Ch 11. Bootstrap cont'd

Ana-Maria Staicu

11/23/2019

We will discuss applications of bootstrap for constructing confidence intervals (CIs) and hypothesis testing.

- Recall defn of CIs: Given an estimate $\widehat{\theta}$ of para θ and an estimated standard error $\widehat{se}$ the ususual $100(1-2\alpha)\%$ CIs are constructed as

$$\widehat{\theta} \pm z_{1-\alpha}\widehat{se};$$

where $z_{1-\alpha}$ is the $(1-\alpha)$ quantile of $N(0,1)$. These CI rely on a normal asymptotic distribution of $\widehat{\theta}$.

Interpretation: under the normality assn, $(\widehat{\theta} - z_{1-\alpha}\widehat{se},\ \widehat{\theta} + z_{1-\alpha}\widehat{se})$ contains the true value θ with probability $(1-\alpha)$. Alternatively, for a one-sided tail.

$$P(\theta \geq \widehat{\theta} - z_{1-\alpha}\widehat{se}) \approx 1-\alpha.$$

Next: How to use bootstrap to construct valid CIs?

Percentile interval

The $100(1-2\alpha)\%$ **bootstrap percentile interval** for θ is defined

$$(\widehat{\theta}_{lo}, \widehat{\theta}_{up}) = (\widehat{\theta}_B^{*(\alpha)}, \widehat{\theta}_B^{*(1-\alpha)})$$

where $\widehat{\theta}_B^{*(\alpha)}$ and $\widehat{\theta}_B^{*(1-\alpha)}$ are the 100αth and $100(1-\alpha)$th empirical percentile of the bootstrap replicates $\widehat{\theta}^{*1}, \ldots, \widehat{\theta}^{*B}$.

Properties/technical details:

- its success depends on how well the distn of $\widehat{\theta}^*$ approx the distn of $\widehat{\theta}$;
- choose large B; for the IID simple case choose $B = 2000$;
- the percentile interval is invariant to transformations of θ.

Percentile interval: justification

Does the percentile CI guarantee that it contains the true para with probab equal to $100(1-2\alpha)$?

- Key assn: Efron's motivation was based on the assn that it exists a monotone increasing transformation $g(\cdot)$ such that

$$P\left(g(\widehat{\theta}) - g(\theta) \leq x\right) = \Phi(x);$$

  heere $\Phi(\cdot)$ does not have to be the CDF of a standard normal. It's sufficient to be symmetric about zero.

- Justification roadmap: Consider one-sided CIs and construct the one sided for θ (real world). Then show that the lower bound coincided with the corresponding percentile of the boostrap replicate samples (boot world).
- Percentile interval doesn not require the explicit form of $g(\cdot)$.

Bias corrected (BC) interval. Intuition

- The percentile intervals work if the distn of $\widehat{\theta}$ doesn't have median equal to θ (*median unbiased*). If that's not the case, then they are biased. Th BC intervals correct this bias.
- Assume that *it exists a monotone increasing transformation* $g(\cdot)$ such that we have

$$P\left(g(\widehat{\theta}) - g(\theta) + z_0 \leq x\right) = \Phi(x);$$

where z_0 accounts for non-symmetry in the distn of $\widehat{\theta}$.

- The constant z_0 is calculated as

$$P(\widehat{\theta} \leq \theta) = \left(g(\widehat{\theta}) - g(\theta) \leq 0\right) = \Phi(z_0).$$

Bias corrected (BC) interval

The $100(1-2\alpha)\%$ **bias corrected percentile interval** for θ is defined by the $100\alpha_1$th and $100(1-\alpha_2)$th percentiles in the boostrap sample of replicates

$$(\widehat{\theta}_{lo}, \widehat{\theta}_{up}) = (\widehat{\theta}_B^{*(\alpha_1)}, \widehat{\theta}_B^{*(1-\alpha_2)})$$

where $\alpha_1 = \Phi\left(2z_0 + \Phi^{-1}(\alpha)\right)$ and $1-\alpha_2 = \Phi\left(2z_0 + \Phi^{-1}(1-\alpha)\right)$.

- Justification: similar to the previous arguments. Work on the board.
- BC intervals are invariant to transformations of θ.

Bias corrected accelerated (BCa) interval (intuition only)

There is a second correction of the percentil intervals, called **bias corrected percentile intervals**. They rely on the assn that *it exists a monotone increasing transformation* $g(\cdot)$ *such that* we have

$$P\left(\frac{g(\widehat{\theta}) - g(\theta)}{1 + ag(\theta)} + z_0 \leq x\right) = \Phi(x)$$

where a - is the acceleration and z_0 as before.

These intervals are more complicated in form. Not discussed here!

Confidence intervals: Reflected percentile

Recall that the bootstrap procedure approximates the distn of $\widehat{\theta} - \theta$ by the empirical distribution of $\widehat{\theta}^* - \widehat{\theta}$.

The $100(1-2\alpha)\%$ **reflected percentile** for θ is defined by

$$\left(2\widehat{\theta} - \widehat{\theta}_b^{*(1-\alpha)},\ 2\widehat{\theta} - \widehat{\theta}_b^{*(\alpha)}\right).$$

where $\widehat{\theta}_B^{*(\alpha)}$ and $\widehat{\theta}_B^{*(1-\alpha)}$ are the 100αth and $100(1-\alpha)$th empirical percentile of $\widehat{\theta}^{*1}, \ldots, \widehat{\theta}^{*B}$.

- Justification: on the board.
- Not invariant to tranformations of θ.
- Other names for this interval: *hybrid percentile* (Shao & Tu, 1995) and *basic interval* (Davison & Hinkley, 1997).

Confidence intervals: bootstrap t confidence intervals

The bootstrap-t interval is based on the assn that the distn of

$$t_n = \frac{\widehat{\theta} - \theta}{\widehat{\sigma}}$$

is approximated by the empirical distn of

$$t_n^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{\widehat{\sigma}^*}.$$

The $100(1-2\alpha)\%$ **bootstrap - t interval** for θ is defined by

$$\left(\widehat{\theta} - t_B^{*(1-\alpha)}\widehat{\sigma}^*,\ \widehat{\theta} - t_B^{*(\alpha)}\widehat{\sigma}^*\right).$$

where $t_B^{*(\alpha)}$ and $t_B^{*(1-\alpha)}$ are the 100α and $100(1-\alpha)$ percentiles in the values $\{t_n^{*1}, \ldots, t_n^{*B}\}$'s.

- It requires double bootstrap: one to calculate $\widehat{\theta}_n^{*b}$ and another to get $\widehat{\sigma}^{*b}$ used in the calculation of t_n^{*b}.
- Justification:on the board.

Bootstrap for hypothesis testing

Consider a general hypothesis testing problem, where the test statistic is specified. To obtain p-value we need to know the distn of the test statistics under the null hypothesis!

- Bootstrap is used to approximate the distribution of the test statistic (call it t_n) under the null hypothesis.
- Key idea: contruct bootstrap samples in a way that ensures that the null hypothesis is valid.
- For every bootstrap draw calculate t_n^{*b}.
- P-value is approximates as

$$p - value = \frac{|\{t_n^{*b} \geq t_n^0 : b = 1, \ldots, B\}|}{B},$$

where t_n^0 is the value of the test statistic for the observed data.

Example: two sample testing via bootstrap

Consider two samples: $X_1, \dots, X_m \sim IID\ F$ and $Y_1, \dots, Y_n \sim IID\ G$ and assume that the two distribution differ solely in their mean, call them μ_F and μ_G respectively.

- Interest in testing $H_0 : \mu_F = \mu_G$.
- Test statistic

$$t_n = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{1/m + 1/n}}$$

  where s_p is the pooled standard error defined as

$$s_p^2 = \{(m-1)s_X^2 + (n-1)s_Y^2\}/(m+n-2)$$

  and s_X^2 and s_Y^2 are sample variances.
- For moderate sample sizes, and assuming F and G are normal, the null distn of t_n is Stundent t with $df = m + n - 2$.
- Alternatively, use bootstrap to approx the null distn of t_n.

Final remarks

- Bootstrap is a computationally intensive method helpful to estimate bias, standard deviation, construct confidence intervals, estimate p-values.
- There is rich literature (monographs + papers) on methods and theoretical properties of bootstrap techniques in many settings, including spatial stats, functional data etc.
- Last word: bootstrap methods can be extremely slow. When implementing boostrap, avoid *for* loops.