Error Propagation and Bootstrapping Cheatsheet

Bootstrapping is used to estimate uncertainty when the sampling distribution is unknown or hard to derive, especially for small samples or complex statistics.

Variance and Standard Error

Variance For N observations of variable x_i, $\mathbf{x}_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,N}\}$, with mean $\bar{x}_i$:

$$\sigma_i^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_{i,n} - \bar{x}_i)^2$$

measures the spread of x_i around its mean.

Standard Error of the Mean (SEM)

$$\mathrm{SEM}_i = \frac{\sigma_i}{\sqrt{N}} \tag{1}$$

The standard error quantifies the uncertainty in the mean's estimate. As N increases, the standard error decreases.

Error propagation Formula

For a function

$$y = f(\mu_1, \mu_2, \ldots, \mu_m)$$

where μ_i are measured values, e.g. means, with variances σ_i^2 and covariances Σ_{ij}. Propagated error is given by:

$$\sigma_y^2 = \sum_{i=1}^{m}\left(\frac{\partial f}{\partial \mu_i}\right)^2 \sigma_i^2 + 2\sum_{i<j}\frac{\partial f}{\partial \mu_i}\frac{\partial f}{\partial \mu_j}\Sigma_{ij}$$

Common formulas

Addition/Subtraction

$$y = \mu_1 \pm \mu_2 \quad \Rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2$$

Multiplication/Division

$$y = \mu_1^{a_1}\mu_2^{a_2}\cdots\mu_m^{a_m} \quad \Rightarrow \quad \left(\frac{\sigma_y}{y}\right)^2 = \sum_{i=1}^{m} a_i^2\left(\frac{\sigma_i}{\mu_i}\right)^2$$

Example.

$$y = \frac{AB}{C^2}$$

$$\left(\frac{\sigma_y}{y}\right)^2 = \left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 + \left(2\frac{\sigma_C}{C}\right)^2$$

If $A = 2.50 \pm 0.03$, $B = 1.20 \pm 0.02$, $C = 3.00 \pm 0.05$: $y = 0.333 \pm 0.013$

Practical Tips

- Only use when $y \gg \sigma_y$ or f is linear otherwise this formula does not hold! e.g. $f(x) = e^{x^2}$ with $x, \sigma_x > 1$
- For nonlinear models or large uncertainties, use ***Monte Carlo*** propagation: simulate many draws of $\{x_i\}$ from their distributions, compute f each time, and take the standard deviation of results.
- Report results as $y \pm \sigma_y$ (68% confidence) unless otherwise stated.

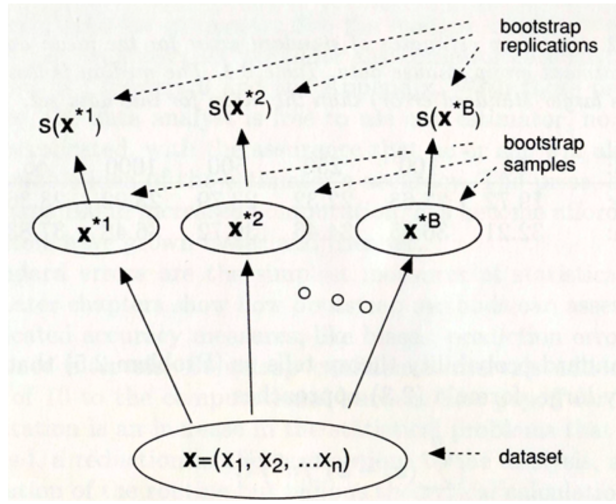

Figure 1: Bootstrapping schematic

Bootstrap Procedure

1. Decide on what measure/statistic/function $s(\mathbf{x})$ to compute from dataset $\mathbf{x}$.
2. Collect N data points x_i to make dataset $\mathbf{x}$.
3. Create B bootstrap sample datasets $\mathbf{x}^{*b}$ by randomly selecting N data points **with replacement** from the original dataset.
4. Compute measure $s(\mathbf{x}^{*b})$ on each sample dataset.
5. Calculate standard error or confidence error

$$\mathrm{SEM}_s^2 = \frac{1}{B-1}\sum_{b=1}^{B}(s(\mathbf{x}^{*b}) - \bar{s}(\mathbf{x}^*))^2$$

where $\bar{s} = B^{-1}\sum_{b=1}^{B} s(\mathbf{x}^{*b})$ is the mean of all bootstrap sample statistics.

Rules of thumb

- Data points should be *representative* – lack bias or missing points.
- Shoot for $N > 20$. Need unique bootstrap samples.
- B should be between 100-5000. Monte Carlo error scales as $1/\sqrt{B}$.
- Check the distribution shape of $s(\boldsymbol{x}^{*,b})$. Does it make sense? Gaussian is good but not necessary.
- Statistic $s(\boldsymbol{x})$ needs to be "smooth" enough — small changes in the data cause small changes in the estimate.