Resampling Methods for Time Series

Overview

1. Bootstrap resampling
2. Effects of dependence
3. Subsampling
4. Further directions

Bootstrap resampling

Main idea Estimate the sampling distribution of a statistic, with particular emphasis on the standard error of the statistic and finding a confidence interval for a statistic.

The idea is to get these properties from the data at hand by analogy to the usual "thought experiment" that motivates the sampling distribution. The sampling distribution of a statistic $\hat{\theta}$ is the distribution of the $\hat{\theta}$ computed from repeated samples X of size n from some distribution F. The trick is to sample from the data itself rather than the population.

Bootstrap sample is a sample drawn *with replacement* from the original sample X, denoted by $X^* \sim F_n$ where F_n is the empirical distribution of the observed sample X, the step function

$$F_n(x) = \tfrac{1}{n} \sum_i H_{X_i}(x) \quad \text{where} \quad H_c(x) = \begin{cases} 0 & \text{if } x \le c, \\ 1 & \text{if } x > c\,. \end{cases} \tag{1}$$

Some call H_c the heaviside function. For now, X^* is the same size as X, so obviously X^* is going to have lots of ties. (If the X_i are unique, then about 63% of the individual values show up in X^*; the other 37% are duplicates.)

Example Suppose we didn't know the CLT and could not figure out the standard error of a mean. Consider the following iterations: for $b = 1, 2, \ldots, B$,

1. Draw a bootstrap sample X_b^* of size n from F_n.
2. Compute the statistic $\hat\theta_b^*$ from X_b^*

Compute the bootstrap standard error as the standard deviation of the $\hat\theta^*$

$$se_B^*(\hat\theta) = \tfrac{1}{B-1}\sum(\hat\theta_b^* - \overline{\hat\theta^*})^2$$

and for a, say, 95% confidence interval for θ, use the 2.5% and 97.5% quantiles of the $\hat\theta^*$. (Just put them in order and pick off the needed percentiles.)

Does this work? Yes. Showing that the percentiles give the right answer is harder, but it is easy for a linear statistic like the mean to figure out what the simulation of the standard error estimates. (Bootstrap resampling is not about simulation, it's about using the empirical distribution F_n in place of F. We simulate because its easy and because we often need this for statistics that are not linear.)

First, define expectation with respect to F_n as

$$\mathbb{E}^*\, X_i^* = X_1\tfrac{1}{n} + X_2\tfrac{1}{n} + \cdots + X_n\tfrac{1}{n} = \overline{X}$$

Using this operator, we see no bias in the mean,

$$\mathbb{E}^*\, \overline{X}^* = (\overline{X} + \overline{X} + \cdots + \overline{X})/n = \overline{X}\ .$$

The squared sampling variation is

$$\mathrm{Var}^*(X_i^*) = \hat\sigma^2 = \tfrac{1}{n}\sum(X_i - \overline{X})^2$$

(Var^* is the variance with respect to F_n) and

$$\mathrm{Var}^*(\overline{X}^*) = \frac{\hat\sigma^2}{n}$$

rather than s^2/n. For confidence intervals, see the associated **R** example.

Key analogy is that resampling from the data X ought to resemble — in the sense of reproducing the sampling distribution — sampling from the population. Let $\hat\theta^*$ denote the value of the statistic when computed from the bootstrap sample X^*. The analogy can then be expressed as

$$\hat\theta^* : \hat\theta \ :: \ \hat\theta : \theta$$

Some math Some math suggests why this might work more generally (we don't need it for linear statistics). Think of the statistic $\hat{\theta}$ as a function of the empirical distribution. For example, the sample mean $\overline{X}$ can be written as

$$\overline{X} = T(F_n) = \int x \, dF_n(x)$$

if you're willing to think in terms of point measures (integrating with respect to a step function). The parameter θ (in this case the mean) has a similar representation

$$\mu = T(F) = \int x \, dF(x) \ .$$

Now consider two linear expansions, thinking of T as an operator on distributions and T' as a derivative (Frechet),

$$\hat{\theta}^* = T(F_n^*) \approx T(F_n) + T'(F_n)(F_n^* - F_n)$$

and

$$\hat{\theta} = T(F_n) \approx T(F) + T'(F)(F_n - F)$$

You get a hint that things are going in a nice direction because both $F_n - F$ and $F_n^* - F_n$ tend to a Brownian bridge as $n \to \infty$. Of course, we also need for the derivatives to be close as well.

Reference Efron and Tibshirani (1993) *An Introduction to the Bootstrap*, Chapman and Hall.

Effects of Dependence

Sampling distribution of mean If the data are a realization of a stationary process, we know that

$$\text{Var}(\overline{X}) = \tfrac{1}{n} \sum (n - |h|)\gamma(h) \ .$$

What's the bootstrap going to do in this case?

Naive resampling fails Sampling with replacement creates bootstrap samples with independent observations,

$$\text{Var}(X_i^*) = \hat{\sigma}^2, \quad \text{Var}^*(\overline{X}^*) = \frac{\hat{\sigma}^2}{n}$$

as before, and $\hat{\sigma}^2$ will be close to $\gamma(0)$. That's not the right answer.

Parametric bootstrap Suppose we know that the underlying process is AR(1). Then we can estimate the parameters and general bootstrap data as

$$X_t^* = \hat{\phi} X_{t-1}^* + w_t^*$$

where $w_t^* \sim G_n$ and G_n is the empirical distribution of the estimated model residuals. This works, but requires that we know the data generating process.

The use of long autoregressions in estimation suggests a workable approach, knowns as a *sieve bootstrap*. The idea is simple. Fit an AR model with a large number of lags and use that model to generate the bootstrap replications. The success clearly depends on how well the AR model captures the dependence and the ratio of the length n to the number of AR coefficients p.

Parameter bias The estimates of the AR coefficients are biased. Hence the "true" model that generates the bootstrap series is shifted away from the actual generating process. The effects of the bias are most pronounced for coefficients of processes with zeros near the unit circle.

Subsampling

Nonparametric Subsampling relies on the same ideas that we used in proving the CLT for a stationary process: arrange the data into blocks, and rely on the blocks becoming less dependent as they get farther apart. The key assumption is that the distribution of the statistic has the form

$$\sqrt{n}(\hat{\theta} - \theta) \sim G$$

Procedure

1. Arrange the time series $X_1, X_2, \ldots, X_n$ into N *overlapping* blocks, each of length b, with overlap ℓ:

$$\{X_1, \ldots, X_b\}, \{X_{1+\ell}, \ldots, X_{b+\ell}, \ldots$$

2. Treat each of the N blocks as if it were the time series of interest, computing the relevant statistic, obtaining $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_N$.

3. Estimate the sampling properties of the estimator computed from the original time series

$$\sqrt{n}(\hat\theta_n - \theta)$$

l from the *rescaled* empirical distribution of the $\tilde\theta_i$,

$$\tilde F_b(x) = \tfrac{1}{N} \sum_i H_{\sqrt{b}(\tilde\theta_i - \hat\theta)}(x)$$

Theory The coverage probabilities implied by the distribution $\tilde F_b$ converges in probability to the correct limits under relatively weak conditions (see reference), so long as the $b \to \infty$ with $b/n \to 0$ (*e.g.*, $b = O(\sqrt{n})$). For iid samples, the theory resembles that used when studying leave-out-several versions of the jackknife and cross-validation. Subsampling provably works in many applications with the type of "weak" dependence associated with ARMA processes.

Not independent The subsampled estimates $\tilde\theta$ are not independent (or even uncorrelated), so we cannot use them directly to estimate the variance, say, of $\hat\theta$. The computation of the empirical distribution $\tilde F_b$ only relies on the expected values of the indicators. Still, it feels odd: How can you get the distribution correct without getting the variance right?

Questions

1. How long to make the blocks?
   A bias/variance trade-off.
2. How much should the blocks overlap?

Reference Politis, Romano, and Wolf (1999) *Subsampling.* Springer.

Further ideas

Prediction intervals Bootstrap prediction intervals avoid assumptions about the distribution of the model errors.