Lecture 4: GLM: Estimation

Maochao Xu

Department of Mathematics
Illinois State University
mxu2@ilstu.edu

Failure times

The data in the following table are the lifetimes (times to failure in hours) of Kevlar epoxy strand pressure vessels at 70% stress level.

Lifetimes of $N = 49$ pressure vessels.

1051	4921	7886	10861	13520
1337	5445	8108	11026	13670
1389	5620	8546	11214	14110
1921	5817	8666	11362	14496
1942	5905	8831	11604	15395
2322	5956	9106	11608	16179
3629	6068	9711	11745	17092
4006	6121	9806	11762	17568
4012	6473	10205	11895	17568
4063	7501	10396	12044	

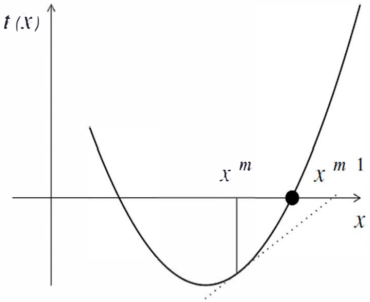
Weibull distribution

The probability density function is

$$f(y; \lambda, \theta) = \frac{\lambda y^{\lambda - 1}}{\theta^{\lambda}} \exp\left\{ -\left(\frac{y}{\theta} \right)^{\lambda} \right\}, \quad y > 0.$$

- Exponential family
- MLE
- Newton-Raphson algorithm
- Fisher scoring

Newton-Raphson algorithm



Comparison: Newton-Raphson and Fisher Scoring

The maximum likelihood estimator is

$$\frac{dl}{d\theta} = U = \sum_{i=1}^{N} \left[\frac{-\lambda}{\theta} + \frac{\lambda y_i^{\lambda}}{\theta^{\lambda+1}} \right].$$

For maximum likelihood estimation using the score function, the estimating equation is

$$\theta^{(m)} = \theta^{(m-1)} - \frac{U^{(m-1)}}{U'^{(m-1)}}.$$

This is the **Newton-Raphson** formula for obtaining the solution.
For maximum likelihood estimation, it is common to approximate U' by its expected value $E(U')$. This is called the method of **Fisher Scoring**. Then, we have

$$\theta^{(m)} = \theta^{(m-1)} + \frac{U^{(m-1)}}{\mathcal{J}^{(m-1)}}.$$

Comparison

Iteration	1	2	3	4
θ	8805.7	9633.9	9876.4	9892.1
$U \times 10^6$	2915.10	552.80	31.78	0.21
$U' \times 10^6$	−3.52	−2.28	−2.02	−2.00
$\mathrm{E}(U') \times 10^6$	−2.53	−2.11	−2.01	−2.00
U/U'	−827.98	−242.46	−15.73	−0.105
$U/\mathrm{E}(U')$	−1152.21	−261.99	−15.81	−0.105

The final estimate is $\theta^{(5)} = 9892.1 - (-0.105) = 9892.2$. This is the maximum likelihood estimate for these data. At this value the log-likelihood function is $l = -480.850$.

Note that **for canonical exponential families the method of scoring and the method of Newton-Raphson coincide.**

Maximum likelihood estimation

Consider independent random variables $Y1, ..., Y_N$ satisfying the properties of a generalized linear model. We wish to estimate parameters $\boldsymbol{\beta}$,

$$E(Y_i) = \mu_i, \quad g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}.$$

For each Y_i, the log-likelihood function is

$$l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i).$$

To obtain the maximum likelihood estimator for the parameter β_j, we have

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}.$$

That is,

$$U_j = \sum_{i=1}^{N} \left[\frac{y_i - \mu_i}{\mathrm{Var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right],$$

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

Fisher Scoring for MLE

The variance–covariance matrix of the U_j's has terms

$$\mathfrak{I}_{jk} = E\left[U_j U_k\right],$$

which form the **information matrix** $\mathfrak{I}$.

$$\begin{aligned}\mathfrak{I}_{jk} &= \mathrm{E}\left\{\sum_{i=1}^{N}\left[\frac{(Y_i-\mu_i)}{\mathrm{var}(Y_i)}x_{ij}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\right]\sum_{l=1}^{N}\left[\frac{(Y_l-\mu_l)}{\mathrm{var}(Y_l)}x_{lk}\left(\frac{\partial \mu_l}{\partial \eta_l}\right)\right]\right\} \\ &= \sum_{i=1}^{N}\frac{\mathrm{E}\left[(Y_i-\mu_i)^2\right]x_{ij}x_{ik}}{\left[\mathrm{var}(Y_i)\right]^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\end{aligned}$$

It can be further simplified to

$$\mathfrak{I}_{jk} = \sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{\mathrm{var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

The estimating Equation for the method of scoring generalizes to

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + \left[\mathfrak{I}^{(m-1)}\right]^{-1}\mathbf{U}^{(m-1)},$$

that is,

$$\mathfrak{J}^{(m-1)}\mathbf{b}^{(m)} = \mathfrak{J}^{(m-1)}\mathbf{b}^{(m-1)} + \mathbf{U}^{(m-1)}. \tag{1}$$

Note that $\mathfrak{J}$ can be written as

$$\mathfrak{J} = \mathbf{X}^T\mathbf{W}\mathbf{X},$$

where $\mathbf{W}$ is the $N \times N$ diagonal matrix with elements

$$w_{ii} = \frac{1}{\operatorname{var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

The expression on the right-hand side of (1) is the vector with elements

$$\sum_{k=1}^{p}\sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{\operatorname{var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 b_k^{(m-1)} + \sum_{i=1}^{N}\frac{(y_i-\mu_i)x_{ij}}{\operatorname{var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)$$

The right-hand side of Equation (1) can be rewritten as

$$\mathbf{X}^T\mathbf{W}\mathbf{z},$$

where $\mathbf{z}$ has elements

$$z_i = \sum_{k=1}^{p} x_{ik} b_k^{(m-1)} + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)$$

with μ_i and $\partial\eta_i/\partial\mu_i$ evaluated at $\mathbf{b}^{(m-1)}$.

Hence the iterative Equation, can be written as

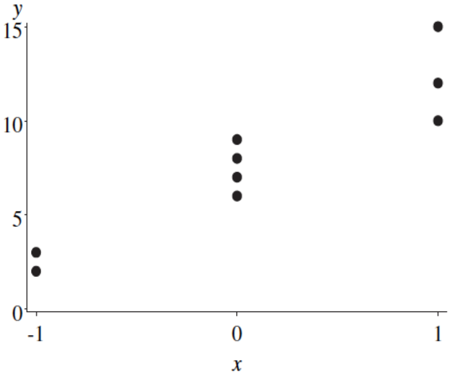
$$\mathbf{X}^T\mathbf{W}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}^T\mathbf{W}\mathbf{z}. \tag{4.25}$$

This is the same form as the normal equations for a linear model obtained by weighted least squares, except that it has to be solved iteratively because, in general, $\mathbf{z}$ and $\mathbf{W}$ depend on $\mathbf{b}$. Thus for generalized linear models, maximum likelihood estimators are obtained by an **iterative weighted least squares** procedure.

Example: Poisson regression

Data for Poisson regression example.

y_i	2	3	6	7	8	9	10	12	15
x_i	-1	-1	0	0	0	0	1	1	1



Example: Poisson regression

From the Figure the initial estimates $b_1^{(1)} = 7$ and $b_2^{(1)} = 5$ are obtained. Therefore,

$$(\mathbf{X}^T\mathbf{W}\mathbf{X})^{(1)} = \begin{bmatrix} 1.821429 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}, \qquad (\mathbf{X}^T\mathbf{W}\mathbf{z})^{(1)} = \begin{bmatrix} 9.869048 \\ 0.583333 \end{bmatrix},$$

$$\text{so} \qquad \mathbf{b}^{(2)} = \left[(\mathbf{X}^T\mathbf{W}\mathbf{X})^{(1)}\right]^{-1}(\mathbf{X}^T\mathbf{W}\mathbf{z})^{(1)}$$

$$= \begin{bmatrix} 0.729167 & 0.4375 \\ 0.4375 & 1.0625 \end{bmatrix} \begin{bmatrix} 9.869048 \\ 0.583333 \end{bmatrix}$$

$$= \begin{bmatrix} 7.4514 \\ 4.9375 \end{bmatrix}.$$

Successive approximations for regression coefficients in the Poisson re-gression example.

m	1	2	3	4
$b_1^{(m)}$	7	7.45139	7.45163	7.45163
$b_2^{(m)}$	5	4.93750	4.93531	4.93530