Bootstrap-t confidence intervals

Mètodes no paramètrics i de remostreig
Grau en Estadística, UB – UPC
Prof. Jordi Ocaña Rebull

Brief reminder of the pivot method: normal distribution mean

$\mathbf{X} = (X_1, \ldots, X_n)$ *iid* random sample from $N(\mu, \sigma^2)$

$$t(\mathbf{X}) = \frac{\sqrt{n}(\bar{X} - \mu)}{\hat{S}} = \frac{(\bar{X} - \mu)}{\hat{S}/\sqrt{n}} = \frac{(\bar{X} - \mu)}{\widehat{\mathrm{se}}_{\bar{X}}} \sim \mathbf{t}(n-1)$$

t is a "pivot", its distribution does not depend on unknown parameters: there are t_L, t_U constants (e.g. $-t_L = t_U = t_{1-\alpha/2}(n-1)$) such that:

$$\Pr\{t_L \le t(\mathbf{X}) \le t_U\} = 1 - \alpha$$

Student's **t** tables

By elementary operations:

$$1-\alpha = \Pr\left\{ t_L \le \frac{\sqrt{n}\left(\bar{X}-\mu\right)}{\hat{S}} \le t_U \right\} =$$

$$\Pr\left\{ \bar{X} - t_U \frac{\hat{S}}{\sqrt{n}} \le \mu \le \bar{X} - t_L \frac{\hat{S}}{\sqrt{n}} \right\} =$$

$$\Pr\left\{ \bar{X} - t_U \widehat{\mathrm{se}}_{\bar{X}} \le \mu \le \bar{X} - t_L \widehat{\mathrm{se}}_{\bar{X}} \right\}$$

Brief reminder of the pivot method: normal distribution mean

- All the preceding formulae and steps are still valid…
  - standard error estimation, (nearly) pivotal character of *t*…
- …except the *t* sampling distribution, no guarantee of **t**(n – 1)
- Bootstrap-t CI: estimate this distribution by a bootstrap process and compute adequate quantiles t_L^* and t_U^* in order to have:

$$1-\alpha \simeq \Pr\left\{ t_L^* \leq \frac{\sqrt{n}\left(\bar{X}-\mu\right)}{\hat{S}} \leq t_U^* \right\}$$

CI for the mean, non-normal data

Bootstrap-t resampling

$$\hat{F}$$

$$\downarrow$$

$$\mathbf{x}_1^* = \left(x_{11}^*, \ldots, x_{1n}^*\right) \mapsto t_1^* = \frac{\sqrt{n}\left(\bar{X}_1^* - \bar{X}\right)}{\hat{S}_1^*}$$

$$\mathbf{x}_2^* = \left(x_{21}^*, \ldots, x_{2n}^*\right) \mapsto t_2^* = \frac{\sqrt{n}\left(\bar{X}_2^* - \bar{X}\right)}{\hat{S}_2^*}$$

$$\vdots$$

$$\mathbf{x}_B^* = \left(x_{B1}^*, \ldots, x_{Bn}^*\right) \mapsto t_B^* = \frac{\sqrt{n}\left(\bar{X}_B^* - \bar{X}\right)}{\hat{S}_B^*}$$

$$t_L^* \qquad t_U^*$$

$$\alpha_1 + \alpha_2 = \alpha$$

$$\overbrace{t_{(1)}^* \le t_{(2)}^* \le \cdots}^{\alpha_1} \Big| \overbrace{\cdots\cdots}^{1-\alpha} \Big| \overbrace{\cdots \le t_{(B-1)}^* \le t_{(B)}^*}^{\alpha_2}$$

- A common election: $\alpha_1 = \alpha_2 = \alpha/2$
- Let $\hat{H}$ be the bootstrap estimate of the t sampling distribution, then

$$\hat{t}_L = \hat{H}^{-1}\left(\frac{\alpha}{2}\right), \quad \hat{t}_U = \hat{H}^{-1}\left(1 - \frac{\alpha}{2}\right)$$

- Themselves approximated by the sample quantiles of $t_1^*, t_2^*, \ldots, t_B^*$

$$\hat{t}_L \simeq t_L^* = t^*_{\left(\frac{\alpha}{2}\right)}, \quad \hat{t}_U \simeq t_U^* = t^*_{\left(1-\frac{\alpha}{2}\right)}$$

Equally tailed bootstrap-t confidence interval

$$\left[\bar{X} - t^{*}_{\left(1-\frac{\alpha}{2}\right)} \frac{\hat{S}}{\sqrt{n}}, \quad \bar{X} - t^{*}_{\left(\frac{\alpha}{2}\right)} \frac{\hat{S}}{\sqrt{n}}\right]$$

Equally tailed bootstrap-t confidence interval

- (Symmetric with respect to the point estimate, here $\bar{X}$)
- Define $\hat{t}_{[1-\alpha]}$ as the value such that

$$1-\alpha = \Pr_{\hat{H}}\left\{\left|\frac{\sqrt{n}\left(\bar{X}^{*}-\bar{X}\right)}{\hat{S}^{*}}\right| \le \hat{t}_{[1-\alpha]}\right\}$$

- Itself approximated by the 1 – α sample quantile of the $\left|t_1^*\right|,\left|t_2^*\right|,\ldots,\left|t_B^*\right|$:

$$\hat{t}_{[1-\alpha]} \simeq t^{*}_{[1-\alpha]} = \left|t^{*}\right|_{(1-\alpha)}$$

Symmetrized bootstrap-t confidence interval

$$\left[\bar{X} - t^*_{[1-\alpha]}\frac{\hat{S}}{\sqrt{n}},\quad \bar{X} + t^*_{[1-\alpha]}\frac{\hat{S}}{\sqrt{n}}\right]$$

that is:

$$\bar{X} \pm t^*_{[1-\alpha]}\frac{\hat{S}}{\sqrt{n}}$$

Symmetrized bootstrap-t confidence interval

- Let θ be a parameter, and $\hat{\theta} = \hat{\theta}(\mathbf{X})$ an estimator of θ computed over some dataset **X** (e.g. multivariate, time series…)
- Let $\widehat{\mathrm{se}}_{\hat{\theta}}$ be an estimator of the standard error of $\hat{\theta}$
- Under mild conditions, the studentized statistic

$$t(\mathbf{X}) = \frac{\hat{\theta}(\mathbf{X}) - \theta}{\widehat{\mathrm{se}}_{\hat{\theta}}}$$

is approximately pivotal

General bootstrap-t confidence intervals

- All the previous statements are applicable to this more general context
- The main problem reduces to the bootstrap estimation of (nearly) constants such that

$$1 - \alpha \simeq \Pr\left\{ t_L^* \le \frac{\hat{\theta}(\mathbf{x}) - \theta}{\widehat{\mathrm{se}}_{\hat{\theta}}} \le t_U^* \right\}$$

General bootstrap-t confidence intervals

Simulate B datasets from the conjectured model producing the "true" data $\mathbf{X}$, adjusting it from $\mathbf{X}$

$\downarrow$

$$\mathbf{X}_1^* \mapsto t_1^* = \frac{\hat\theta(\mathbf{x}_1^*) - \hat\theta(\mathbf{X})}{\widehat{\text{se}}_{\hat\theta}(\mathbf{x}_1^*)} = \frac{\hat\theta_1^* - \hat\theta}{\widehat{\text{se}}_1^*}$$

$$\mathbf{X}_2^* \mapsto t_2^* = \frac{\hat\theta(\mathbf{x}_2^*) - \hat\theta(\mathbf{X})}{\widehat{\text{se}}_{\hat\theta}(\mathbf{x}_2^*)} = \frac{\hat\theta_2^* - \hat\theta}{\widehat{\text{se}}_2^*}$$

$$\vdots$$

$$\mathbf{X}_B^* \mapsto t_B^* = \frac{\hat\theta(\mathbf{x}_B^*) - \hat\theta(\mathbf{X})}{\widehat{\text{se}}_{\hat\theta}(\mathbf{x}_B^*)} = \frac{\hat\theta_B^* - \hat\theta}{\widehat{\text{se}}_B^*}$$

t_L^* t_U^*

$$\alpha_1 + \alpha_2 = \alpha$$

$$\overbrace{t_{(1)}^* \le t_{(2)}^* \le \cdots}^{\alpha_1}\ \overbrace{\cdots\cdots}^{1-\alpha}\ \overbrace{\cdots \le t_{(B-1)}^* \le t_{(B)}^*}^{\alpha_2}$$

General bootstrap-t resampling

- Equally tailed:

$$\left[\hat{\theta} - t^*_{\left(1-\frac{\alpha}{2}\right)}\widehat{\text{se}}_{\hat{\theta}},\quad \hat{\theta} - t^*_{\left(\frac{\alpha}{2}\right)}\widehat{\text{se}}_{\hat{\theta}}\right]$$

- Symmetrized:

$$\left[\hat{\theta} - t^*_{[1-\alpha]}\widehat{\text{se}}_{\hat{\theta}},\quad \hat{\theta} + t^*_{[1-\alpha]}\widehat{\text{se}}_{\hat{\theta}}\right],\quad \text{that is:}$$

$$\hat{\theta} \pm t^*_{[1-\alpha]}\widehat{\text{se}}_{\hat{\theta}}$$

Most common general bootstrap-t confidence intervals