이공학도를 위한
확률과 통계학

- 4장-

Ki woon Moon

Protocol **E**ngineering **L**ab. **Sangmyung** University

2016. 1. 15

Content

- 확률변수의 기대값
- 확률변수의 분산
- 확률변수의 공분산
- 상관계수
- 체비세프 정리

확률변수의 기대값

X가 확률분포 $f(x)$ 를 가지는 확률 변수라 할 때 **X**의 평균 혹은 기대값 $E(x)$는**?**

X가 이산형인 경우

$$\mu = E(X) = \sum_{x} x f(x)$$

X가 연속형인 경우

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

확률변수의 기대값

예제 – 이산형인 경우

7개의 부품으로 구성되어 있는 로트를 검사하려고 할 때**,** 이 로트에 **4**개의 양호한 부품과 **3**개의 결함이 있는 부품이 들어있다고 하면**,** 검사원이 **3**개의 부품을 추출하였을 때 나타나는 양호한 제품의 평균 개수를 구하라**.**

X = 추출된 표본에 포함되는 양호한 부품 수

X의 확률 분포

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}$$

확률변수의 기대값

예제 – 이산형인 경우

X = 추출된 표본에 포함되는 양호한 부품 수

X의 확률 분포

$$\boldsymbol{f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}}$$

$$\boldsymbol{f(0) = \frac{1}{35}}, \quad \boldsymbol{f(1) = \frac{12}{35}}, \quad \boldsymbol{f(2) = \frac{18}{35}}, \quad \boldsymbol{f(3) = \frac{4}{35}}$$

$$\mu = E(x) = 0 * \left(\frac{1}{35}\right) + 1 * \left(\frac{12}{35}\right) + 2 * \left(\frac{18}{35}\right) + 3 * \left(\frac{4}{35}\right) = \frac{12}{7} = 1.7$$

확률변수의 기대값

예제 – 연속형인 경우

어떤 전자장치의 수명 (시간) 을 확률 변수 X라 하고 확률 밀도 함수가

$$\boldsymbol{f(x)} = \begin{cases} \dfrac{\mathbf{20{,}000}}{\boldsymbol{x^3}} & ,\boldsymbol{x} > \mathbf{100} \\ \mathbf{0} & ,\text{다른 곳} \end{cases}$$

이라할 때 기대수명을 구하라.

$$\boldsymbol{\mu} = \boldsymbol{f(x)} = \int_{100}^{\infty} \boldsymbol{x}\frac{\mathbf{20{,}000}}{\boldsymbol{x^3}}\boldsymbol{dx} = \int_{100}^{\infty} \boldsymbol{x}\frac{\mathbf{20{,}000}}{\boldsymbol{x^2}}\boldsymbol{dx} = \mathbf{200}$$

확률변수의 기대값

X가 확률분포 $f(x)$ 를 가지는 확률 변수일 때, 종속된 확률 변수 $g(X)$의 평균 혹은 기대값은**?**

X가 이산형인 경우

$$\mu_{g(x)} = E[g(x)] = \sum_{x} g(x)f(x)$$

X가 연속형인 경우

$$\mu = E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

확률변수의 기대값

예제 – 이산형인 경우

오후 4시에서 5시 사이에 세차장에서 서비스를 받는 차의 수를 X라 할 때 X의 확률 분포를

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

$\boldsymbol{g(x) = 2X - 1}$ 를 종업원이 받는 수당이라고 할 종업원의 기대수익은?

$$E[g(x)] = E(2X-1) = \sum_{x=4}^{9}(2x-1)f(x)$$

확률변수의 기대값

예제 – 이산형인 경우

$$E[g(x)] = E(2X-1) = \sum_{x=4}^{9}(2x-1)f(x)$$

$$E[g(x)] = 7*\left(\frac{1}{12}\right) + 9*\left(\frac{1}{12}\right) + 11*\left(\frac{1}{4}\right) + 13*\left(\frac{1}{4}\right) + 15*\left(\frac{1}{6}\right) + 17*\left(\frac{1}{6}\right)$$

$$= 12.67$$

확률변수의 기대값

예제 – 연속형인 경우

확률 변수 X의 확률 밀도 함수가

$$f(x) = \begin{cases} \frac{x^2}{3} & , \ -1 < x < 2 \\ 0 & , \quad \text{다른 곳} \end{cases}$$

이라할 때 $g(X) = 4X + 3$의 기대값을 구하라.

$$E(4X + 3) = \int_{-1}^{2} \frac{(4x + 3)x^2}{3} dx = \frac{1}{3} \int_{-1}^{2} (4x^3 + 3x^2) dx = 8$$

확률변수의 기대값

X와 **Y**를 결합확률분포 $f(x, y)$ 를 가지는 확률 변수라 할 때,
확률 변수 $\mathrm{g}(X, Y)$의 평균 혹은 기대값은**?**

X,Y가 이산형인 경우

$$\mu_{g(X,Y)} = E[g(X, Y)] = \sum_{x} \sum_{y} g(x, y) f(x, y)$$

X,Y가 연속형인 경우

$$\mu_{g(X,Y)} = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) f(x) dx\, dy$$

확률변수의 기대값

X와 **Y**를 표와 같은 결합확률분포를 가지는확률 변수라 할 때, $g(X,Y) = XY$ 의 평균 혹은 기대값은**?**

f(x,y)		x			행의 합
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{5}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
열의 합		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

확률변수의 기대값

X와 **Y**를 표와 같은 결합확률분포를 가지는확률 변수라 할 때, $g(X,Y)=XY$ 의 평균 혹은 기대값은**?**

$$\mathrm{E(XY)} = \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y)$$

$$\mathrm{E(XY)} = 0*0*f(0,0)+0*1*f(0,1)+1*0*f(1,0)+1*1*f(1,1)+2*0*f(2,0)=f(1,1)=\frac{3}{14}$$

확률변수의 기대값

다음과 같은 결합밀도함수에 대하여 $\mathrm{E}(Y/X)$ 를 구하라.

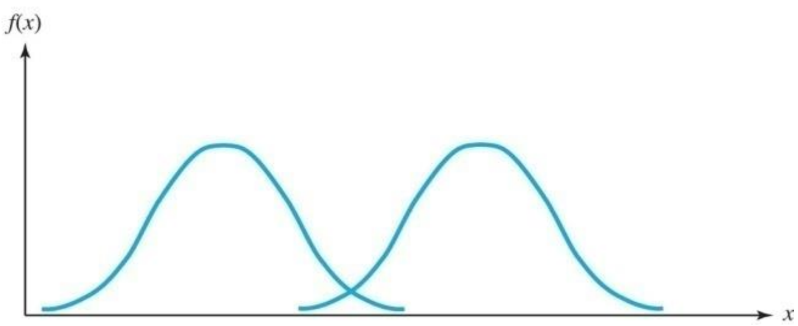
$$\boldsymbol{f(x)} = \begin{cases} \dfrac{\boldsymbol{x(1+3y^2)}}{\boldsymbol{4}} & , \ \boldsymbol{0 < x < 2,\ 0 < y < 1} \\ \boldsymbol{0} & , \quad \text{다른 곳} \end{cases}$$

$$\mathrm{E}\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \frac{\boldsymbol{x(1+3y^2)}}{\boldsymbol{4}} dx\, dy$$

$$= \int_0^1 \frac{y+3y^3}{2} dy = \frac{5}{8}$$

확률변수의 분산

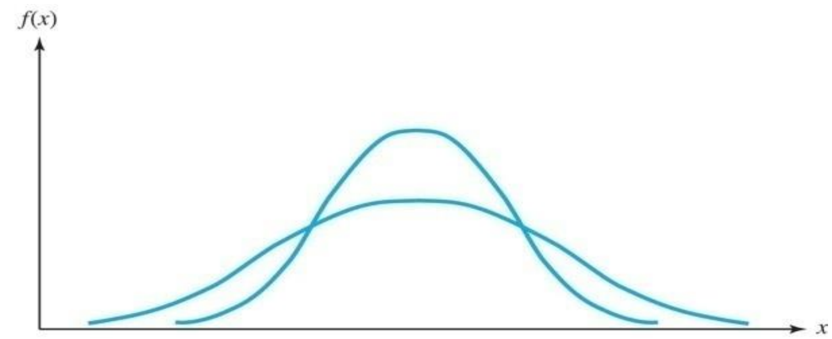
분산은 확률변수에 대한 흩어짐 혹은 변동성을 측정



평균값 다름

평균값에 대하여 밀도함수의 형태

산포 정도가 같음, 따라서 분산은 같음



평균값 같음

분산이 다름

평평하고 퍼져있는 밀도함수의 분산이 큼

확률변수의 기대값 : μ

확률 분변수 X의 분산 or 확률변수 X의 확률분포의 분산분산 : σ^2

확률변수 X의 표준편차 (분산의 제곱근): σ

확률변수의 분산

X를 확률분포 $f(x)$ 와 평균 μ를 가지는 확률 변수라 할 때 **X**의 분산은**?**

X가 이산형인 경우

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (X-\mu)^2 f(x)$$

X가 연속형인 경우

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (X-\mu)^2 f(x)dx$$

확률변수의 기대값

예제 – 이산형인 경우

A와 B의 두 회사에서 사업목적으로 사용된 자동차의 수를 확률변수 X라 하고, 각 회사의 확률 분포는 다음과 같다. B회사에 대한 확률분포의 분산이 A회사의 분산보다 큼을 보여라.

A 회사

x	1	2	3
$f(x)$	0.3	0.4	0.3

B 회사

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

확률변수의 기대값

예제 – 확률변수 X가 이산형인 경우

A 회사

x	1	2	3
$f(x)$	0.3	0.4	0.3

B 회사

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

먼저 기대값을 구하면

$$\mu_A = E(X) = 1 * 0.3 + 2 * 0.4 + 3 * 0.3 = 2.0$$

$$\mu_B = E(X) = 0 * 0.2 + 1 * 0.1 + 2 * 0.3 + 3 * 0.3 + 4 * 0.1 = 2.0$$

확률변수의 기대값

예제 – 확률변수 X가 이산형인 경우

A 회사

x	1	2	3
$f(x)$	0.3	0.4	0.3

B 회사

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

확률변수의 분산을 구하면

$$\sigma^2{}_A = \sum_{x=1}^{3}(x-2)^2 f(x) = (1-2)^2 * 0.3 + (2-2)^2 * 0.4 + (3-2)^2 * 0.3 = 0.6$$

$$\sigma^2{}_B = \sum_{x=0}^{4}(x-2)^2 f(x)$$

$$= (0-2)^2 * 0.2 + (1-2)^2 * 0.1 + (2-2)^2 * 0.3 + (3-2)^2 * 0.3 + (4-2)^2 * 0.1 = 1.6$$

확률변수의 분산

X를 확률분포 $f(x)$ 와 평균 μ를 가지는 확률 변수라 할 때 **X**의 분산은**?**

확률 변수 **X**의 분산 – 간단 공식

$$\sigma^2 = E(X^2) - \mu^2$$

확률변수의 분산

X를 확률분포 $f(x)$ 와 평균 μ를 가지는 확률 변수라 할 때

확률변수 **X**의 분산 – 간단 공식 증명

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x)$$

$$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$$

정의에 의하여 $\mu = \sum_x x f(x)$ 이고, $\sum_x f(x) = 1$ 이므로

$$\boldsymbol{\sigma^2 = \sum_x x^2 f(x) - \mu^2 = E\left(X^2\right) - \mu^2}$$

확률변수의 분산

예제 – 확률변수 X의 분산 (간단 공식 활용)

생산라인으로부터 3개의 부품을 추출하여 검사하였을 때, 결함이 있는 부품수를 확률변수 X라고 하고, 확률변수 X의 확률분포가 다음과 같을 때, 간단공식을 사용하여 분산을 구하라.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

먼저 평균을 구하면

$\mu = E(X) = 0 * 0.51 + 1 * 0.38 + 2 * 0.10 + 3 * 0.01 = 0.61$ 이고,

$E(X^2) = 0 * 0.51 + 1 * 0.38 + 4 * 0.10 + 9 * 0.01 = 0.87$ 이다.

따라서, $\boldsymbol{\sigma^2 = 0.87 - (0.61)^2 = 0.4979}$

확률변수의 분산

X를 확률분포 $f(x)$ 를 가지는 확률변수라 하고, 확률변수 $g(x)$ 의 분산은**?**

X가 이산형인 경우

$$\sigma^2{}_{g(x)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_{x} [g(x) - \mu_{g(X)}]^2 f(x)$$

X가 연속형인 경우

$$\sigma^2{}_{g(x)} = E\left\{[g(X) - \mu_{g(X)}]^2\right\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x) dx$$

확률변수의 공분산

공분산은 두 확률변수 간의 관련성의 척도
두 확률변수 사이의 상관정도, 둘 사이의 정의되는 관계의 밀접도를 측정

$\mu_X = E(X)$ **,** $\mu_Y = E(Y)$일 때,

$g(X,Y) = (X - \mu_X)(Y - \mu_Y)$ 라 하면 μ_{XY} 혹은 $Cov(X,Y)$로 표시

- X의 값이 클 때, Y의 값도 크고, X의 값이 작을 때 Y의 값도 작다면

  $(X - \mu_X)(Y - \mu_Y)$ 는 양의 값을 가짐

- X의 값이 클 때, Y의 값은 작고, X의 값이 작을 때 Y의 값이 크다면

  $(X - \mu_X)(Y - \mu_Y)$ 는 음의 값을 가짐

- 공분산이 0이라면 두 확률변수간에는 아무런 선형관계가 없음
  두 확률변수는 서로 독립적인 관계

확률변수의 공분산

X와 Y를 결합분포 $f(X,Y)$를 가지는 확률변수라고 할 때 **X**와 **Y**의 공분산은**?**

X,Y가 이산형인 경우

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (X-\mu_X)(Y-\mu_Y)f(x,y)$$

X,Y가 연속형인 경우

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (X-\mu_X)(Y-\mu_Y)f(x,y)dx\,dy$$

확률변수의 공분산

μ_x와 μ_y 를 평균으로 하는 두 확률변수 **X**와 **Y**의 공분산은**?**

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

확률변수의 공분산

어떤 상자에서 임의로 **2**개의 볼펜을 꺼낼 때, 청색볼펜의 수 **X**와 적색볼펜의 수 **Y**의 결합확률분포가 다음과 같을 때 **X**와 **Y**의 공분산은**?**

f(x,y)		x			행의 합: $h(x)$
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{5}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
열의 합: $g(x)$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

확률변수의 공분산

f(x,y)		x			h(x)
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{5}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
$g(x)$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

$$\mathrm{E(XY)} = \frac{3}{14}$$

$$\mu_X = \sum_{x=0}^{2} xg(x) = 0 * \frac{5}{14} + 1 * \frac{15}{28} + 2 * \frac{3}{28} = \frac{3}{4}$$

$$\mu_Y = \sum_{y=0}^{2} yh(x) = 0 * \frac{15}{28} + 1 * \frac{3}{7} + 2 * \frac{1}{28} = \frac{1}{2}$$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{3}{14} - \frac{3}{4} * \frac{1}{2} = -\frac{9}{56}$$

확률변수의 상관계수

공분산은 두 확률변수 간의 관련성을 나타내는 척도

확률변순 **X**와 **Y**의 공분산이 σ_{XY} 이고, 표준편차가 σ_X, σ_Y 라고 할 때

$$\rho_{XY} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

를 통하여 두 확률 변수의 상관계수를 구할 수 있음

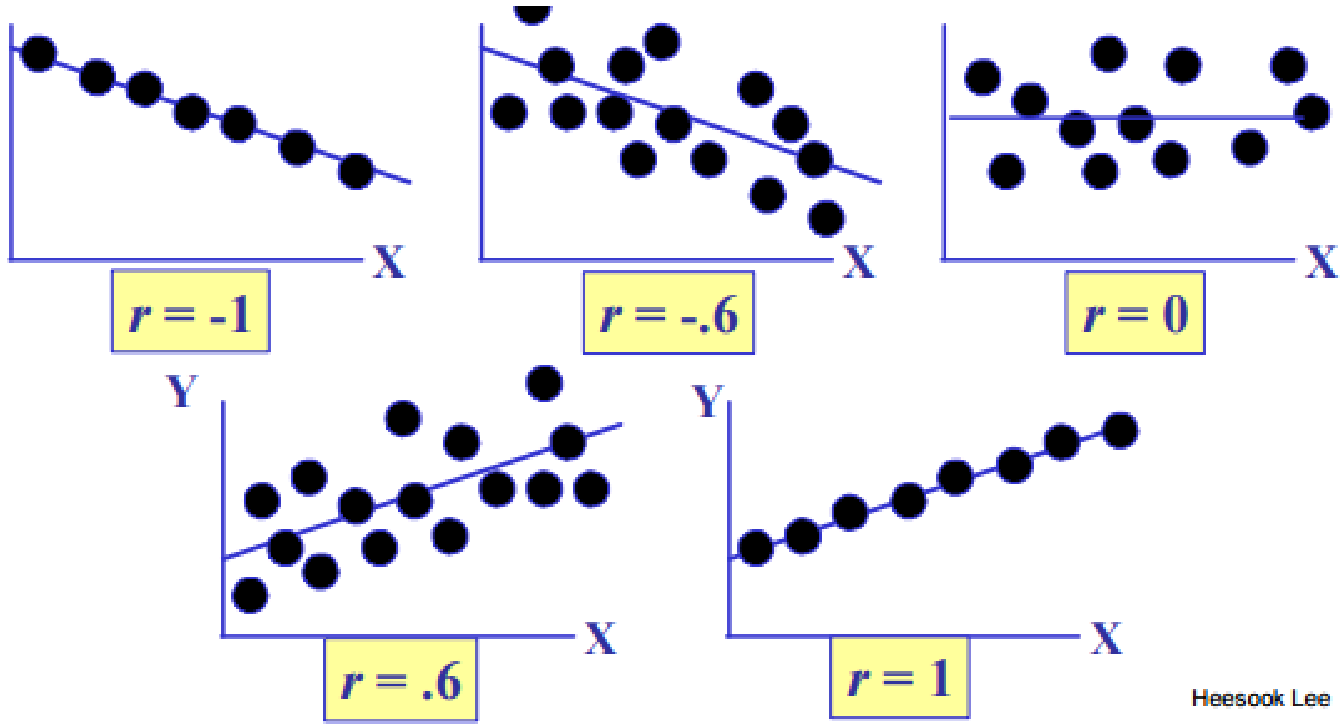
상관계수를 통하여 두 변수의 선형관련성의 정도를 나타냄

확률변수의 상관계수

$$\rho_{XY} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$ 의 값은 [-1,1] 범위로 나타남

양의 상관관계 – 값이 양수일 경우
상관관계가 없음 -값이 0일때
음의 상관관계 – 값이 음수일 경우

확률변수의 상관계수



상관계수는 두 변수의 직선적인 관련성을 나타내는 측도

선형결합된 확률변수의 평균과 분산

확률변수의 기대값을 쉽게 계산할 수 있는 성질

a 와 b 가 상수이면 $E(aX+b) = aE(X) = b$ 이다.

증명

$$E(aX+b) = \int_{-\infty}^{\infty}(ax+b)f(x)dx = a\underbrace{\int_{-\infty}^{\infty} xf(x)dx}_{E(x)} + b\underbrace{\int_{-\infty}^{\infty} f(x)dx}_{1}$$

따라서, $\boldsymbol{E(aX+b) = aE(X) = b}$

따름정리 $a = 0$이면 $E(b) = b$ 이다. $b = 0$ 이면, $E(aX) = aE(X)$ 이다.

선형결합된 확률변수의 평균과 분산

확률변수의 기대값을 쉽게 계산할 수 있는 성질

두 개 이상의 확률변수 X의 함수의 합이나 차의 기대값은 각 함수의 기대값의 합이나 차와 같다.

$$\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{x}) \pm \boldsymbol{h}(\boldsymbol{x})] = \boldsymbol{E}[\boldsymbol{g}(\boldsymbol{x})] \pm \boldsymbol{E}[\boldsymbol{h}(\boldsymbol{x})]$$

증명

$$\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{x}) \pm \boldsymbol{h}(\boldsymbol{x})] = \int_{-\infty}^{\infty} \boldsymbol{E}[\boldsymbol{g}(\boldsymbol{x}) \pm \boldsymbol{h}(\boldsymbol{x})] f(x)dx$$

$$= \int_{-\infty}^{\infty} g(x)f(x)dx \pm \int_{-\infty}^{\infty} h(x)f(x)dx$$

$$= \boldsymbol{E}[\boldsymbol{g}(\boldsymbol{x})] \pm \boldsymbol{E}[\boldsymbol{h}(\boldsymbol{x})]$$

선형결합된 확률변수의 평균과 분산

확률변수의 기대값을 쉽게 계산할 수 있는 성질

결합확률분포를 $f(x,y)$ 를 가지는 두 확률변수 X와 Y가 있을 때, 확률변수 X와 Y의 함수들의 합이나 차의 기대값은 각 함수의 기대값의 합이나 차와 같다.

$$\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y}) \pm \boldsymbol{h}(\boldsymbol{X},\boldsymbol{Y})] = \boldsymbol{E}[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})] \pm \boldsymbol{E}[\boldsymbol{h}(\boldsymbol{X},\boldsymbol{Y})]$$

증명

$$\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y}) \pm \boldsymbol{h}(\boldsymbol{X},\boldsymbol{Y})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) \pm \boldsymbol{h}(\boldsymbol{x},\boldsymbol{y})]f(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)dxdy$$

$$= \boldsymbol{E}[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})] \pm \boldsymbol{E}[\boldsymbol{h}(\boldsymbol{X},\boldsymbol{Y})]$$

선형결합된 확률변수의 평균과 분산

확률변수의 기대값을 쉽게 계산할 수 있는 성질

결합확률분포를 $f(x, y)$ 를 가지는 두 확률변수 X와 Y가 있을 때,
확률변수 X와 Y의 함수들의 합이나 차의 기대값은 각 함수의 기대값의 합이나 차와 같다.

$$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)]$$

따름정리

- $g(X,Y) = g(X)$ 이고, $h(X,Y) = h(Y)$라 하면

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)] \text{ 가 성립}$$

- $g(X,Y) = X$ 이고, $h(X,Y) = Y$ 라 하면

$$E[X \pm Y] = E[X] \pm E[Y] \text{ 가 성립}$$

선형결합된 확률변수의 평균과 분산

확률변수의 기대값을 쉽게 계산할 수 있는 성질

두 확률변수 X와 Y가 서로 독립이라면,

$$\boldsymbol{E(XY) = E(X)E(Y)}$$ 가 성립

증명

결합확률분포 $f(x,y)$를 가지는 두 확률변수의 기대값

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dxdy$$

X와 Y의 주변분포를 $g(x)$, $h(x)$라고 하면 , X와 Y는 독립이므로 $f(x,y) = g(x)h(x)$ 가 성립

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dxdy = \int_{-\infty}^{\infty} xg(x)dx \int_{-\infty}^{\infty} yh(x)dy = E(X)E(Y)$$

선형결합된 확률변수의 평균과 분산

확률변수의 분산이나 표준편차를 쉽게 계산할 수 있는 성질

X와 Y가 결합확률분포 $f(x, y)$를 가지는 확률변수이고 a, b, c 가 상수 일 때

$$\sigma^2{}_{aX+bY+c} = a^2\sigma^2{}_X + b^2\sigma^2{}_Y + 2abc_{XY}$$ 가 성립

증명

$$\sigma^2{}_{aX+bY+c} = E\{[(aX + bY + c) - \mu_{aX+bY+c}]^2\}$$

$$\mu_{aX+bY+c} = E(aX + bY + c) = aE(X) + bE(Y) + c = a\mu_X + b\mu_Y + c$$

$$\sigma^2{}_{aX+bY+c} = E\{[a(X - \mu_X) + b(Y - \mu_Y)]^2\}$$

$$= a^2E[(X - \mu_X)^2] + b^2E[(Y - \mu_Y)^2] + 2abE[(X - \mu_X)(Y - \mu_Y)]$$

$$= a^2\,\sigma^2{}_X + b^2\sigma^2{}_Y + 2abc_{XY}$$

선형결합된 확률변수의 평균과 분산

확률변수의 분산이나 표준편차를 쉽게 계산할 수 있는 성질

X와 Y가 결합확률분포 $f(x, y)$를 가지는 확률변수이고 a, b, c 가 상수 일 때

$$\boldsymbol{\sigma^2}_{\boldsymbol{aX+bY+c}} = \boldsymbol{a^2\sigma^2}_{\boldsymbol{X}} + \boldsymbol{b^2\sigma^2}_{\boldsymbol{Y}} + \boldsymbol{2abc_{XY}}$$ 가 성립

증명

$$\boldsymbol{\sigma^2}_{\boldsymbol{aX+bY+c}} = \boldsymbol{a^2\sigma^2}_{\boldsymbol{X}} + \boldsymbol{b^2\sigma^2}_{\boldsymbol{Y}} + \boldsymbol{2abc_{XY}}$$

$b = 0$ 으로 놓으면 $\boldsymbol{\sigma^2}_{\boldsymbol{aX+c}} = \boldsymbol{a^2\sigma^2}$

$a = 1, b = 0$ 으로 놓으면 $\boldsymbol{\sigma^2}_{\boldsymbol{X+c}} = \boldsymbol{a^2}_{\boldsymbol{X}} = \boldsymbol{\sigma^2}$

$b = 0, c = 0$ 으로 놓으면 $\boldsymbol{\sigma^2}_{\boldsymbol{aX}} = \boldsymbol{a^2\sigma^2}_{\boldsymbol{X}} = \boldsymbol{a^2\sigma^2}$

선형결합된 확률변수의 평균과 분산

확률변수의 분산이나 표준편차를 쉽게 계산할 수 있는 성질

X와 Y를 독립인 확률변수라고 하면, a, b, c 가 상수 일 때

$$\boldsymbol{\sigma^2}_{aX+bY} = a^2\sigma^2{}_X + b^2\sigma^2{}_Y$$ 가 성립

$$\boldsymbol{\sigma^2}_{aX-bY} = a^2\sigma^2{}_X + b^2\sigma^2{}_Y$$ 가 성립

$X_1, X_2, X_3, \cdots, X_n$ 을 독립인 확률변수라 하면,

$$\boldsymbol{\sigma^2}_{a_1X_1+a_2X_2+\cdots+a_nX_n} = a_1{}^2\sigma^2{}_{X_1} + a_2{}^2\sigma^2{}_{X_2} + \cdots + a_n{}^2\sigma^2{}_{X_n}$$ 가 성립

- 확률 변수에 어떠한 상수를 더하거나 빼더라도 분산은 변하지 않음
- 어떠한 상수를 더하거나 빼는 것은 확률변수의 값을
  오른쪽이나 왼쪽으로 이동시킬 뿐이며 변동을 변화하지 않음

체비세프 정리

- 평균을 중심으로 대칭적인 일정한 범위내에 적어도 얼마만큼의 자료가 있는지를 나타냄

- 어떤 실수 k에 대하여 확률변수가 평균으로부터 표준편차의 k배 범위 내의 값을 취할 확률의 예측값을 제공

$$1 - \frac{1}{k^2}$$

확률변수 X가 평균으로부터 표준편차의 K배 범위 내의 값을 취할 확률을 적어도 1-1/k^2

즉 $P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$ **가 성립**

체비세프 정리

확률변수 X가 주어져 있고, 어떤 고등학교 학생들의 키를 확률 변수 X라고 정의하였을 때 우리에게 알려져 있는 정보는 학생들의 키의 평균 μ와 표준편차 σ 를 알고 있을 경우
(키가 정확하게 어떤 분포를 하고 있는지는 알 수 없는 상태)

체비세프 정리를 이용하여

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

$$P(|X - \mu| < 2\sigma) \geqq 1 - \frac{1}{4}$$

$$P(\mu - 2\sigma \leqq X \leqq \mu + 2\sigma) \geqq \frac{3}{4}$$

학생들의 키가 정확하게 어떤 분포를 하고 있는지는 알 수 없지만,
평균에서 표준편차의 2배 떨어진 거리 내에 최소 75%의 학생들이 속해있다는 사실을 알 수 있음

감사합니다.