

This is a review submitted to Mathematical Reviews/MathSciNet.

Reviewer Name: Peng, Cheng

Mathematical Reviews/MathSciNet Reviewer Number: 82361

Address:

Department of Mathematics
West Chester University
West Chester, PA 19383
cpeng@wcupa.edu

Author: Bilodeau, Blair; Foster, Dylan J.; Roy, Daniel M.

Title: Minimax rates for conditional density estimation via empirical entropy.

MR Number: MR4601001

Primary classification:

Secondary classification(s):

Review text:

sectsty graphicx amsmath,amssymb

In classical regression and classification, the primary interest is to estimate the means of response Y in sub-populations defined by the covariate space \mathcal{X} , i.e., $\mu_{\mathcal{X}} = E[Y|\mathbf{X}]$. The general regression problem is defined as $\mu_{\mathcal{X}} = g(\mathbf{X} : \theta)$, θ is the vector of model parameters. Both regression and classification problems have been extensively studied. Estimating the density function of Y in sub-covariate spaces is more general and useful since the density function over covariate spaces contains the information of the conditional means. However, *conditional density estimation* is technically more challenging. The traditional uniform metric entropy that is used to characterize the joint density function will not work for conditional density due to the curse of (high) dimension.

Authors in his article proposed the minimax rates for characterizing conditional density estimation using the *empirical Hellinger entropy*. Let $\mu_{\mathcal{X}}$ be the unknown marginal distribution over covariates, f^* be the true conditional density function, and \hat{f} be the estimator. The estimation of conditional density under *logarithmic loss* is to minimize the following *Kullback-Leibler* (KL) divergence

$$\mathbb{E}_{\mathcal{X} \sim \mu_{\mathcal{X}}} KL(f * (x) || \hat{f}(x))$$

The minimax rates of convergence for these tasks are determined (up to logarithmic factors) by a critical radius ϵ_n satisfying the relationship.

$$\mathcal{H}(\mathcal{F}, \epsilon_n) \asymp n\epsilon_n^2 \tag{1}$$

where \mathcal{H} denotes the entropy for the model class \mathcal{F} under a suitable problem-

specific notion of distance. Instead of using existing entropy such as uniform metric entropy, authors proposed *empirical Hellinger entropy* for the class of conditional density families.

$$\mathcal{H}_{H,2}(\mathcal{F}, \epsilon_n, n) \asymp n\epsilon_n^2$$

Let $\mathcal{R}_n(\mathcal{F})$ be the minimax rate of KL under *empirical Hellinger entropy*. Authors established the upper and lower bounds of $\mathcal{R}_n(\mathcal{F})$ under certain conditions.

They showed that the classical fixed-point relationship in equation (1) governs the minimax rates for conditional density estimation, provided that one adopts empirical Hellinger entropy as the notion of complexity. Furthermore, the empirical Hellinger entropy captures the correct dependence on the complexity of the covariate space and leads to optimal rates, even for high-dimensional, potentially unbounded, covariate spaces.

Authors also related their new development in density estimation method to some existing work such as non-parametric density estimation, non-parametric regression, and sequential prediction, etc.