DELTA METHOD AND BOOTSTRAPPING

AMY RHEE

1. The Delta Method

Motivation for Delta Method: If $\hat{\theta}_n$ is close to θ, then we can use a first-order Taylor approximation:

$$\begin{aligned}
&\frac{g(\hat{\theta}_n) - g(\theta)}{\hat{\theta}_n - \theta} \approx g'(\hat{\theta}_n) \\
\implies &\frac{g(\hat{\theta}_n) - g(\theta)}{g'(\hat{\theta}_n)} \approx \hat{\theta}_n - \theta \\
\implies &\frac{g(\hat{\theta}_n) - g(\theta)}{g'(\hat{\theta}_n)\hat{S}(\hat{\theta}_n)} \approx \frac{\hat{\theta}_n - \theta}{\hat{S}(\hat{\theta}_n)}, \text{ where } \hat{S}(\cdot) \text{ is the estimated standard error}
\end{aligned}$$

The Delta Method: If Y_n satisfies the CLT, meaning that $Y_n \xrightarrow{d} \mathcal{N}(\mu, \sigma^2/n)$, then for a function $g(\cdot)$ that is differentiable and non-zero valued, $g(Y_n) \xrightarrow{d} \mathcal{N}\big(g(\mu), \sigma^2 \frac{g'(\mu)^2}{n}\big)$.

The Delta Method follows from a first-order Taylor expansion:

$$g(\hat{\theta}_n) \approx g(\theta) + g'(\theta)(\hat{\theta}_n - \theta)$$

Then, since $\hat{\theta}_n \xrightarrow{p} \theta$, it follows that $\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = \sqrt{n}(\hat{\theta}_n - \theta)(g'(\theta) + o_P(1))$. Then, application of Slutsky's theorem and the fact that $\sqrt{n}(\hat{\theta}_n - \mu) \xrightarrow{d} \mathcal{N}(\mu, \sigma^2)$ finishes the proof.[1]

Multivariate Delta Method Let $\{\hat{\theta}_n\}$ be a sequence of random vectors. If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, where θ is a constant vector, then $\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, J_g(\theta)'\Sigma J_g(\theta))$ where $J_g(\theta)$ is the Jacobian of g, i.e. the matrix of partials of the entries of g with respect to the entries of θ.

We can also write this for just our estimator: If $\hat{\theta}_n \xrightarrow{d} \mathcal{N}(\theta, \Sigma/n)$, where θ is a constant vector, then $g(\hat{\theta}_n) \xrightarrow{d} \mathcal{N}\big(g(\theta), J_g(\theta)'(\Sigma/n)J_g(\theta)\big)$.

t-statistic: If $\hat{\theta}_n$ is the estimator of our parameter, θ is a known constant (typically the parameter guess under the null hypothesis), and $S(\hat{\theta})$ is the standard error of the estimator,

Date: March 22, 2023.

[1]If we were in the multivariate case, replace $g'(\theta)$ with $\nabla g(\theta)^T$

then

$$t = \frac{\hat{\theta}_n - \theta}{S(\hat{\theta}_n)}$$

Estimating Standard Errors via the Delta Method:

Constructing Confidence Intervals via the Delta Method: Given that $t = \frac{\hat{\theta}_n - \theta}{S(\hat{\theta})} \sim \mathcal{N}(0,1)$, a $(1-\alpha)\%$ confidence interval[2] for θ can be found from

$$1 - \alpha = \mathbb{P}\Big(\hat{\theta}_n - q_{\alpha/2} S(\hat{\theta}_n) < \theta < \hat{\theta}_n + q_{\alpha/2} S(\hat{\theta}_n)\Big)$$

The key is to apply the delta method to get the standard error.

2. Bootstrapping

Motivation for Bootstrapping: Suppose we observe iid observations $\{X_i\}_{i=1}^n$ from some distribution with CDF P_0, and suppose our parameter of interest is $\theta_0 = \theta(P_0)$ (we let θ depend on P_0 to signifiy that it is identified from the distribution of observables). Typically, to perform inference on θ_0, we want to be able to construct a test statistic, etc., but we don't know how to calculate the finite sample distribution $H_n(\cdot; P_0)$. Therefore, we often use asymptotic inference ($n \to \infty$) based on an approximation:

$$H_n(z; P_0) \to H_\infty(z; P_0) \text{ as n} \to \infty$$

The idea of the bootstrap is to approximate $H_n(z; P_0)$ by $H_n(z; \hat{P})$, where $\hat{P}$ is an estimator of P_0. I.e.,

$$H_n(z; \hat{P}) = \mathbb{P}_{\hat{P}}\Big(\frac{\sqrt{n}\hat{\theta}(X_1^*, ..., X_n^*) - \theta(\hat{P})}{\hat{\sigma}(X_1^*, ..., X_n^*)} \leq z\Big)$$

where $\mathbb{P}_{\hat{P}}$ denotes probability statements about $X_1^*, ..., X_n^*$ conditional on $\hat{P}$. While we don't know how to calculate H_n, we can simulate data from $\hat{P}$ are compute H_n by simulation.

Note that in the bootstrap world (with $\mathbb{P}_{\hat{P}}$), we center at $\theta(\hat{P})$ instead of P_0. This is because $\theta(\hat{P})$ is the true parameter in the bootstrap world.

Simulation Steps: To simulate H_n, we could draw a new iid sample $X_1^1, ..., X_n^1$ from $\hat{P}$ and compute:

$$\mathbf{1}\{\frac{\sqrt{n}\hat{\theta}(X_1^1, ..., X_n^1) - \theta(\hat{P})}{\hat{\sigma}(X_1^1, ..., X_n^1)} \leq z\}$$

Repeat this for a large number B of independent draws, and we have:

$$\frac{1}{B}\sum_{i=1}^{B} \mathbf{1}\{\frac{\sqrt{n}\hat{\theta}(X_1^b, ..., X_n^b) - \theta(\hat{P})}{\hat{\sigma}(X_1^b, ..., X_n^b)} \leq z\} \to H_n(z; \hat{P})$$

[2]Recall: CI is a range of values in which the true value of the parameter θ lies with $1 - \alpha$ probability

Nonparametric Bootstrapping: We generate bootstrap samples $X_1^b, ..., X_n^b$ by sampling n draws randomly with replacement from the *original* data $X_1, ..., X_n$. We take $\hat{P}$ as the empirical distribution of the data, i.e.

$$\hat{P}(x) = \frac{1}{n} \sum_i \mathbf{1}\{X_i \leq x\}$$

Parametric Bootstrapping: We generate samples $X_1^b, ..., X_n^b$ by drawing n iid draws from $P(\cdot; \hat{\theta})$. Note that the parameteric bootstrap requires a *correctly specified, fully parametric model*.

Residual Bootstrap This works when we can apply a transformation of the data that reduces to sampling iid residuals. After, we can recenter them and sample them using the nonparametric bootstrap, or using the parametric bootstrap (if we have a correctly specified parametric distribution for the residuals).

For example, consider the AR(1) model

$$y_t = \rho y_{t-1} + \mu - u_t$$

We can estimate ρ and μ, and using $\hat{\rho}$ and $\hat{\mu}$, we can estimate our residuals

$$\hat{u}_t = y_t - \hat{\rho} y_{t-1} - \hat{\mu}$$

Then, to generate the data, we can generate a random sample $u_1^b, ..., u_n^b$ by resampling from the empirical distribution of the re-centered residuals $u_t^* = \hat{u}_t - \frac{1}{n} \sum_i \hat{u}_t$, or from a parametric distribution $F(u; \hat{\theta})$, where $\hat{\theta}$ is a consistent estimator of the true distribution of the residuals $F(u; \theta)$. We then reconstruct the bootstrap sample for $y_1^b, ..., y_n^b$

$$y_t^b = \hat{\rho} y_{t-1}^b + \hat{\mu} + u_t^b$$

Then, we estimate ρ and μ again for each bootstrap sample.