Algebraic computations for asymptotically efficient estimators via information geometry

Kei Kobayashi (Institute of Statistical Mathematics)
with Henry P. Wynn (London School of Economics)

7/April/2011 @ WOGAS3

Motivation for Algebraic Information Geometry

- Differential geometrical objects (e.g. Fisher metric, connections, embedding curvatures and divegences) can sometimes be computed by algebraic computations.
- Statistical objects (e.g. Estimator, Bias term of estimators and Risk) can be computated by algebraic computations.

Most of the existing results on asymptotic theory for algebraic models are focusing on the singularity.

Motivation for Algebraic Information Geometry

- Differential geometrical objects (e.g. Fisher metric, connections, embedding curvatures and divegences) can sometimes be computed by algebraic computations.
- Statistical objects (e.g. Estimator, Bias term of estimators and Risk) can be computated by algebraic computations.

Most of the existing results on asymptotic theory for algebraic models are focusing on the singularity.

Key point

We can do something completely new even for the non-singular classical asymptotic theory.

Exponential Family

Full and Curved Exponential Family for Sufficient Statistics

$$\mathrm{d}P(x|\theta) = \exp(x_i\theta^i - \psi(\theta))\mathrm{d}\nu$$

Full exponential family : $\{\mathrm{d}P(x|\theta) \mid \theta \in \Theta\}$ for an open $\Theta \subset \mathbb{R}^d$.
$x \in \mathbb{R}^d$: a variable representing a sufficient statistics
ν : a carrier measure on $\mathbb{R}^d$
Curved exponential family : $\{\mathrm{d}P(x|\theta) \mid \theta \in \mathcal{V}_\Theta\}$ for a (not necessarily smooth) $\mathcal{V}_\Theta \subset \Theta$.

- We call θ a natural parameter and $\eta = \eta(\theta) := E[x|\theta]$ an expectation parameter.
- $E = E(\Theta) := \{\eta(\theta) \mid \theta \in \Theta\} \subset \mathbb{R}^d$
- $\mathcal{V}_E := \{\eta(\theta) \mid \theta \in \mathcal{V}_\Theta\} \subset E$
- $\eta(\theta) = \nabla_\theta\psi(\theta)$

Algebraic Curved Exponential Family

We say a curved exponential family is algebraic if the following two conditions are satisfied:

1) Θ or E is represented by a real algebraic variety, i.e. $\Theta = \mathcal{V}_\Theta := \mathcal{V}(\langle f_1, \dots, f_k \rangle) = \{\theta \in \mathbb{R}^d | f_1(\theta) = \cdots = f_k(\theta) = 0\}$ or $E = \mathcal{V}_E := \mathcal{V}(\langle g_1, \dots, g_k \rangle)$ for $f_i \in \mathbb{Z}[\theta^1, \dots, \theta^d]$ and $g_i \in \mathbb{Z}[\eta_1, \dots, \eta_d]$.
2) $\theta \mapsto \eta(\theta)$ or $\eta \mapsto \theta(\eta)$ is represented by a polynomial ideal, i.e. $\langle h_1, \dots, h_k \rangle \subset \mathbb{Z}[\theta, \eta]$ for $h_i \in \mathbb{Z}[\theta, \eta]$. Here $\mathbb{Z}[\theta, \eta]$ means $\mathbb{Z}[\theta^1, \dots, \theta^d, \eta_1, \dots, \eta_d]$.

e.g.

- Multivariate Gaussian model with a polynomial relation between the covariances: graphical models, AR(p),...
- Algebraic Poisson regression model
- Algebraic multinomial regression model

Algebraic Estimator

Assume non-singularity at the true parameter $\theta^* \in \mathcal{V}_\Theta$.

$(u, v) \in \mathbb{R}^p \times \mathbb{R}^{d-p}$: local coordinate system around θ^*
s.t. $\{\theta(u, 0) | u \in \exists \mathcal{U} \subset \mathbb{R}^p\} = \mathcal{V}_\Theta$ around θ^*.

The full exponential model defines a MLE map
$(X^{(1)}, \ldots, X^{(N)}) \mapsto \theta(\eta)|_{\eta=\bar{x}}$.

A submodel is given by a coordinate projection map $\theta(u, v) \mapsto u$ which defines a (local) estimator $(X^{(1)}, \ldots, X^{(N)}) \mapsto u$.

We call $\theta(u, v)$ an algebraic estimator if $\theta(u, v) \in \mathbb{Q}(u, v)$.
We can define statistical models and estimators by $\eta(u, v) \in \mathbb{Q}(u, v)$ in the same manner.

Note: MLE for an algebraic model is an algebraic estimator.

Differential Geometrical Objects

Let $w := (u, v)$ and we use indexes $\{i, j, ...\}$ for θ and η, $\{a, b, ...\}$ for u, $\{\kappa, \lambda, ...\}$ for v and $\{\alpha, \beta, ...\}$ for w and Einstein summation notation. We assume

3) $w \mapsto \eta(w)$ or $w \mapsto \theta(w)$ is represented by a polynomial ideal.

Differential Geometrical Objects

Let $w := (u, v)$ and we use indexes $\{i, j, \ldots\}$ for θ and η, $\{a, b, \ldots\}$ for u, $\{\kappa, \lambda, \ldots\}$ for v and $\{\alpha, \beta, \ldots\}$ for w and Einstein summation notation. We assume

3) $w \mapsto \eta(w)$ or $w \mapsto \theta(w)$ is represented by a polynomial ideal.

If conditions 1), 2) and 3) hold then the following quantities are all algebraic (i.e. represented by a polynomial ideal).

- $\eta_i(\theta) = \frac{\partial}{\partial \theta^i} \psi(\theta)$,
- Fisher metric $\underline{G} = (g_{ij})$ w.r.t. θ: $g_{ij}(\theta) = \frac{\partial^2 \psi(\theta)}{\partial \theta^i \partial \theta^j}$,
- Fisher metric $\bar{G} = (g^{ij})$ w.r.t. η: $\bar{G} = \underline{G}^{-1}$,
- Jacobian: $B_{i\alpha}(\theta) := \frac{\partial \eta_i(w)}{\partial w^\alpha}$,
- e-connection: $\Gamma^{(e)}_{\alpha\beta,\gamma} = (\frac{\partial^2}{\partial w^\alpha \partial w^\beta} \theta^i(w))(\frac{\partial}{\partial w^\gamma} \eta_i(w))$,
- m-connection: $\Gamma^{(m)}_{\alpha\beta,\gamma} = (\frac{\partial^2}{\partial w^\alpha \partial w^\beta} \eta_i(w))(\frac{\partial}{\partial w^\gamma} \theta^i(w))$,

Furthermore, if $\psi(\theta) \in \mathbb{Q}(\theta) \cup \log \mathbb{Q}(\theta)$ and $\theta(w) \in \mathbb{Q}(w) \cup \log \mathbb{Q}(w)$, then the quantities are all rational.

Asymptotic Statistical Inference Theory

Under some regularity conditions on the carrier measure, function ψ and the manifolds, the following statistical theory holds (See [Amari(1985)] and [Amari and Nagaoka (2000)]):

- $E_u[(\hat{u}^a - u^a)(\hat{u}^b - u^b)] = N^{-1}[g_{ab} - g_{a\kappa}g^{\kappa\lambda}g_{b\lambda}]^{-1} + O(N^{-2})$. Thus, an estimator is 1-st order efficient iff $g_{a\kappa} = 0$.
- The bias term becomes $E_u[\hat{u}^a - u^a] = (2N)^{-1}b^a(u) + O(N^2)$ where $b^a(u) := \Gamma^{(m)a}{}_{cd}(u)g^{cd}(u)$. Then, the bias corrected estimator $\check{u}^a := \hat{u}^a - b^a(\hat{u})$ satisfies $E_u[\check{u}^a - u^a] = O(N^{-2})$.
- Assume $g_{a\kappa} = 0$, then

$$\Gamma^{(m)}{}_{\kappa\lambda,a}(w) = (\frac{\partial^2}{\partial v^\kappa \partial v^\lambda}\eta_i(w))(\frac{\partial}{\partial u^a}\theta^i(w)) = 0 \qquad (1)$$

  implies second order efficiency after a bias correction, i.e. it becomes optimal among the first-order efficient estimators up to $O(N^{-2})$.

Asymptotic Statistical Inference Theory

Under some regularity conditions on the carrier measure, function ψ and the manifolds, the following statistical theory holds (See [Amari(1985)] and [Amari and Nagaoka (2000)]):

- $E_u[(\hat{u}^a - u^a)(\hat{u}^b - u^b)] = N^{-1}[g_{ab} - g_{a\kappa}g^{\kappa\lambda}g_{b\lambda}]^{-1} + O(N^{-2})$. Thus, an estimator is 1-st order efficient iff $g_{a\kappa} = 0$.
- The bias term becomes $E_u[\hat{u}^a - u^a] = (2N)^{-1}b^a(u) + O(N^2)$ where $b^a(u) := \Gamma^{(m)a}{}_{cd}(u)g^{cd}(u)$. Then, the bias corrected estimator $\check{u}^a := \hat{u}^a - b^a(\hat{u})$ satisfies $E_u[\check{u}^a - u^a] = O(N^{-2})$.
- Assume $g_{a\kappa} = 0$, then

$$\Gamma^{(m)}{}_{\kappa\lambda,a}(w) = (\frac{\partial^2}{\partial v^\kappa \partial v^\lambda}\eta_i(w))(\frac{\partial}{\partial u^a}\theta^i(w)) = 0 \qquad (1)$$

implies second order efficiency after a bias correction, i.e. it becomes optimal among the first-order efficient estimators up to $O(N^{-2})$.

$\Longrightarrow$ All algebraic!

Algebraic Second-Order Efficient Estimators (Vector Eq. Form)

Consider an algebraic estimator $\eta(u, v) \in \mathbb{Z}[u, v]$.

Fact 1

If the degree of η w.r.t. v is 1, then (1) gives the MLE.

In general, (1) implies the following vector equation:

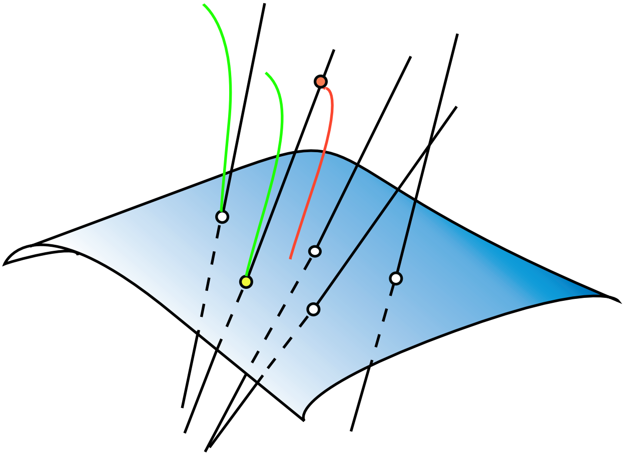
Vector eq. form of the second-order efficient algebraic estimator

$$X = \eta(u, 0) + \sum_{i=p+1}^{d} v_{i-p} e_i(u) + c \cdot \sum_{j=1}^{p} f_j(u, v) e_j(u) \qquad (2)$$

where, for each u,
$\{e_j(u); j = 1, \ldots, p\} \cup \{e_i(u); i = p + 1, \ldots, d\}$ is a complete basis of $\mathbb{R}^d$ s.t. $e_j(u) \in (\nabla_u \eta)^{\perp_{\bar{G}}}$ and $f_j(u, v) \in \mathbb{Z}[u][v]_{\geq 3}$, a polynomial whose degree of v is at least 3, for $j = 1, \ldots, p$.

In (2), a constant $c \in \mathbb{Q}$ is to understand the perturbation.

Image of the Estimators



Algebraic Second-Order Efficient Estimators (Algebraic Eq. Form)

Algebraic eq. form of the second-order efficient algebraic estimator

$$\begin{aligned}(X-\eta(u,0))^\top \tilde{e}_1(u) + h_1(X,u,X-\eta(u,0)) &= 0\\ &\vdots \\ (X-\eta(u,0))^\top \tilde{e}_p(u) + h_p(X,u,X-\eta(u,0)) &= 0\end{aligned} \quad (3)$$

where
$\{\tilde{e}_j(u); j=1,\ldots,p\}$ span $((\nabla_u \eta(u,0))^{\perp_{\bar{G}}})^{\perp_E}$ for every u and $h_j(X,u,t) \in \mathbb{Z}[X,u][t]_3$ for $j=1,\ldots,p$.

Remark 2

$(X-\eta(u,0))^\top \tilde{e}_j(u) = 0$ for $j=1,\ldots,p$ are a set of the estimating equations of MLE.

Algebraic Second-Order Efficient Estimators (Algebraic Eq. Form)

- If a model is defined by an ideal, $I_{\mathcal{M}} := \langle m_1, \ldots, m_{d-p} \rangle$, consisting of all functions vanising on $\mathcal{M}$, then the vector eq. form (3) of the second-order efficient estimators can be represented without u or v:
$(X-\eta)^\top \tilde{e}_j(\eta) + h_j(X, \eta, X-\eta) = 0$ for $j = 1, \ldots, p$ and $I_{\mathcal{M}}$ where $h_j \in \mathbb{Z}[X, \eta][t]_3$.
- As noted below, a unique estimate always exists locally. Denote it $\hat{\hat{\theta}}$,$\hat{\hat{\eta}}$ or $\hat{\hat{u}}$.

Theorem 3

The solution of a vector eq. form (2) of the second-order efficient estimators is given by a set of equations (3).

Proof)

- Since $f_j(u,v) \in \mathbb{Z}[u][v]_{\geq 3}$, there is a $\tilde{f}_j(u,v,\tilde{v}) \in \mathbb{Z}[u,\tilde{v}][v]_3$ with additional p-dim. variables $\tilde{v}$ s.t. $\tilde{f}_j(u,v,v) = f_j(u,v)$.
- Let $\tilde{e}_k \in \{e_i \mid i \in \{1,\dots,d\} \backslash \{k\}\}^{\perp_E}$. This satisfies the condition for $\tilde{e}_j$ for $j = 1,\dots,p$ in the alg. eq. form. Taking the Euclidean inner product of each $\tilde{e}_j$ and the both sides of
  $$X = \eta(u,0) + \sum_{i=p+1}^{d} v_{i-p} e_i(u) + c \cdot \sum_{j=1}^{p} f_j(u,v) e_j(u),$$
  we get $v_i = \tilde{e}_i(u)^\top (X - \eta(u,0))$ and
  $c \cdot f_j(u,v) = \tilde{e}_j(u)^\top (X - \eta(u,0))$.
- Substituting the former equations, the forms of v_is, to the later equations, we get an algebraic eq. form (3).
- Here we used
  $h_j(X,u,t) := \tilde{f}_j(u, (\tilde{e}_i(u)^\top t)_{i=1}^p, (\tilde{e}_i(u)^\top (X - \eta(u,0)))_{i=1}^p)$ for variables $t \in \mathbb{R}^d$ satisfies $h_j(X,u,t) \in \mathbb{Z}[X,u][t]_3$. □

Theorem 4

Every algebraic eq. form (3) gives a second-order efficient estimator.

Proof)

- Represent X in (3) by u and v as $X = \eta(u, v)$, we get $(\eta(u,v) - \eta(u,0))^\top \tilde{e}_j(u) + h_j(\eta(u,v), u, \eta(u,v) - \eta(u,0)) = 0.$
- Partially differentiate this by v twice, $\left(\frac{\partial^2 \eta(u,v)}{\partial v^\lambda \partial v^\kappa}\right)^\top \tilde{e}_j(u)\bigg|_{v=0} = 0$ since each term of $h_j(\eta(u,v), u, \eta(u,v) - \eta(u,0))$ has degree more than 3 of $(\eta_i(u,v) - \eta_i(u,0))_{i=1}^d$ and $\eta(u,v) - \eta(u,0)|_{v=0} = 0$.
- By $\mathrm{span}\{\tilde{e}_j(u); j = 1, \ldots, p\} = ((\nabla_u \eta(u,0))^{\perp_{\bar{G}}})^{\perp_E} = \mathrm{span}\{\bar{G}\partial_{u_a}\eta; a = 1, \ldots, p\}$, we get $\Gamma^{(m)}_{\kappa\lambda a}\Big|_{v=0} = \frac{\partial^2 \eta_i}{\partial v^\lambda \partial v^\kappa} g^{ij} \frac{\partial \eta_j}{\partial u^a}\Big|_{v=0} = 0$
- This means the estimator is second-order efficient. □

Properties of the Estimators (Cont.)

Proposition 5 (Existence and uniqueness of the estimate)

Assume that the Fisher matrix is non-degenerate around $\eta(u^*) \in \mathcal{V}_E$. Then the estimate given by (3) locally uniquely exists for small c, i.e. there is a neighborhood $G(u^*) \subset \mathcal{R}^d$ of $\eta(u^*)$ and $\delta > 0$ such that for every fixed $X \in G(u^*)$ and $-\delta < c < \delta$, a unique estimate exists.

Proof) MLE always exists locally. Furthermore, because of the non-degenerate Fisher matrix, MLE is locally bijective (by the implicit representation theorem). Thus $(u_1, \dots, u_p) \mapsto (g_1(x - \eta(u)), \dots, g_p(x - \eta(u)))$ in (3) is locally bijective. Since $\{g_i\}$ and $\{h_i\}$ are continuous, we can select $\delta > 0$ for (3) to be locally bijective for every $-\delta < c < \delta$.

Flow for the Estimation

Input: $\psi \in \mathbb{Q}(\eta) \cup \log \mathbb{Q}(\eta)$, $m_1, \ldots, m_{d-p} \in \mathbb{Z}[\eta]$ s.t. $\mathcal{V}_E = V(\langle m_1, \ldots, m_{d-p} \rangle)$.

Step1 Compute ψ and $\theta(\eta)$, $G(\eta)$, ($\Gamma^{(m)}(\eta)$ for bias correction)

Step2 Compute $f_{ai} \in \mathbb{Z}[\eta][\xi_{11}, \ldots, \xi_{pd}]_1$ s.t. $f_{aj}(\xi_{11}, \ldots, \xi_{pd}) := \partial_{u^a} m_j$ for $\xi_{bi} := \partial_{u^b} \eta_i$.

Step3 Find $e_{p+1}, \ldots, e_d \in (\nabla_u \eta)^{\perp_{\bar{G}}}$ by eliminating $\{\xi_{aj}\}$ from $\langle e_i(\eta), \partial_{u^a}\eta \rangle_{\bar{G}} = e_{ik}(\eta) g^{kj}(\eta) \xi_{aj} = 0$ and $f_{aj}(\xi_{11}, \ldots, \xi_{pd}) = 0$.

Step4 Select $e_1, \ldots, e_p \in \mathbb{Z}[\eta]$ s.t. $e_1(\eta), \ldots, e_d(\eta)$ are linearly independent over $\mathbb{Q}$.

Step5 Eliminate $v_1, \ldots, v_p$ from
$$X = \eta(u, 0) + \sum_{i=p+1}^{d} v_{i-p} e_i(\eta) + c \cdot \sum_{j=1}^{p} f_j(u, v) e_j(\eta).$$

Onput: $g \in (\mathbb{Z}[\eta][X - \eta]_3)^p$ and $h \in (\mathbb{Z}[\eta][X - \eta]_3)^p$ for
$$g(X - \eta) + c \cdot h(X - \eta) = 0$$

A turning point

From a vector eq. form,

$$X = \eta(u,0) + \sum_{i=p+1}^{d} v_{i-p} e_i(u) + c \cdot \sum_{j=1}^{p} f_j(u,v) e_j(u),$$

we have computed a corresponding algebraic eq. form,

$$\begin{gathered}(X-\eta)^\top \tilde{e}_1(\eta) + h_1(X,\eta,X-\eta) = 0 \\ \vdots \\ (X-\eta)^\top \tilde{e}_p(\eta) + h_p(X,\eta,X-\eta) = 0.\end{gathered}$$

A turning point

From a vector eq. form,

$$X = \eta(u,0) + \sum_{i=p+1}^{d} v_{i-p} e_i(u) + c \cdot \sum_{j=1}^{p} f_j(u,v) e_j(u),$$

we have computed a corresponding algebraic eq. form,

$$(X-\eta)^\top \tilde{e}_1(\eta) + h_1(X,\eta,X-\eta) = 0$$

$$\vdots$$

$$(X-\eta)^\top \tilde{e}_p(\eta) + h_p(X,\eta,X-\eta) = 0.$$

Let's start from an algebraic eq. form!

A turning point

From a vector eq. form,

$$X = \eta(u, 0) + \sum_{i=p+1}^{d} v_{i-p} e_i(u) + c \cdot \sum_{j=1}^{p} f_j(u, v) e_j(u),$$

we have computed a corresponding algebraic eq. form,

$$(X - \eta)^\top \tilde{e}_1(\eta) + h_1(X, \eta, X - \eta) = 0$$

$$\vdots$$

$$(X - \eta)^\top \tilde{e}_p(\eta) + h_p(X, \eta, X - \eta) = 0.$$

Let's start from an algebraic eq. form!

OK. But, how do we select h_is?

A turning point

From a vector eq. form,

$$X = \eta(u,0) + \sum_{i=p+1}^{d} v_{i-p} e_i(u) + c \cdot \sum_{j=1}^{p} f_j(u,v) e_j(u),$$

we have computed a corresponding algebraic eq. form,

$$(X-\eta)^\top \tilde{e}_1(\eta) + h_1(X,\eta,X-\eta) = 0$$
$$\vdots$$
$$(X-\eta)^\top \tilde{e}_p(\eta) + h_p(X,\eta,X-\eta) = 0.$$

Let's start from an algebraic eq. form!

OK. But, how do we select h_is?

We will select h_is in order to decrease the polynomial degree!

Decrease the degree of the estimator

Let $\mathcal{R} := \mathbb{Z}[X, \eta]$ and define

$$\mathcal{I}_3 := \langle \{ (X_i - \eta_i)(X_j - \eta_j)(X_k - \eta_k) \mid 1 \leq i, j, k \leq d \} \rangle$$

as an ideal of $\mathcal{R}$.
Select a monomial order $<$ and set $\eta_1 > \cdots > \eta_d > X_1 > \cdots > X_d$. Let $G_< = \{g_1, \ldots, g_m\}$ be a Gröbner basis of $\mathcal{I}_3$ w.r.t. $<$. Then the residue r_i of h_i by $G_<$ is uniquely determined for each i.

Theorem 6

If the monomial order $<$ is the pure lexicographic,

1. *r_i for $i = 1, \ldots, d$ has degree 2 w.r.t. η, and*
2. *$r_i = 0$ for $i = 1, \ldots, d$ are the estimating equations for a second-order efficient estimator.*

So what?

OK. We can compute second-order efficient estimators with degree 2.

So What?

So what?

OK. We can compute second-order efficient estimators with degree 2.

So What?

So, the homotopy continuation method works!

So what?

OK. We can compute second-order efficient estimators with degree 2.

So What?

So, the homotopy continuation method works!

Great!

...but, what is the homotopy continuation method?

Homotopy continuation method

is an algorithm to solve simultaneous polynomial equations numerically.

Example (2 equations with 2 unknowns)

Input: $f, g \in \mathbb{Z}[x, y]$
Output: The solution of $f(x, y) = g(x, y) = 0$

Homotopy continuation method

is an algorithm to solve simultaneous polynomial equations numerically.

Example (2 equations with 2 unknowns)

Input: $f, g \in \mathbb{Z}[x, y]$
Output: The solution of $f(x, y) = g(x, y) = 0$

Step 1 Select arbitrary polynomials of the form:

$$f_0(x, y) := f_0(x) := a_1 x^{d_1} - b_1 = 0,$$
$$g_0(x, y) := g_0(y) := a_2 y^{d_2} - b_2 = 0$$

where $d_1 = deg(f)$ and $d_2 = deg(g)$. These are easy to solve.

Homotopy continuation method

is an algorithm to solve simultaneous polynomial equations numerically.

Example (2 equations with 2 unknowns)

Input: $f, g \in \mathbb{Z}[x, y]$
Output: The solution of $f(x, y) = g(x, y) = 0$

Step 1 Select arbitrary polynomials of the form:

$$f_0(x, y) := f_0(x) := a_1 x^{d_1} - b_1 = 0,$$
$$g_0(x, y) := g_0(y) := a_2 y^{d_2} - b_2 = 0$$

where $d_1 = deg(f)$ and $d_2 = deg(g)$. These are easy to solve.
Step 2 Take the convex combinations:

$$f_t(x, y) := tf(x, y) + (1 - t)f_0(x, y),$$
$$g_t(x, y) := tg(x, y) + (1 - t)g_0(x, y)$$

then our target becomes the solution for $t = 1$.

Homotopy continuation method (2)

$$f_t(x,y) := tf(x,y) + (1-t)f_0(x,y),$$
$$g_t(x,y) := tg(x,y) + (1-t)g_0(x,y)$$

Step 3 Compute the solution for $t = \delta$ for small δ by the solution for $t = 0$ numerically.

Homotopy continuation method (2)

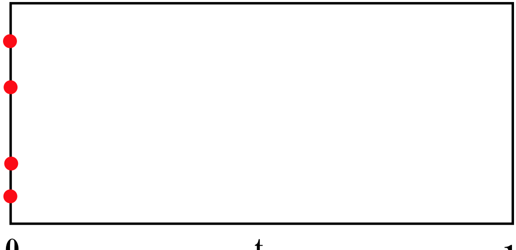
$$f_t(x, y) := tf(x, y) + (1 - t)f_0(x, y),$$
$$g_t(x, y) := tg(x, y) + (1 - t)g_0(x, y)$$

Step 3 Compute the solution for $t = \delta$ for small δ by the solution for $t = 0$ numerically.
Step 4 Repeat this until we get the solution for $t = 1$.

Homotopy continuation method (2)

$$f_t(x, y) := tf(x, y) + (1 - t)f_0(x, y),$$
$$g_t(x, y) := tg(x, y) + (1 - t)g_0(x, y)$$

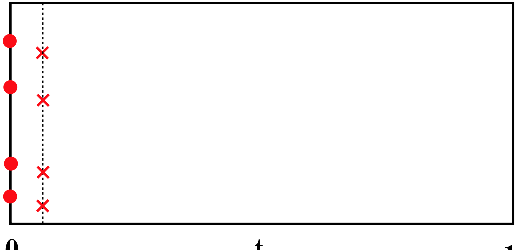
Step 3 Compute the solution for $t = \delta$ for small δ by the solution for $t = 0$ numerically.
Step 4 Repeat this until we get the solution for $t = 1$.
This algorithm is called **the (linear) homotopy continuation method** and justified only if the path connects $t = 0$ and $t = 1$ continuously without an intersection. That can be proved for almost all a and b. [Li(1997)]



Homotopy continuation method (2)

$$f_t(x,y) := tf(x,y) + (1-t)f_0(x,y),$$
$$g_t(x,y) := tg(x,y) + (1-t)g_0(x,y)$$

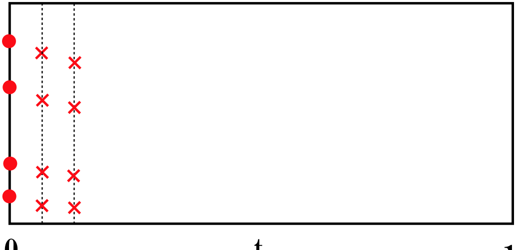
Step 3 Compute the solution for $t=\delta$ for small δ by the solution for $t=0$ numerically.
Step 4 Repeat this until we get the solution for $t=1$.
This algorithm is called **the (linear) homotopy continuation method** and justified only if the path connects $t=0$ and $t=1$ continuously without an intersection. That is proved for almost all a and b. [Li(1997)]



Homotopy continuation method (2)

$$f_t(x, y) := tf(x, y) + (1 - t)f_0(x, y),$$
$$g_t(x, y) := tg(x, y) + (1 - t)g_0(x, y)$$

Step 3 Compute the solution for $t = \delta$ for small δ by the solution for $t = 0$ numerically.
Step 4 Repeat this until we get the solution for $t = 1$.
This algorithm is called **the (linear) homotopy continuation method** and justified only if the path connects $t = 0$ and $t = 1$ continuously without an intersection. That is proved for almost all a and b. [Li(1997)]



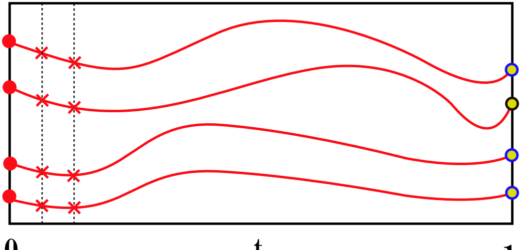
Homotopy continuation method (2)

$$f_t(x,y) := tf(x,y) + (1-t)f_0(x,y),$$
$$g_t(x,y) := tg(x,y) + (1-t)g_0(x,y)$$

Step 3 Compute the solution for $t = \delta$ for small δ by the solution for $t = 0$ numerically.

Step 4 Repeat this until we get the solution for $t = 1$.

This algorithm is called **the (linear) homotopy continuation method** and justified only if the path connects $t = 0$ and $t = 1$ continuously without an intersection. That is proved for almost all a and b. [Li(1997)]



Homotopy continuation method (3)

The number of the paths is the number of the solutions of

$$f_0(x,y) := f_0(x) := a_1 x^{d_1} - b_1 = 0,$$
$$g_0(x,y) := g_0(y) := a_2 y^{d_2} - b_2 = 0.$$

In this case: $d_1 * d_2$.
In general case with m unknowns : $\prod_{i=1}^{m} d_i$.

This causes a serious problem on the computational costs!

In order to solve this problem, Huber and Sturmfels (1995) proposed the nonlinear homotopy continuation methods (or the polyhedral continuation methods). But the degree of the polynomials still affects the computational costs essentially.

Homotopy continuation method (3)

The number of the paths is the number of the solutions of

$$f_0(x, y) := f_0(x) := a_1 x^{d_1} - b_1 = 0,$$
$$g_0(x, y) := g_0(y) := a_2 y^{d_2} - b_2 = 0.$$

In this case: $d_1 * d_2$.
In general case with m unknowns : $\prod_{i=1}^{m} d_i$.

This causes a serious problem on the computational costs!

In order to solve this problem, Huber and Sturmfels (1995) proposed the nonlinear homotopy continuation methods (or the polyhedral continuation methods). But the degree of the polynomials still affects the computational costs essentially.

So, decreasing the degree of 2nd order efficient estimators plays an important role for the homotopy continuation method.

Example: Log Marginal Model

$p_{ij} \in (0,1)$ for $i = 1,2,3$ and $j = 1,2$
Poisson regression: $X_{ij} \sim \mathrm{Po}(Np_{ij})$ i.i.d.
Model constraints:

$$p_{11} + p_{12} + p_{13} = p_{21} + p_{22} + p_{23},$$
$$p_{11} + p_{12} + p_{13} + p_{21} + p_{22} + p_{23} = 1,$$
$$p_{11}p_{13}p_{22}^2 = p_{12}^2 p_{21}p_{23}.$$

- $d = 6, p = 3$
- $\begin{bmatrix} \eta_1 & \eta_2 & \eta_3 \\ \eta_4 & \eta_5 & \eta_6 \end{bmatrix} := N \cdot \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix}$, $\begin{bmatrix} X_1 & X_2 & X_3 \\ X_4 & X_5 & X_6 \end{bmatrix} := \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \end{bmatrix}$
- $\theta^i = \log(\eta_i)$
- $\psi(\theta) = \sum_{i=1}^{6} \exp(\theta^i)$
- $g_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} = \delta_{ij}\eta_i$
- $[u_1, u_2, u_3] := [\eta_1, \eta_3, \eta_5]$

$$e_0 := \begin{bmatrix} \eta_2^2(\eta_4 - \eta_6) \\ -\eta_2^2(\eta_4 - \eta_6) \\ 0 \\ -\eta_3\eta_5^2 - 2\eta_2\eta_4\eta_6 \\ 0 \\ \eta_3\eta_5^2 + 2\eta_2\eta_4\eta_6 \end{bmatrix} \in (\nabla_u \eta)$$

$$[e_1, e_2, e_3] :=$$

$$\left[\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \eta_1(-\eta_1\eta_5^2 + \eta_3\eta_5^2) \\ \eta_2(-\eta_1\eta_5^2 - 2\eta_2\eta_4\eta_6) \\ 0 \\ \eta_4(\eta_2^2\eta_4 - \eta_2^2\eta_6) \\ \eta_5(\eta_2^2\eta_4 + 2\eta_1\eta_3\eta_5) \\ 0 \end{bmatrix}, \begin{bmatrix} \eta_1(\eta_1\eta_5^2 - \eta_3\eta_5^2) \\ \eta_2(\eta_1\eta_5^2 + 2\eta_2\eta_4\eta_6) \\ 0 \\ \eta_4(2\eta_1\eta_3\eta_5 + \eta_2^2\eta_6) \\ 0 \\ \eta_6(\eta_2^2\eta_4 + 2\eta_1\eta_3\eta_5) \end{bmatrix} \right]$$

$$e_1, e_2, e_3 \in (\nabla_u \eta)^{\perp_{\bar{G}}}$$

- An estimating equation for the second order efficient is

$$X - \eta + v_1 \cdot e_1 + v_2 \cdot e_2 + v_3 \cdot e_3 + c \cdot v_1^3 \cdot e_0 = 0$$

- MLE is a root of

$$\{ x_1\eta_2^2\eta_4^2\eta_6 - x_1\eta_2^2\eta_4\eta_6^2 - x_2\eta_1\eta_2\eta_4^2\eta_6 + x_2\eta_1\eta_2\eta_4\eta_6^2 - 2x_4\eta_1\eta_2\eta_4\eta_6^2 - x_4\eta_1\eta_3\eta_5^2\eta_6 + 2x_6\eta_1\eta_2\eta_4^2\eta_6 + x_6\eta_1\eta_3\eta_4\eta_5^2,$$

$$-x_2\eta_2\eta_3\eta_4^2\eta_6 + x_2\eta_2\eta_3\eta_4\eta_6^2 + x_3\eta_2^2\eta_4^2\eta_6 - x_3\eta_2^2\eta_4\eta_6^2 - x_4\eta_1\eta_3\eta_5^2\eta_6 - 2x_4\eta_2\eta_3\eta_4\eta_6^2 + x_6\eta_1\eta_3\eta_4\eta_5^2 + 2x_6\eta_2\eta_3\eta_4^2\eta_6,$$

$$-2x_4\eta_1\eta_3\eta_5^2\eta_6 - x_4\eta_2^2\eta_4\eta_5\eta_6 + x_5\eta_2^2\eta_4^2\eta_6 - x_5\eta_2^2\eta_4\eta_6^2 + 2x_6\eta_1\eta_3\eta_4\eta_5^2 + x_6\eta_2^2\eta_4\eta_5\eta_6,$$

$$\eta_1\eta_3\eta_5^2 - \eta_2^2\eta_4\eta_6, \eta_1 + \eta_2 + \eta_3 - \eta_4 - \eta_5 - \eta_6,$$

$$-\eta_1 - \eta_2 - \eta_3 - \eta_4 - \eta_5 - \eta_6 + 1\}$$

degree = 5*5*5*4*1*1= 500

- A 2nd-order-efficient estimator with degree 2:

$\{-3\,x_1x_2x_4^2x_6\eta_2 + 6\,x_1x_2x_4^2x_6\eta_6 + x_1x_2x_4^2\eta_2\eta_6 - 2\,x_1x_2x_4^2\eta_6^2 + 3\,x_1x_2x_4x_6^2\eta_2 - 6\,x_1x_2x_4x_6^2\eta_4 + 2\,x_1x_2x_4x_6\eta_2\eta_4 - 2\,x_1x_2x_4x_6\eta_2\eta_6 - x_1x_2x_6^2\eta_2\eta_4 + 2\,x_1x_2x_6^2\eta_4^2 + 3\,x_1x_3x_4x_5^2\eta_6 - 2\,x_1x_3x_4x_5\eta_5\eta_6 - 3\,x_1x_3x_5^2x_6\eta_4 + 2\,x_1x_3x_5x_6\eta_4\eta_5 + x_1x_4^2x_6\eta_2^2 - 2\,x_1x_4^2x_6p_2\eta_6 - x_1x_4x_5^2\eta_3\eta_6 - x_1x_4x_6^2\eta_2^2 + 2\,x_1x_4x_6^2\eta_2\eta_4 + x_1x_5^2x_6\eta_3\eta_4 + 3\,x_2^2x_4^2x_6\eta_1 - x_2^2x_4^2\eta_1\eta_6 - 3\,x_2^2x_4x_6^2\eta_1 - 2\,x_2^2x_4x_6\eta_1\eta_4 + 2\,x_2^2x_4x_6\eta_1\eta_6 + x_2^2x_6^2\eta_1\eta_4 - x_2x_4^2x_6\eta_1\eta_2 - 2\,x_2x_4^2x_6\eta_1\eta_6 + x_2x_4x_6^2\eta_1\eta_2 + 2\,x_2x_4x_6^2\eta_1\eta_4 - x_3x_4x_5^2\eta_1\eta_6 + x_3x_5^2x_6\eta_1\eta_4$,

$3\,x_1x_3x_4x_5^2\eta_6 - 2\,x_1x_3x_4x_5\eta_5\eta_6 - 3\,x_1x_3x_5^2x_6\eta_4 + 2\,x_1x_3x_5x_6\eta_4\eta_5 - x_1x_4x_5^2\eta_3\eta_6 + x_1x_5^2x_6\eta_3\eta_4 + 3\,x_2^2x_4^2x_6\eta_3 - x_2^2x_4^2\eta_3\eta_6 - 3\,x_2^2x_4x_6^2\eta_3 - 2\,x_2^2x_4x_6\eta_3\eta_4 + 2\,x_2^2x_4x_6\eta_3\eta_6 + x_2^2x_6^2\eta_3\eta_4 - 3\,x_2x_3x_4^2x_6\eta_2 + 6\,x_2x_3x_4^2x_6\eta_6 + x_2x_3x_4^2\eta_2\eta_6 - 2\,x_2x_3x_4^2\eta_6^2 + 3\,x_2x_3x_4x_6^2\eta_2 - 6\,x_2x_3x_4x_6^2\eta_4 + 2\,x_2x_3x_4x_6\eta_2\eta_4 - 2\,x_2x_3x_4x_6\eta_2\eta_6 - x_2x_3x_6^2\eta_2\eta_4 + 2\,x_2x_3x_6^2\eta_4^2 - x_2x_4^2x_6\eta_2\eta_3 - 2\,x_2x_4^2x_6\eta_3\eta_6 + x_2x_4x_6^2\eta_2\eta_3 + 2\,x_2x_4x_6^2\eta_3\eta_4 + x_3x_4^2x_6\eta_2^2 - 2\,x_3x_4^2x_6\eta_2\eta_6 - x_3x_4x_5^2\eta_1\eta_6 - x_3x_4x_6^2\eta_2^2 + 2\,x_3x_4x_6^2\eta_2\eta_4 + x_3x_5^2x_6\eta_1\eta_4,$

$6\,x_1x_3x_4x_5^2\eta_6 - 4\,x_1x_3x_4x_5\eta_5\eta_6 - 6\,x_1x_3x_5^2x_6\eta_4 + 4\,x_1x_3x_5x_6\eta_4\eta_5 - 2\,x_1x_4x_5^2\eta_3\eta_6 + 2\,x_1x_5^2x_6\eta_3\eta_4 + 3\,x_2^2x_4^2x_6\eta_5 - x_2^2x_4^2\eta_5\eta_6 - 3\,x_2^2x_4x_5x_6\eta_4 + 3\,x_2^2x_4x_5x_6\eta_6 + x_2^2x_4x_5\eta_4\eta_6 - x_2^2x_4x_5\eta_6^2 - 3\,x_2^2x_4x_6^2\eta_5 - x_2^2x_4x_6\eta_4\eta_5 + x_2^2x_4x_6\eta_5\eta_6 + x_2^2x_5x_6\eta_4^2 - x_2^2x_5x_6\eta_4\eta_6 + x_2^2x_6^2\eta_4\eta_5 - 2\,x_2x_4^2x_6\eta_2\eta_5 + 2\,x_2x_4x_5x_6\eta_2\eta_4 - 2\,x_2x_4x_5x_6\eta_2\eta_6 + 2\,x_2x_4x_6^2\eta_2\eta_5 - 2\,x_3x_4x_5^2\eta_1\eta_6 + 2\,x_3x_5^2x_6\eta_1\eta_4,$

$\eta_1\eta_3\eta_5^2 - \eta_2^2\eta_4\eta_6,\ \eta_1 + \eta_2 + \eta_3 - \eta_4 - \eta_5 - \eta_6$,

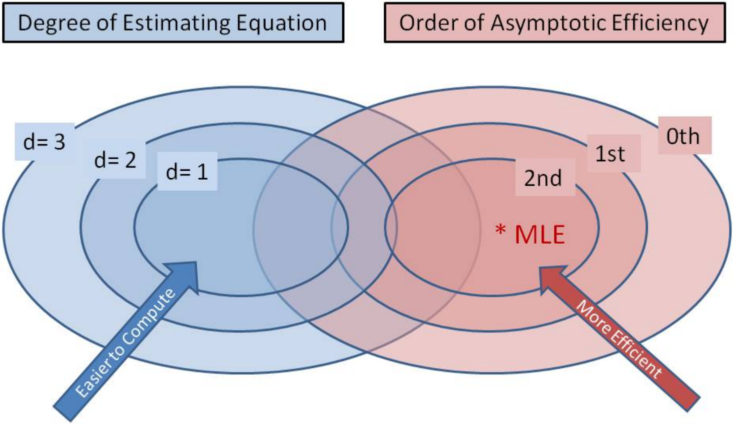
$-\eta_1 - \eta_2 - \eta_3 - \eta_4 - \eta_5 - \eta_6 + 1\}$ degree = 2*2*2*4*1*1= 32

Computational Results by the Homotopy Continuation Methods

- Software for the homotopy methods: HOM4PS2 by Lee, Li and Tsuai.
- $X = (1, 1, 1, 1, 1, 1)$.
- Repeat count: 10.

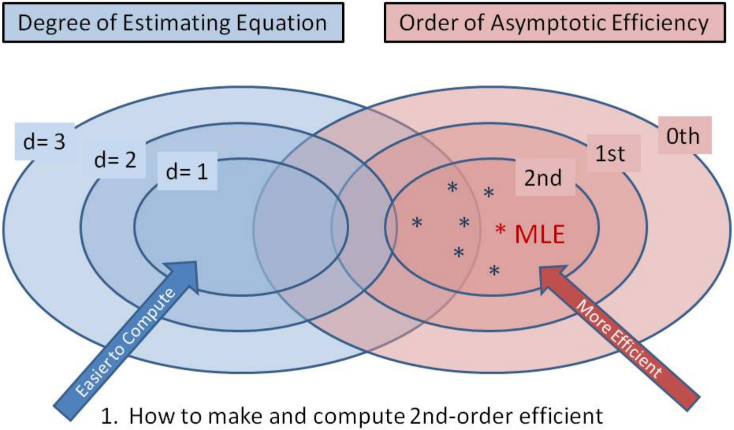
algorithm	estimator	#paths	running time [s] (avg. ± std.)
linear homotopy	MLE	500	1.137 ± 0.073
	2nd eff.	32	0.150 ± 0.047
polyhedral homotopy	MLE	64	0.267 ± 0.035
	2nd eff	24	0.119 ± 0.027

Summary



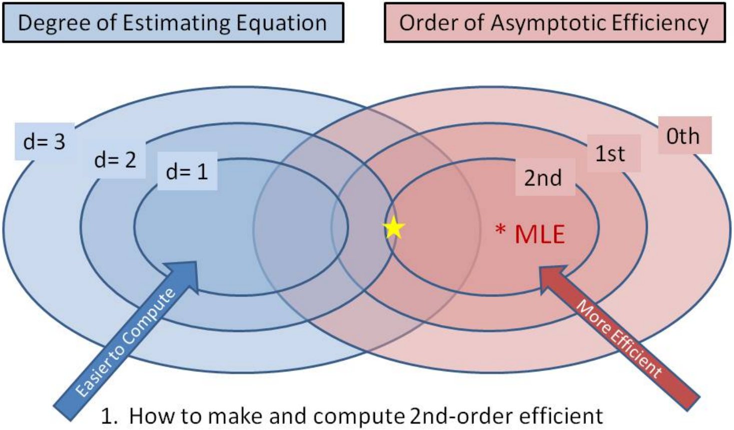
Degree of Estimating Equation
Order of Asymptotic Efficiency
d= 3
d= 2
d= 1
0th
1st
2nd
* MLE
Easier to Compute
More Efficient

Summary



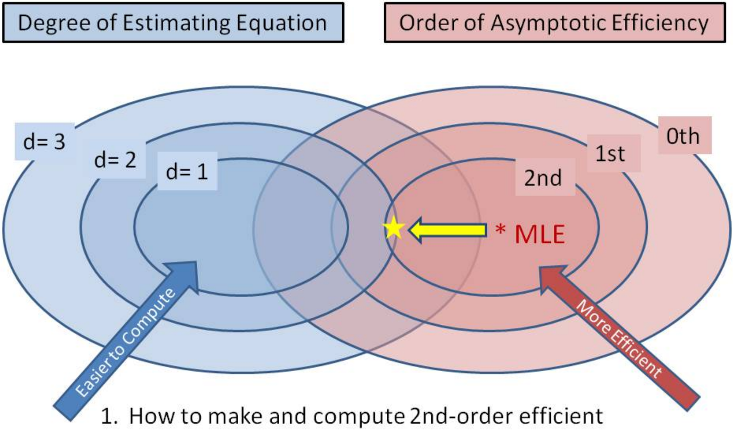
1. How to make and compute 2nd-order efficient estimators algebraically.

Summary



1. How to make and compute 2nd-order efficient estimators algebraically.
2. Existence of a 2nd order efficient estimator with degree 2.

Summary



1. How to make and compute 2nd-order efficient estimators algebraically.
2. Existence of a 2nd order efficient estimator with degree 2.
3. How to compute it from the likelihood equations.

Advantage and Disadvantage of Algebraic Approach

Plus

- Availability of algebraic algorithms
- Exactness of the solutions
- "Differentiability" of the results
- Classifiability of models and estimators.

Minus (Future Works)

- Redundancy of the solutions
- Reality of the varieties
- Singularity of the models
- Globality of the theory

Future works

More Future Works

- Asymptotics based on divergence (Bayesian prediction, model selection etc.).

Explicit form of the estimation $\hat{\hat{u}}$

Explicit form by radicals does not exist in general. However, we can use algebraic approximations e.g.

- Taylor approximation
- Newton-Raphson Methods
- continued fraction
- Laguerre's methods (may contain square roots)

Bias Correction

Fact 7 (Bias term)

The bias correction term $b(\hat{\hat{u}})$ of $\hat{\hat{u}}$ has the same form $b(\hat{u})$ of the MLE $\hat{u}$.

Remark 8

We can select h_i such as the estimating equation becomes unbiased,
i.e. $E_{\eta^*}[g_i(X-\eta) + c \cdot h_i(X, \eta, X-\eta)] = 0$.
The bias of the estimator may be decreased by this.

Example: Periodic Gaussian Model

$$X \sim N(\mu, \Sigma(a)) \text{ with } \mu = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } \Sigma(a) = \begin{bmatrix} 1 & a & a^2 & a \\ a & 1 & a & a^2 \\ a^2 & a & 1 & a \\ a & a^2 & a & 1 \end{bmatrix}$$

and $0 \leq a < 1$.

- $d = 3, p = 1$
- $\log p(x|\theta) =$ $2\,(x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1)\,\theta_2 + 2\,(x_3x_1 + x_4x_2)\,\theta_3 - \psi(\theta)$,
- $\psi(\theta) = -1/2\, \log({\theta_1}^4 - 4\,{\theta_1}^2{\theta_2}^2 + 8\,\theta_1{\theta_2}^2\theta_3 - 2\,{\theta_1}^2{\theta_3}^2$ $-4\,{\theta_2}^2{\theta_3}^2 + {\theta_3}^4) + 2\,\log(2\,\pi)$,
- $\theta(a) = [\frac{1}{1-2a^2+4a^4}, -\frac{a}{1-2a^2+4a^4}, \frac{a^2}{1-2a^2+4a^4}]^\top$,
- $\eta(a) = [-2, -4a, -2a^2]^\top$,
- $(g^{ij}) = \begin{bmatrix} 2\,a^4 + 4\,a^2 + 2 & 8\,a\,(1 + a^2) & 8\,a^2 \\ 8\,a\,(1 + a^2) & 4 + 24\,a^2 + 4\,a^4 & 8\,a\,(1 + a^2) \\ 8\,a^2 & 8\,a\,(1 + a^2) & 2\,a^4 + 4\,a^2 + 2 \end{bmatrix}$

Example: Periodic Gaussian Model (Cont.)

$e_0(a) := [0, -1, a]^\top \in \partial_a \eta(a)$.
$e_1(a) := [3a^2 + 1, 4a, 0]$, $e_2(a) := [-a^2 - 1, 0, 2] \in (\partial_a \eta(a))^{\perp_{\bar{G}}}$.

- An estimating equation for the second order efficient is

$$x - \eta + v_1 \cdot e_1 + v_2 \cdot e_2 + c \cdot v_1^3 \cdot e_0 = 0$$

- By eliminating v_1 and v_2, we get

$$g(a) + c \cdot h(a) = 0$$

  where $g(a) := 8\,(a-1)^2\,(a+1)^2\,(1+2\,a^2)^2 \cdot$
  $(4\,a^5 - 8\,a^3 + 2\,a^3 x_3 - 3\,x_2\,a^2 + 4\,a + 4\,a x_1 + 2\,a x_3 - x_2)$
  and
  $h(a) := (2\,a^4 + a^3 x_2 - a^2 x_3 + 2\,a^2 + a x_2 - 2\,x_1 - x_3 - 4)^3$.
- MLE is a root of
  $4\,a^5 - 8\,a^3 + 2\,a^3 x_3 - 3\,x_2\,a^2 + 4\,a + 4\,a x_1 + 2\,a x_3 - x_2$.
- (Bias correction term of $\hat{\hat{a}}$) $= \frac{\hat{\hat{a}}\left(\hat{\hat{a}}^8 - 4\,\hat{\hat{a}}^6 + 6\,\hat{\hat{a}}^4 - 4\,\hat{\hat{a}}^2 + 1\right)}{\left(1 + 2\,\hat{\hat{a}}^2\right)^2}$.