Stat210B: Theoretical Statistics **Lecture Date: April 6, 2006**

Bootstrap for Confidence Internvals

Lecturer: Michael I. Jordan *Scribe: Alex Simma*

1 Review and Introduction

A quick review of our setting:

- $X \sim F_\theta$, the data is generated from a distribution parametarized by θ
- An estimator $\hat{\Theta}_n$ of θ
- A performance measure $\lambda(F)$
- For the bootstrap case, the performance measure is computed on the empirical measure: $\lambda_n(\hat{F}_n)$

There are different performance metrics. Two possible candidates are:

Different possible performance metrics:

- Variance
- Bias $\lambda_n(F) = E_F(\hat{\theta}_n) - \theta$

It may be surprising that you can estimate bias from just one sample, but we can estimate $E_{\hat{F}_n}(\hat{\theta}^\star_n) - \hat{\theta}_n$ This bias estimate can be used to create a bias-adjusted estimator, but the variance may increase enough to overcome the advantage of reducing the bias.

Remark 1. Bootstrap is, in some way, the ultimate frequentist tool. The core underpinnings of frequentism deal with the behavior of an estimator on multiple samples, but in applied situations, there is only a single sample. However, using the boostrap, it is possible to replicate your sample and perform frequentist analysis on the estimator.

To prove that the bootstrap actually produces accurate estimates, Van der Vaart provides a proof for the mean and then, extends it to other estimators with the functional delta method.

2 Confidence Intervals

Let ζ_α denote the upper α quantile of the distribution of $\frac{\hat{\theta}-\theta}{\hat{\sigma}_n}$ [1] then $P(\hat{\theta}_n - \zeta_\beta \hat{\sigma}_n \leq \theta \leq \hat{\theta}_n - \zeta_\alpha \hat{\sigma}_n) \geq 1 - \beta - \alpha$ which implies $[\hat{\theta}_n - \zeta_b \hat{\sigma_n}, \hat{\theta}_n - \zeta_b \hat{\sigma}_n]$ is a confidence interval of level $1 - \beta - \alpha$

[1] This looks like the t-statistic, but may not have the t distribution. We'd like to get its quantiles.

One way to estimate a confidence interval is through the use of a pivot. A pivot is when $\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$ doesn't depend on the true value of θ, so it's possible to compute the distribution of $\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$

However, what if no pivot exists?

- We can use weak convergence of $\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$ to (say) $\mathcal{N}(0,1)$ to conclude that $\hat{\theta}_n - \theta \xrightarrow{d} \mathcal{N}(0, \hat{\sigma}_n^2)$
- Can use the boostrap. There are 3 variants.
  - Percentile t-method: Bootstrap $\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n}$ by replacing it with $\frac{\hat{\theta}^\star_n - \hat{\theta}_n}{\hat{\sigma}^\star_n}$, which can be Monte-Carlo sampled to get bootstrap quantiles($\hat{\zeta}_\alpha$). The bootstrap quantiles are used in place of the true quantiles(ζ_α)
  - Percentile method: bootstrap $\hat{\theta}_n - \theta$ without normalizing by $\hat{\sigma_n}^2$by computing $\hat{\theta}^\star_n - \hat{\theta}_n$. The division by $\hat{\sigma}^2$ is required to create a pivot, but since the bootstrap doesn't require a pivot, we can avoid it.
  - Efron's method: Directly bootstrap $\hat{\theta}_n$ by computing $\hat{\theta}^\star_n$ This method is transformation-invariant. However, the estimated quantity doesn't include θ, so it seems odd to talk about it.

3 Asymptotic consistency

A confidence interval $[\hat{\theta}_{n,1}, \hat{\theta}_{n,2}]$ is asymptotically consistent at level $1 - \alpha - \beta$ if $\forall P$, $\liminf_{n \to \infty} P[(\hat{\theta}_{n,1} \leq \theta \leq \hat{\theta}_{n,2} | P) \geq 1 - \alpha - \beta$

Lemma 2. *23.3*

Assume $(\hat{\theta}_n - \theta)/\hat{\sigma}_n \xrightarrow{\mathcal{L}} T$ *(some* T *with a continuous cdf) and* $(\hat{\theta}^\star_n - \hat{\theta}_n)/\hat{\sigma}^\star_n \xrightarrow{\mathcal{L}} T$ *(conditionally, in probability) then we have asymptotic consistency.*

Proof. In Van der Vaart □

First, we apply the bootstrap to simple means. We ignore $\hat{\sigma}_n, \hat{\sigma}^\star_n$ because it is easy to get consistent estimators for variances, and by Slutsky's theorem, they go away.

Theorem 3. *23.4*
If the data $X_i \overset{iid}{\sim}$ *dist with* $EX_i = \mu$, $E[(X_i - \mu)(X_i - \mu)] = \Sigma$ *then*

$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ *(by CLT)*

Conditionally on $X_1, X_2, \ldots$ $\sqrt{n}(\bar{X}^\star_n - \bar{X}_n) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ *as.*

Proof.

$$
\begin{aligned}
E(X^\star_1 | \mathbb{P}_n) &= \sum_{i=1}^n \frac{1}{n} X_i = \bar{X}_n \\
E((X^\star_1 - \bar{X}_n)(X^\star_1 - \bar{X}_n) | \mathbb{P}_n) &= \frac{1}{n} \sum_i (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \to \Sigma \, \text{for almost all } X_1, X_2, \ldots
\end{aligned}
$$

At this point, we need to verify the Lindberg condition($E[||X_i^\star||^2 1\{||X_i^\star|| > \epsilon\sqrt{n}\}|\mathbb{P}_n] \to 0\ \forall\epsilon$) to apply the Lindberg CLT. This can be done by replacing the expectation with a sum over $X_1, \ldots, X_n$, selecting a n large enough such that $\epsilon\sqrt{n} > M$ for a M large enough such that the expectation is arbitrarily small. Such an M is guaranteed to exist as a result of tightness. □

Theorem 4. *23.5 Delta method for bootstrap*

Assume ϕ is continuously differentiable in a neighbourhood of θ Let $\hat{\theta}_n \stackrel{a.s}{\to} \theta$. Assume $(\hat{\theta}_n - \theta)/\hat{\sigma}_n \stackrel{\mathcal{L}}{\to} T$ and $(\hat{\theta}_n^\star - \hat{\theta}_n)/\hat{\sigma}_n^\star \stackrel{\mathcal{L}}{\to} T$ (conditionally, a.s) then

$$\begin{aligned}\sqrt{n}(\phi(\hat{\theta}_n) - \phi(\theta)) &\to \phi_\theta'(T) \\ \sqrt{n}(\phi(\hat{\theta}_n^\star) - \phi(\hat{\theta}_n)) &\to \phi_\theta'(T)\end{aligned}$$

Proof. By the Mean Value Theorem, $\phi(\hat{\theta}_n^\star) - \phi(\hat{\theta}_n) = \phi_{\tilde{\theta}_n}^{'}(\hat{\theta}_n^\star - \hat{\theta}_n)$ for an internal $\tilde{\theta}_n$

Continuity of the derivative implies that $\forall\eta > 0$, $\exists\delta > 0$ such that $||\phi_\theta^{'}, h - \phi_\theta^{'} h|| \leq n||h|| \qquad ||\theta' - \theta|| \leq \delta$. Taking a n sufficiently large and δ sufficiently small such that $\sqrt{n}||\hat{\theta}_n^\star - \hat{\theta}_n|| \leq M$ and $||\hat{\theta_n} - \theta|| \leq \delta$ implies that $R_n \triangleq ||\sqrt{n}(\phi(\hat{\theta_n^\star}) - \phi(\hat{\theta}_n)) - \phi_\theta'\sqrt{n}(\phi(\hat{\theta_n^\star}) - \phi(\hat{\theta}_n))|| = ||(\phi_{\hat{\theta} z_n}' - \phi_\theta^{'})\sqrt{n}(\phi(\hat{\theta_n^\star}) - \phi(\hat{\theta}_n))||$

Fix $\epsilon > 0$, fix M large, take η such taht $\eta M < \epsilon$

$$P(R_n > \epsilon|\hat{P}_n) \quad \leq \quad P(\sqrt{n}||\hat{\theta}_n^\star - \hat{\theta}_n|| > M \text{ or } ||\theta_n - \hat{\theta} > \delta)$$

$\hat{\theta}_n \stackrel{a.s}{\to} \theta \implies RHS \stackrel{a.s}{\to} P(||T|| > M)$ which can be made small by taking M large. □

Example:

$S_n^2 = \frac{1}{n}\sum(X_i - \bar{X}_n)^2$ is $\phi(\bar{X}_n, \bar{X_n^2})$ for $\phi(x, y) = y - x^2$ Then thm 23.4 implies consistency for $\bar{X}_n$ and $\bar{X_n^2}$, theorem 23.5 implies consistency of confidence intervals for S_n^2

4 Empirical Bootstrap

We replace the sample mean (in what we've been doing so far) with the empirical distribution. Our setup is:

- $\mathcal{F}$: Donsker class
- $\mathbb{P}_n$ empirical measure
- $X_1^\star, X_2^\star, \ldots, X_n^\star$ sample from $\mathbb{P}_n$
- Bootstrap empirical distribution: $\mathbb{P}_n^\star = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^\star}$
- Bootstrap empirical process:

$$\begin{aligned}\mathbb{G}_n^\star &= \sqrt{n}(\mathbb{P}_n^\star - \mathbb{P}_n) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (M_{n,i} - 1)\delta_{X_i}\end{aligned}$$

where $M_{n,i}$ is the number of times that X_i is redrawn. $M_n \sim Mult(n; \frac{1}{n}, \ldots, \frac{1}{n})$

5 Coming up next

Next lecture will discuss the application of functional delta method to this method.

The main person who's worked on this: Evariste Gin