Yeheng GE

My notes series

Paper Reading

Do not be distracted

Contents

Chapter 1

Cube Root Asymptotics

Overview

This paper gives a functional central limit theory for empirical process.

- many convergence rate $n^{-1/3}$
- Key point is: continuous mapping theorem for the location of maximum point.
-

The paper is a very nice material for understanding the core idea and techniques of empirical process.
Main reference of this paper is in the lecture notes of Pollard "Empirical process: Theory and applications" 1990 version.

0.1 The Mode Estimation Problem

The paper first intuitively gives an example of "mode estimation" and gives its convergence rate $n^{-1/3}$
Suppose $\hat{\theta}_n$ is chosen to maximize

$$\Gamma_n(\theta) = P_n[\theta - 1, \theta + 1] \tag{1.1}$$

is the proportion of observations in an interval of length 2.
If P has a smooth density $p(\dot{})$, the function Γ is approximately parabolic of its optimal value θ_0, which means that

$$\Gamma(\theta) - \Gamma(\theta_0) = \int_{1+\theta_0}^{1+\theta} p(x)dx - \int_{-1+\theta_0}^{-1+\theta} p(x)dx \approx -C(\theta - \theta_0)^2$$

> Take care that $\Gamma(\theta)$ is the expectation of $\Gamma_n(\theta)$. The term above is the "bias" caused by the departure of θ from θ_0.

Then we consider the stochastic term,

$$D_n(\theta) = [\Gamma_n(\theta) - \Gamma_n(\theta_0)] - [\Gamma(\theta) - \Gamma(\theta_0)].$$

For fixed θ, the $D_n(\theta)$ is approximately $N(0, \sigma_\theta^2/n)$ where

$$\sigma_\theta^2 \approx \int_{1+\theta_0}^{1+\theta} p(x)dx + \int_{-1+\theta_0}^{-1+\theta} p(x)dx \approx C\,|\theta - \theta_0|$$

The "add" and "minus" produces different order here. The order produced by "add" is due to the finite density of $p()$. Thus the density integral is proportional to the length of θ

Intuitively, when the bias term $C|\theta - \theta_0|^2$ is large comparing with the stochastic term $C|\theta - \theta_0|$, the θ is far away from the true value θ_0. Thus not maximize the $\Gamma_n(\theta)$.
So the θ could be the solution of $\Gamma_n(\theta)$ if the bias term is the same order or smaller than the stochastic term. It means

$$\begin{aligned} C|\theta - \theta_0|^2 &< Cn^{-1/2}|\theta - \theta_0|^{1/2} \\ C|\theta - \theta_0|^{3/2} &< Cn^{-1/2} \\ C|\theta - \theta_0| &< Cn^{-1/3} \end{aligned}$$

> However, it is just an intuitive explaination. theoretically, we need build error bound uniformly in θ and the normal approximation must hold uniformly over θ.

Note that the variance term σ_θ decreases with $|\theta - \theta_0|$.
If the loss function $g(\theta, \dot{)}$ is differentiable, σ_θ decreases with $|\theta - \theta_0|^2$.
Thus

$$\begin{aligned} C|\theta - \theta_0|^2 &< Cn^{-1/2}|\theta - \theta_0| \\ C|\theta - \theta_0| &< Cn^{-1/2} \end{aligned}$$

It produces the common $n^{-1/2}$ rate.
Thus the variance term σ_θ decreases with $|\theta - \theta_0|$, the non-standard case is a consequence of "shape-edge effect"

0.2 Convergence in distribution and the argmax functional

Chapter 2

Distribution-Invariant Differential Privacy

This paper has very wierd organization.

0.1 The Mode Estimation Problem