Topic 2-2: Likelihoods for Regression Models

Department of Experimental Statistics
Louisiana State University

Date

Motivation Data

- The annual maximum sea levels in Venice for 1931–1981 are given in Pirazzoli (1982).
- The data are as follows:

Year	1	2	3	4	5	6	7	8	9	10	11	12	13
Levels	103	78	121	116	115	147	119	114	89	102	99	91	97
Year	14	15	16	17	18	19	20	21	22	23	24	25	26
Levels	106	105	136	126	132	104	117	151	116	107	112	97	95
Year	27	28	29	30	31	32	33	34	35	36	37	38	39
Levels	119	124	118	145	122	114	118	107	110	194	138	144	138
Year	40	41	42	43	44	45	46	47	48	49	50	51	
Levels	123	122	120	114	96	125	124	120	132	166	134	138	

- In this dataset, we are interested in building a **regression model** for describing the relationship between Year as a *covariate* and See Levels as the *response*.
- What is the main difference between the structure of the dataset and that of datasets in previous section? Do we have response in previous datasets?

Linear Models

- Suppose we observed p covariates from each observation $\mathrm{x}_i, i = 1, \cdots, n$ and the associated response y_i. A normal linear model is the most common regression model for describing the relationship between x_i and y_i that can be expressed as

$$Y_i = x_i^T \beta + e_i, \tag{1}$$

  where $e_1, \cdots, e_n$ are iid $N(0, \sigma^2)$, and $\boldsymbol{\beta}$ are an unknown p-dimensional vector.
- The first component of x_i is usually the constant "1" corresponding to an intercept, the first component of $\boldsymbol{\beta}$.
- In this section, we are interested in constructing likelihood function for unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$ and extend the normal linear model to non-normal models.

Likelihood for Normal Linear Models

- The likelihood of data $\{\mathrm{y}_i, \mathrm{x}_i\}_{i=1}^n$ from the normal linear models (1) is

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma | \{\mathrm{y}_i, \mathrm{x}_i\}_{i=1}^n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \mathrm{x}_i^T \beta)^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{i=1}^n \frac{(y_i - \mathrm{x}_i^T \beta)^2}{2\sigma^2}\right)
\end{aligned}
$$

- Based on the likelihood, the maximum likelihood estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_{MLE} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathrm{Y}$, where $\mathrm{X} = (x_1, \cdots, x_n)^T$ an $n \times p$ matrix, and $\mathrm{X}^T\mathrm{X}$ is assumed to be nonsigular.
- The MLE of $\sigma^2 = \sum_{i=1}^n \hat{e}_i$, where $\hat{e}_i = y_i - \mathrm{x}_i^T \hat{\boldsymbol{\beta}}_{MLE}$ is the residual from i-th data for $i = 1, \cdots, n$. (Check the MLEs by yourself!)

Nonnormal Error Models

- Instead of the assumption of putting the error term from model (1) following a normal distribution, we may assume the error term follows other distributions.
- A popular choice is to assume e_i following a scale-family distribution whose density is

$$\frac{1}{\sigma} f_e \left(\frac{y_i - \mathrm{x}_i^T \boldsymbol{\theta}}{\sigma} \right); \tag{2}$$

  for example, the extreme value density is one of the distribution in the scale family and its density is $f_e(t) = \exp(-t) \times \exp(-\exp(-t))$.
- The likelihood function becomes

$$\prod_{i=1}^{n} \frac{1}{\sigma} f_e \left(\frac{y_i - \mathrm{x}_i^T \boldsymbol{\theta}}{\sigma} \right)$$

Revisit the Sea Level data

-

R Code

-

Additive Errors Nonlinear Model

- The standard nonlinear regression model is very similar to (1) but considering $Y_i = g(\mathrm{x}_i, \boldsymbol{\beta}) + e_i$, where g is a known function with unknown parameters $\boldsymbol{\beta}$.
- Common examples include
  - Exponential growth model $g(\mathrm{x}_i, \boldsymbol{\beta}) = \beta_0 \exp(\beta_1 x_i)$
  - logistic growth model $g(x_i, \boldsymbol{\beta}) = \beta_0(1 + \beta_1 \exp(-\beta_2 x_i))$
- The MLE is from maximizing

$$L(\boldsymbol{\beta}, \sigma | \{\mathrm{y}_i, \mathrm{x}_i\}_{i=1}^n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{i=1}^n \frac{(y_i - g(x_i, \boldsymbol{\beta}))^2}{2\sigma^2}\right)$$

- The MLE of $\boldsymbol{\beta}$ has no closed form but the MLE of σ^2 is $\sum_{i=1}^n \hat{e}_i$, where $\hat{e}_i = y_i - g(x_i, \hat{\boldsymbol{\beta}}_{MLE})$ for $i = 1, \cdots, n$.

Motivation Data for GLM

-

Generalized Linear models

- Generalized linear models introduced by Nelder and Wedderburn (1972) are another important class of nonlinear models that generalizes the normal linear model.
- It assumes the log density of Y_i possess the form:

$$\log f(y_i; \theta_i, \phi) = \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi). \tag{3}$$

- The density of Y_i is almost an exponential family density except for the dispersion term $a_i(\phi)$, where a_i is a known function and ϕ is possibly an unknown parameter. In exponential family language, θ_i is called the natural or canonical parameter. (Explain this language)

Generalized Linear models

Generalized linear models are introduced by Nelder and Wedderburn (1972). They are another important class of nonlinear models that generalizes the normal linear model and is made up of a linear predictor

$$\eta_i = \mathrm{x}_i^T \beta$$

and two functions:

- A link function that describes how the mean, $\mu = E(Y_i)$, depends on the linear predictor

$$g(\mu_i) = \eta_i = \mathrm{x}_i^T \boldsymbol{\beta}$$

- A variance function that describes how the variance depends on the mean

$$Var(Y_i) = V(\mu)a(\phi),$$

where $a(\phi)$ is a function of dispersion parameter ϕ.

Example 1 Normal Linear Models

- In normal linear model (1), $Y_i \sim N(\mathrm{x}_i^T \boldsymbol{\beta}, \sigma^2)$.
- The link function is identity link function, i.e. $g(\cdot) = 1$, and

$$\mu = \mathrm{x}_i^T \boldsymbol{\beta}$$

- The variance function is 1.

Example 2: Logistic Linear Models

- For binary data, Y_i is following a Bernoulli distribution with parameter $E(Y_i) = p_i$ for $i = 1, \cdots, n$.
- A popular link function is called logistic link, $g(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$, i.e,

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathrm{x}_i^T \boldsymbol{\beta}$$

- Variance function $V(p_i) = p_i(1-p_i)$ and $a_i(\phi) = 1$ (because $Var(Y_i)) = p_i(1-p_i)$

General Exponential Family

- The normal and bernoulli density belongs to a family of density functions called general exponential family. If Y_i is from the family, its log density can be expressed as

$$\log f(y_i; \theta_i, \phi) = \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi), \tag{4}$$

where a_i is a known function of dispersion parameter ϕ.

Example of Exponential Family: Normal Distribution

- Consider the normal density

$$
\begin{aligned}
f(y;\mu,\sigma) &= -\log\left(\sqrt{2\pi}\sigma\right) - \frac{(y-\mu)^2}{2\sigma^2} \\
&= \frac{y\mu - \mu^2/2}{\sigma^2} - \log\left(\sqrt{2\pi}\sigma\right) - \frac{y^2}{2\sigma^2}. \qquad (5)
\end{aligned}
$$

- Thus,

$$\theta = \mu, b(\theta) = \frac{\mu^2}{2}, a_i(\phi) = \sigma^2, \text{ and } c(y_i,\phi) = \log(\sigma\sqrt{2\pi}) - \frac{y^2}{2\sigma^2}.$$

Example of Exponential Family: Bernoulli Distribution

- Consider the Bernoulli density

$$f(y;p) = p^y(1-p)^{1-y}$$

and we want to show that the density is one of the densities in the family (4).

- Because

$$\log f(y;p) = y\log p + (1-y)\log(1-p) = y\log(\frac{p}{1-p}) + \log(1-p),$$

we obtain $a_i(\phi) = 1, c(y_i, \phi) = 0, theta = \log\{\frac{p}{1-p}\}$.

- Hence, $p = 1/\{1+\exp(-\theta)\}$, so that

$$b(\theta) = -\log(1-p) = -\log\left\{-\frac{1}{1+\exp(-\theta)}\right\} = \log\{1+\exp(\theta)\}.$$

Mean and Variance of the Exponential Family

- For Y_i from a general exponential faimily dentisty (4)

$$\log f(y_i; \theta_i, \phi) = \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi),$$

its mean and variance satisfies the following relationships:

  - $\mu = E(Y_i) = b'(\theta_i)$
  - $Var(Y_i) = b''(\theta_i)a_i(\phi)$.
  Note that since the variance must be positive, $b(\theta_i)$ is a strictly convex function and $b'(\theta_i)$ is monotone increasing with a unique inverse b'^{-1}.

- The relationship is obtained from the fact that
  - $E\left\{\frac{\partial}{\partial\theta_i} \log f(Y_i; \theta_i, \phi)\right\} = 0$
  - $E\left\{\frac{\partial}{\partial\theta_i} \log f(Y_i; \theta_i, \phi)\right\}^2 = E\left\{\frac{\partial^2}{\partial\theta_i^2} \log f(Y_i; \theta_i, \phi)\right\} = 0.$
  See page 54 of the textbook for details.

Canonical Link Functions for GLM

- Fro a GLM where the response follows a general exponential distribution, we have

$$g(\mu_i) = b'^{-1}(\mu_i) = \mathrm{x}_i^T \boldsymbol{\beta}$$

- The link $g(\cdot) = b'^{-1}(\cdot)$ is called the canonical link.
- Canonical links lead to desirable statistical properties of the GLM, so it tends to be used as a default link.

Likelihood for the GLM

- $$\log L(\boldsymbol{\beta}, \phi | \{y_i, \mathrm{x}_i\}_{i=1}^n) \sum_{i=1}^n \left\{ \frac{y_i \mathrm{x}_i^T \boldsymbol{\beta} - b(\mathrm{x}_i^T \boldsymbol{\beta})}{a_i(\phi)} + c(y_i, \phi) \right\}$$

Examples

-

Generalized Linear Mixed Model

- $$\log f_{Y_i|U_i}(y_i|\mathrm{U}, \mathrm{x}_i, \mathrm{z}_i, \boldsymbol{\beta}, \tau) = \frac{y_i\eta_i - b(\eta_i)}{\tau} + c(y_i, \tau),$$

  where $\eta_i = \mathrm{x}_i^T\boldsymbol{\beta} + \mathrm{z}_i^T\mathrm{U}$ is the linear predictor now enhanced to include the random effects U via the terms $\mathrm{z}_i^T\mathrm{U}$
- In Normal distribution case,

  $$\mathrm{Y} = \mathrm{X}\boldsymbol{\beta} + \mathrm{ZU} + e$$

  here written in the familiar vector form $\mathrm{X} = (\mathrm{x}_1^T, \cdots, \mathrm{x}_n^T)$ and $\mathrm{X} = (\mathrm{z}_1^T, \cdots, \mathrm{z}_n^T)$

  $$\int \prod_{i=1}^{n} f_{Y_i|\mathrm{U}} f(\mathrm{u}) d\mathrm{U}$$

Accelerated Failure Model

- Accelerated Failure Models comprise an important class of regression models for censored data. For the ease of random right censoring, the model is

$$\log T_i = \mathrm{x}_i^T \boldsymbol{\beta},$$

where we observe $Y_i = \min\left(\log T_i, \log R_i\right)$, and R_i is a censoring time that is assumed independent of T_t

- The typical models for the errors e_i are standard normal, logistic, or the logatithm of a standard exponential random variabl whose density is $f_e = e^z e^{-e^z}$.
- The likelihood for parameters $\boldsymbol{\beta}$ and σ with using error density f_e and distribution F_e is

$$L(\boldsymbol{\beta}, \sigma | \{y_i, \delta_i, \mathrm{x}_i\}_{i=1}^n) = \prod_{i=1}^n \left[\frac{1}{\sigma} f_e(r_i)\right]^{\delta_i} \left[1 - F_e(r_i)\right]^{\delta_i},$$

where $\delta_i = I(y_i = \log t_i)$ and $r_i = (y_i - \mathrm{x}_i^T \boldsymbol{\beta})/\sigma$.