Superefficient estimation of the intensity of a stationary Poisson point process via the Stein method

Marianne Clausel
Joint work with Jean-François Coeurjolly and Jérôme Lelong

Laboratoire Jean Kuntzmann (LJK), Grenoble Alpes University
Back to the initial ideas of Charles Stein

Considered as the father of different problems related to optimal estimation, stochastic calculus, ...
Back to the initial ideas of Charles Stein

Considered as the father of different problems related to optimal estimation, stochastic calculus, ... Everything started in the following context

- Let \( X \sim \mathcal{N}(\theta, \sigma^2 I_d) \)
  where \( I_d \) is the \( d \)-dimensional identity matrix.

- Objective: estimate \( \theta \) based on a **single** (for simplicity) observation \( X \).
Back to the initial ideas of Charles Stein

Considered as the father of different problems related to optimal estimation, stochastic calculus, ... Everything started in the following context

- Let $X \sim \mathcal{N}(\theta, \sigma^2 I_d)$
  where $I_d$ is the $d$-dimensional identity matrix.

- Objective: estimate $\theta$ based on a single (for simplicity) observation $X$.

- $\hat{\theta}^{mle} = X$ minimizes $\text{MSE}(\hat{\theta}) = E\left(\|\hat{\theta} - \theta\|^2\right)$ among unbiased estimators.
Considered as the father of different problems related to optimal estimation, stochastic calculus,…

Everything started in the following context:

- Let \( X \sim \mathcal{N}(\theta, \sigma^2 I_d) \)
  where \( I_d \) is the \( d \)-dimensional identity matrix.
- Objective: estimate \( \theta \) based on a single (for simplicity) observation \( X \).

- \( \hat{\theta}^{mle} = X \) minimizes \( \text{MSE}(\hat{\theta}) = \mathbb{E}\left(\|\hat{\theta} - \theta\|^2\right) \) among unbiased estimators.
- Stein (1956)
  \[
  \hat{\theta}_S = \left((1 - b(a + X_i^2)^{-1})X_i\right)_{i=1,\ldots,d} \Rightarrow \text{MSE}(\hat{\theta}_S) \leq \text{MSE}(\hat{\theta}^{mle}) \text{ when } d \geq 3
  \]
- James-Stein (1961)
  \[
  \hat{\theta}_{JS} = X(1 - (d - 2)/\|X\|^2) \Rightarrow \text{MSE}(\hat{\theta}_{JS}) \leq \text{MSE}(\hat{\theta}^{mle}) \text{ when } d \geq 3
  \]
- Stein (1981) key-ingredients for the class: \( \hat{\theta} = X + g(X), \quad g : \mathbb{R}^d \rightarrow \mathbb{R}^d \).
MSE of $\hat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 I_d, \sigma^2 \text{ known})$

\[ \text{MSE}(\hat{\theta}) = \]
MSE of $\hat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 I_d, \sigma^2 \text{ known})$

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\|X - \theta\|^2 + \mathbb{E}\|g(X)\|^2 + 2 \sum_{i=1}^{d} \mathbb{E}((X_i - \theta_i)g_i(X))$$
MSE of $\hat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 I_d, \sigma^2 \text{ known})$)

$$MSE(\hat{\theta}) = E\|X - \theta\|^2 + E\|g(X)\|^2 + 2 \sum_{i=1}^{d} E ((X_i - \theta_i)g_i(X))$$

1. Using an IbP for Gaussian r.v. $E[Zh(Z)] = E[h'(Z)], Z \sim \mathcal{N}(0, 1)$

$$MSE(\hat{\theta}) = MSE(\hat{\theta}^{mle}) + E\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} E\partial_i g_i(X)$$
MSE of $\hat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 I_d, \sigma^2$ known)

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \|X - \theta\|^2 + \mathbb{E} \|g(X)\|^2 + 2 \sum_{i=1}^{d} \mathbb{E} ((X_i - \theta_i)g_i(X))$$

1. Using an IbP for Gaussian r.v. $\mathbb{E}[Zh(Z)] = \mathbb{E}[h'(Z)], Z \sim \mathcal{N}(0, 1)$

$$\text{MSE}(\hat{\theta}) = \text{MSE}(\hat{\theta}_{mle}) + \mathbb{E} \|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} \mathbb{E} \partial_i g_i(X)$$

2. Now choose $g = \sigma^2 \nabla \log f$. Use the well-known fact [based on product and chain–rules] that for $h : \mathbb{R} \rightarrow \mathbb{R}$,

$$(\log(h)')^2 + 2(\log h)'' =$$
MSE of $\widehat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 I_d, \sigma^2$ known)

\[
\text{MSE}(\widehat{\theta}) = E\|X - \theta\|^2 + E\|g(X)\|^2 + 2 \sum_{i=1}^{d} E((X_i - \theta_i)g_i(X))
\]

1. Using an IbP for Gaussian r.v. \(E[Zh(Z)] = E[h'(Z)], Z \sim N(0, 1)\)

\[
\text{MSE}(\widehat{\theta}) = \text{MSE}(\widehat{\theta}^{\text{mle}}) + E\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} E\partial_i g_i(X)
\]

2. Now choose \(g = \sigma^2 \nabla \log f\). Use the well-known fact [based on product and chain–rules ] that for \(h : \mathbb{R} \to \mathbb{R}\),
\[
(\log(h)')^2 + 2(\log h)'' = 4 \frac{(\sqrt{h})''}{\sqrt{h}}.
\]
Get

\[
\text{MSE}(\widehat{\theta}) = \text{MSE}(\widehat{\theta}^{\text{mle}}) + 4\sigma^2 E\left(\frac{\nabla \nabla \sqrt{f(X)}}{\sqrt{f(X)}}\right) \leq \text{MSE}(\widehat{\theta}^{\text{mle}}) \quad \text{if} \quad \nabla \nabla \sqrt{f} \leq 0.
\]
MSE of $\hat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 I_d$, $\sigma^2$ known)

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\|X - \theta\|^2 + \mathbb{E}\|g(X)\|^2 + 2 \sum_{i=1}^{d} \mathbb{E}((X_i - \theta_i)g_i(X))$$

1. Using an IbP for Gaussian r.v. $\mathbb{E}[Zh(Z)] = \mathbb{E}[h'(Z)], Z \sim N(0, 1)$

$$\text{MSE}(\hat{\theta}) = \text{MSE}(\hat{\theta}_{mle}) + \mathbb{E}\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} \mathbb{E}\partial_i g_i(X)$$

2. Now choose $g = \sigma^2 \nabla \log f$. Use the well-known fact [based on product and chain–rules] that for $h : \mathbb{R} \to \mathbb{R}$,

$$(\log(h)')^2 + 2(\log h)'' = 4 \frac{(\sqrt{h})''}{\sqrt{h}}.$$

Get

$$\text{MSE}(\hat{\theta}) = \text{MSE}(\hat{\theta}_{mle}) + 4\sigma^2 \mathbb{E}\left\langle \nabla \nabla \sqrt{f(X)} \right\rangle \leq \text{MSE}(\hat{\theta}_{mle}) \text{ if } \nabla \nabla \sqrt{f} \leq 0.$$ 

Goal: mimic both steps, derive a Stein estimator for the intensity of a Poisson point process [extension Privault-Réveillac (2009), $d = 1$]
The realization $x$, of a spatial point process defined on $S$ and observed in a bounded domain is a finite set of objects $x_i \in S$.

$$x = \{x_1, \cdots, x_n\}.$$
Observation of 97 ants categorized in two species (R package `spatstat`).

The state space is denoted by $S = \mathbb{R}^2 \times \{0, 1\}$ and is equipped with $\max(||\cdot||, d_M)$ for any distance $d_M$ on the mark space $\{0, 1\}$.

**Scientific questions:** Competition inside one species? Between the two species?
Spatial point processes: a formal definition?

- $S$: Polish state space of the point process (equipped with the $\sigma$-algebra of Borel sets $\mathcal{B}$).
- A configuration of points is denoted $x = \{x_1, \cdots, x_n, \cdots\}$. For $B \subset S: x_B = x \cap B$.
- $N_{lf}$: space of locally finite configurations, i.e.

$$\{x, n(x_B) = |x_B| < \infty, \forall B \text{ bounded} \in S\}$$

equipped with

$$N_{lf} = \sigma(\{x \in N_{lf}, n(x_B) = m\}; B \in \mathcal{B}, B \text{ bounded, } m \geq 1).$$

**Definition**

A point process $X$ defined on $S$ is a measurable application defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values on $N_{lf}$.

Measurability of $X \iff N(B) = n(X_B) = |X_B|$ is a r.v. for any bounded $B \in \mathcal{B}$. 
Poisson point processes

Definition of a Poisson point process with intensity $\rho(\cdot)$.

- $\forall m \geq 1$, $\forall$ bounded and disjoint $B_1, \ldots, B_m \subset S$, the r.v. $X_{B_1}, \ldots, X_{B_m}$ are independent.
- $N(B) \sim \mathcal{P}(\int_B \rho(u)du)$
- $\forall B \subset S$, $\forall F \in \mathcal{N}_f$

$$
\mathbb{P}(X_B \in F) = \sum_{n \geq 0} \frac{e^{\int_B \rho(u)du}}{n!} \int_B \cdots \int_B 1\left(\{(x_1, \ldots, x_n) \in F\}\right) \prod_{i=1}^{n} \rho(x_i)dx_i
$$

Notation: $X \sim \text{Poisson}(S, \rho)$. 
Poisson point processes: a few realizations on $[-1, 1]^2$

- $\rho = 200$.
- $\rho = \beta \sin(4\pi u_1 u_2)$.
- $\rho(u) = \beta e^{-u_1 - u_1^2 - 5u_3}$.
Our framework

- Case considered here $\rho(u) \equiv \theta$.
- $X$ homogeneous Poisson point process with intensity $\theta$.
- We assume observing $X$ on $W \subset \mathbb{R}^d$.
- Given $N(W) = n$, we denote $X_1, \ldots, X_n$ the $n$ points in $W$.
- The MLE estimate of the intensity $\theta$ is $\hat{\theta} = N(W)/|W|$.
- Construction of a Stein estimator of $\theta$?
$S$: space of Poisson functionals $F$ defined on $\Omega$ by

$$F = f_0 \mathbf{1}(N(W) = 0) + \sum_{n \geq 1} \mathbf{1}(N(W) = n)f_n(X_1, \ldots, X_n),$$

$f_0 \in \mathbb{R}, f_n \in L^1(W^n, \mathbb{R}^d)$ measurable symmetric functions called form functions of $F$. 
Towards a Stein estimator (1)

- MLE is defined by $\hat{\theta}^{mle} = N(W)/|W|$.
- Aim: define $\hat{\theta}$ of the form $\hat{\theta} = \hat{\theta}^{mle} + \frac{1}{|W|} \zeta$ where $\zeta = \nabla \log(F)$.
- Relation $\text{MSE}(\hat{\theta}) < \text{MSE}(\hat{\theta}^{mle})$ satisfied?

\[
\text{MSE}(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta}^{mle} + \frac{1}{|W|} \nabla \log F - \theta\right)^2\right] \\
= \text{MSE}(\hat{\theta}^{mle}) + \frac{1}{|W|^2} \left(\mathbb{E}[\nabla \log F]^2 + 2\mathbb{E}[(\nabla \log F)(N(W) - \theta|W|)]\right)
\]

$\implies$ Need to transform $2\mathbb{E}[G(N(W) - \theta|W|)]$ with $G = \nabla \log F$ using a IbP formula.
- Notion of derivative?
Malliavin derivatives (1)

Differential operator: let $\pi : W^2 \to \mathbb{R}^d$

$$D^\pi_x F = - \sum_{n \geq 1} 1(N(W) = n) \sum_{i=1}^{n} (\nabla_{x_i} f_n)(X_1, \ldots, X_n) \pi(X_i, x),$$

where

$$S' = \{ F \in S : \exists C > 0 \text{ s.t. } \forall n \geq 1, f_n \in C^1(W^n, \mathbb{R}) \text{ and }$$

$$\|f_n\|_{L^\infty(W^n, \mathbb{R})} + \sum_{i=1}^{n} \|\nabla_{x_i} f_n\|_{L^\infty(W^n, \mathbb{R}^d)} \leq C^n \}. $$

and $\nabla_{x_i} f_n$ gradient of $x_i \mapsto f_n(\ldots, x_i, \ldots)$. 

Lemma [product and chain rules]

For any $x \in W$, for all $F, G \in S'$, $g \in C_b^1(\mathbb{R})$ we have

$$D_x^\pi (FG) = (D_x^\pi F)G + F(D_x^\pi G) \quad \text{and} \quad D_x^\pi g(F) = g'(F)D_x^\pi F.$$
Malliavin derivatives (2)

Lemma [product and chain rules]

For any $x \in W$, for all $F, G \in S'$, $g \in C^1_b(\mathbb{R})$ we have

$$D^\pi_x(FG) = (D^\pi_x F)G + F(D^\pi_x G) \quad \text{and} \quad D^\pi_x g(F) = g'(F)D^\pi_x F .$$

To get an IbP type formula, we need to introduce $\text{Dom}(D^\pi)$ of $S'$ as

$$\text{Dom}(D^\pi) = \left\{ F \in S' : \forall n \geq 1 \text{ and } z_1, \ldots, z_n \in \mathbb{R}^d \right\} \frac{f_{n+1}}{z_{n+1} \in \partial W(z_1, \ldots, z_{n+1}) = f_n(z_1, \ldots, z_n), f_1|_{z \in \partial W}(z) = f_0} , \quad (1)$$

Remark: compatibility conditions important to derive a correct Stein estimator.
Integration by parts formula

**Theorem**

Let $G \in \text{Dom}(\overline{D}^\pi)$, $V : \mathbb{R}^d \to \mathbb{R}$, $V \in C^1(W, \mathbb{R}^d)$

$$E\left[\int_W D_x^\pi G \cdot V(x)dx\right] = E\left[G\left(\sum_{u \in X_W} \nabla \cdot V(u) - \theta \int_W \nabla \cdot V(u)du\right)\right]$$

where $\nabla \cdot V : W \to \mathbb{R}^d$ is defined by $\nabla \cdot V(u) = \int_W V(x)\pi(u, x)dx$. 
Integration by parts formula

**Theorem**

Let \( G \in \text{Dom}(D^\pi) \), \( V : \mathbb{R}^d \to \mathbb{R} \), \( V \in C^1(W, \mathbb{R}^d) \)

\[
\mathbb{E} \left[ \int_W D^\pi_x G \cdot V(x)dx \right] = \mathbb{E} \left[ G \left( \sum_{u \in X_W} \nabla \cdot \mathcal{V}(u) - \theta \int_W \nabla \cdot \mathcal{V}(u)du \right) \right]
\]

where \( \mathcal{V} : W \to \mathbb{R}^d \) is defined by \( \mathcal{V}(u) = \int_W V(x)\pi(u, x)dx \).

**Main application:** let \( \pi(u, x) = u^\top V(x) \), we can find some \( V \) (omit details) such that \( \mathcal{V}(u) = u/d \) and \( \nabla \cdot \mathcal{V}(u) = 1 \). Then

\[
\nabla G = \nabla^{\pi, V} G = -\frac{1}{d} \sum_{n \geq 1} \mathbf{1}(N(W) = n) \sum_{i=1}^{n} \nabla x_i g_n(X_1, \ldots, X_n) \cdot X_i
\]

\[\Rightarrow \quad \mathbb{E}[\nabla G] = \mathbb{E} \left[ G(N(W) - \theta|W|) \right].\]
Towards a Stein estimator (2) : end of the proof

**Theorem**

Let $\hat{\theta} = \hat{\theta}^{mle} + \frac{1}{|W|}\zeta$ where $\zeta = \nabla \log(F)$ is such that $\zeta \in \text{Dom}(\bar{D}^\pi)$ then

$$\text{MSE}(\hat{\theta}) = \text{MSE}(\hat{\theta}^{mle}) + \frac{4}{|W|^2} \mathbb{E} \left( \frac{\nabla \nabla \sqrt{F}}{\sqrt{F}} \right).$$

**Proof:**

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left[ \left( \hat{\theta}^{mle} + \frac{1}{|W|} \nabla \log F - \theta \right)^2 \right]$$

$$= \text{MSE}(\hat{\theta}^{mle}) + \frac{1}{|W|^2} \left( \mathbb{E}[\nabla \log F]^2 + 2\mathbb{E}[\nabla \log F](N(W) - \theta|W|) \right)$$

$$= \text{MSE}(\hat{\theta}^{mle}) + \frac{1}{|W|^2} \left( \mathbb{E}[\nabla \log F]^2 + 2\mathbb{E}[\nabla \nabla \log F] \right)$$

$$= \ldots$$
Theorem

The operator $D^\pi$ is closable and admits a closable adjoint $\delta^\pi$ from $L^2(\Omega, L^2(W, \mathbb{R}^d))$ into $L^2(\Omega)$ and the following duality relation holds:

$$E \left[ \int_W D^\pi_x F \cdot V(x)dx \right] = E \left[ F\delta^\pi(V) \right], \quad \forall F \in \text{Dom}(\overline{D}^\pi), \forall V \in \text{Dom}(\overline{\delta}^\pi).$$

(2)

If $V \in L^\infty(W, \mathbb{R})$ then $\mathcal{V} : W \to \mathbb{R}^d$, defined by

$$\mathcal{V}(u) = \int_W V(x)\pi(u, x)dx,$$

is an element of $C^1(W, \mathbb{R}^d)$ and we have the following explicit expression for $\delta^\pi$:

$$\delta^\pi(V) = \sum_{u \in X^W} \nabla \cdot \mathcal{V}(u) - \theta \int_W \nabla \cdot \mathcal{V}(u) du.$$  

(3)
Step 1 of the proof: Proof of the duality relation in the case $F \in \text{Dom}(D^\pi)$ and $V \in L^\infty(W, \mathbb{R}^d)$

- One has

$$E\left[\int_W D^\pi_x F \cdot V(x) \, dx\right] = -e^{-\theta|W|} \sum_{n \geq 1} \frac{\theta^n}{n!} \sum_{i=1}^{n} \int_{W^{n+1}} \nabla_{z_i} f_n(z_1, \ldots, z_n) \cdot V(x) \, \pi(z_i, x) \, dx \, dz_1 \ldots dz_n$$

$$= -e^{-\theta|W|} \sum_{n \geq 1} \frac{\theta^n}{n!} \sum_{i=1}^{n} \int_{W^{n-1}} g(z^{-i}) \, dz_1 \ldots dz_{i-1} \, dz_{i+1} \ldots dz_n$$

with $g(z^{-i}) = \int_W \nabla_{z_i} f_n(z_1, \ldots, z_n) \cdot V(z_i) \, dz_i$.

- To conclude, use the trace theorem and the symmetry of functions $f_n$. 
Step 2 of the proof: Extension of the IbP formula on $\in \text{Dom}(D^\pi) \otimes L^\infty(W, \mathbb{R}^d)$ [dense subset of $L^2(\Omega, L^2(\mathbb{R}^d, \mathbb{R}))$]

- Define $\delta^\pi$ on $S' \otimes L^\infty(W, \mathbb{R}^d)$ by

  $$\delta^\pi(GV) = G\delta^\pi(V) - \int_{\mathbb{R}^d} D_x^\pi G \cdot V(x)dx.$$

- Use the product rule and extend the IbP on $S' \otimes L^\infty(W, \mathbb{R}^d)$

  $$E\left[ G \int_W D_x^\pi F \cdot V(x)dx \right] = E\left[ \int_W D_x^\pi (FG) \cdot V(x)dx - F \int_W D_x^\pi G \cdot V(x)dx \right]$$

  $$= E\left[ FG\delta^\pi(V) - F \int_W D_x^\pi G \cdot V(x)dx \right]$$

  $$= E\left[ F\delta^\pi(GV) \right].$$
Observe that $\text{Dom}(D^\pi) \otimes L^\infty(W, \mathbb{R}^d)$ is dense in $L^2(\Omega, L^2(W, \mathbb{R}^d))$.

Prove that the operator $D^\pi$ is closable to extend the IbP formula to $L^2(\Omega, L^2(W, \mathbb{R}^d))$. 
Consequences

- Let $\pi(u, x) = u^\top V(x)$.
- Let $V$ (omit details) such that $V(u) = u/d$ and $\nabla \cdot V(u) = 1$.
- Then

\[
\nabla F = \nabla^{\pi, V} F = -\frac{1}{d} \sum_{n \geq 1} 1(N(W) = n) \sum_{i=1}^{n} \nabla_{x_i} f_n(X_1, \ldots, X_n) \cdot X_i
\]

- Hence

\[
E[\nabla F] = E \left[ F(N(W) - \theta|W|) \right].
\]

which is the **IbP formula** needed in the sequel.
Non-uniqueness of the integration by parts formula

- Natural and easier to define a **isotropic** Stein estimator. With

\[
\nabla \log F = -\frac{1}{d} \sum_{n \geq 1} \mathbf{1}(N(W) = n) \sum_{i=1}^{n} \nabla_{x_i}(\log f_n)(X_1, \ldots, X_n) \cdot X_i
\]

\( \log F \) is isotropic \( \Rightarrow \) \( \nabla \log F \) is isotropic (and so will be \( \hat{\theta} \)).

- Other possible choices to get \( \text{div} \mathbf{V}(y) = 1 \):

\( V(x) = (d|W|)^{-1/2} \mathbf{1}(x \in W)1^\top, \pi(y, x) = y^\top V(x) \). New gradient operator :

\[
\nabla \log F = -\sum_{n \geq 1} \mathbf{1}(N(W) = n) \sum_{i=1}^{n} (\text{div}_{x_i} \log f_n)(X_1, \ldots, X_n) \times X_i
\]

- Formula \( \mathbb{E}[\nabla \log F] = \mathbb{E} [\log F(N(W) - \theta|W|)] \) **still** holds. But....

1. non isotropic.
2. can induce some discontinuity problems when computing \( \nabla \log F \) and \( \nabla\nabla \log F \ldots \)
Example in the $d$-dimensional euclidean ball $W = B_d(0, 1)$

- For $1 \leq k \leq n$, $x_{(k),n}$ $k$th closest (wrt $\| \cdot \|$) point of $\{x_1, \ldots, x_n\}$ to zero.
- $X_k$ $k$th closest point to 0 of the PPP $X$ (defined on $\mathbb{R}^d$)
- We define $\varphi(t) = e^{\gamma(1-t)^\kappa} 1(t \leq 1)$, $\gamma \in \mathbb{R}$, $\kappa > 2$.
- Set $F_k = 1(N(W) < k) + \sum_{n \geq k} 1(N(W) = n) \varphi(\|X_{(k),n}\|^2)^2$
- Define

$$\hat{\theta}_k = \hat{\theta}_{mle} + \nabla \log F_k/|W|$$

$$Gain(\hat{\theta}_k) = 1 - \text{MSE}(\hat{\theta}_k)/\text{MSE}(\hat{\theta}_{mle})$$
Explicit expression of the theoretical gain

Theorem

\[ \zeta_k = \nabla \log(F_k) \in \text{Dom}(\overline{D}^\pi) \quad [\varphi > 0, \varphi'(1) = 0] \quad \text{and} \]

\[ \hat{\theta}_k = \hat{\theta}_{mle} - \frac{4d}{|W|} \frac{\varphi'(|\|X(k)||^2)}{\varphi(|\|X(k)||^2)} = \hat{\theta}_{mle} - \frac{4d}{|W|} \left\{ \gamma \kappa \left( 1 - \|X(k)\|^2 \right)^{\kappa-1} \right\} \]

Gain(\hat{\theta}_k) = E[ G(|\|X(k)||^2) ] \quad \text{where} \quad G(t) = -\frac{16}{d^2 \theta |W|} \frac{t (\varphi'(t) + t \varphi''(t))}{\varphi(t)} \]

Need to differentiate properly non–differentiable Poisson functionals as \( x \mapsto \|X(k)\| \).
Proof of the Theorem

Relies on:

Lemma

Let $H^1([0, 1], \mathbb{R})$ be the classical Sobolev space, $\Psi \in H^1([0, 1], \mathbb{R})$. Then,

$$G_k = \sum_{n \geq k} 1(N(W) = n)\Psi(||X_{(k),n}||^2) + \Psi(1)1(N(W) < k) \in \text{Dom}(\overline{D})$$

(4)

and

$$\nabla G_k = -\frac{2}{d} \sum_{n \geq k} 1(N(W) = n) ||X_{(k),n}||^2 \Psi'(||X_{(k),n}||^2).$$

(5)
Numerical experiments: \( \varphi(t) = e^{\gamma(1-t)^{\kappa}}; \kappa = 3; \gamma = -3 \)

- \( m = 50000 \) replications of \( PPP(\mathcal{B}(0, 1), \theta), d = 2. \)

- Empirical vs Monte-Carlo approximations of theoretical gains, for different parameters \( k, \kappa, \gamma. \)

- General comments:
  1. The IbP formula is empirically checked.
  2. The parameters \( k, \kappa, \gamma \) and \( \theta \) are strongly connected. A bad choice can lead to **negative** gains \( [\varphi'(t) + t\varphi''(t) \text{ may be negative for some values of } t] \).
• $m = 50000$ replications of $PPP_\theta(B(0, 1), \theta), d = 2$.

• Monte-Carlo approximations of theoretical gains for different values of $k$. The parameters $\kappa$ and $\gamma$ optimize $\text{Gain}(\hat{\theta}_k)$ for each value of $\theta$.

• General comments:
  1. For any $k$, if we optimize in terms of $\kappa$ and $\gamma$, the gain becomes always positive.
  2. Still, if we want interesting values of gains, $k$ needs to be optimized.
- Simulation based on $m = 50000$ replications.
- For each value of $\theta, d$

$$(k^*, \gamma^*, \kappa^*) = \arg\max_{(k, \gamma, \kappa)} \text{Gain}(\hat{\theta}_k) = \arg\max_{(k, \gamma, \kappa)} \mathbb{E}[\mathcal{G}(\|X(k)\|^2)].$$

<table>
<thead>
<tr>
<th>$\theta = 5$, $d = 1$</th>
<th>MLE</th>
<th>STEIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = 2$</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>5</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\theta = 10$, $d = 1$</th>
<th>MLE</th>
<th>STEIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = 2$</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>10</td>
<td>44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\theta = 20$, $d = 1$</th>
<th>MLE</th>
<th>STEIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = 2$</td>
<td>20</td>
<td>66</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>20</td>
<td>84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\theta = 40$, $d = 1$</th>
<th>MLE</th>
<th>STEIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = 2$</td>
<td>40</td>
<td>125</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>40</td>
<td>169</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\text{Gain (%)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>43.0</td>
</tr>
<tr>
<td>45.6</td>
</tr>
<tr>
<td>46.1</td>
</tr>
<tr>
<td>45.8</td>
</tr>
<tr>
<td>46.0</td>
</tr>
<tr>
<td>46.3</td>
</tr>
<tr>
<td>46.4</td>
</tr>
<tr>
<td>46.5</td>
</tr>
<tr>
<td>47.5</td>
</tr>
<tr>
<td>47.2</td>
</tr>
<tr>
<td>46.9</td>
</tr>
<tr>
<td>48.3</td>
</tr>
</tbody>
</table>
Data-driven estimator: replace $\theta$ by $\hat{\theta}_{mle}$ in the optimization

- Simulation based on $m = 5000$ replications.
- For each value of $\theta$, $d$, let $\Theta(\theta, \rho) = [\theta - \rho \sqrt{\theta/|W|}, \theta + \rho \sqrt{\theta/|W|}]$. Then, we suggest define $\kappa^*$, $\gamma^*$ as the maximum of

\[
\int_{\Theta(\hat{\theta}_{MLE}, \rho)} \text{Gain}(\hat{\theta}_k) d\theta = \frac{16}{d^2|W|} \mathbb{E} \int_{\Theta(\hat{\theta}_{MLE}, \rho)} \frac{G(Y_{(k)})}{\theta} d\theta. \quad (6)
\]

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = 1$</td>
<td>$\rho = 0$</td>
</tr>
<tr>
<td>$\theta = 5$</td>
<td>48.8</td>
</tr>
<tr>
<td>$\theta = 10$</td>
<td>38.6</td>
</tr>
<tr>
<td>$\theta = 20$</td>
<td>39.4</td>
</tr>
<tr>
<td>$\theta = 40$</td>
<td>40.3</td>
</tr>
<tr>
<td>$d = 2$</td>
<td>36.2</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>31.6</td>
</tr>
<tr>
<td>$\theta = 20$</td>
<td>37.3</td>
</tr>
<tr>
<td>$\theta = 40$</td>
<td>27.3</td>
</tr>
<tr>
<td>$\theta = 40$</td>
<td>20.8</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>22.3</td>
</tr>
<tr>
<td>$d = 2$</td>
<td>16.3</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>12.7</td>
</tr>
</tbody>
</table>
A few more comments

- Even if the results are done under the Poisson assumption, if the simulated model
  - is clustered (e.g. Thomas, LGCP) the empirical gains (compared to $N(W)/|W|$) are significant.
  - is regular the empirical gains seems to be close to zero (not really worse than $N(W)/|W|$)

Perspectives

Deriving a general IbP formula for inhomogeneous Poisson point processes or Cox point processes seems reasonable.


$\varphi(t) = e^{\gamma(1-r)^{\kappa}}; \kappa = 3; \gamma = -3$

$\Rightarrow \quad G(t)$ is not positive everywhere but when $t$ is large (i.e. when $\|X(k)\|$ is large, i.e. when $k$ is large), then $G(\cdot)$ is positive and can reach high values.
Comparison with Privault-Réveillac’s estimator when $d = 1$

- Assume $X$ is observed on $\tilde{W} = [0, 2]$.
- Let $X_1$ be the closest point of $X$ to 0, then $\hat{\theta}_{pr}$ is defined for some $\kappa > 0$ by

$$
\hat{\theta}_{pr} = \hat{\theta}_{mle} + \frac{2}{\kappa} \mathbf{1}(N(\tilde{W}) = 0) + \frac{2X_1}{2(1 + \kappa) - X_1} \mathbf{1}(0 < X_1 \leq 2).
$$

Note that $X_1 \sim E(\theta)$.

- The gain writes

$$
\text{Gain}(\hat{\theta}_{pr}) = \frac{2}{\theta \kappa^2} \exp(-2\theta) - \frac{2}{\theta} \mathbb{E}\left(\frac{X_1}{2(1 + \kappa) - X_1} \mathbf{1}(X_1 \leq 2)\right).
$$

Gain optimized in $\kappa$ in terms of $\theta$. 

![Graph showing gain optimized in $\kappa$ in terms of $\theta$.](image)