BEST: International Journal of Humanities, Arts,
Medicine and Sciences (BEST: IJHAMS)
ISSN (P): 2348–0521, ISSN (E): 2454–4728
Vol. 11, Issue 2, Dec 2023, 19–24

A GENERAL MINIMAX LOWER BOUND FOR ESTIMATING AN ARBITRARYNON-SMOOTH FUNCTIONAL

MUKHWANA KOLOLI K.

Department of Mathematics, Kibabii University

ABSTRACT

In this study, an overarching MiniMax lower bound (MLB) was formulated to create the MiniMax Risk. This MiniMax Risk serves as a tool for assessing any non-smooth functional estimations. Both the Minimax lower and upper bounds play a crucial role in determining the fundamental limits and setting benchmarks for evaluating the performance of any statistical estimator. When dealing with nonparametric estimation of statistical functionals, it's essential to establish both lower and upper bounds. Particularly within the realm of MiniMax estimation, the lower bounds hold significant importance. Estimating non-smooth functionals poses distinct properties compared to the estimation of typical smooth functionals. Consequently, standard methods often fall short in providing accurate results for estimating non-smooth functionals. To address these challenges, a set of priors with substantial differences in the functional expected values was created, while minimizing the Chi-square distance between two normal mixtures.

Keywords: Minimax, Lower Bounds, Estimator

INTRODUCTION

The MiniMax framework was used to establish an overarching lower bound (MLB) for developing the MiniMax Risk, particularly applied in estimating non-smooth functional. These bounds significantly contribute to various fields such as science, engineering, and geosciences [1].They are applied in quantifying both lower and upper bounds in estimation, testing, data compression, and L1 distance. They also serve as benchmarks for algorithmic development [2]. Understanding their functional form enables the calculation of statistical quantities and the construction of stochastic models for diverse applications. However, despite substantial research efforts in deriving these bounds and optimal convergence rates for statistical inefficiencies, while smooth functionals have been extensively studied, the estimation of non-smooth functional remains challenging [1],[2].

Recent focus on non-smooth functional estimation in nonparametric settings has gained attention [3],[4]; especially in applications such as clustering, irregularity detection, environmental pollution mapping, gene microarray analysis, and more. These non-smooth functionals pose different convergence rates, requiring distinct estimation techniques compared to standard smooth functionals. Several studies have delved into the challenges of statistical inference when dealing with non-differentiable functionals in models, moment inequalities, and optimal dynamic treatment regimes estimation [2],[3].

Some research related to this thesis focused on the nonparametric estimation of signal norms observed in Gaussian white noise, obtaining both rate and sharp asymptotics for estimators in the MiniMax setup. However, they heavily relied on continuity and normality assumptions for estimator derivation. Studies have shown the difficulties in statistical inference when dealing with nondifferentiable functionals, with implications for unbiased and regular estimators, emphasizing that non-differentiable objects hinder the existence of locally asymptotically unbiased estimator sequences. These limitations have led to stringent constraints on estimators, bias correction methods, and statistical inference procedures, encouraging the exploration of alternative criteria for evaluation.

Moreover, a MiniMax framework for adaptive data analysis was proposed, where the choice of queries in data analysis depends on previous results. The framework provided a sharp MiniMax lower bound for the squared error based on the least favorable adversary and focused on the disparity between actual and estimated values, often quantified by real-valued loss functions such as quadratic error and absolute error functions.

In the realm of MiniMax theories, extracting lower bounds and optimal convergence rates has been a focus. Various techniques were discussed, including those observed in [2], [3], [4], [5],[6],[7] to determine optimal convergence rates, test hypotheses, and estimate functionals under different models and spaces. The MiniMax estimator, aiming to minimize maximum risk, has been a subject of interest, particularly in testing simple null hypotheses against alternatives.

A sharp MiniMax lower bound ws observed on an error that is squaredin the order of $O(\frac{\sqrt{KQ^2}}{n}$ where k denoted the number of queries interrogated, $\frac{Q^2}{n}$ being the ratio of the signal-to-noise per query. The foundation of the lower bound was established by devising a challenging adversary selecting a sequence of queries most susceptible to overfitting. In these aforementioned scenarios, a real-valued loss function precisely delineated the difference between the true value and its estimation and described by the function $L(\theta, \overrightarrow{\theta})$ that estimated the amount of derivation by the prediction derivatives from the actual values. Two loss functions commonly used are quadratic error loss, $L(\theta,-\hat{\theta}) = (\theta - \hat{\theta})^2$ and absolute error loss $L(\theta,-\hat{\theta}) = \left|\theta - \hat{\theta}\right|$. The first has been associated with outliers while the second notbeing differentiable at*a = 0.* The loss function has been preferred ion statistical research because it’s differentiable in the algorithms for optimization.

MiniMax Lower Bound Techniques

The development of MiniMax theories in nonparametric functional estimation heavily relies on the efforts of statistical researchers to derive the MiniMax lower bounds and achieve an optimal convergence rate. Within the literature of statistical inference, various techniques for lower bounds have been explored. For instance, in [4], and [5]; derived the optimal convergence rate by testing a simple null hypothesis against a straightforward alternative. A study [8] estimated quadratic functionals by testing a simple null hypothesis against a composite hypothesis, determining the optimal lower bounds across a wide parameter space. Under the white noise model [7].

According to [9], [10];(1) is the estimator that is used to minimize the maximum risk (MiniMax estimator) that is given as

Sup and is expressed as : $$\sup_{\theta} R(\theta, \hat{\theta}) = \inf_{\hat{\theta}} \sup_{\theta} R(\theta, \hat{\theta}) \quad (1)$$

Where the infimum is over all estimators ˆθ . RHS of (1) is the MiniMax Risk

$$R \equiv R(\theta) = \inf_{\hat{\theta}} \sup_{\theta} R(\theta, \hat{\theta}), \quad (2)$$

Deriving an estimator that reduces the maximum risk is a challenging endeavor, especially considering its dependence on an undisclosed distribution. To address this, they introduced the concept of the MiniMax rate-optimal and an asymptotically MiniMax estimator as a solution for the MiniMax estimator.

However ,the MiniMax rate-optimal estimator with maximum risk (equal to MiniMaxrisk) can be used to provide the MiniMax estimator as :

$$\sup_{\theta\in\Theta} R(\theta,\ddot{\theta}) \asymp \inf_{\hat{\theta}} \sup_{\theta\in\Theta} R(\theta,\ddot{\theta}), \quad n\to\infty$$

where $(.) \asymp (..)$ means that both $(.)/(..)$ and $(..)/(.)$ are both bounded as $n\to\infty$. When an asymptotically MiniMax estimator is used, the MiniMax estimator is

$$\sup_{\theta\in\Theta} R(\theta,\hat{\theta}) \sim \inf_{\hat{\theta}} \sup_{\theta\in\Theta} R(\theta,\hat{\theta}), \quad n\to\infty$$

where $(.) \sim (..)$ means $(.)/(..) \to 1$.

The MiniMax rate-optimal estimator defined in equation (1) shares a strong connection with the best (MiniMax) polynomial approximation issue, which is essentially a convex optimization problem. This link between the two holds significant importance as it underlines the challenge in solving both the MiniMax risk and the convex problem:

$$\sup_{f\in\mathcal{P}} \mathbb{E}_f[F(f)-\hat{F}]^2 \qquad (3)$$

This expression can be converted into a different MiniMax and convex problem that is more efficiently solvable, aiming to minimize the maximum deviation of the polynomial from a given function. An estimator was created based on approximation theory and the application of Hermite polynomials [2]. The investigation focused on scenarios where the means were constrained by a specified value; the resulting estimator was demonstrated to be asymptotically sharp in the MiniMax sense. The derivation of the lower bound involved assessing the difference between the expected functionals under each of the priors µ0 and µ1, along with the functional's variance under the prior µ0. The chi-square distance between two marginal distributions of the observations was also utilized to establish the boundary. They employed a technique to transform the challenging MiniMax and convex problem into a more tractable one, using the MiniMax polynomial to smooth the risk at the origin and derive an optimal estimator.

Nevertheless, the presence of the polynomial factor in estimating the density might lead to challenges, particularly in the tail regions, where the density could be negative and therefore incapable of integration to yield 1. To address this, a study implemented "sufficiently regular" functions within a computational system, utilizing polynomial approximations [3]. It was observed that a polynomial that best approximates a function that has coefficients that are not exactly representable with a finite number of bits [2],[5].The corresponding equation is provided as :

$$\begin{aligned} f_{0,W}(y) &\geq \frac{1}{\sqrt{2\pi}} e^{(-\int \frac{(y-t)^2}{2} w_0(dt))} \\ &= \phi(y) e^{(-\frac{1}{2}W^2 \int t^2 z_0(dt))} \\ &\leq \phi(y) e^{(-\frac{1}{2}W^2)} \end{aligned} \qquad (4)$$

However, implemented polynomial estimations featured coefficients expressed with a finite, often limited number of bits. Consequently, they explored polynomial estimations with a maximum of mi fractional bits in the coefficient of degree i. This strategy facilitated obtaining the optimal polynomial estimation within this constraint. A key issue encountered in polynomial interpolation with polynomial points was oscillation at the interval edges [2].

Optimal polynomial approximations were employed for interpolation, demonstrating zero error between the function and the interpolating polynomial at the given points. Nevertheless, there existed more substantial error between these points. For the function |t| on the interval [−1, 1], its truncated Fourier series provided an estimate based on unbiased approximations of each term [3]. The selection of N was based on a combined consideration of the approximation error from equation (2) and the "stochastic error," which involved estimating the smooth functional equation (2) via noisy observations. It was acknowledged that the Fourier series-based estimator exhibited higher accuracy compared to the polynomial-based estimator.

The method of polynomial approximation was utilized to approximate the function f(t) = |t| with the best polynomial approximation P∗(x) of a continuous function f(t) possessing a minimum of (2k + 2) alternating points. In establishing the MiniMax lower bounds, the collection of these points was instrumental in forming the least favorable priors [6]. These priors were constructed using the Hann-Banach theorem and the Riesz representation theorem..

Methodology

This section delved into the methodologies employed to establish an overall Minimum Bounding Level (MLB) for the estimation of an arbitrary non-smooth functional by testing a set of composite hypotheses. The MLB for estimating the non-smooth functional T(λ) was developed based on a broader MLB formulation. The assessment of the obtained estimator's performance was conducted through the MiniMax Risk. Addressing the challenging MiniMax problem involved transforming it into a solvable MiniMax polynomial, which smoothed the risk at the origin using the finest polynomial approximation. Hermite polynomials were leveraged to create unbiased estimators for each expansion term, employing saddle-point approximation techniques to refine the process.

The overarching MiniMax lower bound was established by partitioning the parameter space Λ into two distinct subsets, Λ0 and Λ1, where H0 : λ ∈ Λ0 represents the null hypothesis and H1 : λ ∈ Λ1 signifies the alternative hypothesis. Priors were constructed as a pair, bounding the chi-square distance between the two. These priors were formulated to exhibit a significant contrast in the expected values of the functional T(λ) and a minimal difference when computing the Chi-square distance between the two mixture models. Mathematical concepts explored in this context encompass Nonparametric Estimation of Density Functions, Polynomial approximation, Hermite polynomials, and the Hilbert space.

Results and Discussion

This section outlines the outcomes of our research. It covers the results concerning the derived general MiniMax lower bound and the specific MiniMax lower bound for estimating the non-smooth functional. A fundamental element in constructing the MiniMax lower bound is the general MiniMax lower bound for estimating an arbitrary functional T(λ) = 1n Σi=1 |λi|.

Consider the estimator of T(λ) based on X as Tˆ(X) = Tˆ, where X represents a random sample following the probability distribution Pλ, with λ∈Λ. The bias of Tˆ(X) is represented as Eλ[Tˆ(X)] − T(X), while the prior distributions associated with Λ0 and Λ1 are respectively denoted as w0 and w1.

Since $k! > (\frac{k}{e})^k$, then

$$\leq \left(1 + e^{\frac{3}{2}W^2}\left(\frac{eW^2}{k_n}\right)^{k_n}\right)^n - 1 \tag{5}$$

Let k_n be the smallest positive integer that satisfies the condition $k_n \geq \frac{logn}{\log\log n} + \frac{\log n}{(\log\log n)^{3/2}}$ then $S_n \to 0$. Let $z_0 \leq \frac{W}{\sqrt{n}}$ and by equation (1) we obtain the MiniMax risk for estimating $T(\lambda) = \frac{1}{n}\sum_{i=1}^{n}|\lambda_i|$ over $\Lambda_n(W) = \{\lambda_i \in \mathbb{R}n : |\lambda_i| \leq W\}$ and $W > 0$ bounded from below as

$$\begin{aligned}\inf_{\hat{T}} \sup_{\lambda\in\Lambda_n(W)} \mathbb{E}\left(\hat{T} - T(\lambda)\right)^2 &\geq \frac{(2W\delta_{k_n} - \frac{W}{\sqrt{n}}S_n)^2}{(S_n+2)^2} \\ &= \beta_*^2 W^2\left(\frac{\log\log n}{\log n}\right)^2(1+o(1))\end{aligned} \tag{6}$$

Where β_* is a Bernstein constant.

CONCLUSIONS

Statistical researchers have dedicated substantial efforts towards establishing MiniMax theories in nonparametric function estimation, involving the derivation of MiniMax lower bounds, upper bounds, and the optimal convergence rate. Within the realm of MiniMax estimation, the lower bounds stand out as particularly significant. In previous applications, these bounds have been presented in simpler scenarios, and the optimal convergence rates for estimating smooth functionals often adhere to parametric rates.

The groundwork for developing the MiniMax lower bound was established by deriving the general MiniMax lower bound, expressed as equation (6). This general MiniMax lower bound was obtained by partitioning the parameter space Λ into two distinct subsets, Λ0 and Λ1, representing the null hypothesis H0: λ∈Λ0 and the alternative hypothesis H1: λ∈Λ1. Two priors, ω0 and ω1, were formulated to exhibit a substantial disparity in the expected functional values, while minimizing the Chi-square distance between two normal mixtures. The MiniMax risk for the resultant estimator was provided in equation

REFERENCES

1. *Goldenshluger A, Lepski OV. Minimax estimation of norms of a probability density: II. Rate-optimal estimation procedures. Bernoulli. 2022 May;28(2):1155-78.*
2. *Cai, T. Tony, and Mark G. Low. "Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional." (2011): 1012-1041.*
3. *Lepski O, Nemirovski A, Spokoiny V. On estimation of the L r norm of a regression function. Probability theory and related fields. 1999 Feb;113:221-53.*

4. *Donoho DL, Liu RC. Geometrizing rates of convergence, III. The Annals of Statistics. 1991 Jun 1:668-701.*
5. *Fan J. On the estimation of quadratic functionals. The Annals of Statistics. 1991 Sep 1:1273-94.*
6. *Bickel PJ, Ritov Y. Estimating integrated squared density derivatives: sharp best order of convergence estimates. Sankhyā: The Indian Journal of Statistics, Series A. 1988 Oct 1:381-93.*
7. *Lepski O, Nemirovski A, Spokoiny V. On estimation of the L r norm of a regression function. Probability theory and related fields. 1999 Feb;113:221-53.*
8. *Bickel PJ, Ritov Y. Estimating integrated squared density derivatives: sharp best order of convergence estimates. Sankhyā: The Indian Journal of Statistics, Series A. 1988 Oct 1:381-93.*
9. *Comminges, Laëtitia, and Arnak S. Dalalyan. "Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression." (2013): 146-190.*
10. *Hirano K, Porter JR. Impossibility results for nondifferentiable functionals. Econometrica. 2012 Jul;80(4):1769-90.*
11. *Brisebarre N, Muller JM. Finding the" truncated" polynomial that is closest to a function. arXiv preprint cs/0307009. 2003 Jul 4.*