Saddlepoint-Based Bootstrap Inference in Dependent Data Settings

Alex Trindade

Dept. of Mathematics & Statistics, *Texas Tech University*

Rob Paige, *Missouri University of Science and Technology*

Indika Wickramasinghe, *Eastern New Mexico University* (former student)

Pratheepa Jeganathan, *Texas Tech University* (current student)

June 2014

Outline

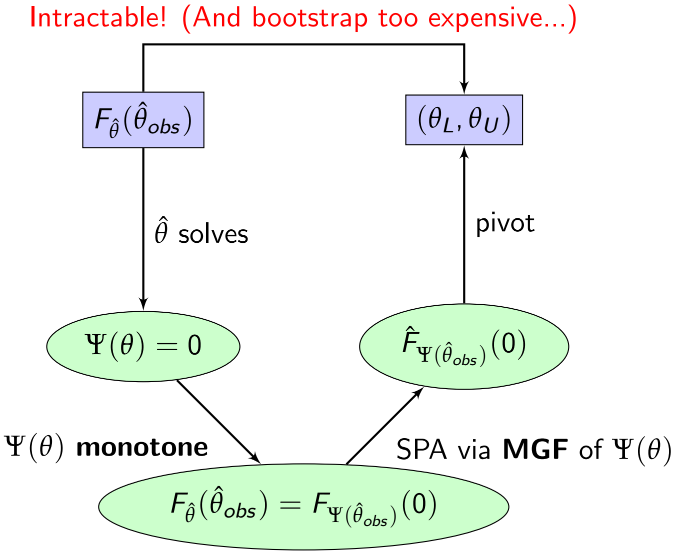
1. Overview of SPBB inference: Saddlepoint-Based Bootstrap
    - An approximate parametric bootstrap for scalar parameter θ
2. Application 1: Spatial Regression Models
    - Classic application of SPBB
    - Better performance than asymptotic-based CIs
3. Application 2: MA(1) Model
    - Challenging; extend methodology in two directions
    - Extension 1: Non-Monotone QEEs
    - Extension 2: Non-Gaussian QEEs

Saddlepoint-Based Bootstrap (SPBB) Inference

Pioneered by Paige, Trindade, & Fernando (SJS, 2009):

- SPBB: an approximate percentile parametric bootstrap;
- replace (slow) MC simulation with (fast) saddlepoint approx (SPA);
- estimators are roots of QEE (*quadratic estimating equation*);
- enjoys near exact performance;
- orders of magnitude faster than bootstrap;
- may be only alternative to bootstrap if no exact or asymptotic procedures;
- Idea:
  - relate distribution of root of QEE $\Psi(\theta)$ to that of estimator $\hat{\theta}$;
  - under normality on data have closed form for MGF of QEE;
  - use to saddlepoint approximate distribution of estimator (PDF or CDF);
  - can pivot CDF to get a CI... numerically!
  - leads to 2nd order accurate CIs, coverage error is $O(n^{-1})$.

SPBB: An Approximate Parametric Bootstrap



Application 1: Spatial Regression Models (Lattice Data)

Jeganathan, Paige, & Trindade (under review)

- Spatial process $\mathbf{y} = [y(s_1), \ldots, y(s_n)]^\intercal$ observed at sites $\{s_1, \ldots, s_n\}$.
- Under stationarity & isotropy, correlation modeled via spatial dependence parameter ρ and spatial weights matrix W.
- 3 main correlation structures for the regression model

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{z}, \qquad \mathbf{z} \sim \mathrm{N}(\mathbf{0}, \sigma^2 V_\rho)$$

  - SAR: $\mathbf{z} = \rho W \mathbf{z} + \boldsymbol{\varepsilon}$, with $V_\rho = (I_n - \rho W)^{-1}(I_n - \rho W^\intercal)^{-1}$
  - CAR: $\mathbb{E}[z(s_i) | z(s_j) : s_j \in N(s_i)] = \rho \sum w_{ij} z(s_j)$, with

$$V_\rho = (I_n - \rho W)^{-1}$$

  - SMA: $\mathbf{z} = \rho W \boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}$, with $V_\rho = (I_n + \rho W)(I_n + \rho W^\intercal)$.

QEEs for ML and REML Estimators of ρ

- IRWGLS estimate for $\mathbf{z}$ (fixed ρ): $\hat{\mathbf{z}} \equiv \mathbf{r} = P_{C(X)^\perp}\mathbf{y}$, with

$$\mathbf{y} \sim \mathrm{N}(X\beta, \sigma^2 V_\rho) \quad \Longrightarrow \quad \mathbf{r} \sim \mathrm{N}(\mathbf{0}, \sigma^2 V_\rho P^T_{C(X)^\perp})$$

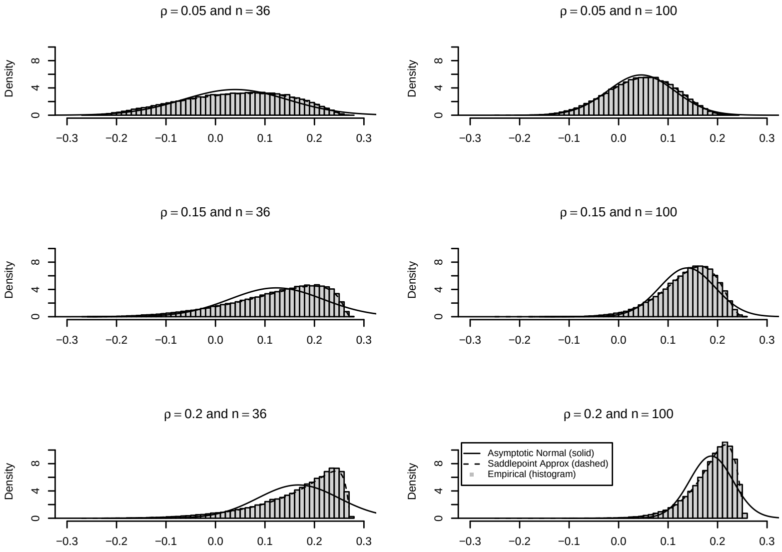
- Leads to following QEEs for estimators $\hat{\rho}_{ML}$ & $\hat{\rho}_{REML}$:

$$\Psi_{ML}(\rho) = \mathbf{r}^\mathsf{T} \left[\mathrm{Tr}\left(V_\rho^{-1}\dot{V}_\rho\right) V_\rho^{-1} - nV_\rho^{-1}\dot{V}_\rho V_\rho^{-1}\right] \mathbf{r}$$

$$\Psi_{REML}(\rho) = \mathbf{r}^T \left[\mathrm{Tr}\left(V_\rho^{-1}\dot{V}_\rho\right) V_\rho^{-1} - \mathrm{Tr}\left(P_{C(X)}\dot{V}_\rho V_\rho^{-1}\right) V_\rho^{-1} - (n-q)\, V_\rho^{-1}\dot{V}_\rho V_\rho^{-1}\right] \mathbf{r}$$

- **Theorem:** ML QEE is biased; REML QEE is unbiased.

Approximations for Distribution of $\hat{\rho}_{ML}$: CAR Model



Empirical Coverage Probabilities and Average Lengths of 95% SPBB & Asymptotic CIs for $\hat{\rho}_{ML}$: CAR Model

Coverages

Sample size	ρ_0=0.05		ρ_0=0.15		ρ_0=0.2	
	SPBB	ASYM	**SPBB**	ASYM	**SPBB**	ASYM
n=16	**0.941**	0.876	**0.934**	0.895	**0.928**	0.866
n=36	**0.959**	0.907	**0.941**	0.908	**0.906**	0.895
n=100	**0.946**	0.925	**0.950**	0.932	**0.946**	0.931

Lengths

Sample size	ρ_0=0.05		ρ_0=0.15		ρ_0=0.2	
	SPBB	ASYM	**SPBB**	ASYM	**SPBB**	ASYM
n=16	**0.480**	0.575	**0.514**	0.628	**0.440**	0.517
n=36	**0.402**	0.443	**0.346**	0.378	**0.286**	0.311
n=100	**0.263**	0.272	**0.214**	0.221	**0.159**	0.165

Real Dataset 1: Mercer & Hall (1911) Wheat Yield

- Yield of grain collected in summer of 1910 over $n = 500$ plots (20×25 grid).
- Mean trend removed via median polish.
- Available in R package `spdep` as "wheat".
- Cressie (1985, 1993): use spatial binary weights matrix W with polynomial neighborhood structure in SAR model.
- **Table:** MLEs and 95% SPBB & ASYM CIs for ρ_0.

Estimation Method	SAR model	CAR model	SMA model
MLE	0.603	0.078	0.077
SPBB 95% CI	**(0.477, 0.726)**	**(0.067, 0.084)**	**(0.055, 0.098)**
ASYM 95% CI	(0.478, 0.727)	(0.069, 0.088)	(0.055, 0.098)

Real Dataset 2: Eire county blood group A percentages

- Available in R package `spdep` as "eire".
- Percentage of a sample with blood type A collected over $n = 26$ counties in Eire.
- Covariates: *towns* (towns/unit area) and *pale* (1=within, 0=beyond).
- Used by Cliff & Ord (1973) to illustrate spatial dependence in SAR and CAR models (binary weights matrix W with neighborhood structure as in `spdep`).

Estimation Method	SAR model	CAR model	SMA model
MLE	0.313	0.078	0.055
SPBB 95% CI	**(-0.190, 0.769)**	**(-0.167, 0.195)**	**(-0.078, 0.209)**
ASYM 95% CI	$(-0.151,\ 0.776)$	$(-0.128,\ 0.283)$	$(-0.083,\ 0.192)$

Application 2: MA(1) Model

Paige, Trindade, & Wickramasinghe (AISM, 2014)

- **Challenging**...; extend methodology in two directions.
- Extension 1: Non-Monotone QEEs
  - Problem: non-monotone QEEs invalidate SPBB
  - Solution: double-SPA & importance sampling
- Extension 2: Non-Gaussian QEEs
  - Problem: key to SPBB is QEE with tractable MGF
  - Solution: elliptically contoured distributions, and some tricks

The MA(1): World's Simplest Model?

- **Model:**

$$X_t = \theta_0 Z_{t-1} + Z_t, \qquad Z_t \sim \text{iid } (0, \sigma^2), \qquad |\theta_0| \leq 1$$

- **Uses:**
  - special case of more general ARMA models;
  - perhaps most useful in testing if data has been over-differenced... if we difference WN we get MA(1) with $\theta_0 = -1$

$$X_t = Z_t \qquad \Longrightarrow \qquad Y_t \equiv X_t - X_{t-1} = Z_t - Z_{t-1}$$

  - connection with unit-root tests in econometrics (Tanaka, 1990, Davis *et al.*, 1995, Davis & Dunsmuir, 1996, Davis & Song, 2011).
- **Inference:** complicated...
  - common estimators (MOME, LSE, MLE) have mixed distributions, point masses at ± 1 and continuous over $(-1, 1)$;
  - LSE & MLE are roots of polynomials of degree $\approx 2n$.

Unification of Parameter Estimators

Theorem

For $|\theta| < 1$, MOME, LSE, and MLE are all roots of QEE, $\Psi(\theta) = \mathbf{x}^\intercal A_\theta \mathbf{x}$, where symmetric matrix A_θ in each case is

- *MOME: (QEE is monotone)*

$$A_\theta = (1 + \theta^2) J_n - 2\theta I_n$$

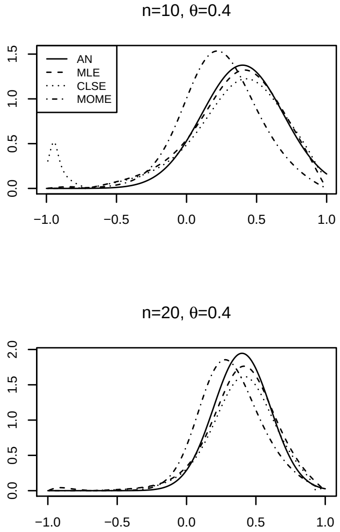
- *LSE: (QEE not monotone...)*

$$A_\theta = \Omega_\theta^{-1} [J_n + 2\theta I_n] \Omega_\theta^{-1}$$

- *MLE: (QEE not monotone...)*

$$A_\theta = function(\theta, I_n, J_n, \Omega_\theta^{-1})$$

SPA densities of estimators: MOME, LSE, MLE, AN



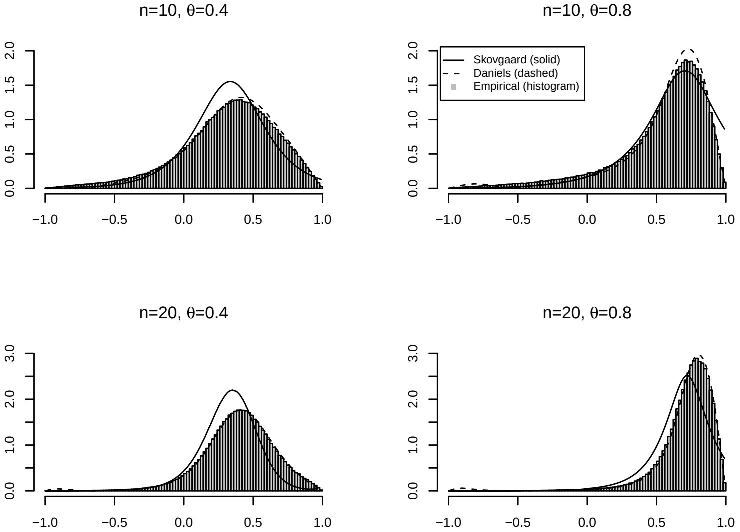
95% CI Coverages & Lengths for MOME (Gaussian Noise)

Settings		Coverage Probability			Average Length		
n	θ_0	**SPBB**	Boot	AN	**SPBB**	Boot	AN
10	0.4	**0.940**	0.432	0.997	**1.484**	1.438	0.561
10	0.8	**0.948**	0.358	0.259	**1.336**	1.653	1.300
20	0.4	**0.953**	0.717	1.000	**1.095**	1.560	0.334
20	0.8	**0.960**	0.524	0.693	**1.005**	1.692	1.616

Extension 1: Non-Monotone Estimating Equations

- Monotonicity of QEE is key (Daniels, 1983).
- Skovgaard (1990) & Spady (1991) give expression for PDF of $\hat{\theta}$ where Jacobian does not require monotonicity of $\Psi(t)$ in t; but involves an intractable conditional expectation...
- Solution: double-SPA & importance sampling.

Example: Density of MLE (Gaussian Noise)



Extension 2: Non-Gaussian QEEs

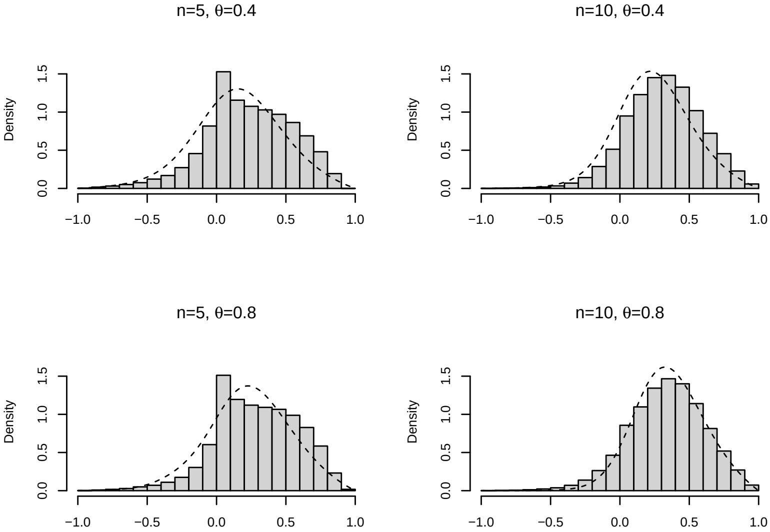
- General Problem with SPBB: need QEEs with tractable MGF...
- One solution: elliptically contoured (EC) distributions.
- Relies on appropriate weighting function $w(t)$ (Provost & Cheong, 2002).

Theorem

With M_N the MVN MGF:

$$M_{EC}(s; \boldsymbol{\mu}, \Sigma) = \int_0^\infty w(t)\, M_N(s; \boldsymbol{\mu}, \Sigma / t)\, dt$$

SPBB approx PDFs of MOME in Laplace MA(1)



Key References

- Butler, R.W. (2007), *Saddlepoint Approximations With Applications*, New York: Cambridge University Press.
- Paige, R.L. and Trindade, A.A. (2008), "Practical Small Sample Inference for Single Lag Subset Autoregressive Models", *J. Statist. Plan. Inf.*, 138, 1934–1949.
- Paige, R.L., Trindade, A.A. and Fernando, P.H. (2009), "Saddlepoint-based bootstrap inference for quadratic estimating equations", *Scand. J. Stat.*, 36, 98–111.
- Paige, R.L., Trindade, A.A., and Wickramasinghe, R.I.P., "Extensions of Saddlepoint-Based Bootstrap Inference", *Ann. Instit. Statist. Math.*, to appear.
- Jeganathan, P., Paige, R.L. and Trindade, A.A., "Saddlepoint-Based Bootstrap Inference for Spatial Dependence in the Lattice Process", under review.

SPBB Details: Key Steps

- Estimator $\hat{\theta}$ of θ_0 solves QEE

$$\Psi(\theta) = \mathbf{x}^\mathsf{T} A_\theta \mathbf{x} = 0$$

- Assume: $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma) \implies$ closed-form for MGF of QEE.
- QEE monotone (e.g., decreasing) in θ implies:

$$F_{\hat{\theta}}(t) = P(\hat{\theta} \leq t) = P\left(\Psi(t) \leq 0\right) = F_{\Psi(t)}(0)$$

- Nuisance parameter $\boldsymbol{\lambda}$: substitute conditional MLE, $\hat{\boldsymbol{\lambda}}_\theta$.
- Now: accurately approximate distribution of $\hat{\theta}$ via SPA

$$F_{\hat{\theta}}(t; \theta_0, \boldsymbol{\lambda}_0) \approx \hat{F}_{\hat{\theta}}\left(t; \theta_0, \hat{\boldsymbol{\lambda}}_{\theta_0}\right) = \hat{F}_{\Psi(t)}\left(0; \theta_0, \hat{\boldsymbol{\lambda}}_{\theta_0}\right)$$

- CI (θ_L, θ_U) produced by pivoting SPA of CDF

$$\hat{F}_{\Psi(\hat{\theta}_{\text{obs}})}(0; \theta_L, \hat{\boldsymbol{\lambda}}_{\theta_L}) = 1 - \frac{\alpha}{2}, \qquad \hat{F}_{\Psi(\hat{\theta}_{\text{obs}})}(0; \theta_U, \hat{\boldsymbol{\lambda}}_{\theta_U}) = \frac{\alpha}{2}$$