Statistical estimation with integral-based loss functions

Akifumi Okuno[1,2]

[1]Inst. Stat. Math., [2]RIKEN AIP

This study considers statistical estimation problems using integral-based loss functions, which encompass the following two specific applications.

(1) **Robust estimation using density-power divergence** discussed in Okuno (2023a). Therein, they minimize the following robust loss function equipped with a parametric density p_θ arbitrarily specified by users (e.g., Gaussian mixture model, Gompertz model):

$$-\frac{1}{\beta}\frac{1}{n}\sum_{i=1}^{n} p_\theta(x_i)^\beta + \frac{1}{1+\beta}\int p_\theta(z)^{1+\beta}\mathrm{d}z.$$

(2) **Higher-order variation regularization** discussed in Okuno (2023b). Therein, they minimize the following penalized loss function equipped with a function f_θ arbitrarily specified by users (e.g., neural network, generalized additive model):

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - f_\theta(x_i)\}^2 + \sum_{k=0}^{K}\eta_k \int \left|\frac{\partial^k}{\partial z^k} f_\theta(z)\right|^q \mathrm{d}z.$$

Due to the difficulty in evaluating the integral terms, much of the existing research has concentrated on (i) straightforward models (e.g., normal density estimation, spline estimation) where the explicit form of these integral terms can be obtained, or (ii) numerical integration, which is computationally intensive. Notably, statisticians have long focused on optimization with full-batch methods (e.g., Newton-Raphson method), and the obsession with full-batch methods has made computations challenging.

However, if we rewind the long history of research all the way back to the beginning, such integral-based loss functions have been known to be simply minimized by stochastic gradient descent (Robbins and Monro, 1951). To bridge the substantial, unseen gap between the practical applications in statistics and the profound advancements in stochastic optimization theory, this talk intentionally sheds light on the potential utility of the stochastic optimization techniques.

References

Okuno, A. (2023a). A stochastic optimization approach to minimize robust density power-based divergences for general parametric density models. *arXiv preprint arXiv:2307.05251*.

Okuno, A. (2023b). A stochastic optimization approach to train non-linear neural networks with a higher-order variation regularization. *arXiv preprint arXiv:2308.02293*.

Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407.