# CLT for the Integrated Square Error of Product Density Estimators

Stella David
Universität Augsburg

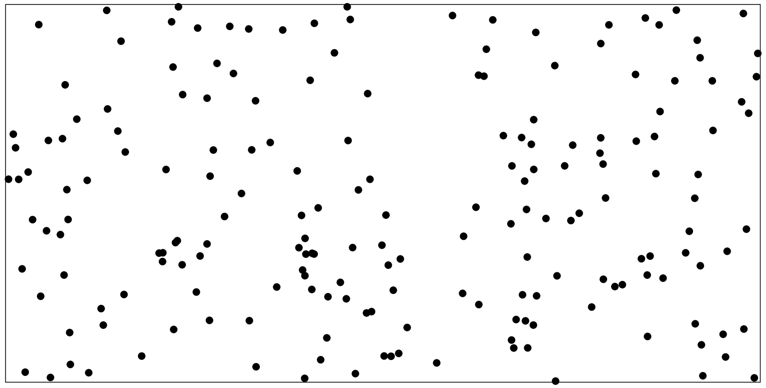
Joint work with Lothar Heinrich

March 27, 2006

# Outline

- Point processes: basics and notation
- Kernel-type product density estimator
- CLT for the integrated square error of the product density estimator
- Outlook

# Point processes – definition



Let $\boldsymbol{N}$ be the set of all locally finite counting measures on $\mathbb{R}^d$, and let $\mathfrak{N}$ be the sigma algebra induced by the sets $\{\psi \in \boldsymbol{N} : \psi(\boldsymbol{B}) = \boldsymbol{n}\}$, $\boldsymbol{B} \in \mathfrak{B}(\mathbb{R}^d)$ (Borel set), $\boldsymbol{n} \in \mathbb{N}_0$.

**Definition:** A point process $\boldsymbol{\Psi}$ in $\mathbb{R}^d$ is a measurable mapping from a probability space $[\boldsymbol{\Omega}, \mathfrak{A}, \mathbb{P}]$ into $[\boldsymbol{N}, \mathfrak{N}]$. Let $\boldsymbol{P} = \mathbb{P} \circ \boldsymbol{\Psi}^{-1}$ denote the probability measure on $[\boldsymbol{N}, \mathfrak{N}]$ induced by $\boldsymbol{\Psi}$, the distribution of $\boldsymbol{\Psi}$, and write $\boldsymbol{\Psi} \sim \boldsymbol{P}$. □

We only consider point processes $\boldsymbol{\Psi}$ that are simple, i.e., $\boldsymbol{\Psi} \in \boldsymbol{N}_s = \{\psi \in \boldsymbol{N} : \psi(\{\boldsymbol{x}\}) \leq 1 \; \forall \boldsymbol{x} \in \mathbb{R}^d\}$.

# Point processes – stationarity and isotropy

**Definition:** A point process $\Psi \sim P$ is stationary if $P$ is translation invariant, i.e.

$$(\Psi(B_1 + x), \ldots, \Psi(B_k + x)) \stackrel{\mathrm{d}}{=} (\Psi(B_1), \ldots, \Psi(B_k))$$

for all $x \in \mathbb{R}^d$, $B_1, \ldots, B_k \in \mathfrak{B}(\mathbb{R}^d)$, $k \geq 1$. $\square$

**Definition:** A point process $\Psi \sim P$ is isotropic if $P$ is rotation invariant, i.e.

$$(\Psi(UB_1), \ldots, \Psi(UB_k)) \stackrel{\mathrm{d}}{=} (\Psi(B_1), \ldots, \Psi(B_k))$$

for all $B_1, \ldots, B_k \in \mathfrak{B}(\mathbb{R}^d)$, $k \geq 1$ and $U \in \mathrm{SO}(d)$. $\square$

**Definition:** If a point process $\Psi$ is both stationary and isotropic, it is called motion invariant. $\square$

We only consider stationary simple point processes.

# Moment measure of order $k$

**Definition:** $\alpha^{(k)}$ (factorial moment measure of order $k$)

For all $B_1, \ldots, B_k \in \mathfrak{B}(\mathbb{R}^d)$:

$$\alpha^{(k)}(B_1 \times \cdots \times B_k) := \mathbb{E} \sum_{\substack{x_1,\ldots,x_k \\ \in \mathrm{supp}(\Psi)}}^{\neq} \mathbb{1}_{B_1}(x_1) \cdots \mathbb{1}_{B_k}(x_k)$$

□

"Factorial": $\alpha^{(k)}(B \times \cdots \times B) = \mathbb{E}\Psi(B)(\Psi(B) - 1) \cdots (\Psi(B) - k + 1)$

$\alpha^{(1)}$ is called intensity measure:

$\alpha^{(1)}(A) = \mathbb{E}\Psi(A) \triangleq$ mean number of points in $A \in \mathfrak{B}(\mathbb{R}^d)$.

$$\Psi \quad \text{stationary} \quad \Rightarrow \quad \alpha^{(1)}(.) = \lambda|.|$$

$\lambda := \mathbb{E}\Psi([0,1]^d) \in (0, \infty)$, the mean number of points in $[0,1]^d$, is called the intensity of the point process $\Psi$.

# Cumulant measure of order $k$

**Definition:** $\gamma^{(k)}$ (cumulant measure of order $k$)

For all $B_1, \ldots, B_k \in \mathfrak{B}(\mathbb{R}^d)$:

$$\gamma^{(k)}(B_1 \times \cdots \times B_k) := \sum_{\ell=1}^{k} (-1)^{\ell-1}(\ell-1)! \sum_{\substack{K_1 \cup \ldots \cup K_\ell \\ = \{1,\ldots,k\}}} \prod_{j=1}^{\ell} \alpha^{(|K_j|)}\Big( \bigtimes_{k_j \in K_j} B_{k_j} \Big)$$

$\square$

"covariance measure" $\gamma^{(2)}$:

For disjoint $A$, $B \in \mathfrak{B}(\mathbb{R}^d)$ we have

$$\begin{aligned} \gamma^{(2)}(A \times B) &= \alpha^{(2)}(A \times B) - \alpha^{(1)}(A)\alpha^{(1)}(B) \\ &= \mathbb{E}\Psi(A)\Psi(B) - \mathbb{E}\Psi(A)\mathbb{E}\Psi(B) \\ &= \mathrm{Cov}[\Psi(A), \Psi(B)]. \end{aligned}$$

# Reduced measures & densities

Let $\Psi$ be a stationary point process with intensity $\lambda$.

**Definition:** $\alpha_{\text{red}}^{(k)}$ (reduced factorial moment measure of order $k$)

$$\alpha^{(k)}(B_1 \times \cdots \times B_k) = \lambda \int_{B_k} \alpha_{\text{red}}^{(k)}((B_1 - x) \times \cdots \times (B_{k-1} - x))\mathrm{d}x$$

□

**Definition:** $\gamma_{\text{red}}^{(k)}$ (reduced cumulant measure of order $k$)

$$\gamma^{(k)}(B_1 \times \cdots \times B_k) = \lambda \int_{B_k} \gamma_{\text{red}}^{(k)}((B_1 - x) \times \cdots \times (B_{k-1} - x))\mathrm{d}x$$

□

**Definition:** The $k$th-order product density $\varrho^{(k)}$ is the Lebesgue density of the $k$th-order reduced factorial moment measure $\alpha_{\text{red}}^{(k)}$. □

**Definition:** The $k$th-order cumulant density $c^{(k)}$ is the Lebesgue density of the $k$th-order reduced cumulant measure $\gamma_{\text{red}}^{(k)}$. □

# Poisson processes and Poisson cluster processes – definition

**Definition:** A point process $\Psi$ is called a Poisson process in $\mathbb{R}^d$ with intensity measure $\Lambda$ iff

- $\Psi(A_1), \ldots, \Psi(A_k)$ are independent for disjoint $A_1, \ldots, A_k \in \mathfrak{B}(\mathbb{R}^d)$ $\forall k \in \mathbb{N}$
- $\Psi(A) \sim \mathrm{Po}(\Lambda(A))$ for every bounded $A \in \mathfrak{B}(\mathbb{R}^d)$ $\square$

Notation: $\Psi \sim \Pi_\Lambda$

$\Psi \sim \Pi_\Lambda$ is stationary $\Leftrightarrow \Lambda(.) = \lambda|.|$ for some $\lambda > 0$

The distribution of a stationary Poisson process with intensity $\lambda$ is denoted by $\Pi_\lambda$.

**Definition:** A stationary Poisson cluster process $\Psi$ in $\mathbb{R}^d$ consists of two components: the primary process $\Psi_\mathrm{p} \sim \Pi_{\lambda_\mathrm{p}}$ $(0 < \lambda_\mathrm{p} < \infty)$ and the secondary process $\Psi_\mathrm{c} \sim P_\mathrm{c}$. Each point $x \in \mathrm{supp}(\Psi_\mathrm{p})$ triggers a point process $\Psi_\mathrm{c}^{[x]} \sim P_\mathrm{c}^{[x]}$ (cluster) that is assumed to be independent of $\Psi_\mathrm{p}$ and $\Psi_\mathrm{c}^{[y]}$, $y \neq x$, and to have the same distribution as the translated process $T_x\Psi_\mathrm{c}$. The condition $\mathbb{E}\Psi_\mathrm{c}(\mathbb{R}^d) < \infty$ guarantees the existence of the Poisson cluster process $\Psi$. $\square$

# Brillinger-mixing point processes

**Definition:** A stationary point process $\boldsymbol{\Psi}$ is called Brillinger-mixing iff

$$\mathbb{E}\boldsymbol{\Psi}^{k}([0,1]^{d}) < \infty \quad \text{and} \quad \|\gamma_{\mathsf{red}}^{(k)}\| := \int_{(\mathbb{R}^d)^{k-1}} |\gamma_{\mathsf{red}}^{(k)}(\mathrm{d}x)| < \infty$$

for all $k \geq 2$. □

For example, stationary Poisson cluster processes with the secondary process $\boldsymbol{\Psi}_{\mathrm{c}}$ satisfying $\mathbb{E}\boldsymbol{\Psi}_{\mathrm{c}}^{k}(\mathbb{R}^{d}) < \infty$ for all $k \geq 1$, are Brillinger-mixing.

# Kernel estimation of the product density

**Definition (Krickeberg 1982 [1]):** Let the stationary point process $\Psi$ be observed in a convex window $W_n$ satisfying $r(W_n) \longrightarrow \infty$ where $r(W_n)$ is the radius of the inscribed sphere of $W_n$. Let $k : \mathbb{R}^d \longrightarrow \mathbb{R}$ be a bounded kernel function with bounded support satisfying $\int_{\mathbb{R}^d} k(x)\mathrm{d}x = 1$. Let the bandwidth $b_n$ satisfy $b_n \longrightarrow 0$ and $b_n^d |W_n| \longrightarrow \infty$. Define

$$\hat{\varrho}_n(t) := \frac{1}{b_n^d |W_n|} \sum_{x,y \in \mathrm{supp}(\Psi)}^{\neq} \mathbb{1}_{W_n}(x) k\left(\frac{y-x-t}{b_n}\right)$$

as a kernel estimator for $\lambda\varrho(t)$, where $\varrho(t) := \varrho^{(2)}(t)$ is the Lebesgue density of $\alpha^{(2)}_{\text{red}}$. □

**Lemma:** Let $\Psi$ be a Brillinger-mixing point process and let $\varrho$ be Lipschitz-continuous in $t \in \mathbb{R}^d$. Then we have

$$\mathbb{E}\hat{\varrho}_n(t) \longrightarrow \lambda\varrho(t).$$

□

# Scaled deviance of the product density estimator (1)

Consider the scaled deviance of the product density estimator,

$$\Delta_n(t) := b_n^{d/2}|W_n|^{1/2}\left(\hat{\varrho}_n(t) - \mathbb{E}\hat{\varrho}_n(t)\right).$$

**Theorem (Heinrich 1988 [2]):** Let $\Psi$ be a stationary Poisson cluster process with intensity $\lambda$. The Lebesgue densities $p^{(2)}$, $p^{(3)}$ and $p^{(4)}$ of the factorial moment measures $\alpha^{(2)}$, $\alpha^{(3)}$ and $\alpha^{(4)}$, respectively, exist und there exist constants $C_1, \ldots, C_4$ such that

$$p(u) := \int_{\mathbb{R}^d} p^{(2)}(u+x, x)\mathrm{d}x \le C_1,$$

$$\int_{(\mathbb{R}^d)^2} p^{(3)}(u, u+x, y)\mathrm{d}(x,y) \le C_2,$$

$$\int_{\mathbb{R}^d} p^{(3)}(u+x, v+x, y)\mathrm{d}x \le C_3,$$

$$\int_{(\mathbb{R}^d)^2} p^{(4)}(u+x, v+y, x, y)\mathrm{d}(x,y) \le C_4$$

for all $u, v \in \mathbb{R}^d$. Furthermore, let the $q$-tuple $(u_1, \ldots, u_q) \in (\mathbb{R}^d)^q$ be chosen such that $u_i \neq u_j$, $u_i \neq -u_j$, $i \neq j$, and every $u_i$, $i = 1, \ldots, q$, is a point of continuity of $p$.

# Scaled deviance of the product density estimator (2)

Recall the scaled deviance of the product density estimator,

$$\Delta_n(t) = b_n^{d/2}|W_n|^{1/2}\left(\hat{\varrho}_n(t) - \mathbb{E}\hat{\varrho}_n(t)\right).$$

**Theorem (Heinrich 1988 [2]), continued:**

Then we have

$$(\Delta_n(u_i))_{i=1}^q \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Sigma_q),$$

where $\mathrm{N}(0, \Sigma_q)$ is a Gaussian vector with covariance matrix $\Sigma_q = (\sigma_{ij})_{i,j=1}^q$, where $\sigma_{ii} = \tau^2 \lambda \varrho(u_i)$, $\tau^2 := \int_{\mathbb{R}^d} k^2(x)\mathrm{d}x$, $i = 1, \ldots, q$, and $\sigma_{ij} = 0$, $i \neq j$.

Furthermore, we have

$$\frac{1}{\tau^2}\sum_{i=1}^q \frac{(\Delta_n(u_i))^2}{\lambda\varrho(u_i)} \xrightarrow{\mathrm{d}} \chi_q^2.$$

□

# Integrated square error (ISE)

The integrated square error (ISE) of the kernel estimator $\hat{\varrho}_n$ on a bounded subset $K \subset \mathbb{R}^d$ satisfying $|K| > 0$ is

$$I_n(K) := \int_K (\hat{\varrho}_n(t) - \lambda\varrho(t))^2 \mathrm{d}t.$$

For a stationary Poisson process with intensity $\lambda$ the second-order product density satisfies $\varrho(t) = \lambda$ for all $t \in \mathbb{R}^d$. This entails

$$I_n(K) = \int_K (\hat{\varrho}_n(t) - \lambda^2)^2 \mathrm{d}t.$$

Because of the asymptotic independence of the components of

$$(\Delta_n(u_i))_{i=1}^q = \left(b_n^{d/2}|W_n|^{1/2}\left(\hat{\varrho}_n(u_i) - \mathbb{E}\hat{\varrho}_n(u_i)\right)\right)_{i=1}^q,$$

for $u_i \neq u_j$ we cannot use Heinrich's result for deriving the asymptotic distribution of the ISE.

# Poisson processes: mean and variance of the ISE

**Lemma:** Let $\boldsymbol{\Psi}$ be a stationary Poisson process with intensity $\boldsymbol{\lambda}$.

Then the expectation of the ISE satisfies, for $\boldsymbol{n} \longrightarrow \infty$,

$$\boldsymbol{b_n^d |W_n| \mathbb{E} \int_K (\hat{\varrho}_n(t) - \lambda^2)^2 \mathrm{d}t = \lambda^2 |K| \int_{\mathbb{R}^d} k^2(y) \mathrm{d}y + O(b_n^d),}$$

and its variance satisfies

$$\boldsymbol{\mathrm{Var}\left( b_n^{d/2} |W_n| \int_K (\hat{\varrho}_n(t) - \lambda^2)^2 \mathrm{d}t \right)}$$

$$\boldsymbol{\longrightarrow 2\lambda^4 \left( |K| + |K \cap \check{K}| \right) \int_{\mathbb{R}^d} \tilde{k}^2(t) \mathrm{d}t,}$$

where $\boldsymbol{\check{K} := \{x \in \mathbb{R}^d : -x \in K\}}$ and $\boldsymbol{\tilde{k}(t) = \int_{\mathbb{R}^d} k(x) k(x+t) \mathrm{d}x.}$ □

# Poisson processes: CLT for the ISE

**Theorem 1:** Let $\Psi$ be a stationary Poisson process with intensity $\lambda$. Let the observation window be of the form $W_n = [0, n)^d$. Let $I_n(K) = \int_K (\hat{\varrho}_n(t) - \lambda^2)^2 \mathrm{d}t$.

Then

$$b_n^{d/2} n^d (I_n(K) - \mathbb{E} I_n(K)) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, 2\lambda^4 \left(|K| + |K \cap \check{K}|\right) \int_{\mathbb{R}^d} \tilde{k}^2(t)\mathrm{d}t\right).$$

The result still holds when $\mathbb{E} I_n(K)$ is replaced with $\dfrac{\lambda^2 |K|}{b_n^{d/2}} \displaystyle\int_{\mathbb{R}^d} k^2(x)\mathrm{d}x$. □

Given a realization of $\Psi \sim P$, Theorem 1 can thus be used for testing complete spatial randomness (test problem $H_0 : P = \Pi_\lambda$ vs. $H_1 : P \neq \Pi_\lambda$ with known intensity $\lambda > 0$).

# Assumptions

**K(d,s)** The kernel function $k$ satisfies

$$\int_{\mathbb{R}^d} x_{i_1} \cdot \ldots \cdot x_{i_\ell} k(x_1, \ldots, x_d) \mathrm{d}(x_1, \ldots, x_d) = 0,$$

for all $i_1, \ldots, i_\ell \in \{1, \ldots, d\}$, $\ell = 1, \ldots, s-1$ (with $s \geq 2$).

**(M1)** The second-order product density $\varrho$ is continuous in $K \oplus b(0, \varepsilon)$ for some $\varepsilon > 0$ with bounded partial derivatives of order $s$.

**(M2)** The third-order cumulant density $c^{(3)}$ and the third-order product density $\varrho^{(3)}$ exist and are bounded.

**(M3)** The fourth-order cumulant density $c^{(4)}$ exists and satisfies

$$\int_{\mathbb{R}^d} |c^{(4)}(x, z, z+y)| \mathrm{d}z \leq C < \infty$$

for all $x, y \in K \oplus b(0, \varepsilon)$ for some $\varepsilon > 0$.

# Brillinger-mixing point processes: mean and variance of the ISE

**Lemma 1:** Let $\boldsymbol{\Psi}$ be a stationary Brillinger-mixing point process in $\mathbb{R}^d$ with intensity $\boldsymbol{\lambda}$. Let the kernel function $\boldsymbol{k}$ satisfy condition **K(d,s)** and let $\boldsymbol{\Psi}$ satisfy the assumptions **(M1)–(M3)**. Then we have, for $n \longrightarrow \infty$,

$$b_n^d|W_n|\,\mathbb{E}\int_K (\hat{\varrho}_n(t) - \lambda\varrho(t))^2 \mathrm{d}t \;\longrightarrow\; \lambda \int_K \varrho(t)\mathrm{d}t \int_{\mathbb{R}^d} k^2(x)\mathrm{d}x.$$

If, in addition, the bandwidth satisfies $b_n^{3d}|W_n| \longrightarrow \infty$ and $b_n^s|W_n| \longrightarrow 0$ (thus $s \geq 3d+1$), then

$$\mathrm{Var}\left(b_n^{d/2}|W_n| \int_K (\hat{\varrho}_n(t) - \lambda\varrho(t))^2 \mathrm{d}t\right)$$
$$\longrightarrow 2\lambda^2 \left(\int_K \varrho^2(t)\mathrm{d}t + \int_{K\cap\check{K}} \varrho^2(t)\mathrm{d}t\right) \int_{\mathbb{R}^d} \tilde{k}^2(t)\mathrm{d}t,$$

where $\check{K} := \{x \in \mathbb{R}^d : -x \in K\}$ and $\tilde{k}(t) = \int_{\mathbb{R}^d} k(x)k(x+t)\mathrm{d}x.$ □

# Poisson cluster processes: CLT for the ISE

**Theorem 2:** Let $\Psi$ be a stationary Poisson cluster process with intensity $\lambda$ and secondary process $\Psi_c$. Assume $\mathbb{E}\Psi_c^8(\mathbb{R}^d) \leq C < \infty$ to hold for some $C > 0$. Let the observation window be of the form $W_n = [0, n)^d$. Let $I_n(K) = \int_K (\hat{\varrho}_n(t) - \lambda\varrho(t))^2 \mathrm{d}t$.

Then we have, under the assumptions of Lemma 1,

$$b_n^{d/2} n^d (I_n(K) - \mathbb{E} I_n(K))$$

$$\xrightarrow{\mathrm{d}} \mathrm{N}\left(0, 2\lambda^2 \left(\int_K \varrho^2(t)\mathrm{d}t + \int_{K \cap \check{K}} \varrho^2(t)\mathrm{d}t\right) \int_{\mathbb{R}^d} \tilde{k}^2(t)\mathrm{d}t\right).$$

The result still holds when $\mathbb{E} I_n(K)$ is replaced with a constant $c_n$ which depends only on the bandwidth $b_n$, the kernel function $k$, the set $K$ and the product density $\varrho$. $\square$

Given a realization of $\Psi \sim P$, Theorem 2 can thus be used for testing $H_0 : P = P_0$ vs. $H_1 : P \neq P_0$, where $P_0$ is the distribution of a Poisson cluster process satisfying the above assumptions with known product density $\varrho$ and intensity $\lambda$.

# CLT for the ISE – Sketch of the Proof

**(1)** Prove the CLT for Poisson cluster processes with bounded cluster radius. In this case a CLT for $m$-dependent point fields (Heinrich 1988 [3]) can be used.

**(2)** In order to prove the CLT for Poisson cluster processes $\Psi$ with secondary process $\Psi_c$ with unbounded cluster radius, use a "truncation method": Let $\Psi^{(\alpha)}$ be the "truncated" Poisson cluster process, where its secondary process $\Psi_c$ is replaced by the truncated cluster $\Psi_c^{(\alpha)} := \Psi_c \cap b(0, \alpha)$.

Due to **(1)**, the CLT for the ISE $I_n^{(\alpha)}(K)$ of $\Psi^{(\alpha)}$ holds.

Showing that for all $\varepsilon > 0$ there exists an $n_0 = n_0(\varepsilon)$ such that

$$\limsup_{\alpha\to\infty} \sup_{n\geq n_0} \mathrm{Var}\left[b_n^{d/2} n^d \left(I_n^{(\alpha)}(K) - I_n(K)\right)\right] \leq \varepsilon$$

entails the CLT for the ISE $I_n(K)$ of $\Psi$.

# Remarks & outlook

- In the setting of Poisson cluster processes, the theory of $m$-dependent point fields can be used for deriving a CLT for the ISE. In the case of Brillinger-mixing point processes this is not possible.

  Idea: Show the ISE's cumulants of order $k \geq 2$ to converge to zero.

- Modifications of the product density estimator (e.g. edge-correction) should also be considered.

- How large does the observation window $W_n$ have to be, i.e., how many points do we need for a satisfactory approximation from the CLT?

  Simulation studies

# References


[1] Krickeberg, K. (1982). *Processus ponctuels en statistique,* École d'Été de Probabilités de Saint-Flour X-1980. Lecture Notes in Mathematics 929, Springer, Berlin.

[2] Heinrich, L. (1988). Asymptotic Gaussianity of Some Estimators for Reduced Factorial Moment Measures and Product Densities of Stationary Poisson Cluster Processes, *Statistics* **19** 1, 87–106.

[3] Heinrich, L. (1988). Asymptotic behaviour of an empirical nearest-neighbour distance function for stationary poisson cluster processes, *Mathematische Nachrichten* **136**, 131–148.