NON-PARAMETRIC MultiTRAJECTORY ESTIMATION

Mikhail Malyutov [1] and Alexandre Tsybakov [2]
[1] *Department of Mathematics, Northeastern University, Boston, U.S.A.*, [2] *Lab. de Probabilités et Modèles Aléatoires, Université Paris 6, Paris, France*

Abstract. The convergence rate for a non-parametric estimation of two distinct smooth trajectories based on pairs of their non-assigned noisy measurements is studied using the kernel estimates for two symmetric functions of observations and the roots of the corresponding parabolic equation. Some simulation results showing the performance of our method are presented.

Keywords: Indistinguishable targets, Symmetric Functions of Measurements, Kernel estimates, roots of random polynomials

1. Introduction and outline of the problem

1.1. Simplified examples

To explain the novelty of the setting and the ideas proposed, we start with a related simplified problem (of interest in the Reliability theory [4]). Suppose, we toss 2 indistinguishable generally biased coins *simultaneously*, and unknown probabilities of heads in a long sequence of independent trials are p_1, p_2 respectively. Suppose that *only the total number of heads* $\sigma(i)$ in trials $i = 1, \ldots, N$, are known. We outline now the Symmetric Functions of Measurements (SFM) method to estimate the set $\{p_1, p_2\}$ of unknown parameters in this and related problems.

Denote $s := p_1 + p_2, \pi := p_1 p_2$ and notice that the set of roots to the quadratic equation $z^2 - sz + \pi = 0$ is exactly $\{p_1, p_2\}$. Next, $s(N) := \sum_{i=1}^{N} \sigma(i)/N$ and N_2/N:=number(outcome 2)$/N$ estimate consistently respectively s and π. Hence the set of roots to the equation $z^2 - s(N) + N_2/N = 0$ estimates consistently the set $\{p_1, p_2\}$ as the number of trials $N \to \infty$. The rate of convergence of the SFM estimates for the set of head probabilities is found in [4] for arbitrary number n of indistinguishable coins flipped simultaneously. If the total number of heads in each of N trials is corrupted by random errors, we can generally propose only the methods based on the EM-algorithm (see [7]) to estimate consistently the set of head probabilities.

A next *static* MultiTrajectory estimation example simplifies the one dealt with in [6]. Suppose in each of N independent experiments we observe 2 real or planar points. Each of them is the noisy measurement $Y_i = a_i + e_i$ of one of fixed centers $a_i, i = 1, 2$ on the line (or on the plane) but *no information is available on assigning random measurements to the centers*. How to estimate the set of the centers consistently?

Introduce $s := a_1 + a_2, \pi := a_1 a_2$ and notice that the set of roots to the quadratic equation $z^2 - sz + \pi = 0$ is exactly $\{a_1, a_2\}$ considered as complex numbers. Denote $\sigma(j) := \sum_{i=1}^{2} Y_i(j), \pi(j) := \prod_{i=1}^{2} Y_i(j)$. Under, say, symmetric gaussian independent errors $e_i(j)$ of j-th measurement $s(N) := \sum_{j=1}^{N} \sigma(j)$ and $\Pi_N := \sum_{j=1}^{N} \pi(j)/N$ estimate consistently respectively s and π. Hence the set of roots to the quadratic equation $z^2 - s(N)z + \Pi(N) = 0$ converges to the set $\{a_1, a_2\}$. The $\sqrt{N}$ rate of convergence is easy to prove given that the centers are different. A more complicated case of multiple roots is treated in [2]. A more general set up of parametric family of trajectories observations corrupted by noise and clutter is studied using a robust version of the EM-algorithm in [7].

1.2. Outline of our method

In this paper we extend the SFM method to the problem where one has noisy observations of two smooth trajectories. An intuitive initial idea is to approximate the trajectories by piecewise constant ones. Then the values of trajectories on the intervals of constancy may be estimated as before. However, such a method has poor convergence properties. To improve the rate of convergence, the moving window (kernel) estimation replaces this naive piecewise approximation approach. More complicated is the estimation of both trajectories near their crossing points, where the roots of the quadratic equation (similar to that considered above) are close to each other. To keep the convergence rate of the method almost the same as for the case of disjoint trajectories, we can first estimate their derivatives by the kernel method applied to SFM, and then restore the trajectories by integration in small neighborhoods of intersection points.

Remark. The SFM method was applied to estimate parameters of MultiTrajectories of fixed or moving (according to a polynomial regression model) targets in [1], [3]. The stepwise algorithm of these papers uses asymptotically infinite divergence of polynomial trajectories, and will likely fail, if the trajectories stay permanently close to each other. The EM-approach of [7] (enabling in addition to estimate the parameters of the random targets-to-observations assignment) and the approach of the present paper seem to be free of this deficiency.

2. Nonparametric setting, non-intersecting trajectories

Here we introduce the MSF estimates in a nonparametric MultiTrajectory setting and estimate their rates of convergence.

Consider the model

$$\begin{aligned} Y_{i1} &= f_1(t_i) + \epsilon_{i1}, \\ Y_{i2} &= f_2(t_i) + \epsilon_{i2}, \quad i = 1, \ldots, n. \end{aligned} \tag{1}$$

Here $f_1(\cdot) : [0,1] \to \mathbf{R}$, $f_2(\cdot) : [0,1] \to \mathbf{R}$, are unknown smooth functions, $t_i = i/n$, and $\epsilon_{i1}, \epsilon_{i2}$ are random variables such that $\epsilon_{11}, \cdots, \epsilon_{n1}$ are i.i.d., $\epsilon_{12}, \cdots, \epsilon_{n2}$ are i.i.d., and the vectors of random variables $(\epsilon_{11}, \cdots, \epsilon_{n1})$ and $(\epsilon_{12}, \cdots, \epsilon_{n2})$ are mutually independent.

We are given unordered pairs of observations $(Y_{11}, Y_{12}), \cdots, (Y_{n1}, Y_{n2})$, such that for each pair of values (Y_{i1}, Y_{i2}) we do not know which value is Y_{i1} and which is Y_{i2}. More accurate estimation would take into account and estimate the parameters of the random assignment mechanism which we will not pursue in this paper. Our problem is to estimate the functions $f_1(\cdot), f_2(\cdot)$ given these observations. We suppose that $f_1(\cdot), f_2(\cdot)$ are smooth, as stated in the following assumption.

Assumption 1. *Let $\beta > 0, L > 0$ and $C_0 > 0$ be finite constants. We assume that the functions $f_1(\cdot), f_2(\cdot)$ belong to $\Sigma(\beta, L, C_0)$, where $\Sigma(\beta, L, C_0)$ is the class of all functions on $[0,1]$ bounded in absolute value by C_0 and such that their derivative of order $\lfloor\beta\rfloor$ satisfies $(\beta - \lfloor\beta\rfloor)$-Hölder condition with constant L, where $\lfloor\beta\rfloor$ is the maximal integer that is strictly less than β.*

Denote by $f_{n1}(x), f_{n2}(x)$ the roots of the quadratic equation

$$Z^2 - s_n(x)Z + \pi_n(x) = 0, \tag{2}$$

if these roots are real, and set $f_{n1}(x) = f_{n2}(x) = 0$ otherwise. Here $s_n(x), \pi_n(x)$ are symmetric functions of measurements introduced below such that, under appropriate conditions

$$s_n(x) \xrightarrow{P} f_1(x) + f_2(x), \quad \pi_n(x) \xrightarrow{P} f_1(x)f_2(x), \quad \text{as } n \to \infty.$$

Hence the quadratic function in (2) converges in probability to the function $F(Z) = Z^2 - (f_1(x) + f_2(x))Z + f_1(x)f_2(x)$ (uniformly in Z on every bounded interval). Clearly, the equation $F(Z) = 0$ has the roots $f_1(x)$ and $f_2(x)$. The roots of (2) converge to $f_1(x), f_2(x)$ in probability as $n \to \infty$ (cf., e.g. [2]) .

Now we define our method. Consider the estimation of f_1, f_2 at arbitrary fixed point $x \in (0,1)$. Define the statistics

$$s_n(x) = \sum_{i=1}^{n} (Y_{i1} + Y_{i2}) W_{ni}(x), \quad \pi_n(x) = \sum_{i=1}^{n} Y_{i1} Y_{i2} W_{ni}(x), \tag{3}$$

where $W_{ni}(x)$ is a weight function such that $\sum_{i=1}^n W_{ni}(x) = 1$ or $\sum_{i=1}^n W_{ni}(x) = 1 + o(1)$, as $n \to \infty$. In the following we consider the kernel weights

$$W_{ni}(x) = \frac{1}{nh} K\left(\frac{t_i - x}{h}\right)$$

where $K : \mathbf{R} \to \mathbf{R}$ is a kernel and $h > 0$ is a bandwidth, but one can also consider other weight functions $W_{ni}(x)$ used in nonparametric estimation problems (see e.g. [5]). We will need the following assumption.

Assumption 2. *The random variables $\epsilon_{i1}, \epsilon_{i2}$ are normal with $E(\epsilon_{i1}) = E(\epsilon_{i2}) = 0$, $E(\epsilon_{i1}^2) = \sigma_1^2 < \infty$, $E(\epsilon_{i2}^2) = \sigma_2^2 < \infty$.*

Theorem. *Let Assumptions 1 and 2 be satisfied, and let $K(\cdot)$ be a compactly supported Lipschitz continuous function such that for $l = \lfloor \beta \rfloor$ we have*

$$\int u^m K(u)du = 0,\ m = 1, \dots, l, \qquad \int K(u)du = 1 \tag{4}$$

(i.e. K is a kernel of order l). Set $h = \alpha n^{-\frac{1}{2\beta+1}}$ for some $\alpha > 0$. If $|f_1(x) - f_2(x)| \geq t_0 n^{-\frac{\beta}{2\beta+1}} \log n$ for some $t_0 > 0$, then

$$\sup_{f_1, f_2 \in \Sigma(\beta, L, C_0)} E_{f_1,f_2}(f_{nj}(x) - f_j(x))^2 \leq C n^{-\frac{2\beta}{2\beta+1}}, j = 1, 2,$$

where $C > 0$ is a finite constant. Here E_{f_1,f_2} denotes the expectation with respect to the joint distribution of $(Y_{i1}, Y_{i2}, i = 1, \dots, n)$ in model (1).

Sketch of proof. Repeating standard argument of nonparametric regression estimation for $s_n(x)$ (see, e.g., [5]), we obtain

$$\sup_{f_1, f_2 \in \Sigma(\beta, L, C_0)} E_{f_1,f_2}\Big((s_n(x) - [f_1(x) + f_2(x)])^2\Big) = O(n^{-\frac{2\beta}{2\beta+1}}),\ n \to \infty. \tag{5}$$

For $\pi_n(x)$ consider separately the bias and variance term. Using the Taylor expansion, the bias of $\pi_n(x)$ can be evaluated as follows

$$\begin{aligned} Bias &= E_{f_1,f_2}(\pi_n(x)) - f_1(x) f_2(x) \\ &= \frac{1}{nh} \sum_{i=1}^n f_1(t_i) f_2(t_i) K\left(\frac{t_i - x}{h}\right) - f_1(x) f_2(x) \\ &= \frac{1}{nh} \sum_{i=1}^n \Big[f_1(x) f_2(x) + (f_1 f_2)'(x)(t_i - x) + \cdots \\ &\quad + \frac{(t_i - x)^l}{(l-1)!} \int_0^1 (f_1 f_2)^{(l)}(x + \tau(t_i - x))(1 - \tau)^{l-1} d\tau \Big] K\left(\frac{t_i - x}{h}\right) - f_1(x) f_2(x) \end{aligned}$$

$$= \frac{1}{nh}\sum_{i=1}^{n}\frac{(t_i-x)^l}{(l-1)!}K\left(\frac{t_i-x}{h}\right)\int_0^1 (f_1f_2)^{(l)}(x+\tau(t_i-x))(1-\tau)^{l-1}d\tau + O\left(\frac{1}{nh}\right)$$
$$= O\left(h^\beta + \frac{1}{nh}\right)$$

where we used that the derivatives $(f_1f_2)^{(m)}$, $m \le l$, are uniformly bounded for $f_1, f_2 \in \Sigma(\beta, L, C_0)$ and that

$$\left|\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{t_i-x}{h}\right)^m K\left(\frac{t_i-x}{h}\right) - \int u^m K(u)du\right| = O\left(\frac{1}{nh}\right).$$

The variance of $\pi_n(x)$ has the form

$$Var = E_{f_1,f_2}\Big((\pi_n(x) - E_{f_1,f_2}[\pi_n(x)])^2\Big) =$$

$$E([\underbrace{\frac{1}{nh}\sum_{i=1}^{n}\xi_{i1}f_2(t_i)K(\frac{t_i-x}{h})}_{Z_1}+$$
$$\underbrace{\frac{1}{nh}\sum_{i=1}^{n}\xi_{i2}f_1(t_i)K(\frac{t_i-x}{h})}_{Z_2}+ \quad + \underbrace{\frac{1}{nh}\sum_{i=1}^{n}\xi_{i1}\xi_{i2}K(\frac{t_i-x}{h})}_{Z_3}]^2.$$

The terms Z_1, Z_2 are treated in similar way, in particular,

$$E(Z_1^2) = \sigma_1^2\frac{1}{n^2h^2}\sum_{i=1}^{n}f_2(t_i)^2K(\frac{t_i-x}{h})$$
$$= O(\frac{1}{nh}),\ n\to\infty,\ \text{since } \sup_x |f_2(x)| \le C_0.$$

For Z_3 we get

$$E(Z_3^2) = \frac{1}{n^2h^2}\sum_{i,k=1}^{n}E(\xi_{i1}\xi_{i2}\xi_{k1}\xi_{k2})K\left(\frac{t_i-x}{h}\right)K\left(\frac{t_k-x}{h}\right)$$
$$= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sigma_1^2\sigma_2^2K^2(\frac{t_i-x}{h})$$
$$= O(\frac{1}{nh}),\quad n\to\infty.$$

Thus

$$\sup_{f_1,f_2\in\Sigma(\beta,L,C_0)} E_{f_1,f_2}[(\pi_n(x) - f_1(x)f_2(x))^2] \tag{6}$$
$$= O(h^{2\beta} + \frac{1}{nh})$$
$$= O(n^{-\frac{2\beta}{2\beta+1}}),\ n\to\infty.$$

It follows from (5), (6) that $s_n(x)$ and $\pi_n(x)$ converge in probability for any x to $s(x) = f_1(x) + f_2(x)$ and $\pi(x) = f_1(x)f_2(x)$ respectively at the rate $n^{-\frac{\beta}{2\beta+1}}$. Hence, the discriminant $(s_n^2(x)/4) - \pi_n(x)$ of the random quadratic equation (2) converges in probability to the discriminant $(f_1(x) - f_2(x))^2/4$ of the equation $F(Z) = 0$ at the same rate. Since we assume that $|f_1(x) - f_2(x)| \geq t_0 n^{-\frac{\beta}{2\beta+1}} \log n$, which is logarithmically larger than the rate of convergence in probability $n^{-\frac{\beta}{2\beta+1}}$, we can guarantee that $(s_n^2(x)/4) - \pi_n(x) > 0$ with probability close to 1 for n large enough. Therefore, with probability close to 1, we have

$$f_{n1}(x) = \frac{s_n(x)}{2} + \sqrt{\frac{s_n^2(x)}{4} - \pi_n(x)}, \tag{7}$$

$$f_{n2}(x) = \frac{s_n(x)}{2} - \sqrt{\frac{s_n^2(x)}{4} - \pi_n(x)}. \tag{8}$$

Using (5), (6) we see that the right hand sides of (7), (8) converge in probability to $f_1(x)$ and $f_2(x)$ respectively at the rate $n^{-\frac{\beta}{2\beta+1}}$. Finally, a uniform integrability argument permits to obtain the convergence of second moments and thus to complete the proof.

3. Regularly intersecting trajectories

Here we extend the above construction for a more complicated case of two smooth ($\beta \geq 2$) possibly intersecting trajectories. Two trajectories will be called regularly intersecting if for all $\Delta > 0$ small enough the inequality $|f_1(x) - f_2(x)| < \Delta$ is satisfied only for x belonging to an interval of length less than $C_1\Delta$, where $C_1 > 0$ is a constant. Here we assume that the trajectories are regularly intersecting and we keep all the conditions and notation of the previous section. This work is still in progress. In particular, we have not yet completed the simulation of the performance of the algorithm outlined below which we apply only in the neighborhoods of the intersection points.

First, note that the derivatives $f_1'(x)$ and $f_2'(x)$ can be estimated consistently at points x such that $|f_1(x) - f_2(x)|$ is not too small. In fact, acting as in the previous section, it is not hard to show that the derivative

$$s_n'(x) = \sum_{i=1}^{n} (Y_{i1} + Y_{i2}) W_{ni}'(x)$$

estimates the function $f_1'(x) + f_2'(x)$ with mean squared error (MSE) of order $n^{-\frac{2(\beta-1)}{2\beta+1}}$ at any point x. Quite similarly,

$$\pi_n'(x) = \sum_{i=1}^{n} Y_{i1} Y_{i2} W_{ni}'(x)$$

estimates $\pi'(x) = (f_1(x)f_2(x))' = f_1'(x)f_2(x) + f_2'(x)f_1(x)$ with MSE of order $n^{-\frac{2(\beta-1)}{2\beta+1}}$. Thus, $f_1'(x)(f_1(x) - f_2(x))$ is estimated with MSE of order $n^{-\frac{2(\beta-1)}{2\beta+1}}$ by

$$v_n(x) = s_n'(x) f_{n1}(x) - \pi_n'(x).$$

Dividing this expression by $f_{n1}(x) - f_{n2}(x)$, we obtain the following estimate of the derivative $f_1'(x)$:

$$f_{n1}^{(1)}(x) = \frac{s_n'(x) f_{n1}(x) - \pi_n'(x)}{f_{n1}(x) - f_{n2}(x)}.$$

Analogously, we define the estimate $f_{n2}^{(1)}(x)$ of the derivative $f_2'(x)$.

Next, we evaluate how close is $f_{nj}^{(1)}(x)$ to $f_j'(x), j = 1, 2$, in the situation where $x = x_n$ is such that $|f_1(x_n) - f_2(x_n)| = \Delta_n > 0$, where $\Delta_n \to 0$, as $n \to \infty$. We will do this only for $j = 1$, since the case $j = 2$ is analogous. Using the above argument and the result of the previous section we obtain that if $\Delta_n > t_0 n^{-\frac{\beta}{2\beta+1}} \log n$,

$$\begin{aligned} |f_{n1}^{(1)}(x_n) - f_1'(x_n)| &= \left| \frac{f_1'(x_n)(f_1(x_n) - f_2(x_n)) + O_p\Big(n^{-\frac{\beta-1}{2\beta+1}}\Big)}{(f_1(x_n) - f_2(x_n)) + O_p\Big(n^{-\frac{\beta}{2\beta+1}}\Big)} - f_1'(x_n) \right| \\ &\leq \frac{O_p\Big(n^{-\frac{(\beta-1)}{2\beta+1}}\Big)}{\Delta_n + O_p\Big(n^{-\frac{\beta}{2\beta+1}}\Big)}. \end{aligned}$$

The last expression tends to 0 in probability only if $\Delta_n \gg n^{-\frac{(\beta-1)}{2\beta+1}}$. For definitiveness, take in what follows $\Delta_n = n^{-\frac{(\beta-1)}{2\beta+1}} \log n$. Then we get

$$|f_{nj}^{(1)}(x_n) - f_j'(x_n)| = O_p\Big(\Delta_n^{-1} n^{-\frac{(\beta-1)}{2\beta+1}}\Big) = O_p(1/\log n), \quad \text{as } n \to \infty. \tag{9}$$

We are now ready to describe our procedure. First, choose a point x_n in a neighborhood of the intersection, i.e. an x_n satisfying $|f_{n1}(x_n) - f_{n2}(x_n)| \approx \Delta_n$, where f_{n1} and f_{n2} are estimators of f_1 and f_2 defined in the previous section. In an $O(\Delta_n)$-neighborhood of x_n we define new adjusted estimators of f_1 and f_2 by the formula

$$\hat{f}_{nj}(x) = f_{nj}(x_n) + f_{nj}^{(1)}(x_n)(x - x_n), \quad j = 1, 2.$$

From (9) and the result of the previous section, using the Taylor expansion of f_j in a neighborhood of x_n (recall that $\beta \geq 2$), we now deduce that

$$\begin{aligned} |\hat{f}_{nj}(x) - f_j(x)| &\leq |f_{nj}(x_n) - f_j(x_n)| + |f_{nj}^{(1)}(x_n) - f_j'(x_n)||x - x_n| \\ &\quad + O\Big((x - x_n)^2\Big) \\ &\leq O_p\Big(n^{-\frac{\beta}{2\beta+1}}\Big) + O_p(1/\log n)\Delta_n + O(\Delta_n^2) \\ &= O_p\Big(n^{-\frac{(\beta-1)}{2\beta+1}}\Big), \quad \text{as } n \to \infty. \end{aligned}$$

Thus, we see that estimation of trajectories around the intersection is possible with the rate $n^{-\frac{(\beta-1)}{2\beta+1}}$ which is slightly slower than the rate $n^{-\frac{\beta}{2\beta+1}}$ obtained aside of the intersection (see the previous section). Such a loss of accuracy seems natural because the problem is more complicated.

4. Simulation Results

We use Java applet prepared by M. Lu to simulate the algorithm described for various parameters of error variance and window size. Here

$$\begin{aligned} f_1(x) &= -\left(x - \frac{1}{2}\right)^2 + \frac{1}{4} \\ f_1(x) &= -\left(x - \frac{1}{2}\right)^2 + \frac{5}{4} \\ K(u) &= \frac{3}{4}(1 - u^2)I(u), \end{aligned}$$

where $x \in$ [0,1] and $I(u) = 1$, if $u \in [0, 1]$, otherwise $I(u) = 0$.

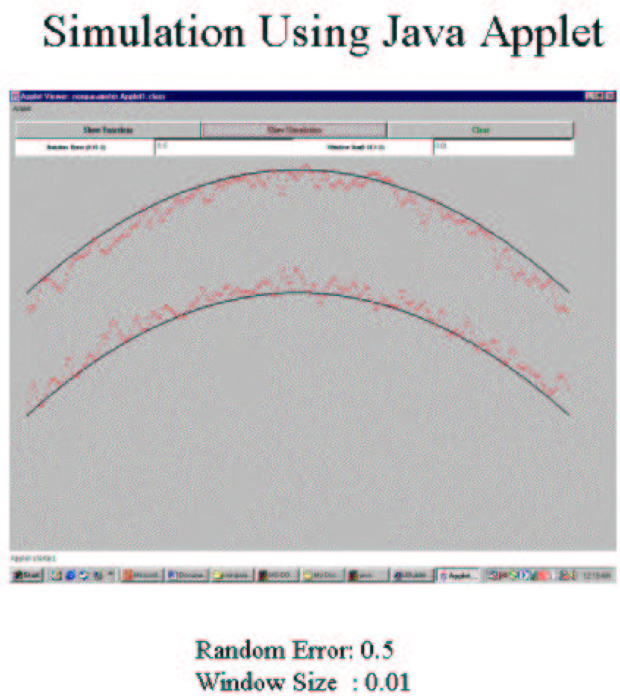

Fig 1 Simulation of the Nonparametric Estimation (M. Lu)

References

1. Bernstein, A.V. Group regression parameter estimation. *Izvestia of USSR Academy of Sciences, Technical Cybernetics*: 137-141, 1973.

2. Bernstein, A.V. Statistical analysis of random polynomials. *Mathematical Methods of Statistics*, **7**, 274-295, 1998.

3. Bernstein, A.V. Statistical analysis of multiple measurements with application to reliability. *Abstracts, 2nd International Conference on Mathematical Methods in Reliability*, Bordeaux, July 4-7, 2000, 191-194, 2000.

4. Bernstein, A.V. and Kagan, A. M. Estimation of failure probabilities from multiple observations. *ibid*, 195-198, 2000.

5. Korostelev, A.P. and A.B. Tsybakov. *Minimax theory of image reconstruction*, Lecture Notes in Statistics, **82**, Springer, N.Y., 1993.

6. Malyutov, M. and I.Tsitovich. Modelling multi-target estimation in noise and clutter. *Proceedings, 12th European Simulation Symposium, ESS 2000 (Simulation in Industry) Sept. 28-30, Hamburg, Germany*, Soc. for Computer Simulation, Delft, Netherlands, 598-600, 2000.

7. Malyutov, M. and Lu, M. Robust Modification of the EM-algorithm for parametric MultiTrajectory estimation in noise and clutter *This volume*, 2003.