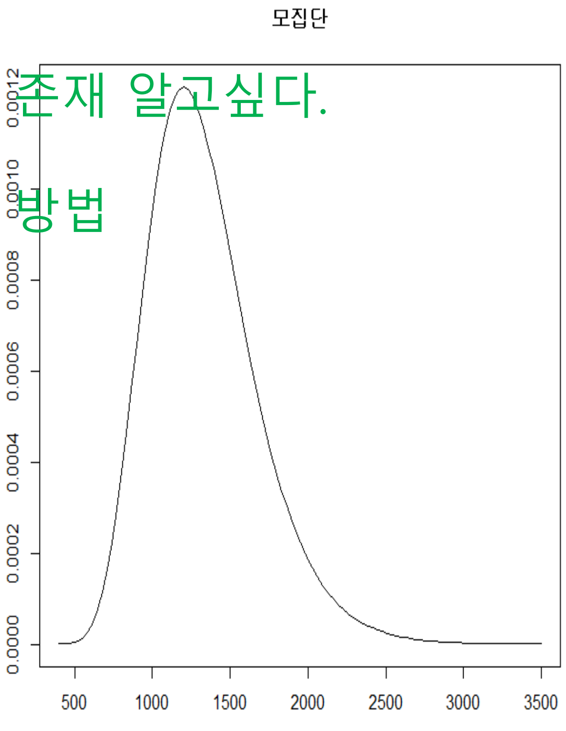
컴퓨터 응용통계

4.4,4.5,4.6,4.7 확률과 확률변수

최경미

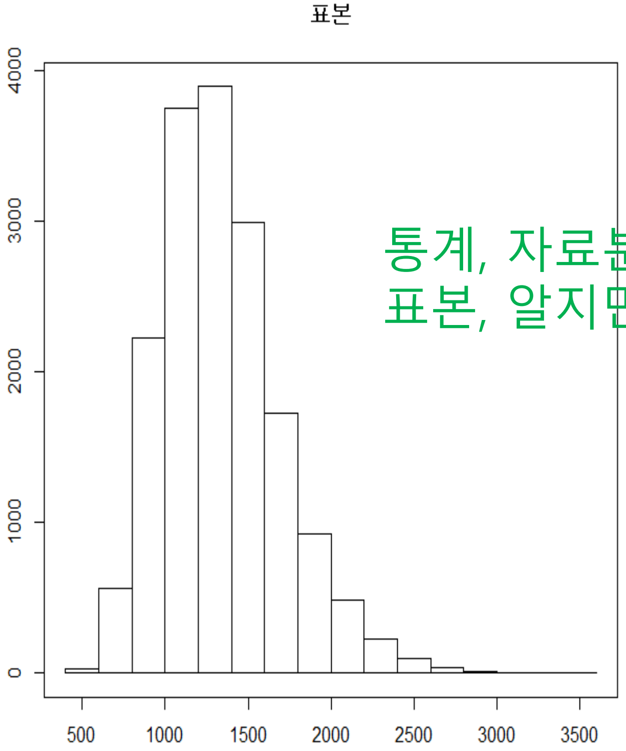
확률과 통계

목표: 모집단 존재 알고싶다.
현실: 모름
확률, 이론적 방법



표본추출

추론



통계, 자료분석
표본, 알지만 변함

모집단과 표본은 쌍둥이, 거울 처럼 존재 한다.

4.4 확률변수와 확률함수

동전 던지기 실험

확률변수(random variable) X

$X(\text{앞면}) = 1,\ X(\text{뒷면}) = 0$

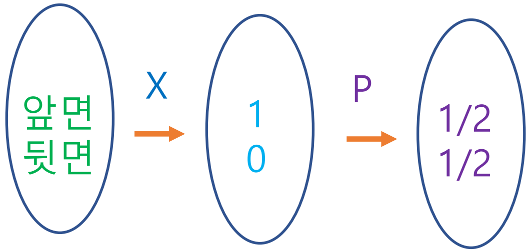
$P(X=1) = P(\text{앞면}) = \frac{1}{2},\ P(X=0) = P(\text{뒷면}) = \frac{1}{2}$

확률분포

분포표(distribution)

동전	X=x	P(X=x)
T	0	1/2
H	1	1/2

합=1



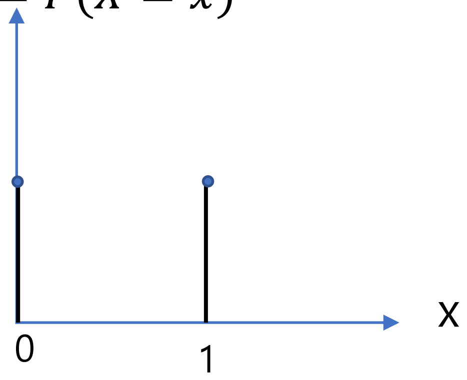
확률변수 값

확률함수(probability density function) $f_X(x) = P(X = x)$

$f_X(1) = P(X=1) = 1/2$

$f_X(0) = P(X=0) = 1/2$

$f_X(x) = P(X = x)$



정의 4.6 확률변수 X는 표본공간 S에서 정의된 실수 함수이다.

모든 확률변수는 고유의 확률함수를 가진다.
확률함수는 확률변수를 유일하게 결정한다.

확률변수의 종류

- 이산확률변수

✓ 확률변수가 가지는 값이 **셀 수 있는 경우**
✓ 실험의 결과가 정성적인 경우
✓ 범주형 자료를 표현할 수 있다

- 연속확률변수

✓ 확률변수가 가지는 값이 **셀 수 없는 경우**

정의 4.7 이산확률변수 X의 확률함수

확률함수 $f(x) = P(X = x)$는 다음의 두 가지 성질을 만족해야 한다.

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$

- 사건 A의 확률은 A의 모든 원소에 해당하는 확률을 더하여 계산할 수 있다.

$$P(X \in A) = \sum_{x \in \mathrm{A}} f(x)$$

예제 4.11 동전 두 개를 던져서 나오는 앞면의 수를 X라고 두자.

표본공간

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

앞면의 수 X

$$X(TT) = 0,\ X(HT) = X(TH) = 1,\ X(HH) = 2$$

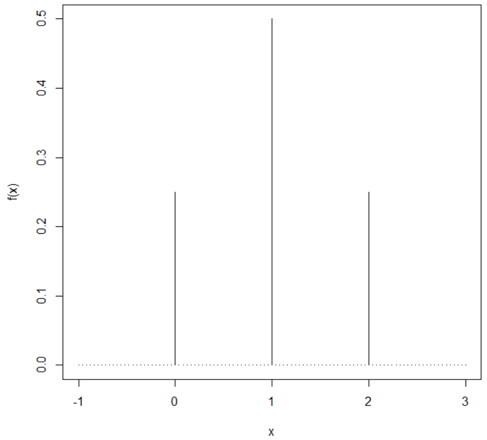
X의 확률

$$P(X=0) = P(TT) = \frac{1}{4},\ P(X=1) = P(\{TH, HT\}) = \frac{1}{2},\ P(X=2) = P(HH) = \frac{1}{4}$$

X의 분포

두 동전의 앞면수

x	0	1	2	총합
$P(X=x)$	1/4	1/2	1/4	1



확률함수 $f(x) = P(X = x)$

$$f_X(0) = \frac{1}{4}, \qquad f_X(1) = \frac{1}{2}, \qquad f_X(2) = \frac{1}{4}$$

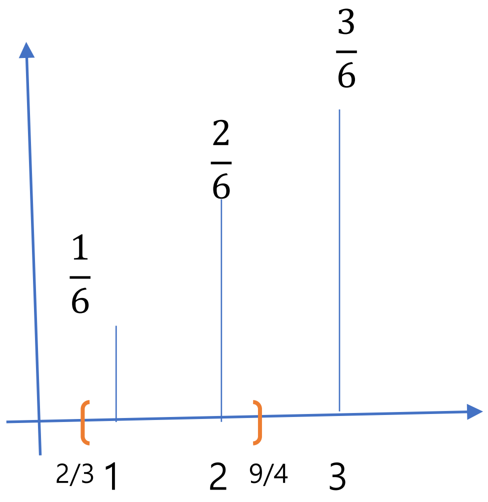
예제 4.12 이산확률변수 X에 대하여 확률함수가 $f(x) = \frac{x}{6}$, for $x = 1,2,3$이고, 나머지 경우에는0으로 주어질 때, $X = 3$일 확률과 $\frac{2}{3} < X < \frac{9}{4}$일 확률을 계산해보자.

$$\mathrm{P}(X = 1) = f(1) = \frac{1}{6}$$

$$\mathrm{P}(X = 2) = f(2) = \frac{2}{6} = \frac{1}{3}$$

$$\mathrm{P}(X = 3) = f(3) = \frac{3}{6} = \frac{1}{2}$$

$$\mathrm{P}\left(\frac{2}{3} < \mathrm{X} < \frac{9}{4}\right) = \mathrm{P}(\mathrm{X} = 1) + \mathrm{P}(\mathrm{X} = 2) = \mathrm{f}(1) + \mathrm{f}(2) = \frac{1}{6} + \frac{2}{6} = \frac{3}{6} = \frac{1}{2}$$

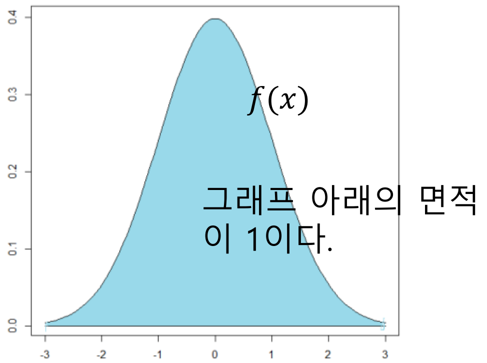


정의 4.7 연속확률변수 X의 확률(밀도)함수 (probability density function; pdf)

확률함수 $f(x)$는 다음의 두 가지 성질을 만족해야 한다.

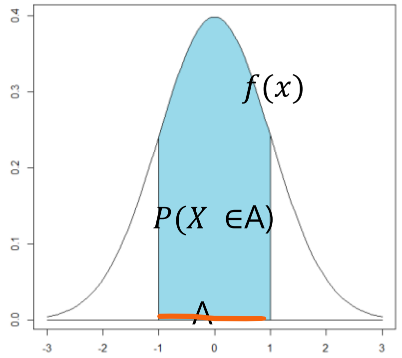
1. $f(x) \geq 0$

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

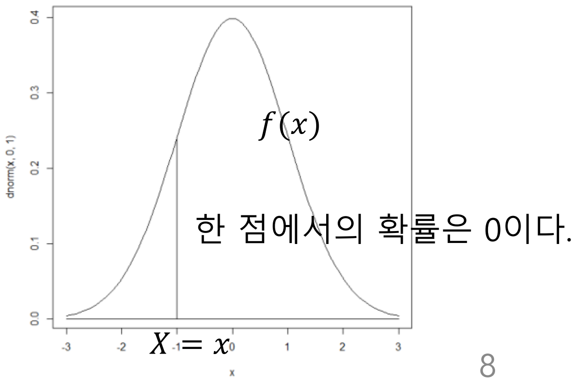


- $f(x) \geq 0$일때, 적분은 그래프 아래의 면적을 나타낸다.
- 사건 A에 대한 확률은 A에서 확률함수 $f(x)$를 적분하여 계산할 수 있다.

  $P(X \in A) = \int_A f(x)\ dx = A$에서 f 아래의 면적



- 연속형 확률함수 $f(x)$ 아래의 전체 면적은 1이 된다.
- 한 점에서의 면적이 0이므로 한 점에서의 확률은 $P(X = x) = 0$이다.
- $P(X \leq x) = P(X < x)$
- 분포함수 (cumulative distribution function; cdf)는 $P(X \leq x) = \int_{-\infty}^{x} f(v)\,dv$로 정의된다.



예제 4.13 X의 확률함수가 $f(x) = cx,\ 0 < x < 1, c > 0$이라고 두자. c를 구하고,X가 $\left[\frac{1}{3}, \frac{1}{2}\right]$에 속하는 확률을 계산해보자.

- 삼각형의 면적은 1이다.

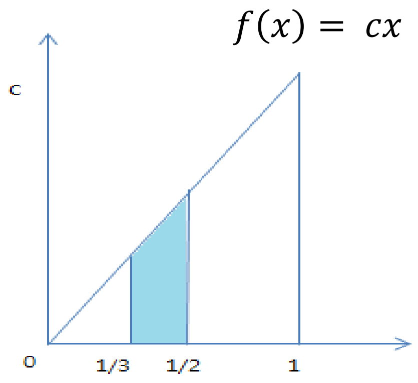
$$\frac{1}{2} \times 1 \times c = \frac{c}{2} = 1$$

$$c = 2$$

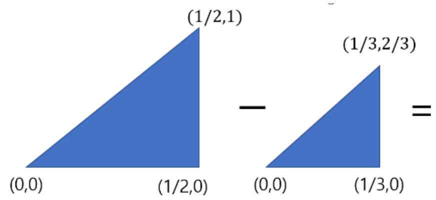
- 적분을 이용하여 삼각형의 면적을 구하자.

$$\int_0^1 cx\,dx = \left[\frac{c}{2}\,x^2\right]_0^1 = \frac{c}{2} = 1$$

따라서 $c = 2$이다.



- 그림 4.5는 $f(x)$를 나타낸다. X가 $\left[\frac{1}{3}, \frac{1}{2}\right]$에 속하는 확률은 이 구간에서 $f(x)$ 아래의 사다리꼴 면적이다.



$$= \frac{1}{4} - \frac{1}{9} = \frac{5}{36}$$

$$P\left(\frac{1}{3} \le X \le \frac{1}{2}\right) = \int_{1/3}^{1/2} 2x\,dx = [\,x^2]_{1/3}^{1/2} = \frac{1}{4} - \frac{1}{9} = \frac{5}{36}$$

4.5 평균과 분산

동전을 한 번 던지면 앞면이 몇 번 나올까?

앞면 나오면 상금 1원, 뒷면 나오면 상금 0원일 때,

기대되는 상금은 얼마일까?

 **또는**  반반?

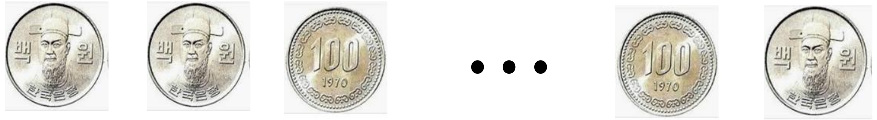
상금 X

$X = 1$ (앞면, 상금 1원), $X = 0$ (뒷면, 상금 0원)

$P(X = 1) = P(X = 0) = \frac{1}{2}$

$1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$ =0.5원. (이런 돈이 없음)

동전을 100번 던지면 앞면이 몇 번 나올까?



$X_1 = 1$ (앞면), $X_2 = 1$ (앞면), $X_3 = 0$ (뒷면), ..., $X_{99} = 0, X_{100} = 1$

상금의 기대값? 대충 50원?

$P(X_i = 1) = (X_i = 0) = \frac{1}{2}$, $i = 1,2,3, \dots, 100$

$100 \times \left(1 \times \frac{1}{2} + 0 \times \frac{1}{2}\right) = 50$원 (이건 말이 됨)

평균은 이론적인 값이다.

기대값 (expected value) 또는 평균 (mean)

정의 4.9 확률변수 X의 평균(mean) 또는 기대값(Expected value)

소문자

$$\text{모평균 } \mu = E[X] = \begin{cases} \sum_{x} xf(x) & (\text{이산인 경우}) \\ \int_{-\infty}^{\infty} xf(x)\,\mathrm{dx} & (\text{연속인 경우}) \end{cases}$$

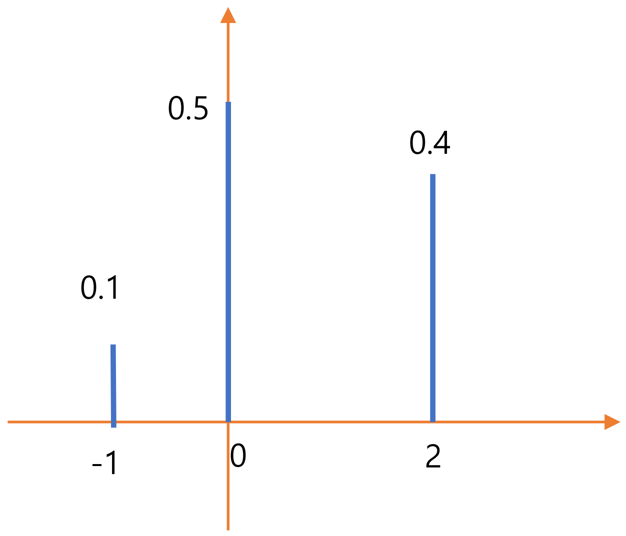
대문자

정리 4.7 확률변수 X의 확률함수가 $f(x)$일 때, $g(X)$의 기대값은 다음과 같다.

$$E[g(X)] = \begin{cases} \sum_{x} g(x)f(x) & (\text{이산인 경우}) \\ \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{dx} & (\text{연속인 경우}) \end{cases}$$

예제 4.14 확률변수 X의 확률함수가 표 4.3과 같이 주어졌다. $Y = X^2$의 기대값을 구하자.

$X = x$	**-1**	**0**	**2**
$p(x)$	0.1	0.5	0.4



X의 기대값 = X의 평균

$$\mu = E[X] = \sum xf(x) = (-1)(0.1) + (0)(0.5) + (2)(0.4) = 0.7$$

$Y = X^2$의 기대값

$$E[Y] = E[X^2] = \sum_x \mathrm{x}^2 f(x) = (-1)^2(0.1) + (0^2)(0.5) + (2^2)(0.4) = 1.7$$

기대값의 선형성

정의: $aX + bY$는 X와 Y의 선형결합이다.

정리 4.8 확률변수 X와 상수 a에 대하여 다음이 성립한다.

① $E[a] = a$ (상수의 기대값은 상수이다.)

증명

$$E[a] = \sum_x af(x) = a\sum_x f(x) = a$$

② $E[a + bX] = a + bE[X]$

(기대값의 선형성이 성립한다.)

(선형결합의 기대값은 기대값의 선형결합이다.)

증명

$$E[a + bX]$$

$$= \sum_x (a + bx)f(x)$$

$$= a\sum_x f(x) + b\sum_x xf(x)$$

$$= a + bE[X]$$

분산(Varinace) σ^2

모집단의 분포에서 중심 μ로부터 자료가 흩어진 정도를 표현

정의 4.10 확률변수 X와 확률함수 $f(x,y)$에 대하여 분산(variance)은 다음과 같이 정의된다.

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 f(x) & \text{(이산인 경우)} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx & \text{(연속인 경우)} \end{cases}$$ 단위 제곱

- $\mu = E[X]$ 상수
- $\sigma^2 = E[(X-\mu)^2]$ (정의, 편차제곱의 기대값)

$= E[X^2 - 2\,\mu\, X + \mu^2]$

$E[X] = \mu,\ E[\mu^2] = \mu^2$

$= E[X^2] - 2\,\mu\, E[X] + E[\mu^2]$ (선형성)

$= E[X^2] - \mu^2$

$= E[X^2] - (E[X])^2$

= 제곱의 기대값 – 기대값의 제곱

편차 $= X - \mu$

$E[X-\mu] = E[X] - E[\mu]$

$= \mu - \mu$

$= 0$

표준편차(standard deviation) σ 는 $\sqrt{\sigma^2}$으로 정의된다. 단위에 제곱 없음

분산의 성질

정리 4.9 확률변수 X와 상수 a에 대하여 다음이 성립한다.

① $Var(a) = 0$

② $Var(aX + b) = a^2 Var(X)$

③ 표준화 변수 $Z = \frac{X-\mu}{\sigma}$에 대하여, $E[Z] = 0$, $Var(Z) = 1$이다.

여기서, $\mu = E[X]$, $\sigma^2 = Var(X)$이다.

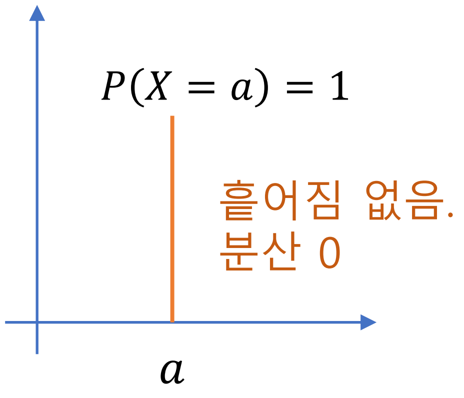
(증명)

①

$X = a$

$E[X] = E[a] = a$

$E[X^2] = E[a^2] = a^2$



$\therefore Var(X) = Var(a) = E[X^2] - (E[X])^2$

$= E[a^2] - (E[a])^2 = a^2 - a^2 = 0$

②

$Var(aX + b)$

$= E\left[\left((aX + \cancel{b}) - (a\mu + \cancel{b})\right)^2\right]$

$= E[a^2(X - \mu)^2]$

$= a^2 Var(X)$ (기대값의 선형성)

③

$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{E[X]-\mu}{\sigma} = 0$

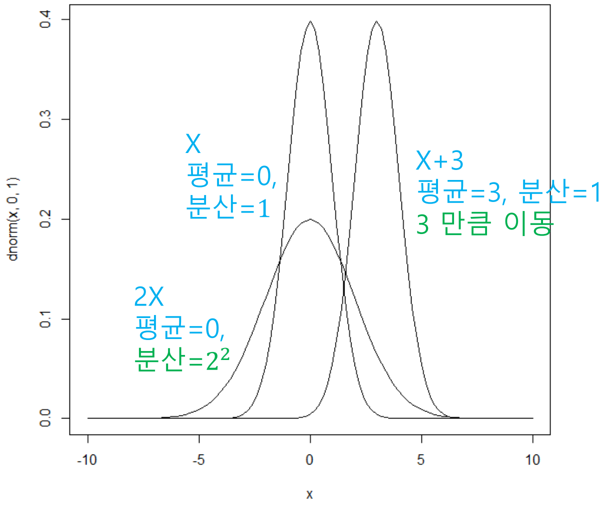
$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right)$

$= Var\left(\frac{1}{\sigma}X + \left(-\frac{\mu}{\sigma}\right)\right)$

$= \left(\frac{1}{\sigma}\right)^2 Var(X)$

$= \left(\frac{1}{\sigma}\right)^2 \sigma^2$

$= 1$



예제 $\sigma^2 = Var(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2 = E[X^2] - \mu^2$

$X = x$	-1	0	2
$p(x)$	0.1	0.5	0.4

예제 4.15확률변수 X의 값에 대한 확률함수가 표4.3과 같이 주어지면, 평균과 분산을 다음과 같이 구할 수 있다.

$$\mu = E[X] = (-1)(0.1) + (0)(0.5) + (2)(0.4) = 0.7$$

$$E[Y] = E[X^2] = \sum_x x^2 f(x) = (-1)^2(0.1) + (0^2)(0.5) + (2^2)(0.4) = 1.7$$

$$\sigma^2 = E[X^2] - \mu^2 = 1.7 - 0.49 = 1.21$$

$$\sigma = \sqrt{1.21} = 1.1$$

예제

예제 4.16 $E[X] = \mu$, $Var(X) = \sigma^2$,$Z = \frac{X-\mu}{\sigma}$, $T = 5Z + 1$라고 두자.

$$E[T] = E[5Z + 1] = 5E[Z] + 1 = 1$$

$$Var(T) = Var(5Z + 1) = Var(5Z) = 5^2 Var(Z) = 25$$

혼자 풀기

1. X의 확률함수가 $f(x) = c, \quad 0 < x < 2, c > 0$이라고 두자. c를 구하고,X가 $\left[\frac{1}{3}, \frac{1}{2}\right]$에 속하는 확률을 계산해보자.

2. $\mathrm{P}(\mathrm{X} = -3) = 0.3, \mathrm{P}(\mathrm{X} = 1) = 0.3, \mathrm{P}(\mathrm{X} = 2) = 0.4$ 일 때, X의 평균 μ와 분산 σ^2은 무엇인가?

3. $\mathrm{P}(\mathrm{X} = -3) = 0.3, \mathrm{P}(\mathrm{X} = 1) = 0.3, \mathrm{P}(\mathrm{X} = 2) = 0.4$ 이고, $Z = \frac{X-\mu}{\sigma}$, $T = 2Z - 1$라고 두자. $E[T]$와 $Var(T)$를 구하라.

4.6 이변수에 대한 결합확률함수

정의 4.11 함수 $f(x,y)$가 두 이산확률변수 X와 Y의 결합확률함수(Joint Probability)가 되려면 다음의 두 가지를 만족해야 한다.

$$f(x,y) = P(X=x, Y=y) \geq 0$$

$$f(x,y) = P(X=x, Y=y) = (\{X=x\} \cap \{Y=y\})$$

$$\sum_x \sum_y f(x,y) = 1$$

	y_1	y_2		y_n	$f_X(x)$
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	$\cdots$	$f(x_1,y_n)$	$f_X(x_1)$
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	$\cdots$	$f(x_2,y_n)$	$f_X(x_2)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
x_m	$f(x_m,y_1)$	$f(x_m,y_2)$	$\cdots$	$f(x_m,y_n)$	$f_X(x_m)$
$f_Y(y)$	$f_Y(y_1)$	$f_Y(y_2)$	$\cdots$	$f_Y(y_n)$	1

X의 주변확률함수

X, Y의 결합확률함수

Y의 주변확률함수

이때, 사건 A의 확률 $P(A)$는 다음과 같이 얻어진다.

$$P((X,Y) \in A) = \sum_{(x,y) \in A} \sum f(x,y)$$

두 이산확률변수 X와 Y의 주변확률함수(marginal pdf)인 $f_X(x)$과 $f_Y(y)$는 다음과 같이 정의된다.

$$f_X(x) = \mathrm{P}(\mathrm{X}=\mathrm{x}) = \sum_y \mathrm{P}(\mathrm{X}=\mathrm{x}, \mathrm{Y}=\mathrm{y}) = \sum_y \mathrm{f}(\mathrm{x},\mathrm{y})$$

$$f_Y(y) = P(Y=y) = \sum_x \mathrm{P}(\mathrm{X}=\mathrm{x}, \mathrm{Y}=\mathrm{y}) = \sum_x \mathrm{f}(\mathrm{x},\mathrm{y})$$

$$\begin{aligned} f_X(x_1) &= \mathrm{P}(\mathrm{X}=\mathrm{x}_1) \\ &= \sum_y \mathrm{P}(\mathrm{X}=\mathrm{x}_1, \mathrm{Y}=\mathrm{y}) \\ &= \sum_y \mathrm{f}(\mathrm{x}_1,\mathrm{y}) \\ &= f(x_1,y_1) + f(x_1,y_2) + \ldots + f(x_1,y_n) \end{aligned}$$

예제 4.17 동전 한 개를 두 번 던지는 실험에서,
X = 첫 번째 던졌을 나오는 앞면의 수
Y = 두 번 던져서 나오는 총 앞면의 수라고 두자.

$Y(TT) = 0$

$Y(HT) = Y(TH) = 1$

$Y(HH) = 2$

$f_Y(0) = \frac{1}{4},\quad f_Y(1) = \frac{1}{2},\quad f_Y(2) = \frac{1}{4}$ (Y의 주변확률함수)

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

$X(TT) = X(TH) = 0$

$X(HT) = X(HH) = 1$

$f_X(0) = \frac{1}{2}, f_X(1) = \frac{1}{2}$ (X의 주변확률함수)

X / Y	0	1	2	$f_X(x)$
0	TT (1/4)	TH (1/4)	0	$f_X(0) = 1/2$
1	0	HT (1/4)	HH (1/4)	$f_X(1) = 1/2$
$f_Y(y)$	$f_Y(0) = 1/4$	$f_Y(1) = 1/2$	$f_Y(2) = 1/4$	1

X의 주변확률함수

Y의 주변확률함수

① $P(X \geq 1, Y \geq 1) = f(1,1) + f(1,2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

② $P(X < Y) = f(0,1) + f(0,2) + f(1,2) = \frac{1}{2}$

독립

정리 4.10 두 확률변수 X와 Y가 독립이기 위한 필요충분 조건은 다음과 같다.

$$f(x,y) \overset{독립}{=} f_X(x)\ f_Y(y)$$

증명.

확률변수 X와 Y가 독립

$$P(X = x, Y = y) = P(\{X = x\} \cap \{Y = y\}) \overset{독립}{=} P(\{X = x\})\ \mathrm{P}(\{Y = y\}) = P(X = x)P(Y = y)$$

$$f(x,y) = P(X = x, Y = y) \overset{독립}{=} P(X = x)P(Y = y) = f_X(x)f_Y(y).$$

Note. A, B 독립 $\Leftrightarrow P(A \cap B) = P(A)P(B)$

예제 4.17 동전 한 개를 두 번 던지는 실험에서,

X = 첫 번째 던졌을 나오는 앞면의 수
Y = 두 번 던져서 나오는 총 앞면의 수
라고 두면, X와 Y 는 독립인가?

$S = \{HH, HT, TH, TT\}$
$X(TT) = X(TH) = 0$
$X(HT) = X(HH) = 1$

$f_X(0) = \frac{1}{2}, f_X(1) = \frac{1}{2}$ (X의 주변확률함수)

X / Y	Y=0	Y=1	Y=2	$f_X(x)$
X=0	TT (1/4)	TH (1/4)	0	1/2
X=1	0	HT (1/4)	HH (1/4)	1/2
$f_Y(y)$	1/4	1/2	1/4	1

X의 주변확률함수

Y의 주변확률함수

$Y(TT) = 0$

$Y(HT) = Y(TH) = 1$

$Y(HH) = 2$

$f_Y(0) = \frac{1}{4}, \quad f_Y(1) = \frac{1}{2}, \quad f_Y(2) = \frac{1}{4}$ (Y의 주변확률함수)

① $P(X \geq 1, Y \geq 1) = f(1,1) + f(1,2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

② $P(X < Y) = f(0,1) + f(0,2) + f(1,2) = \frac{1}{2}$

③ $P(X = 0, Y = 0) = \frac{1}{4}$

$P(X = 0)P(Y = 0) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$

$P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0)$

따라서 두 확률변수는 독립이 아니다.

여섯 개 셀 중에 두 개만 등호 성립, 네 개에서는 등화가 성립하지 않는다.

결합확률변수의 기대값

정의 4.13 두 확률변수 X와 Y와 이들의 결합확률함수 $f(x,y)$에 대하여 $r(X,Y)$의 기대값은 다음과 같이 정의된다.

$$E[r(X,Y)] = \begin{cases} \displaystyle\sum_x \sum_y r(x,y)f(x,y) & (\text{이산형}) \\ \displaystyle\iint r(x,y)f(x,y)\,dx\,dy & (\text{연속형}) \end{cases}$$

기대값의 선형성

정리 4.11 결합확률함수 $f(x,y)$를 갖는 두 확률변수 X와 Y에 대하여 다음 성질이 성립한다.

① $g(X,Y)$와 h(X,Y)의 선형결합의 기대값은 각각 기대값의 선형결합이다.

$$E[ag(X,Y) + bh(X,Y)] = aE[g(X,Y)] + bE[h(X,Y)]$$

선형결합의 기대값은 기대값의 선형결합이다.

X만의 함수 Y만의 함수

② 두 확률변수 X와 Y가 독립이면 $g(X)$와 $h(Y)$의 곱의 기대값은 각각 기대값의 곱이다.

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

(증명) ① $E[ag(X,Y) + bh(X,Y)]$

$$= \sum_x \sum_y \big(ag(x,y) + bh(x,y)\big) f(x,y)$$

$$= a \sum_x \sum_y g(x,y) f(x,y) + \sum_x \sum_y h(x,y) f(x,y)$$

$$= aE[g(X,Y)] + bE[h(X,Y)]$$

②

$E[g(X)h(Y)]$

$$= \sum_x \sum_y \big(g(x)h(y)\big) f(x,y)$$

$$\overset{독립}{=} \sum_x \sum_y g(x)h(y) f_X(x) f_Y(y)$$

$$= \sum_x g(x) f_X(x) \sum_y g(y) f_Y(y)$$

$$= E[g(X)]E[h(Y)]$$

예제

$E[X] = -2,\ E[X^2] = 1, E[Y] = -1, Var(Y) = 3$ 이고, X 와 Y는 독립이라고 두자. $E[(2X-Y)(X+Y)]$의 값은 무엇인가?

(풀이)

$E[(2X-Y)(X+Y)]$

$= 2E[X^2] + E[XY] - E[Y^2]$ (선형성)

$= 2E[X^2] + E[X]E[Y] - (\ Var(Y) +\ (E[Y])^2)$ (독립, 분산의 정의)

$= 2 + (-2)(-1) - (3 + (-1)^2)$

$= 0$

혼자 풀기

$E[X] = -1,\ E[X^2] = 3, E[Y] = -1, Var(Y) = 4$이고, X와 Y는 독립이라고 두자. $E[(X-Y)(X+Y)]$의 값은 무엇인가?

4.7 상관계수

정의 4.14 확률변수 X와 Y의 공분산(covariance)은 다음과 같이 정의된다.

$Cov(X,Y)$

$= E[\,(X - E[X])(Y - E[Y])\,]$ (정의)

$= E[XY - \underbrace{E[X]}_{\text{상수}}Y - X\underbrace{E[Y]}_{\text{상수}} + \underbrace{E[X]}_{\text{상수}}\underbrace{E[Y]}_{\text{상수}}]$ (E[X],E[Y]는 상수)

$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$

$= E[XY] - E[X]E[Y]$

= 곱의 기대값 – 기대값의 곱

- 만약 X가 증가할 때 Y도 증가하면, 둘은 양의 공분산을 갖는다.
- 만약 X가 증가할 때 Y가 감소하면, 음의 공분산을 갖는다.
- 만약 확률변수 X와 Y가 독립이면 $E[XY] = E[X]E[Y]$이므로, $Cov(X,Y) = 0$이 된다.
- 역은 성립하지 않아서, $Cov(X,Y) = 0$이더라도, X와 Y가 독립이 아닐 수 있다.

확률변수 X와 Y의 합의 분산

정리 4.12 확률변수 X와 Y의 합의 분산은 다음과 같다.

$Var(X+Y) \ = \ Var(X)+Var(Y)+2Cov(X,Y)$

(증명)

$Var(X+Y)$

$= E[(X+Y-E[X+Y])^2]$

$= E[(X-E[X]+Y-E[Y])^2]$ $\quad (E[X+Y]=E[X]+E[Y])$

$= E[(X-E[X])^2+(Y-E[Y])^2+2(X-E[X])(Y-E[Y])]$

$= E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2E[(X-E[X])((Y-E[Y])]$

$= Var(X) + Var(Y) + 2\,Cov(X,Y)$

정의 $Var(X)$=$E[(X-E[X])^2]$

정의 $Cov(X,Y)$=$E[(X-E[X])((Y-E[Y])]$

- 두 확률변수가 독립이면, 합의 분산은 분산의 합이 된다.

$Var(X+Y) \ = \ Var(X)+Var(Y)$

상관계수(correlation coefficient)

정의 4.15 확률변수 X와 Y의 상관계수(correlation coefficient)는 다음과 같이 정의된다.

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \qquad -1 \le \rho \le 1$$

여기서 ρ는 변수의 단위에 의존하지 않으며, 두 확률변수가 독립이면 $\rho = 0$이다.

예제 4.19 동전 한 개를 두 번 던지는 실험에서
X = 첫 번째 던졌을 나오는 앞면의 수
Y = 두 번 던져서 나오는 총 앞면의 수
일 때, 두 확률변수의 상관계수를 구해보자.

X / Y	0	1	2	$f_X(x)$
0	TT (1/4)	TH (1/4)	0	$f_X(0) = 1/2$
1	0	HT (1/4)	HH (1/4)	$f_X(1) = 1/2$
$f_Y(y)$	$f_Y(0) = 1/4$	$f_Y(1) = 1/2$	$f_Y(2) = 1/4$	1


$$\mu_X = E[X] = \sum_x x f_X(x) = (0)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{2}\right) = \frac{1}{2}$$

$$E[X^2] = \sum_x x^2 f_X(x) = (0^2)\left(\frac{1}{2}\right) + (1^2)\left(\frac{1}{2}\right) = \frac{1}{2}$$

$$\sigma_X^2 = Var(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$$\mu_Y = E[Y] = \sum_y y f_Y(y) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) = 1$$

$$E[Y^2] = \sum_y y^2 f_Y(y) = (0^2)\left(\frac{1}{4}\right) + (1^2)\left(\frac{1}{2}\right) + (2^2)\left(\frac{1}{4}\right) = \frac{3}{2}$$

$$\sigma_Y^2 = Var(Y) = E[Y^2] - (E[Y])^2 = \frac{3}{2} - 1^2 = \frac{1}{2}$$

$$E[XY] = \sum_x \sum_y xy f_{X,Y}(x,y)$$
$$= (0)(0)\left(\frac{1}{4}\right) + (0)(1)\left(\frac{1}{4}\right) + (0)(2)(0)$$
$$+(1)(0)(0) + (1)(1)\left(\frac{1}{4}\right) + (1)(2)\left(\frac{1}{4}\right) = \frac{3}{4}$$

$$Cov(X,Y) = E[XY] - E[X]E[Y] = \frac{3}{4} - \left(\frac{1}{2}\right)(1) = \frac{1}{4}$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{4}}{\sqrt{\frac{1}{4}}\sqrt{\frac{1}{2}}} = \frac{\sqrt{2}}{2} = 0.707$$

혼자 풀기

6. $P(X = 0, Y = -1\,) = \frac{1}{4}$, $P(X = 1, Y = -1\,) = \frac{1}{4}$, $P(X = 1, Y = 0) = \frac{1}{2}$이다. 다음 중 옳은 것을 모두 고르시오.

a. $E[X] = -\frac{1}{2}, Var(X) = \frac{1}{4}$

b. $E[Y] = \frac{3}{4}, Var(Y) = \frac{3}{16}$

c. $Cov(X, Y) = \frac{1}{8}$

d. $\rho = \frac{1}{\sqrt{3}}$

1 a b   2 b c   3 b d   4 c d

5 위 보기 중 답 없음

과제

- 연습문제 9,10,16,17