

Binary Regression for BV Class Conditional Probability

Rajesh Dachiraju
Hyderabad, India
rajesh.dachiraju@gmail.com

hexagon:2610.00160v2

ABSTRACT. In this article we solve the problem of binary regression when the class conditional probability is a multivariate function of bounded variation.

Contents

Chapter 1. Introduction	1
Chapter 2. Convex functional	3
Chapter 3. Convex Random functional	5
Chapter 4. Binary Regression	7
1. The problem of binary regression	7
2. Computational Fill Distance	15
3. Conclusion	15
4. Data Availability Statement	15
5. Compliance with Ethical Standards	15
Bibliography	17

Introduction

Binary Regression is a statistical method that estimates a relationship between a set of independent random variables and a dependent binary random variable. The dependency is defined as class conditional probability function. The k-nearest neighbour method solves the problem when the class conditional probability function belongs to the class of Lipschitz functions[2] and kernel based methods solve the problem when the class conditional probability belongs to the kernel space[7]. In this article we give a solution when the class conditional probability belongs to a wider class of BV functions(in the Vitali sense). In chapter 2 we define a convex functional, in chapter 3 we define a convex random functional and in chapter 3 we give solution for binary regression problem.

Convex functional

0.1. Definitions. Let $H^k(\Omega)$ denote the Sobolev Hilbert space of the functions defined on the set Ω , $\mathbb{T}^m$ denote the m - dimensional Torus. Defining a function on a torus $\mathbb{T}^m$ means the function is defined over $(0, 1)^m$ and it is periodic with a period $T = (1, 1 \dots 1) \in \{1\}^m$. Let $\mathbb{Z}^m$ denote the set containing all the m -tuples of integers.

Definition 2.1. We define the k -gradient as

$$(2.1) \quad \nabla^k f = \left(\frac{\partial^k f}{\partial x_1^k}, \frac{\partial^k f}{\partial x_2^k}, \dots, \frac{\partial^k f}{\partial x_m^k} \right).$$

We define the functional C_λ as

$$(2.2) \quad C_\lambda(f) = \frac{\lambda^2}{n} \sum_{i=1}^n (f(\mathbf{p}_i) - q_i)^2 + \lambda \|\nabla^k f\|_{L^2(\mathbb{T}^m)}^2 + \|f\|_{L^2(\mathbb{T}^m)}^2,$$

where $k, m \in \mathbb{N}, k > \frac{m}{2}, \lambda \in \mathbb{R}^+$ and $f \in TP_\omega$.

The minimization problem is minimizing the functional defined in Equation 2.2 over the space TP_ω .

It can be shown that the functional in Equation 2.2 is convex and has a unique minimizer in the space $C^0(\mathbb{T}^m) \cap H^k(\mathbb{T}^m)$, the proof of which is given in Appendix A of [1]. The space TP_ω is a linear subspace of $C^0(\mathbb{T}^m) \cap H^k(\mathbb{T}^m)$ and hence it follows that the functional has a unique minimizer in the space TP_ω , the space of trigonometric polynomials of degree less than or equal to ω .

Convex Random functional

A random functional could be

- an ensemble of maps acting from a function space to reals.
- a map acting from a space of random processes to reals.
- an ensemble of maps acting from a space of random processes to reals.

For a convex random functional, over a given space of functions, the convexity is defined on its expected value. The expected value of the functional should satisfy the Jensen's inequality. The minimizer is unique, and the minimizer minimizes the expected value (follows from definition). Here the minimizer is a random process.

Define $\mathbf{q}_i$ to be binary random variables such that $P(\mathbf{q}_i = 1) = \eta_i$ and $P(\mathbf{q}_i = 0) = 1 - \eta_i$. Now the functional defined in Equation 2.2 is a random functional. We now prove that it is a convex random functional.

Theorem 3.1. *The random functional $C_\lambda(f)$ is convex and has a unique minimizer in the space TP_ω .*

Proof.

(3.1)

$$\begin{aligned}
 E[C_\lambda(f)] &= E\left[\frac{\lambda^2}{n} \sum_{i=1}^n (f(\mathbf{p}_i) - \mathbf{q}_i)^2\right] + \lambda E[\|\nabla^k f\|_{L^2(\mathbb{T}^m)}^2] + E[\|f\|_{L^2(\mathbb{T}^m)}^2] \\
 &= \frac{\lambda^2}{n} \sum_{i=1}^n (f(\mathbf{p}_i) - 1)^2 \eta_i + \frac{\lambda^2}{n} \sum_{i=1}^n f(\mathbf{p}_i)^2 (1 - \eta_i) + \lambda \|\nabla^k f\|_{L^2(\mathbb{T}^m)}^2 + \|f\|_{L^2(\mathbb{T}^m)}^2 \\
 &= \frac{\lambda^2}{n} \sum_{i=1}^n (f(\mathbf{p}_i) - \eta_i)^2 + \eta_i (1 - \eta_i) + \lambda \|\nabla^k f\|_{L^2(\mathbb{T}^m)}^2 + \|f\|_{L^2(\mathbb{T}^m)}^2
 \end{aligned}$$

the final expression for $E\{C_\lambda(f)\}$ in Equation 3.1 is convex and has a unique minimizer in $C^0(\mathbb{T}^m) \cap H^k(\mathbb{T}^m)$ using theorem in Appendix A of [1] and also due to $\eta_i > 0$ and $1 - \eta_i > 0$ for $i = 1, 2, \dots, n$. Hence $E\{C_\lambda(f)\}$ has unique minimizer in TP_ω as TP_ω is a linear subspace of $C^0(\mathbb{T}^m) \cap H^k(\mathbb{T}^m)$. So the random functional $C_\lambda(f)$ has a unique minimizer in the space TP_ω . $\square$

To find that we need to derive the Euler Lagrange equation. Given in section 3.1 of [1] is the derivation of E-L equation and we get the random PDE given below

$$(3.2) \quad -\frac{\lambda^2}{n} \sum_{i=1}^n (\mathbf{q}_i - u(\mathbf{p}_i)) \phi(\mathbf{p}_i) + \lambda \int_{\mathbb{T}^m} \nabla^k \phi(\mathbf{x}) \cdot \nabla^k u(\mathbf{x}) d^m \mathbf{x} + \int_{\mathbb{T}^m} \phi(\mathbf{x}) u(\mathbf{x}) d^m \mathbf{x} = 0$$

$$\forall \phi \in TP_{\omega}$$

The PDE is random as $\mathbf{q}_i$ are random variables. This PDE can be solved in a similar way as in case of deterministic case as done in Appendix B of [1] to get the following solution.

$$(3.3) \quad f_{\lambda}(\mathbf{x}) = \sum_{i=1}^n \frac{c_i}{n} w_{\lambda}(\mathbf{x} - \mathbf{p}_i),$$

where

$$(3.4) \quad w_{\lambda}(\mathbf{x}) = P_{\omega} g_{\lambda}(\mathbf{x})$$

$$(3.5) \quad g_{\lambda}(\mathbf{x}) = \sum_{\mathbf{l} \in \mathbb{Z}^m} \frac{1}{1 + \lambda \|\mathbf{l}\|_{2k}^{2k}} \cos(2\pi \mathbf{l} \cdot \mathbf{x}).$$

$\mathbf{c} = [c_1, c_2, \dots, c_n]^T$ is given as

$$(3.6) \quad \mathbf{c} = \left(\frac{1}{n} W_{\lambda} + \frac{1}{\lambda^2} I \right)^{-1} L,$$

where the matrix W_{λ} is given as

$$(3.7) \quad W_{\lambda} = [\gamma_{ij}(\lambda)]_{n \times n}, \gamma_{ij}(\lambda) = w_{\lambda}(\mathbf{p}_i - \mathbf{p}_j)$$

and

$$L = [q_1, q_2, \dots, q_n]^T$$

.

The minimizer f_{λ} is a random process as L is a binary random vector.

Binary Regression

Let X be a random variable uniformly distributed over $(0, 1)^m$. Let Y be a binary random variable such that $P(Y = 1/X = x) = \eta(x)$ and $P(Y = 0/X = x) = 1 - \eta(x)$. Let n samples of X are drawn which are denoted as $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n$. Let the random variable y_i such that the probability $P(y_i = 1) = \eta(\mathbf{p}_i)$ and $P(y_i = 0) = 1 - \eta(\mathbf{p}_i)$. $\eta : (0, 1)^m \rightarrow \mathbb{R}$, $\eta(\mathbf{x}) > 0 \forall \mathbf{x} \in (0, 1)^m$ is called the class conditional probability. In this article we consider η to be the class of BV functions in the Vitali sense.

1. The problem of binary regression

Given n randomly drawn samples $\mathbf{p}_i$ Estimate $\hat{\eta}_n$ such that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{T}^m} E\{(\hat{\eta}_n(\mathbf{x}) - \eta(\mathbf{x}))^2\} d^m \mathbf{x} = 0$$

By this estimation, as given in [2] we get closer to the Baye's classifier and hence achieve Baye's error as data grows.

Definition 4.1. Lets define the functional

$$(4.1) \quad D_\lambda^n(u) = \frac{\lambda^2}{n} \sum_{i=1}^n (u(\mathbf{p}_i) - y_i)^2 + \lambda \|\nabla^k u\|_{L^2(\mathbb{T}^m)}^2 + \|u\|_{L^2(\mathbb{T}^m)}^2,$$

where $k, m \in \mathbb{N}$, $k > \frac{m}{2}$, $\lambda \in \mathbb{R}^+$ and $u \in TP_\omega$.

As we vary the number of data points n and also ω in our analysis in this section, we therefore, the functional $D_\lambda^n(u)$ is denoted as $D_{\lambda, \omega}^n(u)$ and the minimizer u_λ is denoted as $u_{\lambda, \omega}^n$. Note that the suffix n is not a power, but only a notation that the parameter is associated with the functional defined over the set of n data points.

Define denote discrepancy measure ζ_n of the point set $E_n = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$ in the domain $(0, 1)^m$ as

$$(4.2) \quad \zeta_n = D_n^*(\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\})$$

The definition of star discrepancy D_n^* is assumed to be as defined in the book [6]. As points grow dense as $n \rightarrow \infty$, from [6], we have the asymptotic

$$\zeta_n = O\left(\frac{(\ln n)^m}{n}\right)$$

Theorem 4.2. *Considering the definitions in 4.1, If we vary the parameter λ with the number of data points n as $\lambda = \zeta_n^{-\beta}$, where $\beta > 0$ and vary $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_m)$ as $\omega_i = \kappa_i \zeta_n^{-\alpha}$, $i = 1, 2, \dots, m$, where κ_i are a positive constants, and additionally if the following conditions are assumed*

$$(4.3) \quad \begin{aligned} k &> \frac{m}{2} \\ \alpha &> 0 \\ \beta &> 0 \\ 0 < \alpha &< \frac{2}{2k-1} \\ 2 < \beta &< \frac{4k}{2k-1} \\ \frac{\alpha}{\beta} &< \frac{1}{2k-1} \end{aligned}$$

then

$$(4.4) \quad \lim_{n \rightarrow \infty} \mathbf{E}[\|u_{\lambda, \boldsymbol{\omega}}^n - \eta\|_{L^2(\mathbb{T}^m)}] = \lim_{n \rightarrow \infty} O(\zeta_n^{r_0}) = 0$$

Proof. As $u_{\lambda, \boldsymbol{\omega}}^n$ is the minimizer of the functional $D_{\lambda, \boldsymbol{\omega}}^n(f)$ in $TP_{\boldsymbol{\omega}}$, we have

$$(4.5) \quad \mathbf{E}[D_{\lambda, \boldsymbol{\omega}}^n(u_{\lambda, \boldsymbol{\omega}}^n)] \leq \mathbf{E}[D_{\lambda, \boldsymbol{\omega}}^n(P_{\boldsymbol{\omega}}\eta)]$$

Then using the expression for the functional as in Equation 4.1, we have

$$(4.6) \quad \begin{aligned} \frac{\lambda^2}{n} \mathbf{E}\left[\sum_{i=1}^n (u_{\lambda, \boldsymbol{\omega}}^n(\mathbf{p}_i) - \mathbf{y}_i)^2\right] + \lambda \mathbf{E}[\|\nabla^k u_{\lambda, \boldsymbol{\omega}}^n\|_{L^2(\mathbb{T}^m)}^2] + \mathbf{E}[\|u_{\lambda, \boldsymbol{\omega}}^n\|_{L^2(\mathbb{T}^m)}^2] \leq \\ \mathbf{E}\left[\frac{\lambda^2}{n} \sum_{i=1}^n (P_{\boldsymbol{\omega}}\eta(\mathbf{p}_i) - \mathbf{y}_i)^2\right] + \lambda \|\nabla^k P_{\boldsymbol{\omega}}\eta\|_{L^2(\mathbb{T}^m)}^2 + \|P_{\boldsymbol{\omega}}\eta\|_{L^2(\mathbb{T}^m)}^2 \end{aligned}$$

$$(4.7) \quad \begin{aligned} \mathbf{E}\left[\sum_{i=1}^n (u_{\lambda, \boldsymbol{\omega}}^n(\mathbf{p}_i) - \mathbf{y}_i)^2\right] &= \sum_{i=1}^n \mathbf{E}[(u_{\lambda, \boldsymbol{\omega}}^n(\mathbf{p}_i) - 1)^2 \eta(\mathbf{p}_i)] + \sum_{i=1}^n \mathbf{E}[u_{\lambda, \boldsymbol{\omega}}^n(\mathbf{p}_i)^2 (1 - \eta(\mathbf{p}_i))] \\ &= \sum_{i=1}^n \mathbf{E}[(u_{\lambda, \boldsymbol{\omega}}^n(\mathbf{p}_i) - \eta(\mathbf{p}_i))^2] + \eta(\mathbf{p}_i)(1 - \eta(\mathbf{p}_i)) \end{aligned}$$

$$\begin{aligned}
(4.8) \quad & \mathbf{E}\left[\sum_{i=1}^n (P_{\omega}\eta(\mathbf{p}_i) - \mathbf{y}_i)^2\right] \\
&= \sum_{i=1}^n (P_{\omega}\eta(\mathbf{p}_i) - 1)^2 \eta(\mathbf{p}_i) + \sum_{i=1}^n P_{\omega}\eta(\mathbf{p}_i)^2 (1 - \eta(\mathbf{p}_i)) \\
&= \sum_{i=1}^n P_{\omega}\eta(\mathbf{p}_i)^2 + \eta(\mathbf{p}_i)(1 - 2P_{\omega}\eta(\mathbf{p}_i))
\end{aligned}$$

As all terms in the LHS of the inequality 4.6 are positive we have

$$(4.9)$$

$$(4.10)$$

$$\lambda \mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}^2] \leq \mathbf{E}\left[\frac{\lambda^2}{n} \sum_{i=1}^n (P_{\omega}\eta(\mathbf{p}_i) - \mathbf{y}_i)^2\right] + \lambda \|\nabla^k P_{\omega}\eta\|_{L^2(\mathbb{T}^m)}^2 + \|P_{\omega}\eta\|_{L^2(\mathbb{T}^m)}^2$$

$$(4.11)$$

$$\mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}^2] \leq \frac{\lambda}{n} \sum_{i=1}^n P_{\omega}\eta(\mathbf{p}_i)^2 + \eta(\mathbf{p}_i)(1 - 2P_{\omega}\eta(\mathbf{p}_i)) + \|\nabla^k P_{\omega}\eta\|_{L^2(\mathbb{T}^m)}^2 + \frac{1}{\lambda} \|P_{\omega}\eta\|_{L^2(\mathbb{T}^m)}^2$$

Using the Koksma-Hlawka inequality [6] for the convergence of the Riemann integral, as n grows, we have the asymptotic

$$(4.12)$$

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{i=1}^n P_{\omega}\eta(\mathbf{p}_i)^2 + \eta(\mathbf{p}_i)(1 - 2P_{\omega}\eta(\mathbf{p}_i)) - \int_{\mathbb{T}^m} P_{\omega}\eta(\mathbf{x})^2 + \eta(\mathbf{x})(1 - 2P_{\omega}\eta(\mathbf{x})) d^m \mathbf{x} \right| \\
& \leq \zeta_n V_{\mathbb{T}^m} (P_{\omega}\eta^2 + \eta(1 - 2P_{\omega}\eta))
\end{aligned}$$

Which implies the following equations

$$\begin{aligned}
(4.13) \quad & \sum_{i=1}^n P_{\omega}\eta(\mathbf{p}_i)^2 + \eta(\mathbf{p}_i)(1 - 2P_{\omega}\eta(\mathbf{p}_i)) \\
& \leq \int_{\mathbb{T}^m} P_{\omega}\eta(\mathbf{x})^2 + \eta(\mathbf{x})(1 - 2P_{\omega}\eta(\mathbf{x})) d\mathbf{x} + \zeta_n V_{\mathbb{T}^m} (P_{\omega}\eta^2 + \eta(1 - 2P_{\omega}\eta)) \text{ and}
\end{aligned}$$

$$\begin{aligned}
(4.14) \quad & \sum_{i=1}^n P_{\omega}\eta(\mathbf{p}_i)^2 + \eta(\mathbf{p}_i)(1 - 2P_{\omega}\eta(\mathbf{p}_i)) \\
& \geq \int_{\mathbb{T}^m} P_{\omega}\eta(\mathbf{x})^2 + \eta(\mathbf{x})(1 - 2P_{\omega}\eta(\mathbf{x})) d\mathbf{x} - \zeta_n V_{\mathbb{T}^m} (P_{\omega}\eta^2 + \eta(1 - 2P_{\omega}\eta))
\end{aligned}$$

Again as all the terms in the Equation 4.6 are positive, we have

$$(4.15) \quad \mathbf{E}\left[\frac{1}{n} \sum_{i=1}^n (u_{\lambda, \omega}^n(\mathbf{p}_i) - \mathbf{y}_i)^2\right] \leq \mathbf{E}\left[\frac{1}{n} \sum_{i=1}^n (P_{\omega} \eta(\mathbf{p}_i) - \mathbf{y}_i)^2\right] + \frac{1}{\lambda} \|\nabla^k P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2 + \frac{1}{\lambda^2} \|P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2$$

Again using the Koksma-Hlawka inequality [6] for the convergence of the Riemann integral, as n grows, we have

$$(4.16) \quad \left| \sum_{i=1}^n \mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{p}_i) - \eta(\mathbf{p}_i))^2] + \eta(\mathbf{p}_i)(1 - \eta(\mathbf{p}_i)) - \int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2] + \eta(\mathbf{x})(1 - \eta(\mathbf{x}))) d^m \mathbf{x} \right| \leq \zeta_n V_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta))$$

which implies the following equations

$$(4.17) \quad \sum_{i=1}^n \mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{p}_i) - \eta(\mathbf{p}_i))^2] + \eta(\mathbf{p}_i)(1 - \eta(\mathbf{p}_i)) \leq \int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2] + \eta(\mathbf{x})(1 - \eta(\mathbf{x}))) d^m \mathbf{x} + \zeta_n V_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta))$$

and

$$(4.18) \quad \sum_{i=1}^n \mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{p}_i) - \eta(\mathbf{p}_i))^2] + \eta(\mathbf{p}_i)(1 - \eta(\mathbf{p}_i)) \geq \int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2] + \eta(\mathbf{x})(1 - \eta(\mathbf{x}))) d^m \mathbf{x} - \zeta_n V_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta))$$

$$(4.19)$$

Using Equations 4.14, 4.18 and 4.15,

$$(4.20) \quad \int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2] + \eta(\mathbf{x})(1 - \eta(\mathbf{x}))) d^m \mathbf{x} - \int_{\mathbb{T}^m} (P_{\omega} \eta(\mathbf{x})^2 + \eta(\mathbf{x})(1 - 2P_{\omega} \eta(\mathbf{x}))) d^m \mathbf{x} \leq \zeta_n V_{\mathbb{T}^m} (P_{\omega} \eta^2 + \eta(1 - 2P_{\omega} \eta) + \zeta_n V_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta)) + \frac{1}{\lambda} \|\nabla^k P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2 + \frac{1}{\lambda^2} \|P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2$$

Simplifying LHS we get

$$(4.21) \quad \int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2]) d^m \mathbf{x} + \int_{\mathbb{T}^m} (2\eta(\mathbf{x})P_{\omega} \eta(\mathbf{x}) - P_{\omega} \eta(\mathbf{x})^2 - \eta(\mathbf{x})^2) d^m \mathbf{x}$$

Now consider $V_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta))$ and writing the expression for the total variation as an intergal of the absolute of the distributional derivative, we have

(4.22)

$$\begin{aligned}
V_{\mathbb{T}^m}(\mathbf{E}[(u_{\lambda,\omega}^n - \eta)^2] + \eta(1 - \eta)) &= \int_{\mathbb{T}^m} \|D\{\mathbf{E}[(u_{\lambda,\omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2] + \eta(\mathbf{x})(1 - \eta(\mathbf{x}))\}\|_2 d^m \mathbf{x} \\
&\leq \int_{\mathbb{T}^m} \mathbf{E}[|u_{\lambda,\omega}^n(\mathbf{x}) - \eta(\mathbf{x})| \|(\nabla u_{\lambda,\omega}^n(\mathbf{x}) - D\eta(\mathbf{x}))\|_2] d^m \mathbf{x} \\
&\quad + \int_{\mathbb{T}^m} D\|\eta(\mathbf{x})(1 - \eta(\mathbf{x}))\|_2 d^m \mathbf{x} \\
&\leq \int_{\mathbb{T}^m} \mathbf{E}[|u_{\lambda,\omega}^n(\mathbf{x}) - \eta(\mathbf{x})|] d^m \mathbf{x} \int_{\mathbb{T}^m} \mathbf{E}[\|(\nabla u_{\lambda,\omega}^n(\mathbf{x}) - D\eta(\mathbf{x}))\|_2] d^m \mathbf{x} \\
&\quad + \int_{\mathbb{T}^m} D\{\eta(\mathbf{x})(1 - \eta(\mathbf{x}))\} d^m(\mathbf{x}) \\
&\leq \int_{\mathbb{T}^m} \mathbf{E}[|u_{\lambda,\omega}^n(\mathbf{x})| + |\eta(\mathbf{x})|] d^m \mathbf{x} \int_{\mathbb{T}^m} \mathbf{E}[\|(\nabla u_{\lambda,\omega}^n(\mathbf{x}))\| + \|D\eta(\mathbf{x})\|_2] d^m \mathbf{x} \\
&\quad + \int_{\mathbb{T}^m} D\{\eta(\mathbf{x})(1 - \eta(\mathbf{x}))\} d^m(\mathbf{x}) \\
&= (\mathbf{E}[\|u_{\lambda,\omega}^n\|_{L^1(\mathbb{T}^m)}] + \mathbf{E}[\|\eta\|_{L^1(\mathbb{T}^m)}])(\mathbf{E}[\|\nabla u_{\lambda,\omega}^n\|_{L^1(\mathbb{T}^m)}] + \mathbf{E}[\|D\eta\|_{L^1(\mathbb{T}^m)}])
\end{aligned}$$

(Note that, for the function $\psi \in BV(\mathbb{T}^m)$, $D\psi$ is the distributional/weak derivative). As $u_{\lambda,\omega}^n$ is a minimizer in TP_ω , there exists a positive constant K_4 independent of $u_{\lambda,\omega}^n$ such that

$$(4.23) \quad \mathbf{E}[\|u_{\lambda,\omega}^n\|_{L^1(\mathbb{T}^m)}] \leq K_4 \mathbf{E}[\|u_{\lambda,\omega}^n\|_{L^\infty(\mathbb{T}^m)}]$$

Similarly as $\nabla u_{\lambda,\omega}^n$ is smooth, there exists a positive constant K_5 independent of $u_{\lambda,\omega}^n$ such that

$$(4.24) \quad \begin{aligned} &\mathbf{E}[\|\nabla u_{\lambda,\omega}^n\|_{L^1(\mathbb{T}^m)}] \\ &\leq K_5 \mathbf{E}[\|\nabla u_{\lambda,\omega}^n\|_{L^2(\mathbb{T}^m)}] \end{aligned}$$

Using Morrey's inequality [3], there exists a positive constant K_6 independent of $u_{\lambda,\omega}^n$ such that

$$(4.25) \quad \mathbf{E}[\|u_{\lambda,\omega}^n\|_{L^\infty(\mathbb{T}^m)}] \leq K_6 \mathbf{E}[\|\nabla^k u_{\lambda,\omega}^n\|_{L^2(\mathbb{T}^m)}]$$

Using Friedrich's inequality [8] (a generalization of Poincaré-Wirtinger inequality [3]), there exists a positive constant K_7 independent of $u_{\lambda,\omega}^n$ such that

$$(4.26) \quad \mathbf{E}[\|\nabla u_{\lambda,\omega}^n\|_{L^2(\mathbb{T}^m)}] \leq K_7 \mathbf{E}[\|\nabla^k u_{\lambda,\omega}^n\|_{L^2(\mathbb{T}^m)}]$$

Combining Equations 4.23, 4.25 we obtain

$$(4.27) \quad \mathbf{E}[\|u_{\lambda,\omega}^n\|_{L^1(\mathbb{T}^m)}] \leq K_4 K_6 \mathbf{E}[\|\nabla^k u_{\lambda,\omega}^n\|_{L^2(\mathbb{T}^m)}]$$

and Combining Equations 4.24, 4.26 we obtain

$$(4.28) \quad \mathbf{E}[\|\nabla u_{\lambda,\omega}^n\|_{L^1(\mathbb{T}^m)}] \leq K_5 K_7 \mathbf{E}[\|\nabla^k u_{\lambda,\omega}^n\|_{L^2(\mathbb{T}^m)}]$$

Using Equations 4.22, 4.28 and 4.27 we get

$$\begin{aligned}
 (4.29) \quad & V_{\mathbb{T}^m}(\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta)) \\
 & \leq (K_4 K_6 \mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}] \\
 & \quad + \mathbf{E}[\|\eta\|_{L^1(\mathbb{T}^m)}])(K_5 K_7 \mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}] + \mathbf{E}[\|D\eta\|_{L^1(\mathbb{T}^m)}]) \\
 & \quad + \int_{\mathbb{T}^m} D\{\eta(\mathbf{x})(1 - \eta(\mathbf{x}))\} d^m(\mathbf{x}) \\
 & \leq K_8 \mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}^2] + (K_9 \|\eta\|_{L^1(\mathbb{T}^m)} + K_{10} \|\nabla \eta\|_{L^1(\mathbb{T}^m)}) \mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}] \\
 & \quad + \|\eta\|_{L^1(\mathbb{T}^m)} \|D\eta\|_{L^1(\mathbb{T}^m)} \\
 & \quad + \int_{\mathbb{T}^m} D\{\eta(\mathbf{x})(1 - \eta(\mathbf{x}))\} d^m(\mathbf{x})
 \end{aligned}$$

where $K_8 = K_4 K_5 K_6 K_7$, $K_9 = K_5 K_7$ and $K_{10} = K_4 K_5$. Substituting $\lambda = \zeta_n^{-\beta}$ in Equation 4.20 we get

$$\begin{aligned}
 (4.30) \quad & \int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2]) d^m \mathbf{x} - \\
 & \int_{\mathbb{T}^m} (2\eta(\mathbf{x}) P_{\omega} \eta(\mathbf{x}) - P_{\omega} \eta(\mathbf{x})^2 - \eta(\mathbf{x})^2) d^m \mathbf{x} \\
 & \leq \zeta_n V_{\mathbb{T}^m}(P_{\omega} \eta^2 + \eta(1 - 2P_{\omega} \eta)) + \zeta_n V_{\mathbb{T}^m}(\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta)) \\
 & \quad + \zeta_n^{\beta} \|\nabla^k P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2 + \zeta_n^{2\beta} \|P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2
 \end{aligned}$$

As ω grows, and η is a function of bounded variation, we have the asymptotic

$$(4.31) \quad \int_{\mathbb{T}^m} (2\eta(\mathbf{x}) P_{\omega} \eta(\mathbf{x}) - P_{\omega} \eta(\mathbf{x})^2 - \eta(\mathbf{x})^2) d^m \mathbf{x} = O(1/\|\omega\|_2)$$

Substituting $\|\omega\|_2 = \zeta_n^{-\alpha}$ we get

$$(4.32) \quad \int_{\mathbb{T}^m} (2\eta(\mathbf{x}) P_{\omega} \eta(\mathbf{x}) - P_{\omega} \eta(\mathbf{x})^2 - \eta(\mathbf{x})^2) d^m \mathbf{x} = O(\zeta_n^{\alpha})$$

From the definition of total variation, and Fourier analysis we have the asymptotic

$$(4.33) \quad V_{\mathbb{T}^m}(P_{\omega} \eta^2 + \eta(1 - 2P_{\omega} \eta)) = \Theta(1)$$

Using derivative as a Fourier multiplier operator and using Plancherel theorem , we have the asymptotic

$$\begin{aligned}
 (4.34) \quad & \|\nabla^k P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2 = O(\|\omega\|_2^{2k-1}) \\
 & = O(\zeta_n^{-\alpha(2k-1)}) \text{ (after substituting } \|\omega\|_2 = \zeta_n^{-\alpha} \text{)}
 \end{aligned}$$

As η is a BV function, we have the asymptotic

$$(4.35) \quad \|P_{\omega} \eta\|_{L^2(\mathbb{T}^m)}^2 = O(1)$$

Using Equations 4.10, 4.11 and later using Equations 4.31, 4.33, 4.34 and 4.35,
(4.36)

$$\begin{aligned}
\mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}^2] &\leq \lambda \int_{\mathbb{T}^m} P_{\omega} \eta(\mathbf{x})^2 + \eta(\mathbf{x})(1 - 2P_{\omega} \eta(\mathbf{x})) d^m \mathbf{x} + \lambda \zeta_n V_{\mathbb{T}^m} (P_{\omega} \eta^2 + \eta(1 - 2P_{\omega} \eta)) \\
&\quad + \|\nabla^k P_{\omega} \psi\|_{L^2(\mathbb{T}^m)}^2 + \frac{1}{\lambda} \|P_{\omega} \psi\|_{L^2(\mathbb{T}^m)}^2 \\
&= \lambda O(\zeta_n^{\alpha}) + \lambda \zeta_n + O(\zeta_n^{-\alpha(2k-1)}) + \frac{1}{\lambda} O(1) \\
&= O(\zeta_n^{\alpha-\beta}) + O(\zeta_n^{1-\beta}) + O(\zeta_n^{-\alpha(2k-1)}) + O(\zeta_n^{\beta}) \text{ after substituting } \lambda = \zeta_n^{-\beta} \\
&= O(\zeta_n^{\gamma}) \text{ where } \gamma = \min(\alpha - \beta, 1 - \beta, -\alpha(2k - 1), \beta)
\end{aligned}$$

Using Equations 4.29 and 4.36,

$$\begin{aligned}
(4.37) \quad \zeta_n V_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n - \eta)^2] + \eta(1 - \eta)) &= \zeta_n O(\mathbf{E}[\|\nabla^k u_{\lambda, \omega}^n\|_{L^2(\mathbb{T}^m)}^2]) + \zeta_n O(1) \\
&= O(\zeta_n^{\gamma_1})
\end{aligned}$$

where

$$(4.38) \quad \gamma_1 = \gamma + 1 = \min(1 + \alpha - \beta, 2 - \beta, 1 - \alpha(2k - 1), 1 + \beta, 1)$$

Equation 4.34 implies the asymptotic

$$\begin{aligned}
(4.39) \quad \zeta_n^{\beta} \|\nabla^k P_{\omega} \psi\|_{L^2(\mathbb{T}^m)}^2 &= \zeta_n^{\beta} O(\zeta_n^{-\alpha(2k-1)}) \\
&= O(\zeta_n^{\beta-\alpha(2k-1)})
\end{aligned}$$

Equation 4.35 implies the asymptotic

$$(4.40) \quad \zeta_n^{2\beta} \|P_{\omega} \psi\|_{L^2(\mathbb{T}^m)}^2 = O(\zeta_n^{2\beta})$$

Using Equations 4.30, 4.37, 4.33, 4.39 and 4.40 we get

$$\begin{aligned}
(4.41) \quad &\int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2]) d^m \mathbf{x} - \\
&\int_{\mathbb{T}^m} (2\eta(\mathbf{x})P_{\omega}\eta(\mathbf{x}) - P_{\omega}\eta(\mathbf{x})^2 - \eta(\mathbf{x})^2) d^m \mathbf{x} \\
&= O(\zeta_n) + O(\zeta_n^{\gamma_1+1}) + O(\zeta_n^{\beta-\alpha(2k-1)}) + O(\zeta_n^{2\beta}) \\
&\implies \int_{\mathbb{T}^m} (\mathbf{E}[(u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2]) d^m \mathbf{x} = O(\zeta_n^r) + O(\zeta_n^{\alpha}) \\
&= O(\zeta_n^{r_0}) \\
&\implies \mathbf{E}[\int_{\mathbb{T}^m} ((u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2) d^m \mathbf{x}] = O(\zeta_n^{r_0})
\end{aligned}$$

where

$$r_0 = \min(1, \gamma_1 + 1, \beta - \alpha(2k - 1), 2\beta, \alpha)$$

Making $r_0 > 0$ and also noting the previous assumptions, $\alpha > 0$, $\beta > 0$ and $k > \frac{m}{2}$, and after simplification we get the final conditions on α, β, k that are required for

convergence as below

$$\begin{aligned}
 & k > \frac{m}{2} \\
 & \alpha > 0 \\
 & \beta > 0 \\
 (4.42) \quad & 0 < \alpha < \frac{2}{2k-1} \\
 & 2 < \beta < \frac{4k}{2k-1} \\
 & \frac{\alpha}{\beta} < \frac{1}{2k-1}
 \end{aligned}$$

Under these conditions,

$$(4.43) \quad \lim_{n \rightarrow \infty} \mathbf{E} \left[\int_{\mathbb{T}^m} ((u_{\lambda, \omega}^n(\mathbf{x}) - \eta(\mathbf{x}))^2) d^m \mathbf{x} \right] = \lim_{n \rightarrow \infty} O(\zeta_n^{r_0}) = 0$$

□

1.1. Deterministic convergence. Let the probability density function of the scattered data be $\mu : \mathbb{T}^m \rightarrow \mathbb{R}$. Assume μ to be BV. We now prove deterministic convergence, meaning L^2 -norm converges to zero instead of its expectation.

Consider a ball of radius r around every data point and let the i^{th} ball be denoted by S_i .

We now prove that the difference between L^2 -norm and expected value of the L^2 -norm goes to zero as the number of data points go to zero. Let this difference be denoted by D_n . we can show that $D_n = \sum_{i=1}^n \{ (\int_{S_i} (u_i^n - \eta(x))(u_i^n - \eta^*(x)) d\mu - \int_{\mathbb{T}^m} (u^n - \eta)^2 d\mu \}$. Simplifying this yields $\lim_{n \rightarrow \infty} D_n = \int u^2(x) - \sum_{i=1}^n u_i^2$. Using the Koksma–Hlawka inequality this goes to zero as the probability density function of the data points is BV and as it can be considered as the point density measure [4] of the data points. Hence we have

$$\lim_{n \rightarrow \infty} \|u^n - \eta\|_{L^2} = 0$$

Remark 4.3. Although the analysis has been presented on the torus $\mathbb{T}^m$, the results extend immediately to any bounded Lipschitz domain $\Omega \subset \mathbb{R}^m$.

Indeed, after an affine scaling, Ω may be regarded as a Lipschitz subdomain of the unit cube $(0, 1)^m$, which is identified with the torus $\mathbb{T}^m$. The functions and the point density measure are then extended by zero to the complement $\mathbb{T}^m \setminus \Omega$. Since the extension preserves the required regularity on Ω , all estimates established on $\mathbb{T}^m$ apply to the zero-extended functions. Restricting the resulting estimates back to Ω yields the same approximation, stability, and convergence results. Consequently, every theorem in this paper remains valid for arbitrary bounded Lipschitz domains after this scaling and zero-extension procedure.

2. Computational Fill Distance

Let X be a random variable with values in a closed compact $\Omega \subset (0, 1)^m$. Assume Ω has a Lipschitz boundary. Let $p(x)$ be the probability density function of X and assume $p(x) > 0 \forall x \in \Omega$. Now we draw two independent sets of n iid copies of X . We denote them $A_n = \{p_1, p_2 \dots p_n\}$ and $B_n = \{q_1, q_2 \dots q_n\}$.

Definition 4.4. Define ξ_n as the computational fill distance of the data (A_n, B_n) , which is given as

$$\xi_n = \max_{x \in B_n} (d(x, A_n)) = \max_{x \in B_n} \min_{y \in A_n} \|x - y\|_2$$

The rate of convergence of the computational fill distance as the data become dense in a domain is derived here [5] and is given as

$$(4.44) \quad \xi_n = O_P(n^{-1/m} \ln^{1/m} n)$$

In Theorem 4.2 inequalities 4.13 and 4.16 are valid if we replace star discrepancy ζ with computational fill distance ξ , as the rate of convergence of computational fill distance is slower than that of star discrepancy. Hence Theorem 4.2 holds if we replace star discrepancy with computational fill distance.

3. Conclusion

In this paper, we studied the binary regression problem and established convergence results for the proposed estimator under suitable assumptions on the sampling points and the regression function. Error estimates and convergence results were obtained, providing a theoretical justification for the method and clarifying the role of the sampling density in the approximation process.

We further showed that when the feature vectors possess a probability density function of bounded variation, the convergence theory can be strengthened from an expectation-based formulation to deterministic convergence. Although the analysis was presented on the torus $\mathbb{T}^m$, the results extend to arbitrary bounded Lipschitz domains after an affine scaling, embedding into $\mathbb{T}^m$, and zero extension. Thus, the theory is applicable to a broad class of practical data domains.

The results presented here provide a rigorous mathematical foundation for binary regression based on geometric sampling principles. Future work includes extending the analysis to multi-class classification, adaptive sampling strategies, and more general nonparametric learning problems.

4. Data Availability Statement

We declare that no data has been used in this document and none has been made available.

5. Compliance with Ethical Standards

5.1. Funding. This study was not funded by any grant or organization.

5.2. Conflict of Interest. The author Rajesh Dachiraju declares that he has no conflict of interest.

5.3. Ethical approval. This article does not contain any studies with human participants or animals performed by any of the authors.

Bibliography

1. Rajesh Dachiraju, *Approximation of a multivariate function of bounded variation from its scattered data*, 2025.
2. Luc Devroye, László Györfi, and Gábor Lugosi, *A probabilistic theory of pattern recognition*, vol. 31, 01 1996.
3. L.C. Evans and American Mathematical Society, *Partial differential equations*, Graduate studies in mathematics, American Mathematical Society, 1998.
4. *Point density measures: A measure-theoretic framework with applications to geometric probability*, Hexagon preprint hexagon:2610.00203v1, 2026, Contributors: Rajesh Dachiraju.
5. Iosif Pinelis (<https://mathoverflow.net/users/36721/iosif-pinelis>), *A problem on rate of decay of fill distance?*, MathOverflow, URL:<https://mathoverflow.net/q/391340> (version: 2021-05-20).
6. Lauwerens Kuipers and Harald Niederreiter, *Uniform distribution of sequences*, Courier Corporation, 2012.
7. Bernhard Scholkopf and Alexander J. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*, MIT Press, Cambridge, MA, USA, 2001.
8. Weiying Zheng and He Qi, *On Friedrichs–Poincaré-type inequalities*, Journal of mathematical analysis and applications **304** (2005), no. 2, 542–551.