3.2 조건부 확률분포

$f_Y(y) > 0$이고 결합질량함수 $f(x, y)$를 갖는 이산확률변수 X와 Y에 대하여 $Y = y$가 주어졌을 때 X의 확률분포를 조건부 확률분포(conditional probability distribution)라 한다. 그리고 이 조건부 확률에 대응하는 함수

$$f(x|y) = \frac{P(X = x,\ Y = y)}{P(Y = y)} = \frac{f(x, y)}{f_Y(y)},\ x \in S_X$$

를 $Y = y$일 때, X의 조건부 확률질량함수(conditional probability mass function)라 한다.

같은 방법으로 $f_X(x) = P(X = x) > 0$일 때, $X = x$가 주어졌다는 조건 아래서 Y의 조건부 확률질량함수를 다음과 같이 정의한다.

$$f(y|x) = \frac{P(X = x,\, Y = y)}{P(X = x)} = \frac{f(x, y)}{f_X(x)},\ y \in S_Y$$

(예제)

X : 하루 동안 판매한 승용차 수

Y : 특정한 옵션을 구비한 승용차를 판매한 수

Y \ X	0	1	2	$f_Y(y)$
0	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{3}$
1	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
2	0	0	$\frac{1}{6}$	$\frac{1}{6}$
$f_X(x)$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{7}{12}$	1

조건부 확률질량함수 $f(y|x)$

(1) 모든 y에 대하여 $f(y|x) > 0$

(2) $\sum_{\text{모든}\,y} f(y|x) = 1$

$f_Y(y) > 0$이고 결합밀도함수 $f(x, y)$를 갖는 연속확률변수 X와 Y에 대하여 $Y = y$가 주어졌다는 조건 아래서, X의 조건부 밀도함수(conditional probability density function)를

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

로 정의한다.

또한 $Y = y$가 주어졌다는 조건 아래서 X의 조건부 분포함수를 다음과 같이 정의한다.

$$F(x|y) = P(X \leq x \,|\, Y = y) = \begin{cases} \displaystyle\sum_{u \leq x} f(u|y) = \sum_{u \leq x} \frac{P(X = u,\, Y = y)}{P(Y = y)}, & \text{이산형인경우} \\ \displaystyle\int_{-\infty}^{x} f(u|y)\,du = \int_{-\infty}^{x} \frac{f(u, y)}{f_Y(y)}\,du & \text{연속형인경우} \end{cases}$$

따라서 확률변수 X와 Y에 댕여 $Y=y$가 주어졌다는 조건 아래서 $a<X\le b$일 조건부 확률은 다음과 같다.

$$P(a<X\le b|Y=y)=\begin{cases}\displaystyle\sum_{a<x\le b}f(u|y)=\sum_{a<x\le b}\frac{P(X=x,\ Y=y)}{P(Y=y)}, & \text{이산형인경우}\\ \displaystyle\int_a^b f(x|y)dx=\int_a^b\frac{f(x,y)}{f_Y(y)}dx & \text{연속형인 경우}\end{cases}$$

조건부 분포함수를 이용하여 다음과 같이 조건부 확률을 구할 수 있다.

$$P(a<X\le b|Y=y)=F(b|Y=y)-F(a|Y=y)$$

확률변수의 독립(independent)

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = f_Y(y) \; ; \; f(x, y) = f_X(x) f_Y(y)$$

확률변수 X, Y, Z의 확률밀도함수를 각각 $f_X(x), f_Y(y), f_Z(z)$라 할 때, 결합 확률밀도함수 $f(x, y, z)$에 대하여

$$f(x, y, z) = f_X(x) f_Y(y) f_Z(z)$$

가 성립하면 확률변수 X, Y, Z는 독립이라 한다.

세 확률변수 X, Y, Z 중에서 어느 두 확률변수를 택해도 독립이면, 즉

$$f(x, y) = f_X(x) f_Y(y),\ f(x, z) = f_X(x) f_Z(z),\ f(y, z) = f_Y(y) f_Z(z)$$

이면, 확률변수 X, Y, Z는 쌍마다 독립(pairwisely independent)이라 한다.

정리1 다음 (1)~(3)은 동치이다.

(1) X와 Y가 독립이다.

(2) 모든 실수 x, y에 대하여 $F(x, y) = F_X(x) F_Y(y)$

(3) $P(a < X \leq b, c < Y \leq d) = P(a < X \leq b) P(c < Y \leq d)$

두 확률변수 X와 Y가 항등적인 확률분포를 가지는 경우, 즉 모든 실수 x에 대하여

$$f_X(x) = f_Y(x)$$

이면, X와 Y는 항등분포(identically distributed)를 이룬다고 한다.

두 개 이상의 확률변수가 독립이고 항등적인 분포(independent identically distributed; i.i.d.)를 이루는 경우에 확률변수 X와 Y는 독립인 항등분포를 이룬다고 하고 간단히 **i.i.d. 확률변수**라 한다.