PhD in Economics and Statistics - University of Milano-Bicocca

Lecture 3: False Coverage Rate

May 7, 2019

Lecturer: Aldo Solari

The False Coverage Rate criterion was introduced by [1]. Section 2 is adapted from [2].

1 Simultaneous confidence intervals

Suppose that we have m parameters of interest $\theta_1, \ldots, \theta_m$ and we wish to construct a confidence interval for each of them such that *all* the intervals will contain their respective parameters with probability at least $1 - \alpha$, i.e.

$$\mathrm{P}(\bigcap_{i=1}^{m} \{\widetilde{\mathrm{CI}}_i \ni \theta_i\}) \geq 1 - \alpha$$

For example, let $y_1, \ldots, y_m$ are independent with $y_i \sim N(\theta_i, 1)$. In this setting we can construct marginal $1 - \alpha$ confidence intervals

$$\mathrm{CI}_i = [y_i - z_{1-\alpha/2}, y_i + z_{1-\alpha/2}]$$

such that

$$\mathrm{P}(\mathrm{CI}_i \ni \theta_i) \geq 1 - \alpha$$

Simultaneous confidence intervals are given by the Bonferroni confidence intervals

$$\widetilde{\mathrm{CI}}_i = [y_i - z_{1-\alpha/2m}, y_i + z_{1-\alpha/2m}]$$

Bonferroni confidence intervals are simple to use but conservative (i.e. the actual coverage probability is greater than claimed and the confidence intervals are wider than they have to be).

2 False coverage control

Often researchers examine many parameters at once and report confidence intervals only for selected ones. Confidence intervals may have serious reduced coverage probability after selection. One goal would be to achieve *conditional coverage*

$$\mathrm{P}(\mathrm{CI}_i \ni \theta_i | i \in \mathcal{S}) \geq 1 - \alpha$$

where $\mathcal{S}$ is the set of selected parameters. However, conditional coverage following any selection rule cannot, in general, be achieved. To see this, suppose $\theta_i = 0$ for all i. No matter how we construct a CI for θ_i, selecting θ_i if and only if CI_i does not contain 0 results in $\mathrm{P}(\mathrm{CI}_i \ni 0 | i \in \mathcal{S}) = 0$.

In 2005, Benjamini and Yekutieli introduced a relaxation of conditional coverage, named *False Coverage Rate* (FCR).

Definition 2.1. The false coverage rate is defined as

$$\mathrm{FCR} = \mathrm{E}\Big(\frac{V_{\mathrm{CI}}}{R_{\mathrm{CI}} \vee 1}\Big)$$

where R_{CI} is the number of selected parameters and V_{CI} the number of constructed confidence intervals not covering.

Note that without selection, the marginal CIs control the FCR since

$$\mathrm{E}\Big(\frac{\sum_{i=1}^{m} \mathbb{1}\{\theta_i \notin \mathrm{CI}_i\}}{m}\Big) \leq \alpha$$

Let $y_1, \ldots, y_m$ are independent with $y_i \sim N(\theta_i, 1)$. Consider the following selection rules, defined a priori (before seeing the data):

- Unadjusted selection rule:

$$\mathcal{S} = \{i \in \{1, \ldots, m\} : |y_i| \geq |z_{\alpha/2}|\}$$

- Bonferroni selection rule:

$$\mathcal{S} = \{i \in \{1, \ldots, m\} : |y_i| \geq |z_{\alpha/2m}|\}$$

- BH selection rule: calculate $p_i = 2[1 - \Phi(|y_i|)]$ and obtain $R = \max\{i : p_{(i)} \leq i\alpha/m\}$. Then

$$\mathcal{S} = \{i \in \{1, \ldots, m\} : p_i \leq \frac{R\alpha}{m}\}$$

For these (simple) selection rules, the FCR controlling procedure of [1] works as follows:

- Apply the selection rule (Unadjusted, Bonferroni, BH) and obtain $\mathcal{S}$
- Construct

$$\widetilde{\mathrm{CI}}_i = [y_i - z_{1-\tilde{\alpha}}, y_i + z_{1-\tilde{\alpha}}] \qquad i \in \mathcal{S}$$

  where

$$\tilde{\alpha} = \frac{\alpha |\mathcal{S}|}{2m}$$

If the y_i are independent, then this procedure has FCR $\leq \alpha$.

References

[1] Y. Benjamini and D. Yekutieli. False discovery rate–adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, 100(469):71–81, 2005.

[2] E. Candes et al. Stats 300c: Theory of statistics. *Lecture notes*, 2018.