Jackknife Methods

Suppose a sample $X_1, X_2, \ldots, X_n$ is to be used to estimate a population parameter, θ.

We form a statistic T that estimates θ. In order to do much with this estimate, we need to know its variance (or more generally, its sampling distribution); if it is biased, we would like to correct its bias.

Suppose we delete one observation, say X_j, from the sample and compute our estimate from the reduced sample. Let $T_{(-j)}$ denote the estimator computed from the sample with the j^{th} observation removed.

If T is the mean, for example,

$$T_{-j} = \sum_{i \neq j} X_i/(n-1).$$

Note

$$X_j = nT - (n-1)T_{-j}.$$

Jackknife Estimators

The mean of the $T_{(-j)}$,

$$\overline{T}_{(\bullet)} = \frac{1}{r} \sum_{i=1}^{r} T_{(-j)}$$

can be used as an estimator of θ.

The $T_j^* = rT - (r-1)T_{(-j)}$ are called "pseudo-values".

The average of the pseudo-values, $\overline{T}^*$, is called the jackknife estimator.

Note

$$\overline{T}^* = rT + (r-1)\overline{T}_{(\bullet)}.$$

The Jackknife Variance Estimate

The sample variance of the pseudo-values can be used as an estimate of the variance of the jackknife estimate. (The pseudo-values are *not* independent, of course.)

How good is it? Many Monte Carlo studies. ... generally appears conservative.

Jackknifed Bias Reduction

Similar treatment of the pseudo-values leads to an estimate of the bias of the original estimator T:

$$(n-1)(\overline{T}_{(\bullet)} - T)$$

and, hence, to the bias-corrected estimator:

$$nT - (n-1)\overline{T}_{(\bullet)}$$

Bootstrap Methods

Suppose a sample $X_1, X_2, \ldots, X_n$ is to be used to estimate a population parameter, θ.

We form a statistic T that estimates θ. Our interest is then in the sampling distribution of T.

It's often intractable.

The basic idea of the bootstrap is that the true population can be approximated by an infinite population in which each of the n sample points are equally likely.

The parameter is a functional of a population distribution function:

$$\Theta = \int g(x) dF(x)$$

The estimator is often the same functional of the empirical distribution function:

$$T = \int g(x) d\hat{F}(x)$$

Various properties of the distribution of T can be estimated by use of “bootstrap samples”.

- bias
- variance and standard deviation
- other moments (lower-order)

The bootstrap is a “resampling” method, as is the jackknife.

Basic Ideas

The problem in its broadest setting is to find a functional f_t (from some class of functionals) that allows us to relate the distribution function of the sample F_1 to the population distribution function F, that is, such that

$$\mathrm{E}\{f_t(F, F_1)|F\} = 0.$$

Suppose we are trying to estimate $h(\int g(x)dF(x))$. The h presents special problems.

For example, suppose we wish to estimate

$$\Theta = \{\int xdF(x)\}^r$$

Start with

$$\hat{\Theta} = \left\{\int xd\hat{F}(x)\right\}^r = \bar{x}^r$$

Biased.

Correcting for the bias is equivalent to finding t that solves the equation

$$f_t(F, F_1) = T(F_1) - \Theta(F) + t$$

so that f_t has zero expectation with respect to F.

What about repeating this whole process?

Take a sample from "population" with distribution function F_1. Look for f_t so that

$$\mathrm{E}\{f_t(F_1, F_2)|F_1\} = 0.$$

The difference is we know more about this equation because we know more about F_1.

This is the *bootstrap principle.*

What is F_1?

Two approaches:

- nonparametric
  F_1 is the empirical distribution function.
  F_2 is the empirical distribution function of a sample drawn at random with replacement from the finite population with distribution function F_1.
- parametric
  F is assumed known up to a finite set of unknown parameters, λ, then F_1 is F with λ replaced by its sample estimates (of some kind).
  F_2 is similar, except the estimates are the same function of a sample from a population with distribution function F_1 (not just simply drawn with replacement from the original sample).

Example: Bias Reduction

Find f_t (i.e., t) so that

$$\mathrm{E}\{f_t(F, F_1)|F\} = 0$$

or

$$\mathrm{E}\{T(F_1) - \Theta(F) + t|F\} = 0.$$

Change the problem to the sample:

$$\mathrm{E}\{T(F_2) - T(F_1) + t_1|F_1\} = 0,$$

whose solution is

$$t_1 = T(F_1) - \mathrm{E}\{T(F_2)|F_1\},$$

so the bias-reduced estimate is

$$T_1 = 2T(F_1) - \mathrm{E}\{T(F_2)|F_1\}.$$

We may be able to compute $\mathrm{E}\{T(F_2)|F_1\}$.
If so, do so.

If not, estimate by Monte Carlo.

The Mechanics of the Nonparametric Bootstrap

The basic bootstrap procedure is to take m random samples each of size n, and *with replacement* from the given set of data, the original sample $X_1, X_2, \ldots, X_n$; and for each sample, compute an estimate T_j of the same functional form as the original estimator T. The distribution of the T_j's is related to the distribution of T. The variability of T about Θ can be assessed by the variability of T_j about T; the bias of T can be assessed by the mean of $T_j - T$.