Other multinomial with functional probabilities

Example: Logistic Regression

Suppose we observe binary data $Y_1, Y_2, \ldots, Y_n$, which we assume are independent. We also have access to p-dimensional "explanatory" variables, X_i, that are connected to Y_i in some way.

Propose a model for using X_i to predict if $Y_1 = 0$ or $Y_i = 1$.

p_1, p_2, $p_i(x_i)$

$Y_i \sim \mathrm{Bern}(p_i)$, $p_i := p(x_i, \theta)$.

$$L(\theta) = \prod_{i=1}^{n} f_{Y_i|X_i}(y_i \mid x_i; \theta)$$

$$= \prod_{i=1}^{n} p_i(x_i)^{y_i} \left(1 - p_i(x_i)\right)^{1-y_i}$$

$$\ell(\theta) = \sum_{i=1}^{n} y_i \log\left(p_i(x_i)\right) + (1-y_i) \log\left(1 - p_i(x_i)\right)$$

$$= \sum_{i=1}^{n} y_i \log\left(p_i(x_i)\right) + \log\left(1-p_i\right) - y_i \log\left(1-p_i\right)$$

$$= \sum_{i=1}^{n} y_i \log\left(\frac{p_i}{1-p_i}\right) + \log\left(1-p_i\right)$$

$$p_i \underset{\text{"we would like"}}{\approx} \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}$$

logit function: $f(z) = \dfrac{e^z}{1+e^z}$, $z \in \mathbb{R}$

$\downarrow$

$(0,1)$

$$p_i = \frac{e^{(\beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p})}}{1 + e^{\dots}} = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$$

"link function" = $f(z)$ = "logit-link"

GLM Generalized linear models.

$$\ell(\theta) = \sum_{i=1}^{n} y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)$$

$$= \sum_{i=1}^{n} y_i \log\left(\frac{\left(\frac{e^{x_i^T\beta}}{1-e^{x_i^T\beta}}\right)}{1-\left(\frac{e^{x_i^T\beta}}{1-e^{x_i^T\beta}}\right)}\right) + \log\left(1 - \frac{e^{x_i^T\beta}}{1-e^{x_i^T\beta}}\right)$$

$$\underline{1-p_i = \frac{1}{1+e^{x_i^T\beta}}} \; , \quad \underline{\frac{p_i}{1-p_i}} = e^{x_i^T\beta}$$

$$= \sum_{i=1}^{n} y_i \log\left(e^{x_i^T\beta}\right) - \log\left(1+e^{x_i^T\beta}\right)$$

$$l(\theta) = \sum_{i=1}^{n} y_i x_i^T \beta - \log\left(1 + e^{x_i^T \beta}\right)$$

$$\nabla_\beta l(\beta) = \nabla_\beta \sum_{i=1}^{n} y_i x_i^T \beta - \log\left(1 + e^{x_i^T \beta}\right)$$

$$= \sum_{i=1}^{n} y_i x_i - \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} x_i$$

$$= \sum_{i=1}^{n} \underline{y_i x_i} - p_i x_i$$

$$= \underset{n \times p}{X}^T \underset{n \times 1}{y} - X^T \underset{n \times 1}{p}$$

p_i's.

$$= X^T(y-p) \leftarrow \nabla_\beta \ell(\beta)$$

$$\nabla^2_\beta \ell(\beta) = \nabla_\beta \Big(X^T \big(y - p(\beta)\big) \Big)$$

$$= \underline{-X^T \Gamma X}$$

$$\Gamma = \begin{pmatrix} p_1(1-p_1) & 0 & \cdots & 0 \\ 0 & p_2(1-p_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & p_n(1-p_n) \end{pmatrix}$$

$$\theta_n = \theta_{n-1} - (H)^{-1} \nabla \ell(\theta)$$

$$\beta_{(n+1)} = \beta_{(n)} - \left(- X^T \Gamma_{\beta_{(n)}} X \right)^{-1} X^T \left(y - p_{\beta_{(n)}} \right)$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

References and Acknowledgements

Casella G, Berger R (2024). *Statistical inference*. Chapman and Hall/CRC.

Le Cam L (1990). “Maximum likelihood: an introduction.” *International Statistical Review/Revue Internationale de Statistique*, pp. 153–171.

Pawitan Y (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

Rice JA (2007). *Mathematical statistics and data analysis*, volume 371. 3 edition. Thomson/Brooks/Cole Belmont, CA.

References and Acknowledgements II

- Compiled on February 5, 2026 using R version 4.5.2.
-
- We acknowledge students and instructors for previous versions of this course / slides.