Chapter02

확률 변수와 확률 분포함수

| 결합확률

FASTCAMPUS
ONLINE
금융공학/퀀트 |

강사. 장순용

I 키포인트

- 결합확률.
- 이변량 확률분포.
- 공분산과 상관계수.
- 독립성과 상관성.
- 변수의 합.

I 이변량 결합확률분포

- 두개의 확률변수 X와 Y에 대한 확률.
- 이변량 결합확률분포: $P(x,y) = P(X = x, Y = y)$.

$\Rightarrow X = x$ AND $Y = y$ 일 확률.

I 이변량 결합확률분포: 공분산 (covariance)

- $Cov(X,Y) = E\left[(X-\mu_x)\left(Y-\mu_y\right)\right]$

$$= \sum_{all\ x_i}\sum_{all\ y_i}(x_i-\mu_x)(y_i-\mu_y)P(x_i,y_i)$$

- 공분산의 간편 수식: $Cov(X,Y) = E[X\,Y] - E[X]\,E[Y]$

$$= \sum_{all\ x_i}\sum_{all\ y_i} x_i\, y_i P(x_i,y_i) - \mu_x\mu_y$$

- $Var(X) = Cov(X,X)$ $\Leftarrow$ 분산과 공분산의 연결

I 이변량 결합확률분포: 상관계수 (correlation)

- $Cor(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y}$
- 상관계수의 값은 -1과 1사이의 수치이다.
- 상관계수는 선형관계의 방향과 강도를 나타냄.

$\rightarrow Cor(X,Y) > 0$: X와 Y사이에 양의 선형관계가 있음.

$\rightarrow Cor(X,Y) < 0$: X와 Y사이에 음의 선형관계가 있음.

$\rightarrow Cor(X,Y) = 0$: X와 Y사이에 선형관계가 **없음**.

I 독립성과 상관성

- 독립성: $P(X,Y) = P(X)\,P(Y)$.

  $\Rightarrow Cov(X,Y) = E[X\,Y] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0.$

  $Cor(X,Y) = 0$. 그러므로 "상관성 없음"을 내포함.

- 상관계수: $Cor(X,Y)$.

  → 상관계수는 -1과 1 사이의 수치이다.

  → "상관성이 없다" = "상관계수 0". 하지만 독립성을 내포하지는 않는다.

I 독립성과 상관성

예: X와 Y 동전 두 개를 동시에 던지는 경우를 가정해 본다. 표본공간은 HH, HT, TH, TT로 이루어져 있다. 독립성을 확인해 본다.

개개 확률분포함수는 다음과 같다:

$$\left.\begin{aligned} P(X=H) &= \frac{1}{2} \\ P(X=T) &= \frac{1}{2} \end{aligned}\right\} P(X)$$

$$\left.\begin{aligned} P(Y=H) &= \frac{1}{2} \\ P(Y=T) &= \frac{1}{2} \end{aligned}\right\} P(Y)$$

I 독립성과 상관성

예: X와 Y 동전 두 개를 동시에 던지는 경우를 가정해 본다. 표본공간은 HH, HT, TH, TT로 이루어져 있다. 독립성을 확인해 본다.

그러므로 다음을 쉽게 확인할 수 있다. 즉, X와 Y는 서로 독립관계이다.

$$P(X = H, Y = H) = \frac{1}{4} = P(X = H) \times P(Y = H)$$

$$P(X = H, Y = T) = \frac{1}{4} = P(X = H) \times P(Y = T)$$

$$P(X = T, Y = H) = \frac{1}{4} = P(X = T) \times P(Y = H)$$

$$P(X = T, Y = T) = \frac{1}{4} = P(X = T) \times P(Y = T)$$

I 독립성과 상관성

예: −1, 0, 1에서 동일한 확률 1/3 을 갖는 확률변수 X와 $Y = X^2$과 같이 정의되는 Y변수가 있다. 그러면 결합확률은 다음과 같다. 물음에 답하여라.

$$P(X = -1, Y = 1) = \frac{1}{3} \quad , \quad P(X = 0, Y = 0) = \frac{1}{3} \quad , \quad P(X = 1, Y = 1) = \frac{1}{3}$$

1). X와 Y가 서로 독립인지 확인해 보아라.

X의 확률분포함수는 다음과 같다:

$$P(X = -1) = \frac{1}{3} \quad , \quad P(X = 0) = \frac{1}{3} \quad , \quad P(X = 1) = \frac{1}{3}$$

그리고 Y의 확률분포함수는 다음과 같다:

$$P(Y = 0) = \frac{1}{3} \quad , \quad P(Y = 1) = \frac{2}{3}$$

I 독립성과 상관성

예: −1, 0, 1에서 동일한 확률 1/3 을 갖는 확률변수 X와 $Y = X^2$과 같이 정의되는 Y변수가 있다. 그러면 결합확률은 다음과 같다. 물음에 답하여라.

$$P(X = -1, Y = 1) = \frac{1}{3}, \quad P(X = 0, Y = 0) = \frac{1}{3}, \quad P(X = 1, Y = 1) = \frac{1}{3}$$

1). X와 Y가 서로 독립인지 확인해 보아라.

그런데 다음을 확인할 수 있다.

$$P(X = -1, Y = 1) = \frac{1}{3} \neq P(X = -1) \times P(Y = 1) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

$$P(X = 0, Y = 0) = \frac{1}{3} \neq P(X = 0) \times P(Y = 0) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$$

$$P(X = 1, Y = 1) = \frac{1}{3} \neq P(X = 1) \times P(Y = 1) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

$\Rightarrow P(X,Y) \neq P(X)P(Y)$ 이므로 X와 Y는 서로 독립이 아닌 종속 관계이다.

I 독립성과 상관성

예: −1, 0, 1에서 동일한 확률 1/3 을 갖는 확률변수 X와 $Y = X^2$과 같이 정의되는 Y변수가 있다. 그러면 결합확률은 다음과 같다. 물음에 답하여라.

$$P(X = -1, Y = 1) = \frac{1}{3} \quad , \quad P(X = 0, Y = 0) = \frac{1}{3} \quad , \quad P(X = 1, Y = 1) = \frac{1}{3}$$

2). 이제는 X와 Y사이의 상관계수를 계산해 보아라.

$$E[X] = -1 \times P(X = -1) + 0 \times P(X = 0) + 1 \times P(X = 1) = -\frac{1}{3} + 0 + \frac{1}{3} = 0$$

$$E[Y] = 0 \times P(Y = 0) + 1 \times P(Y = 1) = 0 + 1 \times \frac{2}{3} = \frac{2}{3}$$

$$E[X\,Y] = E[X\,X^2] = E[X^3] = -1 \times P(X = -1) + 0 \times P(X = 0) + 1 \times P(X = 1) = 0$$

그러므로 $Cov(X, Y) = E[X\,Y] - E[X]\,E[Y] = 0 - 0 \times \frac{2}{3} = 0$

$\Rightarrow$ 공분산이 0이니 상관계수도 0이다! 서로 독립은 아닌데 상관계수가 0인 경우 발견!

I 두 확률 변수의 합 (이변량)

- 확률변수 X와 Y의 합의 경우를 생각해 본다.
- 평균 (기대값): $E[X+Y] = E[X] + E[Y]$
- 분산: $Var(X+Y) = E\left[\left(X+Y-\mu_x-\mu_y\right)^2\right]$

$$= E\left[\left((X-\mu_x)+(Y-\mu_y)\right)^2\right]$$

$$= E\left[(X-\mu_x)^2+(Y-\mu_y)^2+2(X-\mu_x)(Y-\mu_y)\right]$$

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

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

$\Leftarrow$ X와 Y가 서로 독립인 경우.

I 여러 확률 변수의 합 (다변량)

- 평균 (기대값): $E[X_1+X_2+X_3+\cdots] = E[X_1]+E[X_2]+E[X_3]+\cdots$
- 분 산 : $Var(X_1+X_2+X_3+\cdots) = Var(X_1)+Var(X_2)+Var(X_3)+\cdots$

$$+2Cov(X_1,X_2)+2Cov(X_1,X_3)+2Cov(X_2,X_3)+\cdots$$

- 변수들이 서로 독립적이라면 공분산 항은 필요 없음:

$$Var(X_1+X_2+X_3+\cdots) = Var(X_1)+Var(X_2)+Var(X_3)+\cdots$$

I 현대 포트폴리오 이론 (MPT: Modern Portfolio Theory)

포트폴리오 (Portfolio)

= Port + Folio

(가지고 가다) + (서류/증권)

≃ "서류철" *or* "서류가방"

I 현대 포트폴리오 이론 (MPT: Modern Portfolio Theory)

- N개의 자산으로 포트폴리오를 구성 (서브인덱스 $i = 1, 2, \cdots, N$).
- 개개 자산이 포트폴리오에서 차지하는 비율은 w_i.

  $\rightarrow w_i$ 는 $w_1 + w_2 + \cdots + w_N = 1$ 제약 조건 충족.
- 개개 자산의 수익률을 확률변수 R_i로 나타내면, 포트폴리오 수익률 R_p는:

$$R_P = w_1 R_1 + w_2 R_2 + \cdots + w_N R_N$$

- 포트폴리오의 리스크는 수익률의 표준편차 σ_P 또는 그것의 제곱인 분산 $\sigma_P{}^2$으로 나타낼 수 있다.

I 현대 포트폴리오 이론 (MPT: Modern Portfolio Theory)

- 포트폴리오의 평균 수익률은 R_P의 기대값이다:

$$E[R_P] = w_1 E[R_1] + w_2 E[R_2] + \cdots + w_N E[R_N]$$

- 개개 자산이 서로 독립적인 경우 포트폴리오의 리스크는 다음과 같다:

$$\sigma_P{}^2 \approx w_1{}^2 \sigma_1{}^2 + w_2{}^2 \sigma_2{}^2 + \cdots + w_N{}^2 \sigma_N{}^2$$

- 만약에 $w_i = 1/N$이고 $\sigma_1{}^2 \approx \sigma_2{}^2 \approx \cdots \approx \sigma_N{}^2$라면:

$$\sigma_P{}^2 \approx N \frac{1}{N^2} \sigma^2 = \frac{\sigma^2}{N}$$

N이 커지면 0으로 수렴

I 끝.

감사합니다.

Fast campus

FAST CAMPUS
ONLINE
장순용 강사.