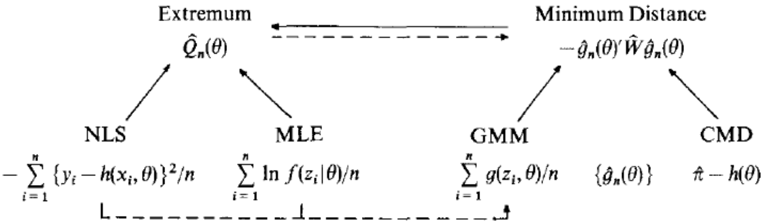
Estimation Principles

Raffaele Saggio

UBC

November 3, 2021

A Map of the World



How to estimate stuff?

- Common framework to analyzing many problems in econometrics summarized in Newey and McFadden (1994)
- Define quantity of interest θ_0 as maximizer of population criterion fn $Q_0(\theta)$
- Define estimator as maximizer of sample criterion fn $\hat{Q}_N(\theta)$
  - "Extremum" estimator (Amemiya, 1973)
  - θ of fixed dimension!
    - Need different tools when dim (θ) grows with N

How to estimate stuff?

- Common framework to analyzing many problems in econometrics summarized in Newey and McFadden (1994)
- Define quantity of interest θ_0 as maximizer of population criterion fn $Q_0(\theta)$
- Define estimator as maximizer of sample criterion fn $\hat{Q}_N(\theta)$
  - "Extremum" estimator (Amemiya, 1973)
  - θ of fixed dimension!
    - Need different tools when dim (θ) grows with N

- Basic results that hold subject to "usual" regularity conditions:
  - Consistency: $\hat{\theta} \xrightarrow{p} \theta_0$
  - Normality: $\sqrt{N}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}^{-1}\boldsymbol{V}\boldsymbol{H}^{-1}\right)$

Preliminaries

The data are an *iid* sample $\{\boldsymbol{Z}_i\}_{i=1}^{N}$ from some d.f. $F_Z(.)$

Parameter of interest is:

$$\theta_0 = \arg\max_{\theta \in \Theta} Q_0(\theta)$$

where $Q_0(\theta)$ is some population criterion function.

Assume θ_0 is a singleton (point-identification)

Guess your estimator!

Example I?:

$$Q_0(\theta) = E\left[\left(Y_i - X_i'\theta\right)^2\right]$$

Guess your estimator!

Example I?:

$$Q_0(\theta) = E\left[\left(Y_i - X_i'\theta\right)^2\right]$$

Example II?:

$$Q_0(\theta) = E\left[Y_i \ln \Phi\left(X_i'\theta\right) + (1 - Y_i) \ln\left(1 - \Phi\left(X_i'\theta\right)\right)\right]$$

Guess your estimator!

Example I?:

$$Q_0(\theta) = E\left[\left(Y_i - X_i'\theta\right)^2\right]$$

Example II?:

$$Q_0(\theta) = E\left[Y_i \ln \Phi\left(X_i'\theta\right) + (1 - Y_i)\ln\left(1 - \Phi\left(X_i'\theta\right)\right)\right]$$

Example III?

$$Q_0(\theta) = E\left[\left(Y_i - X_i'\theta\right) Z_i\right]' E\left[Z_i Z_i'\right]^{-1} E\left[\left(Y_i - X_i'\theta\right) Z_i\right]$$

Estimator

Estimate θ by max'ing sample criterion fn:

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \widehat{Q}_N(\theta)$$

Estimator

Estimate θ by max'ing sample criterion fn:

$$\widehat{\theta} = \underset{\theta \in \Theta}{\arg\max}\ \widehat{Q}_N(\theta)$$

Example I:

$$\hat{Q}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - X_i'\theta\right)^2$$

Example II:

$$\hat{Q}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} Y_i \ln\Phi\left(X_i'\theta\right) + (1 - Y_i)\ln\left(1 - \Phi\left(X_i'\theta\right)\right)$$

Example III:

$$\hat{Q}_N(\theta) = \left[\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - X_i'\theta\right) Z_i\right]' \left[\frac{1}{N}\sum_{i=1}^{N} Z_i Z_i'\right]^{-1} \left[\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - X_i'\theta\right) Z_i\right]$$

Consistency

When does $\hat{\theta} \xrightarrow{p} \theta_0$?

Intuition:

- Need for θ_0 to be "well separated"
- Need $\hat{Q}_N(\theta)$ "close" to $Q_0(\theta)$ for maximizer to be close

Standard <u>sufficient</u> conditions:

- $Q_0(.)$ continuous
- Θ compact
- $\sup_{\theta \in \Theta} \left| \widehat{Q}_N(\theta) - Q_0(\theta) \right| \xrightarrow{p} 0$ (uniform convergence)

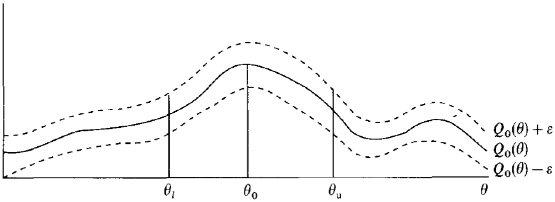
Uniform Convergence

Pointwise convergence usually easy to establish via LLN:

$$\left|\widehat{Q}_N(\theta) - Q_0(\theta)\right| \xrightarrow{p} 0 \ \forall \theta \in \Theta$$

Uniform convergence is stronger, need to trap "worst-case" deviation of fn from limit:

$$\sup_{\theta \in \Theta} \left|\widehat{Q}_N(\theta) - Q_0(\theta)\right| \xrightarrow{p} 0$$



On finite grid, pointwise → uniform convergence. What can go wrong w/ continuous Θ?

Folding tent example

Consistency

Theorem

If the following conditions hold:

i) Θ *is compact*

ii) $Q_0(\theta)$ *is uniquely maximized at* θ_0

iii) $Q_0(\theta)$ *is continuous*

iv) $\widehat{Q}_N(\theta)$ *converges uniformly to* $Q_0(\theta)$

Then,

$$\widehat{\theta} \xrightarrow{p} \theta_0$$

Proof

Asymptotic Distribution

Theorem

If $\hat{\theta} \xrightarrow{p} \theta_0$ *and the following conditions hold:*

i) θ_0 *is in the interior of* Θ

ii) $\widehat{Q}_N(\theta)$ *is twice differentiable in a neighborhood* $\mathcal{N}$ *of* θ_0

iii) $\sqrt{N}\nabla_\theta \widehat{Q}_N(\theta_0) \xrightarrow{d} N(0, \boldsymbol{V})$

iv) there is an $\boldsymbol{H}(\theta)$ *that is continuous at* θ_0 *such that* $\sup_\theta \left\| \nabla_{\theta\theta} \widehat{Q}_N(\theta) - \boldsymbol{H}(\theta) \right\| \xrightarrow{p} 0$

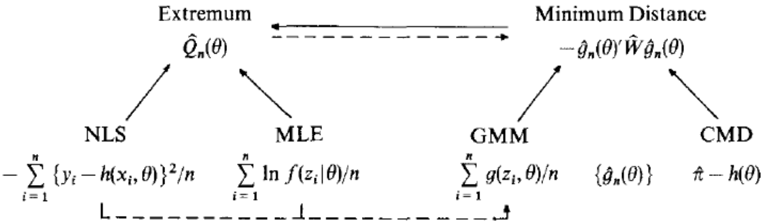
v) $\boldsymbol{H} \equiv \boldsymbol{H}(\theta_0)$ *is nonsingular*

Then,

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}^{-1}\boldsymbol{V}\boldsymbol{H}^{-1}\right)$$

Sketch of Proof

Special Cases



Classical Minimum Distance (CMD)

You have an economic model that says that certain observed moments in the data obeys a certain structure $\boldsymbol{g}(\theta_0)$.

$$\boldsymbol{\pi} - \boldsymbol{g}(\theta_0) = 0 \tag{1}$$

Classical Minimum Distance (CMD)

You have an economic model that says that certain observed moments in the data obeys a certain structure $\boldsymbol{g}(\theta_0)$.

$$\boldsymbol{\pi} - \boldsymbol{g}(\theta_0) = 0 \tag{1}$$

Objective fn (minimize instead of max):

$$\widehat{Q}_N(\theta) = [\widehat{\boldsymbol{\pi}} - \boldsymbol{g}(\theta)]' \widehat{\boldsymbol{W}} [\widehat{\boldsymbol{\pi}} - \boldsymbol{g}(\theta)]$$

- $\widehat{\boldsymbol{\pi}} \xrightarrow{p} \boldsymbol{\pi}$ is a vector of reduced form sample moments
- $\boldsymbol{g}(\theta)$ is a structural function
- $\widehat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W}$ is a symmetric weighting matrix

Classical Minimum Distance (CMD)

You have an economic model that says that certain observed moments in the data obeys a certain structure $\boldsymbol{g}(\theta_0)$.

$$\boldsymbol{\pi} - \boldsymbol{g}(\theta_0) = 0 \tag{1}$$

Objective fn (minimize instead of max):

$$\widehat{Q}_N(\theta) = [\widehat{\boldsymbol{\pi}} - \boldsymbol{g}(\theta)]' \, \widehat{\boldsymbol{W}} \, [\widehat{\boldsymbol{\pi}} - \boldsymbol{g}(\theta)]$$

- $\widehat{\boldsymbol{\pi}} \xrightarrow{p} \boldsymbol{\pi}$ is a vector of reduced form sample moments
- $\boldsymbol{g}(\theta)$ is a structural function
- $\widehat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W}$ is a symmetric weighting matrix

Classical Minimum Distance (CMD)

You have an economic model that says that certain observed moments in the data obeys a certain structure $\boldsymbol{g}(\theta_0)$.

$$\boldsymbol{\pi} - \boldsymbol{g}(\theta_0) = 0 \tag{1}$$

Objective fn (minimize instead of max):

$$\widehat{Q}_N(\theta) = [\widehat{\boldsymbol{\pi}} - \boldsymbol{g}(\theta)]' \widehat{\boldsymbol{W}} [\widehat{\boldsymbol{\pi}} - \boldsymbol{g}(\theta)]$$

- $\widehat{\boldsymbol{\pi}} \xrightarrow{p} \boldsymbol{\pi}$ is a vector of reduced form sample moments
- $\boldsymbol{g}(\theta)$ is a structural function
- $\widehat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W}$ is a symmetric weighting matrix

Optimal weighting matrix: $\boldsymbol{W} = Var(\hat{\boldsymbol{\pi}})^{-1}$ (OMD)

Derivation of asymptotic variance

FOC:

$$\nabla_\theta \boldsymbol{g}\left(\widehat{\theta}\right)' \widehat{\boldsymbol{W}}\left[\widehat{\boldsymbol{\pi}} - \boldsymbol{g}\left(\widehat{\theta}\right)\right] = 0$$

Derivation of asymptotic variance

FOC:

$$\nabla_\theta \boldsymbol{g}\left(\widehat{\theta}\right)' \widehat{\boldsymbol{W}}\left[\widehat{\pi} - \boldsymbol{g}\left(\widehat{\theta}\right)\right] = 0$$

Mean value expansion:

$$\boldsymbol{g}\left(\widehat{\theta}\right) = \boldsymbol{g}\left(\theta_0\right) + \boldsymbol{G}\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_0\right)$$

where $\boldsymbol{G}\left(\theta\right) = \nabla_\theta \boldsymbol{g}\left(\theta\right)$. Substitute in to get:

$$\boldsymbol{G}\left(\widehat{\theta}\right)' \widehat{\boldsymbol{W}}\left[\widehat{\pi} - \boldsymbol{g}\left(\theta_0\right) - \boldsymbol{G}\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_0\right)\right] = 0$$

Derivation of asymptotic variance

FOC:

$$\nabla_\theta \boldsymbol{g}\left(\widehat{\theta}\right)' \widehat{\boldsymbol{W}}\left[\widehat{\pi} - \boldsymbol{g}\left(\widehat{\theta}\right)\right] = 0$$

Mean value expansion:

$$\boldsymbol{g}\left(\widehat{\theta}\right) = \boldsymbol{g}\left(\theta_0\right) + \boldsymbol{G}\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_0\right)$$

where $\boldsymbol{G}(\theta) = \nabla_\theta \boldsymbol{g}(\theta)$. Substitute in to get:

$$\boldsymbol{G}\left(\widehat{\theta}\right)' \widehat{\boldsymbol{W}}\left[\widehat{\pi} - \boldsymbol{g}\left(\theta_0\right) - \boldsymbol{G}\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_0\right)\right] = 0$$

Rearranging, normalizing, and taking limits yields:

$$\boldsymbol{G}\left(\theta_0\right)' \boldsymbol{W}\sqrt{N}\left[\widehat{\pi} - \boldsymbol{g}\left(\theta_0\right)\right] = \boldsymbol{G}\left(\theta_0\right)' \boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)\sqrt{N}\left(\widehat{\theta} - \theta_0\right) + o_p(1)$$

Standard Errors

Solve for $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ to get:

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)=\left[\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)\right]^{-1}\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\sqrt{N}\left[\widehat{\boldsymbol{\pi}}-\boldsymbol{g}\left(\theta_0\right)\right]+o_p\left(1\right)$$

By assumption:

$$\sqrt{N}\left[\widehat{\boldsymbol{\pi}}-\boldsymbol{g}\left(\theta_0\right)\right]\xrightarrow{d}N\left(0,\boldsymbol{V}_{\pi}\right)$$

Standard Errors

Solve for $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ to get:

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)=\left[\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)\right]^{-1}\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\sqrt{N}\left[\widehat{\boldsymbol{\pi}}-\boldsymbol{g}\left(\theta_0\right)\right]+o_p\left(1\right)$$

By assumption:

$$\sqrt{N}\left[\widehat{\boldsymbol{\pi}}-\boldsymbol{g}\left(\theta_0\right)\right]\xrightarrow{d}N\left(0,\boldsymbol{V}_\pi\right)$$

Slutzky: $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ is (asymptotically) a linear combination of normals w/ variance taking usual sandwich form:

$$\left[\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{H}}\right]^{-1}\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{V}_\pi\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{V}}\left[\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{H}}\right]^{-1}$$

Standard Errors

Solve for $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ to get:

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)=\left[\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)\right]^{-1}\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\sqrt{N}\left[\widehat{\boldsymbol{\pi}}-\boldsymbol{g}\left(\theta_0\right)\right]+o_p\left(1\right)$$

By assumption:

$$\sqrt{N}\left[\widehat{\boldsymbol{\pi}}-\boldsymbol{g}\left(\theta_0\right)\right]\xrightarrow{d}N\left(0,\boldsymbol{V}_{\pi}\right)$$

Slutzky: $\sqrt{N}\left(\widehat{\theta}-\theta_0\right)$ is (asymptotically) a linear combination of normals w/ variance taking usual sandwich form:

$$\Bigg[\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{H}}\Bigg]^{-1}\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{V}_{\pi}\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{V}}\Bigg[\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{H}}\Bigg]^{-1}$$

Standard errors: replace unknown quantities with sample analogues (i.e. $\boldsymbol{G}\left(\widehat{\theta}\right)$ for $\boldsymbol{G}\left(\theta_0\right)$, $\hat{\boldsymbol{V}}_{\pi}$ for $\boldsymbol{V}_{\pi}$)

Optimal Weighting

With optimal weights $\boldsymbol{W} = \boldsymbol{V}_{\pi}{}^{-1}$, asymptotic variance reduces to:

$$\left[\boldsymbol{G}\left(\theta_0\right)' \boldsymbol{V}_{\pi}{}^{-1} \boldsymbol{G}\left(\theta_0\right)\right]^{-1}$$

Optimal Weighting

With optimal weights $\boldsymbol{W} = \boldsymbol{V_\pi}^{-1}$, asymptotic variance reduces to:

$$\left[\boldsymbol{G}\left(\theta_0\right)' \boldsymbol{V_\pi}^{-1} \boldsymbol{G}\left(\theta_0\right)\right]^{-1}$$

Specification testing: optimal weighting yields null distribution for minimized value of criterion function

$$\hat{Q}_N^{OMD}\left(\hat{\theta}\right) \sim \chi^2\left(J - K\right)$$

Answers question: does my model explain the data up to sampling error? (i.e. could <u>population</u> $R^2 = 1$?)

Practical Problems w/ OMD

Optimal weighting behaves poorly when $\boldsymbol{W}$ is large relative to sample size (Altonji and Segal, 1996)

- Variant of "Kakwani bias" in FGLS
- Problem emerges from correlation between weights and moments

Practical Problems w/ OMD

Optimal weighting behaves poorly when $\boldsymbol{W}$ is large relative to sample size (Altonji and Segal, 1996)

- Variant of "Kakwani bias" in FGLS
- Problem emerges from correlation between weights and moments

Potential solutions:

- Use inefficient (but tractable) $\boldsymbol{W}$: equal (or diagonal) weight matrix (e.g., Abowd and Card, 1992)
- Jackknife / split sample estimation of $\boldsymbol{W}$ (Kezdi, Solon, and Hahn, 2002)
- Parameterize $\boldsymbol{W} = \boldsymbol{W}(\delta)$ and estimate δ along with θ via "CUGMM" (Hansen, Heaton, and Yaron, 1996; Hausman et al, 2011)
- Generalized Empirical Likelihood (Newey and Smith, 2004)

Spec test without optimal weighting

- Is it possible to test our model if we don't have (want) to use optimal weighting matrix?

- Newey (1985): yes! Consider estimation of θ that is based upon

$$[\hat{\pi} - g(\theta)]' \mathbf{A} [\hat{\pi} - g(\theta)] \tag{2}$$

  where $\mathbf{A}$ is a possibly stochastic matrix.

- Newey showed that

$$N[\hat{\pi} - g(\theta)]' \mathbf{R}^{-} [\hat{\pi} - g(\theta)] \sim \chi^2_{J-K} \tag{3}$$

  where $\mathbf{R}^{-}$ is the generalized inverse of the matrix $\mathbf{R} = \mathbf{M}\mathbf{V}_{\pi}\mathbf{M}'$ with

$$\mathbf{M} = \mathbf{I} - \mathbf{G}(\hat{\theta})(\mathbf{G}(\hat{\theta})'\mathbf{A}\mathbf{G}(\hat{\theta}))^{-1}\mathbf{G}(\hat{\theta})'\mathbf{A} \tag{4}$$

- Need generalized inverse because $\mathbf{R}$ does not have full rank (presence of $\mathbf{M}$).

On the frontier

- Important assumption is "correct" specification.

$$\sqrt{N}(\hat{\pi} - g(\theta_0)) \sim^a \mathcal{N}(0, \mathbf{V}_\pi) \tag{5}$$

On the frontier

- Important assumption is "correct" specification.

$$\sqrt{N}(\hat{\pi} - g(\theta_0)) \sim^a \mathcal{N}(0, \mathbf{V}_\pi) \tag{5}$$

- What happens under misspecification?
- Chamberlain (1994): be careful with weights! Randomness of weights increases noise in resulting estimator.
- Active frontier in econometrics today: what if my model is wrong? Can my standard errors account for that?
  - Armstrong and Kolesar (2018); Bonhomme and Weidner (2019).

GMM

Criterion function (min once again):

$$\widehat{Q}_{GMM}(\theta) = \widehat{\boldsymbol{g}}(\theta)' \widehat{\boldsymbol{W}} \widehat{\boldsymbol{g}}(\theta)$$

- $\widehat{\boldsymbol{g}}(\theta) = \frac{1}{N}\sum_i \boldsymbol{f}(\boldsymbol{Z}_i, \theta)$ is a $J \times 1$ vector of moment conditions
- CMD nested by separable case where $\boldsymbol{f}(\boldsymbol{Z}_i, \theta) = \boldsymbol{\pi}(\boldsymbol{Z}_i) + \tilde{\boldsymbol{f}}(\theta)$
- MM nested by just-id case where $J = dim(\theta)$

Identification

Population moment restrictions:

$$E\left[\boldsymbol{f}\left(\boldsymbol{Z}_i, \theta\right)\right] = 0 \text{ iff } \theta = \theta_0$$

Examples:

- Instrumental variables (orthogonality condition):

$$E\left[\left(Y_i - X_i'\beta\right) Z_i\right] = 0$$

- Rational expectations restrictions (e.g., Hall, 1978) (real interest rate $= 0$)

$$u'(C_t) = \beta\,\mathrm{E}[u'(C_{t+1})|\Omega_t]$$

Derivation of Asymptotic Variance

FOC:

$$\widehat{\boldsymbol{G}}\left(\widehat{\theta}\right)'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{g}}\left(\widehat{\theta}\right)=0$$

Mean value expansion:

$$\widehat{\boldsymbol{g}}\left(\widehat{\theta}\right)=\widehat{\boldsymbol{g}}\left(\theta_0\right)+\widehat{\boldsymbol{G}}\left(\overline{\theta}\right)\left(\widehat{\theta}-\theta_0\right)$$

Substitute in to get:

$$\widehat{\boldsymbol{G}}\left(\widehat{\theta}\right)'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{g}}\left(\theta_0\right)+\widehat{\boldsymbol{G}}\left(\widehat{\theta}\right)'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{G}}\left(\overline{\theta}\right)\left(\widehat{\theta}-\theta_0\right)=0$$

Therefore

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)=\left(\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)\right)^{-1}\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\sqrt{N}\widehat{\boldsymbol{g}}\left(\theta_0\right)+o_p\left(1\right)$$

Derivation of Asymptotic Variance

By assumption:

$$E\left[\sqrt{N}\widehat{\boldsymbol{g}}\left(\theta_0\right)\right] = 0$$

Hence, for $\boldsymbol{V_f} = E\left[\boldsymbol{f}\left(\boldsymbol{Z}_i, \theta_0\right)\boldsymbol{f}\left(\boldsymbol{Z}_i, \theta_0\right)'\right]$, we have

$$AVAR\left(\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\sqrt{N}\widehat{\boldsymbol{g}}\left(\theta_0\right)\right) = \boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{V_f}\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)$$

Consequently $AVAR\left(\sqrt{N}\left(\widehat{\theta} - \theta_0\right)\right)$ is:

$$\left(\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{H}}\right)^{-1}\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{V_f}\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{V}}\left(\underbrace{\boldsymbol{G}\left(\theta_0\right)'\boldsymbol{W}\boldsymbol{G}\left(\theta_0\right)}_{\boldsymbol{H}}\right)^{-1}$$

Plugin sample analogues to get standard errors

Optimal Weighting

When $\boldsymbol{W} = \boldsymbol{V_f}^{-1}$ (optimal weighting) AVAR simplifies to:

$$\left(\boldsymbol{G}\left(\theta_0\right)' \boldsymbol{V_f}^{-1} \boldsymbol{G}\left(\theta_0\right)\right)^{-1}$$

2-step estimation approach (ala 3SLS):

1. get inefficient starting estimates $\hat{\theta}^{(1)}$ from choosing $\hat{\boldsymbol{W}} = \boldsymbol{I}$.
2. minimize $Q_{GMM}\left(\theta\right)$ using $\hat{\boldsymbol{W}} = \boldsymbol{W}\left(\hat{\theta}^{(1)}\right)$

Keep going?: further steps provide no (first order) asymptotic advantage but often perform better in finite samples

CUGMM (Hansen, Heaton, and Yaron, 1996)

As w/ OMD, problems emerge when weight matrix has too many unknown parameters

One solution: "continuously update" weight matrix $\boldsymbol{W}(\theta)$ and minimize

$$\widehat{Q}_{CUGMM}(\theta) = \widehat{\boldsymbol{g}}(\theta)' \widehat{\boldsymbol{W}}(\theta) \widehat{\boldsymbol{g}}(\theta)$$

Better asymptotic performance than GMM but can be difficult to find minimum.

Empirical Likelihood

GMM a bit like a quadratic approx to a log likelihood
EL: maximize non-parametric likelihood fn subject to moment restrictions

$$\max_{\{\pi_i\},\theta} \prod_{i=1}^{N} \pi_i$$

$$s.t. \quad \sum_{i=1}^{N} \pi_i \boldsymbol{f}\left(\boldsymbol{Z}_i, \theta\right) = 0$$

$$\sum_{i=1}^{N} \pi_i = 1$$

Difficult optimization problem:

- w/o constraints $\hat{\pi}_i = \frac{1}{N}$ (returns the EDF)
- w/ constraints $\{\hat{\pi}_i\}$ give efficient estimates of joint distribution of data and $\hat{\theta}$ higher order unbiased (Newey and Smith, 2004)
- Generally difficult to compute (extension to CUGMM).

Maximum Likelihood

You are willing to specify the *entire* conditional distribution of the outcome given covariates.

Maximum Likelihood

You are willing to specify the *entire* conditional distribution of the outcome given covariates. Objective fn:

$$\widehat{Q}_{ML}(\theta) = \frac{1}{N}\sum_i l(\boldsymbol{Z}_i, \theta)$$

FOC:

$$\frac{1}{N}\sum_i \boldsymbol{s}\left(\boldsymbol{Z}_i, \widehat{\theta}\right) = 0$$

- ML: if model is correct $\rightarrow E[\boldsymbol{s}(\boldsymbol{Z}_i, \theta_0)] = 0$
- No weighting problems!
- Likelihood chooses the right moments for you! (embedded in the score function)
  - "Efficient Method of Moments" on FOCs (Gallant and Tauchen, 1996)
- ML is prone to misspecification. Even getting a minor feature of the entire distribution wrong can make the ML estimator inconsistent.
- Can you think of a relevant counter-example?

Asymptotic variance

"Influence function" representation – 1st order effect of adding an obs on estimator:

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)=\widehat{\boldsymbol{H}}\left(\theta_0\right)^{-1}\frac{1}{\sqrt{N}}\sum_i \boldsymbol{s}\left(\boldsymbol{Z}_i,\theta_0\right)+o_p\left(1\right)$$

Variance of IF gives asymptotic variance:

$$\boldsymbol{H}\left(\theta_0\right)^{-1}\boldsymbol{V}\boldsymbol{H}\left(\theta_0\right)^{-1}$$

where $\boldsymbol{V}=\boldsymbol{E}\left[\boldsymbol{s}\left(\boldsymbol{Z}_i,\theta_0\right)\boldsymbol{s}\left(\boldsymbol{Z}_i,\theta_0\right)'\right]$

Asymptotic variance

"Influence function" representation – 1st order effect of adding an obs on estimator:

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)=\widehat{\boldsymbol{H}}\left(\theta_0\right)^{-1}\frac{1}{\sqrt{N}}\sum_i \boldsymbol{s}\left(\boldsymbol{Z}_i,\theta_0\right)+o_p\left(1\right)$$

Variance of IF gives asymptotic variance:

$$\boldsymbol{H}\left(\theta_0\right)^{-1}\boldsymbol{V}\boldsymbol{H}\left(\theta_0\right)^{-1}$$

where $\boldsymbol{V}=\boldsymbol{E}\left[\boldsymbol{s}\left(\boldsymbol{Z}_i,\theta_0\right)\boldsymbol{s}\left(\boldsymbol{Z}_i,\theta_0\right)'\right]$

Information matrix equality:

$$\boldsymbol{V}=\boldsymbol{H}\left(\theta_0\right)$$

When this holds, asymptotic variance reduces to Cramer-Rao lower bound $\boldsymbol{H}\left(\theta_0\right)^{-1}$.

Nothing beats the CRLB $\rightarrow$ nothing beats a properly specified likelihood model.

Simulation Estimators

Often difficult to write down the likelihood

- Particularly difficult for multinomial choice problems where regions of integration can quickly become intractable
- Or for parametric models with lots of unobserved heterogeneity

Simulation Estimators

Often difficult to write down the likelihood

- Particularly difficult for multinomial choice problems where regions of integration can quickly become intractable
- Or for parametric models with lots of unobserved heterogeneity

Simulation methods

- Can use computer to compute likelihood via simulation (MSL)
- With continuous variables often easier to simulate moment conditions (MSM)
- Great (free) introduction by Train (2009) or Chapter 12 of Cameron and Trivedi.

Simulation Estimators

Often difficult to write down the likelihood

- Particularly difficult for multinomial choice problems where regions of integration can quickly become intractable
- Or for parametric models with lots of unobserved heterogeneity

Simulation methods

- Can use computer to compute likelihood via simulation (MSL)
- With continuous variables often easier to simulate moment conditions (MSM)
- Great (free) introduction by Train (2009) or Chapter 12 of Cameron and Trivedi.

Warning: can easily waste a lot of time on this stuff

- Tendency to use these methods when you don't fully understand the model / identification
- Important to "warm up" by estimating simpler models that you understand!

Maximum Simulated Likelihood

$$\hat{\theta}_{MSL} \equiv \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \ln \hat{L}_M \left(\boldsymbol{Z}_i, \theta \right)$$

- $\hat{L}_M \left(\boldsymbol{Z}_i, \theta \right) = \hat{L} \left(\boldsymbol{Z}_i, \theta, u_{i1}, u_{i2}, ..., u_{iM} \right)$ is simulated likelihood
- Fully parametric: drawing $\left\{ u_{im} \right\}_{m=1}^{M}$ from <u>known</u> distribution
- Hold $\left\{ \boldsymbol{u}_i \right\}_{i=1}^{N}$ fixed when searching over θ to avoid "chatter"
- Need large M to ensure consistency:
  $\lim_{M \to \infty} P \left(\left| \hat{L}_M \left(\boldsymbol{Z}_i, \theta \right) - L \left(\boldsymbol{Z}_i, \theta \right) \right| > \varepsilon \right) = 0$

Example (Probit)

Probit likelihood:

$$L(Y_i, X_i, \beta) = \Phi\left(X_i'\beta\right)^{Y_i}\left[1 - \Phi\left(X_i'\beta\right)\right]^{1-Y_i}$$

Simulated analogue:

$$\begin{aligned}\hat{L}_M(Y_i, X_i, \beta) &= \left(\frac{1}{M}\sum_{m=1}^{M} 1\left[X_i'\beta - u_{im} > 0\right]\right)^{Y_i} \\ &\quad \times \left(\frac{1}{M}\sum_{m=1}^{M} 1\left[X_i'\beta - u_{im} < 0\right]\right)^{1-Y_i}\end{aligned}$$

where $\left\{u_{im} \overset{iid}{\sim} N(0,1)\right\}_{i=1,m=1}^{N,M}$

- Unbiased simulator: $E\left[\hat{L}_M(Y_i, X_i, \beta)\right] = L(Y_i, X_i, \beta)$
- But $\hat{L}_M(Y_i, X_i, \beta)$ nondifferentiable for finite M (step function)

Example (Random Coefficient Logit)

Suppose

$$Y_i = 1\left[b_i X_i + \varepsilon_i > 0\right]$$

where $\varepsilon_i \sim Logistic(0,1)$ and $b_i \sim N\left(\mu, \sigma^2\right)$

- Could simulate likelihood by taking draws from (ε_i, b_i).
- Better to integrate out ε_i analytically to obtain smooth objective fn:

$$\begin{aligned}\hat{L}_M\left(Y_i, X_i, \mu, \sigma\right) \equiv & \left(\frac{1}{M}\sum_{m=1}^{M}\Lambda\left(\left(\mu+\sigma u_{im}\right)X_i\right)\right)^{Y_i} \\ & \times\left(1-\frac{1}{M}\sum_{m=1}^{M}\Lambda\left(\left(\mu+\sigma u_{im}\right)X_i\right)\right)^{1-Y_i}\end{aligned}$$

Asymptotics

Pakes and Pollard (1989): general asymptotic framework for simulation estimators ("stochastic equicontinuity")

Gourieroux and Monfort (1991): if $M, N \to \infty$ and $\sqrt{N}/M \to 0$, then $\widehat{\theta}_{MSL} \xrightarrow{p} \theta_0$ and

$$\sqrt{N}\left(\widehat{\theta}_{MSL} - \theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}\left(\theta_0\right)^{-1} \boldsymbol{V} \boldsymbol{H}\left(\theta_0\right)^{-1}\right)$$

Asymptotics

Pakes and Pollard (1989): general asymptotic framework for simulation estimators ("stochastic equicontinuity")

Gourieroux and Monfort (1991): if $M, N \to \infty$ and $\sqrt{N}/M \to 0$, then $\widehat{\theta}_{MSL} \xrightarrow{p} \theta_0$ and

$$\sqrt{N}\left(\widehat{\theta}_{MSL} - \theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}\left(\theta_0\right)^{-1} \boldsymbol{V} \boldsymbol{H}\left(\theta_0\right)^{-1}\right)$$

Practical advice: start with $M = 100$ to get initial estimates then check stability as M increases

Asymptotics

Pakes and Pollard (1989): general asymptotic framework for simulation estimators ("stochastic equicontinuity")

Gourieroux and Monfort (1991): if $M, N \to \infty$ and $\sqrt{N}/M \to 0$, then $\widehat{\theta}_{MSL} \xrightarrow{p} \theta_0$ and

$$\sqrt{N}\left(\widehat{\theta}_{MSL} - \theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}\left(\theta_0\right)^{-1} \boldsymbol{V} \boldsymbol{H}\left(\theta_0\right)^{-1}\right)$$

Practical advice: start with $M = 100$ to get initial estimates then check stability as M increases

Std errors: plug in simulation estimators of $\boldsymbol{H}\left(\hat{\theta}_{MSL}\right), \boldsymbol{V}\left(\hat{\theta}_{MSL}\right)$

Method of Simulated Moments

$$\widehat{\theta}_{MSM} \equiv \arg\min_{\theta} \widehat{\boldsymbol{g}}_M(\theta)' \widehat{\boldsymbol{W}} \widehat{\boldsymbol{g}}_M(\theta)$$

- $\widehat{\boldsymbol{g}}_M(\theta) = \frac{1}{N} \sum_i \widehat{\boldsymbol{f}}_M(\boldsymbol{Z}_i, \theta)$ is simulated moment condition
- McFadden (1989): for <u>fixed</u> M, as $N \to \infty$, $\widehat{\theta}_{MSM} \xrightarrow{p} \theta_0$ and $AVAR\left(\sqrt{N}\left(\widehat{\theta}_{MSM} - \theta_0\right)\right) =$

$$\left(\boldsymbol{G}(\theta_0)' \boldsymbol{W}\boldsymbol{G}(\theta_0)\right)^{-1} \boldsymbol{G}(\theta_0)' \boldsymbol{W}\boldsymbol{V}_{\boldsymbol{f},M} \boldsymbol{W}\boldsymbol{G}(\boldsymbol{\theta_0}) \left(\boldsymbol{G}(\theta_0)' \boldsymbol{W}\boldsymbol{G}(\theta_0)\right)^{-1}$$

Which moments to match?

If model is fully parametric, best to match the scores: "method of simulated scores" Hajivassiliou and McFadden (1998)

- Problem: typically hard to get unbiased simulator of scores
- Example (binary choice):

$$E\left[Y_i \frac{\frac{\partial}{\partial\theta} P\left(Y_i = 1|X_i\right)}{P\left(Y_i = 1|X_i\right)} - \left(1 - Y_i\right)\frac{\frac{\partial}{\partial\theta} P\left(Y_i = 1|X_i\right)}{1 - P\left(Y_i = 1|X_i\right)}\right] = 0$$

Which moments to match?

If model is fully parametric, best to match the scores: "method of simulated scores" Hajivassiliou and McFadden (1998)

- Problem: typically hard to get unbiased simulator of scores
- Example (binary choice):

$$E\left[Y_i \frac{\frac{\partial}{\partial\theta} P\left(Y_i = 1|X_i\right)}{P\left(Y_i = 1|X_i\right)} - \left(1 - Y_i\right) \frac{\frac{\partial}{\partial\theta} P\left(Y_i = 1|X_i\right)}{1 - P\left(Y_i = 1|X_i\right)}\right] = 0$$

- Can you see a problem with this?

MSL vs MSM

Rules of thumb:

- parametric discrete choice model $\Rightarrow$ MSL
  - MSL chooses the "right" moments
  - Works well for moderate M
- semi-parametric model or model w/ mixed continuous-discrete choices $\Rightarrow$ MSM
  - don't want to match moments you don't believe
  - too hard to simulate multidimensional density of continuous variables

MSL vs MSM

Rules of thumb:

- parametric discrete choice model $\Rightarrow$ MSL
  - MSL chooses the "right" moments
  - Works well for moderate M
- semi-parametric model or model w/ mixed continuous-discrete choices $\Rightarrow$ MSM
  - don't want to match moments you don't believe
  - too hard to simulate multidimensional density of continuous variables

Eisenhauer, Heckman, and Mosso (2015, IER):

- Empirical horserace of MSL vs MSM for modern dynamic discrete choice model
- Conclusion: $MSL > MSM$ (assuming model is correct)

Appendix

Proof

For any ε we have with probability approaching 1:

a) $\widehat{Q}_N\left(\widehat{\theta}\right) > \widehat{Q}_N(\theta_0) - \varepsilon$ (by virtue of maximizer)

b) $Q_0\left(\widehat{\theta}\right) > \widehat{Q}_N\left(\widehat{\theta}\right) - \varepsilon$ (by uniform convergence)

c) $\widehat{Q}_N(\theta_0) > Q_0(\theta_0) - \varepsilon$ (by uniform convergence)

Therefore:

$$Q_0\left(\widehat{\theta}\right) \overset{b)}{>} \widehat{Q}_N\left(\widehat{\theta}\right) - \varepsilon \overset{a)}{>} \widehat{Q}_N(\theta_0) - 2\varepsilon \overset{c)}{>} Q_0(\theta_0) - 3\varepsilon.$$

Proof

For any ε we have with probability approaching 1:

a) $\widehat{Q}_N\left(\widehat{\theta}\right) > \widehat{Q}_N(\theta_0) - \varepsilon$ (by virtue of maximizer)

b) $Q_0\left(\widehat{\theta}\right) > \widehat{Q}_N\left(\widehat{\theta}\right) - \varepsilon$ (by uniform convergence)

c) $\widehat{Q}_N(\theta_0) > Q_0(\theta_0) - \varepsilon$ (by uniform convergence)

Therefore:

$$Q_0\left(\widehat{\theta}\right) \overset{b)}{>} \widehat{Q}_N\left(\widehat{\theta}\right) - \varepsilon \overset{a)}{>} \widehat{Q}_N(\theta_0) - 2\varepsilon \overset{c)}{>} Q_0(\theta_0) - 3\varepsilon.$$

Let $\mathcal{N}$ be an open neighborhood of θ_0 and $\mathcal{N}^c$ its complement.
Note that $\sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta) = Q_0(\theta^*) < Q_0(\theta_0)$.

- Choose $3\varepsilon = Q_0(\theta_0) - \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta)$.
- It follows that $Q_0\left(\widehat{\theta}\right) > \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta)$ w.p.a.1.

Hence $\widehat{\theta} \in \mathcal{N}$ $\square$

Proof Sketch

Start w/ FOC:

$$\nabla_\theta \widehat{Q}_N \left(\widehat{\theta}\right) = 0$$

- Mean value expansion:

$$\nabla_\theta \widehat{Q}_N \left(\widehat{\theta}\right) = \nabla_\theta \widehat{Q}_N \left(\theta_0\right) + \widehat{\boldsymbol{H}}\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_0\right)$$

  where $\overline{\theta} \in \left[\widehat{\theta}, \theta_0\right]$ and $\widehat{\boldsymbol{H}}\left(\theta\right) \equiv \bigtriangledown_{\theta\theta} \widehat{Q}_N \left(\theta\right)$.
- Rearranging we get:

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) = -\widehat{\boldsymbol{H}}\left(\overline{\theta}\right)^{-1}\left(\sqrt{N}\nabla_\theta \widehat{Q}_N \left(\theta_0\right)\right)$$

Proof Sketch

Start w/ FOC:

$$\nabla_\theta \widehat{Q}_N\left(\widehat{\theta}\right) = 0$$

- Mean value expansion:

$$\nabla_\theta \widehat{Q}_N\left(\widehat{\theta}\right) = \nabla_\theta \widehat{Q}_N\left(\theta_0\right) + \widehat{\boldsymbol{H}}\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_0\right)$$

  where $\overline{\theta} \in \left[\widehat{\theta}, \theta_0\right]$ and $\widehat{\boldsymbol{H}}\left(\theta\right) \equiv \triangledown_{\theta\theta} \widehat{Q}_N\left(\theta\right)$.

- Rearranging we get:

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) = -\widehat{\boldsymbol{H}}\left(\overline{\theta}\right)^{-1}\left(\sqrt{N}\nabla_\theta \widehat{Q}_N\left(\theta_0\right)\right)$$

- Want to show:

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) = -\boldsymbol{H}\left(\theta_0\right)^{-1}\left(\sqrt{N}\nabla_\theta \widehat{Q}_N\left(\theta_0\right)\right) + o_p\left(1\right)$$

Proof Sketch

Start w/ FOC:

$$\nabla_\theta \widehat{Q}_N\left(\widehat{\theta}\right) = 0$$

- Mean value expansion:

$$\nabla_\theta \widehat{Q}_N\left(\widehat{\theta}\right) = \nabla_\theta \widehat{Q}_N\left(\theta_0\right) + \widehat{\boldsymbol{H}}\left(\overline{\theta}\right)\left(\widehat{\theta} - \theta_0\right)$$

  where $\overline{\theta} \in \left[\widehat{\theta}, \theta_0\right]$ and $\widehat{\boldsymbol{H}}\left(\theta\right) \equiv \bigtriangledown_{\theta\theta} \widehat{Q}_N\left(\theta\right)$.

- Rearranging we get:

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) = -\widehat{\boldsymbol{H}}\left(\overline{\theta}\right)^{-1}\left(\sqrt{N}\nabla_\theta \widehat{Q}_N\left(\theta_0\right)\right)$$

- Want to show:

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) = -\boldsymbol{H}\left(\theta_0\right)^{-1}\left(\sqrt{N}\nabla_\theta \widehat{Q}_N\left(\theta_0\right)\right) + o_p\left(1\right)$$

- Need $\left\|\widehat{\boldsymbol{H}}\left(\overline{\theta}\right) - \boldsymbol{H}\left(\theta_0\right)\right\| \xrightarrow{p} 0$

Proof Sketch

Triangle inequality:

$$
\begin{aligned}
\left\|\widehat{\boldsymbol{H}}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\theta_0\right)\right\| &\leq \underbrace{\left\|\widehat{\boldsymbol{H}}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\bar{\theta}\right)\right\|}_{\text{Noise in Hessian}}+\underbrace{\left\|\boldsymbol{H}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\theta_0\right)\right\|}_{\text{Noise in parameter}} \\
&\leq \sup_{\theta\in\Theta}\left\|\widehat{\boldsymbol{H}}\left(\theta\right)-\boldsymbol{H}\left(\theta\right)\right\|+\left\|\boldsymbol{H}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\theta_0\right)\right\|
\end{aligned}
$$

Proof Sketch

Triangle inequality:

$$
\begin{aligned}
\left\|\widehat{\boldsymbol{H}}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\theta_0\right)\right\| &\leq \underbrace{\left\|\widehat{\boldsymbol{H}}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\bar{\theta}\right)\right\|}_{\text{Noise in Hessian}}+\underbrace{\left\|\boldsymbol{H}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\theta_0\right)\right\|}_{\text{Noise in parameter}} \\
&\leq \sup_{\theta\in\Theta}\left\|\widehat{\boldsymbol{H}}\left(\theta\right)-\boldsymbol{H}\left(\theta\right)\right\|+\left\|\boldsymbol{H}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\theta_0\right)\right\|
\end{aligned}
$$

- By assumption $\sup_{\theta\in\Theta}\left\|\widehat{\boldsymbol{H}}\left(\theta\right)-\boldsymbol{H}\left(\theta\right)\right\| \xrightarrow{p} 0$ (kills 1st term)
- Because $\hat{\theta} \xrightarrow{p} \theta_0$, it must also be that $\bar{\theta} \xrightarrow{p} \theta_0$ (squeeze theorem).
- Since $\boldsymbol{H}\left(.\right)$ is continuous, it follows from the continuous mapping theorem that $\left\|\boldsymbol{H}\left(\bar{\theta}\right)-\boldsymbol{H}\left(\theta_0\right)\right\| \xrightarrow{p} 0$ (kills 2nd term)
- Therefore $\widehat{\boldsymbol{H}}\left(\bar{\theta}\right) \xrightarrow{p} \boldsymbol{H}\left(\theta_0\right)$

Proof Sketch

Recall

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right)=-\widehat{\boldsymbol{H}}\left(\overline{\theta}\right)^{-1}\left(\sqrt{N}\nabla_{\theta}\widehat{Q}_N\left(\theta_0\right)\right)$$

Proof Sketch

Recall

$$\sqrt{N}\left(\widehat{\theta}-\theta_0\right) = -\widehat{\boldsymbol{H}}\left(\overline{\theta}\right)^{-1}\left(\sqrt{N}\nabla_\theta \widehat{Q}_N\left(\theta_0\right)\right)$$

- We established $\widehat{\boldsymbol{H}}\left(\overline{\theta}\right) \xrightarrow{p} \boldsymbol{H}\left(\theta_0\right)$
- By continuous mapping theorem, $\widehat{\boldsymbol{H}}\left(\overline{\theta}\right)^{-1} \xrightarrow{p} \boldsymbol{H}\left(\theta_0\right)^{-1}$
- And by assumption: $\sqrt{N}\nabla_\theta \widehat{Q}_N\left(\theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{V}\right)$
- Therefore, by Slutsky,
  $\sqrt{N}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}\left(\theta_0\right)^{-1}\boldsymbol{V}\boldsymbol{H}\left(\theta_0\right)^{-1}\right)$ □

Example: "folding tent"

$$Q_N(\theta) = \begin{cases} \max\{4N\theta - 2, 0\} & \text{if } \theta \le \frac{3}{4N} \\ \max\{-4N\theta + 4, 0\} & \text{if } \theta > \frac{3}{4N} \end{cases}$$

Example: "folding tent"

$$Q_N(\theta) = \begin{cases} \max\{4N\theta - 2, 0\} & \text{if } \theta \leq \frac{3}{4N} \\ \max\{-4N\theta + 4, 0\} & \text{if } \theta > \frac{3}{4N} \end{cases}$$

Example: "folding tent"

$$Q_N(\theta) = \begin{cases} \max\{4N\theta - 2, 0\} & \text{if } \theta \leq \frac{3}{4N} \\ \max\{-4N\theta + 4, 0\} & \text{if } \theta > \frac{3}{4N} \end{cases}$$

- $\Theta = [0, 1]$ is compact

Example: "folding tent"

$$Q_N(\theta) = \begin{cases} \max\{4N\theta - 2, 0\} & \text{if } \theta \leq \frac{3}{4N} \\ \max\{-4N\theta + 4, 0\} & \text{if } \theta > \frac{3}{4N} \end{cases}$$

- $\Theta = [0, 1]$ is compact
- $Q_N(.)$ is continuous

Example: "folding tent"

$$Q_N(\theta) = \begin{cases} \max\{4N\theta - 2, 0\} & \text{if } \theta \le \frac{3}{4N} \\ \max\{-4N\theta + 4, 0\} & \text{if } \theta > \frac{3}{4N} \end{cases}$$

- $\Theta = [0, 1]$ is compact
- $Q_N(.)$ is continuous
- $\lim_{N\to\infty} Q_N(\theta) = 0 \ \forall \theta \in [0, 1]$ (pointwise convergence)

Example: "folding tent"

$$Q_N(\theta) = \begin{cases} \max\{4N\theta - 2, 0\} & \text{if } \theta \leq \frac{3}{4N} \\ \max\{-4N\theta + 4, 0\} & \text{if } \theta > \frac{3}{4N} \end{cases}$$

- $\Theta = [0, 1]$ is compact
- $Q_N(.)$ is continuous
- $\lim_{N\to\infty} Q_N(\theta) = 0 \ \forall \theta \in [0, 1]$ (pointwise convergence)
- But $\max_{\theta\in\Theta} |Q_N(\theta) - 0| = 1 \ \forall N \in \mathbb{N}$!