MULTINOMIAL DISTRIBUTION

- THIS DISTRIBUTION IS A GENERALIZATION OF BERNOULLI DISTRIBUTION.
- THE VALUE OF THE RANDOM VARIABLE (RV) X IS ONE OF K MUTUALLY EXCLUSIVE AND EXHAUSTIVE STATES.

IF $X = x$; $x \in \{s_1, s_2, \ldots, s_K\}$ WITH PROBABILITIES $p_1, p_2, \ldots, p_K$ RESPECTIVELY.

- $$s_i = \begin{cases} 1 & \text{IF } x = s_i \\ 0 & \text{IF } x \neq s_i \end{cases}$$

$$\sum_{i=1}^{K} s_i = 1$$

- THE ASSOCIATED PROBABILITY DISTRIBUTION IS :

$$b_X(x; p_1, p_2, \ldots, p_K) = \prod_{i=1}^{K} p_i^{s_i} \quad ; \quad \sum_{i=1}^{K} p_i = 1$$

- GIVEN A SET OF INDEPENDENT IDENTICALLY DISTRIBUTED (IID) INSTANCES OF X ; $x_1, x_2, \ldots, x_n$;

ESTIMATE p_i ; $1 \leq i \leq K$.

SOLUTION :

$$L(p_1, p_2, \ldots, p_K ; x_1, x_2, \ldots, x_n) = \prod_{t=1}^{n} \prod_{i=1}^{K} p_i^{s_i^t} \quad ; \quad \sum_{i=1}^{K} p_i = 1$$

$$\ln L(p_1, p_2, \ldots, p_K ; x_1, x_2, \ldots, x_n) = \sum_{t=1}^{n} \sum_{i=1}^{K} s_i^t \ln p_i \longleftarrow \text{OBJECTIVE FUNCTION}$$

NEED TO MAXIMIZE $\ln L(\cdot, \cdot, \cdot)$ WITH THE CONSTRAINT $\sum_{i=1}^{K} p_i = 1$

WE USE METHOD OF LAGRANGE MULTIPLIER

$$\mathcal{L}(p_1, p_2, \ldots, p_k ; x_1, x_2, \ldots, x_n ; \lambda)$$

$$= \sum_{t=1}^{n} \sum_{i=1}^{k} s_i^t \ln p_i + \lambda\left(1 - \sum_{i=1}^{k} p_i\right)$$

$$\frac{\partial \mathcal{L}}{\partial p_i} = \sum_{t=1}^{n} s_i^t \cdot \frac{1}{p_i} - \lambda = 0 \Rightarrow p_i = \frac{1}{\lambda} \sum_{t=1}^{n} s_i^t \; ; \quad 1 \le i \le k$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_{i=1}^{k} p_i = 0$$

$$\therefore \; 1 = \frac{1}{\lambda} \sum_{i=1}^{k} \sum_{t=1}^{n} s_i^t = \frac{1}{\lambda} \sum_{t=1}^{n} \underbrace{\sum_{i=1}^{k} s_i^t}_{\Rightarrow = 1} \Rightarrow \lambda = n$$

$$\therefore \; p_i = \frac{1}{n} \sum_{t=1}^{n} s_i^t$$

THAT IS: $\hat{p}_i = \frac{1}{n} \sum_{t=1}^{n} s_i^t \; ; \quad 1 \le i \le k$

THE OBJECTIVE FUNCTION IS CONCAVE, AND THE CONSTRAINT IS LINEAR. THEREFORE $\hat{p}_i$'s MAXIMIZE THE OBJECTIVE FUNCTION, SUBJECT TO THE GIVEN CONSTRAINTS. $\square$

MAXIMUM LIKELIHOOD ESTIMATION FOR LINEAR REGRESSION

- GIVEN A SET OF DATA POINTS (x_i, y_i) ; $1 \le i \le n$
- y_i IS DRAWN FROM A (DISTRIBUTION) RV $N(\mu_{y_i}, \sigma^2)$; $1 \le i \le n$

  NORMAL RV
- ASSUME $\mu_{y_i} = \beta_0 + \beta_1 x_i$; WHERE β_0, β_1 ARE CONSTANTS; $1 \le i \le n$
- THEREFORE y_i IS DRAWN FROM DENSITY FUNCTION

  $f_Y(y; \beta_0 + \beta_1 x_i, \sigma^2)$; $1 \le i \le n$

  PDF OF A NORMAL RV
- THE GOAL IS TO ESTIMATE $\beta_0, \beta_1, \sigma^2$
- ASSUME y_i's ARE INDEPENDENT OF EACH OTHER.

THE ML ESTIMATOR IS

$$L(\beta_0, \beta_1, \sigma^2; x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n)$$

$$= \prod_{i=1}^{n} f_Y(y_i; \beta_0 + \beta_1 x_i, \sigma^2)$$

$$\ln L = \sum_{i=1}^{n} \ln f_Y(y_i; \beta_0 + \beta_1 x_i, \sigma^2)$$

- THE MLE OF β_0, β_1, AND σ^2 IS OBTAINED BY MAXIMIZING $L(\cdot)$
- $f_Y(y_i; \beta_0 + \beta_1 x_i, \sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\dfrac{1}{2\sigma^2}\left\{y_i - (\beta_0 + \beta_1 x_i)\right\}^2\right]$

$\Rightarrow y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, WHERE $\varepsilon_i \sim N(0, \sigma^2)$

- THIS IS A CLASSICAL LINEAR REGRESSION MODEL.

$$\therefore \ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left\{y_i - (\beta_0 + \beta_1 x_i)\right\}^2$$

$$\frac{\partial \ln L}{\partial \beta_0} = -\frac{1}{2\sigma^2}\cdot 2(-1)\sum_{i=1}^{n}\{y_i-(\beta_0+\beta_1 x_i)\} = 0 \qquad \text{— (1)}$$

$$\frac{\partial \ln L}{\partial \beta_1} = -\frac{1}{2\sigma^2}\cdot 2 \sum_{i=1}^{n}\{y_i-(\beta_0+\beta_1 x_i)\}(-x_i) = 0 \qquad \text{— (2)}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2}\frac{(-1)}{\sigma^4}\sum_{i=1}^{n}\{y_i-(\beta_0+\beta_1 x_i)\}^2 = 0 \qquad \text{— (3)}$$

THESE ARE THREE EQUATIONS IN THREE UNKNOWNS: $\beta_0, \beta_1, \sigma^2$.

EQUATION (1) LEADS TO $\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i) = 0$

LET $\sum_{i=1}^{n} x_i = n\bar{x}$; $\sum_{i=1}^{n} y_i = n\bar{y}$

$\therefore\ n\bar{y} - n\beta_0 - n\beta_1\bar{x} = 0 \Rightarrow \beta_0 = \bar{y} - \beta_1\bar{x}$ — (4)

EQUATION (2) LEADS TO

$$\sum_{i=1}^{n} x_i y_i - \beta_0 n\bar{x} - \beta_1\sum_{i=1}^{n} x_i^2 = 0$$

USE OF EQUATION (4) LEADS TO

$$\sum_{i=1}^{n} x_i y_i - (\bar{y}-\beta_1\bar{x})n\bar{x} - \beta_1\sum_{i=1}^{n} x_i^2 = 0$$

$$\beta_1\left\{n\bar{x}^2 - \sum_{i=1}^{n} x_i^2\right\} = n\bar{x}\bar{y} - \sum_{i=1}^{n} x_i y_i$$

$$\beta_1 = \left\{n\bar{x}\bar{y} - \sum_{i=1}^{n} x_i y_i\right\} \Big/ \left\{n\bar{x}^2 - \sum_{i=1}^{n} x_i^2\right\}$$

NOTE THAT

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$\therefore \beta_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

LET $\hat{\beta}_0$, $\hat{\beta}_1$, AND $\hat{\sigma}^2$ BE THE ML ESTIMATES

$$\therefore \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \left\{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right\} \Big/ \left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right\}^2 \quad \text{(VIA EQUATION 3)}$$

NOTE THAT THESE ARE ALSO THE LEAST SQUARES ESTIMATES, WHICH ARE OBTAINED BY MINIMIZING $\sum_{i=1}^{n} \varepsilon_i^2$.

$\square$