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 $\theta$

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
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

\[ F_\hat{\theta}(\hat{\theta}_{obs}) \]

\( \hat{\theta} \) solves

\[ \Psi(\theta) = 0 \]

\( \Psi(\theta) \) monotone

\[ F_\hat{\theta}(\hat{\theta}_{obs}) = F_\Psi(\hat{\theta}_{obs})(0) \]

\( \hat{\theta}_{obs} \) solves

\[ \hat{\theta}_{obs} \]

(\( \theta_L, \theta_U \))

pivot

SPA via MGF of \( \Psi(\theta) \)

Intractable! (And bootstrap too expensive...)

alex.trindade@ttu.edu

(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)
Spatial process \( y = [y(s_1), \ldots, y(s_n)]^\top \) observed at sites \( \{s_1, \ldots, s_n\} \).

Under stationarity & isotropy, correlation modeled via spatial dependence parameter \( \rho \) and spatial weights matrix \( W \).

3 main correlation structures for the regression model

\[
y = X\beta + z, \quad z \sim N(0, \sigma^2 V_\rho)
\]

- **SAR:** \( z = \rho W z + \epsilon \), with \( V_\rho = (I_n - \rho W)^{-1}(I_n - \rho W^\top)^{-1} \)
- **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:** \( z = \rho W \epsilon + \epsilon \), with \( V_\rho = (I_n + \rho W)(I_n + \rho W^\top) \).
QEEs for ML and REML Estimators of $\rho$

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

$$y \sim N(X\beta, \sigma^2 V_\rho) \implies r \sim N(0, \sigma^2 V_\rho P_{C(X)\perp}^T)$$

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

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

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

- **Theorem:** ML QEE is biased; REML QEE is unbiased.
Approximations for Distribution of $\hat{\rho}_{ML}$: CAR Model

$\rho = 0.05$ and $n = 36$

$\rho = 0.15$ and $n = 36$

$\rho = 0.2$ and $n = 36$

$\rho = 0.05$ and $n = 100$

$\rho = 0.15$ and $n = 100$

$\rho = 0.2$ and $n = 100$

Asymptotic Normal (solid)
Saddlepoint Approx (dashed)
Empirical (histogram)

Figure: Empirical distribution of $\hat{\rho}$ in the CAR model (histogram) overlaid with the asymptotic normal (solid line) and saddlepoint approximations (dashed line). The Monte Carlo approximation to the true density displayed in the histograms is based on $10^5$ simulated replicates.

alex.trindade@ttu.edu

SPBB Inference Under Dependence

June 2014 7 / 21
### Empirical Coverage Probabilities and Average Lengths of 95% SPBB & Asymptotic CIs for $\hat{\rho}_{ML}$: CAR Model

#### Coverages

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\rho_0=0.05$</th>
<th>$\rho_0=0.15$</th>
<th>$\rho_0=0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPBB</td>
<td>ASYM</td>
<td>SPBB</td>
</tr>
<tr>
<td>$n=16$</td>
<td>0.941</td>
<td>0.876</td>
<td>0.934</td>
</tr>
<tr>
<td>$n=36$</td>
<td>0.959</td>
<td>0.907</td>
<td>0.941</td>
</tr>
<tr>
<td>$n=100$</td>
<td>0.946</td>
<td>0.925</td>
<td>0.950</td>
</tr>
</tbody>
</table>

#### Lengths

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\rho_0=0.05$</th>
<th>$\rho_0=0.15$</th>
<th>$\rho_0=0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPBB</td>
<td>ASYM</td>
<td>SPBB</td>
</tr>
<tr>
<td>$n=16$</td>
<td>0.480</td>
<td>0.575</td>
<td>0.514</td>
</tr>
<tr>
<td>$n=36$</td>
<td>0.402</td>
<td>0.443</td>
<td>0.346</td>
</tr>
<tr>
<td>$n=100$</td>
<td>0.263</td>
<td>0.272</td>
<td>0.214</td>
</tr>
</tbody>
</table>
Yield of grain collected in summer of 1910 over \( n = 500 \) plots (20 \( \times \) 25 grid).

Mean trend removed via median polish.

Available in R package spdep as “wheat”.


**Table:** MLEs and 95% SPBB & ASYM CIs for \( \rho_0 \).

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>SAR model</th>
<th>CAR model</th>
<th>SMA model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.603</td>
<td>0.078</td>
<td>0.077</td>
</tr>
<tr>
<td><strong>SPBB 95% CI</strong></td>
<td>(0.477, 0.726)</td>
<td>(0.067, 0.084)</td>
<td>(0.055, 0.098)</td>
</tr>
<tr>
<td><strong>ASYM 95% CI</strong></td>
<td>(0.478, 0.727)</td>
<td>(0.069, 0.088)</td>
<td>(0.055, 0.098)</td>
</tr>
</tbody>
</table>
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`).

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>SAR model</th>
<th>CAR model</th>
<th>SMA model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.313</td>
<td>0.078</td>
<td>0.055</td>
</tr>
<tr>
<td><strong>SPBB 95% CI</strong></td>
<td>(-0.190, 0.769)</td>
<td>(-0.167, 0.195)</td>
<td>(-0.078, 0.209)</td>
</tr>
<tr>
<td><strong>ASYM 95% CI</strong></td>
<td>(-0.151, 0.776)</td>
<td>(-0.128, 0.283)</td>
<td>(-0.083, 0.192)</td>
</tr>
</tbody>
</table>
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:**
  \[ X_t = \theta_0 Z_{t-1} + Z_t, \quad Z_t \sim \text{iid } (0, \sigma^2), \quad |\theta_0| \leq 1 \]

- **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 \( \theta_0 = -1 \)
    \[ X_t = Z_t \implies Y_t = X_t - X_{t-1} = Z_t - Z_{t-1} \]

- **Inference:** complicated...
  - common estimators (MOME, LSE, MLE) have mixed distributions, point masses at \( \pm 1 \) and continuous over \((-1, 1)\);
  - LSE & MLE are roots of polynomials of degree \( \approx 2n \).
Theorem

For $|θ| < 1$, MOME, LSE, and MLE are all roots of QEE, $Ψ(θ) = x^T A_θ x$, where symmetric matrix $A_θ$ in each case is

- **MOME**: (QEE is monotone)
  
  $$A_θ = (1 + θ^2) J_n - 2θ I_n$$

- **LSE**: (QEE not monotone...)
  
  $$A_θ = Ω_θ^{-1} [J_n + 2θ I_n] Ω_θ^{-1}$$

- **MLE**: (QEE not monotone...)
  
  $$A_θ = \text{function}(θ, I_n, J_n, Ω_θ^{-1})$$
SPA densities of estimators: MOME, LSE, MLE, AN

- $n=10, \theta=0.4$
- $n=10, \theta=0.8$
- $n=20, \theta=0.4$
- $n=20, \theta=0.8$

alex.trindade@ttu.edu

(Dept. of Mathematics & Statistics, Texas Tech University, Rob Paige, Missouri University of Science and Technology, Indika Wickramasinghe, Eastern New Mexico University, Pratheepa Jeganathan, Texas Tech University, former student, current student)

SPBB Inference Under Dependence
June 2014 14 / 21
### 95% CI Coverages & Lengths for MOME (Gaussian Noise)

<table>
<thead>
<tr>
<th>Settings</th>
<th>Coverage Probability</th>
<th>Average Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPBB</td>
<td>Boot</td>
</tr>
<tr>
<td>n</td>
<td>θ₀</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.4</td>
<td>0.940</td>
</tr>
<tr>
<td>10</td>
<td>0.8</td>
<td>0.948</td>
</tr>
<tr>
<td>20</td>
<td>0.4</td>
<td>0.953</td>
</tr>
<tr>
<td>20</td>
<td>0.8</td>
<td>0.960</td>
</tr>
</tbody>
</table>
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)

- n=10, θ=0.4
- n=10, θ=0.8
- n=20, θ=0.4
- n=20, θ=0.8
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:*

\[
M_{EC} (s; \mu, \Sigma) = \int_0^\infty w(t) M_N (s; \mu, \Sigma/t) \, dt
\]
SPBB approx PDFs of MOME in Laplace MA(1)

$n=5, \theta=0.4$

$n=10, \theta=0.4$

$n=5, \theta=0.8$

$n=10, \theta=0.8$


SPBB Details: Key Steps

- Estimator $\hat{\theta}$ of $\theta_0$ solves QEE
  \[ \Psi(\theta) = x^T A_\theta x = 0 \]

- Assume: $x \sim N(\mu, \Sigma) \implies$ closed-form for MGF of QEE.

- QEE monotone (e.g., decreasing) in $\theta$ implies:
  \[ F_{\hat{\theta}}(t) = P(\hat{\theta} \leq t) = P(\Psi(t) \leq 0) = F_{\Psi(t)}(0) \]

- Nuisance parameter $\lambda$: substitute conditional MLE, $\hat{\lambda}_\theta$.

- Now: accurately approximate distribution of $\hat{\theta}$ via SPA
  \[ F_{\hat{\theta}}(t; \theta_0, \lambda_0) \approx \hat{F}_{\hat{\theta}}\left(t; \theta_0, \hat{\lambda}_{\theta_0}\right) = \hat{F}_{\Psi(t)}\left(0; \theta_0, \hat{\lambda}_{\theta_0}\right) \]

- CI ($\theta_L, \theta_U$) produced by pivoting SPA of CDF
  \[ \hat{F}_{\Psi(\hat{\theta}_{\text{obs}})}(0; \theta_L, \hat{\lambda}_{\theta_L}) = 1 - \frac{\alpha}{2}, \quad \hat{F}_{\Psi(\hat{\theta}_{\text{obs}})}(0; \theta_U, \hat{\lambda}_{\theta_U}) = \frac{\alpha}{2} \]