Lecture - 18

Instructor: Dr. Arabin Kumar Dey

February 18, 2012

The bootstrap, invented by Efron in 1979, is a method for estimating standard errors and computing confidence intervals.

Let the $\theta = T(F)$ be an interesting parameter of a distribution function F, which is the mean parameter of the distribution F, where $T(\cdot)$ is a functional of F. One simple example is $T(F) = \int xdF(x)$, which is the mean parameter of the distribution F. Another example is $T(F) = \int (x - \int ydF(x))^2 dF(x)$, which is the variance parameter.

Let $X_1, \cdots, X_n$ be an i.i.d. sample from F, and we use $\hat{F}_n$ to denote the empirical distribution which puts mass $\frac{1}{n}$ on each of X_i's. An estimate of the parameter $\theta = \int xdF(x)$ is the sample mean

$$\hat{\theta} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Now suppose we want to know the variance of our estimators $\hat{\theta}$ usually depends on the unknown distribution F. For example, when $\hat{\theta}$ is the sample mean, we have

$$V_F(\hat{\theta}) = \frac{\sigma^2}{n}$$

where

$$\sigma^2 = \int (x - \int ydF(y))^2 dF(x).$$

The basic bootstrap has two steps :

1. Estimate $V_F(\hat{\theta})$ with $V_{\hat{F}_n}(\hat{\theta})$.

2. Approximate $V_{\hat{F}_n}$ using simulation.

Note that for simple estimator $\hat{\theta}$ we may be able to directly calculate $V_{\hat{F}_n}(\hat{\theta})$ without using simulation. For example, when $\hat{\theta} = \hat{X}_n = \sum_{i=1}^n \frac{X_i}{n}$, we have

$$V_{\hat{F}_n}(\hat{\theta}) = \frac{\hat{\sigma}^2}{n} = \frac{1}{n}\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

However, in more complicated cases it is not easy to write down the formula of $V_{\hat{F}_n}(\hat{\theta})$ and we need to resort to simulation.

1 Simulation

Suppose that $Y \sim H$ and we want to estimate $E(h(Y))$. We can draw i.i.d. samples $Y_1, Y_2, \cdots, Y_B$ from H and use the sample mean

$$\frac{1}{B}\sum_{j=1}^B h(Y_j) \to E(h(Y)) = \int h(y)dH(y).$$

as $B \to \infty$. In particular, we will use the following result that

$$\frac{1}{B}\sum_{j=1}^B (Y_j - \bar{Y})^2 = \frac{1}{B}\sum_{j=1}^B Y_j^2 - (\frac{1}{B}\sum_{i=1}^n Y_j)^2 \xrightarrow{P} V(Y).$$

In bootstrap we want to estimate $V_{\hat{F}_n}(\hat{\theta})$ which stands for "the variance of $\hat{\theta}$ if the true population distribution is $\hat{F}_n$", and recall that $\hat{\theta} = T(X_1, X_2, \cdots, X_n))$. Now think of $\hat{\theta}$ as Y in the above example (i.e. $Y = T(X_1^*, \cdots, X_n^*))$ and the distribution G of Y in this case is the empirical distribution of all samples $(X_1^*, \cdots, X_n^*)$ whose elements are drawn i.i.d., from $\hat{F}_n$. As a result we have the following bootstrap procedure:

(1) For $b = 1, 2, \cdots, B$:

1. Draw $X_1^*, \cdots, X_n^* \sim \hat{F}_n$
2. Compute $\hat{\theta}^*{}_b = T(X_1^*, \cdots, X_n^*)$

Compute $v_{boot} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b^* - \frac{1}{B}\sum_{c=1}^{B}\hat{\theta}_c^*)^2$

Bagging* : So far, we have investigated the bootstrap as a means to access estimation accuracy. An interesting question is whether the bootstrap can improve accuracy. Bagging is an attempt to do this. Bagging is a acronym meaning "bootstrap aggregation". The idea is simple. Suppose we are estimating some quantity, e.g., the optimal portfolio weights to achieve an expected return of 0.012. We have one estimate from the original sample, and this estimate is often used. However, we also have B additional estimates, one from each of the bootstrap samples. The bagging estimate is the average of all of these bootstrap estimates.

2 Confidence Interval:

There are different types of bootstrap confidence intervals available in the literature. We are going to discuss only two types.

- Bootstrap percentile confidence interval
- Bootstrap-t confidence interval

3 Bootstrap percentile confidence interval:

1. Draw $X_1^*, \cdots, X_n^* \sim \hat{F}_n$

2. Compute $\hat{\theta}^*{}_b = T(X_1^*, \cdots, X_n^*)$ for $b = 1(1)B$

3. Compute non-parametric confidence interval based on $\hat{\theta}^*{}_b$ points.

4. Let $\hat{\theta^*}_{(r)}$ is the r-th order statistics.

5. Therefore the $100(1-\alpha)\%$ confidence interval will be $[\hat{\theta^*}_{(k)}, \hat{\theta^*}_{(k')}]$ where $k = [\frac{\alpha}{2}b]$ if $\frac{\alpha}{2}b$ is integer and $k = [\frac{\alpha}{2}b] + 1$ if $\frac{\alpha}{2}b$ is not an integer. Similarly, $k^{'} = [(1-\frac{\alpha}{2})b]$ if $(1-\frac{\alpha}{2})b$ is integer and $k^{'} = [(1-\frac{\alpha}{2})b] + 1$ if $(1-\frac{\alpha}{2})b$ is not an integer.

4 Bootstrap-t confidence interval:

Confidence interval for mean :

See Rupert page 328.

We can construct a $(1-\alpha)\%$ confidence interval as $[\bar{X} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}]$ when X_i's are i.i.d. sample from $N(\mu, \sigma^2)$. But the problem will occur if we are not sampling from normal distribution, but rather some other distribution. In that case the following bootstrap confidence interval can be constructed. Let $\bar{X}_{boot,b}$ and $s_{boot,b}$ be the sample mean and standard deviation of the b-th resample, $b = 1, \cdots, B$. Define

$$t_{boot,b} = \frac{\bar{X} - \bar{X}_{boot,b}}{\frac{s_{boot,b}}{\sqrt{n}}}$$

Notice that $t_{boot,b}$ is defined in the same way as t except for two changes. First, $\bar{X}$ and s in t are replaced by $\bar{X}_{boot,b}$ and $s_{boot,b}$. Second, μ in t is replaced by $\bar{X}$ in $t_{boot,b}$. The last point is a bit subtle, and you should stop to think about it. A resample is taken using the original sample as the population. Thus, for the resample, the population mean is $\bar{X}$!.

Because the resamples are independent of each other, the collection $t_{boot,1}, t_{boot,1}, \cdots$ can be treated as a random sample from the distribution of the $t - statistic$. After B values of $t_{boot,b}$ have been calculated, one from each resample, we find the $100(1-\frac{\alpha}{2})\%$ and $100(1-\frac{\alpha}{2})\%$ percentiles of this collection of $t_{boot,b}$ values. Call these percentiles t_L and t_U. More specifically, we find t_U and t_L as we described earlier. We sort all the B values from smallest to largest. Then we calculate the $B\alpha/2$ and round to the nearest integer. Suppose the result is K_L. Then the K_L-th sorted value of $t_{boot,b}$ is t_L. Similarly, let K_U be $B(1-\frac{\alpha}{2})$ rounded to the nearest integer and then t_U is the K_Uth sorted value of $t_{boot,b}$. Finally we can make the bootstrap confidence interval for μ as $(\bar{X}+t_L\frac{s}{\sqrt{n}}, \bar{X}+t_U\frac{s}{\sqrt{n}})$. We get two advantages through bootstrap:

- We do not need to know the population distribution.
- We do not need to calculate the distribution of t-statistic using probability theory.

See more examples and applications in Rupert.

4.1 Jackknife Estimator:

Jackknifing, which is similar to bootstrapping, is used in statistical inference to estimate the bias and standard error (variance) of a statistic, when a random sample of observations is used to calculate it. The basic idea behind the jackknife estimator lies in systematically recomputing the statistic estimate leaving out one or more observations at a time from the sample set. Therefore, in delete-1 jackknife the resamples for the sample (X_1, X_2, X_3) can be given by , (X_2, X_3), (X_1, X_3) and (X_1, X_2).

Suppose, $\hat{\theta}_b^* \quad b = 1(1)n$ are the estimators based on jackknife resamples of size (n-1) from the original sample of size n. Then jackknife estimate of θ can be written as

$$\hat{\theta^*_{avg}} = \frac{\sum_{b=1}^{n} \hat{\theta}_b^*}{n}$$

and standard error of $\hat{\theta}$ can be given as $[\frac{n-1}{n}\sum_{b=1}^{n}(\theta_b^* - \hat{\theta^*_{avg}})^2]^{\frac{1}{2}}$

Remark : delete-1 jackknife is similar to cross-validation which is discussed later.