Gaussian approximation and bootstrap confidence set

Consider a parametric problem for an independent sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$: $\boldsymbol{Y} \sim I\!P \in (I\!P_{\boldsymbol{\theta}})$. The MLE is computed by maximizing $L(\boldsymbol{\theta}) = \log(dI\!P_{\boldsymbol{\theta}}/d\mu)$:

$$\widetilde{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta})$$

where $\ell_i(Y_i, \boldsymbol{\theta})$ is the log-likelihood for one observation Y_i. The bootstrap log-likelihood $L^{\flat}(\boldsymbol{\theta})$ and the bootstrap estomate $\widetilde{\boldsymbol{\theta}}^{\flat}$ are defined by reweighting the log-likelihood

$$\widetilde{\boldsymbol{\theta}}^{\flat} = \operatorname*{argmax}_{\boldsymbol{\theta}} L^{\flat}(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta}) w_i^{\flat}$$

with independent random weights $w_i^{\flat}$. The bootstrap confidence width $z^{\flat}$ is defined by the condition

$$I\!P^{\flat}\Bigl(L^{\flat}(\widetilde{\boldsymbol{\theta}}^{\flat}) - L^{\flat}(\widetilde{\boldsymbol{\theta}}) > z^{\flat}\Bigr) = \alpha$$

The bootstrap confidence set is

$$\mathcal{E}(z^{\flat}) = \bigl\{\boldsymbol{\theta} \colon L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}) \le z^{\flat}\bigr\}$$

Bootstrap validity requires to evaluate the resulting coverage probability for the true parameter $\boldsymbol{\theta}^{*}$:

$$\bigl| I\!P\bigl(\boldsymbol{\theta}^{*} \in \mathcal{E}(z^{\flat})\bigr) - \alpha \bigr| \le \diamondsuit_n$$

Existing results ensure quite slow approximation rate (AR) $(n/p^3)^{-1/8}$, where p is the parameter dimension.

Workpackages:

1. Improve the AR to $(n/p^a)^{-1/2}$ or even $(n/p^a)^{-1}$. In particular, study the impact of the dimension p. The current guess is $a = 3/2$.

2. Explore some particular case including Generalized Linear Models and Instrumental Variables

3. Efficient implementation for special problems and applications to benchmark data.

Contact: Alexey Naumov, Arshak Minasyan, VS