Poisson regression

Counts data

Examples of counts data:

- Number of hospitalizations over a period of time
- Number of passengers in a bus station
- Blood cells number in a blood sample
- Number of typos in a book
- . . . . . .

Example: tortoise species data

The Galapagos Islands off the coast of Ecuador are great locations for studying the factors that influence the development and survival of different life species. The data set provides counts for the total number of tortoise species, and the number of species that occur only on that one island (the endemics) (Johnson and Raven, 1973).

Example: tortoise species data

This data set also contains the following geographic variables:

- Area: area in square km;
- Elevation: elevation in meters;
- Nearest: distance from nearest island;
- Scruz: distance from Santa Cruz (which is near the center of the Galapagos);
- Adjacent: area of adjacent island in square km.

Poisson distribution for counts data

- Poisson distribution can be defined via a counting process with the following properties:
  1. The expected number of events occurring in an interval of time is proportional to the length of the interval.
  2. The probability that two events occurring in an infinitely small interval is 0.
  3. The number of events occurring in separate intervals are independent.
- Poisson is a good approximation of Binomial distributed data when the total number of trials is large and small success probability.

Poisson regression

Assume that the response Y_i is a count, where Y_i could taking values 0,1,2,$\cdots$. The distribution of Y_i may be modelled by the Poisson distribution with mean μ_i. That is

$$Y_i \sim \text{Poisson}(\mu_i),$$

which has the pmf $f(y) = \exp(-\mu)\mu^y/y!$ for $y = 0, 1, 2\cdots$.

Here $\mu > 0$.

Link function

One common link function used for the Poisson regression is the log function. That is

$$\log(\mu_i) = X_i^T \beta,$$

where X_i is a p-dim predictor and β is a p-dim unknown coefficients. The link function implies that

$$\mu_i = \exp(X_i^T \beta).$$

Maximum likelihood estimator

The log-likelihood function of β is

$$\ell(\beta) = \log\{\prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i^{Y_i}}{Y_i!}\} = \sum_{i=1}^{n} Y_i \log(\mu_i) - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} \log(Y_i!)$$

$$= \sum_{i=1}^{n} Y_i X_i^T \beta - \sum_{i=1}^{n} \exp(X_i^T \beta) - \sum_{i=1}^{n} \log(Y_i!).$$

The the MLE for β is

$$\hat{\beta} = \arg\max_{\beta} \Big[\sum_{i=1}^{n} Y_i X_i^T \beta - \sum_{i=1}^{n} \exp(X_i^T \beta) \Big].$$

Score function and hessian matrix

- The score function is

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} \{Y_i - \exp(X_i^T \beta)\} X_i.$$

- The MLE $\hat{\beta}$ is a solution of $\partial \ell(\beta)/\partial \beta = 0$.
- The Hessian matrix is

$$\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^T} = -\sum_{i=1}^{n} X_i X_i^T \exp(X_i^T \beta) = -X^T V X,$$

where $X = (X_1, \cdots, X_n)^T$ is an $n \times p$ design matrix and $V = \text{diag}\{\exp(X_1^T \beta), \cdots, \exp(X_n^T \beta)\}$.

Asymptotic normality of $\hat{\beta}$

Applying the large sample theory of the maximum likelihood estimator $\hat{\beta}$, we have

$$\hat{\beta} - \beta \sim N(0, (X^T V X)^{-1}).$$

Wald type inference for β could be based on the asymptotic normality.

Deviance

- The log-likehood for μ_i in a saturated model is

$$\ell(\mu_i) = \sum_{i=1}^{n} \{ Y_i \log(Y_i) - Y_i \} + Const..$$

- The log-likelihood for μ_i is the full model with

  $\mu_i = \exp(X_i^T \beta)$ is

$$\ell(\beta) = \sum_{i=1}^{n} \{ Y_i \log(\hat{\mu}_i) - \hat{\mu}_i \} + Const..$$

  where $\hat{\mu}_i = \exp(X_i^T \hat{\beta})$ and $\hat{\beta}$ is the MLE of β.

- The deviance is then defined as

$$D = 2 \sum_{i=1}^{n} \{ Y_i \log(Y_i / \hat{\mu}_i) - (Y_i - \hat{\mu}_i) \}.$$

Some remarks

- The likelihood ratio type inference could be conducted based on the deviance.
- The analysis of deviance can be done as that in logistic regression model.
- The model diagnostic and residual plots could be also done similarly as those in logistic regression model.

Over or under dispersion

- In poisson regression model, we assume that

$$E(Y_i) = \text{Var}(Y_i) = \mu_i.$$

Note that the mean and variance are the same. This might not be flexible in practice.

- A generalization of the Poisson regression model is

$$E(Y_i) = \mu_i \text{ and } \text{Var}(Y_i) = \phi\mu_i,$$

where ϕ is the dispersion parameter.

Quasi-likelihood

- Similar to the logistic regression model, the quasi log-likelihood for β can be defined as

$$Q(\beta) = \sum_{i=1}^{n} \int_{Y_i}^{\mu_i} \frac{Y_i - \mu}{\phi V(\mu)} d\mu$$

  where $V(\mu) = \mu$ and $\mu_i = \exp(X_i^T \beta)$.
- The estimation of β is the same as the usual poisson regression without dispersion parameter.
- The asymptotic normality of $\hat{\beta}$ is $\hat{\beta} - \beta \sim N(0, \phi(X^T V X)^{-1})$.

Estimation of dispersion parameter

The dispersion parameter ϕ can be estimated by

$$\hat{\phi} = \frac{\sum_{i=1}^{n}(Y_i - \hat{\mu}_i)/\hat{\mu}_i}{n-p}.$$

where $\hat{\mu}_i = \exp(X_i^T \hat{\beta})$.