Semiparametric Additive Isotonic Regression

Guang Cheng
SAMSI & Duke University

- **Introduction**
  - Background in non/semi-parametric isotonic regression;
- **Main Results**
  - Motivation: Curse of dimensionality;
  - Semiparametric additive isotonic regression;
- **Simulations**
  - Iterative algorithm;
  - Simulation results;
- **Discussions**

Introduction

Model Setting:

$$Y = H(W) + \epsilon, \tag{1}$$

where $H(\cdot)$ is monotonic in each of the coordinates of $W \in R^J$.

Applications:

In *Spatial Epidemiology*, $Y \sim$ the risk of disease; $W \sim$ distance from different classes of radiation to workers.

In *Biostatistics*, $Y \sim$ the proportion of some organism killed by a drug; $W_1 \sim$ the dosage of that drug, $W_2 \sim$ exposure time to that drug.

Isotonic Estimator:

$$\widehat{H}_n \equiv \arg\min \frac{1}{n} \sum (Y_i - H(W_i))^2 \quad \text{w.r.t. monotonicity constrain.}$$

When $J = 1$, we call it *oracle estimator*, denoted as $\widehat{H}_n^{OR}$.

Max-min formula:

$$\widehat{H}_n^{OR}(W_{(i)}) = \max_{s \le i} \min_{t \ge i} \frac{\sum_{j=s}^{t} Y_{[j]}}{t - s + 1}, \tag{2}$$

$$n^{1/3} \left[\frac{2p_W(w)}{\sigma^2 \dot{H}_0(w)} \right]^{1/3} (\widehat{H}_n^{OR}(w) - H_0(w)) \xrightarrow{d} SGCM(B(t) + t^2), \tag{3}$$

where B is a two-sided Brownian motion.

SGCM: slope of the greatest convex minorant.

Remark

- Comparing to other nonparametric estimation methods, the isotonic regression approach is entirely data-driven due to the built-in monotonicity constraint;
- The isotonic estimator $\widehat{H}_n^{OR}(w)$ can be any monotone function passing through $\widehat{H}_n^{OR}(w_{(i)})$s;
- In practice, pool-adjacent-violators algorithm is often used for fitting isotonic regression;
- More details can be referred to Brunk (1970).

Semiparametric Isotonic Regression:

$$Y = X'\beta + H(W) + \epsilon. \tag{4}$$

- $X'\beta \sim$ adjustments for covarites;
- $H(\cdot) \sim$ model nuisance covarites.

Similarly,

$$(\hat{\beta}_n, \hat{H}_n^s) \equiv \arg\min \left(\frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i'\beta - H(W_i))^2 \right). \tag{5}$$

In terms of the asymptotic behaviors of $(\hat{\beta}_n, \widehat{H}_n^s)$,

is model (4) good when J is arbitrarily large?

Main Results

Lemma 1: Given that ϵ has the sub-exponential tail, i.e. $E(\exp(\gamma|\epsilon|)) < C$ for some $\gamma, C > 0$, and $E(X - E(X|W))^{\otimes 2}$ is strictly positive definite, we have

for $J = 2$

$$\|\hat{H}_n^s(W) - H_0(W)\|_2 = O_P(n^{-1/4}(\log n)^2),$$

and for $J \geq 3$

$$\|\hat{H}_n^s(W) - H_0(W)\|_2 = O_P\left(n^{-\frac{1}{4(J-1)}} \log n\right),$$

where $\|\cdot\|_2$ is L_2 norm.

Remark:

- Empirical Processes technique is employed to derive Lemma 1;
- Curse of dimensionality under shape constrains; (Stone 1985, spline estimates, under smoothness conditions);
- Convergence Rates in lemma 1 depend on the condition of ϵ, e.g. $E|\epsilon|^r < \infty \Rightarrow o_P\left(n^{-1/(4J-4)+1/r}\right)$ for any fixed r;
- $\hat{\beta}_n$ is not asymptotically normal since H does not belong to the P-Donsker class when $J > 1$.

Semiparametric Additive Isotonic Regression:

$$Y = X'\beta + \sum_{j=1}^{J} h_j(W_j) + \epsilon$$

is proposed such that

- Circumvent the curse of dimensionality;
- $\hat{\beta}_n$ is asymptotically normal;
- $\hat{h}_j$ has the same asymptotic distribution of $\hat{H}_n^{OR}$ (oracle property).

Note that $(\hat{\beta}_n, \hat{h}_1, \ldots, \hat{h}_J) \equiv \arg\min \left(\frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i'\beta - \sum_{j=1}^{J} h_j(W_{ij}))^2 \right)$

Reference: *Morton-Jones, Diggle, Parker, Dickinson, and Binks (2000)*

Assumptions:

A1. Smoothness conditions on h_j and the density for W_j:

$$\begin{aligned}\inf_{|w_j - w_j'| \geq \delta} |h_j(w_j) - h_j(w_j')| &\quad \gtrsim \quad \delta^\gamma, \\ \sup_{0 \leq w_j, w_j' \leq 1} |p_{W_j}(w_j) - p_{W_j}(w_j')| &\quad \lesssim \quad |w_j - w_j'|^\rho,\end{aligned}$$

for any $\delta > 0$ and some constants $\rho, \gamma > 0$;

A2. Lipschitz continuous condition on $E(X|W_j)$:

$$\|E(X|W_j = w_j) - E(X|W_j = w_j')\| \lesssim |w_j - w_j'|;$$

A3. $E(X - \sum_{j=1}^J E(X|W_j))^{\otimes 2}$ is strictly positive definite.

Note that the parameter identifiability condition in the additive model is weaker since

$$E(X - \sum_{j=1}^{J} E(X|W_j))^{\otimes 2} \geq E(X - E(X|W))^{\otimes 2}.$$

Such relaxation allows the interaction term to enter the model:

$$Y = (W_1 W_2)\beta + h_1(W_1) + h_2(W_2) + \epsilon.$$

If $E(W_1) = E(W_2) = 0$ and $W_1 \perp W_2$, we have

$$\begin{aligned} E(X - E(X|W))^{\otimes 2} &= 0 \\ E(X - \sum_{j=1}^{J} E(X|W_j))^{\otimes 2} &= Var(W_1)Var(W_2) \end{aligned}$$

Theorem 1: Assuming the conditions (A1)-(A3) and the subexponential tail of ϵ, we have

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \sigma^2[E(X - \sum_{j=1}^{J} E(X|W_j))^{\otimes 2}]^{-1}),$$

$$\hat{h}_j(w) - \widehat{H}_n^{OR}(w) = o_P(n^{-1/3}) \text{ for } w \in (0,1)$$

if W is pairwise independent.

References: *Mammen and Yu (2006) and Huang (2002)*

In the semiparametric additive models with $h_j \in L_2(P(W_j))$, the local semiparametric efficiency bound is proved to be:

$$\sigma^{-2} E(X - E(X|\mathcal{F}))^{\otimes 2} = \sigma^{-2} \inf_{\sum f_j \in \mathcal{F}} E(X - \sum_{j=1}^{J} f_j(W_j))^{\otimes 2},$$

where $\mathcal{F}$ is the sum of closed L_2 subspaces.

Therefore,

- The asymptotic variance of $\hat{\beta}_n$ is even smaller than that of local asymptotic efficient estimator since we estimate the projection of Y onto the sum of closed subspace S_1, and closed convex cones $S_{21}, \ldots, S_{2J}$ formed by $h_j(W_j)$s;
- If the monotonicity of h_j is ignored, $\hat{\beta}_n$ will become asymptotically less efficient.

Simulations

Iterative Algorithm:

$$\text{Set} \quad S_n(\beta, h_1, \ldots, h_J) = \left(\frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i'\beta - \sum_{j=1}^{J} h_j(W_{ij}))^2 \right)$$

Let $\beta^{(0)} = 0$ (or $\beta^{(0)} : Y \sim X$) and $k = 0$.

$\Lambda-$**step**: Minimize $S_n(\beta^{(k)}, \Lambda)$ with respect to Λ to obtain $\Lambda^{(k)} \equiv (h_1^{(k)}, \ldots, h_J^{(k)})$ using the backfitting algorithm;

$\beta-$**step**: Minimize $S_n(\beta, \Lambda^{(k)})$ with respect to β (OLS: $(Y - \Lambda^{(k)}(W)) \sim X$). Set $k = k + 1$, and let $\beta^{(k)}$ be the minimizer. Go back to Λ-step.

Remark

- Convergence of the backfitting algorithm in the case that $S_{21}, \ldots, S_{2J}$ are closed convex cones is proved in Dykstra (1983);
- Backfitting estimate $\hat{h} \equiv (\hat{h}_1, \ldots, \hat{h}_J)$ is not unique if there exists a tuple of vectors $(f_1, \ldots, f_J)$ such that $\sum_{j=1}^{J} f_j(w_{ij}) = 0$ for any $i = 1, \ldots, n$ and $f_j + h_j$ is monotone.

Simulation Results:

- $W_1 \sim Unif[-1, 1]$;
- $W_2 \sim$ truncated normal within $[-1, 1]$;
- $\epsilon \sim 0.5N(1, 0.5^2) + 0.5N(-1, 1^2)$;
- $h_1(w_1) = w_1 \exp(-w_1^2/2)$ and $h_2(w_2) = \sin(\pi w_1/2)$;
- $\beta_0 = 1$.

Sample size: 100, 300, 600;

For each sample size, 100 datasets

For each dataset, 500 iterations.

Simulation results for β:

M1. Semiparametric additive isotonic model:

$$Y = (W_1 W_2)\beta + h_1(W_1) + h_2(W_2) + \epsilon;$$

M1(h_1). h_1 is assumed to be known:

$$Y^{(1)} = (W_1 W_2)\beta + h_2(W_2) + \epsilon;$$

M1(h_2). h_2 is assumed to be known:

$$Y^{(2)} = (W_1 W_2)\beta + h_1(W_1) + \epsilon;$$

M1(h_1, h_2). h_1 and h_2 are assumed to be known:

$$Y^{(12)} = (W_1 W_2)\beta + \epsilon;$$

Table 1. *Comparison between* $M1$-$M_1(h_1, h_2)$*(*$\beta_0 = 1$*)*

n	$M1$	$M1(h_1)$	$M1(h_2)$	$M1(h_1, h_2)$
100	0.985 (0.358)	0.987 (0.242)	1.009 (0.237)	0.992 (0.169)
300	0.989 (0.245)	1.005 (0.145)	0.996 (0.137)	0.998 (0.089)
600	0.991 (0.189)	1.006 (0.141)	0.999 (0.103)	0.999 (0.067)

Simulation results for h_1:

M1. Semiparametric additive isotonic model:

$$Y = (W_1 W_2)\beta + h_1(W_1) + h_2(W_2) + \epsilon;$$

M2. Nonparametric additive isotonic model:

$$Y = h_1(W_1) + h_2(W_2) + \epsilon;$$

M3. Oracle Model:

$$Y = h_1(W_1) + \epsilon;$$

Table 2. *Comparison between $M1$, $M2$ and $M3$ (MISE)*

n	$M1$	$M2$	$M3$	$M1/M2$	$M1/M3$
100	0.138	0.113	0.111	1.220	1.240
300	0.081	0.072	0.073	1.122	1.105
600	0.056	0.045	0.046	1.075	1.044

Discussions

- Generalize to $H(Y) = X'\beta + \sum_{j=1}^{J} h_j(W_j) + \epsilon$ with unknown $H(\cdot)$;
- Model selection problem;
- How to relax the pairwise independence assumption of W?

Reference

Brunk, H.D. (1970). Estimation of Isotonic Regression (with discussion) In: Puri, M.L. (Ed.), Nonparametric Techniques in Statistical Inference. Cambridge University Press, Cambridge.

Dykstra, R. (1983). An Algorithm for Restricted Least Squares Regression *Journal of the American Statistical Association* **78** 837–842.

Huang, J. (2002). A note on estimating a partly linear model under monotonicity constraints. *Journal of Statistical Planning and Inference* **107** 345–351.

MAMMEN, E. AND YU, KYUSANG (2007). Additive Isotone Regression *IMS Lecture Notes*

MORTON-JONES, T., DIGGLE, P., PARKER, L., DICKINSON, H.O. AND BINKS, K. (2000). Additive isotonic regression models in epidemiology. *Statistics in medicine* **19** 849–859.

STONE, C.J. (1985). Additive Regression and other nonparametric models *Annals of Statistics* **13** 689–705.