AI 기초

2019-2020

강봉주

조건부 확률과 독립

조건부 확률과 독립

[조건부 확률]

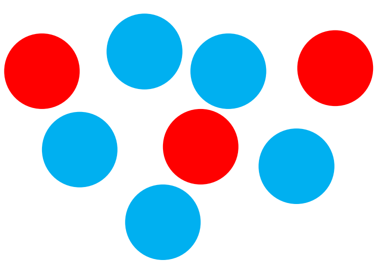
어떤 확률 시행에서 관심이 있는 결과는 특정한 표본 공간의 부분 집합에만 있다고 하자. 즉, $\omega_1 \subset \Omega$. 이런 경우를 ω_1 $(P(\omega_1) > 0)$ 을 표본 공간으로 하는 새로운 확률을 정의하는 문제이다.

조건부 확률과 독립

[조건부 확률]

예제)

주머니 속에 빨간 돌이 3개 파란 돌이 5개 있을 때, 임의로 빼낸 돌은 다시 넣지 않고(비복원) 연속적으로 돌을 꺼낸다고 하자. 이 때 첫번째 꺼낸 돌이 빨간 돌(ω_1)일 때 두번째 꺼낸 돌(ω_2)이 파란 돌일 확률은?

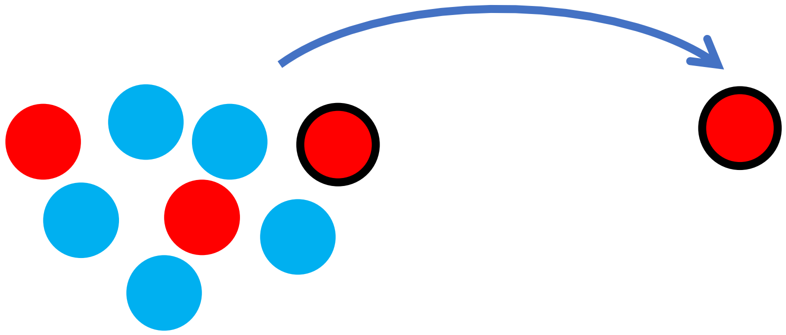


조건부 확률과 독립

[조건부 확률]

예제)

첫번째 꺼낸 돌이 빨간 돌(ω_1)일 확률



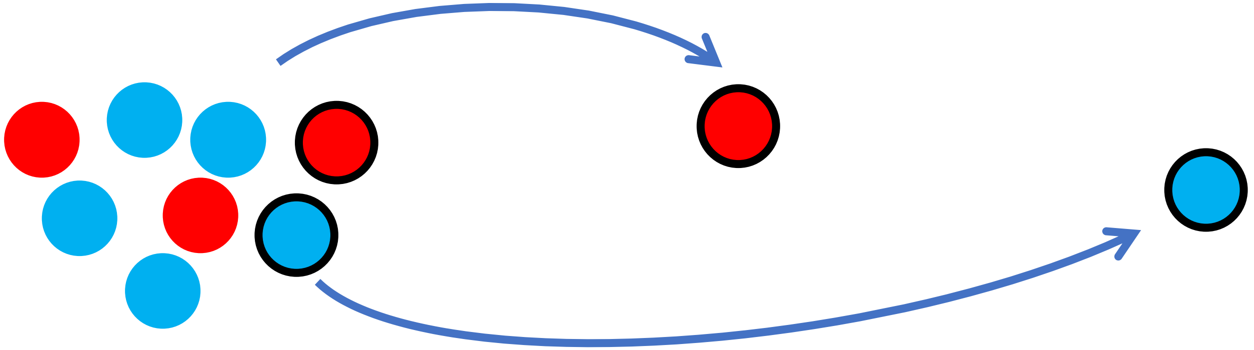
$$\Pr(\omega_1) = \frac{\binom{3}{1}}{\binom{8}{1}} = \frac{3}{8}$$

조건부 확률과 독립

[조건부 확률]

예제)

첫번째 꺼낸 돌이 빨간 돌(ω_1)이고 두번째 꺼낸 돌(ω_2)이 파란 돌일 확률



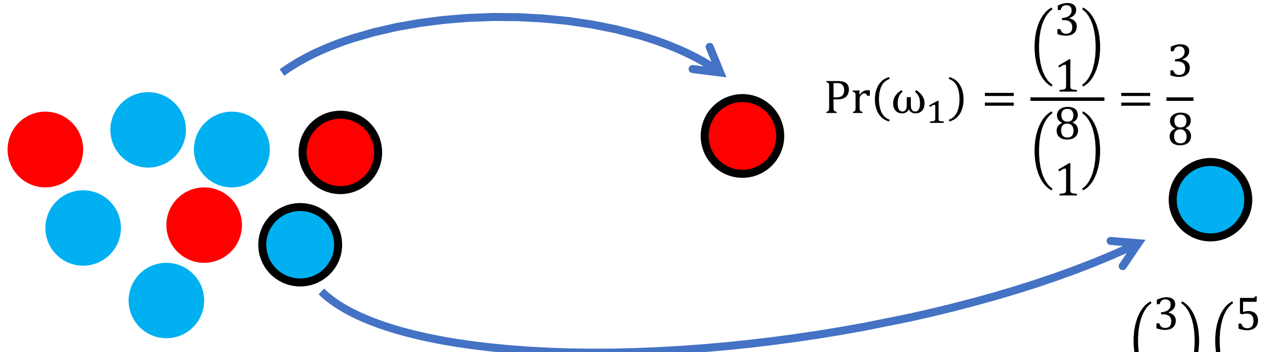
$$\Pr(\omega_1 \cap \omega_2) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{1}\binom{7}{1}} = \frac{15}{56}$$

조건부 확률과 독립

[조건부 확률]

예제)

첫번째 꺼낸 돌이 빨간 돌(ω_1)일 때 두번째 꺼낸 돌(ω_2)이 파란 돌일 확률



$$\Pr(\omega_1) = \frac{\binom{3}{1}}{\binom{8}{1}} = \frac{3}{8}$$

$$\Pr(\omega_1 \cap \omega_2) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{1}\binom{7}{1}} = \frac{15}{56}$$

$$\Pr(\omega_2|\omega_1) = \frac{\Pr(\omega_1 \cap \omega_2)}{\Pr(\omega_1)} = \frac{\frac{15}{56}}{\frac{3}{8}} = \frac{5}{7}$$

조건부 확률과 독립

[조건부 확률]

예제)

첫번째 꺼낸 돌이 빨간 돌(ω_1)인 경우에 표본 공간은 돌이 7개 이 중에서 파란 돌을 꺼낼 확률

$$\Pr(\omega_2|\omega_1) = \frac{\binom{5}{1}}{\binom{7}{1}} = \frac{5}{7}$$

조건부 확률과 독립

[조건부 확률]

- 조건부 확률의 정의

$$A_1, A_2 \subset \mathcal{X}, \qquad P(A_1) > 0$$

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$$

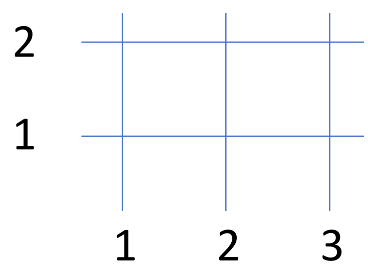
조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

$$f(x_1, x_2) = \frac{x_1 + x_2}{21}, x_1 = 1, 2, 3, \qquad x_2 = 1, 2$$

$x_2 = 1, 2$



$x_1 = 1, 2, 3$

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

In

```
# 결합 확률 정의
from pgmpy.factors.discrete import JointProbabilityDistribution as JPD

prob = list()
for i in np.arange(1, 4):
    for j in np.arange(1,3):
        prob.append((i+j)/21)

fxy = JPD(['X1', 'X2'],[3, 2], prob)
print(fxy)
```

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

```
Out      +-------+-------+------------+
         ¦ X1    ¦ X2    ¦   P(X1,X2) ¦
         +=======+=======+============+
         ¦ X1(0) ¦ X2(0) ¦     0.0952 ¦
         +-------+-------+------------+
         ¦ X1(0) ¦ X2(1) ¦     0.1429 ¦
         +-------+-------+------------+
         ¦ X1(1) ¦ X2(0) ¦     0.1429 ¦
         +-------+-------+------------+
         ¦ X1(1) ¦ X2(1) ¦     0.1905 ¦
         +-------+-------+------------+
         ¦ X1(2) ¦ X2(0) ¦     0.1905 ¦
         +-------+-------+------------+
         ¦ X1(2) ¦ X2(1) ¦     0.2381 ¦
         +-------+-------+------------+
```

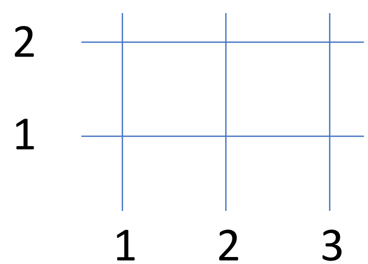
조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

$$f(x_1, x_2) = \frac{x_1 + x_2}{21}, x_1 = 1, 2, 3, \qquad x_2 = 1, 2$$

$x_2 = 1, 2$



$x_1 = 1, 2, 3$

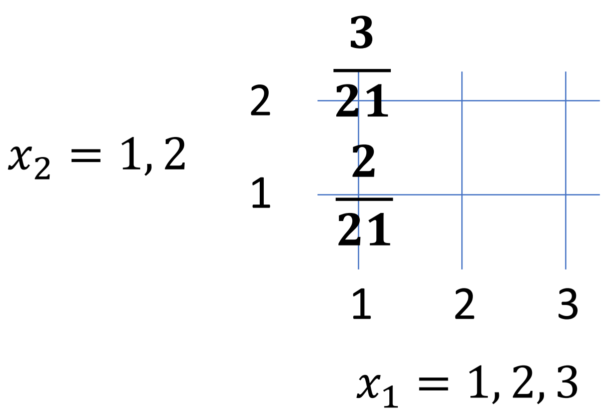
$\Pr(X_1 = 1)$=?

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

$$f(x_1, x_2) = \frac{x_1 + x_2}{21}, x_1 = 1, 2, 3, \qquad x_2 = 1, 2$$



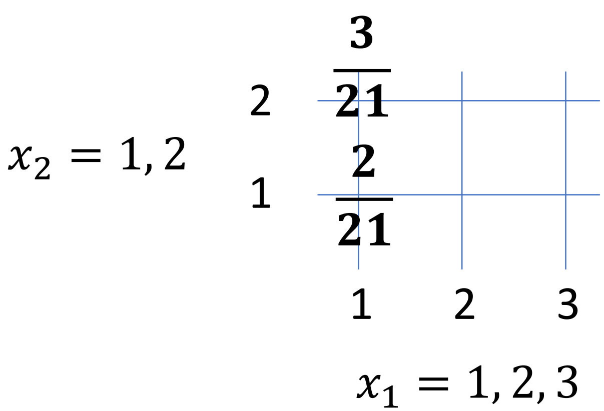
$$\Pr(X_1 = 1) = \frac{2}{21} + \frac{3}{21} = \frac{5}{21}$$

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

$$f(x_1, x_2) = \frac{x_1 + x_2}{21}, x_1 = 1, 2, 3, \qquad x_2 = 1, 2$$



$$\Pr(X_1 = 1) = f(1,1) + f(1,2)$$
$$= \sum_{x_2} f(1, x_2)$$
$$= f_1(1)$$

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수
- X_1 확률변수의 주변(marginal) 확률밀도함수

$$f_1(x_1) = \sum_{x_2} f(x_1, x_2) = f(x_1, 1) + f(x_1, 2) = \frac{x_1 + 1}{21} + \frac{x_1 + 2}{21} = \frac{2x_1 + 3}{21}$$

결합 확률 밀도 함수가 주어지면 주변 분포를 구할 수 있다.

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수
- X_1 확률변수의 주변(marginal) 확률밀도함수

In

```
# 주변 확률 구하기
fx = fxy.marginal_distribution(['X1'], inplace=False)
print(fx)
```

Out

```
+-------+---------+
¦ X1    ¦   P(X1) ¦
+=======+=========+
¦ X1(0) ¦  0.2381 ¦
+-------+---------+
¦ X1(1) ¦  0.3333 ¦
+-------+---------+
¦ X1(2) ¦  0.4286 ¦
+-------+---------+
```

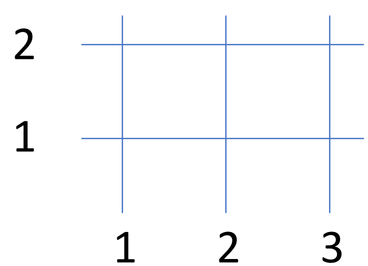
조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

$$f(x_1, x_2) = \frac{x_1 + x_2}{21}, x_1 = 1, 2, 3, \qquad x_2 = 1, 2$$

$x_2 = 1, 2$



$x_1 = 1, 2, 3$

$\Pr(X_2 | X_1 = 1) = ?$

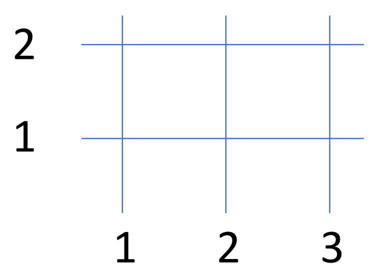
조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수

$$f(x_1, x_2) = \frac{x_1 + x_2}{21}, x_1 = 1, 2, 3, \qquad x_2 = 1, 2$$

$x_2 = 1, 2$



$x_1 = 1, 2, 3$

$$\Pr(X_2 | \mathrm{X}_1 = 1) = \frac{f(x_2, 1)}{f_1(1)}$$

$$= \frac{\frac{x_2 + 1}{21}}{\frac{5}{21}} = \frac{x_2 + 1}{5}$$

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수
- $\Pr(X_2|\mathrm{X}_1 = 1) = \frac{f(x_2, 1)}{f_1(1)}$

In
```
prob = fxy.conditional_distribution([('X1', 0)], inplac
e=False)
print(prob)
```

Out
```
+-------+---------+
¦ X2    ¦   P(X2) ¦
+=======+=========+
¦ X2(0) ¦  0.4000 ¦
+-------+---------+
¦ X2(1) ¦  0.6000 ¦
+-------+---------+
```

조건부 확률과 독립

[주변 분포(marginal distribution)와 조건부 분포(conditional distribution)]

- 2개의 확률변수 X_1, X_2에 대한 확률밀도함수를 $f(x_1, x_2)$
- 하나가 아니므로 결합(joint) 확률밀도함수
- $X_1 = x_1$ 일 때 X_2의 조건부 확률밀도함수

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}$$

조건부 확률과 독립

[**상관계수(correlation coefficient)**]

- X, Y의 2개의 확률변수의 결합확률밀도함수가 $f(x, y)$
- 주변 기대값과 분산은 $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$
- X, Y의 공분산(covariance)은 σ_{XY}

$$\sigma_{XY} = E(X - \mu_X)(Y - \mu_Y)$$

$$= E(XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y)$$

$$= E(XY) - \mu_Y E(X) - \mu_X E(X) + \mu_X \mu_Y$$

$$= E(XY) - \mu_X \mu_Y$$

조건부 확률과 독립

[상관계수(correlation coefficient)]

- X, Y의 2개의 확률변수의 결합확률밀도함수가 $f(x, y)$
- 주변 기대값과 분산은 $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$
- X, Y의 공분산(covariance)은 σ_{XY}
- 상관계수(correlation coefficient)는 ρ_{XY}

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

조건부 확률과 독립

[**상관계수(correlation coefficient)**]

예제) 다음의 분포에서 상관계수를 구하세요.

(x, y)	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)
$f(x, y)$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$

조건부 확률과 독립

[**상관계수(correlation coefficient)**]

예제) 다음의 분포에서 상관계수를 구하세요.

(x, y)	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)
$f(x, y)$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$

$$E(X) = \sum_x x \sum_y f(x, y) = 1\left(\frac{2}{15} + \frac{4}{15} + \frac{3}{15}\right) + 2\left(\frac{1}{15} + \frac{1}{15} + \frac{4}{15}\right) = \frac{21}{15}$$

$$E(X^2) = \sum_x x^2 \sum_y f(x, y) = 1\left(\frac{2}{15} + \frac{4}{15} + \frac{3}{15}\right) + 2^2\left(\frac{1}{15} + \frac{1}{15} + \frac{4}{15}\right) = \frac{33}{15}$$

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \frac{11}{5} - \left(\frac{7}{5}\right)^2 = \frac{6}{25}, \qquad \sigma_X = \frac{\sqrt{6}}{5}$$

조건부 확률과 독립

[상관계수(correlation coefficient)]

예제) 다음의 분포에서 상관계수를 구하세요.

(x, y)	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)
$f(x, y)$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$

$$E(Y) = \sum_y y \sum_x f(x, y) = 1\left(\frac{2}{15} + \frac{1}{15}\right) + 2\left(\frac{4}{15} + \frac{1}{15}\right) + 3\left(\frac{3}{15} + \frac{4}{15}\right) = \frac{34}{15}$$

$$E(Y^2) = \sum_y y^2 \sum_x f(x, y) = 1\left(\frac{2}{15} + \frac{1}{15}\right) + 2^2\left(\frac{4}{15} + \frac{1}{15}\right) + 3^2\left(\frac{3}{15} + \frac{4}{15}\right) = \frac{86}{15}$$

$$\sigma_Y^2 = E(Y^2) - \mu_Y^2 = \frac{86}{15} - \left(\frac{34}{15}\right)^2 = \frac{134}{225}, \sigma_Y = \frac{\sqrt{134}}{15}$$

조건부 확률과 독립

[**상관계수(correlation coefficient)**]

예제) 다음의 분포에서 상관계수를 구하세요.

(x, y)	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)
$f(x, y)$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$

$$E(XY) = \sum_x \sum_y xyf(x, y) = 1 \times 1 \times \frac{1}{15} + \cdots + 2 \times 3 \times \frac{4}{15} = \frac{49}{15}$$

$$\sigma_{XY} = \frac{49}{15} - \frac{21}{15} \times \frac{34}{15} = \frac{21}{225}$$

$$\rho_{XY} = \frac{\frac{21}{225}}{\frac{\sqrt{6}}{5} \times \frac{\sqrt{134}}{15}} = 0.247$$

조건부 확률과 독립

[상관계수(correlation coefficient)]

과제) 다음의 분포에서 상관계수를 프로그램으로 구하세요.

(x, y)	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)
$f(x, y)$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$

조건부 확률과 독립

[독립]

- 2개의 확률변수 X_1, X_2에 대한 결합확률밀도함수를 $f(x_1, x_2)$
- 조건부확률밀도함수를 이용한 결합확률밀도함수

$$f(x_1, x_2) = f(x_2|x_1)f_1(x_1)$$

조건부 확률과 독립

[독립]

- 2개의 확률변수 X_1, X_2에 대한 결합확률밀도함수를 $f(x_1, x_2)$
- 조건부확률밀도함수를 이용한 결합확률밀도함수
- $f(x_1, x_2) = f(x_2|x_1)f_1(x_1)$
- 조건부확률밀도함수 $f(x_2|x_1)$가 x_1에 의존하지 않을 때

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_2|x_1)f_1(x_1)\,dx_1 = f(x_2|x_1)\int_{-\infty}^{\infty} f_1(x_1)\,dx_1 = f(x_2|x_1)$$

따라서 결합확률밀도함수는 다음과 같이 정의된다.

$$f(x_1, x_2) = f_2(x_2)f_1(x_1)$$

조건부 확률과 독립

[독립]

- 2개의 확률변수 X_1, X_2가 서로 독립이기 위한 필요 충분 조건

$$f(x_1, x_2) = f_2(x_2) f_1(x_1)$$

조건부 확률과 독립

[독립]

예제) 다음과 같이 결합 확률밀도함수가 정의된 두개의 확률 변수 X_1, X_2에 대하여 독립 여부를 판정하세요.

$$f(x_1, x_2) = x_1 + x_2, \qquad 0 < x_1, x_2, < 1$$

조건부 확률과 독립

[**독립**]

예제) 다음과 같이 결합 확률밀도함수가 정의된 두개의 확률 변수 X_1, X_2에 대하여 독립 여부를 판정하세요.

$$f(x_1, x_2) = x_1 + x_2, \qquad 0 < x_1, x_2, < 1$$

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 = x_1 + \frac{1}{2}, \qquad f_2(x_2) = x_2 + \frac{1}{2}$$

$$\Rightarrow f(x_1, x_2) \neq f_1(x_1) f_2(x_2)$$

조건부 확률과 독립

[독립]

- 2개의 확률변수가 독립인 경우의 중요한 성질
- $u(X_1)$은 X_1만의 함수이고, $v(X_2)$는 X_2만의 함수일 때

$$E\big(u(X_1)v(X_2)\big) = E\big(u(X_1)\big)E\big(v(X_2)\big)$$

$$E\big((X-\mu_X)(Y-\mu_Y)\big) = E(X-\mu_X)E(Y-\mu_Y) = 0$$