

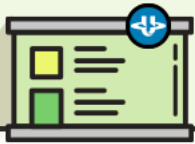
03

강

통계적 추론

확률분포 2

한국방송통신대학교 통계·데이터과학과
이공희 교수



학습내용

- ① 결합 확률분포를 이해한다.
- ② 공분산과 상관계수를 이해한다.
- ③ 조건부 확률분포를 이해한다.
- ④ 독립성을 이해한다.

01

결합 확률분포



1 이산형 결합 확률분포

● 결합 확률질량함수와 결합누적분포함수

- $f(x_i, y_j) = P(X = x_i, Y = y_j)$

- $x_1 < x_2 < \cdots < x_s < \cdots, \quad y_1 < y_2 < \cdots < y_t < \cdots$

$$F(x_s, y_t) = P(X \leq x_s, Y \leq y_t) = \sum_{j=1}^t \sum_{i=1}^s f(x_i, y_j)$$

1

이산형 결합 확률분포

● 결합 확률질량함수의 성질

- $0 \leq f(x_i, y_j) \leq 1$
- $\sum_j \sum_i f(x_i, y_j) = 1$
- $f_X(x_i) = \sum_j f(x_i, y_j), f_Y(y_j) = \sum_i f(x_i, y_j)$
- $P(x_a \leq X \leq x_b, y_c \leq Y \leq y_d) = \sum_{j=c}^d \sum_{i=a}^b f(x_i, y_j)$

1

이산형 결합 확률분포

예

3개의 공(1~3)이 든 상자에서 2개의 공을 뽑을 때, 첫 번째 공의 숫자를 X , 두 번째 공의 숫자를 Y 라 할 때 결합 확률질량함수는?

1

이산형 결합 확률분포

예

3개의 공(1~3)이 든 상자에서 2개의 공을 뽑을 때, 첫 번째 공의 숫자를 X , 두 번째 공의 숫자를 Y 라 할 때 결합 확률질량함수는?

2

연속형 결합 확률분포

● 결합 누적분포함수와 결합 확률밀도함수

- $F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(s, t) ds dt$

- $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$

2

연속형 결합 확률분포

● 결합 확률밀도함수의 성질

- $f(x, y) \geq 0$

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

- $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$

- $P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) dx dy$

2

연속형 결합 확률분포

예 $f(x, y) = 6x^2y$ 일 때 $P(0 \leq X \leq \frac{3}{4}, \frac{1}{3} \leq Y \leq 1)$ 값은?

2

연속형 결합 확률분포

예 $f(x, y) = 6x^2y$ 일 때 $P(0 \leq X \leq \frac{3}{4}, \frac{1}{3} \leq Y \leq 1)$ 값은?

3

기댓값

● 기댓값

■ 이산형 확률변수:

$$E(g(X, Y)) = \sum_x \sum_y g(x, y) f(x, y)$$

■ 연속형 확률변수

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

3

기댓값

● 기댓값의 성질

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

3

기댓값

예

다음 결합분포에서 $X+Y$ 의 분포와 기댓값은?

Y \ X	-1	1	행의 합
0	0.4	0.1	0.5
1	0.1	0.4	0.5
열의 합	0.5	0.5	1

3

기댓값

예 다음 결합분포에서 $X+Y$ 의 분포와 기댓값은?

4

적률생성함수

● 적률생성함수

- $M_{X,Y}(t_1, t_2) = E(e^{t_1X+t_2Y})$
- $M_X(t_1) = M_{X,Y}(t_1, 0)$
- $M_Y(t_2) = M_{X,Y}(0, t_2)$

5

다차원 결합 확률분포

- 결합 확률질량함수

- $0 \leq f(x_1, x_2, \dots, x_k) \leq 1$

- $\sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} f(x_1, x_2, \dots, x_k) = 1$

- $P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \dots, a_k \leq X_k \leq b_k)$

$$= \sum_{a_1 \leq x_1 \leq b_1} \sum_{a_2 \leq x_2 \leq b_2} \cdots \sum_{a_k \leq x_k \leq b_k} f(x_1, x_2, \dots, x_k)$$

5

다차원 결합 확률분포

● 결합 확률밀도함수

$$\blacksquare 0 \leq f(x_1, x_2, \dots, x_k)$$

$$\blacksquare \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_k) d_{x_1} d_{x_2} \cdots d_{x_k} = 1$$

$$\blacksquare P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \dots, a_k \leq X_k \leq b_k)$$

$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_k}^{b_k} f(x_1, x_2, \dots, x_k) d_{x_1} d_{x_2} \cdots d_{x_k}$$

02

공분산



1

공분산

● 공분산(covariance)의 정의

- $$\begin{aligned} Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

1

공분산의 정의

● 공분산의 성질

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(aX + b, cY + d) = acCov(X, Y)$
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$

1

공분산의 정의

● 공분산의 성질

- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$
- $Var(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 Var(X_i) \pm 2 \sum_{i < j} a_i a_j Cov(X_i, X_j)$

1

공분산

예 다음 결합분포에서 $Cov(X, Y)$ 는?

Y \ X	-1	1	행의 합
0	0.4	0.1	0.5
1	0.1	0.4	0.5
열의 합	0.5	0.5	1

2

상관계수

● 상관계수(correlation)의 정의

$$\blacksquare \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

2

상관계수

- 상관계수의 성질
 - $-1 \leq \text{Corr}(X, Y) \leq 1$
 - $\text{Corr}(aX + b, cY + b) = ac/|ac| \text{Corr}(X, Y)$

2

상관계수

● 상관계수의 성질

- $P(Y = aX + b) = 1$

$$\rightarrow \text{Corr}(X, Y) = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$$

2

상관계수

예 다음 결합분포에서 $Cov(X, Y)$ 는?

Y \ X	-1	1	행의 합
0	0.4	0.1	0.5
1	0.1	0.4	0.5
열의 합	0.5	0.5	1

03

조건부 확률분포



1

조건부 확률분포

● 이산형 확률변수

$$\blacksquare f_{Y|X} = P(Y = y|X = x) = \frac{f(x,y)}{f_X(x)}, \quad y = 0, 1, 2, \dots$$

1

조건부 확률분포

예

불량품 1개와 정상제품 4개 들어있는 바구니
에서 3개의 제품을 복원추출

X : 불량품 총수

Y : 마지막 제품이 불량품인지 여부

$P(Y|X = 1)$ 의 값은?

1

조건부 확률분포

예 $P(Y|X = 1)$ 의 값은?

1

조건부 확률분포

● 연속형 확률변수

- $f_{Y|X} = \frac{f(x,y)}{f_X(x)}, \quad -\infty < y < \infty$
- $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$

1

조건부 확률분포

예

두 확률 변수 X, Y 의 결합확률 밀도함수가 다음과 같을 때 $f_{Y|X}(y|x)$ 는?

$$f(x, y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{그 밖에} \end{cases}$$

1

조건부 확률분포

예

두 확률 변수 X, Y 의 결합확률 밀도함수가 다음과 같을 때 $f_{Y|X}(y|x)$ 는?

2

조건부 기댓값

● 조건부 기댓값

■ 이산형 확률변수:

$$E(Y|X = x) = \sum_{y=0}^{\infty} y f_{Y|X}(y|x)$$

■ 연속형 확률변수

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

2

조건부 기댓값

예 조건부 확률분포 $f_{Y|X}(y|x)$ 가 다음과 같을 때 $E(Y|X = x)$ 는?

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & , x < y < 1 \\ 0 & , \text{그 밖에} \end{cases}$$

2

조건부 기댓값

예 조건부 확률분포 $f_{Y|X}(y|x)$ 가 다음과 같을 때 $E(Y|X = x)$ 는?

2

조건부 기댓값

- 조건부 기댓값의 성질
 - $Var[E(Y|X)] \leq Var(Y)$

2

조건부 기댓값

● 조건부 기댓값의 성질

- $E[E(Y|X)] = E(Y)$
- $E(aY + b|X = x) = aE(Y|X = x) + b$

2

조건부 기댓값

예 $f_{Y|X} = 2 \cdot \frac{y}{x^2} I_{(0,x)}(y), 0 < x < 1$ 일 때
 $E(Y^2 - \frac{4}{3}XY + \frac{4}{9}X^2 | X = x)$ 는?

04

독립성



1

독립성의 정의

● 두 사건의 독립

- $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$

1

독립성의 정의

● 확률변수간 독립

- $f(x, y) = f_{Y|X}(y|x)f_X(x)$
- $f_{Y|X}(y|x)$ 가 x 에 의존하지 않음 $\Rightarrow f_{Y|X}(y|x) = f_Y(y)$
 $\Rightarrow f(x, y) = f_X(x)f_Y(y)$

1

독립성의 정의

- 두 확률변수의 독립
 - $f(x, y) = f_X(x)f_Y(y)$
 - $F(x, y) = F_X(x)F_Y(y)$

2

독립성과 공분산

- 독립된 확률변수의 성질
 - X, Y 독립 $\Rightarrow Cov(X, Y) = 0, Corr(X, Y) = 0$
 - $Var(X \pm Y) = Var(X) + Var(Y)$

2

독립성과 공분산

증명 X, Y 독립 $\Rightarrow Cov(X, Y) = 0$

2

독립성과 공분산

예

결합분포가 다음과 같을 때 X, Y 가 독립인지 살펴보고 $Cov(X, Y)$ 를 구하라.

Y \ X	-1	0	1
0	0	1/3	0
1	1/3	0	1/3

2

독립성과 공분산

예

결합분포가 다음과 같을 때 X, Y 가 독립인지 살펴보고 $Cov(X, Y)$ 를 구하라.

3

독립성과 적률생성함수

● 독립성과 적률생성함수

- X, Y 독립 $\Rightarrow M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$

4

다변량 확률분포와 독립성

● 독립성

■ X_1, X_2, \dots, X_n 독립

- $f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n)$

- $Var(X_1 + X_2 \cdots + X_n) = \sum_{i=1}^n Var(X_i)$



정리하기

■ 결합 확률질량함수 :

$$f(x_i, y_j) = P(X = x_i, Y = y_j)$$

■ 결합 확률밀도함수 :

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(s, t) ds dt$$

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$



정리하기

- 두 확률변수가 같이 변하는 정도의 측도 : 공분산과 상관계수

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

- $X = x$ 의 조건하에서 Y 의 조건부 확률질량(밀도)함수 :

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

- 두 확률변수의 독립 :

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

04

강

다음시간안내

이산형 확률분포
