건 용의자 20명 = 남자 12명 + 여자 8명
  = 머리긴사람 10명 + 머리짧은사람 10명

- 범인이 남자일 확률
  - 남자 집합 (사건)이 범인 (선택된 표본)이 속해 있다는 주장의 신뢰도
  - $P(A) = 0.6$

- 범인이 머리카락이 길 확률
  - 머리카락이 긴 사람의 집합 (사건) B에 범인 (선택된 표본)이 속해 있다는 주장의 신뢰도
  - $P(B) = .5$

결합확률과 조건부확률

베이지안 확률론은 두 사건 A와 B의 관계를 알고 있다면 사건 B가 발생하였다는 사실로부터 기존에 알고 있는 사건 A에 대한 확률 $P(A)$를 좀 더 ~~정~~ 정확한 확률로 바꿀 수 있는 방법을 알려준다. 이를 위해서는 결합확률과 조건부확률이라는 두 가지 개념을 정의해야 한다.

결합확률 (joint probability)

- 사건 A와 B가 동시에 발생할 확률.

$$P(A \cap B) \quad \text{or} \quad P(A, B)$$

주변확률 (marginal probability)

- 결합확률과 대비되는 개념
- 결합되지 않는 개별 사건의 확률 $P(A)$ 또는 $P(B)$를 주변확률(marginal probability)이라고 한다.

$$P(A), \quad P(B)$$

조건부확률 (conditional probability)

- B가 사실일 경우에 대한 사건 A의 확률을 사건 B에 대한 사건 A의 조건부확률을 이라고 하며, 다음과 같이 표기한다.

$$P(A \mid B)$$

$$= \text{new } P(A) \text{ if } P(B) = 1$$

$$= \frac{P(A,B)}{P(B)}$$

$P(A,B)$ —— B 표본공간 내에서 $A \cap B$에 대한 확률.

$P(B)$ —— 표본공간을 B로 제한

조건부확률이 $\frac{P(A,B)}{P(B)}$ 로 정의된 근거:

① 사건 B가 사실 = 모든 가능한 표본은 사건 B에 포함

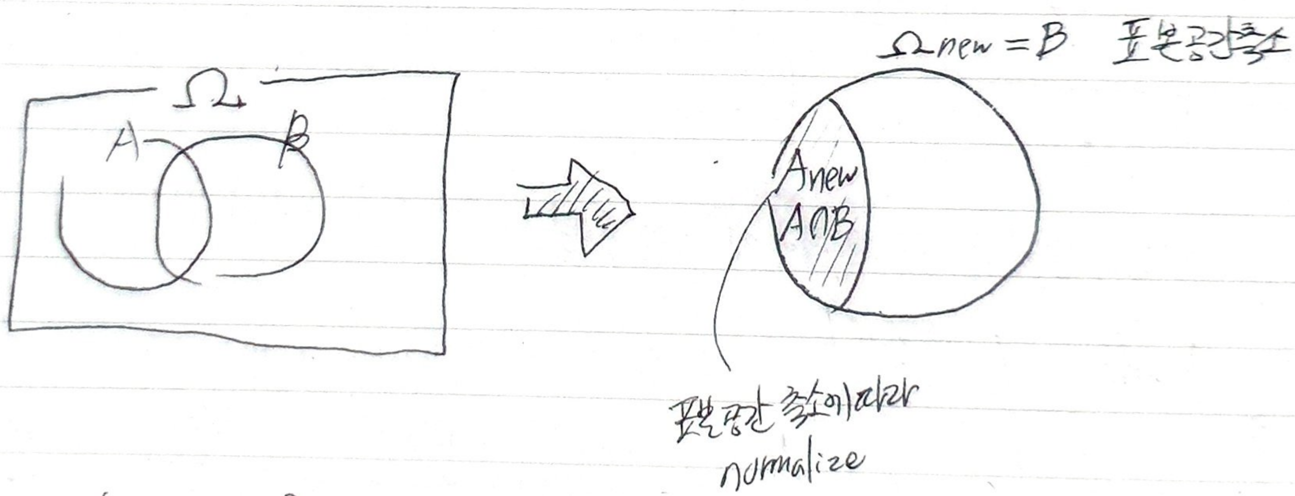
→ 새로운 실질표본공간 $\boxed{\Omega_{new} \to B}$

② 사건 A의 원소는 모두 사건 B의 원소도 되므로 사실상 사건 $A \cap B$의 원소가 된다.

즉, 새로운 실질적 $\boxed{A_{new} \to A \cap B}$ 가 된다.

③ 따라서 사건 A의 확률 즉, 신뢰도는 원래의 신뢰도(결합확률)를 새로운 표본공간의 신뢰도(확률)로 정규화(normalize)한 값

$$P(A|B) = \frac{P(A_{new})}{P(\Omega_{new})} = \frac{P(A,B)}{P(B)}$$



조건부확률 $P(A|B)$

① $P(A)$ given $P(B)=1$

② 표본이 사건 B에 속한다는 새로운 사실을 알게 되었을 때,

③ 이 표본이 사건 A에 속한다는 사실의 정확성(신뢰도)이 어떻게 변하는지 알려준다.

[다시 범인찾기문제]

$P(A)$ = 범인 남자

$P(B)$ = 범인 머리카락 길다

$P(A|B)$ = 범인의 머리카락이 길다는 사실을 알게되었을 때, 계산되는 $P(A)$의 값

$\boxed{\text{필요한 정보} = P(A, B)}$

(case1) 12명의 남자 중 머리카락이 긴 사람이 3명

	B	B^c	
A	3	9	12
A^c	7	1	8
	10	10	

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{3/20}{10/20} = \frac{3}{10}$$

(case 2) 12명의 남자 중 머리카락이 긴 사람 6명

	B	B^c	
A	6	6	12
A^c	4	4	8
	10	10	

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{6/20}{10/20} = \frac{6}{10}$$

* 새로운 정보 (사건B)와 상관없이 사건 A (남자가 범인일 확률) 확률은 변함없다. 이 경우, 사건 A가 사건 B와 서로 독립 (independent)이라고 한다.

NO

DATE.

$(\Omega = 20)$

	A	A^c	
B	3	7	10
B^c	9	1	10
	12	8	20

연습문제 6.5.1

case 1 (~~머리카락 긴 사람들 중~~ 남자 12명 중 머리 긴 사람 3명)

에 대해 다음 확률을 구하여라.

(1) $P(A|B^c)$ 범인 짧은머리일때 남자가 범인일 확률 $= \dfrac{P(A\cap B^c)}{P(B^c)} = \dfrac{9}{10}$

(2) $P(A^c|B)$ 범인 긴머리일때 여자가 범인일 확률 $= \dfrac{P(A^c\cap B)}{P(B)} = \dfrac{7}{10}$

(3) $P(A^c|B^c)$ 범인 짧은머리일때 여자가 범인일 확률 $= \dfrac{P(A^c\cap B^c)}{P(B^c)} = \dfrac{1}{10}$

(4) $P(B|A)$ ~~범인 긴머~~ 남자가 범인일때 긴머리일 확률 $= \dfrac{P(B\cap A)}{P(A)} = \dfrac{3}{12}$

(5) $P(B|A^c)$ 여자가 범인일때 긴머리일 확률 $= \dfrac{7}{8}$

(6) $P(B^c|A)$ 남자가 범인일때 짧은머리일 확률 $= \dfrac{9}{12}$

(7) $P(B^c|A^c)$ 여자가 범인일때 짧은머리일 확률 $= \dfrac{1}{8}$

독립

- 사건 A와 사건 B의 결합확률의 값이 다음과 같은 관계가 성립하면 두 사건 A와 B는 서로 독립(independent)이라고 정의한다.

$$P(A,B) = P(A)P(B)$$

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

$\therefore$ $P(A|B) = P(A)$ if A, B are independent

원인과 결과, 근거와 추론, 가정과 조건부 결론

조건부확률 $P(A|B)$에서 사건(주장/명제) B, A는 각각

- '가정(B)과 그 가정에 따른 조건부 결론(A)' 또는
- '원인(B)과 결과(A)' 또는
- '근거(B)와 추론(A)'

으로 생각할 수 있다.

결합확률의 정의를 바꿔쓰면 다음과 같이 되는데..

$$\boxed{P(A,B) = P(A|B)P(B)}$$

이 식은 다음과 같은 관점에서 볼 수 있다.

A, B가 모두 발생할 확률

= B라는 사건 확률과, B 사건이 발생한 경우 다시 A가 발생할 확률의 곱.

= (B가 진실일 확률) × (B가 진실일 때 A도 진실일 확률)

NO.

DATE.

예제

위 식을 응용하면 다음과 같은 수식도 성립한다.

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

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

연습문제 6.5.4

B, C가 독립일때, 다음이 성립함을 증명하여라.

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

$$\begin{aligned} P(A,B|C)P(C) &= P(A,B,C) \\ &= P(A|B,C)P(B,C) \\ &= P(A|B,C)P(B)P(C) \end{aligned}$$

$$\therefore P(A,B|C) = P(A|B,C)P(B)$$

다음 수식을 증명하라.

(1) $P(A, B, C, D) = P(A, B \mid C, D)\, P(C, D)$

$P(A, B, C, D) = P((A \cap B) \cap (C \cap D))$

$= P(A, B \mid C, D)\, P(C, D)$

(2) $P(A, B \mid C)\, P(C) = P(A \mid B, C)\, P(B, C)$

$P(A, B, C) = P(C \mid A, B)\, P(A, B)$

$= P(B \mid A, C)\, P(A, C)$

$= P(A \mid B, C)\, P(B, C)$

$= P(B, C \mid A)\, P(A)$

$\vdots$

(3) $P(A, B, C \mid D, E) = \dfrac{P(A, B \mid C, D, E)\, P(C, D \mid E)\, P(E)}{P(D, E)}$

$P(A, B, C \mid D, E) = \dfrac{P(A, B, C, D, E)}{P(D, E)}$

$= \dfrac{P(A, B \mid C, D, E)\, P(C, D, E)}{P(D, E)}$

$= \dfrac{P(A, B \mid C, D, E)\, P(C \mid D, E)\, P(D, E)}{P(D, E)}$

NO.

DATE.

사슬 법칙 (연쇄법칙, chain rule)

조건부확률과 결합확률의 관계를 확장하면 복수의 사건 $X_1, X_2, \ldots, X_N$에 대한 조건부 확률을 다음처럼 쓸 수 있다. 이를 사슬법칙(chain rule)이라고 한다.

$$P(X_1, X_2) = P(X_1)P(X_2 \mid X_1)$$

$$P(X_1, X_2, X_3) = P(X_3 \mid X_1, X_2)P(X_1, X_2)$$

$$= P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2)$$

$$P(X_1, X_2, X_3, X_4) = P(X_4 \mid X_1, X_2, X_3)P(X_1, X_2, X_3)$$

$$= P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2)P(X_4 \mid X_1, X_2, X_3)$$

$\vdots$

$$P(X_1, \ldots, X_N) = \cancel{P(X_4 \mid X_1, X_2, X_3)P(X_1, X_2, X_3)}$$

$$\cancel{= P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2)P}$$

$$\boxed{P(X_1)\prod_{i=2}^{N} P(X_i \mid X_1, \ldots, X_{i-1})}$$

- 그래프모델, 텍스트 분석, 언어모델에서 활용

아버지 가 방에 계신다

X_1 X_2 X_3 X_4 ...

[X_1부터 X_{i-1}까지의 단어로 구성된 문장이 존재할 확률]

확률변수

확률적인 숫자 값을 출력하는 변수를 확률변수(random variable)라고 한다. 확률변수에 대한 더 수학적인 정의는 이후에 다시 공부할 것이다. 여기에서는 주사위처럼 어떤 숫자가 나올 수 있지만, 정확히 어떤 숫자가 나올지 예측할 수 없는 기계라고 생각하면 된다. 확률변수는 보통 X, Y 처럼 알파벳 대문자로 표기한다.

위의 범인 찾기 문제에서는 두 확률변수 X, Y를 정의할 수 있다. 확률변수 X는 성별을 나타내고 확률변수 Y는 머리카락이 긴지 짧은지를 나타낸다.

- $X=0$ 인 경우가 사건 A (남자인 사건)
- $X=1$ 인 경우가 사건 A^c (여자인 사건)
- $Y=0$ 인 경우가 사건 B (머리카락이 긴 사건)
- $Y=1$ 인 경우가 사건 B^c (머리카락이 짧은 사건)

확률변수는 확률분포를 그 안에 내포하고 있어서, 그 확률분포에 따라 숫자를 출력할 수 있다. 확률변수 X가 가진 확률을 확률변수의 확률 $P(X)$라고 한다. 위 예에서 확률변수 X는 사건 A와 사건 A^c을 가질 수 있고 각 사건에 할당된 확률은 다음과 같은 표로 나타낼 수 있다.

확률변수 X의 값	확률변수 X의 사건	각 사건에 할당된 확률
$X=0$	A	$P(X=0)=P(A)=\frac{12}{20}$
$X=1$	A^c	$P(X=1)=P(A^c)=\frac{8}{20}$

X, Y 결합확률

X값	Y값	X사건	Y사건	각 사건 조합에 할당된 확률
X=0	Y=0	A	B	$P(X=0, Y=0) = P(A,B) = \frac{3}{20}$
X=0	Y=1	A	B^c	$P(X=0, Y=1) = P(A,B^c) = \frac{9}{20}$
X=1	Y=0	A^c	B	$P(X=1, Y=0) = P(A^c,B) = \frac{7}{20}$
X=1	Y=1	A^c	B^c	$P(X=1, Y=1) = P(A^c,B^c) = \frac{1}{20}$

pgmpy 패키지

Pgmpy (Probablistic Graphical Models in Python) 패키지를 사용하면 이산확률모델을 쉽게 구현할수 있다.

```
pip install pgmpy
```