The Stata Journal (2006)
6, Number 2, pp. 153–155

Maximum simulated likelihood: Introduction to a special issue

Maximum simulated likelihood (MSL) makes previously intractable estimators computationally feasible. Faster computers and new simulation techniques are moving MSL estimators into the mainstream toolbox. This issue of the *Stata Journal* contains several papers on this subject. I organized this special issue to bring together work being done by people inside and outside StataCorp and to highlight Stata's tools for implementing MSL methods.

Many likelihoods require you to evaluate high-dimensional integrals. Although modern adaptive-quadrature methods have extended the set of feasible problems, many real-world problems involve approximating integrals that cannot be approximated by quadrature methods in a reasonable amount of time.

Although simulation techniques can be used to approximate the high-dimensional integrals, only the speed of modern computers and recent advances in simulation methods have made this task practicable. Estimators obtained by maximizing likelihoods that are approximated by simulation techniques are known as MSL estimators.[1]

Several authors have shown that MSL estimators have the same large-sample properties as maximum likelihood, as long as the number of repetitions, R, used to approximate the integral grows faster than the square root of the number of observations in the sample, $\sqrt{N}$.[2]

To provide some intuition for the restriction that $\sqrt{N}/R \rightarrow 0$, suppose that the likelihood of interest $L(\mathbf{Z}, \boldsymbol{\theta})$ depends on an integral that has no closed form, where $\mathbf{Z}$ is the data and $\boldsymbol{\theta}$ is a vector of parameters. MSL estimators generally obtain $\widehat{L}(\mathbf{Z}, \boldsymbol{\theta})$, an unbiased approximation to $L(\mathbf{Z}, \boldsymbol{\theta})$, by averaging R draws from the underlying distribution. However, $\ln\{\widehat{L}(\mathbf{Z}, \boldsymbol{\theta})\}$ is not unbiased for $\ln\{L(\mathbf{Z}, \boldsymbol{\theta})\}$, and $\ln\{\widehat{L}(\mathbf{Z}, \boldsymbol{\theta})\}$ is the objective function that MSL maximizes. Although this bias does not affect the consistency of the MSL estimates, it does affect the large-sample variance–covariance of the estimator, unless $\sqrt{N}/R \rightarrow 0$.

When you are analyzing a specific dataset with N observations, the requirement that $\sqrt{N}/R \rightarrow 0$ implies that you can choose any real number $a > .5$, yielding $R = \lfloor N^a \rfloor$, where $\lfloor\, \rfloor$ returns the greatest integer less than or equal to N^a. This ambiguity is addressed to various extents by most of the MSL-related papers in this issue.

1. Some authors, such as Gouriéroux and Monfont (1996), call them simulated maximum-likelihood estimators.
2. See Pakes and Pollard (1989) and McFadden (1989) for the original results. Also see Cameron and Trivedi (2005), Gouriéroux and Monfont (1996), and Train (2003) for excellent introductions to the statistical theory for maximum likelihood and MSL.

Much of the MSL-related literature addresses how to obtain unbiased approximations to $L(\mathbf{Z}, \boldsymbol{\theta})$. A key discovery has been that some simulators can approximate an integral with fewer repetitions if certain correlated deterministic sequences are used instead of pseudorandom numbers. Three of the six MSL papers in this issue discuss Stata implementations of the most important methods. Cappellari and Jenkins (2006) and Gates (2006) discuss the most commonly used method for approximating the multivariate normal distribution by simulation and their distinct Stata implementations. These two papers also discuss how this simulator relies on sequences of numbers between 0 and 1 to approximate the multivariate normal. Cappellari and Jenkins (2006) and Drukker and Gates (2006) discuss the deterministic sequences of correlated numbers that Bhat (2001) and Train (2000) have shown to dramatically reduce the number of simulations required to approximate some likelihoods.

Although Gates (2006), Drukker and Gates (2006), and Cappellari and Jenkins (2006) cover some of the same topics, their treatments and accompanying commands are different. Cappellari and Jenkins (2006) provide more examples of how to use the methods, whereas Gates (2006) and Drukker and Gates (2006) provide more detail on how and why the methods work. The Gates routines provide some important functionality not available in the Cappellari–Jenkins version.[3]

Haan and Uhlendorff (2006) compare the computational burden of an MSL estimator to a standard maximum likelihood estimator that approximates the integral by adaptive quadrature. Although the MSL estimator does surprisingly well for such a low-dimensional problem, their simulation methodology shows how researchers can determine an R for their datasets. The idea is to start with a small R, say, $R = \lfloor N^{.55} \rfloor$, and then repeat the estimation with higher values of R until the point estimates and the log likelihood settle down. This method is analogous to the one used by [XT] **quadchk** to determine the correct number of quadrature points.

Because of the high-dimensional integrals, MSL is often applied to construct estimators for the parameters of multiple discrete-outcome models. Deb and Trivedi (2006) and Stewart (2006) present Stata commands that implement MSL estimators in this vein. Deb and Trivedi (2006) discuss their MSL estimator for the parameters of a negative binomial regression model with a multinomial, endogenous treatment and their `mtreatnb` command, which implements the estimator in Stata. Stewart (2006) presents his `redpace` command for estimating the parameters of random-effects dynamic probit models with autocorrelated errors by MSL. Stewart (2006) also presents the method of choosing R discussed above and some related results showing how much the point estimates can vary with seed.

David M. Drukker, Guest Editor
StataCorp
College Station, TX
ddrukker@stata.com

3. In particular, the program discussed in Gates (2006) produces Hammersley as well as Halton sequences, and the simulator discussed in Drukker and Gates (2006) offers derivatives not available in the Cappellari–Jenkins version.

1 References

Bhat, C. R. 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research, Part B* 35: 677–693.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cappellari, L., and S. P. Jenkins. 2006. Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *Stata Journal* 6: 156–189.

Deb, P., and P. K. Trivedi. 2006. Maximum simulated likelihood estimation of a negative binomial regression model with multinomial endogenous treatment. *Stata Journal* 6: 246–255.

Drukker, D. M., and R. Gates. 2006. Generating Halton sequences using Mata. *Stata Journal* 6: 214–228.

Gates, R. 2006. A Mata Geweke–Hajivassiliou–Keane multivariate normal simulator. *Stata Journal* 6: 190–213.

Gouriéroux, C., and A. Monfont. 1996. *Simulation-Based Econometric Methods*. Oxford: Oxford University Press.

Haan, P., and A. Uhlendorff. 2006. Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood. *Stata Journal* 6: 229–245.

McFadden, D. 1989. A method of simulated moments for estimation of discrete reponse models with numerical integration. *Econometrica* 57: 995–1026.

Pakes, A., and D. Pollard. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57: 1027–1057.

Stewart, M. 2006. Maximum simulated likelihood estimation of random-effects dynamics probit models with autocorrelated errors. *Stata Journal* 6: 256–272.

Train, K. 2000. Halton sequences for mixed logit. Economics Department, University of California, Berkeley, Working Paper E00-278.
http://repositories.cdlib.org/iber/econ/E00-278/.

———. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.