Posterior Distribution of Non-differentiable Functions

Toru Kitagawa (UCL), José-Luis Montiel-Olea (Columbia), Jonathan Payne (NYU)

September 9, 2016

Motivation

- $\star$ This note was motivated by previous work on set-id SVARs.
  ($\lambda \in [\underline{g}(\theta)\ ,\ \overline{g}(\theta)]$, $\lambda \in \mathbb{R}$, $\theta \in \mathbb{R}^p$, $\underline{g}, \overline{g}$ nondifferentiable)

Motivation

- ⋆ This note was motivated by previous work on set-id SVARs.
  ($\lambda \in [\underline{g}(\theta)\ ,\ \overline{g}(\theta)]$, $\lambda \in \mathbb{R}$, $\theta \in \mathbb{R}^p$, $\underline{g}, \overline{g}$ nondifferentiable)

- ⋆ Most of the statistical analysis in the SVAR literature is Bayesian,
  (but most of the earlier work on set-identified models is Freq.)

Motivation

- ⋆ This note was motivated by previous work on set-id SVARs.
  ($\lambda \in [\underline{g}(\theta)\ ,\ \overline{g}(\theta)]$, $\lambda \in \mathbb{R}$, $\theta \in \mathbb{R}^p$, $\underline{g}, \overline{g}$ nondifferentiable)

- ⋆ Most of the statistical analysis in the SVAR literature is Bayesian,
  (but most of the earlier work on set-identified models is Freq.)

- ⋆ Standard Bayesian analysis of set-id models is not 'robust'.
  (*credibility* for diff. λ-priors need not be above $1 - \alpha$)

Motivation

- $\star$ This note was motivated by previous work on set-id SVARs.
  ($\lambda \in [\underline{g}(\theta)\, , \, \overline{g}(\theta)]$, $\lambda \in \mathbb{R}$, $\theta \in \mathbb{R}^p$, $\underline{g}, \overline{g}$ nondifferentiable)

- $\star$ Most of the statistical analysis in the SVAR literature is Bayesian,
  (but most of the earlier work on set-identified models is Freq.)

- $\star$ Standard Bayesian analysis of set-id models is not 'robust'.
  (*credibility* for diff. λ-priors need not be above $1-\alpha$)

- $\star$ 'Robust' Bayes credible sets require the posterior of $\underline{g}(\theta)$, $\overline{g}(\theta)$.
  (this approach also shows up in Kline & Tamer [16])

Motivation

- ⋆ This note was motivated by previous work on set-id SVARs.
  ($\lambda \in [\underline{g}(\theta)\ ,\ \overline{g}(\theta)]$, $\lambda \in \mathbb{R}$, $\theta \in \mathbb{R}^p$, $\underline{g}, \overline{g}$ nondifferentiable)
- ⋆ Most of the statistical analysis in the SVAR literature is Bayesian,
  (but most of the earlier work on set-identified models is Freq.)
- ⋆ Standard Bayesian analysis of set-id models is not 'robust'.
  (*credibility* for diff. λ-priors need not be above $1-\alpha$)
- ⋆ 'Robust' Bayes credible sets require the posterior of $\underline{g}(\theta)$, $\overline{g}(\theta)$.
  (this approach also shows up in Kline & Tamer [16])
- ⋆ We wanted to understand the behavior of these posteriors
  (at least in situations where the sample size is large)

This project

⋆ We study the posterior distribution of $g(\theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^p$.
(large sample from a parametric model and a fixed prior for θ)

This project

- ⋆ We study the posterior distribution of $g(\theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^p$.
  (large sample from a parametric model and a fixed prior for θ)
- ⋆ We describe the relation between the posterior of $g(\theta)$ and:

This project

- ⋆ We study the posterior distribution of $g(\theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^p$.
  (large sample from a parametric model and a fixed prior for θ)

- ⋆ We describe the relation between the posterior of $g(\theta)$ and:

  - ⋆ The distribution of the (plug-in) ML estimator, $g(\hat{\theta}_n)$.

This project

- ⋆ We study the posterior distribution of $g(\theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^p$.
  (large sample from a parametric model and a fixed prior for θ)
- ⋆ We describe the relation between the posterior of $g(\theta)$ and:
  - ⋆ The distribution of the (plug-in) ML estimator, $g(\hat{\theta}_n)$.
  - ⋆ The bootstrap distribution of $g(\hat{\theta}_n)$.

This project

- ⋆ We study the posterior distribution of $g(\theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^p$.
  (large sample from a parametric model and a fixed prior for θ)
- ⋆ We describe the relation between the posterior of $g(\theta)$ and:
  - ⋆ The distribution of the (plug-in) ML estimator, $g(\hat{\theta}_n)$.
  - ⋆ The bootstrap distribution of $g(\hat{\theta}_n)$.
- ⋆ To do so, we will work under the following assumptions:

This project

- ⋆ We study the posterior distribution of $g(\theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^p$.
  (large sample from a parametric model and a fixed prior for θ)

- ⋆ We describe the relation between the posterior of $g(\theta)$ and:
  - ⋆ The distribution of the (plug-in) ML estimator, $g(\hat{\theta}_n)$.
  - ⋆ The bootstrap distribution of $g(\hat{\theta}_n)$.

- ⋆ To do so, we will work under the following assumptions:
  - ⋆ $g : \mathbb{R}^p \to \mathbb{R}$ is Lipschitz continuous

This project

- ⋆ We study the posterior distribution of $g(\theta) \in \mathbb{R}$, $\theta \in \mathbb{R}^p$.
  (large sample from a parametric model and a fixed prior for θ)
- ⋆ We describe the relation between the posterior of $g(\theta)$ and:
  - ⋆ The distribution of the (plug-in) ML estimator, $g(\hat{\theta}_n)$.
  - ⋆ The bootstrap distribution of $g(\hat{\theta}_n)$.
- ⋆ To do so, we will work under the following assumptions:
  - ⋆ $g : \mathbb{R}^p \to \mathbb{R}$ is Lipschitz continuous
  - ⋆ A Bernstein von-Mises Theorem holds for θ

Main Result and Implications

⋆ Posterior distribution of $g(\theta)$∼Bootstrap distribution of $g(\widehat{\theta}_n)$
(as the sample size grows large and the BvM theorem kicks in)

Main Result and Implications

⋆ Posterior distribution of $g(\theta)$∼Bootstrap distribution of $g(\widehat{\theta}_n)$
(as the sample size grows large and the BvM theorem kicks in)

$\Longrightarrow$ Bayesians can use the bootstrap dist. of $g(\widehat{\theta}_n)$ as a posterior.
(May be simpler than MCMC; "poor man's posterior")

MAIN RESULT AND IMPLICATIONS

$\star$ Posterior distribution of $g(\theta)$∼Bootstrap distribution of $g(\widehat{\theta}_n)$
(as the sample size grows large and the BvM theorem kicks in)

$\Longrightarrow$ Bayesians can use the bootstrap dist. of $g(\widehat{\theta}_n)$ as a posterior.
(May be simpler than MCMC; "poor man's posterior")

$\Longrightarrow$ Posterior credible sets for $g(\theta)$ need not be valid conf. sets.
(even if this relation holds true for θ)
(Dumbgen [93], Andrews [00] and Santos & Fang [16]).

Outline

1. Illustrative example: $g(\theta) = |\theta|$
2. Main Result: Asymptotic Behavior of Posterior
3. Corollary 1: Directional Differentiability
4. Corollary 2: Bootstrap fails $\implies$ Posterior fails
5. Conclusion

Outline

1. Illustrative example: $g(\theta) = |\theta|$
2. Main Result: Asymptotic Behavior of Posterior
3. Corollary 1: Directional Differentiability
4. Corollary 2: Bootstrap fails $\implies$ Posterior fails
5. Conclusion

Outline

1. Illustrative example: $g(\theta) = |\theta|$
2. Main Result: Asymptotic Behavior of Posterior
3. Corollary 1: Directional Differentiability
4. Corollary 2: Bootstrap fails $\implies$ Posterior fails
5. Conclusion

Outline

1. Illustrative example: $g(\theta) = |\theta|$
2. Main Result: Asymptotic Behavior of Posterior
3. Corollary 1: Directional Differentiability
4. Corollary 2: Bootstrap fails $\implies$ Posterior fails
5. Conclusion

Outline

1. Illustrative example: $g(\theta) = |\theta|$
2. Main Result: Asymptotic Behavior of Posterior
3. Corollary 1: Directional Differentiability
4. Corollary 2: Bootstrap fails $\implies$ Posterior fails
5. Conclusion

1. Illustrative example: $g(\theta) = |\theta|$

Set-up

⋆ $X_i \sim \mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$ is unknown.

Set-up

⋆ $X_i \sim \mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$ is unknown.

⋆ The Maximum Likelihood Estimator for θ is given by:

$$\widehat{\theta}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i,$$

Set-up

⋆ $X_i \sim \mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$ is unknown.

⋆ The Maximum Likelihood Estimator for θ is given by:

$$\widehat{\theta}_n \equiv \frac{1}{n} \sum_{i=1}^{n} X_i,$$

⋆ Transformation: $g(\theta) = |\theta|$

Set-up

- $\star$ $X_i \sim \mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$ is unknown.
- $\star$ The Maximum Likelihood Estimator for θ is given by:

$$\widehat{\theta}_n \equiv \frac{1}{n} \sum_{i=1}^{n} X_i,$$

- $\star$ Transformation: $g(\theta) = |\theta|$
- $\star$ There is a prior P over the parameter θ.

Posterior for $|\theta|$

⋆ Suppose that the prior is $\theta \sim \mathcal{N}(0, (1/\lambda^2))$

Posterior for $|\theta|$

- ⋆ Suppose that the prior is $\theta \sim \mathcal{N}(0, (1/\lambda^2))$
- ⋆ Closed-form expression for the posterior of θ:

$$\theta_n^{P*}|X^n \sim \mathcal{N}\Big(\frac{n}{n+\lambda^2}\widehat{\theta}_n\,,\ \frac{1}{n+\lambda^2}\Big)$$

Posterior for $|\theta|$

- ⋆ Suppose that the prior is $\theta \sim \mathcal{N}(0, (1/\lambda^2))$

- ⋆ Closed-form expression for the posterior of θ:

$$\theta_n^{P*} | X^n \sim \mathcal{N}\Big(\frac{n}{n+\lambda^2}\widehat{\theta}_n \, , \, \frac{1}{n+\lambda^2}\Big)$$

- ⋆ Posterior draws for θ_n^{P*} can be expressed as:

$$\theta_n^j = \frac{Z_j^*}{\sqrt{n+\lambda^2}} + \frac{n}{n+\lambda^2}\widehat{\theta}_n, \quad Z_j^* \sim \mathcal{N}(0,1) \perp \text{prior,data.}$$

Posterior for $|\theta|$

- ⋆ Suppose that the prior is $\theta \sim \mathcal{N}(0, (1/\lambda^2))$

- ⋆ Closed-form expression for the posterior of θ:

$$\theta_n^{P*} | X^n \sim \mathcal{N}\Big(\frac{n}{n+\lambda^2}\widehat{\theta}_n\,,\ \frac{1}{n+\lambda^2}\Big)$$

- ⋆ Posterior draws for θ_n^{P*} can be expressed as:

$$\theta_n^j = \frac{Z_j^*}{\sqrt{n+\lambda^2}} + \frac{n}{n+\lambda^2}\widehat{\theta}_n, \quad Z_j^* \sim \mathcal{N}(0,1) \perp \text{prior,data.}$$

- ⋆ Centered and scaled posterior draws of $g(\theta_n^j)$ are given by:

$$\boxed{\sqrt{n}(g(\theta_n^j) - g(\hat{\theta}_n)) = \sqrt{n}\Big(\Big|\frac{Z_j^*}{\sqrt{n+\lambda^2}} + \frac{n}{n+\lambda^2}\widehat{\theta}_n\Big| - \Big|\widehat{\theta}_n\Big|\Big)}$$

BOOTSTRAP FOR $|\widehat{\theta}_n|$

$\star$ We use the parametric bootstrap for simplicity.

BOOTSTRAP FOR $|\widehat{\theta}_n|$

- ⋆ We use the parametric bootstrap for simplicity.
- ⋆ Bootstrap distribution: $\theta_n^{B*}|X^n \sim \mathcal{N}(\widehat{\theta}_n, 1/n)$

Bootstrap for $|\widehat{\theta}_n|$

- ⋆ We use the parametric bootstrap for simplicity.
- ⋆ Bootstrap distribution: $\theta_n^{B*}|X^n \sim \mathcal{N}(\widehat{\theta}_n, 1/n)$
- ⋆ Bootstrap draws for θ_n^{B*} can be expressed as:

$$\theta_n^j = \frac{Z_j^{**}}{\sqrt{n}} + \widehat{\theta}_n, \quad Z_j^{**} \sim \mathcal{N}(0,1) \perp \text{data}.$$

Bootstrap for $|\widehat{\theta}_n|$

- ⋆ We use the parametric bootstrap for simplicity.
- ⋆ Bootstrap distribution: $\theta_n^{B*}|X^n \sim \mathcal{N}(\widehat{\theta}_n, 1/n)$
- ⋆ Bootstrap draws for θ_n^{B*} can be expressed as:

$$\theta_n^j = \frac{Z_j^{**}}{\sqrt{n}} + \widehat{\theta}_n, \quad Z_j^{**} \sim \mathcal{N}(0,1) \perp \text{data}.$$

- ⋆ Centered and scaled bootstrap draws of $g(\theta_n^j)$ are given by:

$$\boxed{\sqrt{n}(g(\theta_n^j) - g(\hat{\theta}_n)) = \sqrt{n}\Big(\Big|\frac{Z_j^{**}}{\sqrt{n}} + \widehat{\theta}_n\Big| - \Big|\widehat{\theta}_n\Big|\Big)}$$

Asymptotic equivalence

⋆ The centered and scaled posterior $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$:
(for prior $\theta \sim \mathcal{N}(0, \lambda^{-2})$)

$$\left(\left|\sqrt{\frac{n}{n+\lambda^2}} Z^* + \frac{n}{n+\lambda^2}\sqrt{n}\widehat{\theta}_n\right| - \left|\sqrt{n}\widehat{\theta}_n\right|\right), \quad Z^* \sim \mathcal{N}(0,1) \perp X^n$$

Asymptotic equivalence

- ⋆ The centered and scaled posterior $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$:
  (for prior $\theta \sim \mathcal{N}(0, \lambda^{-2})$)

$$\left(\left|\sqrt{\frac{n}{n+\lambda^2}}Z^* + \frac{n}{n+\lambda^2}\sqrt{n}\widehat{\theta}_n\right| - \left|\sqrt{n}\widehat{\theta}_n\right|\right), \quad Z^* \sim \mathcal{N}(0,1) \perp X^n$$

- ⋆ The centered and scaled distribution of $\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n))$:

$$\left(\left|Z^{**} + \sqrt{n}\widehat{\theta}_n\right| - \left|\sqrt{n}\widehat{\theta}_n\right|\right), \quad Z^{**} \sim \mathcal{N}(0,1) \perp X^n$$

Asymptotic equivalence

⋆ The centered and scaled posterior $\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n))$:
(for prior $\theta \sim \mathcal{N}(0, \lambda^{-2})$)

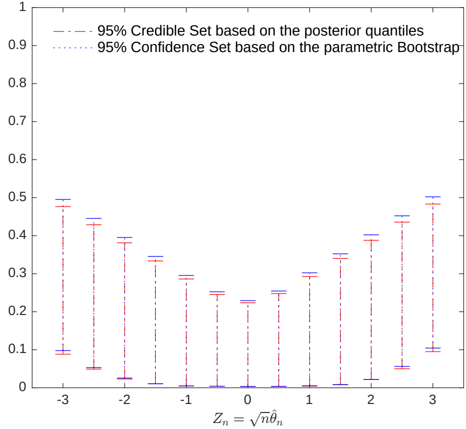
$$\left(\left|\sqrt{\frac{n}{n+\lambda^2}} Z^* + \frac{n}{n+\lambda^2}\sqrt{n}\widehat{\theta}_n\right| - \left|\sqrt{n}\widehat{\theta}_n\right|\right), \quad Z^* \sim \mathcal{N}(0,1) \perp X^n$$

⋆ The centered and scaled distribution of $\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n))$:

$$\left(\left|Z^{**} + \sqrt{n}\widehat{\theta}_n\right| - \left|\sqrt{n}\widehat{\theta}_n\right|\right), \quad Z^{**} \sim \mathcal{N}(0,1) \perp X^n$$

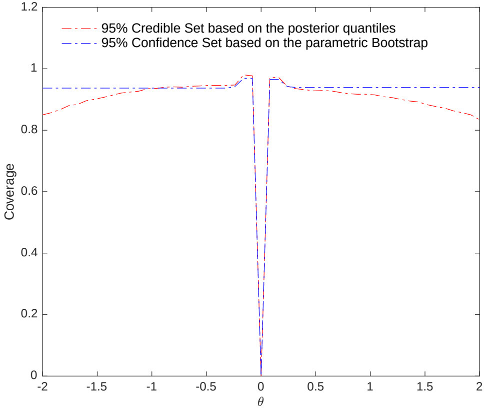
⋆ And we show that the 'distance' between these two r.v.'s $\to 0$.
(as the sample size goes to infinity)

Quantiles of $|\theta|$: $N(0, 1/5)$ Prior & Bootstrap



95% Credible set vs. 95% Bootstrap Confidence Set
($n = 100$, 10000 draws of Z^* and Z^{**})

Coverage of $|\theta|$: $N(0, 1/5)$ Prior & Bootstrap



95% Credible set and 95% Bootstrap Confidence Set
(n=100, 1000 data sets for each θ)

2. Main Result: Asymptotic Behavior of the Posterior of $g(\theta)$

Bounded Lipschitz Distance

⋆ Let ϕ_n^* and ψ_n^* be RVs whose distribution depends on X^n.

Bounded Lipschitz Distance

- Let ϕ_n^* and ψ_n^* be RVs whose distribution depends on X^n.
- Define the set:

$$\text{BL}(1) \equiv \Big\{ f : \mathbb{R}^p \to \mathbb{R} \mid \sup_{a \in \mathbb{R}^k} |f(a)| \leq 1 \text{ and } |f(a_1) - f(a_2)| \leq ||a_1 - a_2|| \quad \forall a_1, a_2 \Big\}.$$

Bounded Lipschitz Distance

- Let ϕ_n^* and ψ_n^* be RVs whose distribution depends on X^n.
- Define the set:

$$\mathrm{BL}(1) \equiv \Big\{ f : \mathbb{R}^p \to \mathbb{R} \mid \sup_{a \in \mathbb{R}^k} |f(a)| \leq 1 \text{ and } |f(a_1) - f(a_2)| \leq ||a_1 - a_2|| \quad \forall a_1,\ a_2 \Big\}.$$

- The Bounded Lipschitz distance between the distributions induced by ϕ_n^* and ψ_n^* (conditional on the data X^n) is:

$$\beta(\phi_n^*, \psi_n^*;\ X^n) \equiv \sup_{f \in BL(1)} \Big| \mathbb{E}[f(\phi_n^*)|X^n] - \mathbb{E}[f(\psi_n^*)|X^n] \Big|.$$

Bounded Lipschitz Distance

- ⋆ Let ϕ_n^* and ψ_n^* be RVs whose distribution depends on X^n.
- ⋆ Define the set:

$$\text{BL}(1) \equiv \Big\{ f : \mathbb{R}^p \to \mathbb{R} \mid \sup_{a \in \mathbb{R}^k} |f(a)| \leq 1 \text{ and } |f(a_1) - f(a_2)| \leq ||a_1 - a_2|| \quad \forall a_1,\, a_2 \Big\}.$$

- ⋆ The Bounded Lipschitz distance between the distributions induced by ϕ_n^* and ψ_n^* (conditional on the data X^n) is:

$$\beta(\phi_n^*, \psi_n^*;\, X^n) \equiv \sup_{f \in BL(1)} \Big| \mathbb{E}[f(\phi_n^*)|X^n] - \mathbb{E}[f(\psi_n^*)|X^n] \Big|.$$

- ⋆ ϕ_n^* and ψ_n^* converge in Bounded Lipschitz distance in probability if:

$$\beta(\phi_n^*, \psi_n^*;\, X^n) \xrightarrow{p} 0, \qquad \text{as } n \to \infty$$

Assumptions

A1: $g : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous.

Assumptions

A1: $g : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous.

A2: $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \mathcal{I}^{-1}(\theta))$.

Assumptions

A1: $g : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous.

A2: $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \mathcal{I}^{-1}(\theta))$.

A3: The centered and scaled random variables:

$$Z_n^{P*} = \sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n) \Big| X^n$$

and

$$Z_n^{B*} = \sqrt{n}(\theta_n^{B*} - \widehat{\theta}_n) \Big| X^n$$

converge (in the Bounded Lipschitz distance in probability) to the asymptotic distribution of the ML estimator:

$$Z \sim N(0, \mathcal{I}^{-1}(\theta_0))$$

Assumptions

A1: $g : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous.

A2: $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \mathcal{I}^{-1}(\theta))$.

A3: The centered and scaled random variables:

$$Z_n^{P*} = \sqrt{n}\Big(\frac{1}{\sqrt{n+\lambda^2}} Z + \frac{n}{n+\lambda^2}\widehat{\theta}_n - \widehat{\theta}_n\Big) \,\Big|\, X^n$$

and

$$Z_n^{B*} = \sqrt{n}\Big(\frac{1}{\sqrt{n}} Z + \widehat{\theta}_n - \widehat{\theta}_n\Big) \,\Big|\, X^n$$

converge (in the Bounded Lipschitz distance in probability) to the asymptotic distribution of the ML estimator:

$$Z \sim N(0, \mathcal{I}^{-1}(\theta_0))$$

Assumptions

A1: $g : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous.

A2: $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \mathcal{I}^{-1}(\theta))$.

A3: The centered and scaled random variables:

$$Z_n^{P*} = \sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n) \Big| X^n$$

and

$$Z_n^{B*} = \sqrt{n}(\theta_n^{B*} - \widehat{\theta}_n) \Big| X^n$$

converge (in the Bounded Lipschitz distance in probability) to the asymptotic distribution of the ML estimator:

$$Z \sim N(0, \mathcal{I}^{-1}(\theta_0))$$

(Bernstein Von-Mises Theorem + Bootstrap Consistency)

Assumptions

A1: $g : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous.

A2: $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} Z \sim N(0, \mathcal{I}^{-1}(\theta))$.

A3: The centered and scaled random variables:

$$Z_n^{P*} = \sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n) \Big| X^n$$

and

$$Z_n^{B*} = \sqrt{n}(\theta_n^{B*} - \widehat{\theta}_n) \Big| X^n$$

converge (in the Bounded Lipschitz distance in probability) to the asymptotic distribution of the ML estimator:

$$Z \sim N(0, \mathcal{I}^{-1}(\theta_0))$$

(Bernstein von-Mises Theorem + Bootstrap Consistency)

Main Result:

If Assumptions 1 to 3 hold, then:

The Bounded Lipschitz distance between

$$\sqrt{n}(g(\theta_n^{P*}) - g(\widehat{\theta}_n)) \Big| X^n$$

and

$$\sqrt{n}(g(\theta_n^{B*}) - g(\widehat{\theta}_n)) \Big| X^n$$

goes in probability to zero.

PROOF (SKETCH):

- ⋆ Posterior and bootstrap distributions can be written as: (after centering and scaling)

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)),$$
$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

PROOF (SKETCH):

- ⋆ Posterior and bootstrap distributions can be written as:
  (after centering and scaling)

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)),$$

$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

- ⋆ Recall that:

PROOF (SKETCH):

- ⋆ Posterior and bootstrap distributions can be written as: (after centering and scaling)

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)),$$
$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

- ⋆ Recall that:
  - ⋆ Z_n^{P*} and Z_n^{B*} both converge to a common limit Z

Proof (Sketch):

- ⋆ Posterior and bootstrap distributions can be written as: (after centering and scaling)

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)),$$
$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

- ⋆ Recall that:
  - ⋆ Z_n^{P*} and Z_n^{B*} both converge to a common limit Z
  - ⋆ g is Lipschitz continuous

Proof (Sketch):

- Posterior and bootstrap distributions can be written as:
  (after centering and scaling)

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)),$$
$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

- Recall that:
  - Z_n^{P*} and Z_n^{B*} both converge to a common limit Z
  - g is Lipschitz continuous
- We show that both distributions converge to a common limit:

$$\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

Proof (Sketch):

- Posterior and bootstrap distributions can be written as: (after centering and scaling)

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)),$$
$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) = \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

- Recall that:
  - Z_n^{P*} and Z_n^{B*} both converge to a common limit Z
  - g is Lipschitz continuous
- We show that both distributions converge to a common limit:

$$\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

- To do so, we use the Lipschtz property of:

$$\Delta_n(a) \equiv \sqrt{n}(g(\theta_0 + a/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

3. Corollary 1: Directional Differentiability

Corollary 1

- ⋆ **A4:** g is Directionally Differentiable.

Corollary 1

- ⋆ **A4:** g is Directionally Differentiable.
- ⋆ If Assumptions 1 to 4 hold, both random variables:

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) \mid X^n$$

and

$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) \mid X^n$$

converge (in the BL metric in probability) to:

$$g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n) \mid X^n$$

where $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta)$ and $Z = N(0, \mathcal{I}^{-1}(\theta))$.

Corollary 1

- ⋆ **A4:** g is Directionally Differentiable.
- ⋆ If Assumptions 1 to 4 hold, both random variables:

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) \mid X^n$$

and

$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) \mid X^n$$

converge (in the BL metric in probability) to:

$$g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n) \mid X^n$$

where $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta)$ and $Z = N(0, \mathcal{I}^{-1}(\theta))$.

- ⋆ So, for the case $g(\theta) = |\theta|$, the limit is:

$$|Z + Z_n| - |Z_n| \mid X^n$$

Corollary 1

- ⋆ **A4:** g is Directionally Differentiable.
- ⋆ If Assumptions 1 to 4 hold, both random variables:

$$\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) \mid X^n$$

and

$$\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) \mid X^n$$

converge (in the BL metric in probability) to:

$$g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n) \mid X^n$$

where $Z_n = \sqrt{n}(\widehat{\theta}_n - \theta)$ and $Z = N(0, \mathcal{I}^{-1}(\theta))$.

- ⋆ So, for the case $g(\theta) = |\theta|$, the limit is:

$$|Z + Z_n| - |Z_n| \mid X^n$$

- ⋆ Posterior of $g(\theta)$ need not be $\mathcal{N}$, even when posterior for θ is.

Proof (Sketch)

⋆ Because g is Lipschitz we have the approximation:

$$\begin{aligned}
&\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \\
&\qquad = \sqrt{n}(g(\theta_0 + Z_n^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n})) \\
&\qquad \approx \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n}))
\end{aligned}$$

Proof (Sketch)

- Because g is Lipschitz we have the approximation:

$$\begin{aligned}&\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \\ &\quad = \sqrt{n}(g(\theta_0 + Z_n^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n})) \\ &\quad \approx \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n}))\end{aligned}$$

- We can rewrite this as perturbations around θ_0:

$$\begin{aligned}&\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n})) \\ &\quad = \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0)) \\ &\quad\quad - \sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0))\end{aligned}$$

Proof (Sketch)

- Because g is Lipschitz we have the approximation:

$$\begin{aligned}&\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \\ &\qquad = \sqrt{n}(g(\theta_0 + Z_n^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n})) \\ &\qquad \approx \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n}))\end{aligned}$$

- We can rewrite this as perturbations around θ_0:

$$\begin{aligned}&\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n})) \\ &\qquad = \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0)) \\ &\qquad\quad - \sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0))\end{aligned}$$

- Since g is directionally differentiable, for large n:

$$\begin{aligned}&\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0)) - \sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0)) \\ &\qquad \approx g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)\end{aligned}$$

Proof (Sketch)

- ⋆ Because g is Lipschitz we have the approximation:

$$\begin{aligned}\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))& \\ &= \sqrt{n}(g(\theta_0 + Z_n^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n})) \\ &\approx \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n}))\end{aligned}$$

- ⋆ We can rewrite this as perturbations around θ_0:

$$\begin{aligned}\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + Z_n/\sqrt{n}))& \\ &= \sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0)) \\ &\quad - \sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0))\end{aligned}$$

- ⋆ Since g is directionally differentiable, for large n:

$$\begin{aligned}\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0)) - \sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0))& \\ \approx g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)&\end{aligned}$$

- ⋆ Same idea as in Santos & Fang [16].

4. Corollary 2: Credible sets for $g(\theta)$

Confidence and Credible sets

⋆ Let $q^*_\alpha(X^n)$ denote the quantile: (for $* \in \{B, P\}$)

$$q^*_\alpha(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(g(\theta^*_n) \leq c \mid X^n) \geq \alpha\}$$

Confidence and Credible sets

⋆ Let $q^*_\alpha(X^n)$ denote the quantile: (for $* \in \{B, P\}$)

$$q^*_\alpha(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(g(\theta^*_n) \leq c \mid X^n) \geq \alpha\}$$

⋆ $(1-\alpha)$% two-sided confidence set for $g(\theta)$:

$$CS^B_n(1-\alpha) \equiv \left[q^B_{\alpha/2}(X^n)\,,\; q^B_{1-\alpha/2}(X^n)\right]$$

Confidence and Credible sets

- ⋆ Let $q_\alpha^*(X^n)$ denote the quantile: (for $* \in \{B, P\}$)

$$q_\alpha^*(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(g(\theta_n^*) \leq c \mid X^n) \geq \alpha\}$$

- ⋆ $(1-\alpha)\%$ two-sided confidence set for $g(\theta)$:

$$CS_n^B(1-\alpha) \equiv \left[q_{\alpha/2}^B(X^n)\,,\; q_{1-\alpha/2}^B(X^n)\right]$$

- ⋆ $(1-\alpha)\%$ quantile based two-sided credible set for $g(\theta)$:

$$CS_n^P(1-\alpha) \equiv \left[q_{\alpha/2}^P(X^n)\,,\; q_{1-\alpha/2}^P(X^n)\right]$$

Confidence and Credible sets

⋆ Let $q^*_\alpha(X^n)$ denote the quantile: (for $* \in \{B, P\}$)

$$q^*_\alpha(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(g(\theta^*_n) \leq c \mid X^n) \geq \alpha\}$$

⋆ $(1-\alpha)\%$ two-sided confidence set for $g(\theta)$:

$$CS^B_n(1-\alpha) \equiv \left[q^B_{\alpha/2}(X^n)\,,\; q^B_{1-\alpha/2}(X^n)\right]$$

⋆ $(1-\alpha)\%$ quantile based two-sided credible set for $g(\theta)$:

$$CS^P_n(1-\alpha) \equiv \left[q^P_{\alpha/2}(X^n)\,,\; q^P_{1-\alpha/2}(X^n)\right]$$

⋆ $CS^*_n(1-\alpha)$ fails to cover the parameter $g(\theta)$ at θ by at least $d_\alpha\%$ ($d_\alpha > 0$) if:

$$\limsup_{n\to\infty} \mathbb{P}_\theta\left(g(\theta) \in CS^*_n(1-\alpha)\right) \leq 1-\alpha-d_\alpha, \qquad (1)$$

where $\mathbb{P}_\theta$ refers to the distribution of X_i under parameter θ.

Additional Assumption

⋆ Let $F_\theta(y|Z_n)$ denote the c.d.f. of the random variable

$$Y \equiv g_\theta'(Z + Z_n) \mid Z_n$$

Additional Assumption

⋆ Let $F_\theta(y|Z_n)$ denote the c.d.f. of the random variable

$$Y \equiv g'_\theta(Z + Z_n) \mid Z_n$$

⋆ **A5:** The c.d.f. $F_\theta(y|Z_n)$ is Lipschitz continuous with a constant k that does not depend on Z_n

Back to the Example

⋆ The p.d.f. of $Y \equiv g_0'(Z + Z_n) = |Z + Z_n|$ is:

$$h_0(y|Z_n) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y - Z_n)^2\Big) + \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y + Z_n)^2\Big)$$

$$\leq \sqrt{\frac{2}{\pi}}$$

Back to the Example

⋆ The p.d.f. of $Y \equiv g_0'(Z + Z_n) = |Z + Z_n|$ is:

$$h_0(y|Z_n) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y - Z_n)^2\Big) + \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y + Z_n)^2\Big) \leq \sqrt{\frac{2}{\pi}}$$

⋆ This implies that for $y_1 \geq y_2$:

$$F_\theta(y_1|Z_n) - F_\theta(y_2|Z_n) = \int_{y_2}^{y_1} h(y|Z_n)dy \leq (y_1 - y_2)\sqrt{\frac{2}{\pi}}$$

Back to the Example

⋆ The p.d.f. of $Y \equiv g_0'(Z + Z_n) = |Z + Z_n|$ is:

$$\begin{aligned} h_0(y|Z_n) &= \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y - Z_n)^2\Big) \\ &\quad + \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(y + Z_n)^2\Big) \\ &\leq \sqrt{\frac{2}{\pi}} \end{aligned}$$

⋆ This implies that for $y_1 \geq y_2$:

$$F_\theta(y_1|Z_n) - F_\theta(y_2|Z_n) = \int_{y_2}^{y_1} h(y|Z_n)dy \leq (y_1 - y_2)\sqrt{\frac{2}{\pi}}$$

⋆ An analogous argument works for $y_1 < y_2$ and **A5.** holds

Additional Corollary

⋆ Suppose Assumptions 1 – 5 hold

Additional Corollary

- ⋆ Suppose Assumptions $1 - 5$ hold
- ⋆ Suppose that the nominal $(1 - \alpha)\%$ bootstrap confidence set fails to cover $g(\theta)$ at θ by at least $d_\alpha\%$

Additional Corollary

- ⋆ Suppose Assumptions 1 − 5 hold
- ⋆ Suppose that the nominal $(1-\alpha)\%$ bootstrap confidence set fails to cover $g(\theta)$ at θ by at least $d_\alpha\%$
- ⋆ Then, for any $0 < \epsilon < d_\alpha$,

$$\limsup_{n\to\infty} \mathbb{P}_\theta \left(g(\theta) \in CS_n^P(1-(\alpha+\epsilon)) \right) \leq 1-\alpha-d_\alpha$$

(i.e. the credible set:

$$CS_n^P(1-(\alpha+\epsilon))\% = \left[q^P_{(\alpha+\epsilon)/2}(X^n)\,,\; q^P_{1-(\alpha+\epsilon)/2}(X^n) \right]$$

fails to cover $g(\theta)$ at θ by at least $(d_\alpha - \epsilon)\%$)

Conclusion

⋆ We studied the posterior distribution of $g(\theta)$, $\theta \in \mathbb{R}^p$

Conclusion

- ⋆ We studied the posterior distribution of $g(\theta)$, $\theta \in \mathbb{R}^p$
- ⋆ Main result: Posterior distribution $\sim$ Bootstrap distribution

Conclusion

- $\star$ We studied the posterior distribution of $g(\theta)$, $\theta \in \mathbb{R}^p$
- $\star$ Main result: Posterior distribution $\sim$ Bootstrap distribution
- $\star$ **I1:** Bayesians can use bootstrap for posterior inference

Conclusion

- ⋆ We studied the posterior distribution of $g(\theta)$, $\theta \in \mathbb{R}^p$
- ⋆ Main result: Posterior distribution $\sim$ Bootstrap distribution
- ⋆ **I1:** Bayesians can use bootstrap for posterior inference
- ⋆ **I2:** Bootstrap c.set fails for $g(\theta) \implies$ Posterior fails as well

Conclusion

- ⋆ We studied the posterior distribution of $g(\theta)$, $\theta \in \mathbb{R}^p$
- ⋆ Main result: Posterior distribution $\sim$ Bootstrap distribution
- ⋆ **I1:** Bayesians can use bootstrap for posterior inference
- ⋆ **I2:** Bootstrap c.set fails for $g(\theta) \implies$ Posterior fails as well
- ⋆ We hope this note gives a better idea of the limits of Bayesian inference when used for frequentist anaysis!

THANK YOU