The Weak-Feature-Impact Phase Transition of the NPMLE in Monotone Binary Regression

Dario Kieffer
joint work with Angelika Rohde

Mathematisches Institut
Albert-Ludwigs-Universität Freiburg

Vienna, September 18, 2025

DFG Research Unit FOR 5381

Overview

Introduction

- Feature-label pair $(X, Y) \in \mathbb{R} \times \{0, 1\}$ related via

$$\Phi_0(X) = \mathbb{P}(Y = 1|X).$$

- Φ_0 not necessarily continuous but **monotonically increasing**.

"The larger X, the more probable is label 1"

Statistical problem

Based on $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} (X, Y)$, infer on Φ_0.

- If – except for monotonicity – no additional information on Φ_0 is available, nonparametric MLE (NPMLE) is the natural choice.

The NPMLE $\hat{\Phi}_n$

$$\hat{\Phi}_n \in \operatorname*{Argmax}_{\Phi \in \mathcal{F}_\uparrow} \sum_{i=1}^{n} \log(p_\Phi(X_i, Y_i)),$$

where $p_\Phi(x, y) := \Phi(x)^y (1 - \Phi(x))^{1-y}$ for $\Phi \in \mathcal{F}_\uparrow$, i.e.

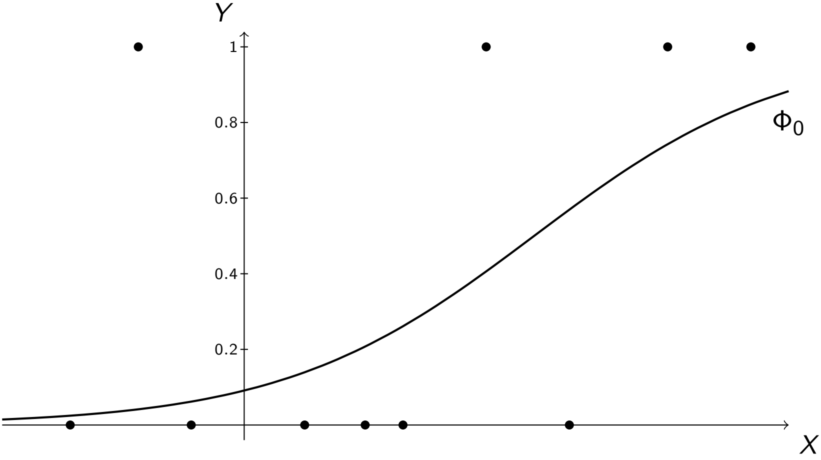
$$\hat{\Phi}_n \in \operatorname*{Argmax}_{\Phi \in \mathcal{F}_\uparrow} \sum_{i=1}^{n} Y_i \log \Phi(X_i) + (1 - Y_i) \log(1 - \Phi(X_i)).$$

- Existence $\surd$, Uniqueness at the sample points (in case of pairwise difference) $\surd$ (cf. [Groeneboom and Wellner (1992)]).
- By [Grenander (1956)], $\big(\hat{\Phi}_n(X_{(1)}), \ldots, \hat{\Phi}_n(X_{(n)})\big)$ is given via left-hand derivative of the greatest convex minorant of

$$\left\{ \left(i, \sum_{j \leq i} Y_{(j)} \right) : i \in \{1, \ldots, n\} \right\}.$$

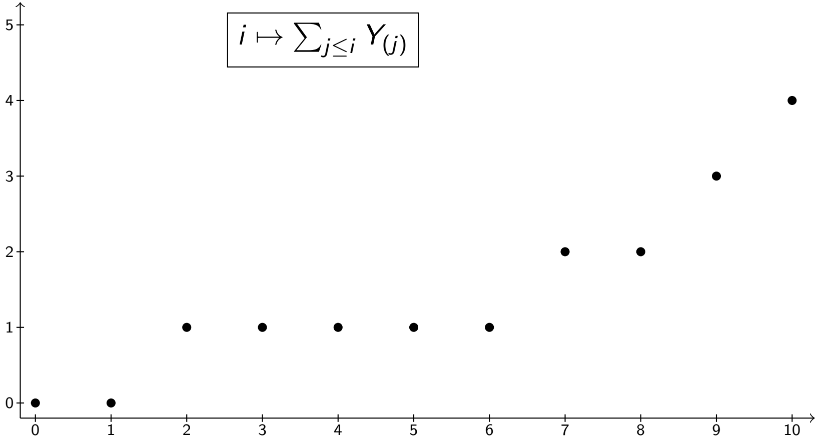
Graphical derivation of $\hat{\Phi}_n$

Observations $(X_1, Y_1), \ldots, (X_{10}, Y_{10})$



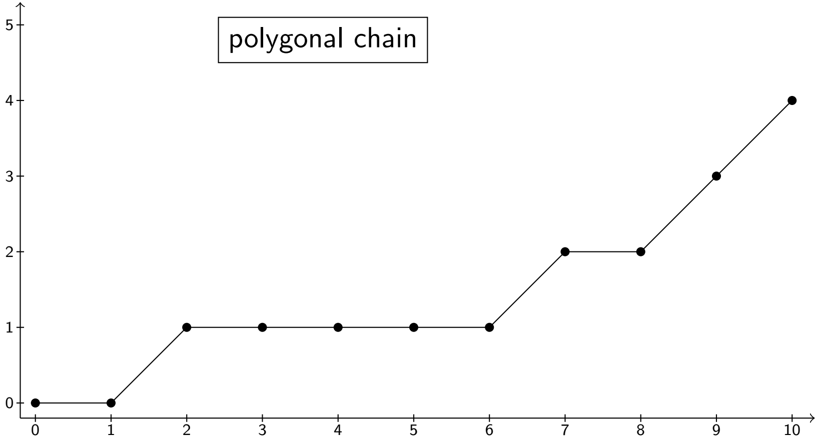
Graphical derivation of $\hat{\Phi}_n$

$Y_{(1)}, \ldots, Y_{(10)}$ ordered based on the order statistic $X_{(1)}, \ldots, X_{(10)}$



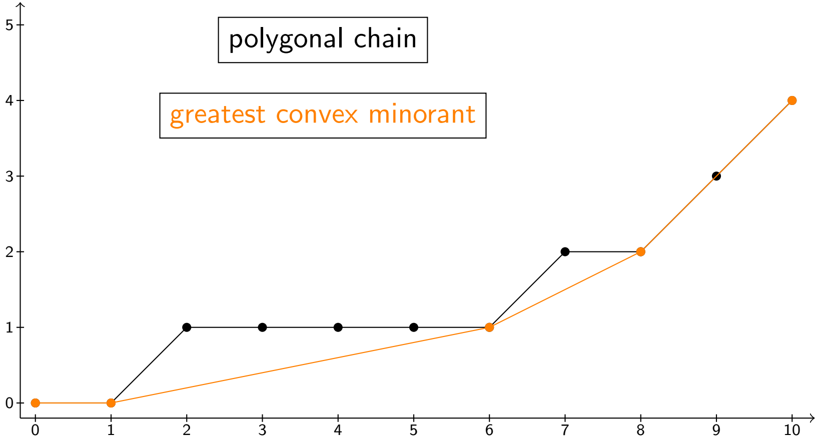
Graphical derivation of $\hat{\Phi}_n$

$Y_{(1)}, \ldots, Y_{(10)}$ ordered based on the order statistic $X_{(1)}, \ldots, X_{(10)}$



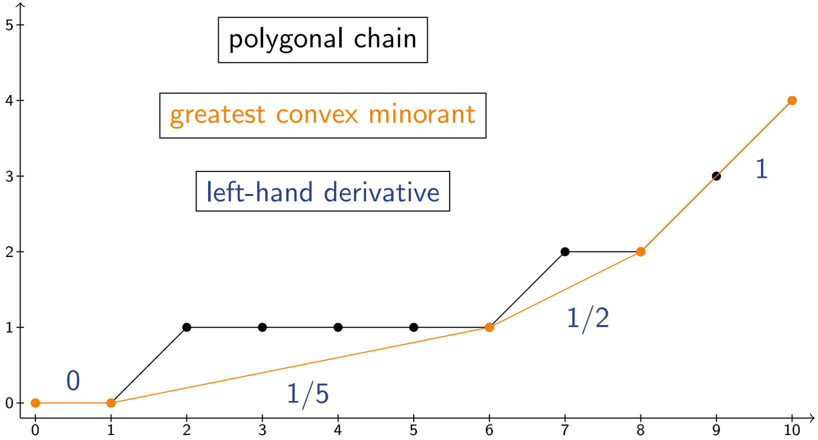
Graphical derivation of $\hat{\Phi}_n$

$Y_{(1)}, \ldots, Y_{(10)}$ ordered based on the order statistic $X_{(1)}, \ldots, X_{(10)}$

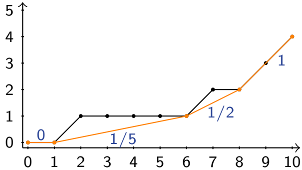


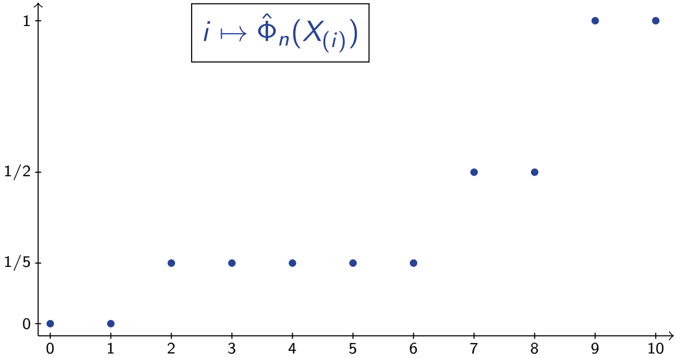
Graphical derivation of $\hat{\Phi}_n$

$Y_{(1)}, \ldots, Y_{(10)}$ ordered based on the order statistic $X_{(1)}, \ldots, X_{(10)}$

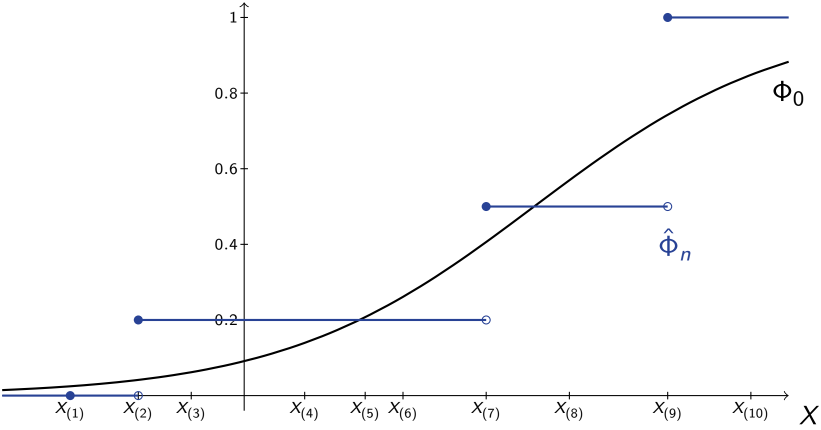


Graphical derivation of $\hat{\Phi}_n$





Graphical derivation of $\hat{\Phi}_n$



State of the art

- Pointwise convergence rate and limiting distribution [Grenander (1957), Rao (1969), Wright (1981), Groeneboom (1985), Groeneboom and Jongbloed (2014), ...]
  - For locally flat functions [Carolan and Dykstra (1999)]
  - In the misspecified case [Patilea (2001), Jankowski (2014)]
  - Unified approach [Westling and Carone (2020)]
  - Generalization to n-dependent monotonically increasing functions with possibly locally changing shape [Mallick, Sarkar and Kuchibhotla (2023)]
- L^p-error (rate of convergence and limiting distribution) [Groeneboom, Hooghiemstra and Lopuhaä (1999), Kulikov and Lopuhaä (2005), Durot (2002, 2007, 2008), ...]
- Minimax-optimality [Cator (2011), Chatterjee, Guntuboyina and Sen (2015), ...]
- Berry-Esseen bounds for Chernoff-type limits [Han and Kato (2022)]
- Bootstrap [Sen, Banerjee and Woodroofe (2010), Cattaneo, Jansson and Nagasawa (2024)]

State of the art — The pointwise limiting distribution

Results concerning the pointwise limiting distributions depend heavily on whether Φ_0 is (locally) strictly increasing or flat:

Φ_0 locally strictly increasing

$\Phi_0'(x) \neq 0$:

- $n^{1/3}$-consistency
- Chernoff limit

$\Phi_0^{(\beta)}(x)$ first derivative $\neq 0$

- $n^{\beta/(2\beta+1)}$-consistency
- Chernoff-type limit

Nonparametric regime

Φ_0 locally flat around x:

- $n^{1/2}$-consistency
- Left-derivative of GCM of Brownian motion at $F_X(x)$ as limit

Parametric regime

Weak feature-impact

What happens if the impact of the features on the labels is **weak**?

Motivation: A weak feature-label relationship occurs frequently in practical applications.
For example, privacy preserving requirements may diminish the isolated effect of the features on the labels considerably.

- The extremal case of no impact corresponds to Φ_0 being constant.
- A very steep increase of Φ_0 from 0 to 1 or even a jump function is what one might consider as fully related.

Weak feature-impact

What happens if the impact of the features on the labels is **weak**?

Motivation: A weak feature-label relationship occurs frequently in practical applications.
For example, privacy preserving requirements may diminish the isolated effect of the features on the labels considerably.

- The extremal case of no impact corresponds to Φ_0 being constant.
- A very steep increase of Φ_0 from 0 to 1 or even a jump function is what one might consider as fully related.

Here, weakness is colloquially understood as "close to flatness" of the feature-label relationship $x \mapsto \mathbb{P}(Y = 1|X = x)$.

Weak feature-impact

- Simulations indicate that the flatter the slope, the later the established limiting distributions kick in.
  - $\Rightarrow$ Large-sample-size asymptotic does not provide a suitable description of the finite sample situation.

Question: How can this small-sample effect be explained on a rigorous mathematical basis?

$\Rightarrow$ Weakness in the sense of "almost flatness" of a feature-label relationship has to be put into relation with the sample size to make its presence visible.

Slightly rephrased question: How does the transition from the nonparametric to the parametric regime look like in terms of distributional approximation?

The weak-feature-impact scenario

Let $(X_1, Y_1^n), \ldots, (X_n, Y_n^n) \overset{iid}{\sim} (X, Y^n) \in \mathbb{R} \times \{0,1\}$, related via

$$\mathbb{P}\big(Y^n = 1 \big| X\big) = \Phi_0(\delta_n X) =: \Phi_n(X)$$

for some monotonically increasing function Φ_0 and *level of feature impact* $\delta_n \searrow 0$.

- Each Φ_n inherits qualitative properties of Φ_0, such as (strict) monotonicity or (k-fold) differentiability.
- Φ_n changes its steepness but not its shape.
- In case Φ_0 is continuously differentiable,

$$\Phi_n'(x) = \delta_n \Phi_0'(\delta_n x) = \delta_n\big(\Phi_0'(0) + o(1)\big).$$

- If $\Phi_0' > 0$, $(\delta_n)_{n\in\mathbb{N}}$ characterizes the speed in which the derivative of $x \mapsto \mathbb{P}(Y^n = 1 | X = x)$ approaches zero, **uniformly on compacts**.
- Note that a weak feature-label relation is a global property and hence cannot be modeled locally solely.

Pointwise theory

Minimax lower bound over restricted classes

- **Crucial aspect:** Level of feature impact controls the gradient of the feature-label relation uniformly on compacts, both from above **and** from below.

$$
\begin{aligned}
\mathcal{F}_\delta :=& \Big\{ \Phi \in \mathcal{F}_\uparrow \mid \text{ 'steepness' of } \Phi \text{ between } \delta/2 \text{ and } \delta \Big\} \\
=& \Big\{ \Phi \in \mathcal{F}_\uparrow \mid \underset{\substack{\uparrow \\ \text{Lipschitz} \\ \text{seminorm}}}{\|\Phi\|_{\mathcal{X},L}} \leq \delta \text{ and } \inf_\nu \underset{\substack{\uparrow \\ \text{modulus} \\ \text{of continuity}}}{\omega_\nu^{\mathcal{X}}}(\Phi)/\nu \geq \delta/2 \Big\}, \quad \delta \in [0,1].
\end{aligned}
$$

- Φ_0 continuously differentiable and $\Phi_0'(0) \in (1/2, 1)$

$$\Downarrow$$

$\Phi_n = \Phi_0(\delta_n \bullet) \in \mathcal{F}_{\delta_n}$ for n sufficiently large.

Minimax lower bound over restricted classes

Theorem (Pointwise lower bound)

For any x_0 contained in the interior of $\mathcal{X}$, there exists a positive constant $C > 0$, such that

$$\liminf_{n\to\infty} \inf_{\delta\in[0,\frac{1}{4T}]} \inf_{T_n^\delta(x_0)} \sup_{\Phi\in\mathcal{F}_\delta} P_\Phi^{\otimes n}\left(\left(\sqrt{n} \wedge \left(\frac{n}{\delta}\right)^{1/3}\right)\left|T_n^\delta(x_0) - \Phi(x_0)\right| \geq C\right) > 0,$$

where the infimum is running over all estimators

$$T_n^\delta(x_0) = T_n^\delta\big(x_0, (x_1, y_1), \ldots, (x_n, y_n)\big).$$

Minimax lower bound over restricted classes

Theorem (Pointwise lower bound)

For any x_0 contained in the interior of $\mathcal{X}$, there exists a positive constant $C > 0$, such that

$$\liminf_{n\to\infty} \inf_{\delta\in[0,\frac{1}{4T}]} \inf_{T_n^\delta(x_0)} \sup_{\Phi\in\mathcal{F}_\delta} P_\Phi^{\otimes n}\left(\left(\sqrt{n} \wedge \left(\frac{n}{\delta}\right)^{1/3}\right)\left|T_n^\delta(x_0)-\Phi(x_0)\right| \geq C\right) > 0,$$

where the infimum is running over all estimators

$$T_n^\delta(x_0) = T_n^\delta\big(x_0, (x_1, y_1), \ldots, (x_n, y_n)\big).$$

- The lower bound exhibits an **elbow** at $\delta \sim \frac{1}{\sqrt{n}}$.
- Result remains true for continuously differentiable functions by smoothing out kinks in the hypotheses.

Asymptotics of the NPMLE

Questions:

- Does the NPMLE in the weak-feature-impact scenario attains the minimax lower bound on the convergence rate?
- If it attains the lower bound, how is the limiting distribution affected by the elbow in the convergence rate?
- Is there a phase transition in the limiting distribution?

Convergence rates & limiting distributions

Theorem

For $\beta \in \mathbb{N}$, let x_0 be an interior point of $\mathcal{X}$ and assume Φ_0 to be β-times continuously differentiable in a neighborhood of zero with the βth derivative being the first non-vanishing derivative in zero.

(i) *(Slow regime) If $n\delta_n^{2\beta} \longrightarrow \infty$,*

Chernoff-type limit

$$\left(\frac{n}{\delta_n}\right)^{\beta/(2\beta+1)} \big(\hat{\Phi}_n(x_0) - \Phi_n(x_0)\big) \longrightarrow_{\mathcal{L}} f_\beta^{*,\ell}(0) \quad \text{as } n \longrightarrow \infty.$$

(ii) *(Fast regime) If $n\delta_n^{2\beta} \longrightarrow 0$,*

does not depend on β

$$\sqrt{n}\big(\hat{\Phi}_n(x_0) - \Phi_n(x_0)\big) \longrightarrow_{\mathcal{L}} g_{\beta,0}^{*,\ell}(F(x_0)) \quad \text{as } n \longrightarrow \infty.$$

(iii) *(Boundary case) Let the inverse F_X^{-1} be Hölder continuous to the exponent $\alpha > 1/2$. If $n\delta_n^{2\beta} \longrightarrow c \in (0, \infty)$,*

$$\sqrt{n}\big(\hat{\Phi}_n(x_0) - \Phi_n(x_0)\big) \longrightarrow_{\mathcal{L}} g_{\beta,c}^{*,\ell}(F(x_0)) \quad \text{as } n \longrightarrow \infty.$$

Convergence rate

Does the NPMLE in the weak-feature-impact scenario attains the minimax lower bound on the convergence rate?

- If Φ_0 is differentiable at 0 with $\Phi_0'(0) > 0$,

$$\left(\sqrt{n} \wedge \left(\frac{n}{\delta_n}\right)^{1/3}\right) \left| \hat{\Phi}_n(x_0) - \Phi_n(x_0) \right| \;=\; \mathcal{O}_{\mathbb{P}}(1),$$

  i.e. the NPMLE attains the lower bound adaptively in the weak-feature-impact scenario.

Limiting distribution — Phase transition

How is the limiting distribution affected by the elbow in the convergence rate?
Is there a phase transition in the limiting distribution?

- The distributional behavior of $\hat{\Phi}_n$ changes at the critical level

$$\delta_n \sim \frac{1}{\sqrt{n}} \text{ of feature impact:}$$

  - Slow regime $\delta_n \sqrt{n} \longrightarrow \infty$ — Chernoff limit (but faster rate of convergence)
  - Boundary case $\delta_n \sqrt{n} \longrightarrow c > 0$ — New limit
  - Fast regime $\delta_n \sqrt{n} \longrightarrow 0$ — Limit for locally flat functions

Comparison to literature

Comparison to [Mallick, Sarkar and Kuchibhotla (2023)]:

- We describe a global property, they describe a local property (focused on pointwise results).
- We characterize a new "intermediate" limiting distribution which cannot be learned from their results.
- They only consider the slow regime, but the class of arising limiting distributions is richer than here, as they allow for changing the qualitative properties of the elements of the sequence $(\Phi_n)_{n\in\mathbb{N}}$.

$\Rightarrow$ Their results and our results are non-nested

Sketch of proof

Central ingredient: The switch relation

Based on the characterization of the NPMLE as the left-hand derivative of the greatest convex minorant, we obtain the **switch relation**, which implies that the process

$$U_n : [0,1] \to \mathbb{R},$$
$$U_n(a) := \underset{x \in \mathbb{R}}{\operatorname{argmin}^+} \left\{ \frac{1}{n} \sum_{i=1}^{n} Y_i^n \mathbb{1}_{\{X_i \leq x\}} - a \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \right\}$$

behaves like a generalized inverse to $\hat{\Phi}_n$:

Lemma (Groeneboom (1985))

For every $x_0 \in \mathcal{X}$ and every $a \in \mathbb{R}$, we have almost surely

$$\hat{\Phi}_n(x_0) > a \quad \Longleftrightarrow \quad U_n(a) < F_n^{-1}(F_n(x_0)).$$

Sketch of proof

- For arbitrary $v \in \mathbb{R}$, start from

$$\mathbb{P}\big(r_n\big(\hat{\Phi}_n(x_0) - \Phi_n(x_0)\big) \leq v\big)$$
$$= \mathbb{P}\Big(\hat{\Phi}_n(x_0) \leq \Phi_n(x_0) + \frac{v}{r_n}\Big)$$

and apply the switch relation

$$= \mathbb{P}\Bigg(\operatorname*{argmin}^+_{s\in[-T,T]} \Bigg\{\frac{1}{n}\sum_{i=1}^{n}\big(Y_i^n - \Phi_n(x_0)\big)\mathbb{1}_{\{X_i \leq s\}} - \frac{v}{r_n}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i \leq s\}}\Bigg\} \geq F_n^{-1}(F_n(x_0))\Bigg).$$

Sketch of proof

- Multiplication by a positive constant or adding a constant inside the argmin does not change the location of the argmin.

  $\Rightarrow$ allows to generate suitable norming sequences
  (Note that the labels are **n-dependent**).
- Suitably transform the argument s.
- Prove weak convergence of the process inside the argmin.
- Apply the argmax-continuous-mapping theorem, provided the real number on the right-hand side is a continuity point of the limit (cf. [Cattaneo, Jansson, Nagasawa (2024)]).
- Switch back (utilizing a general switch relation for the continuous limit) to get the v back to the right-hand side.

$\square$

Theory of the L^1-error

L^1-error under weak feature impact

Recall: A weak feature-label relation is a **global** property.

$\Rightarrow$ Reasonable to study the statistical behavior of the L^1-error

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi(t)| dt$$

in the weak-feature-impact scenario.

L^1-error — State of the art

Rate of consistency

- If Φ_0 is strictly isotonic, then

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt = \mathcal{O}_{\mathbb{P}}(n^{-1/3}).$$

- If Φ_0 is constant, then

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

$\rightarrow$ Local adaptation for the global estimation problem in terms of oracle inequalities [Chatterjee et al. (2015) and Bellec (2018)].

L^1-error — State of the art

Rate of consistency

- If Φ_0 is strictly isotonic, then

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt = \mathcal{O}_{\mathbb{P}}(n^{-1/3}).$$

- If Φ_0 is constant, then

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

$\rightarrow$ Local adaptation for the global estimation problem in terms of oracle inequalities [Chatterjee et al. (2015) and Bellec (2018)].

Limiting distribution

If Φ_0 is strictly isotonic, then

$$n^{1/6}\left(n^{1/3}\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt - \mu\right) \longrightarrow_{\mathcal{L}} N \sim \mathcal{N}(0, \sigma^2)$$

provided Φ_0 possesses a Hölder continuous derivative and X a continuous strictly positive Lebesgue density on $[-T, T]$.

L^1-error — State of the art

Rate of consistency

- If Φ_0 is strictly isotonic, then

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt = \mathcal{O}_{\mathbb{P}}(n^{-1/3}).$$

- If Φ_0 is constant, then

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt = \mathcal{O}_{\mathbb{P}}(n^{-1/2}).$$

$\rightarrow$ Local adaptation for the global estimation problem in terms of oracle inequalities [Chatterjee et al. (2015) and Bellec (2018)].

Limiting distribution

If Φ_0 is strictly isotonic, then

$$n^{1/6}\left(n^{1/3}\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_0(t)| dt - \mu\right) \longrightarrow_{\mathcal{L}} N \sim \mathcal{N}(0, \sigma^2)$$

provided Φ_0 possesses a Hölder continuous derivative and X a continuous strictly positive Lebesgue density on $[-T, T]$.

No limiting distribution for the flat case

Minimax lower bound over restricted classes

Recall the restricted classes

$$\mathcal{F}_\delta := \Big\{\Phi \in \mathcal{F}_\uparrow \mid \text{ 'steepness' of } \Phi \text{ between } \delta/2 \text{ and } \delta\Big\}$$
$$= \Big\{\Phi \in \mathcal{F}_\uparrow \mid \underbrace{\|\Phi\|_{\mathcal{X},L}}_{\substack{\text{Lipschitz}\\ \text{seminorm}}} \leq \delta \text{ and } \inf_\nu \underbrace{\omega_\nu^{\mathcal{X}}(\Phi)}_{\substack{\text{modulus}\\ \text{of continuity}}}/\nu \geq \delta/2\Big\}, \quad \delta \in [0,1].$$

Theorem (L^1-lower bound)

$$\liminf_{n\to\infty} \inf_{\delta \in [0, \frac{1}{4T}]} \inf_{T_n^\delta} \sup_{\Phi \in \mathcal{F}_\delta} \left(\sqrt{n} \wedge \left(\frac{n}{\delta}\right)^{1/3}\right) \mathbb{E}_\Phi^{\otimes n}\left[\int_{-T}^{T} \left|T_n^\delta(t) - \Phi(t)\right| dt\right] > 0,$$

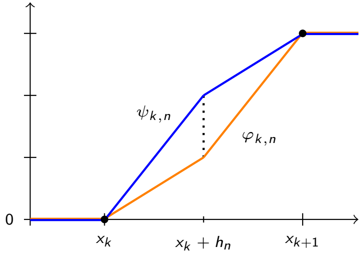
where the infimum is running over all estimators

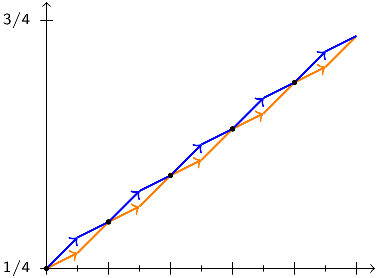
$$T_n^\delta = T_n^\delta\big(\bullet, (x_1, y_1), \ldots, (x_n, y_n)\big).$$

Minimax lower bound over restricted classes

Hypotheses construction in the slow regime

The proof is based on Assouad's hypercube technique.





$\varphi_{n,k}$ and $\psi_{n,k}$ are the base functions to construct the hypotheses. Note that the pointwise distance between these two functions at $x_k + h_n$ is of order $(n/\delta)^{-1/3}$, with $h_n \sim (n\delta^2)^{-1/3}$.

Adaptivity in the weak-feature-impact scenario

Proposition

Suppose that Φ_0 is continuously differentiable with $\Phi_0'(0) > 0$. Then

$$\left(\sqrt{n} \wedge \left(\frac{n}{\delta_n}\right)^{1/3}\right)\mathbb{E}\left[\int_{-T}^{T} \left|\hat{\Phi}_n(t) - \Phi_n(t)\right| dt\right] = \mathcal{O}(1)$$

in the weak-feature-impact scenario.

Sketch of proof

On basis of Fubini's theorem and partial integration, the idea is

- to rewrite

$$\begin{aligned}&\mathbb{E}\int_{-T}^{T}\left|\hat{\Phi}_n(t)-\Phi_n(t)\right|dt\\&=\int_{-T}^{T}\int_0^1\mathbb{P}\left(\hat{\Phi}_n(t)-\Phi_n(t)>x\right)dx\,dt+\int_{-T}^{T}\int_0^1\mathbb{P}\left(\Phi_n(t)-\hat{\Phi}_n(t)>x\right)dx\,dt,\end{aligned}$$

- to employ the switch relation in the probabilities inside the integrals to move over to the inverse process,
- and to derive by means of the slicing device and the Dvoretzky-Kiefer-Wolfowitz inequality a tail bound for the latter.

Here, the level of feature impact δ_n, i.e. the exact dependence on the derivative Φ_n', starts to matter.

Sketch of proof

- Whereas NPMLE and inverse process both scale at the rate $n^{1/3}$ in the classical asymptotics, their convergence rates do not coincide in the weak-feature-impact scenario any longer: The inverse process scales pointwise at the rate

$$(n\delta_n^2)^{1/3}.$$

- It is insightful to contrast its rate with the convergence rate

$$(n/\delta_n)^{1/3}$$

of the NPMLE in the slow regime.

Sketch of proof

- In the parametric regime ($n\delta_n^2 = \mathcal{O}(1)$), arguing by means of the inverse process is subtle as it is not everywhere convergent any longer.
- However, the interval of non-convergence turns out to have a length of order δ_n only. Combining this with sufficiently fast convergence outside this interval bounds the expected L^1-error in the fast regime.

$\square$

Limiting distribution theory

- Let Φ_0 be differentiable in a neighborhood of zero with $\Phi_0'(0) > 0$.
- Let p_X be continuously differentiable on $[-T, T]$ (one-sided at $-T, T$) and assume that Φ_0' is Hölder-continuous in a neighborhood of zero.

Theorem (Limiting distribution of the L^1-error)

(i) *(Slow regime) If $n\delta_n^2 \longrightarrow \infty$, then*

$$(n\delta_n^2)^{1/6}\Bigg(\underbrace{\left(\frac{n}{\delta_n}\right)^{1/3}\int_{-T}^{T}|\hat{\Phi}_n(t)-\Phi_n(t)|dt}_{\textit{stabilized } L^1\textit{-error}}-\underset{\substack{\uparrow\\ \mathcal{O}(1)}}{\mu_n}\Bigg) \longrightarrow_{\mathcal{L}} N \sim \mathcal{N}(0,\sigma^2)$$

$(n\delta_n^2)^{1/3} =$ convergence rate of the inverse process
(collapes at the phase transition!)

Limiting distribution theory

- Let Φ_0' be continuous in a neighborhood of zero.

Theorem (Limiting distribution of the L^1-error – continued)

(ii) *(Fast regime) If $n\delta_n^2 \longrightarrow 0$,*

$$\sqrt{n}\int_{-T}^{T} |\hat{\Phi}_n(x) - \Phi_n(x)| dP_X(x) \longrightarrow_{\mathcal{L}} \max_{s\in[-T,T]} A(s),$$

where $(A(s))_{s\in[-T,T]}$ is a continuous, centered Gaussian process with $A(-T) = -A(T)$ and

$$\mathrm{Cov}(A(s), A(t)) = \Phi_0(0)(1-\Phi_0(0))(1-2|F_X(s)-F_X(t)|)$$

for $s, t \in [-T, T]$.

Sketch of proof — Slow regime

- Rewrite

$$\int_{-T}^{T} |\hat{\Phi}_n(t) - \Phi_n(t)| dt = \int_{-T}^{T} \left(\hat{\Phi}_n(t) - \Phi_n(t)\right)_+ dt + \int_{-T}^{T} \left(\Phi_n(t) - \hat{\Phi}_n(t)\right)_+ dt.$$

- Use Cavalieri's principle to each of the summands on the RHS

$$\int_{-T}^{T} \int_{0}^{\Phi_n(T)-\Phi_n(t)} \mathbb{1}_{\{\hat{\Phi}_n(t) > \Phi_n(t)+u\}} du\, dt \ + \ \dots.$$

- Employ the switch relation and prove the approximation

$$\int_{\Phi_n(-T)}^{\Phi_n(T)} \left(\Phi_n^{-1}(a) - F_n^{-1} \circ \underset{\substack{\uparrow \\ \text{Inverse process}}}{\tilde{U}_n(a)}\right) \mathbb{1}_{\{F_n^{-1} \circ \tilde{U}_n(a) < \Phi_n^{-1}(a)\}} da \ + \ o_{\mathbb{P}}(n^{-1/2}) \ + \ \dots$$

Sketch of proof — Slow regime

- Bring a first strong Gaussian approximation (KMT construction) into play — F_n is the empirical distribution function!
- Incorporate the approximating Brownian bridges into the inverse process.
- Bring some further strong Gaussian approximation into play, namely standard Brownian motions W_n given $X_1, \ldots, X_n$, conditionally independent of the Brownian bridges of the KMT approximation, that approximate an arising partial sum process (given $X_1, \ldots, X_n$) [Sakhanenko (1985)].
- Technical arguments concerning the n-dependence of the arising integrals and the integrand.
- Employ the Bernstein blocking method for proving convergence in distribution, conditional on $X_1, \ldots, X_n$, in probability.
- Conclude the unconditional convergence in distribution. $\square$

Sketch of proof — Fast regime

- In the parametric regime ($n\delta_n^2 = \mathcal{O}(1)$), the inverse process is not convergent any longer.

  $\Rightarrow$ Developing new strategy, not based on the switch relation.
- First, we utilize uniform consistency on compacts to prove

$$\sqrt{n}\int_{-T}^{T}|\hat{\Phi}_n(x)-\Phi_n(x)|dP_X(x) = \sqrt{n}\int_{-T}^{T}|\hat{\Phi}_n(x)-\Phi_n(x)|dP_n(x) + o_{\mathbb{P}}(1).$$

- Next, we show that we may replace Φ_n by the constant $\Phi_0(0)$ in the L^1-distance within an error of negligible order.

Sketch of proof — Fast regime

- Now, as the NPMLE is an increasing function

$$\begin{aligned}
&\int_{-T}^{T} |\hat{\Phi}_n(x) - \Phi_0(0)| dP_n(x) \\
&\quad = \sup_{s\in[-T,T]} \left\{ \int_s^T \big(\hat{\Phi}_n(x) - \Phi_0(0)\big) dP_n(x) \right. \\
&\qquad\qquad \left. - \int_{-T}^{s} \big(\hat{\Phi}_n(x) - \Phi_0(0)\big) dP_n(x) \right\} \\
&\quad = \sup_{s\in[-T,T]} \left\{ \int_{-T}^{T} \big(\hat{\Phi}_n(x) - \Phi_0(0)\big)\big(1 - 2\mathbb{1}_{\{x\le s\}}\big) dP_n(x) \right\}.
\end{aligned}$$

- Exploiting the characterization of the NPMLE as local average between two jumping points allows to approximate the expression by a supremum over a centered partial sum process.

$\square$

Conclusion

Summary — Rates of convergence

- New minimax lower bounds over restricted classes of the type

$$\mathcal{F}_\delta := \left\{ f \in \mathcal{F}_\uparrow \;\middle|\; \text{'steepness' of } f \text{ between } \delta/2 \text{ and } \delta \right\}$$

pointwise and in L^1:

$$\frac{1}{\sqrt{n}} \vee \left(\frac{\delta}{n}\right)^{1/3}.$$

- Elbow at $\delta \sim \frac{1}{\sqrt{n}}$, separating parametric and nonparametric regime.
- Matching pointwise and L^1-rates of convergence of the NPMLE in the weak-feature-impact scenario

$$\frac{1}{\sqrt{n}} \vee \left(\frac{\delta_n}{n}\right)^{1/3}.$$

Summary — Limiting distributions

Phase transition corresponding to the elbow at $\delta_n \sim \frac{1}{\sqrt{n}}$.

- Pointwise limiting distributions:
  - Slow regime: Chernoff distribution
(but faster rate of convergence)
  - Boundary case: New limit
  - Fast regime: Limit for flat functions
- Limiting distribution of the stabilized L^1-error:
  - Slow regime: Normal distribution
(but slower rate of convergence)
  - Fast regime: New limit

Outlook

- L^∞-error,
- Weak feature impact in higher dimension:
  - Generalized additive model

    $$\log\left(\frac{\mathbb{P}(Y=1|X=x)}{\mathbb{P}(Y=0|X=x)}\right)=\sum_{j=1}^{d} f_j(x_j), \quad x=(x_1,\dots,x_d)\in\mathbb{R}^d,$$

    f_j isotonic for all j,
  - Semiparametric modeling of single index structure

    $$\mathbb{P}(Y=1|X=x)=\psi\big(\langle\beta,x\rangle\big)$$

    ψ isotonic,
- Bootstrapping at a phase transition.

Thank you for your attention!