Approximation Methods Project - "Density Estimation in RKHS with Application to Korobov Spaces in High Dimensions" Paper Summary

1 Introduction

The paper [1] present and analayses a kernel-based method to estimate a probability density function from independent samples. Given a density function $f \in N_K$ on a domain Ω with a corresponding kernel K, we get an approximation of the form $\hat{f} = \sum_{j=1}^{N} c_j K(x_j, \cdot)$ where the coefficients c_j are determined by the samples $Y_1, ..., Y_M$ and the points x_j are predetermined points (and are chosen in such a way that will "hopefully" give a good approximation for all the density functions in N_K) i.e. $\hat{f} \in V_N$ where $V_N = Span\{K(x_j, \cdot) | j = 1, ..., N\}$.

2 The Kernel Based Approximation Method

2.1 Reproducing kernel Hilbert space

We assume that the kernel K is positive definite and is of the form:

$$K(x, y) = \sum_{l=0}^{\infty} \beta_l \varphi_l(x) \varphi_l(y) \quad , x, y \in \Omega \tag{1}$$

where $\beta_l \to 0$, $\beta_l > 0$ such that $\{\beta_l \varphi_l(x)\}$ are an orthonormal basis in N_k and $\{\varphi_l(x)\}$ are an orthonormal basis in $L^2(\Omega)$. Under these conditions we get:

$$\langle f, g \rangle_K = \sum_{l=0}^{\infty} \frac{\langle f, \varphi_l(x) \rangle_{L^2} \langle g, \varphi_l(x) \rangle_{L^2}}{\beta_l} \tag{2}$$

We also assume that $\int_\Omega \sqrt{K(x,x)}\, dx < \infty$ and $\int_\Omega K(x,x)\, dx < \infty$ which ensures that $\hat{f} \in L^1(\Omega)$ and that $N_k \subseteq L^2(\Omega)$. The paper also introduces a contiumm of nested hilbert spaces:

$$N_K^\tau = \{v \in L^2(\Omega) | \|v\|_{N_K^\tau} < \infty\} \quad , \quad \|v\|_{N_K^\tau}^2 = \sum_{l=0}^{\infty} \beta_l^{-\tau} |\langle v, \varphi_l \rangle_{L^2(\Omega)}|^2 \tag{3}$$

We get that $N_K^{\tau_2} \subseteq N_K^{\tau_1}$ for $\tau_1 \leq \tau_2$ and higher values of τ corresponds to spaces with smoother functions.

2.2 Density Estimator

The paper presents the ideal minimization problem:

$$\hat{f} = arg \min_{v \in V_N} 0.5 \|v - f\|_{L^2(\Omega)}^2 + 0.5 \lambda \|v\|_K^2 \tag{4}$$

where λ is a regularization parameter. However this minimization problem is not practical since f is unknown.We instead solve a similar minimization problem:

$$\hat{f_Y} = arg \min_{v \in V_N} 0.5 \|v\|_{L^2(\Omega)}^2 + 0.5 \lambda \|v\|_K^2 - \frac{1}{M} \sum_{m=1}^{M} v(Y_m) \tag{5}$$

Which is an equivalent to finding $\hat{f_Y} \in V_N$ which satisfies: $\langle \hat{f_Y}, v \rangle_\lambda = \frac{1}{M} \sum_{m=1}^{M} v(Y_m) \quad v \in V_N$ where $\langle f, g \rangle_\lambda = \langle f, g \rangle_{L^2(\Omega)} + \lambda \langle f, g \rangle_K$ This is easily solvable since N_K is an RKHS so the solution to the problem is unique and is of the form : $\sum_{j=1}^{N} c_j K(x_j, \cdot)$. The coefficients c_j satisfy:

$$Ac = b \quad A_{jk} = \langle K(x_j, \cdot), K(x_k, \cdot) \rangle_{L^2(\Omega)} + \lambda K(x_j, x_k) \quad b_j = \frac{1}{M} \sum_{m=1}^{M} K(x_j, Y_m) \tag{6}$$

We also get that the solution satisfies $\mathbb{E}[\hat{f_Y}] = \hat{f}$ (i.e. we get "on average" the solution to the ideal minimization problem). It is important to note that neither $\hat{f_Y}$ nor $\hat{f}$ are necessarily a valid density function since there are no constraints on $\hat{f}$ to satisfy: $\int_\Omega \hat{f}(x)dx = 1 \quad , \quad \hat{f}(x) \geq 0$

3 Estimation Error

The error of the estimator is measured in terms of the MISE (mean integraded square error) which consists of the estimator bias squared and the estimator variance term:

$$\mathbb{E}[\int_\Omega (\hat{f}_Y(x) - f(x))^2 dx] = \|\hat{f} - f\|^2_{L^2(\Omega)} + \mathbb{E}[\|\hat{f}_Y - \hat{f}\|^2_{L^2(\Omega)}] \tag{7}$$

The paper then analyses and bounds each term. For the bias term the paper proves that:

$$\|\hat{f} - f\|^2_{L^2(\Omega)} \leq \|P_N f - f\|^2_{L^2(\Omega)} + \lambda \|f\|^2_K \tag{8}$$

Where P_N is the projection operator onto V_N. The first term $\|P_N f - f\|^2_{L^2(\Omega)}$ is determined by the "richness" of the space V_N and is affected by the dimension N and the location of the points x_j. The second term $\lambda \|f\|^2_K$ is affected by the regularization parameter λ i.e. we "pay" by a bias error for increasing the regularization error. The paper also shows that for density functions $f \in N^2_K$ we get:

$$\|\hat{f} - f\|^2_{L^2(\Omega)} \leq 3\|P_N f - f\|^2_{L^2(\Omega)} + 8\lambda^2 \|f\|^2_K \tag{9}$$

i.e. for smoother density functions we get a bias squared error that converges faster in λ. For the variance term the paper proves that for $\tau \in (0, 1]$ such that N^τ_K is a RKHS and that $\langle K_\tau(\cdot, \cdot), f\rangle_{L^2(\Omega)} < \infty$ we get:

$$\mathbb{E}[\|\hat{f}_Y - \hat{f}\|^2_{L^2(\Omega)}] \leq \frac{\langle K_\tau(\cdot, \cdot), f\rangle_{L^2(\Omega)}}{M\lambda^\tau} \tag{10}$$

It is important to note that the MISE error bound does not depend on the dimension d of the density function (as long as the projection error doesn't scale with d)

4 Example of Density Functions in Korobov Spaces

The paper anylses the method further in the case of Korobov Spaces. for a d-dimensional density function f defined on the hyper cube $\Omega = [0, 1]^d$ the korobov kernel is of the form:

$$K_\alpha(x, y) = \sum_{h \in \mathbb{Z}^d} r(h, \gamma)^{-1} e^{2\pi i h(x-y)} \tag{11}$$

where α is a smoothness parameter (all mixed derivatives of order $\alpha/2$ are in $L^2(\Omega)$ - roughly speaking), $\gamma_u, \quad u \subseteq \mathbb{N}$ are weights and

$$r(h, \gamma) = \begin{cases} 1 & h = (0, ..., 0) \\ \gamma^{-1}_{supp(h)} \prod_{j \in supp(h)} |h_j|^\alpha & else \end{cases} \tag{12}$$

which can also be written as:

$$K_\alpha(x, y) = 1 + \sum_{\emptyset \neq u \subseteq \{1,..,d\}} \gamma_u (\frac{(2\pi)^\alpha}{(-1)^{\alpha/2+1}\alpha!})^{|u|} \prod_{j \in u} B_\alpha(|\{x_j - y_j\}|) \tag{13}$$

where $B_\alpha(x)$ is the Bernoulli polynomial and $\{x\}$ denotes the fractional part of x. The paper presents a method for bounding the projection bias error: by using the "component by component" algorithm [2] a vector $z \in \{1, ..., d\}^d$ can be computed such that the points $x_j = \{\frac{jz}{N}\}$ satisfy:

$$\|P_N f - f\|^2_{L^2(\Omega)} \leq C\|f\|_{kor,\alpha} \frac{1}{N^{\alpha/4-\delta}} \tag{14}$$

for all $0 < \delta < \frac{\alpha}{4}$ and where C is constant that depends on α, δ, d but can be bound independently of d if the weights γ satisfy some condition.

The paper also shows that for $\tau > \frac{1}{\alpha}$ the hilbert space $N^{\tau}_{kor,\alpha}$ is a RKHS and therefore in the case of korobv space the variance error term can be bound by $C\frac{\|f\|_{L^2(\Omega)}}{M\lambda^{\tau}}$. All in all, we get the following bounds for MISE error in korobov space:

$$\mathbb{E}[\int_{\Omega}(\hat{f}_Y(x) - f(x))^2 dx] \leq C(\|f\|_{kor,\alpha}\frac{1}{N^{\alpha/4-\delta}} + \lambda\|f\|^2_{kor,\alpha} + \frac{\|f\|_{L^2(\Omega)}}{M\lambda^{\tau}}) \tag{15}$$

Morever if $f \in N^2_{kor,\alpha}$ we get:

$$\mathbb{E}[\int_{\Omega}(\hat{f}_Y(x) - f(x))^2 dx] \leq \tilde{C}(\|f\|_{kor,\alpha}\frac{1}{N^{\alpha/4-\delta}} + \lambda^2\|f\|^2_{N^2_{kor,\alpha}} + \frac{\|f\|_{L^2(\Omega)}}{M\lambda^{\tau}}) \tag{16}$$

By choosing the optimal λ and N the paper shows that the MISE error convergence rate can get arbitrarily close to $M^{-1/(1+\frac{1}{\alpha})}$: for $\delta \in (0, \frac{\alpha}{4})$, $\epsilon \in (0, 1-\frac{1}{\alpha}]$, if $f \in N_{kor,\alpha}$ we choose $\lambda^* = M^{-1/(1+\frac{1}{\alpha}+\epsilon)}$, $N^* = \mathcal{O}(M^{1/(0.5\alpha(1+\epsilon)+0.5-0.5\delta(1+\frac{1}{\alpha}+\epsilon)})$ and $\tau^* = \frac{1}{\alpha} + \epsilon$ we get:

$$\mathbb{E}[\int_{\Omega}(\hat{f}_Y(x) - f(x))^2 dx] \leq C\|f\|^2_{kor,\alpha}M^{-1/(1+\frac{1}{\alpha}+\epsilon)} \tag{17}$$

Similarly for $f \in N^2_{kor,\alpha}$ by choosing the optimal λ , N and τ we get:

$$\mathbb{E}[\int_{\Omega}(\hat{f}_Y(x) - f(x))^2 dx] \leq C\|f\|^2_{N^2_{kor,\alpha}}M^{-1/(1+\frac{1}{2\alpha}+\epsilon)} \tag{18}$$

The last bounds are extremely important since it shows that (for the specific case of korobov spaces) the density estimator has a MISE convergence rate very close to the optimal rate of M^{-1} (shown in [3]), and the smoother the density function (i.e. larger alpha for which the function is in the corresponding korobov space) the closer the error convergence rate is to M^{-1}. The paper also shows that the converagence rate of $M^{-1/(1+\frac{1}{\alpha})}$ is asymptotically minimax (i.e. any possible density estimator of functions in $N_{kor,\alpha}$ will not get an error convergence rate that is better $M^{-1/(1+\frac{1}{\alpha})}$), and similarly for $f \in N^2_{kor,\alpha}$.

5 Research Question

Since the density estimator does not necessarily satisfy the the "pdf constraints" the question that arises is what is the solution to the minimization problem of the density estimator under these constraints, and what would be the MISE error of such an estimator? Alternatively, what would be the MISE error of the same estimator that's presented in the paper but with a simple normalization?

It is easy to incorporate the constraint: $\int_{\Omega}\hat{f}(x)dx = 1$. If we denote $S = \int_{\Omega}K(x, x_i)dx$ then the constraint is equivalent to $\sum_{j=1}^{N} c_j = \frac{1}{S}$ which can then be solved by using least squares for the linear system: $A'c_{[1:N-1]} = b'$ where $A'_{i,j} = A_{i,j} - A_{i,N}$, $A' \in \mathbb{R}^{M\times N-1}$ and $b' = b_i - \frac{1}{S}A_{i,N}$. However this solution does not "take care" of the non-negativity constraint and a further analysis on the MISE error of such solution is needed.

References

[1] F. Nobile Y.Kazashi. Density estimation in rkhs with application to korobov spaces in high dimensions. *SIAM Journal on Numerical Analysis*, 61(2):1080–1102, 2023.

[2] D.Nuyens I.H.Sloan R.Cools, F.Kuo. Fast component-by-component construction of lattice algorithms for multivariate approximation with pod and spod weights. *Math. Comp.*, 90:787–812, 2021.

[3] J. M. Steele D. W. Boyd. Lower bounds for nonparametric density estimation rates. *Ann. Statist.*, 6:932–934, 1978.