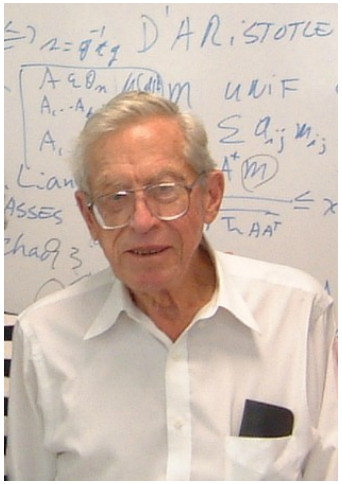
Superefficient estimation of the intensity of a stationary Poisson point process via the Stein method

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

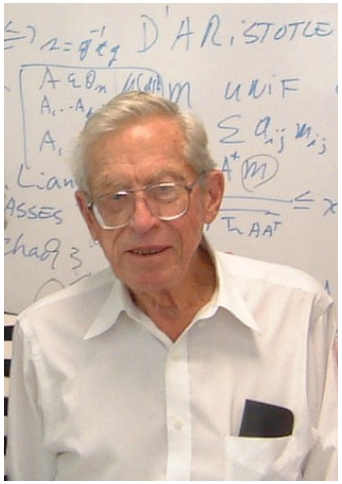
Laboratoire Jean Kuntzmann (LJK), Grenoble Alpes University

Back to the initial ideas of Charles Stein



Considered as the father of different problems related to optimal estimation, stochastic calculus,. . .

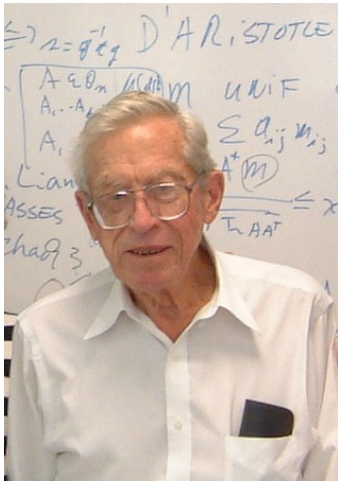
Back to the initial ideas of Charles Stein



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

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

Back to the initial ideas of Charles Stein

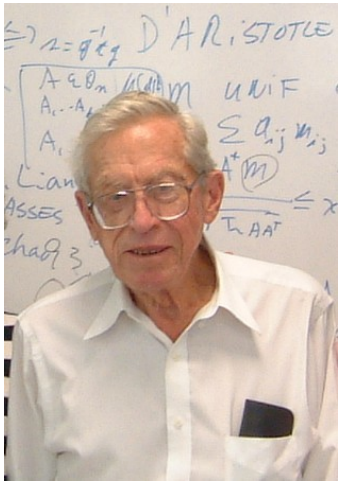


Considered as the father of different problems related to optimal estimation, stochastic calculus,...
Everything started in the following context

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

- $\widehat{\theta}^{mle} = X$ minimizes $\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}\left(\|\widehat{\theta} - \theta\|^2\right)$ among unbiased estimators.

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

- $\widehat{\theta}^{mle} = X$ minimizes $\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}\left(\|\widehat{\theta} - \theta\|^2\right)$ among unbiased estimators.
- Stein (1956)

  $$\widehat{\theta}_S = \left((1 - b(a + X_i^2)^{-1})X_i\right)_{i=1,\dots,d} \Rightarrow \mathrm{MSE}(\widehat{\theta}_S) \leq \mathrm{MSE}(\widehat{\theta}^{mle}) \text{ when } d \geq 3$$

- James-Stein (1961)

  $$\widehat{\theta}_{JS} = X(1 - (d-2)/\|X\|^2) \Rightarrow \mathrm{MSE}(\widehat{\theta}_{JS}) \leq \mathrm{MSE}(\widehat{\theta}^{mle}) \text{ when } d \geq 3$$

- Stein (1981) key-ingredients for the class: $\widehat{\theta} = X + g(X), \quad g : \mathbb{R}^d \to \mathbb{R}^d$.

MSE of $\widehat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_d$, σ^2 known)

$\mathrm{MSE}(\widehat{\theta}) =$

MSE of $\widehat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_d$, σ^2 known)

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

MSE of $\widehat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_d$, σ^2 known)

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

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

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{MSE}(\widehat{\theta}^{mle}) + \mathrm{E}\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} \mathrm{E}\nabla g_i(X)$$

MSE of $\widehat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_d$, σ^2 known)

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

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

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{MSE}(\widehat{\theta}^{mle}) + \mathrm{E}\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} \mathrm{E}\nabla g_i(X)$$

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

MSE of $\widehat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_d$, σ^2 known)

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

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

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{MSE}(\widehat{\theta}^{mle}) + \mathrm{E}\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} \mathrm{E}\nabla g_i(X)$$

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

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{MSE}(\widehat{\theta}^{mle}) + 4\sigma^2 \mathrm{E}\left(\frac{\nabla\nabla\sqrt{f(X)}}{\sqrt{f(X)}}\right) \leq \mathrm{MSE}(\widehat{\theta}^{mle}) \quad \text{if} \quad \nabla\nabla\sqrt{f} \leq 0.$$

MSE of $\widehat{\theta} = X + g(X)$ ($X \sim \mathcal{N}(\theta, \sigma^2 \mathbf{I}_d$, σ^2 known)

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

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

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{MSE}(\widehat{\theta}^{mle}) + \mathrm{E}\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^{d} \mathrm{E}\nabla g_i(X)$$

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

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{MSE}(\widehat{\theta}^{mle}) + 4\sigma^2 \mathrm{E}\left(\frac{\nabla\nabla\sqrt{f(X)}}{\sqrt{f(X)}}\right) \leq \mathrm{MSE}(\widehat{\theta}^{mle}) \quad \text{if} \quad \nabla\nabla\sqrt{f} \leq 0.$$

Goal: mimic both steps, derive a Stein estimator for the intensity of a Poisson point process [extension Privault-Réveillac (2009), $d = 1$]

Point processes

- S : Polish state space of the point process (equipped with the σ-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
$\mathcal{N}_{lf} = \sigma(\{x \in N_{lf},\ n(x_B) = m\}; B \in \mathcal{B}, B \text{ bounded}, m \geq 1)$.

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 $X \Longleftrightarrow N(B) = |X_B|$ 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 $B_1, \cdots, B_m \subset S$, the r.v. $X_{B_1}, \cdots, X_{B_m}$ are independent.
- $N(B) \sim \mathcal{P}(\int_B \rho(u)du)$
- $\forall B \subset S, \forall F \in N_{lf}$

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

Notation : $X \sim \text{Poisson}(S, \rho)$.

Our framework

- Case considered here $\rho(u) \equiv \theta$.
- X homogeneous Poisson point process with intensity θ .
- 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 θ is $\widehat{\theta} = N(W)/|W|$.
- Construction of a Stein estimator of θ?

Poisson functionals

$\mathcal{S}$: space of Poisson functionals F defined on Ω by

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

$f_0 \in \mathbb{R}$, $f_n : W^n \to \mathbb{R}^d$ measurable symmetric functions called form functions of F.

Towards a Stein estimator (1)

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

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

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

$\Longrightarrow$ Need to transform $2\mathrm{E}[G(N(W) - \theta|W|)]$ with $G = \nabla \log F$ using a IbP formula.

- Notion of derivative ?

Malliavin derivatives (1)

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

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

where $\mathcal{S}' = \{F \in \mathcal{S} :$ 'the f_n are cont. diff. in any variable x_i'$\}$ and where $\nabla_{x_i} f_n$ gradient of $x_i \mapsto f_n(\ldots, x_i, \ldots)$.

Malliavin derivatives (2)

Lemma [product and chain rules]

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

$$D_x^\pi(FG) = (D_x^\pi F)G + F(D_x^\pi G) \qquad \text{and} \qquad D_x^\pi g(F) = g'(F) D_x^\pi F \ .$$

Malliavin derivatives (2)

Lemma [product and chain rules]

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

$$D_x^\pi(FG) = (D_x^\pi F)G + F(D_x^\pi G) \qquad \text{and} \qquad D_x^\pi g(F) = g'(F) D_x^\pi F \ .$$

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

$$\mathrm{Dom}(D^\pi) = \Big\{ F \in \mathcal{S}' : \forall n \geq 1 \text{ and } z_1, \ldots, z_n \in \mathbb{R}^d$$

$$\mathbf{f_{n+1}}\Big|_{z_{n+1} \in \partial W}(z_1, \ldots, z_{n+1}) = \mathbf{f_n}(z_1, \ldots, z_n),\ \mathbf{f_1}\Big|_{\mathbf{z} \in \partial \mathbf{W}}(z) = \mathbf{0} \Big\} \ , \qquad (1)$$

<u>Remark</u>: **compatibility conditions** important to derive a correct Stein estimator.

Integration by parts formula

Theorem

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

$$\underbrace{\mathrm{E}\left[\int_W D_x^{\pi} G \cdot V(x)\mathrm{d}x\right]}_{:=\nabla^{\pi,V}F} = \mathrm{E}\left[F\left(\sum_{u\in\mathbf{X}_W} \nabla\cdot\mathcal{V}(u) - \theta\int_W \nabla\cdot\mathcal{V}(u)\mathrm{d}u\right)\right]$$

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

Integration by parts formula

Theorem

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

$$\underbrace{\mathrm{E}\left[\int_W D_x^{\pi} G \cdot V(x)\mathrm{d}x\right]}_{:=\nabla^{\pi,V}F} = \mathrm{E}\left[F\left(\sum_{u\in\mathbf{X}_W} \nabla\cdot\mathcal{V}(u) - \theta\int_W \nabla\cdot\mathcal{V}(u)\mathrm{d}u\right)\right]$$

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

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

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

$$\boxed{\Rightarrow \qquad \mathrm{E}[\nabla G] = \mathrm{E}\left[G(N(W)-\theta|W|)\right].}$$

Towards a Stein estimator (2) : end of the proof

Theorem

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

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

<u>Proof</u>:

$$\begin{aligned}
\mathrm{MSE}(\widehat{\theta}) &= \mathrm{E}\left[\left(\widehat{\theta}^{mle} + \frac{1}{|W|}\nabla \log F - \theta\right)^2\right] \\
&= \mathrm{MSE}(\widehat{\theta}^{mle}) + \frac{1}{|W|^2}\left(\mathrm{E}[(\nabla \log F)^2] + 2\mathrm{E}[(\nabla \log F)(N(W) - \theta|W|)]\right) \\
&= \mathrm{MSE}(\widehat{\theta}^{mle}) + \frac{1}{|W|^2}\left(\mathrm{E}[(\nabla \log F)^2] + 2\mathrm{E}[\nabla\nabla \log F]\right) \\
&= \ldots
\end{aligned}$$

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 $\widehat{\theta}$).

- Other possible choices to get $\operatorname{div} \mathcal{V}(y) = 1$: $V(x) = (d|W|)^{-1/2} \mathbf{1}(x \in W) 1^\top$, $\pi(y, x) = y^\top V(x)$. New gradient operator :

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

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

1. non isotropic.
2. can induce some discontinuity problems when computing $\nabla \log F$ and $\nabla\nabla \log F$ ...

Example in the d-dimensional euclidean ball $W = B_d(0,1)$

- For $1 \leq k \leq n$, $x_{(k),n}$ kth closest (wrt $\|\cdot\|$) point of $\{x_1, \ldots, x_n\}$ to zero.
- X_k kth closest point to 0 of the PPP X (defined on $\mathbb{R}^d$)
- We define

$$\varphi(t) = e^{\gamma(1-t)^\kappa}\, \mathbf{1}(t \leq 1), \qquad \gamma \in \mathbb{R}, \quad \kappa > 2$$

$$F_k = \mathbf{1}(N(W) < k) + \sum_{n \geq k} \mathbf{1}(N(W) = n)\, \varphi(\|X_{(k),n}\|^2)^2.$$

$$Gain(\widehat{\theta}_k) = 1 - \mathrm{MSE}(\widehat{\theta}_k)/\mathrm{MSE}(\widehat{\theta}^{mle})$$

Example in the d-dimensional euclidean ball $W = B_d(0,1)$

- For $1 \le k \le n$, $x_{(k),n}$ kth closest (wrt $\|\cdot\|$) point of $\{x_1, \dots, x_n\}$ to zero.
- X_k kth closest point to 0 of the PPP X (defined on $\mathbb{R}^d$)
- We define

$$\varphi(t) = e^{\gamma(1-t)^\kappa}\, \mathbf{1}(t \le 1), \qquad \gamma \in \mathbb{R}, \quad \kappa > 2$$

$$F_k = \mathbf{1}(N(W) < k) + \sum_{n \ge k} \mathbf{1}(N(W) = n)\, \varphi(\|X_{(k),n}\|^2)^2.$$

$$Gain(\widehat{\theta}_k) = 1 - \mathrm{MSE}(\widehat{\theta}_k)/\mathrm{MSE}(\widehat{\theta}^{mle})$$

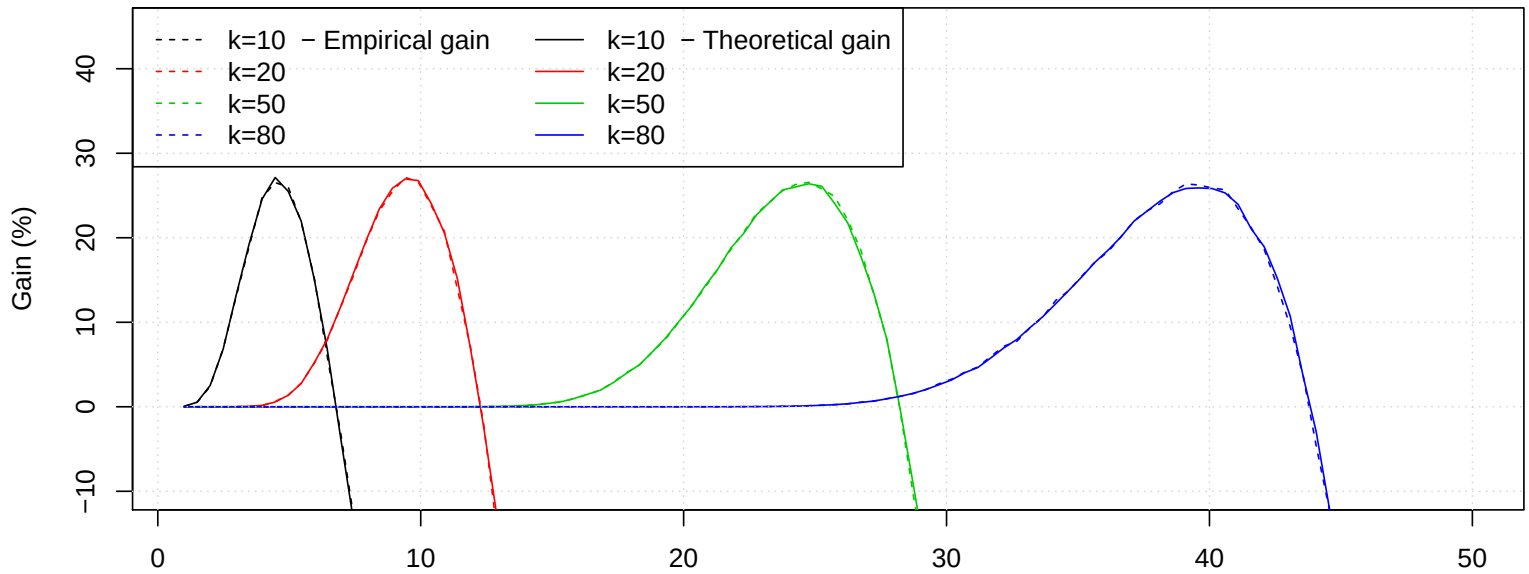
Theorem

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

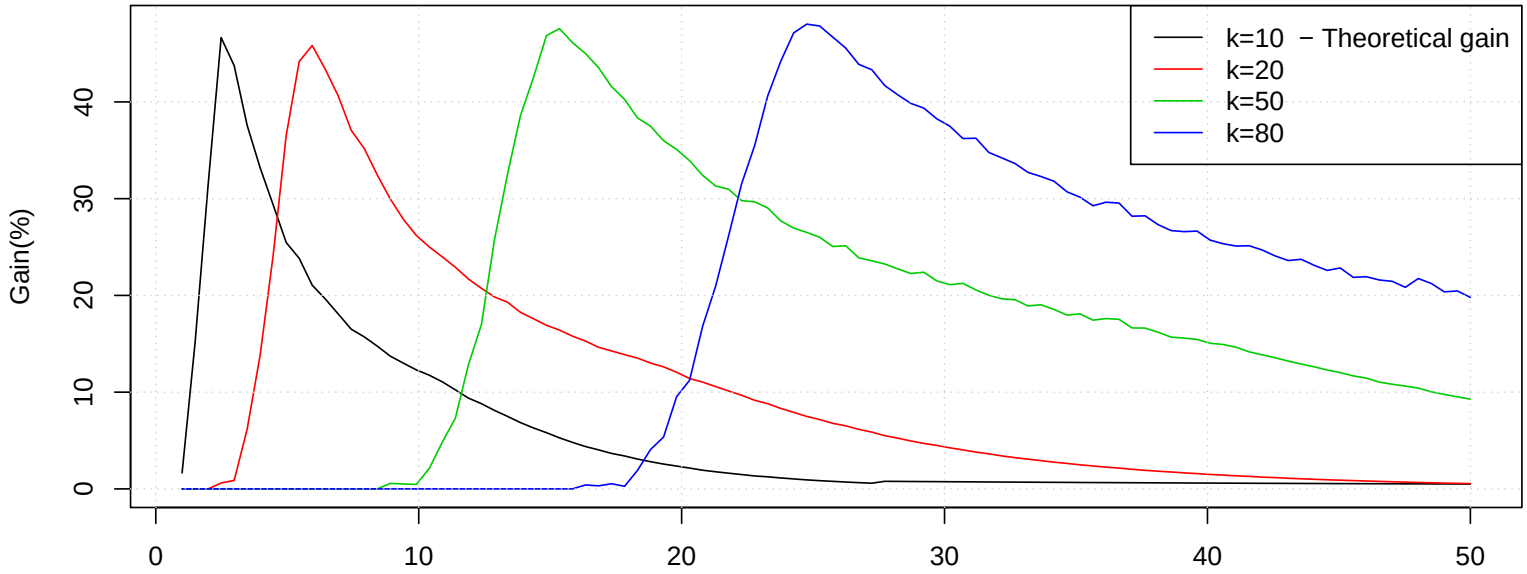
$$\widehat{\theta}_k = \widehat{\theta}^{mle} - \frac{4d}{|W|} \frac{\varphi'(\|X_{(k)}\|^2)}{\varphi(\|X_{(k)}\|^2)} = \widehat{\theta}^{mle} - \frac{4d}{|W|} \left\{ \gamma\kappa \left(1 - \|X_{(k)}\|^2\right)^{\kappa-1} \right\}$$

$$Gain(\widehat{\theta}_k) = \mathrm{E}[\, \mathcal{G}(\|X_{(k)}\|^2)\,] \quad \text{where} \quad \mathcal{G}(t) = -\frac{16}{d^2\theta|W|} \frac{t\,(\varphi'(t) + t\varphi''(t))}{\varphi(t)}$$

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



- $m = 50000$ replications of $PPP_\theta(B(0,1), \theta)$, $d = 2$.
- Empirical and Monte-Carlo approximations of theoretical gains, for different parameters k, κ, γ.
- <u>General comments</u>:
  1. The IbP formula is empirically checked.
  2. The parameters k, κ, γ and θ are strongly connected. A bad choice can lead to **negative** gains [$\varphi'(t) + t\varphi''(t)$ may be negative for some values of t].



- $m = 50000$ replications of $PPP(B_\theta(0,1), \theta)$, $d = 2$.
- Monte-Carlo approximations of theoretical gains for different values of k. The parameters κ and γ optimize $\text{Gain}(\widehat{\theta_k})$ for each value of θ.
- <u>General comments</u>:
  1. For any k, if we optimize in terms of κ and γ, 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 θ, d

$$(k^\star, \gamma^\star, \kappa^\star) = \text{argmax}_{(k,\gamma,\kappa)} \text{Gain}(\widehat{\theta}_k) = \text{argmax}_{(k,\gamma,\kappa)} \text{E}[\mathcal{G}(\|X_{(k)}\|^2)].$$

	MLE			STEIN				Gain (%)
	mean	sd	mse	$k^\star$	mean	sd	mse	
$\theta = 5, d = 1$	5	1.6	2.52	11	4.4	1.0	1.44	43.0
$d = 2$	5	1.3	1.58	18	4.6	0.8	0.86	45.6
$d = 3$	5	1.1	1.19	22	4.6	0.7	0.64	46.1
$\theta = 10, d = 1$	10	2.2	5.03	22	9.2	1.4	2.73	45.8
$d = 2$	10	1.8	3.18	34	9.4	1.2	1.72	46.0
$d = 3$	10	1.5	2.37	44	9.5	1.0	1.27	46.3
$\theta = 20, d = 1$	20	3.1	9.91	42	18.8	2.0	5.31	46.4
$d = 2$	20	2.5	6.38	66	19.1	1.6	3.41	46.5
$d = 3$	20	2.2	4.72	84	19.1	1.3	2.47	47.5
$\theta = 40, d = 1$	40	4.5	20.09	84	38.5	2.9	10.61	47.2
$d = 2$	40	3.6	12.79	125	38.6	2.2	6.78	46.9
$d = 3$	40	3.1	9.58	169	38.8	1.9	4.95	48.3

Data-driven estimator: replace θ by $\widehat{\theta}_{mle}$ in the optimization

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

$$\int_{\Theta(\widehat{\theta}_{MLE},\rho)} \mathrm{Gain}(\widehat{\theta}_k)\mathrm{d}\theta = \frac{16}{d^2|W|}\mathrm{E}\int_{\Theta(\widehat{\theta}_{MLE},\rho)} \frac{\mathcal{G}(Y_{(k)})}{\theta}\mathrm{d}\theta. \tag{2}$$

	Gain (%)			
	$\rho = 0$	$\rho = 1$	$\rho = 1.6449$	$\rho = 1.96$
$\theta = 5, d = 1$	48.8	47.9	36.4	30.1
$d = 2$	38.6	42.4	37.1	31.4
$d = 3$	39.4	42.6	37.0	31.7
$\theta = 10, d = 1$	40.3	43.8	36.7	30.1
$d = 2$	36.2	38.8	33.7	27.9
$d = 3$	31.6	36.6	32.0	28.3
$\theta = 20, d = 1$	37.3	38.6	34.5	28.0
$d = 2$	27.3	33.1	31.0	26.5
$d = 3$	20.8	28.6	28.1	23.8
$\theta = 40, d = 1$	22.3	30.8	29.2	23.9
$d = 2$	16.3	24.0	28.2	24.4
$d = 3$	12.7	19.0	24.5	22.0

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

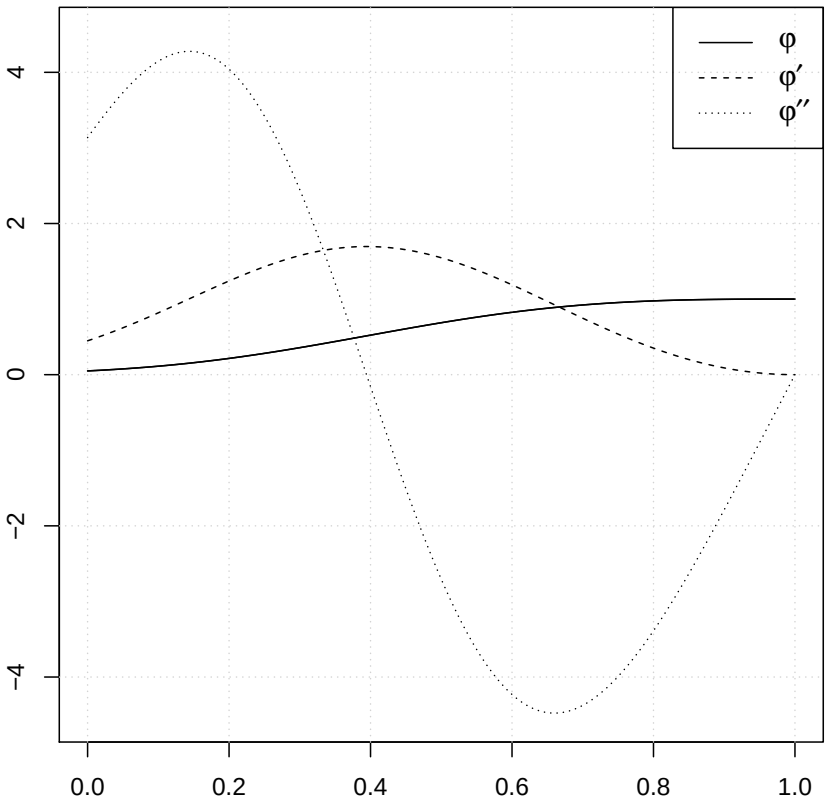
1. Deriving a general IbP formula for inhomogeneous Poisson point processes or Cox point processes seems reasonable.
2. Exploit the IbP for other statistical methodologies.

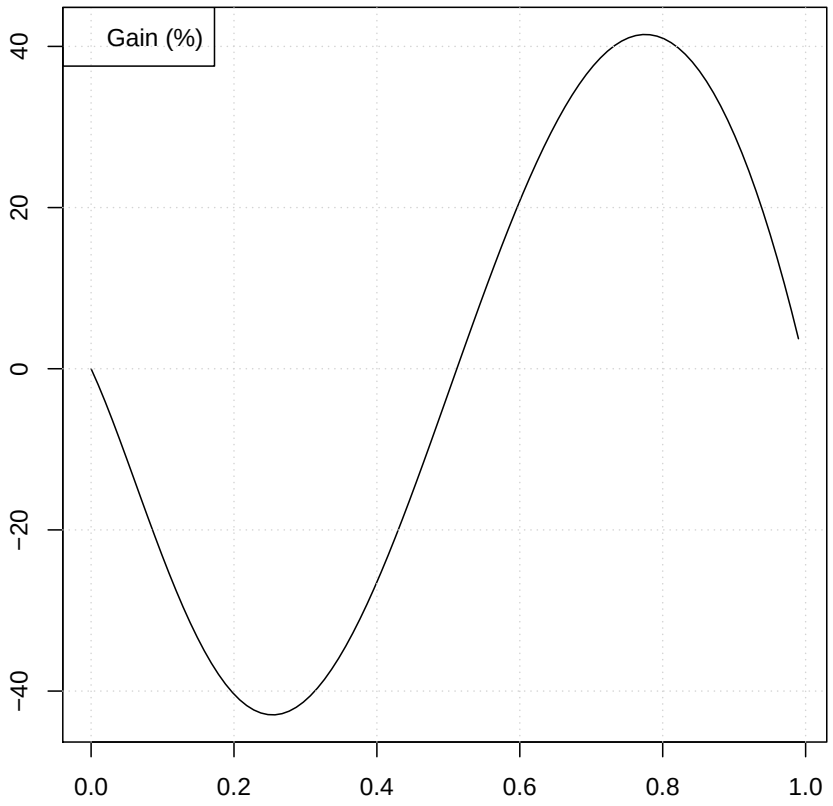
References

- W. James and C. Stein. *Estimation with quadratic loss.* (1961).
- C. Stein. *Estimation of the mean of a multivariate normal distribution.* (1981).
- N. Privault and A. Réveillac. *Stein estimation of Poisson process intensities.* (2009).
- M. Clausel, J.F. Coeurjolly, J. Lelong. *Stein estimation of the intensity of a spatial homogeneous Poisson point process.* (2014) `http://arxiv-web3.library.cornell.edu/abs/1407.4372`

Plots of $\varphi, \varphi', \varphi'', \mathcal{G}$

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





$\Rightarrow$ $\mathcal{G}(t)$ is not positive everywhere but when t is large (i.e. when $\|X_{(k)}\|$ is large, i.e. when k is large), then $\mathcal{G}(\cdot)$ is positive and can reach high values.

Comparison with Privault-Réveillac's estimator when $d = 1$

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

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

  Note that $X_1 \sim \mathcal{E}(\theta)$.
- The gain writes

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

Gain optimized in κ in terms of θ.