

Bootstrapping

Bootstrapping [1] is a simple computational approach to assess the statistical performance of estimators. The paper by Politis [2] is an excellent tutorial on this subject. A variety of applications to signal processing are described by Zoubir and Boashash [3]. The bootstrap has also been extensively used in geophysics, biomedical engineering, control, applied mechanics, and regression analysis.

1 Basic Problem

Given $\underline{Y} = (Y_1, Y_2, \dots, Y_n)$ drawn iid from a (partially or even completely unknown) distribution P , estimate some parameter θ of P . The estimator $\hat{\theta}$ is based on a statistic $T(\underline{Y})$. Then produce some measure of statistical accuracy of the estimator, e.g., its mean, variance, confidence interval, or even full distribution.

For instance, θ could be a moment of P , or the median of P , etc.

For large sample size n and under regularity conditions, asymptotic theory for maximum-likelihood (ML) estimators tells us that the normalized estimation error $\sqrt{n}(\hat{\theta}_{ML} - \theta)$ converges in distribution to $\mathcal{N}(0, 1/I_\theta)$, where I_θ is Fisher information. Since P is unknown, the estimator $\hat{\theta}$ assumed above is generally not the ML estimator. Under regularity conditions, the estimator $\hat{\theta}$ will often be consistent in probability (like the ML estimator), but won't be asymptotically efficient, i.e., its asymptotic variance will exceed $1/I_\theta$. For instance, assume P has mean θ and finite variance. Then the natural estimator $\hat{\theta} = \frac{1}{n}T(\underline{Y}) = \frac{1}{n} \sum_{i=1}^n Y_i$ is consistent but not efficient – unless P is Gaussian with known variance.

In Sections 2 and 3 we consider two approaches to measure the accuracy of the estimator: one that requires acquiring many new samples, and a modification that doesn't. The latter approach is the bootstrapping method.

2 Traditional Monte-Carlo Approach

Assume we draw n iid samples $\underline{Y}^{(k)}$ from P , for $1 \leq k \leq K$. That is, we acquire a total of Kn iid samples from P . For each dataset $\underline{Y}^{(k)}$ we compute an estimate $\hat{\theta}^{(k)}$. For large K , the bias and variance, and cumulative density function (cdf) of the estimator are approximately given by

$$\begin{aligned} \text{Bias}(\hat{\theta}) &\approx \widehat{\text{Bias}}(\hat{\theta}) = \frac{1}{K} \sum_{k=1}^K \hat{\theta}^{(k)} - \theta \\ \text{Var}(\hat{\theta}) &\approx \widehat{\text{Var}}(\hat{\theta}) = \frac{1}{K-1} \sum_{k=1}^K \left(\hat{\theta}^{(k)}\right)^2 - \left(\frac{1}{K} \sum_{k=1}^K \hat{\theta}^{(k)}\right)^2. \end{aligned}$$

(Note that $\widehat{\text{Bias}}(\hat{\theta})$ above is not a valid estimator of the bias because evaluating it requires knowledge of θ !) The cumulative density function (cdf) of $\hat{\theta}$ is given by

$$Q(x) = P[\hat{\theta} \leq x]$$

and may be estimated by the empirical cdf,

$$\hat{Q}(x) = \frac{1}{K} \sum_{k=1}^K 1_{\{\hat{\theta}^{(k)} \leq x\}}.$$

3 Bootstrap

The basic idea of bootstrapping is to use the procedure of Section 2 with the following modification. Instead of drawing $\underline{Y}^{(k)}$ from P , draw it from the empirical distribution

$$\hat{P}(y) = \frac{1}{n} \sum_{i=1}^n 1_{\{Y_i \leq y\}}$$

based on the n samples Y_1, \dots, Y_n . In other words, draw $Y_i^{(k)}$ iid (with replacement) from the uniform distribution over the observed dataset $\{Y_1, \dots, Y_n\}$. We denote by $\underline{Y}^{*(k)}$, $1 \leq k \leq K$, the *resamples* obtained from this procedure and by $\hat{\theta}^{*(k)}$ the estimate computed from $\underline{Y}^{*(k)}$. Then we form the following estimator of the bias and variance of $\hat{\theta}$:

$$\begin{aligned} \text{Bias}^*(\hat{\theta}) &= \frac{1}{K} \sum_{k=1}^K \hat{\theta}^{*(k)} - \theta \\ \text{Var}^*(\hat{\theta}) &= \frac{1}{K} \sum_{k=1}^K \left(\hat{\theta}^{*(k)}\right)^2 - \left(\frac{1}{K} \sum_{k=1}^K \hat{\theta}^{*(k)}\right)^2. \end{aligned}$$

The estimated cdf for the estimator is

$$\hat{Q}^*(x) = \frac{1}{K} \sum_{k=1}^K 1_{\{\hat{\theta}^{*(k)} \leq x\}}.$$

Example. Let P be any distribution over the ten digits $0, 1, \dots, 9$, and θ the median of the distribution. The estimator is simply $\hat{\theta} = \text{med}(Y_1, \dots, Y_n)$. Say that for $n = 5$ we observe $\underline{Y} = (5, 2, 2, 9, 7)$. We use bootstrapping with $K = 2$. Say we obtain the resamples $\underline{Y}^{*(1)} = (9, 2, 9, 5, 2)$ and $\underline{Y}^{*(2)} = (2, 7, 7, 5, 9)$. The corresponding median estimates are $\hat{\theta}^{*(1)} = 5$ and $\hat{\theta}^{*(2)} = 7$. Therefore the bootstrap estimate of the mean of $\hat{\theta}$ is $\frac{1}{2}(5 + 7) = 6$, and its variance is 1.

When does the bootstrap work? For large n , the empirical cdf estimator \hat{P} converges in the sup norm to the actual P :

$$\sup_y |\hat{P}(y) - P(y)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \rightarrow \infty.$$

Thus one may heuristically expect that the bootstrap is approximately equivalent to the procedure of Sec. 2. In fact a little bit more is required, e.g., a sufficient condition is that $\hat{\theta}$ be asymptotically Gaussian.

An example where the bootstrap fails is that of iid uniform Y over the interval $[0, \theta]$, using the estimator $\hat{\theta}(\underline{Y}) = \max_{1 \leq i \leq n} Y_i$.

4 Extensions

The bootstrap principle may also be used in problems where the data Y_i , $1 \leq i \leq n$, are not iid. A modification of the method of Sec. 3 is needed. The approach is best illustrated with an example.

Example. Consider the model

$$Y_i = a + bs_i + W_i, \quad 1 \leq i \leq n$$

where $\theta = (a, b)$ is to be estimated, s_i is a known signal, and W_i is iid noise with mean zero and finite variance. Consider the least-squares estimator

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \sum_{i=1}^n (Y_i - a - bs_i)^2$$

which is consistent under regularity assumptions on the signal. Note this estimator coincides with the ML estimator if the noise distribution is Gaussian with known variance. Define the residual errors

$$E_i = Y_i - \hat{a} - \hat{b}s_i, \quad 1 \leq i \leq n,$$

which follow approximately the same iid distribution as $\{W_i\}$ if \hat{a}, \hat{b} are accurate estimators of a, b . Define the *centered* residuals

$$\tilde{E}_i = E_i - \frac{1}{n} \sum_{j=1}^n E_j, \quad 1 \leq i \leq n.$$

The bootstrap principle is applied as follows.

For $1 \leq k \leq K$, do the following:

1. Draw resamples $E_i^{*(k)}$ iid uniformly with replacement from $\{\tilde{E}_1, \dots, \tilde{E}_n\}$.

2. Generate pseudo-data $Y_i^{*(k)} = \hat{a} + \hat{b}s_i + E_i^{*(k)}$ for $1 \leq i \leq n$.
3. Apply least-squares estimation to obtain

$$(\hat{a}^{*(k)}, \hat{b}^{*(k)}) = \arg \min_{(a,b)} \sum_{i=1}^n (Y_i^{*(k)} - a - bs_i)^2.$$

Then one may obtain the desired estimates of the bias, variance, etc. of the estimator from $(\hat{a}^{*(k)}, \hat{b}^{*(k)})$, $1 \leq k \leq K$.

Related Problems. The same idea applies to problems such as estimation of parameters in a Markov model:

$$Y_i = F_\theta(Y_{i-1}, Y_{i-2}) + W_i, \quad 1 \leq i \leq n$$

where W_i is iid noise with mean zero and finite variance. The prediction function F is parameterized by θ , which can be estimated using nonlinear least-squares:

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=3}^n (Y_i - F_\theta(Y_{i-1}, Y_{i-2}))^2.$$

The bootstrap has also been used to assess the accuracy of spectrum estimators, hypothesis tests, etc.

References

- [1] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, pp. 1—26, Jan. 1979.
- [2] D. Politis, "Computer-intensive methods in statistical analysis," *IEEE Signal Processing Magazine*, pp. 39—55, Jan. 1998.
- [3] A. Zoubir and B. Boashash, "The Bootstrap and its Application in Signal Processing," *IEEE Signal Processing Magazine*, pp. 56—76, Jan. 1998.