다변량 분포

정의

- 다변량 분포

1. 두개 이상의 확률변수가 결합된 분포를 의미한다.
2. 표본공간 e에서 확률변수 $x_{1,}x_2 \cdots x_n$이 있을 때

D = $\begin{bmatrix} x_{1(c)} \\ \cdots \\ x_{n(c)} \end{bmatrix}$인 벡터를 확률 벡터라고 한다.

3. 이 때, A를 D의 부분집합이라고 한다면 이를 $P_{x_{1,}x_2 \cdots x_n}$(A)로 표기한다.

정의

- 결합 누적 분포 함수

1. 일변량 확률과 마찬가지로, 다변량 분포도 CDF를 정의할 수 있다.
2. $F_{x_1,x_\perp}(x_1, x_2) = \mathrm{P}(\{X_1 \le x_1\} \cap \{X_2 \le x_2\})$ 로 정의한다.

1) 이때, $\mathrm{P}(\{a < X_1 \le b\} \cap \{c < X_2 \le d\}) = F_{x_1,x_\perp}(b, d\,) - F_{x_1,x_\perp}(a, d\,)$

$-F_{x_1,x_\perp}(b, c\,) + F_{x_1,x_\perp}(a, c)$ 이다.

정의

- 결합확률질량함수와 결합확률밀도함수

1. 분포가 이산형이면 결합확률질량함수는

1) $p_{x_1,x_2}(x_1,x_2) = \frac{\partial\{\Sigma\Sigma P_{x_1x_2}(x_1,x_2)\}}{\partial x_1 \partial x_2}$

2. 분포가 연속형이면

1) $f_{x_1,x_2}(x_1,x_2) = \frac{\partial\{\int f(x)\,dx\}}{\partial x_1 \partial x_2}$

3. 이 때, 한쪽 방향만의 분포함수(주변분포함수)를 구할 수 있는데

(1) 분포가 이산형이면 $p_{x_1}(x_1) = \sum_{x_1=1}^{\infty} P_{x_1x_2}(x_1,x_2)$

(2) 분포가 연속형이면 $f_{x_1}(x_1) = \int_{-\infty}^{\infty} f(x)\,dx_2$

정의

- 다변량 분포의 기댓값

1. (x_1, x_2) 이 연속형일때, Y= $g(x_1, x_2)$에 대하여 E(Y)는

1) $\int\int_{-\infty}^{\infty} g(x_1, x_2) f(x_1, x_2)\, dx_1\, dx_2$ 이다.

2. (x_1, x_2)가 이산형일 때, Y= $g(x_1, x_2)$에 대하여 E(Y)는

1) $\Sigma\Sigma g(x_1, x_2) P_{x_1 x_2}(x_1, x_2)$ 이다.

3. 다변량 분포의 적률도 일변량 분포의 적률과 같은 방법으로 구한다.

1) $M_{x_1, x_2}(t_1, t_2) = E(\epsilon^{t_1 x}, e^{t_2 x})$ 라고 할 때

2) $M_{x_1, x_2}(t_1, t_2) = \begin{bmatrix} M_{x_1}(t_1, 0) \\ M_{x_2}(0, t_2) \end{bmatrix}$ 로, 각각의 벡터가 된다.

예제

- 표본공간 e를 동전을 던지는 순서쌍이라고 할 때,
- e={TTT,TTH,THT,HTT,THH,HTH,HHT,HHH}
- $x_{1,}$ = {두번 던지며 나온 H의 수| $x_1 = x_1(c), c \in e\}$
- x_2 = {세번 던지며 나온 H의 수| $x_2 = x_2(c), c \in e\}$

1) $\mathrm{e} \cdot \begin{bmatrix} x_{1(c)} \\ x_{n(c)} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 2 & 3 \\ 0 & 1 & 1 & 1 & 2 & 2 & 2 & 3 \end{bmatrix}$ 으로 표현할 수 있다.

예제

- $p_{x_1,x_2}(x_1,x_2) = \begin{cases} e^{-x_2} & 0 < \mathrm{x} < \mathrm{y} < \infty \\ 0 & else \end{cases}$

1. $M_{x_1,x_2}(t_1,t_2) = \int \int_{-\infty}^{\infty} e^{t_1x_1} e^{t_2x_2} e^{-x_2}\, dx_2\, dx_1$ 이므로

1) $\int \int_{x}^{\infty} e^{t_1x_{1+t_2x_2}-x_2}\, dx_2\, dx_1 = \int_0^{\infty} \left[e^{\infty} \cdot \frac{1}{t_{2-1}} - e^{(t_1+t_2-1)x} \cdot \frac{1}{t_{2-1}} \right] dx$

2) $\int_0^{\infty} \frac{1}{t_{2-1}} \left[1 - e^{(t_1+t_2-1)x} \right] dx = \frac{1}{t_{2-1}} [(e^{\infty} \cdot \frac{1}{t_1-t_{2-1}})$- $(e^{0} \cdot \frac{1}{t_1-t_{2-1}})] = \frac{1}{(t_{2}-1)(t_1-t_{2-1})}$

2. 각각의 확률변수에 대하여 mgf를 확인하면

1) $M_{x_1,x_2}(t_1,t_2) = \begin{bmatrix} M_{x_1}(t_1, 0) \\ M_{x_2}(0, t_2) \end{bmatrix} = \begin{bmatrix} \frac{1}{(1-0)(1-t_1-0\quad)} \\ \frac{1}{(1-t_{2}\quad)(1-0-t_2)} \end{bmatrix} = \begin{bmatrix} \frac{1}{1-t_1} \\ \frac{1}{(1-t_2)^2} \end{bmatrix}$