Estimating quantities via bootstrap.

When studying Wald's test, we ran into the following problem: assume that someone gives us an estimator $\hat{\theta}_n$ to estimate a certain parameter $\theta_\star$, that we do not know. This person also gives us a theoretical result guaranteeing that $\hat{\theta}_n$ is *asymptotically normal* in the following sense:

$$\frac{\hat{\theta}_n - \theta_\star}{\sqrt{\mathrm{Var}_{\theta_\star}(\hat{\theta}_n)}} \longrightarrow_{n\to\infty}^{\text{in distribution}} \mathcal{N}(0,1).$$

We want to use this result e.g. for hypothesis testing, or to build a confidence interval.

Given an observation $X_1, \ldots, X_n$, we can compute $\hat{\theta}_n(X_1, \ldots, X_n)$. But the variance term $\mathrm{Var}_{\theta_\star}(\hat{\theta}_n)$ cannot be computed from data, because it corresponds to the variance of $\hat{\theta}_n$ under the *true* distribution with parameter $\theta_\star$.

Good cases: when we have a simple expression Let us first consider a simple example, the case of the empirical mean $\hat{m}_n$. It is defined as

$$\hat{m}_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

In this case, using the fact that the X_i's are independent, we have the theoretical formula

$$\mathrm{Var}_{\theta_\star}\left[\hat{m}_n\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}_{\theta_\star}(X_i) = \frac{1}{n}\mathrm{Var}_{\theta_\star}(X).$$

The term $\mathrm{Var}_{\theta_\star}(X)$ is still not accessible, but we may estimate it from data, e.g. by using the empirical variance

$$\widehat{\mathrm{Var}}_n := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{m}_n\right)^2.$$

Let us note that even this case is not completely trivial. In particular, ask yourselves: *why (besides the name) does this give us a good estimate of the variance?*

General case: the bootstrap recipe Next, let us present the bootstrap method, without justification. Let $X_1, \ldots, X_n$ be our data set. Let us fix an integer m.

1. For $k = 1$ to m, pick n data points $Y_{1,k}, \ldots, Y_{n,k}$ from the data set (independently, **with** replacement). Call this the "k-th bootstrap data set" $D_k = (Y_{1,k}, \ldots, Y_{n,k})$.

2. For every $k = 1$ to m, compute the value $T_{n,k} := \hat{\theta}_n\left(Y_{1,k}, \ldots, Y_{n,k}\right)$.

3. Compute the following quantity

$$\frac{1}{m}\sum_{k=1}^{m}\left(T_{n,k} - \frac{1}{m}\sum_{j=1}^{m} T_{n,j}\right)^2 .$$

And return this as an estimate for the variance of $\hat{\theta}_n$.

Some justification, 1: the law of large numbers Let us recall the law of large numbers. Under very weak assumptions, if $Y_1, \ldots, Y_m$ are m independent random variables, all distributed as some fixed random variable Y, the following convergence holds

$$\frac{1}{m}\sum_{k=1}^{m} Y_k \longrightarrow_{m\to\infty} \mathbb{E}[Y] \text{ (in probability) } .$$

For simulation purposes, we would read this informally as follows: if I want to compute $\mathbb{E}[Y]$, but all I can do is sampling Y many times, I should sample it m times (with m large) and compute the empirical mean of my data set

$$\frac{1}{m}\sum_{k=1}^{m} Y_k,$$

which should give me a good numerical estimate for $\mathbb{E}[Y]$.

An important point to keep in mind, is that the law of large number is valid in many different situations. Let us for example consider the case where the random variable Y is defined as follows:

1. First, sample (A, B), a couple of two random variables with given joint distribution.

2. Then, compute $Y = h(A, B)$ where h is some fixed function of two variables.

How could we compute $\mathbb{E}[Y]$ numerically? Well, if we can sample the couple (A, B) as many times as we want, we should

1. Independently draw m realisations $(A_1, B_1), \ldots (A_m, B_m)$,

2. Compute each time the quantity $Y_k = h(A_k, B_k)$ (for $k = 1 \ldots m$),

3. And finally compute the empirical mean

$$\frac{1}{m}\sum_{k=1}^{m} Y_k \approx \mathbb{E}[Y].$$

There is (almost) no restriction on the way Y is defined. In particular, if Y is defined as a function of n (rather than 1 or 2) random variables $A_1, \ldots, A_n$, by some formula

$$Y = g(A_1, \ldots, A_n),$$

then a way to numerically compute $\mathbb{E}[Y]$ is to do the following:

1. Draw a large number m of realisations of the n-tuple, denote them by

$$(A_{1,1}, \ldots, A_{n,1}),\ (A_{1,2}, \ldots, A_{n,2}), \ldots (A_{1,m}, \ldots A_{n,m})$$

   (each parenthesis is a n-tuple, and we have m of them)

2. Compute $Y_k := g(A_{1,k}, \ldots, A_{n,k})$ for each $k = 1, \ldots, m$.

3. Finally, compute the empirical mean

$$\frac{1}{m} \sum_{k=1}^{m} Y_k \approx \mathbb{E}[Y].$$

Some justification, 2: re-sampling from the data set The bootstrap method is based on the previous observation: in order to compute, say $\mathbb{E}_{\theta_\star}[\hat{\theta}_n]$, where $\hat{\theta}_n$ is a function of n variables, we should

1. Generate m realisations (called above the "bootstrap data sets") of these n variables.

2. Compute the value of $\hat{\theta}_n$ for each realisation.

3. Compute an empirical mean.

Steps 2 and 3 are fine, but for step 1 there is an issue: each variable in each realisation should be distributed independently, and according to the true pdf, but we do not have access to it!! The only thing we have access to, is the initial data set...

Instead of sampling from the true distribution, that we cannot access, we will thus sample from the "empirical distribution", as provided by the data set. Sampling from the data set is equivalent to choosing a data point uniformly at random, and this is exactly what we do in the first step of the "bootstrap recipe".

Why "bootstrap"? In total, we need $m \times n$ points drawn from the real distribution. We only have n data points to start with... but we then pretend that they give a good description of the real distribution, and re-sample from them, to produce more fictive data points. Fortunately, it works...