Distributional Approximation of Regression M-Estimator

Debraj Das

North Carolina State University

16^{th} November, 2016

Model

Multiple linear regression model :

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i, \qquad i = 1, 2, ..., n$$

where

- $y_1, \ldots, y_n$ are responses.
- $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are known non random design vectors.
- $\epsilon_1, \ldots, \epsilon_n$ are iid random variables.
- β is the $p \times 1$ vector of parameters (p is fixed).

M-Estimator

$\bar{\boldsymbol{\beta}}_n$ is the M-estimator of the parameter $\boldsymbol{\beta}$ corresponding to the objective function $\rho(\cdot)$ if

$$\bar{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{t}} \left[\sum_{i=1}^{n} \rho(y_i - \boldsymbol{x}_i' \boldsymbol{t}) \right]$$

Equivalently, if $\rho' = \psi$ then $\psi(\cdot)$ is the score function and $\bar{\boldsymbol{\beta}}_n$ is the solution of the vector equation

$$\sum_{i=1}^{n} \boldsymbol{x}_i \psi(y_i - \boldsymbol{x}_i' \boldsymbol{t}) = 0.$$

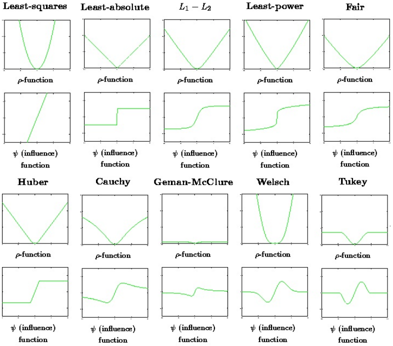
Why Useful?
To develop a unified theory at-least asymptotically.

Common Examples

- Least square estimator: $\rho(x) = x^2/2$ and $\psi(x) = x$.
- θth Quantile regression estimator: if $\mathbf{1}(\cdot)$ is the indicator function then $\rho(x) = (\theta - \mathbf{1}(x < 0))x$ and $\psi(x)$ $= (\theta - 1)\mathbf{1}(x < 0) + \theta\mathbf{1}(x > 0)$.
- LAD regression estimator: $\rho(x) = |x|$ and $\psi(x)$ $= sign(x) = 1, -1, 0$ according as $x > 0$, $x < 0$ and $x = 0$. Can be obtained by assuming $\theta = 1/2$ in the previous case.

A Compact List

type	$\rho(x)$	$\psi(x)$
L_2	$x^2/2$	x
L_1	$\|x\|$	$\operatorname{sgn}(x)$
$L_1 - L_2$	$2(\sqrt{1+x^2/2}-1)$	$\frac{x}{\sqrt{1+x^2/2}}$
L_p	$\frac{\|x\|^\nu}{\nu}$	$\operatorname{sgn}(x)\|x\|^{\nu-1}$
"Fair"	$c^2[\frac{\|x\|}{c} - \log(1+\frac{\|x\|}{c})]$	$\frac{x}{1+\|x\|/c}$
Huber $\begin{cases} \text{if } \|x\| \le k \\ \text{if } \|x\| \ge k \end{cases}$	$\begin{cases} x^2/2 \\ k(\|x\| - k/2) \end{cases}$	$\begin{cases} x \\ k\operatorname{sgn}(x) \end{cases}$
Cauchy	$\frac{c^2}{2}\log(1+(x/c)^2)$	$\frac{x}{1+(x/c)^2}$
Geman-McClure	$\frac{x^2/2}{1+x^2}$	$\frac{x}{(1+x^2)^2}$
Welsch	$\frac{c^2}{2}[1-\exp(-(x/c)^2)]$	$x\exp(-(x/c)^2)$
Tukey $\begin{cases} \text{if } \|x\| \le c \\ \text{if } \|x\| > c \end{cases}$	$\begin{cases} \frac{c^2}{6}(1-[1-(x/c)^2]^3) \\ (c^2/6) \end{cases}$	$\begin{cases} x[1-(x/c)^2]^2 \\ 0 \end{cases}$



Source: `http://research.microsoft.com/en-us/um/people/zhang/INRIA/Publis/Tutorial-Estim/node24.html`

Distributional Approximation Methods

A reasonable good approximation to the exact distribution of the M-estimator is necessary for the purpose of inference on the parameter β, eg.

- for finding confidence intervals.
- for testing hypotheses

Choices:

- Asymptotic Normality:
  Huber (1981)
- Residual Bootstrap:
  Freedman(1981), Lahiri(1992)
- Perturbation Bootstrap:
  Our proposed method

Asymptotic Normality (AN)

Suppose,

- $A_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$.
- $\sigma^2 = \mathbf{E}\psi^2(\epsilon_1)/\mathbf{E}^2\psi'(\epsilon_1)$ when ψ' exists.
- $\sigma^2 = \mathbf{E}\psi^2(\epsilon_1)/(\int \psi(x) f'(x) dx)^2$ when Lebesgue density of ϵ_1 and its derivative exists.

Result: $\left|\mathbf{P}\left(\sqrt{n}\sigma^{-1}\boldsymbol{A}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \in B\right) - \boldsymbol{\Phi}(B)\right| = O(n^{-1/2})$ in an uniform sense. Here B is a subset of $\mathcal{R}^p$.

Residual Bootstrap (RB)

- Suppose, $e_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n$ for $i \in \{1, \ldots, n\}$.
- Draw a random sample (with replacement) $\{e_1^*, \ldots, e_n^*\}$ from $\{e_1, \ldots, e_n\}$.
- Define, $y_i^* = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_n + e_i^*$ for $i \in \{1, \ldots, n\}$.
- RB estimator $\hat{\boldsymbol{\beta}}_n^R$ is defined as the solution of

$$\sum_{i=1}^{n} \boldsymbol{x}_i \Big(\psi(y_i^* - \boldsymbol{x}_i'\boldsymbol{t}) - n^{-1} \sum_{i=1}^{n} \psi(e_i) \Big) = \mathbf{0}$$

Result: Under some conditions on $\psi(\cdot)$, errors and design vectors,

$\Big| \mathbf{P}\Big(\mathbf{f_1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \in B \Big) - \mathbf{CP}\Big(\mathbf{f_2}(\hat{\boldsymbol{\beta}}_n^R - \hat{\boldsymbol{\beta}}_n) \in B \Big) \Big| = O_p(n^{-1})$ in an uniform sense. Here B is a subset of $\mathcal{R}^p$ and $\mathbf{f_1}(\cdot)$ & $\mathbf{f_2}(\cdot)$ are known functions.

Perturbation Bootstrap (PB)

- $\{G_1^*, \ldots, G_n^*\}$ is a iid sample from **Beta**$(1/2, 3/2)$.
- PB estimator $\hat{\boldsymbol{\beta}}_n^P$ is defined as the solution of

$$\sum_{i=1}^{n} \boldsymbol{x}_i \psi(y_i - \boldsymbol{x}_i' \boldsymbol{t}) G_i^* = \mathbf{0}$$

Result: Under some conditions on $\psi(\cdot)$, errors and design vectors,

$\left|\mathbf{P}\Big(\mathbf{f_3}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \in B\Big) - \mathbf{CP}\Big(\mathbf{f_4}(\hat{\boldsymbol{\beta}}_n^P - \hat{\boldsymbol{\beta}}_n) \in B\Big)\right| = O_p(n^{-1})$ in an uniform sense. Here B is a subset of $\mathcal{R}^p$ and $\mathbf{f_3}(\cdot)$ & $\mathbf{f_4}(\cdot)$ are known functions.

Another PB for Least Square ($\psi(x) = x$)

- $\{G_1^*, \ldots, G_n^*\}$ is a iid sample from **Beta**$(1/2, 3/2)$.
- Define, $z_i = \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}_n + 4e_i(G_i^* - 1/4)$ for $i \in \{1, \ldots, n\}$.
- PB estimator $\hat{\boldsymbol{\beta}}_n^P$ is defined as the solution of

$$\sum_{i=1}^{n} \boldsymbol{x}_i(z_i - \boldsymbol{x}_i' \boldsymbol{t}) = \mathbf{0}$$

Result: Under some conditions on $\psi(\cdot)$, errors and design vectors,

$\left| \mathbf{P}\Big(\mathbf{f_3}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \in B\Big) - \mathbf{CP}\Big(\mathbf{f_4}(\hat{\boldsymbol{\beta}}_n^P - \hat{\boldsymbol{\beta}}_n) \in B\Big) \right| = O_p(n^{-1})$ in an uniform sense. Here B is a subset of $\mathcal{R}^p$ and $\mathbf{f_3}(\cdot)$ & $\mathbf{f_4}(\cdot)$ are the same functions as in the previous slide.

Remarks

- Bootstrap methods have much better accuracy than asymptotic normal approximation.
- RB and PB corrects for skewness whereas normal approximation does not.
- In case of bootstrap methods one just need to repeat the procedure several times (say $n \log n$ times) and then sort them and find desired quantile.
- PB is clearly easy to implement than RB. The alternative PB for the LS requires to find LS estimator depending on some pseudo observations.

Simulation Study

Framework:

- $p = 10$ and $\beta = (5, 6.5, -3, 2, -7.5, -3.5, 4, -1, 9, 4)'$.
- Design vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ generated from $MVN(\mathbf{0}, \boldsymbol{\Sigma})$ where $\Sigma_{i,j} = 0.5^{|i-j|}$.
- Errors $\epsilon_1, \ldots, \epsilon_n$ generated separately from
  - $N(0, 1)$
  - $Laplace(0, 1/\sqrt{2})$
  - $Gumbel(-0.45, 0.78)$
  - $0.5 * Gumbel(-0.75, 0.78) + 0.5 * Gumbel(-0.15, 0.78)$
- $\rho(x) = x^2/2$ or $\psi(x) = x$.

Comparison: We compare the empirical coverages of 95 % confidence regions obtained by the three methods.

Error $\sim N(0, 1)$

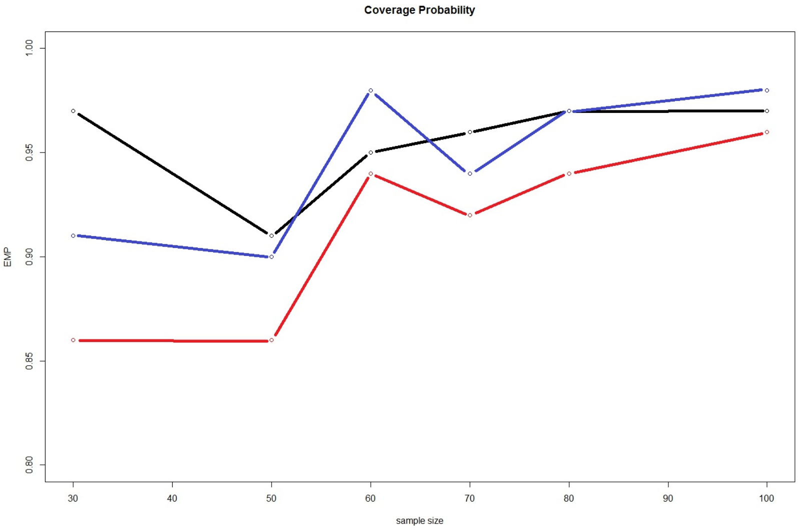


Figure 1: AN vs RB vs PB

Error $\sim$ *Laplace*$(0, 1/\sqrt{2})$

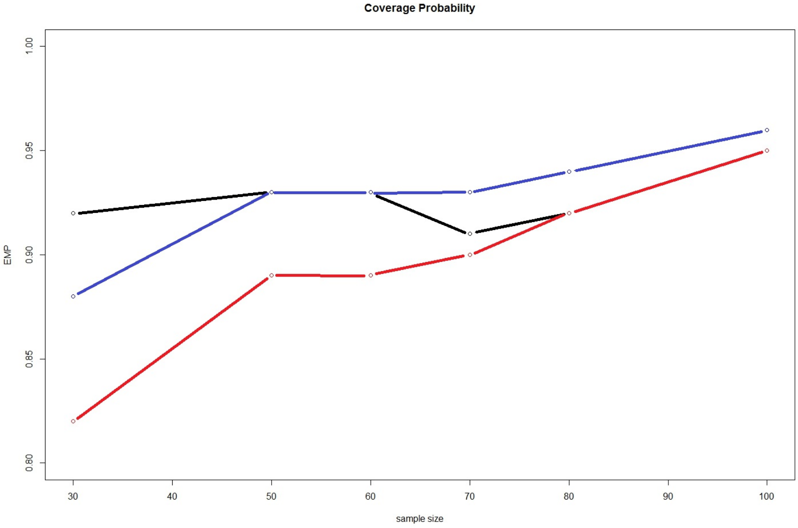


Figure 2: AN vs RB vs PB

Error $\sim$ $Gumbel(-0.45, 0.78)$

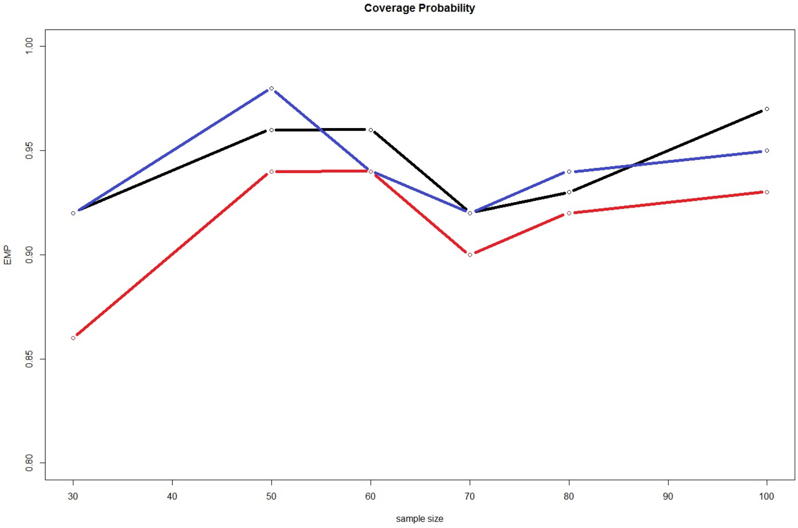


Figure 3: AN vs RB vs PB

$$0.5 * Gumbel(-0.75, 0.78) + 0.5 * Gumbel(-0.15, 0.78)$$

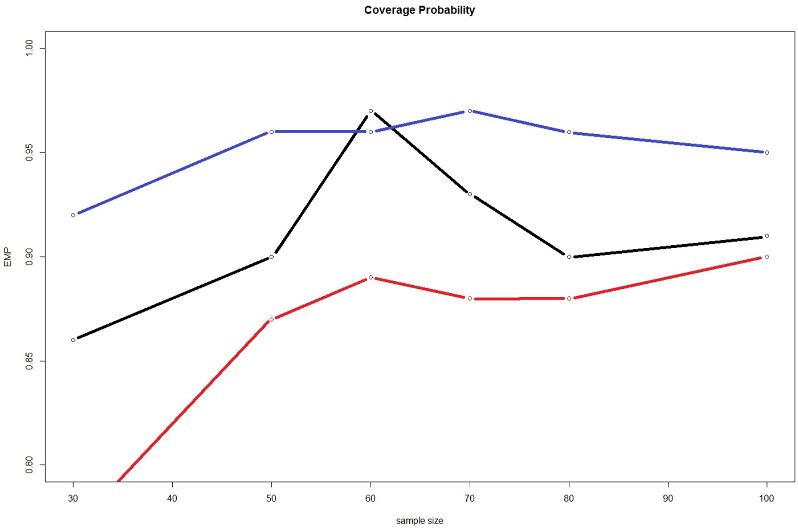


Figure 4: AN vs RB vs PB

References

- BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434-451.
- DAS, D. and LAHIRI, S. N. (2016). Second Order Correctness of Perturbation Bootstrap M-Estimator of Multiple Linear Regression Parameter. Preprint. Available at arXiv:1605.01440.
- FREEDMAN, D. A. (1981). Bootstrapping Regression Models. *Ann. Statist.* **96** 1218–1228.
- HUBER, P. (1981). *Robust Statistics*. Wiley, New York.
- JIN, Z. , YING, Z. and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika*. **88** 381-390 .
- LAHIRI, S. N. (1992). Bootstrapping M-estimators of a multiple linear regression parameter. *Ann. Statist.* **20** 1548-1570.

Thank you