5601 Notes: The Subsampling Bootstrap

Charles J. Geyer

November 21, 2002

1 Web Page

This handout accompanies the web pages

```
http://www.stat.umn.edu/geyer/5601/examp/subboot.html
http://www.stat.umn.edu/geyer/5601/examp/subtoot.html
```

2 History

The term "bootstrap" was coined by Efron (1979). He described both the nonparametric and parametric bootstrap. In particular, his nonparametric bootstrap is the procedure of resampling *with replacement* from the original sample *at the same sample size* that is by far the most commonly used bootstrap procedure.

It wasn't long before people experimented with resampling at different sample sizes. But the key discovery in that area came later. Politis and Romano (1994) described resampling *without replacement* from the original sample *at smaller than the original sample size.*

This is different enough from Efron's idea that in their book (Politis, et al., 1999) they don't call it "bootstrap" but just plain "subsampling".

Whatever you call it, here's why it is such an important innovation.

- Politis and Romano's subsampling bootstrap takes samples *without replacement* of size b from the original sample of size n, generally with $b \ll n$ (read "b much less than n"). Such samples are themselves samples of size b from the true unknown distribution F of the original sample.

- Efron's nonparametric bootstrap takes samples *with replacement* of size n from the original sample of size n (both sample sizes the same). Such samples are samples of size n from the empirical distribution $\widehat{F}_n$ associated with the original sample.

Each of these procedures does the Wrong Thing.

- The Right Thing is samples of the right size n from the right distribution F.

1

- The Politis and Romano thing is samples of the wrong size $b \ll n$ from the right distribution F.

- The Efron thing is samples of the right size n from the wrong distribution $\widehat{F}_n$.

Both Wrong Things are wrong. We would like to do the Right Thing but we can't. (More precisely, we have exactly one such sample, the original data, and can't get more. Scientists may get more data, but that's of no interest to us statisticians.)

So which Wrong Thing do we want to do? Both have pluses and minuses. The Efron procedure is older, more widely used, and familiar to more people. It is also easier to use, at least in simple situations. But the Politis and Romano procedure has the great virtue of working in situations where the Efron bootstrap fails. The two main classes of such situations are presented in the following sections.

3   Stationary Time Series

A *time series* is a sequence X_1, X_2, ..., X_n of *dependent* (note *not* independent) random variables.

A times series is *stationary* if every consecutive block

$$X_{i+1}, X_{i+2}, \ldots, X_{i+b} \tag{1}$$

of length b has the same (marginal) distribution. Roughly speaking, what actually happens changes over time, but the *probability distribution of what happens* does not change over time.

The ordinary (Efron) nonparametric bootstrap doesn't work for time series or any other form of dependent data. If the data are not i. i. d. then it makes no sense whatsoever to obtain i. i. d. bootstrap samples from $\widehat{F}_n$ (or any other distribution for that matter).

The (Politis and Romano) subsampling bootstrap does work for stationary time series. Under the stationarity assumption the $n - b - 1$ consecutive blocks (1) of length b are identically distributed. Hence the estimators corresponding to such blocks

$$\theta^*_{bi} = t(X_{i+1}, X_{i+2}, \ldots X_{i+b})$$

are identically distributed (not i. i. d., just i. d., since they are dependent) and analogous (in the "bootstrap world") to

$$\hat{\theta}_n = t(X_1, X_2, \ldots X_n)$$

(in the "real world"). The only problem we have to deal with is that

$$\mathrm{se}(\hat{\theta}_n) \approx \frac{c}{\sqrt{n}}$$

for some positive constant c (assuming our estimator obeys the "square root law") whereas

$$\mathrm{se}(\theta_b^*) \approx \frac{c}{\sqrt{b}}$$

for the same positive constant c (but different denominator). Thus we need to scale

$$\mathrm{se}(\hat{\theta}_n) \approx \mathrm{se}(\theta_b^*) \cdot \sqrt{\frac{b}{n}}$$

to get from $\mathrm{se}(\theta_b^*)$, which we can estimate by subsampling (as the standard deviation of the $n - b + 1$ quantities θ_{bi}^*), to $\mathrm{se}(\hat{\theta}_n)$, which is the quantity we need to make a confidence interval for $\hat{\theta}_n$ and which we otherwise have no way to estimate.

4   Extreme Values

Suppose X_1, X_2, ..., X_n are i. i. d. Uniform$(0, \theta)$ random variables. Since the larger the sample the more the largest values crowd up against θ, the natural estimator of θ is the maximum data value

$$\hat{\theta}_n = X_{(n)} = \max(X_1, X_2, \ldots, X_n).$$

This is in fact the maximum likelihood estimate.

The main statistical interest in this estimator is that it is a counter example to both the "square root law" and the "usual asymptotics" of maximum likelihood.

- The "rate" is n rather than $\sqrt{n}$.

- The asymptotic distribution is not normal.

More precisely,
$$n(\theta - \hat{\theta}_n) \xrightarrow{\mathcal{D}} \mathrm{Exponential}(1/\theta) \tag{2}$$

But to use the subsampling bootstrap, we need only know that the actual rate is n. We do not need to know the actual asymptotic distribution.

Actually, we do not even need to know the rate. By looking at the distribution of θ_b^* for different subsample sizes b we can get an estimate of the rate (described in Section 6 below). But for now we'll assume we know the rate.

5   Confidence Intervals

The fundamental idea of the subsampling bootstrap is that

$$\tau_n(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathrm{Something}, \tag{3}$$

3

where "Something" denotes any distribution whatsoever. Then, trivially,

$$\tau_b(\hat{\theta}_b - \theta) \xrightarrow{\mathcal{D}} \text{Something}, \tag{4}$$

converges to the same "Something," because whether we index by n or b is merely a matter of notation. Usually, we write (4) as

$$\tau_b(\theta_b^* - \theta) \xrightarrow{\mathcal{D}} \text{Something} \tag{5}$$

to distinguish the estimator $\hat{\theta}_n$ for the full data and the estimator θ_b^* for a subsample, but it is the key feature of the subsampling bootstrap that $\hat{\theta}_b$ and θ_b^* are equal in distribution (because the subsampling is done without replacement as discussed in Section 2).

The basic assumptions of the subsampling bootstrap are

$$
\begin{aligned}
b &\to \infty \\
\frac{b}{n} &\to 0 \\
\tau_b &\to \infty \\
\frac{\tau_b}{\tau_n} &\to 0
\end{aligned}
\tag{6}
$$

where n is the sample size and b the subsample size.

Under these assumptions

$$\tau_b(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} 0 \tag{7}$$

just because we would need to multiply by τ_n rather than τ_b to get a nonzero limit and τ_b/τ_n goes to zero.

Subtracting (7) from (5) gives

$$\tau_b(\theta_b^* - \hat{\theta}_n) \xrightarrow{\mathcal{D}} \text{Something}, \tag{8}$$

where "Something" denotes the same distribution as in (3).

To summarize where we have gotten, the subsampling bootstrap is based on the assumptions (6) and the convergence in distribution (3). It then follows from asymptotic theory that (8) describes the same asymptotics as (3).

It does not matter what the limiting distribution is because we approximate it using the subsampling bootstrap. Suppose the limiting distribution, the "Something" in (3) has distribution function F. We don't know the functional form of F but we can approximate it by the empirical distribution function F_b^* of the left hand side of (8) using bootstrap subsampling to simulate θ_b^*.

We know that for large n

$$F^{-1}(\alpha/2) < \tau_n(\hat{\theta}_n - \theta) < F^{-1}(1 - \alpha/2) \tag{9}$$

occurs with probability approximately $1 - \alpha$. That's what the convergence in distribution statement (3) means when F is the distribution function of the

"Something" on the right hand side. $F^{-1}(\alpha/2)$ is the $\alpha/2$ quantile of this distribution and $F^{-1}(1-\alpha/2)$ is the $1-\alpha/2$ quantile. Thus if Y is a random variable having this distribution and the distribution is continuous, the probability that

$$F^{-1}(\alpha/2) < Y < F^{-1}(1 - \alpha/2) \tag{10}$$

is $1 - \alpha$. Since we are assuming Y and $\tau_n(\hat{\theta}_n - \theta)$ have approximately the same distribution for large n, (9) has approximately the same probability as (10).

Of course, we don't know F, but F_b^* converges to F, so for large b and n, we have

$$F_b^{*-1}(\alpha/2) < \tau_n(\hat{\theta}_n - \theta) < F_b^{*-1}(1 - \alpha/2) \tag{11}$$

with probability close to $1 - \alpha$. Rearranging (11) to put θ in the middle by itself gives

$$\hat{\theta}_n - \tau_n^{-1} F_b^{*-1}(1 - \alpha/2) < \theta < \hat{\theta}_n - \tau_n^{-1} F_b^{*-1}(\alpha/2) \tag{12}$$

which is the way subsampling bootstrap confidence intervals are done.

In practice, we don't explicitly calculated empirical c. d. f.'s and invert them. We use the R `quantile` function to directly calculate quantiles. Assuming we already have calculated the estimator `theta.hat` for the original data having sample size `n` and a vector `theta.star` of estimators for bootstrap subsamples of size `b` and have previously defined a function `tau` that calculates the "rate" and a significance level `alpha`, the following three R statements calculate the confidence interval.

```
z.star <- tau(b) * (theta.star - theta.hat)
crit.val <- quantile(z.star, probs = c(1 - alpha / 2, alpha / 2))
theta.hat - crit.val / tau(n)
```

6 Estimating the Rate

We aren't always lucky enough to know the rate of convergence τ_b. But even if we don't, we can estimate the rate from looking at the distribution of θ_b^* for different subsample sizes b. The method described here is that of Chapter 8 of Politis, et al. (1999).

Suppose $\tau_n = n^\beta$ for some constant β. This is the usual case. It includes the "square root law" $(\beta = 1/2)$ found in the usual asymptotics and most other examples of interest. Under this supposition, what we need to do is estimate the unknown constant β.

To do this we take subsamples at different sizes. For each sample size b we look at the distribution of $\theta_b^* - \hat{\theta}_n$. For each such distribution we determine a number of differences of quantiles. On the web page, the quantiles are determined as order statistics. Suppose $\theta_{b,i}^*$ denotes the i-th bootstrap estimator for a subsample of size b and $\theta_{b,(i)}^*$ the i-th order statistic for the estimators for subsamples of size b. Then we define

$$y_{b,i} = \theta_{b,(l_i)}^* - \theta_{b,(k_i)}^*$$

5

where l_i and k_i are some numbers chosen rather haphazardly. However they are chosen, the scaling of the distributions by n^β means

$$y_{b,i} \approx Cn^{-\beta}$$

for some constant C or for that matter

$$\bar{y}_b \approx Cn^{-\beta}$$

where $\bar{y}_b$ is the average of the $y_{b,i}$ for each b. Taking logs gives

$$\log(\bar{y}_b) \approx \log(C) - \beta \log(b)$$

thus we can estimate β by regressing $\log(\bar{y}_b)$ on $\log(b)$. We don't need a regression routine, because this is simple linear regression

$$\hat{\beta} = -\frac{\mathrm{cov}\{\bar{y}_b, \log(b)\}}{\mathrm{var}\{\log(b)\}}$$

does the job.

As the example on the web page shows, this actually works if n and b are large enough.

References

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1–26.

Politis, D. N.; Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22, 2031–2050.

Politis, D. N.; Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer Verlag.