STAT576
Empirical Process Theory

Pingbang Hu

August 21, 2023

Abstract

This is a graduate-level theoretical statistics course taught by Sabyasachi Chatterjee at University of Illinois Urbana-Champaign, aiming to provide an introduction to empirical process theory with applications to statistical M-estimation, non-parametric regression, classification and high dimensional statistics.

While there are no required textbooks, some books do cover (almost all) part of the material in the class, e.g., Van Der Vaart and Wellner's *Weak Convergence and Empirical Processes* [VW96].

This course is taken in Fall 2023, and the date on the covering page is the last updated time.

Contents

Chapter 1

Introduction

Lecture 1: Introduction to Mathematical Statistics

1.1 Overview of Empirical Process Theory

21 Aug. 9:00

Given inputs i.i.d. data points $X_1, \ldots, X_n$, the empirical CDF is the function

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}.$$

The classical result is that, fixing t, $F_n(t) \to F(t)$ almost surely, while $\sqrt{n}(F_n(t) - F(t)) \to \mathcal{N}(0, F(t)(1 - F(t)))$ in distribution. On the other hand, by the Glivenko-Cantelli theorem, $\sup_{t \leq \mathbb{R}} |F_n(t) - F(t)| \overset{n\to\infty}{\to} 0$ almost surely.

In this example, the empirical process is $\{F_n(t)\}_{t\in\mathbb{R}}$. More generally, let χ be the domain, $\mathbb{P}$ be a distribution on χ, and $\mathcal{F}$ be the class of function from $\chi \to \mathbb{R}$. And we're interested in

$$G_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}\left[f(X)\right]$$

where $X \sim X_1, \ldots, X_n$. Then, $\{G_n(f) \colon f \in \mathcal{F}\}$ is called the empirical process.

Now, two questions arises:

1. Uniform Law of Large Number: As $n \to 0$, whether

$$\sup_{f\in\mathcal{F}} |G_n(f)| \to 0,$$

   and if, at what rate?

2. Uniform Central Limit Theorem: In general, $\sqrt{n}G_n(f) \to \mathcal{N}(0, \mathrm{Var}\left[f(X)\right])$ in distribution. Consider

   - $X_1, \ldots, X_n$ i.i.d. from $\mathcal{U}(0, 1)$.
   - $\mathcal{F} = \left\{\mathbb{1}_{[-\infty,t]} :: t \in \mathbb{R}\right\}$
   - $U_n(t) = \sqrt{n}(F_n(t) - t)$.

   Then $U_n(t) \to \mathcal{N}(0, t - t^2)$, $(U_n(t_1), \ldots, U_n(t_k)) \to \mathrm{MVN}(0, \Sigma)$ where $\Sigma_{ij} = \min(t_i, t_j)$. Check!

1.1.1 M-Estimation

We're going to focus on the class of estimators called "M-estimator", which is

$$\hat{\theta} = \operatorname*{arg\,max}_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} M_\theta(X_i),$$

where Θ is the parameter space, and $M_\theta \colon \chi \to \mathbb{R}$. For example,

- $M_\theta(X) = -\log P_\theta(X)$, then $\hat{\theta}$ is the MLE;
- $M_\theta(X) = (X-\theta)^2$, $|X-\theta|$, or $-\mathbb{1}_{|X-\theta|\le 1}$, which is the location estimator

Now, consider $\theta_0 = \arg\max_{\theta\in\Theta} \mathbb{E}\left[M_\theta(X_1)\right]$ where $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathbb{P}$. Two questions are:

- does $\hat{\theta}$ converge to θ_0, i.e., $d(\hat{\theta}, \theta_0) \to 0$ for some metric d?
- What is the rate?

Consider $M_n(\theta) = \frac{1}{n}\sum_{i=1}^n M_\theta(X_i)$, then

$$\begin{aligned}
\mathbb{P}(d(\hat{\theta}, \theta_0) > \epsilon) &\le \mathbb{P}\left(\sup_{\theta:\, d(\theta,\theta_0)>\epsilon} M_n(\theta_0) - M_n(\theta) \ge 0\right) \\
&= \mathbb{P}\left(\sup_{\theta:\, d(\theta,\theta_0)>\epsilon} (M_n(\theta_0) - M(\theta_0) - [M_n(\theta) - M(\theta)]) \ge \inf_{\theta:\, d(\theta,\theta_0)>\epsilon} (M(\theta) - M(\theta_0))\right) \\
&\le \mathbb{P}\left(2\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \ge \inf_{\theta:\, d(\theta,\theta_0)>\epsilon} (M(\theta) - M(\theta_0))\right).
\end{aligned}$$

Appendix

Bibliography

[VW96] Aad W. Van Der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York, NY: Springer, 1996. ISBN: 978-1-4757-2547-6 978-1-4757-2545-2. DOI: 10.1007/978-1-4757-2545-2. (Visited on 08/21/2023).