Delta Method

Suppose $x_1, \dots, x_n \overset{iid}{\sim} F(\underset{\sim}{\theta})$, $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_d)$

Thm: The MLE $\hat{\theta}$ of θ is

$$\hat{\underset{\sim}{\theta}} \dot{\sim} MVN_d(\underset{\sim}{\theta}, I_E^{-1}(\underset{\sim}{\theta}))$$

where

$$I_E(\underset{\sim}{\theta}) = \begin{bmatrix} e_{11}(\underset{\sim}{\theta}) & \cdots & e_{1d}(\underset{\sim}{\theta}) \\ \vdots & & \\ e_{d1}(\underset{\sim}{\theta}) & \cdots & e_{dd}(\underset{\sim}{\theta}) \end{bmatrix}$$

with

$$e_{ij}(\underset{\sim}{\theta}) = -E\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ell(\underset{\sim}{\theta})\right].$$

Note: $I_E(\underset{\sim}{\theta})$ is referred to as the expected information matrix.

95% CI for θ_i $\qquad \hat{\theta}_i \dot{\sim} N(\theta_i, \psi_{ii})$

ψ_{ii} is the ii term in $I_E^{-1}(\underset{\sim}{\theta})$.

if ψ_{ii} were known then

$$\hat{\theta}_i \pm z_{\alpha/2} \sqrt{\psi_{ii}}$$

Since θ is unknown, it is usual to approximate $I_E(\underset{\sim}{\theta})$ using the observed information matrix

$$I_0(\underset{\sim}{\theta}) = \begin{bmatrix} -\frac{\partial^2}{\partial\theta_1^2} l(\underset{\sim}{\theta}) & \cdots & -\frac{\partial^2}{\partial\theta_1 \partial\theta_d} l(\underset{\sim}{\theta}) \\ \vdots & & \\ -\frac{\partial^2}{\partial\theta_d \partial\theta_1} l(\underset{\sim}{\theta}) & \cdots & -\frac{\partial^2}{\partial\theta_d^2} l(\underset{\sim}{\theta}) \end{bmatrix}$$

and evaluated at $\hat{\underset{\sim}{\theta}}$.

<u>approximate 95% CI for θ_i</u>

$\tilde{\psi}_{ij}$ is the ij term of $I_0^{-1}(\underset{\sim}{\theta})\Big|_{\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}}$

$$\hat{\theta}_i \pm z_{\alpha/2} \sqrt{\tilde{\psi}_{ii}}$$

We may be interested in $\phi = g(\underset{\sim}{\theta})$.
Recall the invariance property of MLEs. $\hat{\phi} = g(\hat{\underset{\sim}{\theta}})$

<u>Thm</u> (delta method)

Let $\hat{\underset{\sim}{\theta}}$ be the MLE of $\underset{\sim}{\theta}$, with variance-covariance matrix V_θ. Then if $\phi = g(\theta)$ is a scalar function, the MLE of ϕ satisfies

$$\hat{\phi} \,\dot{\sim}\, N(\phi, V_\phi)$$

where

$$V_\phi = (\nabla\phi)' \, V_\theta \, (\nabla \phi)$$

with

$$\nabla \phi = \left[\frac{\partial \phi}{\partial \theta_1}, \ldots, \frac{\partial \phi}{\partial \theta_d}\right]'$$

evaluated at $\hat{\underset{\sim}{\theta}}$.

<u>Note</u>: In the same way that the approximate normality of $\hat{\theta}$ can be used to obtain confidence intervals for components of $\underset{\sim}{\theta}$, the delta method enables the approximate normality of $\hat{\phi}$, CI's for ϕ.

Example:

Simulate time to failure of a ~~___~~ sample of $n = 32$ engine components with different levels of corrosion.

The goal of the analysis is to ascertain how the failure time is affected by the corrosion level.

data $\{ (w_1, t_1), \ldots, (w_n, t_n) \}$.

t_i = i^{th} failure time
w_i = i^{th} corrosion level for engine i, fixed

The t_i's are assumed to ~~random~~ be realizations of random variables dependent on w_i.

$T \sim \exp(\lambda)$ $\quad f_T(t) = \lambda e^{-\lambda t} \quad t > 0$

$E[T] = \frac{1}{\lambda}$

Suppose $\lambda = a w^b$, $E[T] = a^{-1} w^{-b}$.

Determine the MLE of $\underset{\sim}{\theta} = (a, b)$

Likelihood

$$L(a,b) = f(t_1, \dots, t_n / \lambda)$$
$$= \prod_{i=1}^{n} f(t_i / \lambda)$$
$$= \prod_{i=1}^{n} \left[\lambda e^{-\lambda t_i} \right]$$
$$= \prod_{i=1}^{n} \left[a w_i^b \cdot e^{-a w_i^b t_i} \right]$$

log-Likelihood

$$\ell(a,b) = \log L(a,b)$$
$$= \log \left[\prod_{i=1}^{n} a w_i^b \, e^{-a w_i^b t_i} \right].$$
$$= \sum_{i=1}^{n} \left[\log \left(a w_i^b \, e^{-a w_i^b t_i} \right) \right]$$
$$= \sum_{i=1}^{n} \left[\log(a) + b \log(w_i) - a w_i^b t_i \right]$$

$$= n \log(a) + b \sum_{i=1}^{n} \log(w_i) - a \sum_{i=1}^{n} w_i^{b} t_i \quad *$$

$$\frac{\partial}{\partial a} l(a,b) = \frac{n}{a} - \sum_{i=1}^{n} w_i^{b} t_i = 0$$

$$\frac{\partial}{\partial b} l(a,b) = \sum_{i=1}^{n} \log(w_i) - a \sum_{i=1}^{n} w_i^{b} t_i \log(w_i) = 0$$

Recall $\frac{d}{da} x^a = \frac{d}{da} e^{a \log x} = e^{a \log x} \cdot \log x$

$$= x^a \log x.$$

Since the equations have no analytical solution, numerical techniques are required to maximize $*$.

The mean lifetime curve

$$E[T] = \hat{a} \, w^{\hat{b}}$$

The observed information matrix in this example

$$I_0(a,b) = \begin{bmatrix} -\frac{\partial^2}{\partial a^2} \ell(a,b) & -\frac{\partial}{\partial a \partial b} \ell(a,b) \\ -\frac{\partial^2}{\partial a \partial b} \ell(a,b) & -\frac{\partial}{\partial b^2} \ell(a,b) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{n}{a^2} & \sum w_i^b t_i \log(w_i) \\ \sum w_i^b t_i \log(w_i) & a \sum w_i^b t_i (\log(w_i))^2 \end{bmatrix}$$

It is convenient and often more accurate to work with the observed information matrix.

Use the delta method to develop a CI for $\phi = E[T] = a^{-1} w_0^{-b}$, w_0 fixed

By the invariance property of MLEs

$$\hat{\phi} = \hat{a} \, w_0^{\hat{b}}$$

by the delta method

$$\text{Var}(\hat{\phi}) \simeq (\nabla \phi)' V \nabla \phi$$

where
$$(\nabla \phi)' = \left[-a^{-2} w_0^{-b} , -a^{-1} w_0^{-b} \log(w_0) \right].$$

evaluated at $(\hat{a}, \hat{b})$. and

$$V = I_0(a,b) \Big|_{(a,b)=(\hat{a},\hat{b})}.$$

Test if $H_0: b=0$, $H_1: b \neq 0$

Fit the restricted model

$$T \sim Exp(a)$$

$$\ell(a) = n \log(a) - a \sum t_i$$

$$\hat{a} = \frac{n}{\sum t_i}$$

Deviance : $D = -2 \log(\Lambda) = -2 \log \left(\frac{f(\underset{\sim}{t}|a)}{f(\underset{\sim}{t}|a,b)} \right)$

$$= 2 \{ \ell(a,b) - \ell(a) \}.$$

compare with $\chi_1^2(.95)$