Bootstrap Confidence Intervals for Location-Scale Models

When a continuous random variable X has a probability density function of the form,

$$g(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}),$$

where $f(z)$ has mean 0 and standard deviation 1, it is called a location-scale family of densities. In this case X may be expressed as $X = \mu + \sigma Z$, where Z is a random variable with probability density $f(z)$. Our goal is to obtain bootstrap interval estimates for μ and σ, from a sample $X_1, X_2, ..., X_n$ with sample mean $\overline{X}$ and sample standard deviation S.

We will learn that there are many different methods available for constructing bootstrap confidence intervals, but one of the most accurate methods is available for location-scale families due to the concept of a pivot quantity, defined below.

When $f(z)$ is normal, then the t statistic,

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t distribution with $n-1$ degrees of freedom. Thus we have

$$P(-t_{.975} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{.975}) = .95$$

which yields a 95% confidence interval of

$$\overline{X} - t_{.975}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{.975}\frac{S}{\sqrt{n}}.$$

The t statistic defined above is a pivot quantity, which means that it does not depend on the values of μ and σ. If X is not normally distributed, then the t statistic is still a pivot quantity, known as a t-pivot, but it does not have a t distribution. The key idea for constructing bootstrap confidence intervals for μ in location-scale families is to obtain a bootstrap distribution of the t-pivot, and then to use it as above to get the confidence interval.

Bootstrap confidence interval for μ in a location-scale model

1. Compute $\overline{X}$ and S from the data.

2. Obtain a bootstrap sample of size n, and compute $\overline{X}_b$, S_b, and the t-pivot quantity

$$t_b = \frac{\overline{X}_b - \overline{X}}{S_b/\sqrt{n}}$$

3. Repeat step 2 many times to obtain a bootstrap distribution of t_b.
4. Now a $100(1-\alpha)\%$ confidence interval can be obtained via:

$$\overline{X} - t_{b,1-\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} - t_{b,\alpha/2}\frac{S}{\sqrt{n}}.$$

Note that this interval may not be symmetric about $\overline{X}$.

Bootstrap confidence interval for σ^2 in a location-scale model

Bootstrap intervals for σ^2 in location-scale families are constructed in a similar way, using the idea of a χ^2-pivot:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}.$$

When $f(z)$ is normal, then χ^2 follows a χ^2 distribution with $n-1$ degrees of freedom, and the probability statement

$$P(\chi^2_{.025} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{.975}) = .95$$

is used to yield the 95% confidence interval

$$\frac{(n-1)S^2}{\chi^2_{.975}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{.025}}.$$

Again, if X is not normally distributed, the χ^2-pivot no longer follows a χ^2 distribution, but is still a pivot quantity. We can then use the bootstrap to create a bootstrap distribution χ^2_b to obtain a confidence interval for σ^2, as shown in the text.