3강 확률분포(2)

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

01 결합확률분포

1. 이산형 결합확률분포

- 결합확률질량함수와 결합누적분포함수
  - $f(x_i, y_j) = P(X = x_i, Y = y_j)$
  - $x_1 < x_2 < \cdots < x_s < \cdots, \quad y_1 < y_2 < \cdots < y_s < \cdots$

  $$F(x_s, y_t) = P(X \le x_s, Y \le y_t) = \sum_{j=1}^{t} \sum_{i=1}^{s} f(x_i, y_j)$$
- 결합확률질량함수의 성질
  - $0 \le f(x_i, y_j) \le 1$
  - $\sum_j \sum_i f(x_i, y_j) = 1$
  - $f_x(x_i) = \sum_j f(x_i, y_j), \quad f_y(y_j) = \sum_i f(x_i, y_j)$
  - $$P(x_a \le X \le x_b, y_c \le Y \le y_d) = \sum_{j=c}^{d} \sum_{i=a}^{b} f(x_i, y_j)$$
- 예) 3개의 공(1~3)이 든 상자에서 2개의 공을 뽑을 때, 첫 번째 공의 숫자를 X, 두 번째 공의 숫자를 Y라 할 때 결합확률질량함수는?
- 풀이)

X Y	1	2	3
1	0	1/6	1/6
2	1/6	0	1/6
3	1/6	1/6	0

$$f(x,y) = \begin{cases} \frac{1}{6}, & x \neq y \\ 0, & x = y \end{cases}$$

2. 연속형 결합확률분포

- 결합누적분포함수와 결합확률밀도함수
    - $F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t) dsdt$
    - $f(x,y) = \dfrac{\partial^2}{\partial x \partial y} F(x,y)$
- 결합확률밀도함수의 성질
    - $f(x,y) \geq 0$
    - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dxdy = 1$
    - $f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx$
    - $P(a \leq X \leq b, c \leq Y \leq d) = \int_{c}^{d} \int_{a}^{b} f(x,y) dxdy$
- 예)
  $f(x,y) = 6x^2 y$
  일 때,
  $P(0 \leq X \leq \dfrac{3}{4}, \dfrac{1}{3} \leq Y \leq 1)$
  의 값은?
- 풀이)
  $$P(0 \leq X \leq \frac{3}{4}, \frac{1}{3} \leq Y \leq 1) = \int_{\frac{1}{3}}^{1} \int_{0}^{\frac{3}{4}} 6x^2 y dxdy = \int_{\frac{1}{3}}^{1} \left[6y \cdot \frac{1}{3} x^3 \right]_{0}^{\frac{3}{4}} dy$$

  $$= \int_{\frac{1}{3}}^{1} 2y \cdot (\frac{3}{4})^3 dy = \frac{27}{32} \int_{\frac{1}{3}}^{1} ydy = \frac{27}{32} \left[\frac{1}{2} \cdot (1^2 - (\frac{1}{3})^2) \right]$$

  $$= \frac{27}{32} \cdot \frac{1}{2} \cdot \frac{8}{9} = \frac{3}{8}$$

3. 기댓값

- 기댓값
    - 이산형 확률변수:
      $E(g(x,y)) = \sum_{x} \sum_{y} g(x,y) f(x,y)$

- 연속형 확률변수:
  $$E(g(x,y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dydx$$
- 기댓값의 성질
  - $E(g(X,Y) \pm h(X,Y)) = E(g(X,Y)) \pm E(h(X,Y))$
- 예) 다음 결합분포에서 X+Y의 분포와 기댓값은?

X Y	-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

- 풀이)
  - X+Y의 분포는 다음과 같다.

	-1	0	1	2
f(x+y)	0.4	0.1	0.1	0.4

$$E(X,Y) = -1 \cdot 0.4 + 0 \cdot 0.1 + 1 \cdot 0.1 + 2 \cdot 0.4 = 0.5$$

$$E(X,Y) = E(X) + E(Y) = 0 + 0.5 = 0.5$$

4. 적률생성함수

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

5. 다차원 결합확률분포

- 결합확률질량함수
  - $0 \le f(x_1, x_2, \cdots, x_k) \le 1$
  - $\sum_{x_1}\sum_{x_2}\cdots\sum_{x_k} f(x_1, x_2, \cdots, x_k) = 1$
  - $P(a_1 \le X_1 \le b_1, a_2 \le X_2 \le b_2, \cdots, a_k \le X_k \le b_k) = \sum_{a_1 \le X_1 \le b_1}\sum_{a_2 \le X_2 \le b_2}\cdots\sum_{a_k \le X_k \le b_k} f(x_1, x_2, \cdots, x_k)$
- 결합확률밀도함수

- $0 \leq f(x_1, x_2, \cdots, x_k)$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \cdots, x_k) d_{x_1} d_{x_2} \cdots d_{x_k} = 1$
- $P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \cdots, a_k \leq X_k \leq b_k) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_k}^{b_k} f(x_1, x_2, \cdots, x_k) d_{x_1} d_{x_2} \cdots d_{x_k}$

02 공분산

1. 공분산

- 공분산(covariance)의 정의
  - $Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$
- 공분산의 성질 1
  - $Cov(X, Y) = Cov(Y, X)$
  - $Cov(aX + b, cY + d) = ac \cdot Cov(X, Y)$
  - $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- 공분산의 성질 2
  - $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$
  - $Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) \pm 2 \sum_{i<j} a_i a_j Cov(X_i, X_j)$
- 예) 다음 결합분포에서 Cov(X,Y)는?

X **Y**	**-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

- 풀이)

$$E(X) = 0, \quad E(Y) = 0.5$$

$$E(XY) = -1 \cdot 0.1 + 1 \cdot 0.4 = 0.3$$

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 0.3 - 0 \cdot 0.5 = 0.3$$

2. 상관계수

- 상관계수(correlation)의 정의
  - $Corr(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$
- 상관계수의 성질 1
  $-1 \le Corr(X,Y) \le 1$
  - - 상관계수는 단위에 의존하지 않는다.
  - $Corr(aX+b, cY+d) = \dfrac{ac}{|ac|}Corr(X,Y)$
- 상관계수의 성질 2
  - $P(Y = aX + b) = 1$

  $\rightarrow Corr(X,Y) = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$
- 예) 다음 결합분포에서 Corr(X,Y)는?

X Y	**-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

$$Var(X) = E(X^2) - (E(X))^2 = (-1)^2 \cdot 0.5 + 1^2 \cdot 0.5 - 0 = 1$$

$$Var(Y) = E(Y^2) - (E(Y))^2 = 1^2 \cdot 0.5 - (0.5)^2 = 0.25$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{0.3}{\sqrt{1}\sqrt{0.25}} = \frac{0.3}{0.5} = 0.6$$

03 조건부 확률분포

1. 조건부 확률분포

- 이산형 확률변수
  - $f_{Y|X} = P(Y = y|X = x) = \dfrac{f(x,y)}{f_X(x)}, \quad y = 0, 1, 2, \cdots$

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

- 예) 불량품 1개와 정상제품 4개 들어있는 바구니에서 3개의 제품을 복원추출
  X: 불량품 총수
  Y: 마지막 제품이 불량품인지 여부
  P(Y|X=1)의 값은?
- 풀이)
  $$P(Y=0|X=1) = \frac{P(Y=0|X=1)}{P(X=1)} = \frac{2(\frac{4}{5})^2\frac{1}{5}}{3(\frac{4}{5})^2\frac{1}{5}} = \frac{2}{3}$$

  $$P(Y=1|X=1) = \frac{P(Y=1|X=1)}{P(X=1)} = \frac{(\frac{4}{5})^2\frac{1}{5}}{3(\frac{4}{5})^2\frac{1}{5}} = \frac{1}{3}$$

	0	1
f(y|x=1)	2/3	1/3

- 연속형 확률변수
  - $f_{Y|X} = \dfrac{f(x,y)}{f_X(x)}, \quad -\infty < y < \infty$
  - $f_X(x) = \displaystyle\int_{-\infty}^{\infty} f(x,y)dy$
- 예) 두 확률 변수 X,Y의 결합확률밀도함수가 다음과 같을 때
  $f_{Y|X}(y|x)$
  는?
  $$f(x,y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & otherwise \end{cases}$$
- 풀이)
  $$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$
  - 이므로
    $f_X(x)$
    을 우선 구한다.
  - $f_X(x) = \displaystyle\int_{-\infty}^{\infty} f(x,y)dy = \int_x^1 2dy = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & otherwise \end{cases}$
  - $f_{Y|X}(y|x) = \dfrac{f(x,y)}{f_X(x)} = \begin{cases} \frac{2}{2(1-x)} & x < y < 1 \\ 0 & otherwise \end{cases}$

2. 조건부 기댓값

- 조건부 기댓값
  - 이산형 확률변수
    $$E(Y|X=x) = \sum_{y=0}^{\infty} y \cdot f_{Y|X}(y|x)$$
  - 연속형 확률변수
    $$E(Y|X=x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)dy$$
- 예) 조건부 확률분포가 다음과 같을 때 E(Y|X=x)는?
  $$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x}, & x < y < 1 \\ 0 & otherwise \end{cases}$$
- 풀이)
  $$E(Y|X=x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)dy$$
  $$= \int_x^1 y\frac{1}{1-x}dy = \frac{1}{1-x}\int_x^1 ydy = \frac{1}{1-x}\left[\frac{1}{2}y^2\right]_x^1$$
  $$= \frac{1}{2}\frac{1}{1-x}(1-x^2) = \frac{1+x}{2}, \quad 0 < x < 1$$
- 조건부 기댓값의 성질
  - *E[E(Y|X)] = E(Y)*
  - *E(aY+b|X=x) = aE(Y|X=x)+b*
- 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?
- 풀이)
  $$E(Y^2 - \frac{4}{3}XY + \frac{4}{9}X^2|X=x) = E(Y^2|X=x) - \frac{4}{3}E(Y|X=x) + \frac{4}{9}x^2$$
  $$E(Y^2|X=x) = \int_0^x y^2 \cdot \frac{2y}{x^2}dy = \frac{2}{x^2}\left[\frac{1}{4}y^4\right]_0^x = \frac{x^2}{2}$$

$$E(Y|X=x)=\int_0^x y\cdot\frac{2y}{x^2}dy=\frac{2}{x^2}\left[\frac{1}{3}y^3\right]_0^x=\frac{2}{3}x$$

$$E(Y^2-\frac{4}{3}XY+\frac{4}{9}X^2|X=x)=\frac{x^2}{2}-\frac{4}{3}x\cdot\frac{2}{3}x+\frac{4}{9}x^2=\frac{1}{18}x^2,\quad 0<x<1$$

04 독립성

1. 독립성의 정의

- 두 사건의 독립
  - *P(A|B) = P(A) ⇔ P(A∩B) = P(A)P(B)*
- 확률변수간 독립
  - $f(x,y)=f_{Y|X}(y|x)f_X(x)$
    $f_{Y|X}(y|x)$
  - 가 x에 의존하지 않음 ⇒
    $f_{Y|X}(y|x)=f_Y(y)$

    $\Rightarrow f(x,y)=f_X(x)f_Y(y)$
- 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 ⇒ *Cov(X,Y)=0, Corr(X,Y)=0*
  - *Var(X±Y) = Var(X) + Var(Y)*
- 증명) X,Y 독립 ⇒ Cov(X,Y)=0
- 풀이)
  $Cov(X,Y)=E(XY)-E(X)E(Y)$

$$=\int_{-\infty}^{\infty}xyf(x,y)dxdy-\int_{-\infty}^{\infty}xf_X(x)dx\int_{-\infty}^{\infty}yf_Y(y)dy$$

$$=\int_{-\infty}^{\infty}xyf_X(x)f_Y(y)dxdy-\int_{-\infty}^{\infty}xf_X(x)dx\int_{-\infty}^{\infty}yf_Y(y)dy$$

$$= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy - \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy = 0$$

- 예) 결합분포가 다음과 같을 때 X,Y가 독립인지 살펴보고 Cov(X,Y)를 구하라.

X **Y**	**-1**	**0**	**1**
0	0	1/3	0
1	1/3	0	1/3

- 풀이)
  - Cov(X,Y) = E(XY) - E(X)E(Y) = 0 - 0*(2/3) = 0
  - P(X=1, Y=1) = 1/3 이고, P(X=1) = 1/3, P(Y=1) = 2/3 이므로 P(X=1, Y=1) ≠ P(X=1)P(Y=1) 이다.
    - 따라서, X,Y는 독립이 아니다.
  - X,Y가 독립이면 Cov(X,Y)=0 이지만, Cov(X,Y)=0 이라고 해서 X,Y가 반드시 독립은 아니다.

3. 독립성과 적률생성함수

- 독립성과 적률생성함수
  - X,Y 독립 ⇒ $M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$
- 독립성 $X_1, X_2, \cdots, X_n$
  - 독립이면,
    - $f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^{n} f(x_i)$
    - $Var(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i)$

05 정리하기

- 결합확률질량함수
  - $f(x_i, y_j) = P(X = x_i, Y = y_j)$
- 결합확률밀도함수
  - $F(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t) ds dt$

- $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$
- 두 확률변수가 같이 변하는 정도의 측도: 공분산과 상관계수
  - $Cov(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y)$
  - $Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$
- X=x의 조건하에서 Y의 조건부 확률질량(밀도)함수
  - $f_{Y|X}(Y=y|X=x) = \frac{f(x,y)}{f_X(x)}$
- 두 확률변수의 독립
  - $f(x,y) = f_X(x)f_Y(y)$