A Minimum Distance Weighted Likelihood Method of Estimation

K.Arun Kumar

Indian Statistical Institute

August 2, 2014

The Problem and Disparities

Problem: To estimate the "best" fitting parameter θ_g in the model family $\{f_\theta : \theta \in \Theta\}$ using i.i.d observations $X_1, X_2, \ldots, X_n$ from a continuous density g.

"best" is in the sense that $\theta_g := \text{argmin}_{\theta \in \Theta} \rho(g, f_\theta)$ for some distance ρ. Disparity is a distance of the form

$$\int C\left(\frac{g}{f_\theta} - 1\right) f_\theta dx$$

for some convex function C. We will denote this form of ρ by ρ_C.
Example of Disparity: Hellinger Distance $\int(\sqrt{g} - \sqrt{f_\theta})^2 dx$; $C(t) = (\sqrt{t+1} - 1)^2$.

Minimum Disparity Estimation

Natural Estimator: $\hat{\theta}_n := \text{argmin}_{\theta \in \Theta} \rho_C(\hat{g}_n, f_\theta)$, for some non-parametric density estimator $\hat{g}_n$, usually the kernel density estimator.

- Under appropriate differentiable conditions on C, $\hat{\theta}_n$ solves the equation
$$\int A\left(\frac{\hat{g}_n}{f_\theta} - 1\right) \nabla f_\theta dx = 0,$$
with $A(t) = C'(t)(t+1) - C(t)$.
- Notational convenience: Define $\delta + 1 = g/f_\theta$ and $\delta_n + 1 = \hat{g}_n/f_\theta$.

The Difficulty

- Notice that the objective function and so the estimating function both involve integrals to be computed at each iteration for finding the estimator numerically.
- It can be numerically challenging to deal with integrals on infinite support and/or with multivariate integrals.
- Cheng & Vidyashankar (2006) in this respect propose the use of the MCMC integration technique. This suggestion brings in the new problem of "how many samples should be chosen?"
- Can we use the samples already at hand in simplifying matters?

New Idea

- YES! Notice that

$$\rho_C(g, f_\theta) = \int C(\delta) f_\theta dx = \int \frac{C(\delta) + \delta}{1 + \delta} g dx.$$

- Since the observations are from g, a natural estimator of ρ_C is given by

$$\frac{1}{n} \sum_{i=1}^{n} \frac{C(\delta_n(X_i)) + \delta_n(X_i)}{1 + \delta_n(X_i)}$$

.

- Using this as the objective function, we get the estimating equation as

$$\frac{1}{n} \sum_{i=1}^{n} \frac{A(\delta(X_i)) + 1}{\delta(X_i) + 1} u_\theta(X_i) = 0.$$

This we call minimum distance weighted likelihood estimating equation.

Literature Review

- Joe (1989) gave expression for bias and variance of

$$\frac{1}{n}\sum_{i=1}^{n} J(\hat{f}_n(X_i))$$

which was proposed as an estimator for $\int J(f)fdx$.

- Hall (1987) (1993) proved asymptotic normality in the special case $J(f) = \ln f$.
- Sricharan et al. (2012) (2012) gave different estimates of non-linear density functionals using k-NN density estimate and proved their asymptotic normality.

Asymptotic Normality

- Using techniques from empirical processes, by showing that the class

$$\left\{ A\left(\frac{h}{f_\theta}-1\right)\frac{f_\theta}{h}u_\theta : \sup_x |h(x)-g(x)| < \delta_n \right\}$$

  satisfies Glivenko-Cantelli condition as $n \to \infty$ for some $\delta_n \downarrow 0$, we prove that $\hat{\theta}_n \xrightarrow{P} \theta_g$.
- Here we use that $\sup_x |\hat{g}_n(x) - g(x)| \xrightarrow{P} 0$ as $n \to \infty$. The above result does not require any boundedness conditions related to A. It only requires continuity of A and finiteness of some integral.
- Using results of Van der Vaart & Wellner (2000), we prove that for $g = f_{\theta_0}$ for some $\theta_0 \in \Theta$, $\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(0, I^{-1}(\theta_0))$.

Numerical Study

We consider the model $(1-\varepsilon)N(0,1)+\varepsilon N(0,25)$ with different choices of ε and sample size 100 replicated 100 times and calculated MSE of the mean estimator we get with Hellinger distance (HD), symmetric chi-squared (SCS) disparity and negative exponential disparity (NED).

Table: MSE of Weighted Disparity Estimator

Error(ε)	HD	SCS	NED
0%	0.009338745	0.01471686	0.01801712
5%	0.012974419	0.01668110	0.01652158
10%	0.014531107	0.01481695	0.01427600
20%	0.016173109	0.01595819	0.01587566
30%	0.017963730	0.01858150	0.01710205
40%	0.041553957	0.03473304	0.03020073
50%	0.079105797	0.04595299	0.04897471

Comparison with ?MARK)

- Markatou et al. (1998) proposed robust estimation by weighted likelihood estimating function. They mention as a special case a set of weights which can correspond to disparity estimating equations.
- The major difference with our work is that our estimating function has an objective function. But their estimating function cannot be written as the derivative of an objective function.
- Using our approach we can also estimate the objective function as the mean over the random sample at hand. So, our approach allows for non-differentiable continuous C.

Robustness Properties

- As was mentioned in Markatou et al. (1998), the estimator we get using our method are robust to outliers. Finite sample breakdown properties are yet to be done.

References

Hall, P. (1987), Estimation of integrated square density derivatives, *SPL*, 6(2), 109-115.

Joe, H. (1989), Estimation of Entropy and Other Functionals of Multivariate Density, *AISM*, 41(4), 683-697.

Markatou, M., Basu, A. and Lindsay, B. G. (1998). Weighted likelihood equations with bootstrap root search. *JASA*, 93, 740–750.

Sricharan, K., Raich, R., and Hero, A.O. (2012), Estimation of non-linear functionals of densities with confidence, *IEEE Trans. Infor. Theory*, 58(7), 4135-4159.

Van der Vaart, A. W. and Wellner, J. A. (2000), Preservation theorems for glivenko- cantelli and uniform glivenko-cantelli classes. In Evarist Gine, David M. Mason, and Jon A. Wellner, editors, *High Dimensional Probability II*, volume 47 of *Progress in Probability*, pages 115–133.

Cheng, A. and Vidyashankar, A. N., Minimum Hellinger distance estimation for randomized play the winner design. *Journal of Statistical Planning and Inference*, 136 (6), 1875–1910.