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 $\boldsymbol{x}_i, i = 1, \cdots, n$ and the associated response y_i. A normal linear model is the most common regression model for describing the relationship between $\boldsymbol{x}_i$ 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 $\boldsymbol{x}_i$ 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 $\{\boldsymbol{y}_i, \boldsymbol{x}_i\}_{i=1}^n$ from the normal linear models (1) is

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

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

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

  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

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

Revisit the Sea Level data

-

R Code

-

Additive Errors Nonlinear Model

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

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

- 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 with parameters θ_i possess the form:

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

- $a_i(\cdot)$ is a known function and ϕ is possibly an unknown parameter called dispersion parameters.
- Densities satisfying form (3) belong a generalized exponential family. For the students who are not familiar with exponential faimily, please check Appendix B of Section 2 in the Textbook.
- GLM can be used for regression modeling of various type responses, including continuous and categorical responses.

Example of the Exponential Family: Normal Distribution

- Consider Y is a continuous random variable the normal density

$$
\begin{aligned}
f(y;\mu,\sigma) &= -\log\left(\sqrt{2\pi}\sigma\right) - \frac{(y-\mu)^2}{2\sigma^2} \\
&= \frac{y\mu - \mu^2/2}{\sigma^2} - \log\left(\sqrt{2\pi}\sigma\right) - \frac{y^2}{2\sigma^2}. \qquad (4)
\end{aligned}
$$

- Thus,

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

Example of the Exponential Family: Bernoulli Distribution

- Consider Y is a binary random variable from Bernoulli density

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

where $E(Y) = p$.

- Because

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

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

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

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

Example of the Exponential Family: Poisson Distribution

- Consider Y is a random variable from Poisson density

$$f(y;p) = \frac{\theta^y e^{-\lambda}}{y!},$$

where $E(Y) = \theta$.

- Because

$$\log f(y;p) = y \log \lambda - \lambda - \log(y!),$$

we obtain

$$\theta_i = \log(\lambda_i), b(\theta_i) = e^{\theta_i} = \lambda_i, a_i(\phi) = 1, c(y_i, \phi) = -\log(y_i!).$$

Mean and Variance of the Exponential Family

- For Y_i from a general exponential faimily dentisty (3)

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

  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, besides selecting a distribution to model response variable Y, we need to choose a link function $g(\cdot)$ to connect the mean of Y with the linear predictor $\boldsymbol{x}_i\boldsymbol{\beta}$, i.e.,

$$g(E(Y)) = \boldsymbol{x}_i\boldsymbol{\beta}$$

- In general, the link can be any function who is invertible. Then, for Y is from density (3), we know

$$g(b'(\theta_i)) = g(E(Y_i)) = \boldsymbol{x}_i\boldsymbol{\beta}.$$

- If we further choose $g(\cdot) = b'^{-1}(\cdot)$, then

$$g(E(Y_i)) = g(b'(\theta_i)) = \theta_i = \boldsymbol{x}_i\boldsymbol{\beta}$$

- 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 in GLM.

Examples of Canonical Links

Density of Y	$E(Y)$	θ	$g(E(Y))$
Normal	$E(Y) = \mu$	$\theta = \mu$	$\mu = \boldsymbol{x}^T\boldsymbol{\beta}$
Bernoulli	$E(Y) = p$	$\theta = \log(\frac{p}{1-p})$	$\log(\frac{p}{1-p}) = \boldsymbol{x}^T\beta$
Poisson	$E(Y) = \lambda$	$\theta = \log(\lambda)$	$\log(\lambda) = \boldsymbol{x}^T\boldsymbol{\beta}$

Likelihood for the GLM

- By using the canonical link in GLM, the log likelihood of the dataset $\{(y_i, x_i)\}_{i=1}^n$ is

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

Generalized Linear Mixed Model (GLMM)

- So far we have considered only fixed effect models in which the input $\boldsymbol{x}$ are fixed in advance. For example, in the Venice dataset, the input year is "fixed" from 1 to 51. If instead the year effect is said to have arisen from a statistical distribution, then we have a **random effects** model.
- GLMM extends GLM to include not only fixed effects but also random effect terms in the inputs. Such model considers Y_i whose log density can be expressed as

$$\log f_{Y_i|\boldsymbol{U}}(y_i|\boldsymbol{U}, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta}, \phi) = \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)$$

  and

$$\theta_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{U},$$

  where $\boldsymbol{U}$ is a random effect vector term following a density $f_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{v})$ with parameters $\boldsymbol{v}$ and $\boldsymbol{z}_i$ are known nonrandom predictors

GLMM (Cont.)

- In the normal distribution case, GLMM becomes the usual linear mixed model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{e}$, where $\mathrm{X} = (\mathrm{x}_1^T, \cdots, \mathrm{x}_n^T)$ and $\mathrm{Z} = (\mathrm{z}_1^T, \cdots, \mathrm{z}_n^T)$.
- Suppose $Y_1, \cdots, Y_n$ are conditionally independent from (17), then the likelihood of GLMM is

$$L(\boldsymbol{\beta}, \phi, \boldsymbol{v}|\{y_i, x_i, z_i\}_{i=1}^n) = \int \prod_{i=1}^n \log f_{Y_i|\boldsymbol{U}}(y_i|\boldsymbol{U}, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\beta}, \phi) f_{\boldsymbol{U}}(\boldsymbol{u}|\boldsymbol{v}) d\boldsymbol{u}$$

- Note integration over $\boldsymbol{u}$ is required because U is not ovserved, and the likelihood is the density of the observed data.
- In general, the likelihood is not easy to calculate or maximize. Breslow and Clayton (1993) and McCullagh (1997) give some of the important approaches for finding the maximum likelihood estimators, but this remains an active area of research.

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

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

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 variable 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

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

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

Example for Accelerated Failure Model

-