다항 분포

정의

- 어떤 확률실험을 독립적으로 n회 반복할 때

1. 이 실험들은 K개의 서로 배반인 c_1, $c_2, \cdots, c_k$개의 경우의 수를 갖고 있고, 이 경우의수의 시행 횟수는 n번으로 모두 똑같다.

1) x_1을 c_1에 속하는 실험값으로 두고, x_2, x_3 ,⋯을 각각 $c_2, \cdots, c_k$에 대응하는 실현값이라고 놨을 때,

$$(1)\ f(x_1, x_2, x_3) = \begin{cases} \frac{n!}{x_1!x_2!\cdots x_{k-1}!(n-x_1-\cdots-x_{k-1})!} {p_1}^{x_1} \cdots {p_{k-1}}^{x_{k-1}} (1 - p_1 - \cdots - p_k)^{(n-x_1-\cdots-x_{k-1})} \\ x,\ else \end{cases}$$

(2) 즉, 이는 베르누이 분포의 일반화 형식이다.

예제

- 삼항분포 X = x_1, Y = y_2 일때, MGF를 구하라

1. PDF $f(x_1, x_2, x_3)$를 먼저 구하면

1) $f(x_1, x_2, x_3) = \begin{cases} \frac{n!}{x!y!(n-x-y)!} {p_x}^x {p_y}^y (1 - {p_x}^x - {p_y}^y)^{(n-x-y)} \\ x,\ else \end{cases}$

2) 위 식을 간단하게 만들면

(1) $f(x_1, x_2, x_3) = \begin{cases} [p_x + p_y + (1 - {p_x}^x - {p_y}^y)]^n \\ x,\ else \end{cases}$

2. 적률생성함수 mgf를 구하면

1) M(t1,0) = $[p_x e^{t_1} + p_y + (1 - {p_x}^x - {p_y}^y)]^n = [p_x e^{t_1} + (1 - p_x)]^n$

2) M(0,t2) = $[p_x + p_y e^{t_2} + (1 - {p_x}^x - {p_y}^y)]^n = [p_y e^{t_1} + (1 - p_y)]^n$

예제

- 위에서 구한 삼항분포의 조건부 분포 $p_{y|x}(y|x)$를 구하고, 이를 이항분포와 비교하라

1. $p_{y|x}(y|x) = \dfrac{\frac{n!}{x!y!(n-x-y)!} {p_x}^x {p_y}^y \left(1-{p_x}^x - {p_y}^y\right)^{(n-x-y)}}{\frac{n!}{(n-x)!} \cdot p^x (1-p)^{(n-x)}} = \dfrac{(n-x)!}{y!(n-x-y)!} \cdot \left(\dfrac{p_y}{1-p_x}\right)^y \cdot \left(\dfrac{(1-p_x-p_y)}{1-p_x}\right)^{(n-x-y)}$

1) k= $n-x$, $p = \frac{p_y}{1-p_x}$로 재정의 하면, $\frac{k!}{y!(k-y)!} \cdot p^y \cdot (1-p)^{(k-y)}$

2) 이는 이항분포로 볼 수 있다.

3) 따라서

(1) $E(y|x) = kp = (n-x)\left(\frac{p_y}{1-p_x}\right)$이고

(2) $var(y|x) = kp(1-p) = (n-x)\left(\frac{p_y}{1-p_x}\right)\left(\frac{(1-p_x-p_y)}{1-p_x}\right)$

예제

- 위에서 구한 삼항분포의 조건부 분포 $p_{y|x}(y|x)$를 구하고, 이를 이항분포와 비교하라

2. 위와 같은 논리로,

1) $E(x|y) = kp = (n-x)(\frac{p_x}{1-p_y})$

2) $var(x|y) = kp(1-p) = (n-x)(\frac{p_x}{1-p_y})(\frac{(1-p_x-p_y)}{1-p_y})$

3) 이 때, 상관계수를 선형방정식 꼴로 구하면

(1) $b^2 = \rho^2 \frac{\sigma_y^2}{\sigma_x^2}$ 에서 각각의 $\rho = (\frac{p_x}{1-p_y})$ 또는 $(\frac{p_y}{1-p_x})$ 이므로, 둘을 곱하면

(2) $\rho^2 = (\frac{p_x}{1-p_y})(\frac{p_y}{1-p_x})$ 이고, $\rho = \sqrt{(\frac{p_x}{1-p_y})(\frac{p_y}{1-p_x})}$