Empirical Likelihood

marian.grendar@savba.sk[1]

Robust 2010, Králíky.cz, 31.1.-5.2.2010

[1]KM FPV UMB BB

Plan

- Likelihood & contrast: parametric model
- Likelihood & contrast: moment condition model
  - Estimating Equations
  - Empirical Estimating Equations (E^3)
  - Which contrast?
  - Existence problems of the E^3 approach

Likelihood and contrast: parametric model

Setup

Setting:

Chance: r.v. $X \in \mathscr{X} \subseteq \mathrm{R}^d$, with cdf $Q_r(x)$

Model: $\Phi(\Theta) = \{Q(x;\theta) : \theta \in \Theta\}$;
parametric space: $\Theta \subseteq \mathrm{R}^K$, K finite.

Data: $X_1^n = X_1, \ldots, X_n$, iid Q_r.

Contrast, Generalized Minimum Contrast estimators

Contrast (f-divergence):

$$D_\phi(Q \,||\, Q_r) = \mathrm{E}_{Q_r} \phi\left(\frac{dQ(x;\theta)}{dQ_r}\right),$$

where $\phi(\cdot)$ is a convex function, with minimum at 1.

The *Generalized Minimum Contrast* (GMC) estimator

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} D_\phi(Q \,||\, \tilde{Q}_r).$$

where $\tilde{Q}_r$ is a nonparametric estimator of Q_r.

Likelihood, Maximum Likelihood as GMC

1) Let $\phi(x) = -\log(x)$.

2) Let $\tilde{Q}_r$ be the empirical cdf $\hat{Q}_r(x) = \frac{\sum_{i=1}^{n} I(X_i \leq x)}{n}$.

Then $D_\phi(Q \,||\, \hat{Q}_r)$ is the *log-likelihood function* and

$$\hat{\theta}_{\mathrm{ML}} = \arg\inf_{\theta \in \Theta} D_\phi(Q \,||\, \hat{Q}_r)$$

is the *Maximum Likelihood estimator* $\hat{\theta}_{\mathrm{ML}}$ of θ.

Hellinger contrast, Beran estimator

Other choices of $\phi(\cdot)$ require a smooth nonparametric estimate $\tilde{Q}_r$ of Q_r (e.g., kernel density).

If $\phi(x) = -2(\sqrt{x} - 1)$, then the GMC estimator is

$$\hat{\theta} = \arg\inf_{\theta \in \Theta} 2\left(1 - \mathrm{E}\sqrt{q(x;\theta)}\sqrt{\tilde{q}_r(x)}\right),$$

i.e., *Beran's* minimum Hellinger distance *estimator*.

Likelihood Ratio Test

Partition Θ into Θ_0 and Θ_0^c.
Test of $H_0 : \theta \in \Theta_0$.

Log-Likelihood Difference Statistic:

$$\lambda(\theta; X_1^n) = \inf_{\theta \in \Theta_0} D(Q \,||\, \hat{Q}_r) - \inf_{\theta \in \Theta} D(Q \,||\, \hat{Q}_r)$$

Wilks Thm: Under some regularity conditions, asymptotically

$$-2\,\lambda(\theta; X_1^n) \sim \chi_K^2.$$

Other discrepancies, cf. eg. Broniatowski & Keziou (2009).

Likelihood and contrast: moment condition model

Estimating Equations

Setup:

Chance: r.v. $X \in \mathscr{X} \subseteq \mathrm{R}^d$, with cdf $Q_r \in \mathscr{Q}(\mathscr{X})$, where $\mathscr{Q}(\mathscr{X})$ is the set of all cdf's on $\mathscr{X}$.

Model:
Estimating functions: $u(X;\theta) : \mathscr{X} \times \Theta \to \mathrm{R}^J$, where $\theta \in \Theta \subseteq \mathrm{R}^K$; K can be, in general, different than J.
Estimating equations (EE):

$$\Phi(\theta) = \{Q \in \mathscr{Q}(\mathscr{X}) : \mathrm{E}_Q u(X;\theta) = 0\}.$$

Model: $\Phi(\Theta) \; = \bigcup_{\theta\in\Theta} \Phi(\theta)$.

Estimating Equations: examples

Examples:

Ex. 1: $\mathscr{X} = \mathrm{R}$, $\Theta = [0, \infty)$, $u(X;\theta) = X - \theta$.

Ex. 2: (Brown & Chen) $\mathscr{X} = \mathrm{R}$, $\Theta = \mathrm{R}$,
$u(X;\theta) = \{X - \theta, \mathrm{sgn}(X - \theta)\}$.

Ex. 3: (Qin & Lawless) $\mathscr{X} = \mathrm{R}$, $\Theta = \mathrm{R}$,
$u(X;\theta) = \{X - \theta, X^2 - (2\theta^2 + 1)\}$.

Empirical Estimating Equations

To connect the model $\Phi(\Theta)$ with the data X_1^n, replace the model $\Phi(\Theta)$ by its empirical, data-based analogue
$\Phi_n(\Theta) = \bigcup_{\theta\in\Theta} \Phi_n(\theta)$, where

$$\Phi_n(\theta) = \left\{ Q_n \in \mathscr{Q}(X_1^n) : \mathrm{E}_{Q_n} u(X;\theta) = 0 \right\}$$

are the empirical estimating equations.

Empirical Estimating Equations (E^3) approach to estimation and inference replaces the set $\Phi(\Theta)$ of cdf's supported on $\mathscr{X}$ by the set $\Phi_n(\Theta)$ of cdf's that are supported on the data X_1^n.

An estimate $\hat{\theta}$ of θ_r is obtained by means of a rule (e.g., GMC) that selects $\hat{Q}_n(x;\hat{\theta})$ from $\Phi_n(\Theta)$.

E^3-*GMC estimator*

Data: $X_1^n = X_1, \ldots, X_n \sim Q_r$.

Use GMC to select $\hat{Q}_n(x;\hat{\theta})$ from $\Phi_n(\Theta)$:

$$\hat{Q}_n(x;\hat{\theta}) = \arg \inf_{Q_n(x;\theta)\in\Phi_n(\Theta)} D_\phi(Q_n \,||\, \hat{Q}_r) \qquad (1)$$

GMC rule is used to select a member of the E^3 model set $\Phi_n(\theta)$, that is closest to the empirical cdf $\hat{Q}_r$, in the sense of $D_\phi(\cdot\,||\,\cdot)$.

Other rules: Cressie Read class of divergences – Generalized Empirical Likelihood class of estimators.

E^3-*GMC estimator: convex dual form*

The θ part of the optimization problem (1):

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} \inf_{Q_n(x) \in \Phi_n(\theta)} \mathrm{E}_{\hat{Q}_r} \phi \left(\frac{dQ}{d\hat{Q}_r} \right), \tag{2}$$

The convex dual form of (2):

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} \sup_{\mu \in \mathrm{R}, \lambda \in \mathrm{R}^J} \left[\mu - \mathrm{E}_{\hat{Q}_r} \phi^*(\mu + \lambda' u(x;\theta)) \right], \tag{3}$$

where $\phi^*(y) = \sup_x xy - \phi(x)$ is the Legendre Fenchel transformation of $\phi^*(x)$.

Maximum Empirical Likelihood as E^3*-GMC estimator*

To get *Maximum Empirical Likelihood* (MEL) use
$\phi(x) = -\log x$

$$\hat{\theta}_{\mathrm{MEL}} = \arg \inf_{\theta \in \Theta} \sup_{\lambda \in \mathrm{R}^J} \mathrm{E}_{\hat{Q}_r} \log(1 + \lambda' u(x;\theta)).$$

MEL selects among the data-supported cdf's from the model $\Phi_n(\Theta)$ the one with the highest value of the likelihood.
MEL = ML on $\Phi_n(\Theta)$.

Asymptotic properties of MEL

Qin & Lawless '94: Under some regularity conditions, which include the assumption that the model is correctly specified (i.e., $Q_r \in \Phi(\Theta)$)

$$n^{1/2}(\hat{\theta}_{\mathrm{EML}} - \theta_r) \to_d n(0, \Sigma),$$

where

$$\Sigma = \left[\mathrm{E}\frac{\partial u'}{\partial \theta}(\mathrm{E}uu')^{-1}\mathrm{E}\frac{\partial u}{\partial \theta} \right],$$

and θ_r solves $\mathrm{E}_{Q_r} u(X;\theta) = 0$.

Empirical Likelihood Ratio

A GMC test of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$ is based on the GMC statistic

$$\lambda(\theta; X_1^n) = \inf_{Q \in \Phi_n(\theta_0)} D_\phi(Q \,||\, \hat{Q}_r) - \inf_{Q \in \Phi_n(\theta_1)} D_\phi(Q \,||\, \hat{Q}_r).$$

Thm: If θ_0 satisfies $\mathrm{E}u(X; \theta_0) = 0$, then under some regularity conditions,

$$k_n(\phi)\lambda(X_1^n; \theta_0) \to \chi_K^2.$$

Empirical Likelihood Ratio test: take $\phi(\cdot) = -\log(\cdot)$ and $k_n(\phi) = 2$.

E^3 & MEL: 'engineering' form

E^3 model $\Phi_q(\Theta) = \bigcup_{\theta\in\Theta} \Phi_q(\theta)$ where

$$\Phi_q(\theta) = \{q(x;\theta) : \sum_{i=1}^{n} q(x_i;\theta)u(x_i;\theta) = 0; \sum_{i=1}^{n} q(x_i;\theta) = 1; \\ 1 \geq q(\cdot;\theta) \geq 0\};$$

i.e., the set of probability mass functions that are supported on the data, and satisfy the estimating equations.

MEL selects:

$$\hat{q}(\cdot;\hat{\theta})_{\mathrm{MEL}} = \arg \sup_{q(\cdot;\theta)\in\Phi_q(\Theta)} \frac{1}{n}\sum_{i=1}^{n} \log q(x_i,\theta).$$

Random selection of recent works on EL

- Adjusted Empirical Likelihood and its properties
- Efficient nonparametric estimation of causal effects in randomized trials with noncompliance
- Using Empirical Likelihood to combine data: application to food risk assessment
- Optimally combined censored and uncensored datasets
- Extending the scope of Empirical Likelihood
- Generalized Empirical Likelihood-based model selection criteria for moment conditions models
- Adjusted Exponentially Tilted Likelihood with applications to brain morphology
- Combining quantitative trait loci analyses and microarray data: an empirical likelihood approach
- Empirical Likelihood based diagnostics for heteroscedasticity in partial linear models

Questions

1) Are all the MEL-like methods equal?

2) How to extend MEL and MEL-like methods into a Bayesian method?

Answer: Bayesian infinite dimensional consistency under misspecification.

Bayesian infinite dimensional consistency

A prior Π is put on $\Phi(\Theta)$; (it induces a prior $\Pi(\theta)$ over Θ). The prior Π combines with the data X_1^n to define the posterior:

$$\Pi_n(A|X_1^n) = \frac{\int_A e^{-l_n(Q)}\Pi(dQ)}{\int_\Phi e^{-l_n(Q)}\Pi(dQ)},$$

where $l_n(Q) \triangleq -\mathrm{E}_{\hat{Q}_r} \log \frac{dQ}{d\hat{Q}_r}$, and $A \subseteq \Phi$.

Bayesian infinite-dimensional consistency: the objective – to determine the distribution(s) on which the posterior Π concentrates as n gets large.

Bayesian consistency under misspecification

If the model is not correctly specified, i.e., $Q_r \notin \Phi(\Theta)$, then θ_r can be defined as the value corresponding to the distribution $\hat{Q}(Q_r)$ on which the posterior concentrates.

Bayesian Law of Large Numbers

BLLN. (g & Judge, 09) Under some regularity conditions the posterior concentrates on the union of weak ϵ-balls that are centered at the L-projections $\hat{Q}$ of Q_r on Φ.

The L-projection $\hat{Q}$ of Q_r on Φ

$$\hat{Q} = \arg \inf_{Q \in \Phi} L(Q \,||\, Q_r),$$

where $L(Q \,||\, Q_r)$ is the L-divergence of Q wrt Q_r

$$L(Q \,||\, Q_r) = -\mathrm{E}_{Q_r} \log \frac{dQ}{dQ_r}.$$

The BLLN is an extension of Schwartz' consistency theorem, to the case of misspecified model.

Answers

Recall that E^3-GMC is

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} \inf_{Q_n(x) \in \Phi_n(\theta)} \mathrm{E}_{\hat{Q}_r} \phi\left(\frac{dQ}{d\hat{Q}_r}\right).$$

1) BLLN implies that MEL (i.e., $\phi(x) = -\log x$) is consistent under misspecification; other GMC and GEL methods are not.

2) MEL and Bayesian Maximum A-Posteriori (MAP) estimator asymptotically coincide.

Existence problems of E^3

- Convex hull problem
- Empty Set Problem (g & Judge, 09)

Convex hull constraint

Test of hypothesis $H_0 : \theta = \theta_0$.
For any sample X_1^n such that $\Phi_n(\theta_0) = \emptyset$ no EL test exists.

Ex: $\mathscr{X} = \mathrm{R}$, $\Theta = \mathrm{R}$, $u(X;\theta) = X - \theta$.
The MEL estimator is $\mathrm{E}_{\hat{Q}_r} X$, i.e., the sample mean.
Test the point hypothesis $\Theta_0 = \{\theta_0\}$.
If θ_0 lays outside the convex hull of the data
(i.e., MEL estimate $\hat{q}(\cdot;\theta_0)$ does not exist under the restriction),
then no EL (or EL-like) test can be constructed.

A way out: Modified E^3 *(*mE^3*)*

Modification: the non-negativity constraints on q are dropped out.

Modified Empirical Estimating Equations (mE^3) model:
$\Phi(\theta)$ is replaced with

$$\Phi_q^m(\theta) = \left\{ q(x;\theta) : \sum_{i=1}^{n} q(x_i;\theta) u(x_i;\theta) = 0; \sum_{i=1}^{n} q(x_i;\theta) = 1 \right\}.$$

So that the model $\Phi_q^m(\Theta) = \bigcup_{\theta\in\Theta} \Phi_q^m(\theta)$.

Among others, the *Euclidean Empirical Likelihood* (i.e., GMC with $\phi(x) = 1/2(x^2 - 1)$) allows the negative weights q.

Empty Set Problem

$\Phi_n(\Theta)$ is data-dependent.

Empty Set Problem (ESP): There are E^3 models for which the set $\Phi_n(\Theta) = \emptyset$, for some X_1^n.

Affine Empty Set Problem (aESP): There are the modified mE^3 models for which the set $\Phi_q^m(\Theta) = \emptyset$, for some X_1^n.

Some examples of models with ESP

QL = Qin & Lawless, '94

- QL, Example 1
- QL, Example 2
- QL, Example 3
- restricted parameter space

QL, Example 1

Estimating functions:

$$u_1(X;\theta) = X - \theta,$$
$$u_2(X;\theta) = X^2 - (2\theta^2 + 1).$$

So that

$$\Phi_q(\theta) = \Big\{ q(x;\theta) : \sum_{i=1}^{n} q(x_i;\theta)u_1(x_i;\theta) = 0;$$
$$\sum_{i=1}^{n} q(x_i;\theta)u_2(x_i;\theta) = 0; \sum_{i=1}^{n} q(x_i;\theta) = 1; q(x_i;\theta) \geq 0, 1 \leq i \leq n \Big\}.$$

QL, Example 1, cont'd

Condition on data for which $\Phi_q(\Theta)$ is empty:
if the data set X_1^n is such that the LHS in

$$\sum_{i=1}^{n} q(x_i;\cdot)x_i^2 - 2\left(\sum_{i=1}^{n} q(x_i;\cdot)x_i\right)^2 = 1. \tag{4}$$

can attain the RHS' value 1 for no q.

The maximal value of LHS in Eq. (4) can be attained for such q that the only non-zero elements of $q(x;\cdot)$ are $q_{(1)} \triangleq q(x_{(1)};\cdot)$ and $q_{(n)} \triangleq q(x_{(n)};\cdot)$; there $x_{(1)}$ denotes the lowest, $x_{(n)}$ the largest value in X_1^n.

QL, Example 1, cont'd

Without the non-negativity constraints the maximal value of the LHS of Eq. (4), denoted v, is attained for

$$\hat{q}^{m}_{(1)} = \frac{\frac{x_{(1)}^2 - x_{(n)}^2}{4(x_{(1)} - x_{(n)})} - x_{(n)}}{x_{(1)} - x_{(n)}},$$

if $x_{(1)} \neq x_{(n)}$. Then $\hat{q}^{m}_{(n)} = 1 - \hat{q}^{m}_{(1)}$.

If the data $X_1^n \sim r_X(x;\theta)$ are such that v is smaller than 1 (i.e., the RHS of Eq. (4)) then $\Phi^m_q(\Theta)$ as well as $\Phi_q(\Theta)$ is empty for such data.
Consequently, for such data there is no mE^3-based or E^3-based estimator.

QL, Example 1, cont'd

For a given $r_X(x;\theta)$ the probability of data set with ESP is $\Pr(v < 1)$. It can be estimated by MC.

Q&L: $Q_r(x;\theta) = n(0,1)$, and $n = 15$. Then the probability is 0.0173 (estimated by 10000 MC runs). Thus, in 17 of 1000 samples of size $n = 15$ drawn from $n(0,1)$ it is meaningless to look for EL, or any other E^3-based (or mE^3-based) estimate.

QL, Example 2

Bivariate observations $(X, Y)_1^n$, such that $\mathrm{E}(X) = \mathrm{E}(Y) = \theta$, $\theta \in \Theta = \mathrm{R}$.

Bivariate estimating function $u(x, y; \theta) = (X - \theta, Y - \theta)$.

$\Phi_q(\Theta)$ will be empty for every sample, such that $X_i - Y_i > 0$, or $X_i - Y_i < 0$, for all $i = 1, \dots, n$, as Q&L note.

If $X \sim n(-0.3, 0.1)$ and $Y \sim n(0, 0.1)$, then for $n = 10$ the probability that $\Phi_q(\Theta)$ is empty is 0.056.

QL, Example 3

$\Theta = \{a\}$, where $a \in \mathrm{R}$ is known.

Estimating function: $u(X) = X - a$.

ESP: $\Phi_q(\Theta)$ is empty for any data set that contains only values greater (smaller) than a.

Illustration: $r_X(x) = [0.025,\ 0.025,\ 0.15,\ 0.8]$ is a pmf on $\mathscr{X} = \{1, 2, 3, 4\}$. Let $a = 2.0$, $n = 40$. The probability that $\Phi_q(2)$ is empty is 0.129.

Restricted parameter space

Example: $u(X;\theta) = X - \theta$, where $\theta \in \Theta = [0, \infty)$.

The set $\Phi_q(\Theta)$ is empty for any data set $X_1^n < 0$.

A model without ESP

Brown & Chen, '98: estimation of a location parameter by a data-based combination of the mean and the median.

E^3 model:

$$\Phi_q(\theta) = \Big\{ q(x;\theta) : \sum_{i=1}^{n} q(x_i;\theta)(x_i - \theta) = 0;$$

$$\sum_{i=1}^{n} q(x_i;\theta)\mathrm{sgn}(x_i - \theta) = 0; \sum_{i=1}^{n} q(x_i;\cdot) = 1; q(x_i;\cdot) \geq 0, 1 \leq i \leq n \Big\},$$

and $\theta \in \Theta = \mathrm{R}$. No problem of the empty set. In this case MEL always exists.
(However, EL c.i./test need not exist for every θ).

Summary

- There are E^3 models that are subject to ESP.
- There are mE^3 models with the affine ESP. For such models the escape route of lifting up the non-negativity constraints does not work.
- There are models which are free of ESP.

Implications of ESP for E^3 *based methods*

- If the E^3 and mE^3-based methods are to be used also in the future applications, the models should be checked on case-by-case basis for ESP, aESP.

References

- Bickel, P. J., Klaassen, C.A.J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- Brown, B. M., and Chen, S. X. (1998). Combined and Least Squares Empirical Likelihood. *Ann. Inst. Statist. Math.* 90:443-450.
- Corcoran, S. A. (2000). Empirical exponential family likelihood using several moment conditions. *Stat. Sinica*. 10:545-557.
- Grendár, M. and Judge, G. (2009). Asymptotic Equivalence of Empirical Likelihood and Bayesian MAP. *Ann. Statist.* 37/5A:2445-2457.

References (cont'd)

- Grendár, M. and Judge, G. (2009). Empty Set Problem of Maximum Empirical Likelihood Methods. *Electron. J. Statist.* 3:1542-1555.
- Kitamura, Y. (2007). Nonparametric Likelihood: efficiency and robustness. *Jap. Econ. Rev.* 58/1:26-45.
- Owen, A. B. (2001). *Empirical Likelihood*. New York:Chapman-Hall/CRC.
- Qin, J., and Lawless, J. (1994). Empirical Likelihood and General Estimating Equations. *Ann. Statist.* 22:300-325.