Deviance of Binomial Distribution

Monday, March 13, 2023 2:53 PM

$$\hat{\ell}(\hat{p}_{max}; y) = \sum_{i=1}^{N}\left[Y_i \log\left(\frac{Y_i}{n_i}\right) - Y_i \log\left(\frac{n_i - Y_i}{n_i}\right) + n_i \log\left(\frac{n_i - Y_i}{n_i}\right) + \cancel{\log\binom{n_i}{Y_i}}\right]$$

$$\hat{\ell}(\hat{p}^*; y) = \sum\left[Y_i \log\left(\frac{\hat{Y}_i}{n_i}\right) - Y_i \log\left(\frac{n_i - \hat{Y}_i}{n_i}\right) + n_i \log\left(\frac{n_i - \hat{Y}_i}{n_i}\right) + \cancel{\log\binom{n_i}{Y_i}}\right]$$

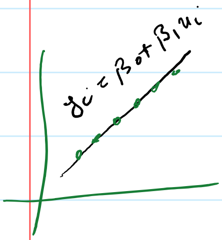
$$\Rightarrow D = 2\left[\ell(\hat{p}_{max}) - \ell(\hat{p}^*)\right] = 2\sum\left[Y_i \log\left(\frac{Y_i}{\hat{Y}_i}\right) - Y_i \underline{\log\left(\frac{n - Y_i}{n - \hat{Y}_i}\right)}\right.$$

$$\left. + n_i \underline{\log\left(\frac{n_i - Y_i}{n_i - \hat{Y}_i}\right)}\right]$$

$$= 2\sum\left[Y_i \log \frac{Y_i}{\hat{Y}_i} + (n_i - Y_i) \log\left(\frac{n - Y_i}{n - \hat{Y}_i}\right)\right]$$

Error term and residuals

Monday, March 13, 2023 2:53 PM



$$Y_i = \boxed{\beta_0 + \beta_1 X_i} + \varepsilon_i$$

fixed term

error

↳ fixed

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$E(Y_i) = \beta_0 + \beta_1 X_i + E(\varepsilon_i) = \beta_0 + \beta_1 X_i$$

$$Var(Y_i) = Var(\varepsilon_i) = \sigma^2$$

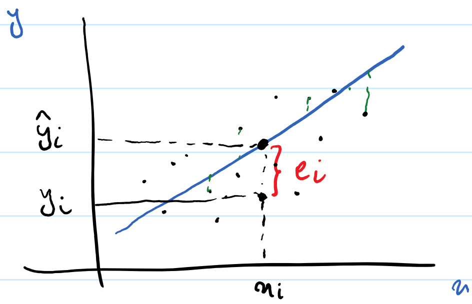
$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

↳ $\in$ exponential family of dist

---

error terms $\varepsilon_i \sim N(0, \sigma^2)$

residuals $e_i = y_i - \hat{y}_i$



e_i is the observed value for ε_i

Estimation of Beta

Monday, March 13, 2023 2:53 PM

$$RSS = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

$$\begin{cases} \dfrac{\partial RSS}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = -2\left[\sum y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum x_i\right] \\ \dfrac{\partial RSS}{\partial \hat{\beta}_1} = -2 \sum \left[x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)\right] = -2\left[\sum x_i y_i - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i^2\right] \end{cases}$$

$$\frac{\partial^2 RSS}{\partial \hat{\beta}_0^2} = (-2)(-n) = 2n$$

$$\Rightarrow \frac{\partial RSS}{\partial \hat{\beta}_0} = 0 \Rightarrow \hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\Rightarrow \frac{\partial RSS}{\partial \hat{\beta}_1} = 0 \Rightarrow \sum x_i y_i - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i^2 = 0$$

$$\Rightarrow \sum x_i y_i - \frac{\sum x_i \sum y_i - \hat{\beta}_1 (\sum x_i)^2}{n} - \hat{\beta}_1 \sum x_i^2 = 0$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \checkmark$$

$$\hat{\beta}_0 \sum x_i \quad \left(\frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n}\right) \sum x_i = \frac{\sum y_i \sum x_i - \hat{\beta}_1 (\sum x_i)^2}{n}$$

$$H = \begin{pmatrix} \dfrac{\partial^2 RSS}{\partial \hat{\beta}_0^2} & \dfrac{\partial^2 RSS}{\partial \hat{\beta}_1 \partial \hat{\beta}_0} \\ \dfrac{\partial^2 RSS}{\partial \hat{\beta}_0 \partial \hat{\beta}_1} & \dfrac{\partial^2 RSS}{\partial \hat{\beta}_1^2} \end{pmatrix} = 2 \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

H is positive definite since $|H| = 4\left(n \sum x_i^2 - (\sum x_i)^2\right) \geq 0$

since $\dfrac{\sum x_i^2}{n} \geq \left(\dfrac{\sum x_i}{n}\right)^2$

$|cA| = c^p |A|$ $\quad$ A is $p \times p$ matrix, $c \neq 0$ constant

Unbiasedness of Beta and its variance

Monday, March 13, 2023 2:53 PM

$$E(\hat{\beta}_1) = E\left(\frac{\sum (x_i-\bar{x})(Y_i-\bar{Y})}{\sum (x_i-\bar{x})^2}\right) = \frac{1}{\sum (x_i-\bar{x})^2}\left[\sum (x_i-\bar{x}) E(Y_i-\bar{Y})\right]$$

$E(\bar{Y})$

$$= \frac{1}{\sum (x_i-\bar{x})^2}\left[\sum (x_i-\bar{x})\left[\underbrace{(\cancel{\beta_0}+\beta_1 x_i)}_{E(Y_i)} - \underbrace{(\cancel{\beta_0}+\beta_1 \bar{x})}_{E(\bar{Y})}\right]\right]$$

$$= \frac{1}{\sum (x_i-\bar{x})^2} \sum (x_i-\bar{x})\, \beta_1 (x_i-\bar{x})$$

$$= \frac{\sum (x_i-\bar{x})^2 \beta_1}{\sum (x_i-\bar{x})^2} = \beta_1$$

Similarly $E(\hat{\beta}_0) = \beta_0$

$$Var(\hat{\beta}_1) = Var\left(\frac{\sum (x_i-\bar{x})(Y_i-\bar{Y})}{\sum (x_i-\bar{x})^2}\right)$$

try to prove

$$= \frac{\sum (x_i-\bar{x})^2\, Var(Y_i-\bar{Y})}{\left[\sum (x_i-\bar{x})^2\right]^2}$$

$$= \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i-\bar{x})^2}\right]\sigma^2 \qquad \sigma^2 = Var(Y_i) = Var(\epsilon_i)$$

Boston Example

Monday, March 13, 2023 2:53 PM

In Boston example

$$\hat{y}_i = \underbrace{34.55}_{\hat{\beta}_0} - \underbrace{0.95}_{\hat{\beta}_1} x_i$$

p-value of $\hat{\beta}_0 = 2e-16 < 0.05 \Rightarrow \beta_0 = 0$ is rejected

p-value of $\hat{\beta}_1 = 2e-16 < 0.05 \Rightarrow \beta_1 = 0$ is rejected

$R^2 = 54.41\%$

Matrix form

Monday, March 13, 2023 2:53 PM

$$Y = X\beta + \varepsilon$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \underset{N \times p}{\begin{bmatrix} 1 & X_{12} & \cdots & X_{1p} \\ 1 & X_{22} & & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{N2} & \cdots & X_{Np} \end{bmatrix}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

$$Y_i = \beta_1 + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

$$i = 1, \ldots, N$$

Score function and Information matrix

Monday, March 13, 2023 4:57 PM

$Y_i = \beta_1 + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$

$E(Y_i) = X\beta = g(\mu_i) = \eta_i$

$\hookrightarrow$ link function

score function $U = \dfrac{\partial l}{\partial \beta}$

- in normal distribution $X\beta = \mu_i = \eta_i$

normal is part of exp family and in exp we have

$$f(Y_1 \dots Y_N \mid \theta_1, \dots, \theta_N) \overset{iid}{=} \prod f(Y_i \mid \theta_i)$$

$$= \prod_{i=1}^{N} \exp\left(Y_i b(\theta_i) + c(\theta_i) + d(Y_i) \right) \quad \text{canonical form}$$

$$\underset{\text{log-likelihood function}}{l} = \ln(L(\theta; Y)) = \ln \prod \exp\left(Y_i b(\theta_i) + c(\theta_i) + d(Y_i)\right)$$

$$= \sum Y_i b(\theta_i) + \sum c(\theta_i) + \sum d(Y_i)$$

$E(Y_i) = \mu_i = -\dfrac{c'(\theta_i)}{b'(\theta_i)}$

$Var(Y_i) = \dfrac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^3}$

chain rule

$$U_i = \frac{\partial l}{\partial \beta_j} = \underbrace{\frac{\partial l}{\partial \theta_i}}_{1} \times \underbrace{\frac{\partial \theta_i}{\partial \mu_i}}_{2} \times \underbrace{\frac{\partial \mu_i}{\partial \beta_j}}_{3}$$

$$\frac{\partial l}{\partial \theta_i} = \sum Y_i b'(\theta_i) + \sum c'(\theta_i)$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \left(\frac{\partial \mu_i}{\partial \theta_i}\right)^{-1} = \left(\left(-\frac{c'(\theta_i)}{b'(\theta_i)}\right)'\right)^{-1}$$

$$= \left[\frac{-c''(\theta_i)}{b'(\theta_i)} + \frac{c'(\theta_i)\, b''(\theta_i)}{[b'(\theta_i)]^2}\right]^{-1}$$

$$= \left[b'(\theta_i)\, Var(Y_i)\right]^{-1}$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial(\beta_1 + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{\partial \beta_j} = x_{ij} \checkmark$$

$$U_i = \sum_{i=1}^{N} \frac{y_i - \mu_i}{Var(Y_i)} X_{ij}$$

$$I_{jk} = E(U_j U_k)$$

Distribution of hat(beta)

Tuesday, March 14, 2023 12:19 PM

If $Y \sim N_a(\mu, \Sigma)$ then $AY \sim N_b(A\mu, A\Sigma A^T)$ if A is a matrix of $b \times a$ dimension

$$A_{b\times a} Y_{a\times 1} \longrightarrow AY \text{ is } b\times 1$$

$$\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{Y}$$

$\hookrightarrow N(X\beta, \sigma^2 \underset{\sim}{I})$

$\downarrow$ identity matrix

Derivative of vectors

Tuesday, March 14, 2023 12:26 PM

Rules

$$\frac{\partial x^T A}{\partial x} = A \qquad \frac{\partial A x}{\partial x} = A^T \qquad \frac{\partial x^T A x}{\partial x} = (A + A^T)x$$

So if A is symmetric ; i.e. $A = A^T$, then $\frac{\partial x^T A x}{\partial x} = 2Ax$

Besides; note that

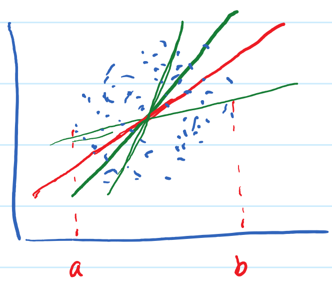
$\underset{1\times p}{\beta^T} \underbrace{\underset{p\times N}{X^T} \underset{N\times 1}{Y}}_{p\times 1}$ it is of dimension 1×1 → constant

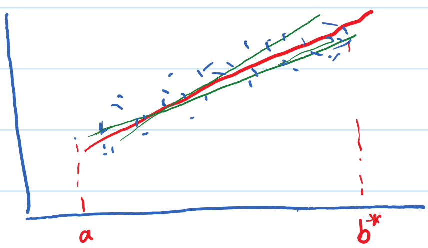
$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 1\times 3 + 2\times 4 = 11$$

$\beta^T X^T Y = Y^T X \beta$

- $X^T X$ is invertable since it is a postive-definite matrix

Slide 92





Tuesday, March 14, 2023 12:46 PM

Slide 94

$\underset{\sim}{r} = (I - H)\underset{\sim}{y}$

we know that $\underset{\sim}{y} \sim N(\underset{\sim}{X}\beta, \sigma^2 I)$

$\Rightarrow E(\underset{\sim}{r}) = (I-H)E(\underset{\sim}{y}) = (I-H)\underset{\sim}{X}\beta = (I - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{X}\beta$

$= \underset{\sim}{X}\beta - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{X}\beta = \underset{\sim}{X}\beta - \underset{\sim}{X}\beta = \underset{\sim}{0}$

$Var(\underset{\sim}{r}) = (I-H)\,Var(\underset{\sim}{y})\,(I-H)^T$

$= (I-H)\sigma^2 I(I-H) = \sigma^2 (I-H)^2 = \sigma^2 (I-H)$

Excercise

$H = X(X^TX)^{-1}X^T$ is idempotent which means that $H^2 = H$.

Show that $(I-H)$ is also idempotent

Tuesday, March 14, 2023 6:49 PM

$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad E(Y) = X\beta$

$\Rightarrow E(\hat{\beta}) = (X^T X)^{-1} X^T E(Y) = \underbrace{(X^T X)^{-1} X^T X}_{I} \beta = \beta$ ✓

$\hat{\beta}$ is unbaised

$\cancel{Var(\hat{\beta}) = (}$

$Var(Y) = \Sigma$ with some matrix A

$Var(AY) = A\, Var(Y) A^T = A \Sigma A^T$

Tuesday, March 14, 2023 6:49 PM

If we drop the i^{th} observation $\hat{\beta}^{-i}$ $\Rightarrow \hat{Y}_i^{-1} = X\hat{\beta}^{-1}$

If we include the i^{th} observation $\hat{Y}_i = X\hat{\beta}$

Minimal Model

Tuesday, March 14, 2023 6:49 PM

In minimal model $y_i = \beta_0 + \varepsilon_i$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{X} \beta_0 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

$X^T X = N$

$X^T y = \sum y_i$

$$\hat{\beta} = (X^T X)^{-1} X^T y = \frac{1}{N} \sum y_i = \bar{y}$$

Tuesday, March 14, 2023 8:36 PM

ANOVA

Model 1

treatments A, B, S

K observations in each treatment

$$
\begin{bmatrix} Y_{A1} \\ \vdots \\ Y_{AK} \\ Y_{B1} \\ \vdots \\ Y_{BK} \\ Y_{S1} \\ \vdots \\ Y_{SK} \end{bmatrix}
=
\underbrace{
\begin{array}{c}
\begin{array}{ccc} A & B & S \end{array} \\
\begin{bmatrix}
1 & 0 & 0 \\
\vdots\ 1 & \vdots\ 0 & \vdots\ 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
\vdots\ 0 & \vdots\ 1 & \vdots\ 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\vdots\ 0 & \vdots\ 0 & \vdots\ 1 \\
0 & 0 & 1
\end{bmatrix}
\end{array}
}_{X}
\begin{bmatrix} \mu_A \\ \mu_B \\ \mu_S \end{bmatrix}
$$