[Numerical Method finding MLE]

Def) $f(y;\lambda,\theta) = \frac{\lambda y^{\lambda-1}}{\theta^{\lambda}} e^{-(y/\theta)^{\lambda}}$

Weibull distribution

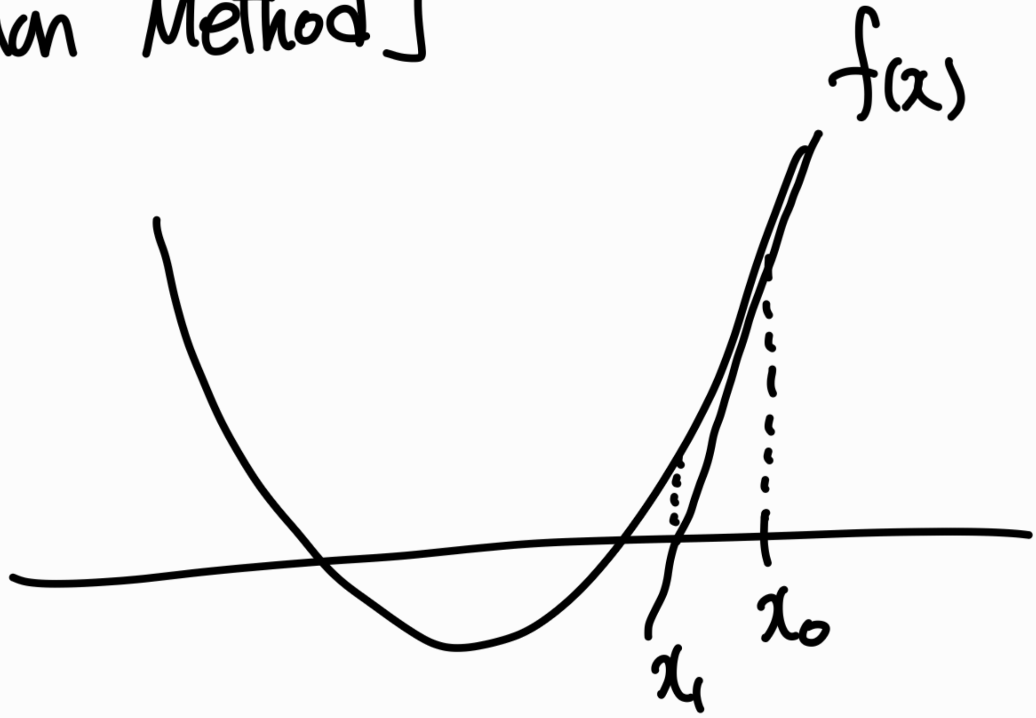
For a fixed λ,

$$l = -\lambda \log\theta - \sum \left(\frac{y_i}{\theta}\right)^{\lambda} + \text{const}$$

$$\frac{dl}{d\theta} = \frac{-\lambda}{\theta} + \sum \lambda \left(\frac{y_i}{\theta}\right)^{\lambda-1} \frac{y_i}{\theta^2}$$

$$\Rightarrow \quad \hat{\theta} = \frac{\sum y_i^{\lambda}}{N}$$

[Newton Method]



Input : x_0 (Initial point)

Algorithm : $x_m = x_{m-1} - \dfrac{f(x_{m-1})}{f'(x_{m-1})}$

Output : x_∞ is a solution to $f(x)=0$

Example : $f(\theta) = \dfrac{dl}{d\theta} = U$

$\Rightarrow \hat{\theta} = \theta_\infty = \text{Newton}(\theta_0)$

Obs) If $U' \simeq E(U')$, $E(U')$

$= -\mathrm{Var}\,U \underset{\uparrow}{=} \boxed{} \overset{\text{let}}{=} -J$

exp prop

$$\Rightarrow \quad \theta_m = \theta_{m-1} - \frac{U|_{\theta=\theta_{m-1}}}{U'|_{\theta=\theta_{m-1}}}$$

$$\Rightarrow \quad \theta_m = \theta_{m-1} + \frac{U|_{\theta=\theta_{m-1}}}{J|_{\theta=\theta_{m-1}}}$$

<Note> $\mathrm{std}(\theta_\infty) \propto \frac{1}{J}$

In fact, $= \sqrt{1/J}$

$Y_i \sim$ cano exp $\begin{cases} EY_i = \mu_i = -\dfrac{c'(\theta_i)}{b'(\theta_i)} \\ g(\mu_i) = x_i^T\beta = \eta_i \end{cases}$

$$l = \sum y_i b(\theta_i) + c(\theta_i) + d(\theta_i)$$

$$\Rightarrow \frac{dl}{d\beta_j} = \sum y_i \frac{db}{d\beta_j} + \frac{dc}{d\beta_j} + \frac{d\,d}{d\beta_j}$$

Obs: $\dfrac{db_i}{d\beta_j} = \dfrac{db_i}{d\theta_i}\dfrac{d\theta_i}{d\mu_i}\dfrac{d\mu_i}{d\eta_i}\dfrac{d\eta_i}{d\beta_j}$

$$= b'(\theta_i)\, b'(\theta_i) \operatorname{Var} y_i \, x_j \frac{d\mu_i}{d\eta_i}$$

$$d\mu_i = d\theta_i \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{(b'(\theta))^2}$$

$$d\eta_i = d\beta_j x_j$$

$$= \sum \frac{(y_i - \mu_i)}{\operatorname{Var} y_i} x_{ij} \frac{\partial \mu_i}{\partial \eta_i} \overset{\text{let}}{=} U_j$$

$\Rightarrow$ We have a random vector $U = \begin{bmatrix} U_1 \\ \vdots \\ U_p \end{bmatrix}$

$$\Rightarrow \hat{\beta} = \beta_\infty = \operatorname{Newton}(\beta_0)$$

$$\beta_m = \beta_{m-1} + \frac{U|_{\beta=\beta_{m-1}}}{J|_{\beta=\beta_{m-1}}}$$

does not make sense

$$= \beta_{m-1} + J^{-1} U$$

, provided J^{-1} exist

$$\Rightarrow J \beta_m = J \beta_{m-1} + U$$

Multiply by J $\Rightarrow$ $X^T W X \beta_m = X^T W X \beta_{m-1} + U \overset{\text{let}}{=} X^T W z$

obs) $[J]_{jk} \stackrel{\text{def}}{=} EU_jU_k - EU_jEU_k$

$$= \sum \frac{x_{ij}x_{ik}}{\operatorname{Var} Y_i} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

$\Rightarrow \quad J = X^T W X$ where $W =$

We have the normal equation to the (canonical) generalized model

$$X^T W X \beta = X^T W z$$

(special case) $W = I \Rightarrow X^T X \beta = X^T y$

It can be solved by iterating an algorithm

[Statistical Inference]

Def) $W = (\hat{\theta} - \theta)^T \operatorname{Var}\hat{\theta} \, (\hat{\theta} - \theta) \overset{d}{\sim} \chi^2(p)$

$$U(\beta) = U(\hat{\beta}) + \nabla U \, (\beta - \hat{\beta})$$

$$= -J(\beta - \hat{\beta})$$

$U(\hat{\beta}) = 0$

$\nabla U \simeq EU' = -J$

If J is const, $E - J(\beta - \hat{\beta})$

$$= -JE(\beta - \hat{\beta}) = EU = 0$$

$$\Rightarrow \quad E\hat{\beta} = \beta \quad \cdots (1)$$

$$\mathrm{Var}\,\hat{\beta} = E(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T$$

$$= E\, J^{-1} U\, U^T J^{-1}$$

$$= J^{-1} E\, U U^T J = J^{-1} J J^{-1} \underset{(2)}{=} J^{-1}$$

From (1) and (2), the wald statistic

$$(\hat{\beta}-\beta)^T J (\hat{\beta}-\beta) \sim \chi^2(p)$$

[Checking Model Adequacy]

Saturated Model	H_1	M_1	β_{max}	$\hat{\beta}_{max}$
Proposed Model	H_0	M_0	β	$\hat{\beta}$

Def) $D = 2\ell(\hat{\beta}_{max}) - 2\ell(\hat{\beta})$ (Deviance)

obs) $$D = \underbrace{2\ell(\hat{\beta}_{max}) - 2\ell(\beta_{max})}_{\chi^2(N)}$$

$$+ \underbrace{2\ell(\beta_{max}) - 2\ell(\beta)}_{\text{Const}}$$

$$+ \underbrace{2\ell(\beta) - 2\ell(\hat{\beta})}_{\chi^2(p)}$$

Recall: $\ell(\beta) - \ell(\hat{\beta}) = -\frac{1}{2}(\hat{\beta}-\beta)^T J(\hat{\beta})(\hat{\beta}-\beta)$

Fact: $D \overset{d}{\sim} \chi^2(N-p, \upsilon)$

Example 1: $Y_1 \sim \text{Binomial}(n_1, p_1)$

$Y_2 \sim \text{Binomial}(n_2, p_2)$

proposed model: $p_1 = p_2$

$$\Rightarrow \quad \hat{p} = \frac{y_1 + y_2}{n_1 + n_2}$$

$$l = \log \binom{n_1}{y_1} + y_1 \log p_1 + (n_1 - y_1) \log p_1$$

$$+ \log \binom{n_2}{y_2} + y_2 \log p_2 + (n_2 - y_2) \log p_2$$

$$l_{prop} = l \Big|_{p_1 = p_2 = \hat{p}}$$

$$\Rightarrow D = 2 l_{full} - 2 l_{prop} \overset{d}{\sim} \chi^2(1, \upsilon)$$

$$l_{full} = l \Big|_{p_1 = \frac{y_1}{n_1} \; p_2 = \frac{y_2}{n_2}}$$

Example 2 : $Y_i \sim N(\mu_i, \sigma^2)$

$$l(\mu) = -\frac{1}{2\sigma^2}(y-\mu_i)^2 + \text{const}$$

Full Model : $l = \text{const}$ $(\because \hat{\mu}_i = y_i)$

Proposed Model = Linear Model

(i.e., $\hat{\mu}_i = x_i^T\beta$)

$$l_{prop} = -\frac{1}{2\sigma^2}\Sigma(y_i - x_i^T\hat{\beta})^2 + \text{const}$$

where $\hat{\beta} = (X^TX)^{-1}X^Ty$

$$D = 2l_{full} - 2l_{prop} = \frac{1}{\sigma^2}\Sigma(y_i - x_i^T\hat{\beta})^2$$

$$\overset{d}{\sim} \chi^2(N-p, v)$$

Example 3:

Proposed 1: $\beta \in \mathbb{R}^q$

Proposed 2: $\beta \in \mathbb{R}^p$ $\quad (q < p)$

$$\Delta D \overset{\text{let}}{=} D_0 - D_1$$

$$= 2\ell(\hat{\beta}_{max}) - 2\ell(\beta_1) - 2\ell(\hat{\beta}_{max}) + 2\ell(\beta_2)$$

$$= 2\ell(\beta_2) - 2\ell(\beta_1)$$

Assume Proposed 2 $\approx$ Full Model,
(i.e., $D_1 \sim \chi^2(N-p)$)

So, under H_0, $\Delta D \sim \chi^2(p-q)$

In this case, we don't know σ^2

and so we introduce an alternative

in order to avoid calculating unknown σ^2

$$F = \frac{\frac{\Delta D}{p-q}}{\frac{D_1}{N-q}} \sim F(p-q, N-q, \nu)$$

Under H_0, $F \sim F(p-q, N-q, 0)$