



통계학

1. [50pts] (10 pts each)

* 관련 개념 : 이산확률분포, 결합확률분포, 조건부분포, 총확률의법칙, 적률생성함수(mgf)

* 출제 단위 : 기초 수리통계학 (21주차~31주차)

* 난이도 : 상

* 코멘트 : 제시된 pmf 내의 pattern을 활용하여 주변확률분포와 조건부분포를 구하는 문제. 매클로린급수와 이항정리를 역으로 적용하는 것이 수월해야 하며, pmf의 개형을 보고 어떠한 분포인지 유추할 수 있는 능력이 중요하다. 또한 주어진 X,Y의 support를 잘 보고 계산을 수행해야 하므로 여러모로 까다로운 문제이다. 계산이 막힌다면 차분하게 다시 해보거나, 2번 문제를 먼저 풀고 오는 것이 좋을 수 있다.

Let X,Y have the joint pmf

$$f(x, y) = \frac{C}{x!(y-x)!}, \quad y = 0, 1, 2, \dots; \quad x = 0, 1, \dots, y$$

(a) Determine the constant C

$$\text{Sol)} \quad f_Y(y) = \sum_{x=0}^y f(x, y) = \sum_{x=0}^y \frac{C}{x!(y-x)!} = \frac{C}{y!} \sum_{x=0}^y \frac{y!}{x!(y-x)!} = \frac{C}{y!} \sum_{x=0}^y \binom{y}{x} = \frac{C}{y!} 2^y$$

$$\sum_{y=0}^{\infty} f_Y(y) = \sum_{y=0}^{\infty} \frac{C}{y!} 2^y = C \sum_{y=0}^{\infty} \frac{e^{-2} 2^y}{y!} \cdot e^2 = C \cdot e^2 = 1 \quad \Rightarrow \quad C = e^{-2}$$

(5pts) 식을 올바르게 세움

(5pts) $C = e^{-2}$ 를 정확하게 계산



(b) Find the marginal pmf of Y .

$$\text{Sol) } f_Y(y) = \frac{c}{y!} 2^y = \frac{e^{-2} 2^y}{y!} \sim \text{Poi}(2)$$

(5pts) $f_Y(y)$ 를 정확하게 계산

(5pts) $\text{Poi}(2)$ 분포를 기재

(c) Find the conditional pmf of X given $Y = y$.

$$\begin{aligned} \text{Sol) } P(X = x|Y = y) &= \frac{f(x,y)}{f_Y(y)} = \frac{\frac{e^{-2}}{x!(y-x)!}}{\frac{e^{-2} 2^y}{y!}} = \frac{y!}{x!(y-x)!} \left(\frac{1}{2}\right)^y = \frac{y!}{x!(y-x)!} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{y-x} \\ &\Rightarrow X|Y \sim B\left(Y, \frac{1}{2}\right) \end{aligned}$$

(5pts) $P(X = x|Y = y)$ 를 정확하게 계산

(5pts) $B\left(Y, \frac{1}{2}\right)$ 분포를 기재 ($X|Y$ 형태까지 정확하게 기재)

(d) Find the marginal pmf of X using the above results.

$$\begin{aligned} \text{Sol) } f_X(x) &= \sum_{y=0}^{\infty} f(x,y) = \sum_{y=x}^{\infty} \frac{e^{-2}}{x!(y-x)!} = \frac{e^{-2}}{x!} \sum_{y=x}^{\infty} \frac{1}{(y-x)!} = \frac{e^{-2}}{x!} \cdot e = \frac{e^{-1} 1^x}{x!} \\ &\Rightarrow X \sim \text{Poi}(1) \end{aligned}$$

(5pts) $f_X(x)$ 를 정확하게 계산

(5pts) $\text{Poi}(1)$ 분포를 기재



(e) Compute the joint mgf of X and Y .

Sol) $M_{X,Y}(t_1, t_2) = E(e^{t_1 X + t_2 Y}) = E(E(e^{t_1 X + t_2 Y} | Y)) = E(e^{t_2 Y} E(e^{t_1 X} | Y))$

Since $X|Y \sim B\left(Y, \frac{1}{2}\right)$, we have $E(e^{t_1 X} | Y) = M_{X|Y}(t_1) = \left(\frac{1}{2}e^{t_1} + 1 - \frac{1}{2}\right)^Y$

Thus, $M_{X,Y}(t_1, t_2) = E\left(e^{t_2 Y} \left(\frac{1}{2}e^{t_1} + \frac{1}{2}\right)^Y\right) = E\left(\left(\frac{1}{2}(e^{t_1} + 1)e^{t_2}\right)^Y\right) = E(k^Y)$

Finally, $E(k^Y) = \sum_{y=0}^{\infty} k^y \frac{e^{-2} 2^y}{y!} = e^{-2} \sum_{y=0}^{\infty} \frac{(2k)^y}{y!} = e^{-2+2k}$

Therefore, $M_{X,Y}(t_1, t_2) = e^{-2+(e^{t_1}+1)e^{t_2}}$

(5pts) $M_{X,Y}(t_1, t_2)$ 정의식 기재

(5pts) 답이 정확함



2. [50pts] (10 pts each)

* 관련 개념 : 단순선형회귀분석, 최소제곱법, 재모수화(reparameterization), 원점회귀

* 출제 단위 : 회귀분석 기초(13~16), 회귀분석 심화(2014년도 기출)

* 난이도 : 중

* 코멘트 : 원점회귀와 재모수화 두 개념이 합쳐진 문제. 주어진 조건을 잘 받아들이고 최소제곱법을 활용하면 수월하게 풀 수 있다. 회귀분석 기본기인 최소제곱법, 합(Σ) 기호 연산 등이 잘 연마되어 있어야 실수와 소요시간을 줄일 수 있다.

다음과 같은 단순 회귀 모형이 있다.

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad \epsilon_i \sim iid N(0, \sigma^2)$$

Y_i 는 집의 매매 가격(\$1,000 단위), X_i 는 넓이(square feet 단위)이며, 20개의 관측치를 이용하여 다음의 결과를 얻었다.

$$\begin{aligned} \sum_i Y_i &= 10000 & \sum_i (Y_i - \bar{Y})^2 &= 200000 & \sum_i (X_i - \bar{X})(Y_i - \bar{Y}) &= 600000 \\ \sum_i X_i &= 45000 & \sum_i (X_i - \bar{X})^2 &= 4500000 \end{aligned}$$

(a) $Y_i = \alpha + \beta X_i + \epsilon_i$ 의 회귀식에서 α 와 β 의 최소자승 추정치를 구하라. 최소자승 추정치인 $\hat{\beta}$ 가 집 크기와 매매 가격 사이의 관계에서 무엇을 의미하는가?

$$\text{Sol) } \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = 500 - \frac{2}{15} \frac{45000}{20} = 200, \quad \hat{\beta} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2} = \frac{600000}{4500000} = \frac{2}{15}$$

$\hat{\beta}$ 는 넓이 당 매매가격의 상승분의 추정량이다.

(5pts) 최소자승치 정확하게 유도

(5pts) $\hat{\beta}$ 의 의미 정확하게 기재

(b) R^2 을 계산하고, R^2 값이 무엇을 의미하는지 설명하라.



$$\text{Sol) } R^2 = \frac{SSR}{SST} = \frac{\hat{\beta}^2 S_{xx}}{S_{yy}} = \left(\frac{2}{15}\right)^2 \frac{4500000}{200000} = \frac{2}{5}$$

R^2 는 전체 변동량 중 추정회귀식이 설명하는 변동의 비율이다.

(5pts) R^2 값을 정확하게 계산

(5pts) R^2 의 의미 정확하게 기재

- (c) 만약 절편 항을 포함하지 않고, $(Y_i - \bar{Y})$ 를 $(X_i - \bar{X})$ 에 대해 회귀할 때, 정규방정식을 이용하여 기울기 계수의 최소자승추정치를 구하여라.

$$\text{Sol) } \tilde{Y}_i \equiv (Y_i - \bar{Y}), \tilde{X}_i \equiv (X_i - \bar{X}) \Rightarrow \tilde{Y}_i = \beta \tilde{X}_i + \epsilon_i$$

$$\frac{\partial}{\partial \beta} \sum (\tilde{Y}_i - \beta \tilde{X}_i)^2 = \sum (\tilde{Y}_i - \beta \tilde{X}_i)(-2\tilde{X}_i) \Rightarrow \tilde{\beta} = \frac{\sum \tilde{X}_i \tilde{Y}_i}{\sum \tilde{X}_i^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \hat{\beta} = \frac{2}{15}$$

(5pts) 최소제곱법 정확하게 활용하여 식 유도

(5pts) 최소자승치를 정확하게 계산

- (d) 절편항을 포함하지 않고, Y_i 를 $(X_i - \bar{X})$ 에 대해 회귀한다고 했을 때, 정규방정식을 이용하여 기울기 계수의 최소자승추정치를 구하여라

$$\text{Sol) } Y_i = \beta \tilde{X}_i + \epsilon_i \Rightarrow \tilde{\beta} = \frac{\sum \tilde{X}_i Y_i}{\sum \tilde{X}_i^2} = \frac{\sum (X_i - \bar{X}) Y_i}{\sum (X_i - \bar{X})^2} = \hat{\beta} = \frac{2}{15}$$

(5pts) 최소제곱법 정확하게 활용하여 식 유도

(5pts) 최소자승치를 정확하게 계산



- (e) 절편항을 포함하지 않고, $(Y_i - \bar{Y})$ 를 X_i 에 대해 회귀한다고 했을 때, 정규방정식을 이용하여 기울기 계수의 최소자승추정치를 구하여라

$$\text{Sol) } \tilde{Y}_i = \beta X_i + \epsilon_i \Rightarrow \tilde{\beta} = \frac{\sum X_i(Y_i - \bar{Y})}{\sum X_i^2} = \frac{\sum X_i(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2 + \frac{1}{n}(\sum X_i)^2} = \frac{600000}{4500000 + \frac{1}{20}(45000)^2} = \frac{4}{705} \neq \beta$$

(5pts) 최소제곱법 정확하게 활용하여 식 유도

(5pts) 최소자승치를 정확하게 계산



BLIS 편입

본 자료는 BLIS 편입에서 수험생들을 위해 만든 자료로써, 무단배포 및 상업적 이용을 금합니다.

https://blog.naver.com/blis_edu