Closure Relations

Adrian Baddeley

Joint work with G. Nair, A. Mira, R.S. Anderssen, F. DeHoog, P. Grabarnik

Statistical theory often says we “should” act in a particular way

Statistical theory often says we “should” act in a particular way

- minimum variance estimation

Statistical theory often says we “should” act in a particular way

- minimum variance estimation
- maximum likelihood

Statistical theory often says we “should” act in a particular way

- minimum variance estimation
- maximum likelihood
- minimum Bayes risk

Statistical theory often says we “should” act in a particular way

- minimum variance estimation
- maximum likelihood
- minimum Bayes risk

Markov chains do not have such a “compulsory” interpretation

In a nutshell

Suppose $\pi\, P = \pi$

In a nutshell

Suppose $\pi\, P = \pi$

where π is a vector of probabilities and P is a transition matrix.

In a nutshell

Suppose $\pi\, P = \pi$
where π is a vector of probabilities and P is a transition matrix. [Assuming . . .],

- the Markov chain with transition probabilities P converges in law to π

In a nutshell

Suppose $\pi\, P = \pi$
where π is a vector of probabilities and P is a transition matrix. [Assuming . . .],

- the Markov chain with transition probabilities P converges in law to π
  $\longrightarrow$ MCMC methods for simulating from π

In a nutshell

Suppose $\pi\, P = \pi$
where π is a vector of probabilities and P is a transition matrix. [Assuming . . .],

- the Markov chain with transition probabilities P converges in law to π
  $\longrightarrow$ MCMC methods for simulating from π
- for any vector v,

$$\pi v = \pi\,[Pv]$$

In a nutshell

Suppose $\pi\, P = \pi$
where π is a vector of probabilities and P is a transition matrix. [Assuming . . .],

- the Markov chain with transition probabilities P converges in law to π
  $\longrightarrow$ MCMC methods for simulating from π
- for any vector v,
  $$\pi v = \pi\,[Pv]$$
  $\longrightarrow$ Recurrence relation between moments of π

In a nutshell

Suppose $\pi\, P = \pi$
where π is a vector of probabilities and P is a transition matrix. [Assuming . . .],

- the Markov chain with transition probabilities P converges in law to π
  $\longrightarrow$ MCMC methods for simulating from π
- for any vector v,
  $$\pi v = \pi\,[Pv]$$
  $\longrightarrow$ Recurrence relation between moments of π
- π is characterised by
  $$\pi[(P - I)v] = 0 \qquad \text{for all } v$$

In a nutshell

Suppose $\pi\, P = \pi$
where π is a vector of probabilities and P is a transition matrix. [Assuming . . .],

- the Markov chain with transition probabilities P converges in law to π
  $\longrightarrow$ MCMC methods for simulating from π
- for any vector v,
  $$\pi v = \pi\,[Pv]$$
  $\longrightarrow$ Recurrence relation between moments of π
- π is characterised by
  $$\pi[(P - I)v] = 0 \qquad \text{for all } v$$
  $\longrightarrow$ Stein's method for approximation by π

In a nutshell

Suppose $\pi\, P = \pi$
where π is a vector of probabilities and P is a transition matrix. [Assuming . . .],

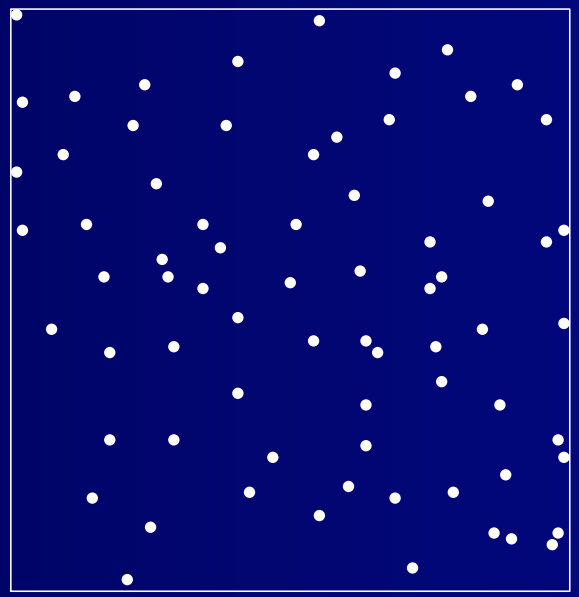
- the Markov chain with transition probabilities P converges in law to π
  $\longrightarrow$ MCMC methods for simulating from π
- for any vector v,
  $$\pi v = \pi\,[Pv]$$
  $\longrightarrow$ Recurrence relation between moments of π
- π is characterised by
  $$\pi[(P - I)v] = 0 \qquad \text{for all } v$$
  $\longrightarrow$ Stein's method for approximation by π
  $\longrightarrow$ Estimating equations for fitting π_θ to data

Plan

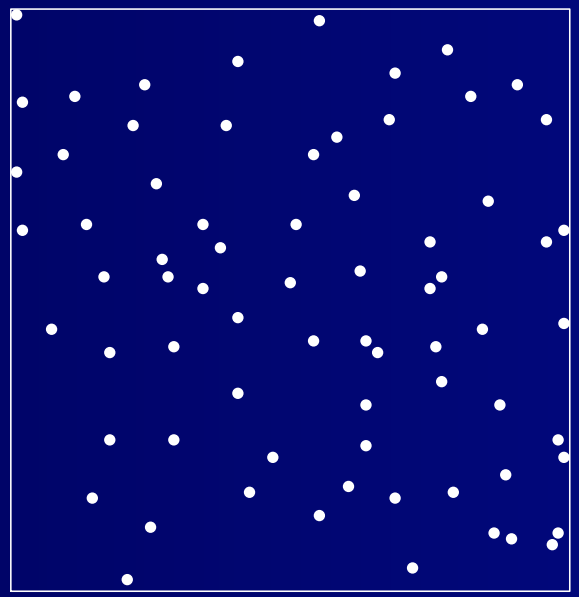
1. Introduction
2. Spatial point process models
3. MCMC for spatial point processes
4. Recurrence relations for moments
5. Estimating equations

2. Spatial point process models

Point pattern dataset



Point pattern dataset



$$\mathbf{x} = \{x_1, \ldots, x_n\}, \quad x_i \in D, \quad n \geq 0$$

Point process model

Model $\mathbf{x}$ as a realisation of a *random point process* $\mathbf{X}$ whose possible outcomes are (finite) point patterns in D.

Poisson point process model

The homogeneous *Poisson* point process Π with intensity $\lambda > 0$ in D is characterised by

Poisson point process model

The homogeneous *Poisson* point process Π with intensity $\lambda > 0$ in D is characterised by

1. total number of points $N = n(\Pi)$ has a Poisson distribution with mean $\mu = \lambda|D|$

Poisson point process model

The homogeneous *Poisson* point process Π with intensity $\lambda > 0$ in D is characterised by

1. total number of points $N = n(\Pi)$ has a Poisson distribution with mean $\mu = \lambda|D|$
2. Given $N = n$, the n points are stochastically independent and uniformly distributed in D.

Poisson point process model

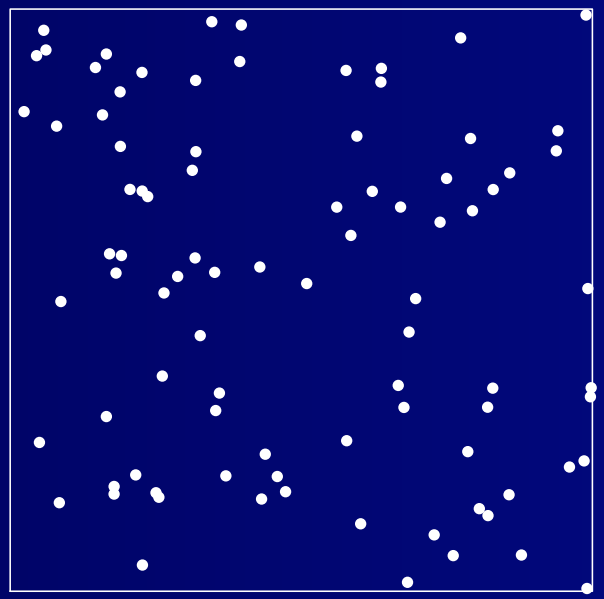
The homogeneous *Poisson* point process Π with intensity $\lambda > 0$ in D is characterised by

1. total number of points $N = n(\Pi)$ has a Poisson distribution with mean $\mu = \lambda|D|$
2. Given $N = n$, the n points are stochastically independent and uniformly distributed in D.

For any real-valued function $h(\mathbf{x})$ of the configuration $\mathbf{x}$,

$$\mathbb{E}\, h(\Pi) = e^{-\lambda|D|} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \int_D \cdots \int_D h(\mathbf{x})\, \mathrm{d}x_1\, \mathrm{d}x_2 \ldots\, \mathrm{d}x_n$$

Poisson process



Probability density

We can specify a finite point process model by its *probability density* $f(\mathbf{x})$

Probability density

We can specify a finite point process model by its *probability density* $f(\mathbf{x})$

relative to the distribution of the Poisson point process with unit intensity in D

Probability density

We can specify a finite point process model by its *probability density* $f(\mathbf{x})$

relative to the distribution of the Poisson point process with unit intensity in D

$$\begin{aligned}\mathbb{E}\, h(\mathbf{X}) &= \mathbb{E}[h(\Pi) f(\Pi)] \\ &= e^{-|D|} \sum_{n=0}^{\infty} \frac{1}{n!} \int_D \cdots \int_D h(\mathbf{x}) f(\mathbf{x}) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \, \mathrm{d}x_n\end{aligned}$$

Gibbs model

A *Gibbs point process model* has a probability density of the form

$$f(\mathbf{x}) = \exp V(\mathbf{x})$$

where

$$V(\mathbf{x}) = V_0 + \sum_i V_1(x_i) + \sum_{i<j} V_2(x_i, x_j) + \ldots$$

is called the *potential*.

Gibbs model

A *Gibbs point process model* has a probability density of the form

$$f(\mathbf{x}) = \exp V(\mathbf{x})$$

where

$$V(\mathbf{x}) = V_0 + \sum_i V_1(x_i) + \sum_{i<j} V_2(x_i, x_j) + \ldots$$

is called the *potential*.

V_0 is the normalising constant, and $V_1, V_2, V_3, \ldots$ are functions with values in $[-\infty, \infty)$ called the *potentials of order* $1, 2, 3, \ldots$.

Gibbs model

A *Gibbs point process model* has a probability density of the form

$$f(\mathbf{x}) = \exp V(\mathbf{x})$$

where

$$V(\mathbf{x}) = V_0 + \sum_i V_1(x_i) + \sum_{i<j} V_2(x_i, x_j) + \ldots$$

is called the *potential*.
V_0 is the normalising constant, and $V_1, V_2, V_3, \ldots$ are functions with values in $[-\infty, \infty)$ called the *potentials of order* $1, 2, 3, \ldots$.

The potentials $V_2, V_3, \ldots$ introduce *dependence/interaction between points*.

Hard core process

Assume $\exp V_1(u) \equiv \beta > 0$ and

$$\exp V_2(u,v) = \begin{cases} 0 & \text{if } \|u-v\| \leq R \\ 1 & \text{otherwise} \end{cases}$$

where $R > 0$ is a threshold distance, and $V_k \equiv 0$ for $k > 2$. Then

$$f(\mathbf{x}) = \begin{cases} \alpha\beta^{n(\mathbf{x})} & \text{if } \|x_i - x_j\| > R \text{ for all } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha = \exp(V_0)$. This is equivalent to a Poisson process of intensity β, conditioned on the event that $\|x_i - x_j\| > R$ for all $i \neq j$. Known as a *hard core* process.

Hard core process

Assume $\exp V_1(u) \equiv \beta > 0$ and

$$\exp V_2(u,v) = \begin{cases} 0 & \text{if } \|u-v\| \leq R \\ 1 & \text{otherwise} \end{cases}$$

where $R > 0$ is a threshold distance, and $V_k \equiv 0$ for $k > 2$. Then

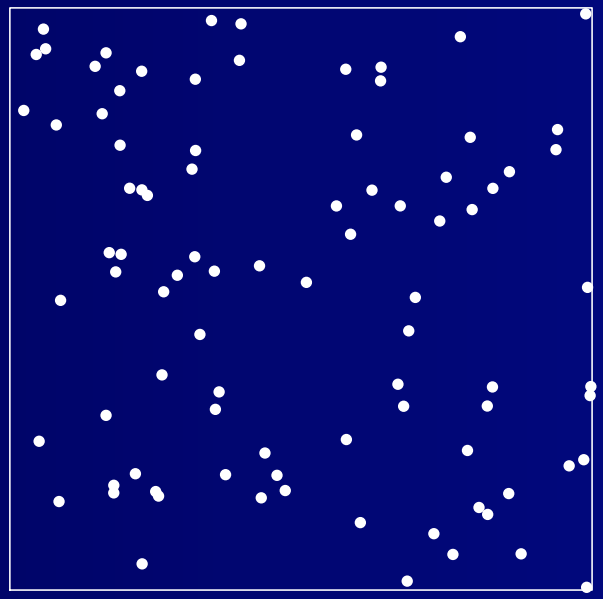
$$f(\mathbf{x}) = \begin{cases} \alpha\beta^{n(\mathbf{x})} & \text{if } \|x_i - x_j\| > R \text{ for all } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha = \exp(V_0)$. This is equivalent to a Poisson process of intensity β, conditioned on the event that $\|x_i - x_j\| > R$ for all $i \neq j$. Known as a *hard core* process.

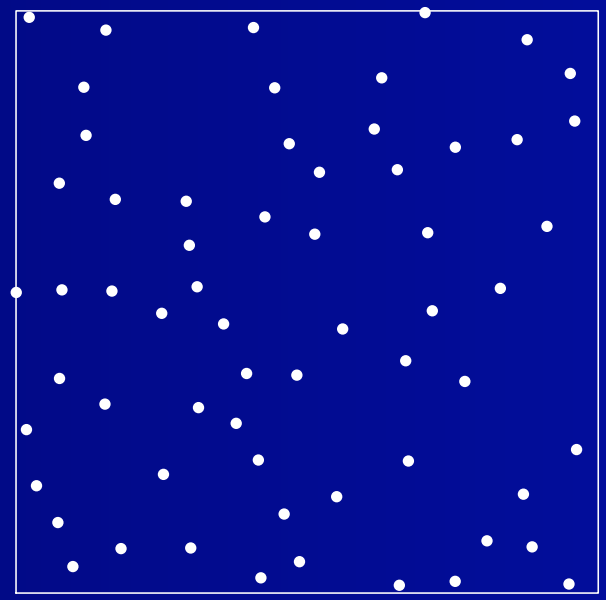
Note $\alpha = e^{(1-\beta)|D|}/p(\beta, R)$ where $p(\beta, R) = \mathbb{P}\left\{\Pi_\beta \text{ has no } R\text{-close pairs}\right\}$ is not tractable

Hard core process

Poisson



Hard core



Strauss process

Let $\exp V_1(u) \equiv \beta > 0$ and

$$\exp V_2(u,v) = \begin{cases} \gamma & \text{if } \|u-v\| \le R \\ 1 & \text{otherwise} \end{cases}$$

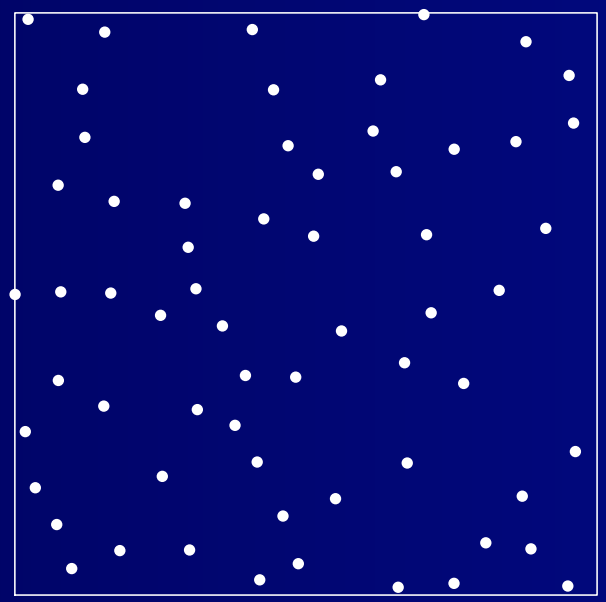
where $0 \le \gamma \le 1$ is the "interaction strength". Then

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{s(\mathbf{x})}$$

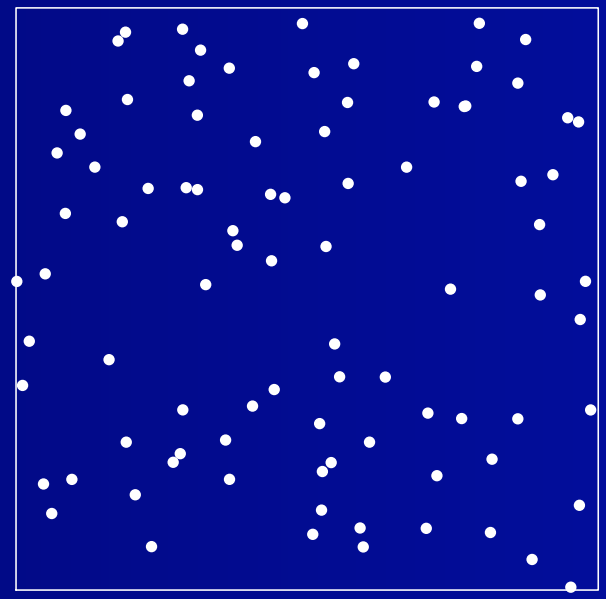
where

$$s(\mathbf{x}) = \sum_{i<j} \mathbf{1}\{\|x_i - x_j\| \le R\}.$$

Hard core



Strauss



Plan

3. MCMC for spatial point processes

Gibbs sampler for finite system

For a finite system of random variables $\mathbf{X} = (X_1, \ldots, X_n)$, the Gibbs sampler is a Markov chain with states $\mathbf{x} = (x_1, \ldots, x_n)$ in which each transition involves choosing a site i and updating x_i according to the one-site conditional probabilities

$$\mathbb{P}\left\{X_i = x_i \mid X_j = x_j, j \neq i\right\}$$

Gibbs sampler for point process

For a finite spatial point process $\mathbf{X}$, the analogue of the Gibbs sampler is a spatial birth-death process:

Gibbs sampler for point process

For a finite spatial point process $\mathbf{X}$, the analogue of the Gibbs sampler is a spatial birth-death process: a continuous-time Markov process whose states are point patterns $\mathbf{x}$,

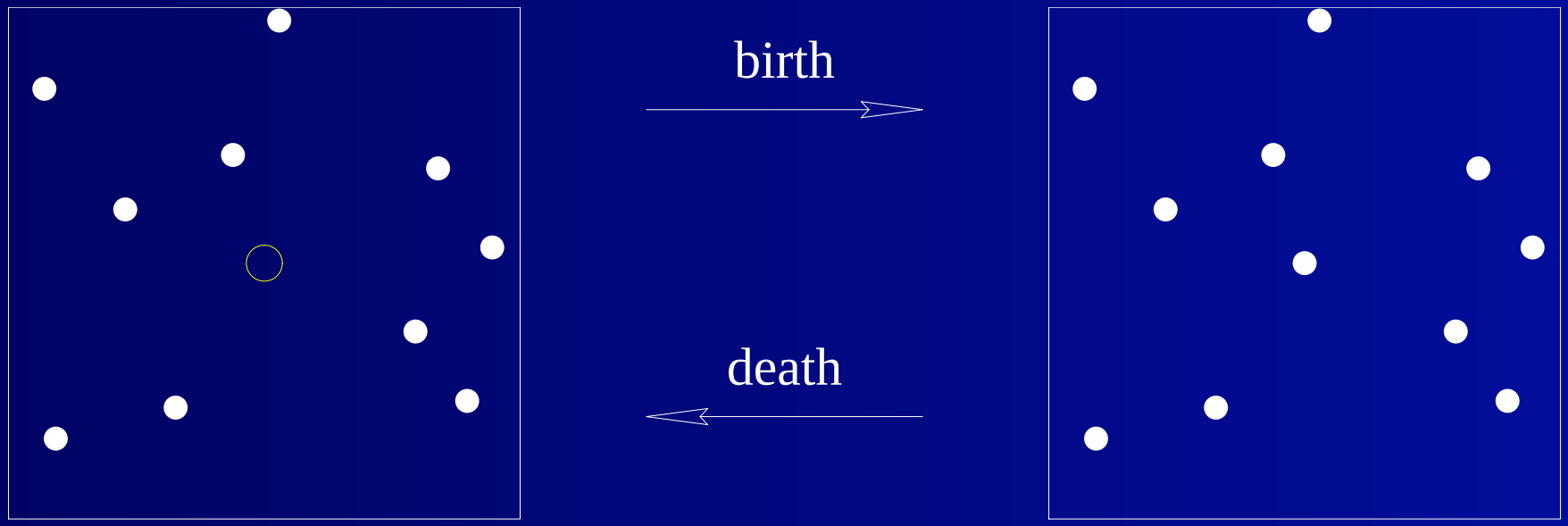
Gibbs sampler for point process

For a finite spatial point process $\mathbf{X}$, the analogue of the Gibbs sampler is a spatial birth-death process: a continuous-time Markov process whose states are point patterns $\mathbf{x}$, in which each transition is either instantaneous deletion ("death")

$$\mathbf{x} \mapsto \mathbf{x} \setminus \{x_i\}$$

or instantaneous addition ("birth")

$$\mathbf{x} \mapsto \mathbf{x} \cup \{u\}$$


birth
death

Gibbs sampler for point process

For detailed balance,

$$\frac{\text{birth}}{\text{death}} = \frac{b(\mathbf{x} \mapsto \mathbf{x} \cup \{u\})}{d(\mathbf{x} \cup \{u\} \mapsto \mathbf{x})} = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})}$$

Gibbs sampler for point process

For detailed balance,

$$\frac{\text{birth}}{\text{death}} = \frac{b(\mathbf{x} \mapsto \mathbf{x} \cup \{u\})}{d(\mathbf{x} \cup \{u\} \mapsto \mathbf{x})} = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})}$$

Without loss of generality,

$$\begin{aligned} d() &\equiv 1 \\ b(\mathbf{x} \mapsto \mathbf{x} \cup \{u\}) &= \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} \end{aligned}$$

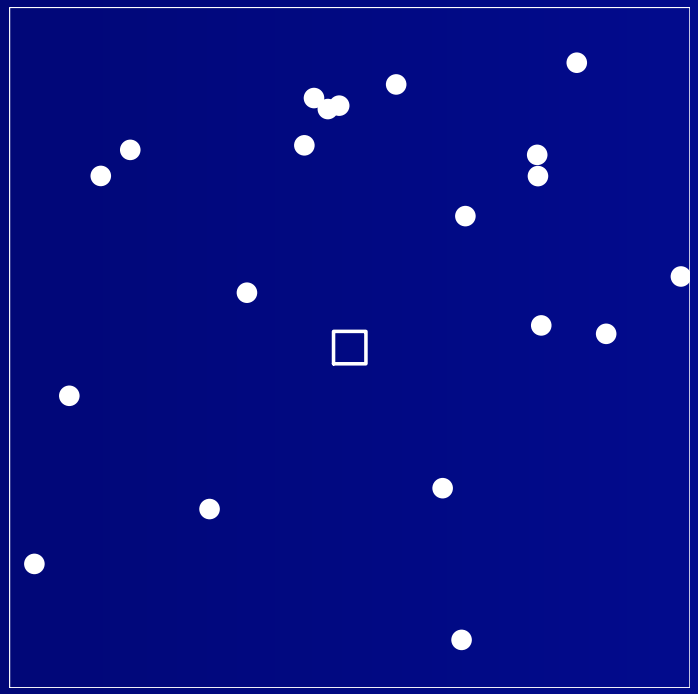
so each point has an independent exponential(1) lifetime

Conditional intensity of point process

For a finite point process with probability density $f(\mathbf{x})$, the Papangelou *conditional intensity* at a location $u \notin \mathbf{x}$ is

$$\Lambda(u, \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})}$$

Conditional intensity of point process



Roughly speaking, $\Lambda(u, \mathbf{x})\ \mathrm{d}u$ is the probability that $\mathbf{X}$ has a point in an infinitesimal neighbourhood of u, given complete information about the realisation of $\mathbf{X}$ outside this neighbourhood.

Conditional intensity of hard core process

For a hard core process,

$$\Lambda(u, \mathbf{x}) = \begin{cases} \beta & \text{if } \|u - x_i\| > R \text{ for all } i \\ 0 & \text{otherwise} \end{cases}$$

Conditional intensity of hard core process

For a hard core process,

$$\Lambda(u, \mathbf{x}) = \begin{cases} \beta & \text{if } \|u - x_i\| > R \text{ for all } i \\ 0 & \text{otherwise} \end{cases}$$

In the Gibbs sampler, new points are added at a uniform rate β at all locations permitted by the hard core constraint.

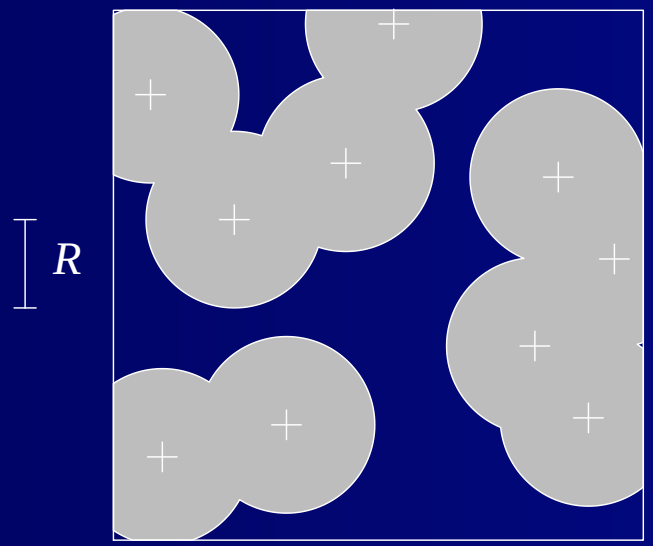
Conditional intensity of Strauss process

For a Strauss process,

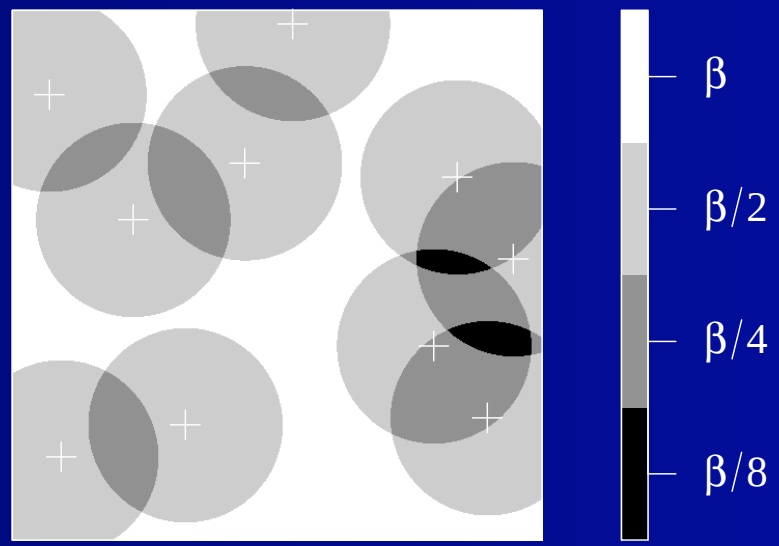
$$\Lambda(u, \mathbf{x}) = \begin{cases} \beta\gamma^{t(u,\mathbf{x})} & \text{if } \|u - x_i\| > R \text{ for all } i \\ 0 & \text{otherwise} \end{cases}$$

where $t(u, \mathbf{x}) = \sum_i \mathbf{1}\{\|x_i - u\| \leq R\}$ is the number of points of $\mathbf{x}$ close to u.

Hardcore



Strauss



Conditional intensity $\Lambda(u, \mathbf{x})$ as a function of u for fixed realisation $\mathbf{x}$

Plan

4. Recurrence relations for moments

Moments

The *intensity function* $\lambda(u)$ of a point process $\mathbf{X}$ is the spatially-varying expected number of points per unit area, defined so that, for all $B \subseteq D$,

$$\mathbb{E}n(\mathbf{X} \cap B) = \int_B \lambda(u)\,\mathrm{d}u.$$

Goal:

For a Gibbs point process $\mathbf{X}$ with given potentials V_k, find the intensity

Moments

Goal:

For a Gibbs point process $\mathbf{X}$ with given potentials V_k, find the intensity

This problem is analytically intractable

Moments

Intractability of the moments of Gibbs point processes is a major obstacle, because the moments are needed in

Moments

Intractability of the moments of Gibbs point processes is a major obstacle, because the moments are needed in

- statistical physics (“equation of state”)

Moments

Intractability of the moments of Gibbs point processes is a major obstacle, because the moments are needed in

- statistical physics (“equation of state”)
- probability modelling/prediction

Moments

Intractability of the moments of Gibbs point processes is a major obstacle, because the moments are needed in

- statistical physics (“equation of state”)
- probability modelling/prediction
- fitting models to point pattern data

Moments

Intractability of the moments of Gibbs point processes is a major obstacle, because the moments are needed in

- statistical physics (“equation of state”)
- probability modelling/prediction
- fitting models to point pattern data
- model diagnostics

Moments

Intractability of the moments of Gibbs point processes is a major obstacle, because the moments are needed in

- statistical physics (“equation of state”)
- probability modelling/prediction
- fitting models to point pattern data
- model diagnostics
- . . .

Moments

Intractability of the moments of Gibbs point processes was the original motivation for the invention of *Markov Chain Monte Carlo* methods

N. Metropolis et al (1953)
Equation of state calculations by fast computing machines.
Journal of Chemical Physics **21** 1087–1092

“Homogeneous” case

Let D be a rectangle, and identify opposite edges, making it a torus. Assume V_k are invariant under translation on the torus, so the point process is “homogeneous” (“stationary”).

Then the intensity λ is constant, and $\mathbb{E}n(\mathbf{X} \cap B) = \lambda\,|B|$ for $B \subseteq D$.

Let’s try to determine the constant λ.

Georgii-Nguyen-Zessin identity

For any functional $h(\mathbf{x})$ and any fixed location $u \in D$

$$\mathbb{E}h(\mathbf{X}) = \mathbb{E}\left[\frac{\Lambda(u, \mathbf{X} \setminus \{u\})}{\lambda} h(\mathbf{X} \setminus \{u\}) \,\middle|\, u \in \mathbf{X}\right]$$

where λ is the (constant) intensity, and the right hand side is formally defined as an expectation with respect to the Palm distribution of $\mathbf{X}$ at u.

Georgii-Nguyen-Zessin identity

For any functional $h(\mathbf{x})$ and any fixed location $u \in D$

$$\mathbb{E}h(\mathbf{X}) = \mathbb{E}\left[\frac{\Lambda(u, \mathbf{X} \setminus \{u\})}{\lambda} h(\mathbf{X} \setminus \{u\}) \,\middle|\, u \in \mathbf{X}\right]$$

where λ is the (constant) intensity, and the right hand side is formally defined as an expectation with respect to the Palm distribution of $\mathbf{X}$ at u.

Taking $h(\mathbf{x}) \equiv 1$ gives

$$\lambda = \mathbb{E}\left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X}\right].$$

GNZ identity

$$\lambda = \mathbb{E}\left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X}\right]$$

GNZ identity

$$\lambda = \mathbb{E}\left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X}\right]$$

For example, for the Strauss process,

$$\Lambda(u, \mathbf{x}) = \beta \gamma^{t(u,\mathbf{x})}$$

where $t(u, \mathbf{x}) = \sum_i \mathbf{1}\{\|x_i - u\| \leq R\}$.

GNZ identity

$$\lambda = \mathbb{E}\left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X}\right]$$

For example, for the Strauss process,

$$\Lambda(u, \mathbf{x}) = \beta \gamma^{t(u,\mathbf{x})}$$

where $t(u, \mathbf{x}) = \sum_i \mathbf{1}\{\|x_i - u\| \leq R\}$. So the intensity of the Strauss process satisfies

$$\lambda = \beta \mathbb{E}\left[\gamma^{t(u,\mathbf{X}\setminus\{u\})} \mid u \in \mathbf{X}\right].$$

GNZ identity

$$\lambda = \mathbb{E}\left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X}\right]$$

For example, for the Strauss process,

$$\Lambda(u, \mathbf{x}) = \beta \gamma^{t(u,\mathbf{x})}$$

where $t(u, \mathbf{x}) = \sum_i \mathbf{1}\{\|x_i - u\| \leq R\}$. So the intensity of the Strauss process satisfies

$$\lambda = \beta \mathbb{E}\left[\gamma^{t(u,\mathbf{X}\setminus\{u\})} \mid u \in \mathbf{X}\right].$$

Here the conditional intensity $\Lambda(u, \mathbf{x})$ is known explicitly.

GNZ identity

$$\lambda = \mathbb{E}\left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X}\right]$$

For example, for the Strauss process,

$$\Lambda(u, \mathbf{x}) = \beta \gamma^{t(u,\mathbf{x})}$$

where $t(u, \mathbf{x}) = \sum_i \mathbf{1}\{\|x_i - u\| \leq R\}$. So the intensity of the Strauss process satisfies

$$\lambda = \beta \mathbb{E}\left[\gamma^{t(u,\mathbf{X}\setminus\{u\})} \mid u \in \mathbf{X}\right].$$

Here the conditional intensity $\Lambda(u, \mathbf{x})$ is known explicitly. But the right hand side is an expectation with respect to the unknown (Palm) distribution of the point process, so both sides are intractable.

Idea for approximation

Idea:

Approximate the RHS by an expectation with respect to a **Poisson** process with the same unknown intensity λ

Idea for approximation

Idea:

Approximate the RHS by an expectation with respect to a **Poisson** process with the same unknown intensity λ

i.e.

$$\lambda \approx \mathbb{E}_{\mathsf{Pois}(\lambda)} \left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X} \right]$$

Idea for approximation

Idea:

Approximate the RHS by an expectation with respect to a **Poisson** process with the same unknown intensity λ

i.e.

$$\lambda \approx \mathbb{E}_{\mathsf{Pois}(\lambda)} \left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X} \right]$$

by Slivnyak's Theorem this is equivalent to

$$\lambda \approx \mathbb{E}_{\mathsf{Pois}(\lambda)} \left[\Lambda(u, \mathbf{X}) \right]$$

Solve for λ to obtain the "Poisson-saddlepoint approximation" λ^{PS}

Strauss process

$$\lambda \approx \mathbb{E}_{\mathsf{Pois}(\lambda)}\left[\Lambda(u, \mathbf{X})\right]$$

For the Strauss process

$$\begin{aligned}\mathbb{E}_{\mathsf{Pois}(\lambda)}\left[\Lambda(u, \mathbf{X})\right] &= \mathbb{E}_{\mathsf{Pois}(\lambda)}\left[\beta\,\gamma^{t(u,\mathbf{X})}\right] \\ &= \beta\exp(-\lambda(1-\gamma)\pi R^2).\end{aligned}$$

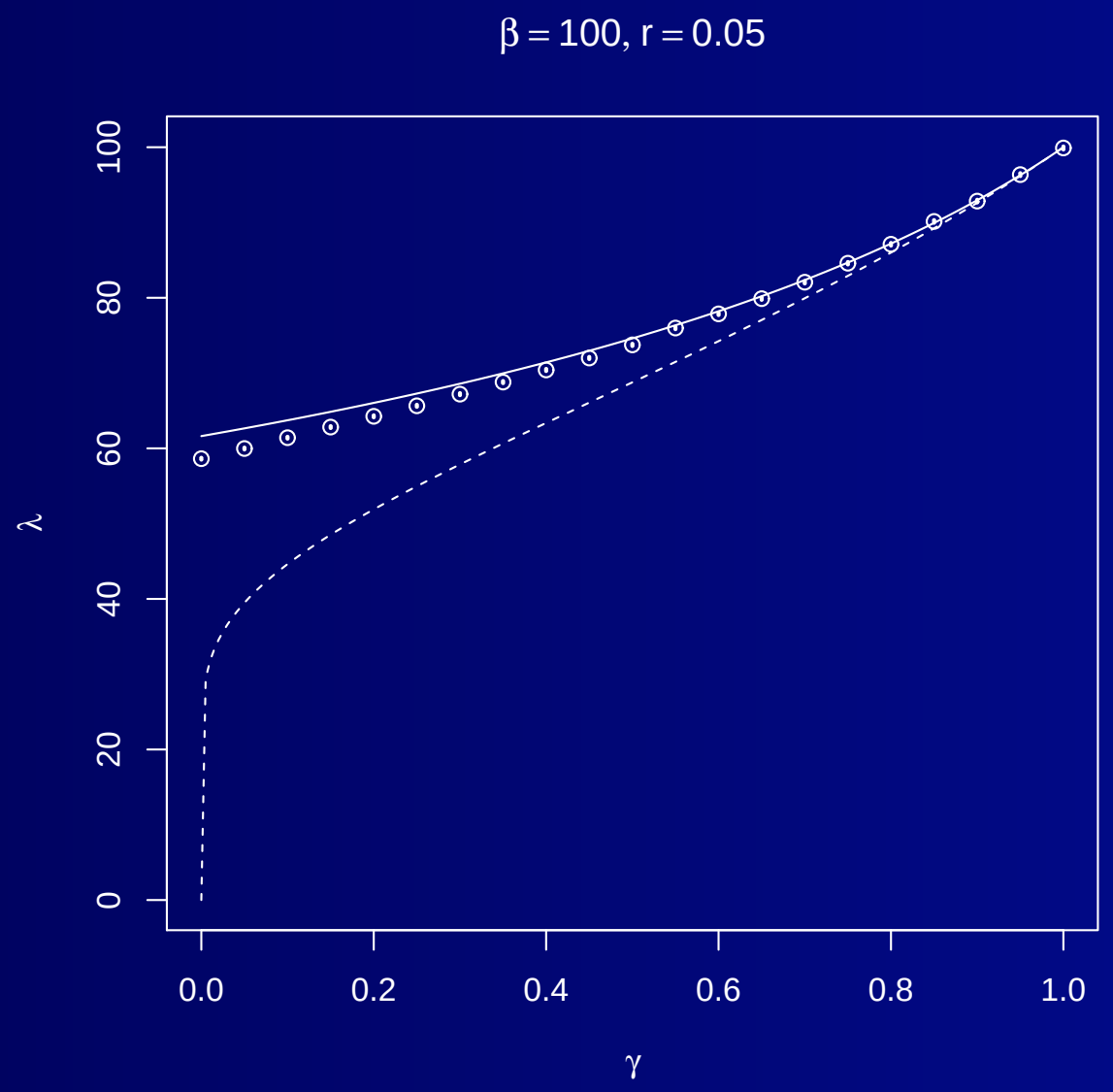
Thus we solve

$$\lambda = \exp(-\lambda G)$$

where $G = (1-\gamma)\pi R^2$. The solution is

$$\lambda^{\mathsf{PS}} = \frac{W(\beta G)}{G}$$

where W is the inverse function of $x \mapsto xe^x$, known as *Lambert's W-function*.



○ exact values

— Poisson-saddlepoint

– – mean field

Similarity to mean field approximation

Possion-saddlepoint:

$$\lambda \approx \mathbb{E}_{\mathsf{Pois}(\lambda)}\left[\Lambda(u, \mathbf{X})\right]$$

Mean field:

$$\lambda \approx \exp \mathbb{E}_{\mathsf{Pois}(\lambda)}\left[\log \Lambda(u, \mathbf{X})\right]$$

Inhomogeneous case

Remove the “homogeneity” assumptions.

Now we want to find the *intensity function* $\lambda(u), u \in D$ defined by

$$\mathbb{E}\, n(\mathbf{X} \cap B) = \int_B \lambda(u)\, \mathrm{d}u.$$

Inhomogeneous case

The GNZ identity gives

$$\lambda(u) = \mathbb{E}\left[\Lambda(u, \mathbf{X} \setminus \{u\}) \mid u \in \mathbf{X}\right]$$

Again we approximate this by

$$\lambda(u) = \mathbb{E}_{\mathsf{Pois}(\lambda(\cdot))}\left[\Lambda(u, \mathbf{X})\right]$$

where the expectation is with respect to a Poisson process with intensity function $\lambda(\cdot)$.

Pairwise interaction

Consider a general pairwise interaction process (where the only condition is $V_k \equiv 0$ for $k > 2$).

Then the Poisson-saddlepoint approximation $\lambda^{\mathsf{PS}}(\cdot)$ is the solution of the integral equation

$$\lambda(u) = \beta(u) \exp(-\int_D (1 - \gamma(u,v))\lambda(v)\, \mathrm{d}v)$$

where $\beta(u) = \exp V_1(u)$ and $\gamma(u,v) = \exp V_2(u,v)$.

Pairwise interaction

Consider a general pairwise interaction process (where the only condition is $V_k \equiv 0$ for $k > 2$).

Then the Poisson-saddlepoint approximation $\lambda^{\mathsf{PS}}(\cdot)$ is the solution of the integral equation

$$\lambda(u) = \beta(u) \exp(-\int_D (1 - \gamma(u,v))\lambda(v)\, \mathrm{d}v)$$

where $\beta(u) = \exp V_1(u)$ and $\gamma(u,v) = \exp V_2(u,v)$.

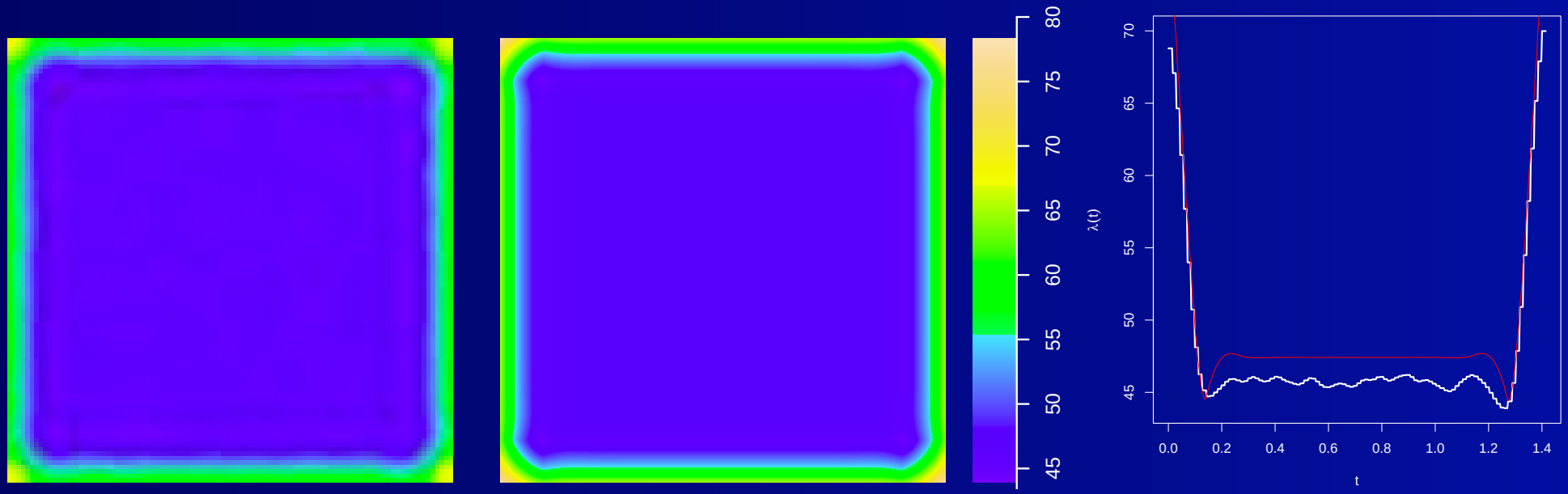
Solve by (under-relaxed) Picard iteration

Example 1: Simulation in bounded domain

Simulation of the stationary Strauss process in a bounded domain D does not yield a stationary process, due to “edge effects”. How bad are the edge effects?

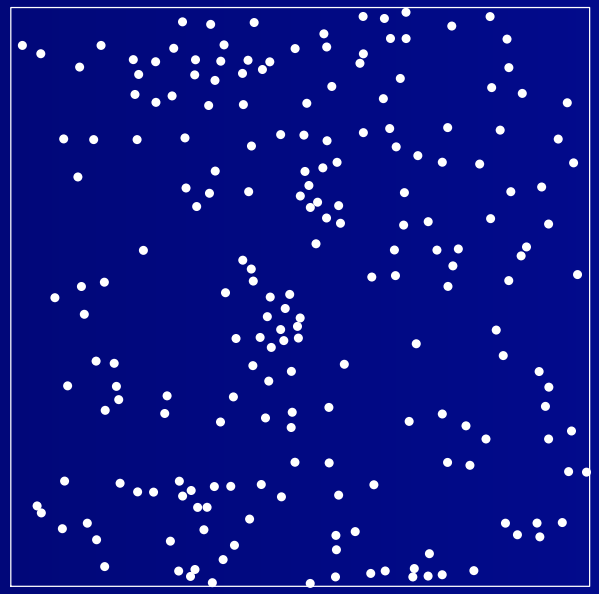
Example 1: Simulation in bounded domain

Simulation of the stationary Strauss process in a bounded domain D does not yield a stationary process, due to “edge effects”. How bad are the edge effects?



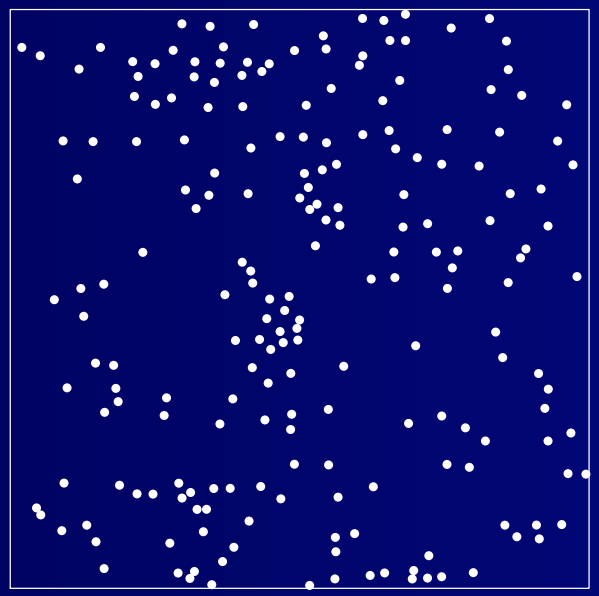
Intensity of finite Strauss process in rectangle

Example 2: data analysis

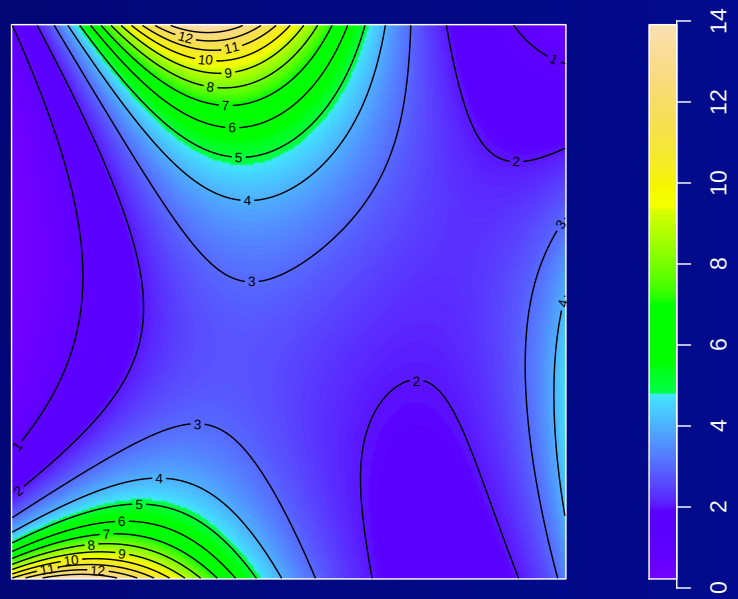


Japanese Pines data

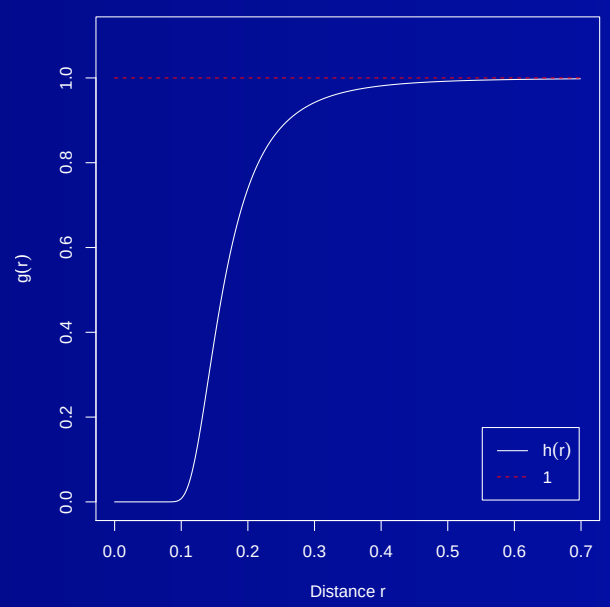
Example 2: data analysis



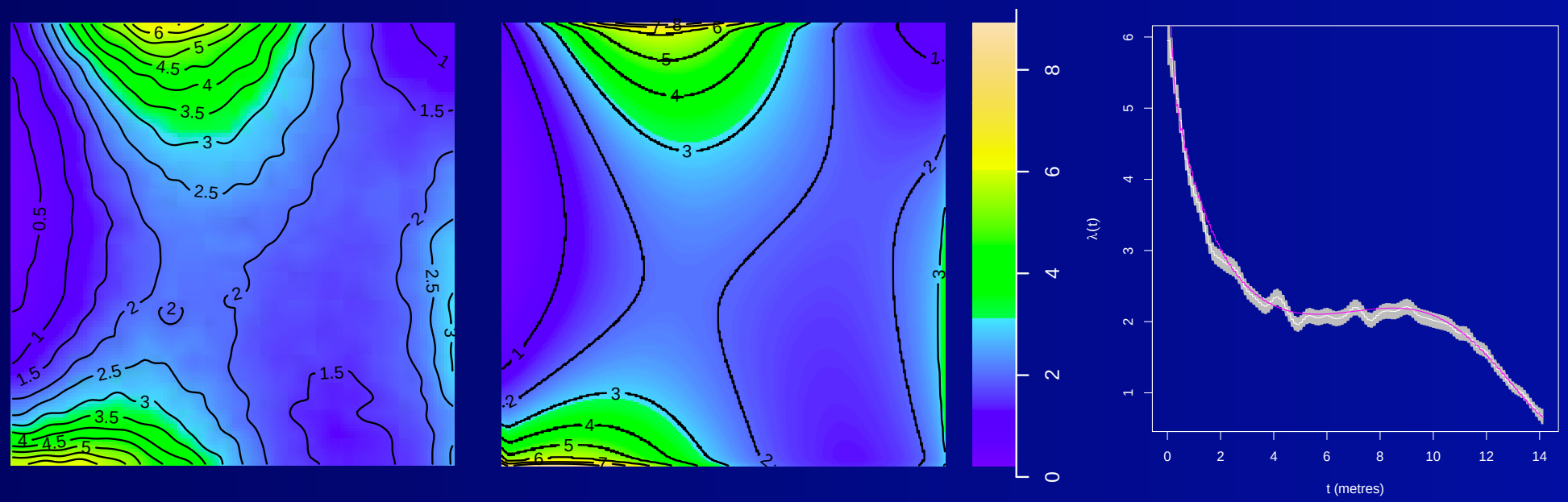
data



$\exp V_1$

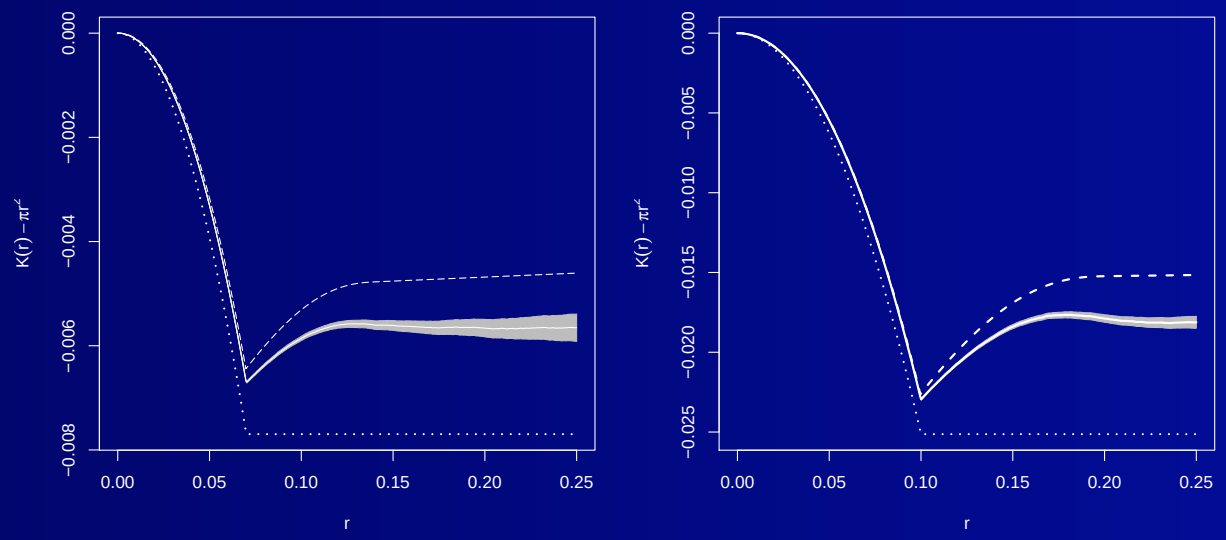


$\exp V_2$



Intensity estimated by MCMC and by Poisson-saddlepoint approximation

Second moments



K-function of the Strauss process: true value (grey shading), Poisson-saddlepoint approximation (dashed lines), sparse approximation (dotted lines).

Summary

Summary

- Accurate approximations to the first two moments of a Gibbs process

Summary

- Accurate approximations to the first two moments of a Gibbs process
- Many potential applications

Summary

- Accurate approximations to the first two moments of a Gibbs process
- Many potential applications

  model fitting, model validation, prediction, conditional prediction, Bayesian inference, improvements to MCMC.

Summary

- Accurate approximations to the first two moments of a Gibbs process
- Many potential applications

  model fitting, model validation, prediction, conditional prediction, Bayesian inference, improvements to MCMC.

- Potential for fast maximum likelihood estimation

Summary

- Accurate approximations to the first two moments of a Gibbs process
- Many potential applications

  model fitting, model validation, prediction, conditional prediction, Bayesian inference, improvements to MCMC.

- Potential for fast maximum likelihood estimation
- Available in open source software `spatstat`

References

- A. Baddeley and G. Nair.
  Fast approximation of the intensity of Gibbs point processes.
  Electronic Journal of Statistics **6** (2012) 1155–1169.
- A. Baddeley and G. Nair.
  Approximating the moments of a spatial point process.
  Stat **1** (2012) 18-30.
- R.S. Anderssen *et al.*
  Solution of an integral equation arising in spatial point process theory.
  J. Integral Equations & Applicns. **26** (2014) 437–453.

Plan

5. Estimating equations

Goal

Goal:

find *computationally tractable* parameter estimators for fitting Gibbs point process models to point pattern data

Maximum likelihood

Assume parametric model, with probability density

$$f(\mathbf{x}, \theta)$$

where θ is the parameter, whose value we want to estimate. The **<u>maximum likelihood estimate</u>** (MLE) of θ from data $\mathbf{x}$ is

$$\widehat{\theta} = \text{argmax}_\theta \; f(\mathbf{x}, \theta).$$

Likelihood score equations

Under regularity conditions, the MLE is the solution of

$$U(\widehat{\theta}, \mathbf{x}) = 0$$

where $U(\theta, \mathbf{x}) = \frac{\partial}{\partial \theta} \log f(\mathbf{x}, \theta)$ is the *score function*.

Likelihood score equations

Under regularity conditions, the MLE is the solution of

$$U(\widehat{\theta}, \mathbf{x}) = 0$$

where $U(\theta, \mathbf{x}) = \frac{\partial}{\partial \theta} \log f(\mathbf{x}, \theta)$ is the *score function*.
The score satisfies

$$\mathbb{E}_\theta \left[U(\theta, X) \right] = 0$$

MLE for exponential family

The likelihoods are typically in regular *exponential family* form

$$f(\mathbf{x}, \theta) = \frac{\exp\{\theta^{\intercal} V(\mathbf{x})\}}{Z(\theta)}$$

where

$\mathbf{x}$	is the data,
θ	the parameter,
$V(\mathbf{x})$	a statistic (easily computable),
$Z(\theta)$	the normalising constant.

MLE for exponential family

The score is

$$U(\mathbf{x}, \theta) = V(\mathbf{x}) - \mathbb{E}_\theta[V(X)]$$

so the MLE is the solution $\widehat{\theta}$ of

$$\mathbb{E}_\theta[V(X)] = V(\mathbf{x})$$

MLE for exponential family

Problem: $Z(\theta)$ is intractable

$\Rightarrow$ moments $\mathbb{E}_\theta[V(X)]$ are intractable

$\Rightarrow$ maximum likelihood is analytically intractable

$\Rightarrow$ finite sample behaviour is poorly understood

Pseudolikelihood

Besag (1974, 1977)

Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a vector of discrete random variables. Define the pseudolikelihood

$$\mathsf{PL}\left(\theta, \mathbf{x}\right) = \prod_{i=1}^{n} \mathbb{P}_\theta \left\{ X_i = x_i \mid \left(X_j = x_j,\ j \neq i\right) \right\}$$

the product of the conditional likelihoods of each X_i given the other variables. Estimate θ by

$$\tilde{\theta} = \mathsf{argmax}_{\theta \in \Theta} \mathsf{PL}\left(\theta, \mathbf{x}\right).$$

Pseudolikelihood for point process

Besag (1978)

Suppose $\mathbf{X}$ is a spatial point process with Papangelou conditional intensity $\Lambda_\theta(u, \mathbf{X})$. Define the pseudolikelihood

$$\mathsf{PL}\left(\theta, \mathbf{x}\right) = \left[\prod_{i=1}^{n} \Lambda_\theta(x_i, \mathbf{x})\right] \exp\left(-\int_D \Lambda(u, \mathbf{x})\, \mathrm{d}u\right).$$

Again estimate θ by

$$\tilde{\theta} = \mathsf{argmax}_{\theta \in \Theta} \mathsf{PL}\left(\theta, \mathbf{x}\right).$$

Maximum pseudolikelihood

The MPLE for a Gibbs model is typically

- easy to compute
- consistent and asymptotically normal
- biased and inefficient in small samples

MCMC methods

MCMC maximum likelihood inference is elegant, statistically efficient, but requires large computing resources.

Pseudolikelihood is cheap to compute, but is statistically inefficient.

We need something in between. . .

Pseudolikelihood estimating equations

Suppose $\mathbf{X} = (X_1, \ldots, X_N)$ has regular exponential family distribution with density

$$f(\mathbf{x}; \theta) = \frac{\exp\{\theta^{\mathsf{T}} V(\mathbf{x})\}}{Z(\theta)}.$$

Then, writing X_{-i} for $(X_j,\ j \neq i)$,

$$\mathbb{P}_\theta \left\{ X_i = x_i \mid X_{-i} = \mathbf{x}_{-i} \right\} = \frac{\exp\{\theta^{\mathsf{T}} V(\mathbf{x})\}}{\sum_{\mathbf{x}'} \exp\{\theta^{\mathsf{T}} V(\mathbf{x}')\}}$$

where $\mathbf{x}'$ runs over all states satisfying $\mathbf{x}'_{-i} = \mathbf{x}_{-i}$.

Pseudolikelihood

We find easily

$$\frac{\partial}{\partial\theta}\log \mathsf{PL}\left(\theta, \mathbf{x}\right) = \sum_{i=1}^{n}\left(V(x) - \mathbb{E}_\theta\left[V(X) \mid X_{-i} = \mathbf{x}_{-i}\right]\right)$$

so $\tilde{\theta}$ is the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta\left[V(X) \mid X_{-i} = \mathbf{x}_{-i}\right] = V(x)$$

This is an unbiased estimating equation.

Estimating equations (reminder)

Consider a parameter estimator $\widehat{\theta}$ which is the solution of an *estimating equation*

$$g(\mathbf{x}, \theta) = 0$$

for a given function g, where $\mathbf{x}$ is the data.

Estimating equations (reminder)

Consider a parameter estimator $\widehat{\theta}$ which is the solution of an *estimating equation*

$$g(\mathbf{x}, \theta) = 0$$

for a given function g, where $\mathbf{x}$ is the data.
We call g *unbiased* if, for all θ,

$$\mathbb{E}_\theta \left[g(X, \theta) \right] = 0$$

Estimating equations (reminder)

Consider a parameter estimator $\widehat{\theta}$ which is the solution of an *estimating equation*

$$g(\mathbf{x}, \theta) = 0$$

for a given function g, where $\mathbf{x}$ is the data.

We call g *unbiased* if, for all θ,

$$\mathbb{E}_\theta \left[g(X, \theta) \right] = 0$$

This does not mean the estimator $\widehat{\theta}$ is unbiased, but it will usually be consistent.

Estimating equations (reminder)

Example 1: The MLE is the solution of the estimating equation

$$U(\mathbf{x};\theta) = 0$$

where U is the score.

Estimating equations (reminder)

Example 1: The MLE is the solution of the estimating equation

$$U(\mathbf{x};\theta) = 0$$

where U is the score.

The score is an unbiased estimating function:

$$\mathbb{E}_\theta[U(X;\theta)] = 0.$$

Estimating equations (reminder)

Example 2: The method-of-moments estimator is the value of θ satisfying $g(\mathbf{x}, \theta) = 0$ where

$$g(\mathbf{x}, \theta) = S(\mathbf{x}) - \mathbb{E}_\theta[S(X)]$$

where $S(\mathbf{x})$ is the observed value of a chosen statistic S, and $\mathbb{E}_\theta[S(X)]$ is the population moment.

Estimating equations (reminder)

Example 2: The method-of-moments estimator is the value of θ satisfying $g(\mathbf{x}, \theta) = 0$ where

$$g(\mathbf{x}, \theta) = S(\mathbf{x}) - \mathbb{E}_\theta[S(X)]$$

where $S(\mathbf{x})$ is the observed value of a chosen statistic S, and $\mathbb{E}_\theta[S(X)]$ is the population moment.

The method-of-moments estimating function is an unbiased estimating function:

$$\mathbb{E}_\theta G(\mathbf{x}, \theta) = \mathbb{E}_\theta[S(\mathbf{X}) - \mathbb{E}_\theta[S(X)]] = 0.$$

Estimating equations

Godambe-Heyde theory of estimating functions generalises Fisher-Cramér asymptotic theory of MLE, making it possible to

- calculate asymptotic variance of estimators
- choose the optimal estimating function within a class
- construct an associated ‘quasilikelihood’

Estimating equations

Godambe-Heyde theory of estimating functions generalises Fisher-Cramér asymptotic theory of MLE, making it possible to

- calculate asymptotic variance of estimators
- choose the optimal estimating function within a class
- construct an associated ‘quasilikelihood’

It doesn’t tell us *how* to

- find a class of unbiased estimating functions
- trade off statistical efficiency against computational cost.

Equilibrium principle

Standard fact: if $\mathbf{Y}$ is a stationary Markov process with equilibrium distribution π, and if X has distribution π, then

$$\mathbb{E}\left[(\mathcal{A}\,S)(X)\right] = 0$$

for essentially all statistics $S = S(X)$,
where $\mathcal{A}$ is the *generator* of $\mathbf{Y}$.

Equilibrium principle

Discrete case

Define generator $\mathcal{A}$ by

$$(\mathcal{A}\,S)(\mathbf{x}) \;=\; \mathbb{E}[S(Y_{n+1}) - S(Y_n) \;\mid\; Y_n = \mathbf{x}]$$

Equilibrium principle

Discrete case

Define generator $\mathcal{A}$ by

$$\begin{aligned}(\mathcal{A}\,S)(\mathbf{x}) &= \mathbb{E}[S(Y_{n+1}) - S(Y_n) \mid Y_n = \mathbf{x}] \\ &= \mathbb{E}[S(Y_{n+1}) \mid Y_n = \mathbf{x}] - S(\mathbf{x})\end{aligned}$$

Equilibrium principle

Discrete case

Define generator $\mathcal{A}$ by

$$
\begin{aligned}
(\mathcal{A}\,S)(\mathbf{x}) &= \mathbb{E}[S(Y_{n+1}) - S(Y_n) \mid Y_n = \mathbf{x}] \\
&= \mathbb{E}[S(Y_{n+1}) \mid Y_n = \mathbf{x}] - S(\mathbf{x}) \\
&= \sum_{y \in \mathcal{X}} (S(\mathbf{y}) - S(\mathbf{x}))P(\mathbf{x}, \mathbf{y})
\end{aligned}
$$

Equilibrium principle

Then

$$\begin{aligned}
\mathbb{E}_\pi[(\mathcal{A}\,S)(X)] &= \mathbb{E}_\pi[S(Y_{n+1}) - S(Y_n)] \\
&= \mathbb{E}_\pi[S(Y_{n+1})] - \mathbb{E}_\pi[S(Y_n)] \\
&= 0.
\end{aligned}$$

Equilibrium principle

Time-invariance estimating equations

Baddeley (1995, 2000), Kessler & Sørensen (1995), Hansen & Scheinkman (1995)

Find a Markov chain sampler $\mathbf{Y}^{(\theta)} = (Y_t^{(\theta)},\ t \geq 0)$ for the distribution of X under θ.

Let $\mathcal{A}_\theta$ be the generator of $\mathbf{Y}^{(\theta)}$. Then

$$\mathbb{E}_\theta\left[(\mathcal{A}_\theta\, S)(X)\right] = 0$$

for "any" statistic S.

Equilibrium principle

Choose a statistic S.

Given the observed data $\mathbf{x}$, solve for θ in

$$(\mathcal{A}_\theta S)(\mathbf{x}) = 0.$$

This is an unbiased estimating equation for θ.

Solution $\hat{\theta}_T$ = *equilibrium estimator*.

Equilibrium principle

Example 1

Take any parametric model for X.

Let $Y_1^{(\theta)}, Y_2^{(\theta)}, \ldots$ be **i.i.d. realisations** of X under θ. Then $\mathbf{Y}$ has discrete time generator

$$(\mathcal{A}_\theta\, S)(\mathbf{x}) = \mathbb{E}_\theta S(X) - S(\mathbf{x})$$

Equilibrium principle

Example 1

Take any parametric model for X.

Let $Y_1^{(\theta)}, Y_2^{(\theta)}, \ldots$ be **i.i.d. realisations** of X under θ. Then $\mathbf{Y}$ has discrete time generator

$$(\mathcal{A}_\theta\ S)(\mathbf{x}) = \mathbb{E}_\theta S(X) - S(\mathbf{x})$$

Then $\widehat{\theta}$ is the solution of

$$\mathbb{E}_\theta S(X) = S(\mathbf{x})$$

i.e. the method-of-moments estimator.
(In regular exponential families this gives the MLE)

Equilibrium principle

Example 2

Finite system of r.v.'s $\mathbf{X} = (X_1, \ldots, X_M)$.

Let $\mathbf{Y}$ be the single-site **Gibbs sampler**.

Generator:

$$(\mathcal{A}_\theta S)(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_\theta[S(X) \mid X_{-i} = \mathbf{x}_{-i}] - S(\mathbf{x})$$

Equilibrium principle

Example 2

Finite system of r.v.'s $\mathbf{X} = (X_1, \ldots, X_M)$.

Let $\mathbf{Y}$ be the single-site **Gibbs sampler**.

Generator:

$$(\mathcal{A}_\theta S)(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_\theta[S(X) \mid X_{-i} = \mathbf{x}_{-i}] - S(\mathbf{x})$$

Equilibrium equation $(\mathcal{A}_\theta S)(\mathbf{x}) = 0$ is **equivalent** to

maximum pseudolikelihood normal equations

if we choose $S \equiv V$ in exponential family $f(\mathbf{x}; \theta) \propto \exp(\theta^\mathsf{T} V(\mathbf{x}))$.

Equilibrium principle

Example 3

Point process $\mathbf{X}$ with conditional intensity $\lambda(u, \mathbf{X})$.

Let $\mathbf{Y}$ be the associated spatial birth-and-death process with unit death rate.

Generator is

$$\begin{aligned}(\mathcal{A}_\theta S)(\mathbf{x}) &= \int_W [S(\mathbf{x} \cup \{u\}) - S(\mathbf{x})]\lambda(u, \mathbf{x})\, \mathrm{d}u \\ &\quad + \sum_i [S(\mathbf{x} \setminus \{x_i\}) - S(\mathbf{x})]\end{aligned}$$

Equilibrium principle

Example 3

Point process $\mathbf{X}$ with conditional intensity $\lambda(u, \mathbf{X})$.

Let $\mathbf{Y}$ be the associated spatial birth-and-death process with unit death rate.

Generator is

$$\begin{aligned}(\mathcal{A}_\theta S)(\mathbf{x}) &= \int_W [S(\mathbf{x} \cup \{u\}) - S(\mathbf{x})]\lambda(u, \mathbf{x})\, \mathrm{d}u \\ &\quad + \sum_i [S(\mathbf{x} \setminus \{x_i\}) - S(\mathbf{x})]\end{aligned}$$

Time-invariance equation $(\mathcal{A}_\theta S)(\mathbf{x}) = 0$ is **equivalent** to

maximum pseudolikelihood normal equations

if we choose $S \equiv V$ in exponential family $f(\mathbf{x}; \theta) \propto \exp(\theta^\mathsf{T} V(\mathbf{x}))$.

Equilibrium principle

Familiar examples

X	$\mathbf{Y}$	$S(\mathbf{x})$	$\hat{\theta}_T$
any	i.i.d.	any	method of moments
any	i.i.d.	$V(\mathbf{x})$	MLE
finite system of r.v.'s	1-site Gibbs sampler	$V(\mathbf{x})$	max. pseudolikelihood
Markov random field	1-site Gibbs sampler	$V(\mathbf{x})$	max. pseudolikelihood
spatial point process	spatial birth-death	any	Takacs-Fiksel estimator
spatial point process	spatial birth-death	$V(\mathbf{x})$	spatial MPLE
random censoring	i.i.d. (lifetimes)	e.d.f.	reduced sample estimator

Equilibrium principle

Unfamiliar/new estimators

X	$\mathbf{Y}$	$S(\mathbf{x})$	$\hat{\theta}_T$
continuous r.v.'s	Ornstein-Uhlenbeck diffusion	any	variational estimator
Markov random field	block-update Gibbs sampler	$V(\mathbf{x})$	generalised pseudolik.
spatial point process	multiple birth-death	$V(\mathbf{x})$	generalised pseudolik.
spatial point process	Barker dynamics	$V(\mathbf{x})$	logistic composite lik.
Markov random field	Metropolis-Hastings	$V(\mathbf{x})$	weird new estimator

Questions

What does it mean?

Questions

What does it mean?

- recipe for deriving estimators

Questions

What does it mean?

- recipe for deriving estimators
- common structure for different techniques of parameter estimation
  (MLE, method of moments, MPLE, Takacs-Fiksel, minimum contrast, variational,. . .)

Questions

What does it mean?

- recipe for deriving estimators
- common structure for different techniques of parameter estimation
  (MLE, method of moments, MPLE, Takacs-Fiksel, minimum contrast, variational,. . .)
- algebraic form of estimating equations corresponds to transition structure of $\mathbf{Y}$

Questions

What does it mean?

- recipe for deriving estimators
- common structure for different techniques of parameter estimation (MLE, method of moments, MPLE, Takacs-Fiksel, minimum contrast, variational,. . .)
- algebraic form of estimating equations corresponds to transition structure of $\mathbf{Y}$
- when recipe is applied to ‘simple’ MCMC samplers $\mathbf{Y}$ it yields ‘simple’ *ad hoc* estimators

References

- A. Baddeley.
  Time-invariance estimating equations.
  Bernoulli **6** (2000) 783–808.
- A. Baddeley and D. Dereudre.
  Variational estimators for the parameters of Gibbs point process models.
  Bernoulli **19** (2013) 905–930.
- A. Baddeley, J.-F. Coeurjolly, E. Rubak and R. Waagepetersen
  Logistic Regression for Spatial Gibbs Point Processes.
  Biometrika **101** (2014) 377–392.

Adrian.Baddeley@curtin.edu.au