응용통계연구 제 6권 2호
1993년 9월 pp.319-328.

Convergence Rate of Newton-Raphson Method

Kwan Jeh Lee[1]

Abstract

The actual convergence rate of Newton-Raphson iteration method at each step is studied under the regularity conditons for the limiting distribution: The convergence rate of it is accelerated with good starting values. Hence we can decide a number of iterations according to our purposes.

Introduction

The setting we start from is one in which it is too difficult to solve the likelihood equations. That is, it is too difficult to obtain the maximum likelihood estimators(MLEs). Therefore we calculate approximate roots of the equation. These approximate roots are asymptotically equivalent to MLEs in the sense that we can use them as MLEs are used in forming the likelihood ratio test statisitics.

In numerical analysis there are several methods by which we can find approximate values of the desired root. In the present study we will study the Newton-Raphson method. The closeness between an approximate value and the MLE can be measured by the probability of the distance between them, which goes to zero at a certain rate. In 1958 Stuart compared this method and the Method of Scoring and provided the numerical examples. In 1966 Barnett also discussed the performance of the Newton-Raphson Method. However they did not find the actual convergence rate for which the probability of the distance goes to zero. We study their actual convergence rates and find that the convergence rate of the Newton-Raphson Method is accelerated with some starting values.

1. Newton-Raphson Method

Under the regularity conditions for the limiting distribution of the MLE, let $\underline{T}=(T_1,\cdots,T_s)$ to be an $n^{1/\alpha}$-consistent estimator of a true parameter $\underline{\theta}=(\theta_1,\cdots,\theta_s)$if

$$\underline{T}=\underline{\theta}+O_p(n^{-\alpha}).$$

Notice from now on that the regularity conditions in this study means the regularity conditions for the limiting distribution of the MLE, and that for simplicity we will use $O_p(\cdot)$ instead of $\underline{O}_p(\cdot)$or $O_p(\cdot)\cdot\underline{1}$. The Newton-Raphson approximation for multiparameters is calculated as follow.

$$\underline{\delta_1}=\underline{T}-[l_{..}(\underline{T})]^{-1}\cdot l_{.}(\underline{T}),$$

where $l(\theta)$ is a log-likelihood function, and $l_{.}(T)$and $l_{..}(T)$ are the first and the second

1) Dept. of Statistics, Dongguk University, 26 3-Ga, Pil-dong, Chung-gu, Seoul, 100-715

derivatives of the log-likelihood function evaluated at $\underline{T}$, that is,

$$l_{.}(\underline{T})=\begin{pmatrix} \frac{\partial}{\partial\theta_1} l(\underline{\theta}) \\ \frac{\partial}{\partial\theta_2} l(\underline{\theta}) \\ \vdots \\ \frac{\partial}{\partial\theta_s} l(\underline{\theta}) \end{pmatrix}_{\underline{\theta}=\underline{T}}$$

and

$$l_{..}(\underline{T})=\begin{pmatrix} \frac{\partial^2}{\partial\theta_1\partial\theta_1} l(\underline{\theta}) & \frac{\partial^2}{\partial\theta_1\partial\theta_2} l(\underline{\theta}) & \cdots & \frac{\partial^2}{\partial\theta_1\partial\theta_s} l(\underline{\theta}) \\ \frac{\partial^2}{\partial\theta_2\partial\theta_1} l(\underline{\theta}) & \frac{\partial^2}{\partial\theta_2\partial\theta_2} l(\underline{\theta}) & \cdots & \frac{\partial^2}{\partial\theta_2\partial\theta_s} l(\underline{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial\theta_s\partial\theta_1} l(\underline{\theta}) & \frac{\partial^2}{\partial\theta_s\partial\theta_2} l(\underline{\theta}) & \cdots & \frac{\partial^2}{\partial\theta_s\partial\theta_s} l(\underline{\theta}) \end{pmatrix}_{\underline{\theta}=\underline{T}}.$$

Let $l_{ij}(\cdot)$ be the (i,j) element of $l_{..}(\cdot)$ and

$$l_{ijk}=\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} l(\underline{\theta}).$$

In order to expand $l_{.}(\underline{T})$ in a Taylor series in $\underline{\hat{\theta}}=(\hat{\theta}_1,\cdots,\hat{\theta}_s)$, we expand its $\underline{i\,th}$ component about $\underline{\hat{\theta}}$, and find

$$\begin{aligned} l_{i}(\underline{T}) &= l_{i}(\underline{\hat{\theta}})+\sum_{j=1}^{s}(T_j-\hat{\theta}_j)l_{ij}(\underline{\hat{\theta}}) \\ &\quad+\frac{1}{2!}\sum_{j=1}^{s}\sum_{k=1}^{s}(T_j-\hat{\theta}_j)(T_k-\hat{\theta}_k)l_{ijk}(\underline{\theta_i^*}) \\ &= \sum_{j=1}^{s}(T_j-\hat{\theta}_j)l_{ij}(\underline{\hat{\theta}}) \\ &\quad+\frac{1}{2!}\sum_{j=1}^{s}\sum_{k=1}^{s}(T_j-\hat{\theta}_j)(T_k-\hat{\theta}_k)l_{ijk}(\underline{\theta_i^*}), \end{aligned}$$

where $\underline{\theta_i^*}$ lies between $\underline{T}$ and $\underline{\hat{\theta}}$, and $l_{.}(\underline{\hat{\theta}})=0$.

By the regularity conditions

$$\begin{aligned} \frac{1}{n}|l'''(\underline{\hat{\theta}_i^*})| &= \left|\frac{1}{n}\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} l(\underline{\hat{\theta}_i^*})\right| \\ &= \left|\frac{1}{n}\sum_{m=1}^{n}\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\log f(x_m|\underline{\hat{\theta}_i^*})\right| \\ &\leq \frac{1}{n}|M_{ijk}(x_1)+\cdots+M_{ijk}(x_n)|, \end{aligned}$$

with probability tending to 1. So we can rewrite

$$l_{.i}(\underline{T}) = \sum_{j=1}^{s}(T_j - \hat{\theta}_j) l_{ij}(\hat{\underline{\theta}}) + \frac{1}{2!}\sum_{j=1}^{2}\sum_{k=1}^{s}(T_j - \hat{\theta}_j)(T_k - \hat{\theta}_k)\cdot O_p(n) .$$

Let

$$\underline{T} - \hat{\underline{\theta}} = O_p(n^{-\beta}).$$

Then

$$\begin{aligned} l_{.i}(\underline{T}) &= \sum_{j=1}^{s}(T_j - \hat{\theta}_j) l_{ij}(\hat{\underline{\theta}}) + \frac{1}{2!}\sum_{j=1}^{s}\sum_{k=1}^{s}\cdot O_p(n^{-\beta})\cdot O_p(n^{-\beta})\cdot O_p(n) \\ &= \sum_{j=1}^{s}(T_j - \hat{\theta}_j) l_{ij}(\hat{\underline{\theta}}) + O_p(n^{-2\beta+1}) . \end{aligned}$$

Expand $l_{ij}(\hat{\underline{\theta}})$ in a Taylor series in $\underline{\theta}$.

$$\begin{aligned} l_{ij}(\hat{\underline{\theta}}) &= l_{ij}(\underline{\theta}) + \sum_{k=1}^{s}(\hat{\theta}_k - \theta_k) l_{ijk}(\underline{\theta}_2^*) \\ &= l_{ij}(\underline{\theta}) + O_p(n^{\frac{1}{2}}), \end{aligned}$$

where $\underline{\theta}_2^*$ lies between $\hat{\underline{\theta}}$ and $\underline{\theta}$ and $\sqrt{n}(\hat{\theta}_k - \theta_k)$ converges in law to a normal distribution. This implies that

$$\hat{\theta}_k - \theta_k = O_p(n^{-\frac{1}{2}}).$$

Thus we can write

$$\begin{aligned} l_{.i}(T) &= \sum_{j=1}^{s}(T_j - \hat{\theta}_j)\{l_{ij}(\underline{\theta}) + O_p(n^{\frac{1}{2}})\} + O_p(n^{-2\beta+1}) \\ &= \sum_{j=1}^{s}(T_j - \hat{\theta}_j) l_{ij}(\underline{\theta}) + O_p(n^{-\min(\beta-\frac{1}{2}, 2\beta-1)}). \end{aligned} \tag{1}$$

Let

$$[l_{..}(\underline{\theta})]^{-1} \equiv \begin{pmatrix} \upsilon_{11}(\underline{\theta}) & \upsilon_{12}(\underline{\theta}) & \cdots & \upsilon_{1s}(\underline{\theta}) \\ \upsilon_{21}(\underline{\theta}) & \upsilon_{22}(\underline{\theta}) & \cdots & \upsilon_{2s}(\underline{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \upsilon_{s1}(\underline{\theta}) & \upsilon_{s2}(\underline{\theta}) & \cdots & \upsilon_{ss}(\underline{\theta}) \end{pmatrix} .$$

Then

$$\sum_{j=1}^{s}\upsilon_{ij}(\underline{\theta}) l_{ji}(\underline{\theta}) = 1, \quad i = 1, 2, \cdots, s,$$

and

$$\sum_{j=1}^{s}\upsilon_{ij}(\underline{\theta}) l_{jk}(\underline{\theta}) = 0, \quad j \neq k.$$

Now let $\phi_i(\underline{\theta})$ be $\underline{i\,th}$ component of $[l_{..}(\underline{\theta})]^{-1} l_{.}(\underline{\theta})$, then

$$\phi_i(\underline{T}) = \sum_{j=1}^{s}\upsilon_{ij}(\underline{T}) l_j(\underline{T}).$$

From the above expansion (1) of $l_{.i}(\underline{T})$ we can rewrite $\phi_i(\underline{T})$ as

$$\begin{aligned}\phi_i(\underline{T}) &= \sum_{j=1}^{s} \upsilon_{ij}(\underline{T}) l_{.j}(\underline{T}) \\ &= \sum_{j=1}^{s} \upsilon_{ij}(\underline{T})\{\sum_{k=1}^{s}(T_k - \hat{\theta}_k) l_{jk}(\underline{\theta}) + O_p(n^{-\min\{\beta-\frac{1}{2},2\beta-1\}})\} \\ &= (T_i - \hat{\theta}_i)\{\sum_{j=1}^{s}\upsilon_{ij}(\underline{T}) l_{ji}(\underline{\theta})\} \\ &\quad + \sum_{k=1,k\neq i}^{s}(T_k - \hat{\theta}_k)\{\sum_{j=1}^{s}\upsilon_{ij}(\underline{T}) l_{jk}(\underline{\theta})\} \\ &\quad + O_p(n^{-\min\{\beta-\frac{1}{2},2\beta-1\}-1}) .\end{aligned}$$

Now expand $\upsilon_{ij}(\underline{T})$ in a Taylor series in $\underline{\theta}$

$$\upsilon_{ij}(\underline{T}) = \upsilon_{ij}(\underline{\theta}) + \sum_{k=1}^{s}(T_k - \theta_k)(\frac{\partial}{\partial\theta_k}\upsilon_{ij}(\underline{\theta_4^*})),$$

where $\underline{\theta_4^*}$ lies between $\underline{T}$ and $\underline{\theta}$. Thus

$$\begin{aligned}\upsilon_{ij}(\underline{T}) l_{ji}(\underline{\theta}) &= \{\upsilon_{ij}(\underline{\theta}) + \sum_{k=1}^{s}(T_k - \theta_k)(\frac{\partial}{\partial\theta_k}\upsilon_{ij}(\underline{\theta_4^*}))\} l_{ji}(\underline{\theta}) \\ &= \upsilon_{ij}(\underline{\theta}) l_{ji}(\underline{\theta}) + \sum_{k=1}^{s}(T_k - \theta_k)(\frac{\partial}{\partial\theta_k}\upsilon_{ij}(\underline{\theta_4^*})) l_{ji}(\underline{\theta}) .\end{aligned}$$

Therefore

$$\sum_{j=1}^{s}\upsilon_{ij}(\underline{T}) l_{ji}(\underline{\theta}) = 1 + \sum_{j=1}^{s}\sum_{k=1}^{s}(T_k - \theta_k)(\frac{\partial}{\partial\theta_k}\upsilon_{ij}(\underline{\theta_4^*})) l_{ji}(\underline{\theta}).$$

Similarly, for $i \neq k$

$$\sum_{j=1}^{s}\upsilon_{ij}(\underline{T}) l_{jk}(\underline{\theta}) = \sum_{j=1}^{s}\sum_{l=1}^{s}(T_l - \theta_l)(\frac{\partial}{\partial\theta_l}\upsilon_{ij}(\underline{\theta_4^*})) l_{ji}(\underline{\theta}).$$

We can rewrite $\phi_i(\underline{T})$ as follows.

$$\begin{aligned}\phi_i(\underline{T}) &= (T_i - \hat{\theta}_i)\{1 + \sum_{j=1}^{s}\sum_{k=1}^{s}(T_k - \theta_k)(\frac{\partial}{\partial\theta_k}\upsilon_{ij}(\underline{\theta_4^*})) l_{ji}(\underline{\theta})\} \\ &\quad + \sum_{k=1,k\neq i}^{s}(T_k - \hat{\theta}_k)\{\sum_{j=1}^{k}\sum_{l=1}^{s}(T_l - \theta_l)(\frac{\partial}{\partial\theta_l}\upsilon_{ij}(\underline{\theta_4^*})) l_{ji}(\underline{\theta})\} \\ &\quad + O_p(n^{-\min\{\beta-\frac{1}{2},2\beta-1\}-1}) \\ &= (T_i - \hat{\theta}_i) + O_p(n^{-\alpha-\beta}) + O_p(n^{-\min\{\beta-\frac{1}{2},2\beta-1\}-1}) .\end{aligned}$$

Therefore

$$\begin{aligned}\delta_i &= T_i - \{(T_i - \hat{\theta}_i) + O_p(n^{-\alpha-\beta}) + O_p(n^{-\min\{\beta-\frac{1}{2},2\beta-1\}-1})\} \\ &= \hat{\theta}_i + O_p(n^{-\alpha-\beta}) + O_p(n^{-\min\{\beta-\frac{1}{2},2\beta-1\}-1}) .\end{aligned}$$

Note the following special cases.

Case 1. When $\alpha = 1/2$, that is, $\underline{T} = \underline{\theta} + O_p(n^{-\frac{1}{2}})$, $\beta = 1/2$. Thus we have

$$\underline{\delta}_1 = \hat{\underline{\theta}} + O_p(n^{-1}).$$

Case 2. When $\alpha = 1$, that is, $\underline{T} = \underline{\theta} + O_p(n^{-1})$, again $\beta = 1/2$ and we have

$$\underline{\delta}_1 = \hat{\underline{\theta}} + O_p(n^{-1}).$$

The convergece rate is not improved by *n-consistent* estimator for multiparameter case. In order words, the approximation with $\sqrt{n}$*-consistent* estimator as a starting value has the same convergence rate as that with *n-consistent* estimator, to the MLE.

Case 3. When $\underline{T} = \underline{\theta} + O_p(n^{-\frac{1}{2}})$, then by Case 1,

$$\underline{\delta}_1 = \hat{\underline{\theta}} + O_p(n^{-1}).$$

Now using $\underline{\delta}_1$ as a starting value, that is, we have the second step approximation as follows.

$$\underline{\delta}_2 = \underline{\delta}_1 - [l_{..}(\underline{\delta}_1)]^{-1} l_{.}(\underline{\delta}_1) \ .$$

In order to calculate it, consider

$$\begin{aligned} \underline{\delta}_1 - \underline{\theta} &= (\underline{\delta}_1 - \hat{\underline{\theta}}) + (\hat{\underline{\theta}} - \underline{\theta}) \\ &= O_p(n^{-\frac{1}{2}}) \ . \end{aligned}$$

Thus $\alpha = 1/2$ and by Case 1 $\beta = 1$. Now we can calculate $\underline{i\ th}$ elements of $\underline{\delta}_2$. That is

$$\begin{aligned} \delta_{2i} &= \delta_{1i} - \{(\delta_{1i} - \hat{\theta}_i) + O_p(n^{-\alpha-\beta}) + O_p(n^{-\min\{\beta-\frac{1}{2}, 2\beta-1\}-1})\} \\ &= \hat{\theta}_i + O_p(n^{-\frac{1}{2}-1}) + O_p(n^{-\min\{1-\frac{1}{2}, 2\cdot 1-1\}-1}) \\ &= \hat{\theta}_i + O_p(n^{-\frac{3}{2}}) + O_p(n^{-\frac{3}{2}}) \ . \end{aligned}$$

Hence

$$\underline{\delta}_2 = \hat{\underline{\theta}} + O_p(n^{-\frac{3}{2}}).$$

Let β be the rate of *k-step* N-R approximation to the MLE $\hat{\underline{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_s)$ if

$$\underline{\delta}_k = \hat{\underline{\theta}} + O_p(n^{-\beta}).$$

Therefore we have the following theorem.

Theorem 1 If we iterate the Newton-Raphson approximation k times with a $n^{1/\alpha}$-consistent estimator of a true parameter $\underline{\theta} = (\theta_1, \cdots, \theta_s)$, the rate of the approximation to the MLE is

$$\beta = \frac{1}{2}(k+1),$$

that is,

$$\underline{\delta}_k = \hat{\underline{\theta}} + O_p(n^{-\frac{1}{2}(k+1)}),$$

where α is greater than, or equal to $1/2$, and $k = 1, 2, 3, \cdots$.

Proof. We already considered the cases where $k=1,2$ and $\alpha=1/2$. In each iteration β increases by $1/2$. This completes the proof. □

When we start the method with a consistent estimator $\underline{T}$ at a slower rate such as $\underline{T}=\underline{\theta}+o_p(n^{-\frac{1}{4}})$, we have 1/2-equivalent estimators to the MLE at the first iteration of the Newton-Raphson method and can apply Theorem 1 from the second iteration of the method.

Corollary 1 If we iterate the Newton-Raphson approximation k times with a consistent estimator $\underline{T}$ such as $\underline{T}=\underline{\theta}+o_p(n^{-\tau})$, $1/4\le\tau$, of a true parameter $\underline{\theta}=(\theta_1,\cdots,\theta_s)$, the rate of the approximation to the MLE is

$$\beta=1/2(k+1),$$

that is,

$$\underline{\delta_k}=\underline{\hat{\theta}}+O_p(n^{-\frac{1}{2}(k+1)}),$$

where $k=2, 3,\cdots$. And for k = 1, the one-step approximation of the Newton-Rahpson method is

$$\underline{\delta_1}=\underline{\hat{\theta}}+o_p(n^{-\frac{1}{2}}).$$

3. Concluding Remarks

One of the competing iteration methods for the N-R method is the Method of Scoring: Suppose that the conditions and the definitions are hold. A multivariate version of one-step approximation of the Method of Scoring with $\underline{T_0}=(T_{01},T_{02},\cdots,T_{0k})^T$, a $\sqrt{n}$-consistent estimator, as a starting value is given by

$$\underline{\eta_2}=\underline{T_0}+\frac{1}{n}[I(\underline{T_0})]^{-1}l'(T_0),$$

where

$$I(\underline{T_0})=\begin{pmatrix} I_{11}(\underline{\theta}) & I_{12}(\underline{\theta}) & \cdots I_{1k}(\underline{\theta}) \\ I_{21}(\underline{\theta}) & I_{22}(\underline{\theta}) & \cdots I_{2k}(\underline{\theta}) \\ \vdots & \vdots & \vdots \\ I_{k1}(\underline{\theta}) & I_{k2}(\underline{\theta}) & \cdots I_{kk}(\underline{\theta}) \end{pmatrix}_{\underline{\theta}=\underline{T_0}}$$

and

$$\begin{aligned} I_{ij}(\underline{T_0}) &= E_{\underline{\theta}}[\frac{\partial}{\partial\theta_i}\log f(\underline{x}|\underline{\theta})\frac{\partial}{\partial\theta_j}\log f(\underline{x}|\underline{\theta})]_{\underline{\theta}=\underline{T_0}} \\ &= -E_{\underline{\theta}}[\frac{\partial}{\partial\theta_i\partial\theta_j}\log f(\underline{x}|\underline{\theta})]_{\underline{\theta}=\underline{T_0}} \end{aligned}$$

Then we can easily have

$$\underline{\eta_2}=\underline{\hat{\theta}}+o_p(n^{-1/2}).$$

Hence, the convergence rate of the Method of Scoring is slower than that of the N-R method under the regularity conditions. A more study on some additional conditions to the regularity conditions applied in this study is felt desiralbe, since in Stuart(1958) two methods were compared numerically and concluded that the Method of Scoring converges faster than the N-R method.

4 Examples

Now we consider two statistical models, whose solutions of the likelihood equations are not feasible. Let δ_1 be one-step iteration approximation of the Newton-Raphson method with $\sqrt{n}$-consistent estimator as a starting variable. Then

$$\delta_1 = \hat{\theta} + O_p\left(\frac{1}{n}\right)$$

from the consequence of this study.

Example 1. (Logistic Model - Location parameter) Suppose $x_1, \cdots, x_n$ are *iid* according to the logistic density

$$f(x|\theta) = \frac{e^{-(x-\theta)}}{\{1+e^{-(x-\theta)}\}^2}.$$

The likelihood function is

$$L(\theta|\underline{x}) = \prod_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}^2}.$$

The log-likelihood function is given by

$$\begin{aligned} l(\theta) &\equiv \log L(\theta|\underline{x}) \\ &= \prod_{i=1}^{n} \log \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}^2} \\ &= \sum_{i=1}^{n} \log \left\{ \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}^2} \right\} \\ &= \sum_{i=1}^{n} \{\log e^{-(x_i-\theta)}\} - \log \{1+e^{-(x_i-\theta)}\}^2 \\ &= -\sum_{i=1}^{n}(x_i-\theta) - 2\sum_{i=1}^{n} \log \{1+e^{-(x_i-\theta)}\}. \end{aligned}$$

Thus the likelihood equation becomes

$$\begin{aligned} 0 &= l'(\theta) \\ &= \frac{\partial}{\partial\theta}\left\{-\sum_{i=1}^{n}(x_i-\theta) - 2\sum_{i=1}^{n}\log\{1+e^{-(x_i-\theta)}\}\right\} \\ &= -\sum_{i=1}^{n}(-1) - 2\sum_{i=1}^{n}\frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}} \\ &= n - 2\sum_{i=1}^{n}\frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}}. \end{aligned}$$

After some simplication it becomes

$$\sum_{i=1}^{n} \frac{1}{1+e^{(x_i-\theta)}} = \frac{n}{2}.$$

The left side is an increasing function of θ which is zero at $\theta=-\infty$ and n at $\theta=+\infty$. Therefore the likelihood equation has a unique root $\hat{\theta}$,which is the MLE since $l'(\theta)>0$ as $\theta<\hat{\theta}$ and $l'(\theta)<\theta$ as $\theta>\hat{\theta}$. It is not easy to solve this equation. For later use calculate the 2nd derivative of $l(\theta)$

$$\begin{aligned} l''(\theta) &= \frac{\partial^2}{\partial\theta^2} l(\theta) \\ &= -2\sum_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}^2}. \end{aligned}$$

Notice the following theorem.

Theorem 2 Let $x_1,\cdots,x_n$ be iid with distribution $F(x-\theta)$. Suppose that $F(0)=1/2$ and that at zero F has a density $f(0)>0$. Then

$$\sqrt{n}(\tilde{x}_n-\theta) \overset{\mathcal{L}}{\rightarrow} N(0,\frac{1}{4f^2(0)}),$$

where $\tilde{x}_n$ is a sample median.

See, for example, Lehmann's TPE p.353. Since the location logistic model satisfies the conditions of the theorem above,

$$\sqrt{n}(\tilde{x}_n-\theta) \overset{\mathcal{L}}{\rightarrow} N(0,\frac{1}{4f^2(0)}),$$

so that $\tilde{x}_n$ is $\sqrt{n}$-consistent. Therefore one-step Newton-Raphson approximation is

$$\begin{aligned} \delta_1 &= \tilde{x}_n - \frac{l'(\tilde{x}_n)}{l''(\tilde{x}_n)} \\ &= \tilde{x}_n - \{n-2\sum_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{1+e^{-(x_i-\theta)}}\} \\ &\quad \cdot \{-2\sum_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}^2}\} \\ &= \tilde{x}_n + \frac{1}{2}\{n-2\sum_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}}\} \cdot \{\sum_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}^2}\}. \end{aligned}$$

Example 2. (Weibull Model - one parameter) Suppose $x_1,\cdots,x_n$ are *iid* with a Weibull density

$$f(x|\theta) = \theta x^{\theta-1} e^{-x^\theta}$$

for $x>0$ and $\theta>0$, then the log-likelihood function of $x_1,\cdots,x_n$ is

$$\begin{aligned} l(\theta) &= n\log\theta + (\theta-1)\sum_{i=1}^{n}\log x_i \\ &\quad - \sum_{i=1}^{n}\exp(\theta\log x_i). \end{aligned}$$

Therefore

$$l'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \exp(\theta \log x_i),$$

and

$$l''(\theta) = -\frac{n}{\theta^2} - \sum_{i=1}^{n} (\log x_i)^2 \exp(\theta \log x_i),$$

which is negative for all $\theta > 0$. As $\theta \to 0$, $l'(\theta) \to \infty$ while for θ sufficiently large, $l'(\theta)$ is negative. Thus the likelihood equation has a unique root and equals the MLE. The one-step Newton-Raphson approximation is

$$\delta_1 = d_o - \frac{l'(d_o)}{l''(d_o)},$$

where d_o is $\sqrt{n}$-consistent estimator and is

$$d_o = \{(\frac{6}{\pi^2})[\sum_{i=1}^{n} (\log x_i)^2 - (\sum_{i=1}^{n} \log x_i)^2/n]/(n-1)\}^{-1/2},$$

which is given by Menon(1963).

References

[1] Barnett, V. D. (1966), "Evaluation of the maximum likelihood estimators where the likelihood equation has multiple roots," *Biometrika*, **53**, pp.151-166.

[2] Kendall, M. G., and Stuart, A. (1979), *The advanced theory of statistics*, Vol. 2. 4th ed., MacMillan, New York.

[3] Lehmann, E. L. (1983), *Theory of point estimation*, Wiley, New York.

[4] Lehmann, E. L. (1986), *Testing statistical hypothesis*, Wiley, New York.

[5] Menon, M. V. (1963), "Estimation of the shape and scale parameters of the Weibull distribution," *Technometrics*, **5**, pp175-182.

[6] Serfling, R. J. (1980), *Approximation theorems of mathematical statistics*, Wiley, New York.

[7] Stuart, A. (1958), "Note 129: Iteration solutions of likelihood equations," *Biometrics*, **14**, pp.128-130.

뉴톤-랩슨 반복법의 점근비율

이 관 제[1)]

요 약

뉴톤-랩슨 반복법이 최우추정량에 접근하는 비율이 초기값에 따라 가속화함을 보았다. 그러므로 최우추정량을 구하기 어려운 경우에 통계적 목적 - Bahadur 효율, 콰지(Quasi) 우도비 검정 통계량의 점근분포, Bartlett 정정계수(correction factor)등 - 에 따라 뉴톤-랩슨 반복의 횟수를 정하여 쓸 수 있다.

1) (100-715) 서울특별시 중구 필동3가 26, 동국대학교 통계학과