Stat 5102 Lecture Slides
Deck 8

Charles J. Geyer
School of Statistics
University of Minnesota

Plug-In and the Bootstrap

The worst mistake one can make in statistics is to confuse the sample and the population or to confuse estimators and parameters. In short, $\widehat{\theta}$ is not θ.

But the plug-in principle (slides 78–84, deck 2 and slides 58–66 and 97, deck 3) seems to say the opposite. Sometimes it is o. k. to just plug in and estimate for an unknown parameter. In particular, it is o. k. to plug in a consistent estimator of the asymptotic variance of a parameter in forming asymptotic confidence intervals for that parameter.

So it is a terrible mistake to confuse a parameter of interest and an estimator for it, but it may not be a mistake to ignore the difference between a nuisance parameter and an estimator for it.

Plug-In and the Bootstrap (cont.)

The "bootstrap" is a cute name for a vast generalization of the plug-in principle.

The name comes from the cliche "pull oneself up by one's bootstraps" which although it describes a literal impossibility actually means succeed by one's own efforts.

In statistics, the hint of impossibility is part of the flavor. The bootstrap seems problematic, but it (usually) works.

The Nonparametric Bootstrap

The bootstrap comes in two flavors, parametric and nonparametric. We'll do the latter first.

The theory of the nonparametric bootstrap is all above the level of this course, so we give a non-theoretical explanation.

The nonparametric bootstrap, considered non-theoretically, is just an analogy.

The Nonparametric Bootstrap (cont.)

	Real World	Bootstrap World
true distribution	F	$\widehat{F}_n$
data	$X_1, \ldots, X_n$ IID F	$X_1^*, \ldots, X_n^*$ IID $\widehat{F}_n$
empirical distribution	$\widehat{F}_n$	F_n^*
parameter	$\theta = t(F)$	$\widehat{\theta}_n = t(\widehat{F}_n)$
estimator	$\widehat{\theta}_n = t(\widehat{F}_n)$	$\theta_n^* = t(F_n^*)$
error	$\widehat{\theta}_n - \theta$	$\theta_n^* - \widehat{\theta}_n$
standardized error	$\frac{\widehat{\theta}_n - \theta}{s(\widehat{F}_n)}$	$\frac{\theta_n^* - \widehat{\theta}_n}{s(F_n^*)}$

Objects on the same line are analogous. The notation $\theta = t(F)$ means θ is some function of the true unknown distribution.

The Nonparametric Bootstrap (cont.)

The notation X_1^*, ..., X_n^* IID $\widehat{F}_n$ means X_1^*, ..., X_n^* are independent and identically distributed from the empirical distribution of the real data.

Sampling from the empirical distribution is just like sampling from a finite population, where the "population" is the real data X_1, ..., X_n. To be IID sampling must be with replacement.

X_1^*, ..., X_n^* are a sample with replacement from X_1, ..., X_n. For short, this is called *resampling*.

The Nonparametric Bootstrap (cont.)

We want to know the sampling distribution of $\widehat{\theta}_n$ or of $\widehat{\theta}_n - \theta$ or of $(\widehat{\theta}_n - \theta)/s(\widehat{F}_n)$.

This sampling distribution depends on the true unknown distribution F of the real data.

It also may be very difficult or impossible to calculate theoretically. Even asymptotic approximation may be difficult, if the parameter $\theta = t(F)$ is a sufficiently complicated function of the true unknown F.

The statistical theory we have covered is quite amazing in what it does, but there is a lot it doesn't do.

The Nonparametric Bootstrap (cont.)

In the "bootstrap world" everything is known. $\widehat{F}_n$ plays the role of the true unknown distribution, and $\widehat{\theta}_n$ plays the role of the true unknown parameter value.

The sampling distribution of θ_n^* or of $\theta_n^* - \widehat{\theta}_n$ or of $(\theta_n^* - \widehat{\theta}_n)/s(F_n^*)$ may still be difficult to calculate theoretically, but it can always be "calculated" by simulation.

See computer examples web page for example.

The Nonparametric Bootstrap (cont.)

Much folklore about the bootstrap is misleading.

The bootstrap is large sample, approximate, asymptotic. It is not an exact method.

The bootstrap analogy works when the empirical distribution $\widehat{F}_n$ is close to the true unknown distribution F. This will usually be the case when the sample size n is large and not otherwise.

Bootstrap Percentile Intervals

The simplest method of making confidence intervals for the unknown parameter is to take $\alpha/2$ and $1 - \alpha/2$ quantiles of the bootstrap distribution of the estimator θ_n^* as endpoints of the $100(1 - \alpha)\%$ confidence interval.

See computer examples web page for example.

The percentile method only makes sense when there is a symmetrizing transformation (some function of $\widehat{\theta}_n$ has an approximately symmetric distribution with the center of symmetry being the true unknown parameter value θ. The symmetrizing transformation does not have to be known, but it does have to exist.

The Parametric Bootstrap

The parametric bootstrap is just like the nonparametric bootstrap except for one difference in the analogy. We use a parametric model $F_{\widehat{\theta}_n}$ rather than the empirical distribution $\widehat{F}_n$ as the analog of the true unknown distribution in the bootstrap world. Thus the analogy looks like

	Real World	Bootstrap World
parameter	θ	$\widehat{\theta}_n$
true distribution	F	$F_{\widehat{\theta}_n}$
data	$X_1, \ldots, X_n$ IID F	$X_1^*, \ldots, X_n^*$ IID $F_{\widehat{\theta}_n}$
estimator	$\widehat{\theta}_n = t(X_1, \ldots, X_n)$	$\theta_n^* = t(X_1^*, \ldots, X_n^*)$
error	$\widehat{\theta}_n - \theta$	$\theta_n^* - \widehat{\theta}_n$
standardized error	$\frac{\widehat{\theta}_n - \theta}{s(X_1, \ldots, X_n)}$	$\frac{\theta_n^* - \widehat{\theta}_n}{s(X_1^*, \ldots, X_n^*)}$

The Parametric Bootstrap (cont.)

Simulation from the parametric model $F_{\widehat{\theta}_n}$ not analogous to finite population sampling and does not resample the data like the nonparametric bootstrap does. Instead we simulate the parametric model. This may be easy (when R has a function to provide such random simulations) or difficult.

See computer examples web page for example.

Nonparametric versus Parametric Bootstrap

The nonparametric bootstrap is nonparametric (surprise!). That means it always does the right thing, except when it doesn't. It doesn't work when the sample size is too small or when the square root law doesn't hold or when the data are not IID or when various technical issues arise that are beyond the scope of this course — the parameter $\theta = t(F)$ is not a nice enough function of the true unknown distribution, but we cannot define the appropriate notion of "nice" nor explain why this matters.

The parametric bootstrap is parametric (surprise!). That means it is always wrong when the model is wrong (does not contain the true unknown distribution). On the other hand, when the parametric bootstrap does the right thing (when the statistical model is correct), it does a much better job at smaller sample sizes than the nonparametric bootstrap.

Nonparametric versus Parametric Bootstrap (cont.)

When the parameter θ is defined in terms of the parametric statistical model and can only be estimated using the parametric model (by maximum likelihood perhaps), the statistical model is needs to be correct for the parameter estimate $\widehat{\theta}_n$ to make sense.

Since we already need the statistical model to be correct, the parametric bootstrap is the logical choice.