STAT 818D Lecture 6
Bootstrap Theory for Mean-Statistics

Readings for this lecture are from Wassermann (2006) Ch.3 (Secs. 3.2, 3.5) and DasGupta (2008) Chapter 29, pp. 461-468. A standard reference for this theory is the "Jackknife and Bootstrap" book of Shao and Tu (1995, Springer).

Today's topics:

- What we want to extract from Bootstrapped Data
- Bootstrap Consistency Limit Theorems for $T_n = \bar{X}_n$
- background concepts on (in-probability or a.s.) convergence of random functions

Comments on `boot` package functions

Bootstrap Confidence Interval simulations can be done with functions `boot.ci`, (and maybe `abc.ci`) in `R` package `boot`

• for `boot.ci`, 1st argument is the output from bootstrapping with package function `boot`, e.g.: `boot(xvec, B, statistic)`

• `statistic` is a user-supplied function of the form e.g.

```
statistic=function(xdat, indices) mean(xdat[indices])
```

• argument `type` in `boot.ci` tells the CI's to compute, from `c("norm","basic", "stud", "perc", "bca")`: default is `"all"`

• or directly use `abc.ci` as in `abc.ci(data, stat2)`

```
stat2=function(xdat, wght=rep(1/length(xdat),
      length(xdat)) sum(xdat*wght)
```

Applicable Features of Bootstrap Distributions

Recall that based on a data sample $\underline{X}_n$ and a statistic $T_n(\underline{X}_n) = T(\mu_n)$, we draw B *iid* bootstrap samples $\underline{X}_n^{*(b)}$ and compute $T_n(\underline{X}_n^{*(b)}) = T(\mu_n^{*(b)})$ for $b = 1, \ldots, B$. We do this to learn:

- Mean, Variances: information about $E(T(\mu_n) - T(P_0))$ and Var$(T(\mu_n))$ from observed data (sampled means & variances) on $T(\mu_n^*) - T(\mu_n)$ for bias correction and variance estimation

- $F^{-1}_{T(\underline{X}_n)-T(P_0)}(p)$ from $F^{-1}_{T(\mu_n^*)-T(\mu_n)\,|\underline{X}_n}(p)$ for Confidence Intervals

- Tail Probabilities $P(T(\mu_n^*) \geq \tau)$ from $P(T(\mu_n^*) \geq \tau \,|\, \underline{X}_n)$ for Hypothesis Test Statistics

Conditional Bootstrap Distributions

Conditional probabilities for $h(\underline{X}_n^{*(b)})$ for a single b can be studied in two ways. One is as discrete random variables conditioned on $\underline{X}_n$ arising from the set of n indices $i_1, \ldots, i_n$ sampled equiprobably with-replacement from $\{1, \ldots, n\}$.

Thus if $h(\underline{X}_n^{*(b)}) \equiv h_0(X_{j_1}^{*(b)}, \ldots, X_{j_k}^{*(b)})$ for a fixed k-tuple $(j_1, \ldots, j_k)$, then

$$E(h(\underline{X}_n^{*(b)}) \,|\, \underline{X}_n) = \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n n^{-k}\, h_0(X_{i_1}, \ldots, X_{i_k})$$

$$= \int \cdots \int h_0(\underline{x})\, d\mu_n(x_1) \cdots d\mu(x_k) \qquad \text{(related to U-statistics)}$$

The second approach is via averaging over $b = 1, \ldots, B$ for large B, using the Law of Large Numbers (next slide).

Conditional Bootstrap Distributions and the Glivenko-Cantelli Law of Large Numbers

All the quantities on the previous slide relate to bootstrap conditional-distribution values (denoted $H_{Boot}(t)$)

$$F_{T_n(\mu_n^*)\,|\,\underline{X}_n}(t) \;=\; P(T_n(\mu_n^*) \le t\,|\,\underline{X}_n) \;=\; \lim_{B\to\infty} \tfrac{1}{B} \textstyle\sum_{b=1}^{B} I_{[T_n(\mu_n^{*(b)})\le t]}$$

The limit is in Probability or w.p.1, conditionally given $\underline{X}_n$, for fixed n. Restatement: w.p.1 for samples $\underline{X}_n$, with respect to P_* of bootstrap draws (indices re-sampled)

$$\lim_{B_0\to\infty} \sup_{B\ge B_0,\, t\in\mathbb{R}} \left| F_{T(\mu_n^*)\,|\,\underline{X}_n}(t) \;-\; \frac{1}{B} \sum_{b=1}^{B} I_{[T(\mu_n^{*(b)})\le t]} \right| \xrightarrow{P_*} 0$$

Use $\|\cdot\|_\infty$ norm or Kolmogorov-Smirnov metric on d.f.'s, along with another useful metric, in proving bootstrap *consistency*.

Metrics and Definition of Consistency

In nonparametric setting where we do Bootstrap, the distribution $H_n(t)$ of $T_n(\mu_n)$ is an ∞-dimensional 'parameter' of interest.

Restrict attention to real-valued T; denote by ρ a metric on the space of distribution functions. Two examples of interest are:

$K(F,G) = \|F-G\|_\infty = \sup_t |F(t)-G(t)|$ **Kolmogorov-Smirnov**

$d_2(F,G) = \inf_{(X,Y):X\sim F,\, Y\sim G} \left(E|Y-X|^2\right)^{1/2}$ **Mallows-Wasserstein**

Definition. (29.2 in DasGupta) The random function $H_{Boot}(t) = P_*(T_n(\mu_n^*) \le t)$ is a consistent estimator of $H_n(t) = P(T_n(\mu_n) \le t)$ (respectively weakly or strongly) for metric ρ if

$\rho(H_{Boot}, H_n) \to 0$ (resp., in Probability or w.p.1) as $n \to \infty$.

Theorems on Consistency of Bootstrap

Assume X_i *iid* with $E(X_i^2) < \infty$. In our main Theorem on consistency, the statistic is an average: but it is centered and scaled so that H_n is asymptotically nondegenerate. Define

$$T_n(\mu_n, P_0) = \sqrt{n} \int x \, d(\mu_n - P_0)(x) = \sqrt{n} \left(\bar{X}_n - E(X_1) \right) \quad \textit{so that}$$

$$T_n(\mu_n^*, \mu_n) = \sqrt{n} \int x \, d(\mu_n^* - \mu_n)(x) = \sqrt{n} \left(\bar{X}_n^* - \bar{X}_n \right)$$

Now $H_{Boot}(t) = P_*(T_n(\mu_n^*, \mu_n) \leq t)$, $H_n(t) = P(T_n(\mu_n, P_0) \leq t)$

and $\sup_t |H_n(t) - \Phi(t/\sigma_X)| \to 0$.

Theorem. H_{Boot} is strongly consistent for H_n in both the K and d_2 metrics, i.e., w.p.1 with respect to $\{X_i\}_{i=1}^{\infty}$,

$$K(H_{Boot}, H_n) + d_2(H_{Boot}, H_n) \to 0 \text{ as } n \to \infty$$

Extensions to Statistics beyond $\bar{X}_n$

The validity of Bootstrap consistency goes well beyond $\bar{X}_n$, holding also for:

- Smooth functions of averages (also in multivariate setting): (Wassermann, Thm. 3.19; DasGupta, Thm. 29.4)

- Sample quantiles (DasGupta Thm. 29.6)

- U-statistics (DasGupta Thm. 29.7)

- A fancier version of these results (beyond the scope of our course) is Wassermann Thm. 3.21

Next time we sketch the proof of our Theorem, then move on to rates of convergence and 'higher-order accuracy'