Maximum likelihood methods *

Jean-Marie Dufour †

First version: November 1995
This version: February 15, 2002, 1:24pm

* This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec).

† Canada Research Chair Holder (Econometrics). Centre de recherche et développement en économique (C.R.D.E.), Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour .

Contents

In this text, we use the same notations and assumptions as in Dufour (1995).

1. Maximum likelihood estimators

1.1 Assumption LIKELIHOOD FUNCTION. *Let* $(\mathcal{Z},\ \mathcal{P})$ *be a statistical model which satisfies the following assumptions:*

(**A1**) $(\mathcal{Z},\ \mathcal{P})$ *is a* μ*-dominated model;*

(**A2**) $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$;

(**A3**) $L(z\,;\theta)\,,\ z \in \mathcal{Z}$, *is the density function* (*with respect to* μ) *associated with* P_θ.

1.2 Definition MAXIMUM LIKELIHOOD ESTIMATOR. *Under the assumptions* (*A1*) *to* (*A3*), *a maximum likelihood (ML) estimator of* θ *is any vector* $\widehat{\theta} \in \Theta$ *such that*

$$L\left(Z\,;\widehat{\theta}\right) = \sup_{\theta\in\Theta} L\left(Z\,;\theta\right)\ .$$

Remark: *In certain cases, there may exist several MLE 's or none.*

1.3 Definition MAXIMUM LIKELIHOOD ESTIMATOR OF A PARAMETER TRANSFORMATION. *If* $\widehat{\theta}$ *is a ML estimator of* θ, *then* $\psi(\widehat{\theta})$ *is defined as a ML estimator of* $\psi(\theta)$.

2. Asymptotic properties of maximum likelihood estimators

Under fairly general regularity conditions, it is possible to show that maximum likelihood estimators (MLE) are consistent, asymptotically normally distributed and efficient. The following are simple examples of such conditions.

2.1. Consistency

In addition to **1.1**, let us now consider the following additional assumptions.

2.1 Assumption *The random vectors* $Y_1, \ldots, Y_n$ *are independent and identically distributed with density* $f(y\,;\theta)$, *with respect to a dominating measure* $d\mu_0(y)$, $\theta \in \Theta \subseteq \mathbb{R}^p$.

2.2 Assumption *The unknown true value* θ_0 *of* θ *is identifiable.*

2.3 Assumption *The log-likelihood function*

$$l_n(Z_n\,;\theta) \equiv \log[L_n(Z_n\,;\theta)] = \sum_{t=1}^{n} \log f(Y_t\,;\theta)$$

where $Z_n = (Y_1', \, \dots, Y_n')'$, *is continuous with respect to* θ.

2.4 Assumption *The expected value*

$$\underset{\theta_0}{E}[\log f(Y_t\,;\theta)] = \int [\log f(y\,;\theta)]\, f(y\,;\theta_0) d\mu_0$$

exists and is finite.

2.5 Assumption *The parameter space* Θ *is compact.*

2.6 Assumption *The log-likelihood function* $\frac{1}{n} l_n(Z_n;\theta)$ *converges almost surely to* $\underset{\theta_0}{E}\,[\log f(Y_t\,;\theta)]$ *uniformly on* Θ.

2.7 Assumption θ_0 *belongs to a non-empty open subset of* Θ.

2.8 Assumption *There is a neighborhood of* θ_0 *on which* $\frac{1}{n} l_n(Z_n;\theta)$ *converges almost surely to* $\underset{\theta_0}{E}[\log f(Y_t|\theta)]$ *uniformly.*

2.9 Theorem FIRST MLE CONSISTENCY THEOREM. *Under the assumptions* **1.1** *and* **2.1-2.6***, there is a sequence of maximum likelihood estimators which converges almost surely to* θ_0.

2.10 Theorem SECOND MLE CONSISTENCY THEOREM. *Under the assumptions* **1.1***,* **2.1-2.4** *and* **2.7-2.8***, there exists a sequence of local maxima of the log-likelihood function* $l_n(Z_n\,;\theta)$ *which converges almost surely to* θ_0.

2.11 Theorem THIRD MLE CONSISTENCY THEOREM. *Under the assumptions* **1.1***,* **2.1-2.4** *and* **2.7-2.8***, suppose further that the log-likelihood function is differentiable. Then there is a sequence* $\widehat{\theta}_n$, $n \geq n_0$, *of roots of the equation*

$$S_n(Z_n\,;\widehat{\theta}_n) = 0\,,$$

where $S_n(Z_n\,;\theta) = \partial l_n\,(Z_n\,;\theta)\,/\partial\theta$, *which converges in probability to* θ_0.

2.2. Asymptotic normality

2.12 Assumption *The log-likelihood function* $l_n(Z_n\,;\theta)$ *is twice continuously differentiable in an open neighborhood of* θ_0.

2.13 Assumption *The information matrix*

$$I_f\left(\theta_0\right) = \underset{\theta_0}{E}\left[-\frac{\partial^2 \ln f\left(Y_t\,;\theta_0\right)}{\partial\theta\partial\theta'}\right] = \int \left[-\frac{\partial^2 \ln f\left(y\,;\theta_0\right)}{\partial\theta\partial\theta'}\right]\, f(y\,;\theta_0)d\mu_0$$

associated with the density $f(y\,;\theta_0)$ *exists.*

2.14 Assumption *The information matrix* $I_f\left(\theta_0\right)$ *is nonsingular.*

2.15 Theorem ASYMPTOTIC DISTRIBUTION OF MLE. *Under the assumptions* **1.1**, **2.1-2.4** *and* **2.7-2.14**, *we have*

$$\frac{1}{\sqrt{n}}S_n\left(Z_n\,;\theta_0\right) \underset{n\to\infty}{\overset{d}{\to}} N\left[0\,, I_f\left(\theta_0\right)\right]$$

and any consistent sequence $\widehat{\theta}_n$ *of maximum likelihood estimators of* θ *has the properties:*

$$I_f\left(\theta_0\right)\sqrt{n}(\widehat{\theta}_n - \theta_0) - \frac{1}{\sqrt{n}}S_n\left(Z_n\,;\theta_0\right) \underset{n\to\infty}{\overset{p}{\to}} 0\,,$$

$$I_f\left(\theta_0\right)\sqrt{n}(\widehat{\theta}_n - \theta_0) \underset{n\to\infty}{\overset{d}{\to}} N\left[0\,, I_f\left(\theta_0\right)\right]\,.$$

If furthermore the information matrix $I_f\left(\theta_0\right)$ *is non-singular, we have*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \underset{n\to\infty}{\overset{d}{\to}} N\left[0\,,\ I_f\left(\theta_0\right)^{-1}\right]\,,$$

i.e., $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ *converges in distribution to the normal distribution* $N\left[0\,,\ I_f\left(\theta_0\right)^{-1}\right]$.

3. Tests based on ML estimators

Under the assumptions of Theorem **2.15** with $I_f\left(\theta_0\right)$ non-singular, consider the null hypothesis:

$$H_0\ :\psi\left(\theta\right) = 0 \tag{3.1}$$

where $\psi\left(\theta\right) = \left(\psi_1\left(\theta\right), \ldots, \psi_{p_1}\left(\theta\right)\right)'$ is a $p_1 \times 1$ differentiable function of θ such that

$$\text{rank}\left[P\left(\theta\right)\right] = p_1 \text{ for } \ \theta \in N\left(\theta_0\right) \tag{3.2}$$

where $N(\theta_0)$ is an open neighborhood of θ_0 and

$$P(\theta) = \frac{\partial \psi}{\partial \theta'} = \left[\frac{\partial \psi_i(\theta)}{\partial \theta_j}\right]_{\substack{i=1,\dots,p_1 \\ j=1,\dots,p}} . \tag{3.3}$$

rank $[P(\theta)] = P_1$ for $\theta \in N(\theta_0)$, $N(\theta_0) =$ open neighborhood of θ_0.

By definition

$$\begin{aligned} L_n\big(Z_n\,;\hat{\theta}_n\big) &= \max_{\theta\in\Theta} L_n(Z_n\,;\theta) \\ &\Rightarrow \frac{\partial}{\partial\theta}\left[\log L_n(Z_n\,;\theta)\right]_{\theta=\hat{\theta}_n} = 0 \\ S_n\big(Z_n\,;\hat{\theta}_n\big) &= 0 \\ L\big(Z_n\,;\hat{\theta}_n^0\big) &= \max_{\substack{\psi(\theta)=0 \\ \theta\in\Theta}} L_n(Z_n\,;\theta) \ . \end{aligned}$$

To find $\hat{\theta}_n^0$, we consider

$$\begin{aligned} \mathcal{L} &= \log\left[L_n(Z_n\,;\theta)\right] - \psi(\theta)'\lambda \\ &\Rightarrow \frac{\partial}{\partial\theta}\left[\log L_n(Z_n\,;\theta)\right]_{\theta=\hat{\theta}_n^0} = P\big(\hat{\theta}_n^0\big)'\hat{\lambda}_n \\ \psi\big(\hat{\theta}_n^0\big) &= 0 \\ &\Rightarrow S_n\big(Z_n\,;\hat{\theta}_n^0\big) = P\big(\hat{\theta}_n^0\big)'\hat{\lambda}_n \ . \end{aligned}$$

Under $H_0 : \psi(\theta_0) = 0$,

$$\begin{aligned} &\sqrt{n}\,\psi\big(\hat{\theta}_n\big) \xrightarrow[n\to\infty]{d} N\left[0\,, P(\theta_0)\, I_f(\theta_0)^{-1}\, P(\theta_0)'\right] \\ &\frac{1}{\sqrt{n}}\hat{\lambda}_n \xrightarrow[n\to\infty]{d} N\left[0\,, [P(\theta_0)\, I_f(\theta_0)^{-1}\, P(\theta_0)']^{-1}\right] \\ &n\,\psi\big(\hat{\theta}_n^0\big)'\left[P(\theta_0)\, I_f(\theta_0)^{-1}\, P(\theta_0)'\right]^{-1}\psi\big(\hat{\theta}_n\big) \xrightarrow[n\to\infty]{d} \chi^2(p_1) \\ &\frac{1}{n}\hat{\lambda}_n' P(\theta_0)\, I_f(\theta_0)^{-1}\, P(\theta_0)'\,\hat{\lambda}_n \xrightarrow[n\to\infty]{d} \chi^2(p_1) \end{aligned}$$

3.1. Test criteria

1. **Likelihood ratio**

$$LR_n(\psi) = 2\left[\log L_n\big(Z_n\,;\hat{\theta}_n\big) - \log L_n\big(Z_n\,;\hat{\theta}_n^0\big)\right]$$

$$
\begin{aligned}
&= 2\left[\ell_n\big(Z_n\,;\hat{\theta}_n\big) - \ell_n\big(Z_n\,;\hat{\theta}^0_n\big)\right] \\
&= 2\,\log\left[L_n\big(Z_n\,;\hat{\theta}_n\big)/L_n\big(Z_n\,;\hat{\theta}^0_n\big)\right].
\end{aligned}
$$

2. **Wald**

$$
W_n(\psi) = n\,\psi\big(\hat{\theta}_n\big)'\big[P\big(\hat{\theta}_n\big)\widehat{I}_f\big(\hat{\theta}_n\big)P\big(\hat{\theta}_n\big)'\big]^{-1}\psi\big(\hat{\theta}_n\big).
$$

3. **Lagrange multiplier (Rao's score)**

$$
\begin{aligned}
LM_n(\psi) &= \frac{1}{n}\hat{\lambda}_n' P\big(\hat{\theta}^0_n\big)\widehat{I}_f\big(\hat{\theta}^0_n\big)^{-1}P\big(\hat{\theta}^0_n\big)'\hat{\lambda}_n \qquad (3.4)\\
&= \frac{1}{n}S_n\big(Z_n\,;\hat{\theta}^0_n\big)'\widehat{I}_f\big(\hat{\theta}^0_n\big)^{-1}S_n\big(Z_n\,;\hat{\theta}^0_n\big)
\end{aligned}
$$

where

$$
S_n\big(Z_n\,;\hat{\theta}^0_n\big) = P\big(\hat{\theta}^0_n\big)'\hat{\lambda}_n\,. \qquad (3.5)
$$

4. **Neyman's** $\mathrm{C}(\alpha)$ **test _** Let $\tilde{\theta}^0_n$ be any estimator such that

   (a) $\psi\big(\tilde{\theta}^0_n\big) = 0,$

   (b) $\sqrt{n}\big(\tilde{\theta}^0_n - \theta_0\big)$ has an asymptotic distribution ($\tilde{\theta}^0_n$ is root-n consistent under H_0).

   Neyman's $C(\alpha)$ statistic for $\psi(\theta) = 0$ is

$$
\begin{aligned}
PC\big(\tilde{\theta}^0_n\,;\psi\big) &= \frac{1}{n}S_n\big(\tilde{\theta}^0_n\,;Z_n\big)'\widehat{I}_f\big(\tilde{\theta}^0_n\big)^{-1}P\big(\tilde{\theta}^0_n\big)\big[P\big(\tilde{\theta}^0_n\big)\widehat{I}_f\big(\tilde{\theta}^0_n\big)^{-1}P\big(\tilde{\theta}^0_n\big)'\big]^{-1} \qquad (3.6)\\
&\quad P\big(\tilde{\theta}^0_n\big)\widehat{I}_f\big(\tilde{\theta}^0_n\big)^{-1}S_n\big(\tilde{\theta}^0_n\,;Z_n\big) \qquad (3.7)
\end{aligned}
$$

When $\tilde{\theta}^0_n = \hat{\theta}^0_n$,

$$
PC\big(\tilde{\theta}^0_n\,;\psi\big) = LM_n(\psi)\,.
$$

3.2. Estimators of information matrix

Let $\tilde{\theta}_n$ an estimator of θ. There are three main estimators of the information matrix $\overline{I}(\theta_0)$.

1. Hessian:

$$
\hat{I}_f\big(\tilde{\theta}_n\big)_H = -H\big(\tilde{\theta}\,;Z_n\big) = -\frac{1}{n}\frac{\partial \ell_n\big(\tilde{\theta}_n\,;Z_n\big)}{\partial\theta\partial\theta'}\;; \qquad (3.8)
$$

2. outer product:

$$\hat{I}_f\big(\tilde{\theta}_{n0})\big) = \frac{1}{n}\sum_{t=1}^{n} D_t\big(\tilde{\theta}_n\,; Y_t\big) D_t\big(\tilde{\theta}_n\,; Y_t\big)' \tag{3.9}$$

where

$$D_t\,(\theta\,; Y_t) \;=\; \frac{\partial}{\partial\theta}\,[\log F\,(Y_t\,;\theta)]\ , \tag{3.10}$$

$$I_f\,(\theta) \;=\; V\,[D_t\,(\theta\,; Y_t)]\ ; \tag{3.11}$$

3. expected information:

$$\hat{I}_f\big(\tilde{\theta}_n\big)_E = \hat{I}_f\big(\tilde{\theta}_n\big)\,. \tag{3.12}$$

Under general regularity conditions, $\tilde{\theta}_n \underset{n\to\infty}{\overset{P}{\to}} \theta_0$ entails

$$\underset{n\to\infty}{\text{plim}}\,\hat{I}\big(\tilde{\theta}_n\big)_H = \underset{n\to\infty}{\text{plim}}\,\hat{I}_1\big(\tilde{\theta}_n\big) = \underset{n\to\infty}{\text{plim}}\,\hat{I}_1\big(\tilde{\theta}_n\big)_E = \overline{I}\,(\theta_0)\ . \tag{3.13}$$

Provided the latter property holds,

$$LR_n\,(\psi)\,,\ W_n\,(\psi)\,,\ LM_n\,(\psi)\,,\ PC\big(\tilde{\theta}_n^0\,;\psi\big)$$

follow under $H_0 : \psi\,(\theta) = 0$

$\chi^2\,(p_1)$ distributions

as $n \to \infty$.

References

DUFOUR, J.-M. (1995): "Statistical Models and Likelihood Functions," Lecture notes, Département de sciences économiques, Université de Montréal.

GOURIÉROUX, C., AND A. MONFORT (1989): *Statistique et modèles économétriques, Volumes 1 et 2*. Economica, Paris.

LEHMANN, E. L. (1983): *Theory of Point Estimation*. John Wiley & Sons, New York.

——— (1986): *Testing Statistical Hypotheses, 2nd edition*. John Wiley & Sons, New York.