

6.5 결합확률과 조건부확률

이 절에서는 결합확률과 조건부확률의 개념을 공부한다. 조건부확률은 데이터로부터 어떤 결론을 얻을 수 있는가를 나타내는 숫자이므로 데이터 분석에 있어서 가장 중요한 개념이다.

범인 찾기 문제

살인 사건이 발생했다고 가정하자. 경찰은 전체 용의자 목록을 가지고 있다. 베이지안 확률론 관점에서 전체 용의자 목록은 바로 표본공간이다. 우리가 알고 싶은 것은 전체 용의자 목록(표본공간)에서 누가 범인(선택된 표본)인가 하는 점이다. 현재 표본공간은 용의자 20명으로 구성되어 있으며 이 중 남자가 12명, 여자가 8명이라고 가정한다.

만약 담당 형사가 범인은 남자라고 생각한다면, '범인이 남자이다'라는 주장은 확률론적 관점에서 남성인 용의자(표본)로만 이루어진 사건(표본공간의 부분 집합)이 된다. 이를 사건 A 라고 하자.

이때 우리가 관심을 가지는 것은 "범인이 남자"라는 사건 A 의 신뢰도 즉, 사건 A 의 확률 $P(A)$ 다. 아무런 추가 정보가 없다면 모든 사람이 범인일 가능성이 같기 때문에 범인이 남자일 확률 $P(A)$ 는 다음과 같이 전체 용의자 수로 남자 용의자 수를 나눈 값이 된다.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{12}{12 + 8} = \frac{12}{20} = 0.6 \quad (6.5.1)$$

반대로 "범인은 여자"라는 사건은 여집합 A^C 로 표현할 수 있고 "범인이 여자일 확률"은 다음처럼 계산한다.

$$P(A^C) = \frac{|A^C|}{|\Omega|} = \frac{8}{12 + 8} = \frac{8}{20} = 0.4 \quad (6.5.2)$$

이때 새로운 사건 B 가 발생했다고 하자. 범인의 것으로 추정되는 긴 머리카락을 발견했다. 이머리카락에서 범인은 머리카락이 길다는 가능성이 제시되었다.

이 새로운 사건 B 는 확률론적으로는 새로운 용의자 목록, 즉 머리카락이 긴 사람의 목록이라는 표본공간의 새로운 부분 집합을 의미한다. 그리고 사건 B 가 발생했다는 것은 이 용의자 목록에 진짜로 범인이 포함되었다는 뜻이다.

현재 표본공간 즉, 전체 용의자 목록에는 머리카락이 긴 사람이 10명, 머리카락이 짧은 사람이 10명이다. 만약 이 사건이 진실이라는 보장이 없다면, 사건 B 에 대한 확률 $P(B)$, 즉 머리카락이 긴 사람이 범인이라는 주장의 신뢰도는 다음과 같다.

$$P(B) = \frac{|B|}{|\Omega|} = \frac{10}{10 + 10} = \frac{10}{20} = 0.5 \quad (6.5.3)$$

반대로 머리카락이 짧은 사람의 사건은 B^C 이고 범인이 머리카락이 짧은 확률은 다음처럼 계산한다.

$$P(B^C) = \frac{|B^C|}{|\Omega|} = \frac{10}{10 + 10} = \frac{10}{20} = 0.5 \quad (6.5.4)$$

지금까지의 상황을 요약하면 다음과 같다.

- 살인 사건 발생. 용의자는 20명
 - 남자 12명, 여자 8명
 - 머리카락이 긴 사람 10명, 머리카락이 짧은 사람 10명

- 범인이 남자일 확률
 - 남자 집합(사건) A 에 범인(선택된 표본)이 속해 있다는 주장의 신뢰도: $P(A) = 0.6$
- 범인이 머리카락이 길 확률
 - 머리카락이 긴 사람의 집합(사건) B 에 범인(선택된 표본)이 속해 있다는 주장의 신뢰도 : $P(B) = 0.5$

결합확률과 조건부확률

베이저안 확률론은 두 사건 A 와 B 의 관계를 알고 있다면 사건 B 가 발생하였했다는 사실로부터 기존에 알고 있는 사건 A 에 대한 확률 $P(A)$ 를 좀 더 정확한 확률로 바꿀 수 있는 방법을 알려준다. 이를 위해서는 결합확률과 조건부확률이라는 두 가지 개념을 정의해야 한다.

결합확률(joint probability)은 사건 A 와 B 가 동시에 발생할 확률이다. 즉, 사건(명제/주장) A 도 진실이고 사건(명제/주장) B 도 진실이므로 사건 A 와 B 의 교집합의 확률을 계산하는 것과 같다.

$$P(A \cap B) \text{ or } P(A, B) \quad (6.5.5)$$

결합확률과 대비되는 개념으로 결합되지 않는 개별 사건의 확률 $P(A)$ 또는 $P(B)$ 를 **주변확률(marginal probability)**라고 한다.

또한 B 가 사실일 경우의 사건 A 에 대한 확률을 사건 B 에 대한 사건 A 의 **조건부확률(conditional probability)**이라고 하며 다음과 같이 표기한다.

$$P(A|B) \quad (6.5.6)$$

위 수식에서 기호 $|$ 는 if를 뜻한다. 즉 다음과 같다.

$$P(A|B) \equiv \text{new } P(A) \text{ if } P(B) = 1 \quad (6.5.7)$$

이 조건부확률의 값은 다음처럼 정의한다.

$$P(A|B) = \frac{P(A, B)}{P(B)} \quad (6.5.8)$$

조건부확률이 위와 같이 정의된 근거는 다음과 같다.

1. 사건 B 가 사실이므로 모든 가능한 표본은 사건 B 에 포함되어야 한다. 즉, 새로운 실질적 표본공간은 $\Omega_{\text{new}} \rightarrow B$ 가 된다.
2. 사건 A 의 원소는 모두 사건 B 의 원소도 되므로 사실상 사건 $A \cap B$ 의 원소가 된다. 즉, 새로운 실질적 $A_{\text{new}} \rightarrow A \cap B$ 가 된다.
3. 따라서 사건 A 의 확률 즉, 신뢰도는 원래의 신뢰도(결합확률)를 새로운 표본공간의 신뢰도(확률)로 정규화(normalize)한 값이라고 할 수 있다.

$$P(A|B) = \frac{P(A_{\text{new}})}{P(\Omega_{\text{new}})} = \frac{P(A, B)}{P(B)} \quad (6.5.9)$$

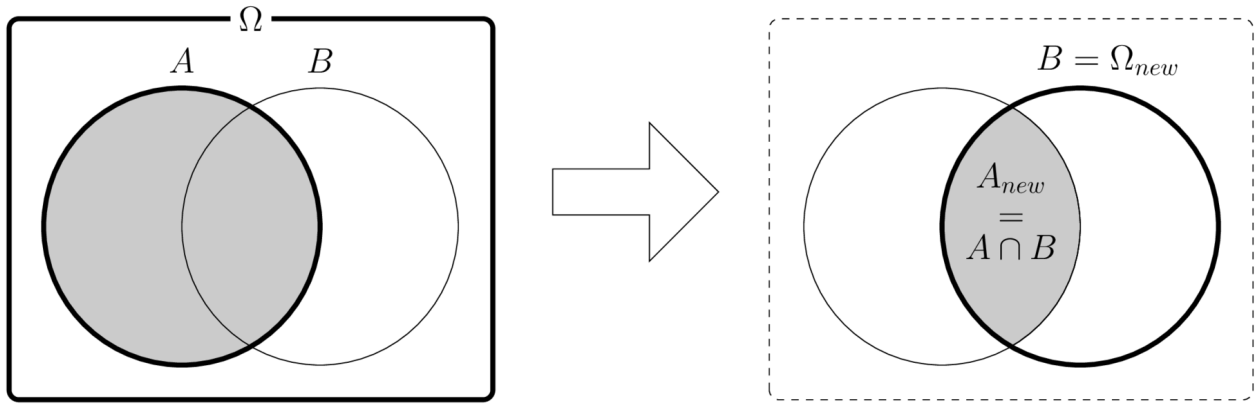


그림 6.5.1 : 조건부확률의 정의

- 조건부확률 $P(A|B)$
 - 사건 B가 발생한 경우의 사건 A의 확률
 - 표본이 이벤트 B에 속한다는 새로운 **사실**을 알게 되었을 때,
 - 이 표본이 사건 A에 속한다는 사실의 정확성(신뢰도)이 어떻게 변하는지를 알려준다.

예를 들어 범인 찾기 문제에서는 조건부확률을 다음처럼 정의한다.

- $P(A)$: 범인이 남자일 확률
- $P(B)$: 범인이 머리카락이 길 확률
- $P(A|B)$: 범인이 머리카락이 길다는 **사실**을 알게 되었을 때, 달라진(갱신된) "범인이 남자일 확률"

조건부확률의 값을 구하기 위해서는 결합확률의 값을 알아야 한다. 그런데 사건 A와 사건 B의 결합확률의 값 $P(A, B)$ 은 기존의 사건 A의 확률 $P(A)$ 나 사건 B의 확률 $P(B)$ 와는 전혀 무관한 별개의 정보이다. 즉, 수학적으로 계산하여 구할 수 있는 값이 아니라 외부에서 주어지지 않으면 안되는 정보인 것이다.

범인 찾기의 경우에도 이미 주어진 정보 $P(A)$, $P(B)$ 와 관계없이 $P(A, B)$ 는 여러 가지 경우가 있을 수 있다.

예를 들어 어떤 경우에는 12명의 남자 중 머리카락이 긴 사람이 다음과 같이 3명일 수도 있고(경우 1)

	범인이 머리카락이 길다: $P(B) = 0.5$	범인이 머리카락이 길지 않다: $P(B^C) = 0.5$	계
범인이 남자다: $P(A) = 0.6$	3명 $P(A, B) = \frac{3}{20}$	9명 $P(A, B^C) = \frac{9}{20}$	12명
범인이 여자다: $P(A^C) = 0.4$	7명 $P(A^C, B) = \frac{7}{20}$	1명 $P(A^C, B^C) = \frac{1}{20}$	8명
계	10명	10명	

또 다른 경우에는 12명의 남자 중 머리카락이 긴 사람이 다음과 같이 6명일 수도 있다(경우 2).

	범인이 머리카락이 길다: $P(B) = 0.5$	범인이 머리카락이 길지 않다: $P(B^C) = 0.5$	계
범인이 남자다: $P(A) = 0.6$	6명 $P(A, B) = \frac{6}{20}$	6명 $P(A, B^C) = \frac{6}{20}$	12명

범인이 여자다: $P(A^C) = 0.4$	4명 $P(A^C, B) = \frac{4}{20}$	4명 $P(A, B^C) = \frac{4}{20}$	8명
계	10명	10명	

이 두 경우에 대해 조건부확률 $P(A|B)$ 를 구해보자.

만약 머리카락이 긴 남자가 3명이라면

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{3/20}{10/20} = \frac{3}{10} \quad (6.5.10)$$

이 된다. 원래 사건 A 의 확률 $P(A)$ 가 0.6 즉 60% 였으므로 범인이 머리카락이 길다는 정보로 인해 남자가 범인일 확률은 절반으로 똑 떨어졌다.

만약 머리카락이 긴 남자가 6명이라면

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{6/20}{10/20} = \frac{6}{10} \quad (6.5.11)$$

이 된다.

이 경우에는 새로운 정보(사건 B)가 주어지든 주어지지 않든 남자가 범인일 확률은 변함없다. 이러한 경우에는 사건 A 가 사건 B 와 서로 독립(**independent**)이라고 한다.

연습 문제 6.5.1

(경우 1)에 대해 다음 확률을 구하라.

- (1) 범인이 머리카락이 짧다면 범인이 남자일 확률 $P(A|B^C)$
- (2) 범인이 머리카락이 길다면 범인이 여자일 확률 $P(A^C|B)$
- (3) 범인이 머리카락이 짧다면 범인이 여자일 확률 $P(A^C|B^C)$
- (4) 범인이 남자라면 범인이 머리카락이 길 확률 $P(B|A)$
- (5) 범인이 여자라면 범인이 머리카락이 길 확률 $P(B|A^C)$
- (6) 범인이 남자라면 범인이 머리카락이 짧을 확률 $P(B^C|A)$
- (7) 범인이 여자라면 범인이 머리카락이 짧을 확률 $P(B^C|A^C)$

연습 문제 6.5.2

(경우 2)에 대해 다음 확률을 구하라.

- (1) 범인이 머리카락이 짧다면 범인이 남자일 확률 $P(A|B^C)$
- (2) 범인이 머리카락이 길다면 범인이 여자일 확률 $P(A^C|B)$
- (3) 범인이 머리카락이 짧다면 범인이 여자일 확률 $P(A^C|B^C)$
- (4) 범인이 남자라면 범인이 머리카락이 길 확률 $P(B|A)$
- (5) 범인이 여자라면 범인이 머리카락이 길 확률 $P(B|A^C)$

(6) 범인이 남자라면 범인이 머리카락이 짧을 확률 $P(B^C|A)$

(7) 범인이 여자라면 범인이 머리카락이 짧을 확률 $P(B^C|A^C)$

독립

수학적으로는 사건 A 와 사건 B 의 결합확률의 값이 다음과 같은 관계가 성립하면 두 사건 A 와 B 는 서로 독립(independent)이라고 정의한다.

$$P(A, B) = P(A)P(B) \quad (6.5.12)$$

독립인 경우 조건부확률과 원래의 확률이 같아짐을 알 수 있다. 즉, B 라는 사건이 발생하든 말든 사건 A 에는 전혀 영향을 주지 않는다는 것이다.

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \quad (6.5.13)$$

원인과 결과, 근거와 추론, 가정과 조건부 결론

조건부확률 $P(A|B)$ 에서 사건(주장/명제) B , A 는 각각

- ‘가정과 그 가정에 따른 조건부 결론’ 또는
- ‘원인과 결과’ 또는
- ‘근거와 추론’

으로 생각할 수도 있다.

또 결합확률의 정의를 바꿔 쓰면 다음과 같이 되는데,

$$P(A, B) = P(A|B)P(B) \quad (6.5.14)$$

이 식은 다음과 같은 관점에서 볼 수 있다.

A, B 가 모두 발생할 확률은 B 라는 사건이 발생할 확률과 그 사건이 발생한 경우 다시 A 가 발생할 경우의 곱

예제

위 식을 응용하면 다음과 같은 수식도 성립한다.

$$P(A, B, C) = P(A|B, C)P(B, C) \quad (6.5.15)$$

확률표기에서 쉼표(comma)가 교집합을 뜻한다는 것을 기억하면 이 식은 쉽게 증명할 수 있다.

$$\begin{aligned} P(A, B, C) &= P(A \cap B \cap C) \\ &= P(A \cap (B \cap C)) \\ &= P(A|B \cap C)P(B \cap C) \\ &= P(A|B, C)P(B, C) \end{aligned} \quad (6.5.16)$$

다음 수식을 증명하라.

(1)

$$P(A, B, C, D) = P(A, B|C, D)P(C, D) \quad (6.5.17)$$

(2)

$$P(A, B|C)P(C) = P(A|B, C)P(B, C) \quad (6.5.18)$$

(3)

$$P(A, B, C|D, E) = \frac{P(A, B|C, D, E)P(C, D|E)P(E)}{P(D, E)} \quad (6.5.19)$$

연습 문제 6.5.4

B, C 가 독립인 사건일 때 다음이 성립함을 증명하라.

$$P(A, B|C) = P(A|B, C)P(B) \quad (6.5.20)$$

사슬 법칙

조건부확률과 결합확률의 관계를 확장하면 복수의 사건 X_1, X_2, \dots, X_N 에 대한 조건부 확률을 다음처럼 쓸 수 있다. 이를 **사슬 법칙(chain rule)**이라고 한다.

$$\begin{aligned} P(X_1, X_2) &= P(X_1)P(X_2|X_1) \\ P(X_1, X_2, X_3) &= P(X_3|X_1, X_2)P(X_1, X_2) \\ &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \\ P(X_1, X_2, X_3, X_4) &= P(X_4|X_1, X_2, X_3)P(X_1, X_2, X_3) \\ &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3) \\ &\vdots \\ P(X_1, \dots, X_N) &= P(X_1) \prod_{i=2}^N P(X_i|X_1, \dots, X_{i-1}) \end{aligned} \quad (6.5.21)$$

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 수는 없는 기계라고 생각하면 된다. 확률변수는 보통 X, Y 처럼 알파벳 대문자로 표기한다.

위의 범인 찾기 문제에서는 두 확률변수 X, Y 를 정의할 수 있다. 확률변수 X 는 성별을 나타내고 확률변수 Y 는 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 X 가 가진 확률을 확률변수의 확률 $P(X)$ 라고 한다. 위 예에서 확률변수 X 는 사건 A 와 사건 A^C 를 가질 수 있고 각 사건에 할당된 확률은 다음과 같은 표로 나타낼 수 있다.

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

결합확률의 확률분포는 각 확률변수가 가질 수 있는 값의 조합으로 나타난다. 예를 들어 확률변수 X 는 사건 A 와 사건 A^C 를 가질 수 있고 확률변수 Y 는 사건 B 와 사건 B^C 를 가질 수 있으므로 두 사건의 조합에 대한 결합 확률은 다음 표와 같이 표현한다. 이 표에서 확률값은 머리카락 긴 남자가 3명인 경우(경우 1)에서 구한 값이다.

X 의 값	Y 의 값	X 의 사건	Y 의 사건	각 사건의 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	$P(X = 1, Y = 1)$
				$= P(A^C, B^C) = \frac{1}{20}$

두 확률변수 X, Y 가 가질 수 있는 모든 사건의 조합에 대해 독립이 성립하면 두 확률변수 X, Y 가 독립이라고 한다. 위 결합확률 표에서 주변확률의 곱을 구해서 결합확률과 비교해보면 확률변수 X, Y 는 독립이 아니라는 것을 알 수 있다.

X 의 값	Y 의 값	X 의 확률	Y 의 확률	주변확률의 곱	결합확률
$X=0$	$Y=0$	$P(X = 0)$	$P(Y = 0)$	$P(X = 0)P(Y = 0)$	$P(X = 0, Y = 0)$
		$= 0.6$	$= 0.5$	$= 0.3$	$= \frac{3}{20}$
$X=0$	$Y=1$	$P(X = 0)$	$P(Y = 1)$	$P(X = 0)P(Y = 1)$	$P(X = 0, Y = 1)$
		$= 0.6$	$= 0.5$	$= 0.3$	$= \frac{9}{20}$
$X=1$	$Y=0$	$P(X = 1)$	$P(Y = 0)$	$P(X = 1)P(Y = 0)$	$P(X = 1, Y = 0)$
		$= 0.4$	$= 0.5$	$= 0.2$	$= \frac{7}{20}$

X 의 값	Y 의 값	X 의 확률	Y 의 확률	주변확률의 곱	결합확률
		$P(X = 0)$	$P(Y = 0)$	$P(X = 0)P(Y = 0)$	$P(X = 0, Y = 0)$
$X=1$	$Y=1$	$= 0.4$	$= 0.5$	$= 0.2$	$= \frac{1}{20}$

연습 문제 6.5.5

(경우 2)에 대해 위와 같은 표를 완성하고 확률변수 X, Y 가 독립임을 보여라.

피지엠패이 패키지

pgmpy(Probabilistic Graphical Models in Python) 패키지를 사용하면 이산확률모형을 쉽게 구현할 수 있다. 피지엠패이 패키지는 다음 명령으로 설치한다.

```
pip install pgmpy
```

피지엠패이 패키지의 JointProbabilityDistribution 클래스는 결합확률 모형을 만드는 데 사용하는 클래스다. 사용법은 다음과 같다.

```
JointProbabilityDistribution(variables, cardinality, values)
```

- variables : 확률변수의 이름 문자열의 리스트. 정의하려는 확률변수가 하나인 경우에도 리스트로 넣어야 한다.
- cardinality : 각 확률변수의 표본 혹은 배타적 사건의 수의 리스트
- values : 확률변수의 모든 표본(조합)에 대한 (결합)확률값의 리스트

variables 에 들어가는 인수가 사건의 이름이 아니라 확률변수의 이름이라는 점에 주의하라. pgmpy에서는 사건의 이름을 명시적으로 지정할 수 없고 입력한 사건의 수가 K 일 때, $0, 1, \dots, K - 1$ 와 같이 숫자로 지정된다.

범인 찾기 예제에서 성별을 나타내는 확률변수 X 와 머리카락 길이를 나타내는 확률변수 Y 의 확률을 다음과 같이 표로 정의할 수 있다.

In [1]:

```
from pgmpy.factors.discrete import JointProbabilityDistribution as JPD

px = JPD(['X'], [2], np.array([12, 8]) / 20)
print(px)
```

```
+-----+-----+
| X      | P(X) |
+-----+-----+
| X(0)   | 0.6000 |
+-----+-----+
| X(1)   | 0.4000 |
+-----+-----+
```

연습 문제 6.5.6

위의 범인 찾기 문제의 예에서 확률변수 Y 의 확률을 JointProbabilityDistribution 클래스 객체 py 로 구현하라.

확률변수 X 와 확률변수 Y 의 결합확률은 다음처럼 정의한다.

In [2]:

```
pxy = JPD(['X', 'Y'], [2, 2], np.array([3, 9, 7, 1]) / 20)
print(pxy)
```

X	Y	P(X,Y)
X(0)	Y(0)	0.1500
X(0)	Y(1)	0.4500
X(1)	Y(0)	0.3500
X(1)	Y(1)	0.0500

연습 문제 6.5.7

위의 범인 찾기 문제의 예에서 남자 12명 중 머리카락이 긴 사람이 6명인 경우(경우 2)의 결합확률 모형을 `JointProbabilityDistribution` 클래스 객체 `pxy2` 로 구현하라.

`JointProbabilityDistribution` 클래스는 결합확률로부터 주변확률을 계산하는 `marginal_distribution()` 메서드, `marginalize()` 메서드와 조건부확률을 계산하는 `conditional_distribution()` 메서드를 제공한다.

`marginal_distribution(values, inplace=True)`

- `values`: 주변확률을 구할 확률변수의 이름 문자열 리스트
- `inplace`: `True` 이면 객체 자신을 주변확률 모형으로 변화시킨다. `False` 면 주변확률 모형 객체를 반환한다.

`marginalize(values, inplace=True)`

- `values`: 어떤 확률변수의 주변확률을 구하기 위해 없앨 확률변수의 이름 문자열 리스트
- `inplace`: `True` 이면 객체 자신을 주변확률 모형으로 변화시킨다. `False` 면 주변확률 모형 객체를 반환한다.

`conditional_distribution(values, inplace=True)`

- `values`: 주변확률을 구할 확률변수의 이름 문자열과 값을 묶은 튜플의 리스트
- `inplace`: `True` 이면 객체 자신을 조건부확률 모형으로 변화시킨다. `False` 면 조건부확률 모형 객체를 반환한다.

`marginal_distribution()` 메서드는 인수로 받은 확률변수에 대한 주변확률분포를 구한다. 다음 코드는 결합확률로부터 주변확률 $P(A)$, $P(A^C)$ 를 계산한다.

In [3]:

```
pmx = pxy.marginal_distribution(['X'], inplace=False)
print(pmx)
```

+-----+	+-----+	
X	P(X)	
+-----+	+-----+	
X(0)	0.6000	
+-----+	+-----+	
X(1)	0.4000	
+-----+	+-----+	

`marginalize()` 메서드는 인수로 받은 확률변수를 주변화(`marginalize`)하여 나머지 확률변수에 대한 주변확률 분포를 구한다. 다음 코드도 앞과 마찬가지로 결합확률로부터 주변확률 $P(A)$, $P(A^C)$ 를 계산한다.

In [4]:

```
pmx = pxy.marginalize(['Y'], inplace=False)
print(pmx)
```

+-----+	+-----+	
X	P(X)	
+-----+	+-----+	
X(0)	0.6000	
+-----+	+-----+	
X(1)	0.4000	
+-----+	+-----+	

다음 코드는 결합확률로부터 주변확률 $P(B)$, $P(B^C)$ 를 계산한다.

In [5]:

```
py = pxy.marginal_distribution(['Y'], inplace=False)
print(py)
```

+-----+	+-----+	
Y	P(Y)	
+-----+	+-----+	
Y(0)	0.5000	
+-----+	+-----+	
Y(1)	0.5000	
+-----+	+-----+	

`conditional_distribution()` 메서드를 사용하면 어떤 확률변수가 어떤 사건이 되는 조건에 대해 조건부확률 값을 계산한다. 다음 코드는 결합확률로부터 조건부확률 $P(B|A)$, $P(B^C|A)$ 를 계산한다.

In [6]:

```
py_on_x0 = pxy.conditional_distribution([( 'X', 0)], inplace=False) # 사건 A에 대한 조건부확률
print(py_on_x0)
```

Y	P(Y)
Y(0)	0.2500
Y(1)	0.7500

다음 코드는 결합확률로부터 조건부확률 $P(B|A^C)$, $P(B^C|A^C)$ 를 계산한다.

In [7]:

```
py_on_x1 = pxy.conditional_distribution([( 'X', 1)], inplace=False)
print(py_on_x1)
```

Y	P(Y)
Y(0)	0.8750
Y(1)	0.1250

다음 코드는 결합확률로부터 조건부확률 $P(A|B)$, $P(A^C|B)$ 를 계산한다.

In [8]:

```
px_on_y0 = pxy.conditional_distribution([( 'Y', 0)], inplace=False) # 사건 B에 대한 조건부확률
print(px_on_y0)
```

X	P(X)
X(0)	0.3000
X(1)	0.7000

check_independence() 메서드를 이용하면 두 확률변수 간의 독립도 확인할 수 있다.

In [9]:

```
pxy.check_independence([ 'X'], [ 'Y'])
```

Out[9]:

False

두 개의 JointProbabilityDistribution 객체끼리 곱하면 두 분포가 독립이라는 가정하에 결합확률을 구한다. 이 값과 원래의 결합확률을 비교하면 독립이 아니라는 것을 알 수 있다.

In [10]:

```
print(px * py)
print(pxy)
```

X	Y	P(X,Y)	
=====	=====	=====	
X(0)	Y(0)	0.3000	
-----	-----	-----	
X(0)	Y(1)	0.3000	
-----	-----	-----	
X(1)	Y(0)	0.2000	
-----	-----	-----	
X(1)	Y(1)	0.2000	
-----	-----	-----	
X	Y	P(X,Y)	
=====	=====	=====	
X(0)	Y(0)	0.1500	
-----	-----	-----	
X(0)	Y(1)	0.4500	
-----	-----	-----	
X(1)	Y(0)	0.3500	
-----	-----	-----	
X(1)	Y(1)	0.0500	
-----	-----	-----	

연습 문제 6.5.8

위에서 구현한 `JointProbabilityDistribution` 클래스 객체 `pxy2`로부터 주변확률 모형 및 조건부확률 모형을 구하라. 또 `check_independence()` 메서드를 이용하여 사건 A, B의 독립을 확인하라.