Tractable Nonparametric Bayesian Inference in Poisson Processes

Ryan Prescott Adams

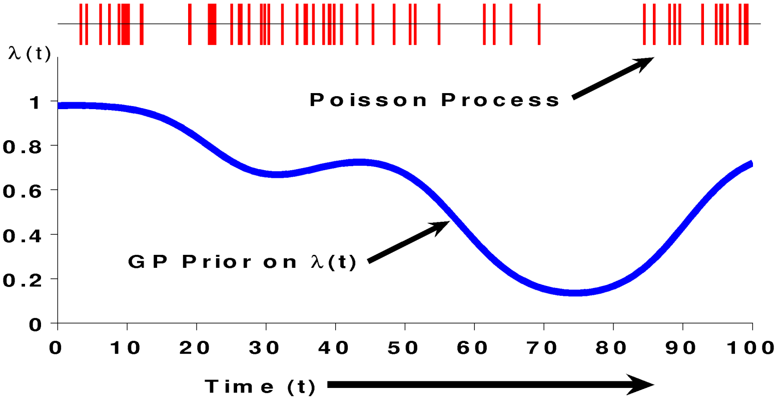
Cavendish Laboratory
University of Cambridge
http://www.inference.phy.cam.ac.uk/rpa23/

11 June 2009

Joint work with Iain Murray and David MacKay

The Main Idea

- Observe a set of events $\mathcal{S} = \{t_k\}_{k=1}^{K}$
- Could be in time or space.
- Model as Poisson with intensity $\lambda(t)$.
- Use a Gaussian process prior on $\lambda(t)$.



Outline

The Poisson Process

The Gaussian Process

The Sigmoidal Gaussian Cox Process

Inference with the SGCP

Extensions

Summary

Outline

The Poisson Process

The Gaussian Process

The Sigmoidal Gaussian Cox Process

Inference with the SGCP

Extensions

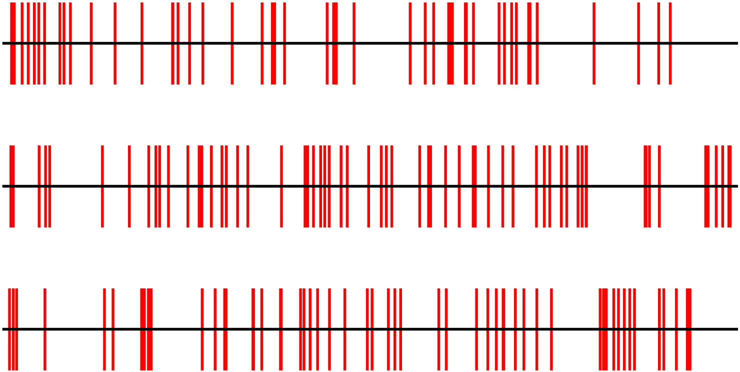
Summary

Point Processes

Random countable subsets of some domain.

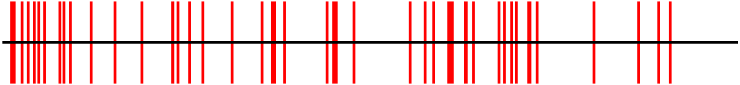
- I'll assume some bounded subset $\mathcal{V}$ of $\mathbb{R}^D$.

"A random locally-finite counting measure."



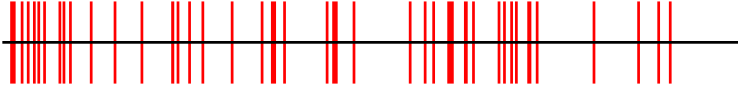
Point Processes

Temporal



Point Processes

Temporal

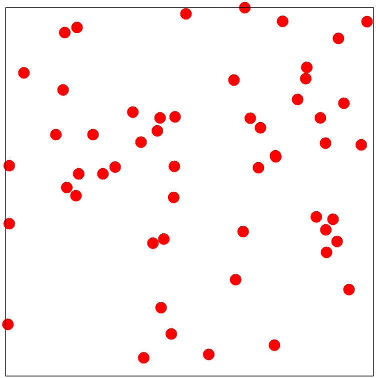


Point Processes

Spatial

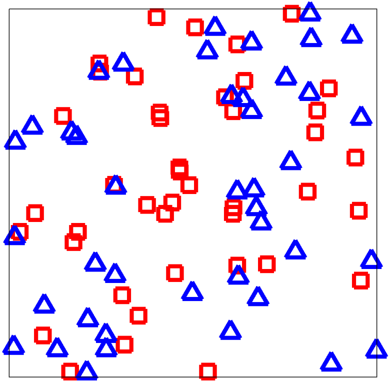
Point Processes

Spatial



Point Processes

Marked



Point Processes

Zappo's Spatiotemporal Marked

http://www.zappos.com/map/

Point Processes, continued

Applications Are Everywhere

- Temporal: neural spikes, credit defaults, bus arrivals, terrorist attacks, ...
- Spatial: galaxies, forests, cities, ...

Different From Density Modeling

- The number of data matter.
- The data are not generally i.i.d.

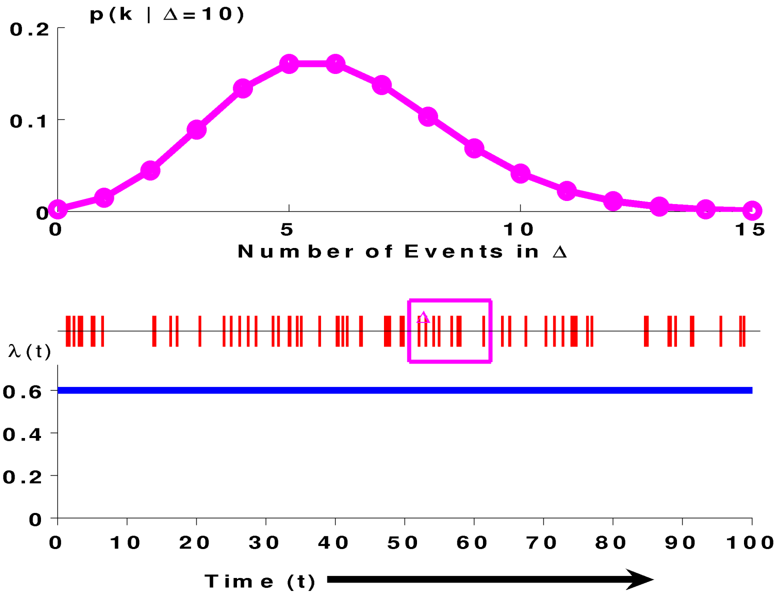
Homogeneous Poisson Process

- Your basic vanilla point process.
- Has a constant intensity $\lambda > 0$.
  - Expected number of events per unit time.
- An interval of length Δ_t has a Poisson-distributed number of events:

$$p(N(\Delta_t) = k \mid \lambda) = \frac{(\lambda \Delta_t)^k}{k!} \exp\{-\lambda \Delta_t\}$$

- Disjoint intervals are independent.
- The time between arrivals is exponentially-distributed with parameter λ.
- For a spatial setting, generalize "interval."

Homogeneous Poisson Process



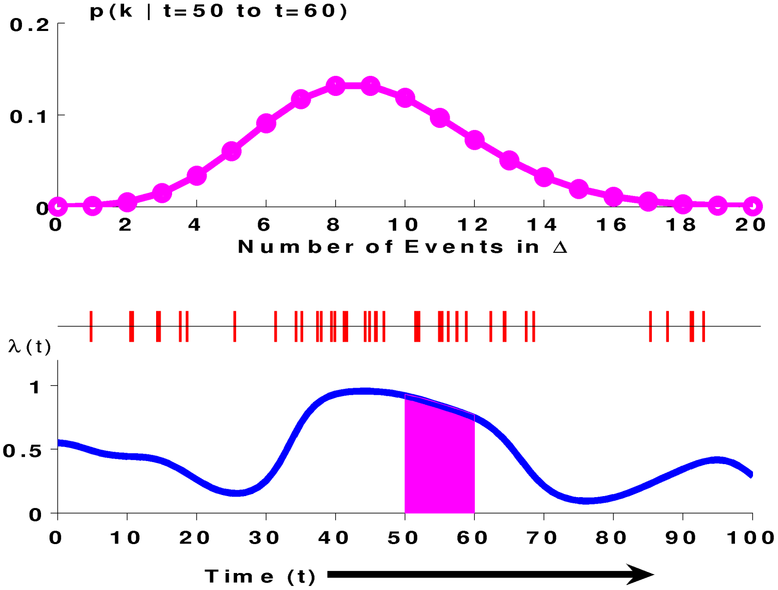
Inhomogeneous Poisson Process

- Intensity is a function of time: $\lambda(t) \geq 0$.
- Disjoint intervals are still independent.
- The number of events in some interval t_1 to t_2 is Poisson distributed with parameter

$$\lambda_{t_1,t_2} = \int_{t_1}^{t_2} \mathrm{d}t \; \lambda(t)$$

- In a spatial setting, generalize the region of integration.

Inhomogeneous Poisson Process



How to Model the Intensity Function?

- Inference in the homogeneous case is easy.
- Varying λ is much more interesting!

Nonparametric Prior on $\lambda(t)$

- What if we don't know much about $\lambda(t)$?
- Get $\lambda(t)$ from a stochastic process.
- Now called a **Cox Process**
- With a GP: **Gaussian Cox Process**

The Big Picture

This talk is about constructing a Gaussian Cox Process that allows tractable inference.

Outline

The Poisson Process

The Gaussian Process

The Sigmoidal Gaussian Cox Process

Inference with the SGCP

Extensions

Summary

Why Gaussian Processes?

- A **nonparametric prior on functions**.
- Popular for Bayesian nonlinear regression and classification.
- Components in larger Bayesian models.

GP Ingredients

- Input space $\mathcal{X}$ (e.g. $\mathbb{R}^D$)
- Output space $\mathcal{Y} = \mathbb{R}$
- Positive-definite covariance function: $K(x, x'; \theta) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- Mean function $m(x; \theta) : \mathcal{X} \to \mathcal{Y}$

Why Gaussian Processes?

GP Mechanics

- Data are input-output pairs: $\mathcal{D} = \{x_n, y_n\}_{n=1}^{N}$
- Turn the inputs into a covariance matrix.
- Use the covariance matrix to construct a joint Gaussian distribution on the outputs.

Useful Because Tractable

- Predictive distributions are Gaussian.
- The marginal likelihood has closed form.

Marginalisation of Gaussians

The Key Ingredient for GPs

- Joint Gaussian $P(y_1, \ldots, y_N) = \mathcal{N}(\boldsymbol{K})$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} & k_{1,4} & k_{1,5} \\ k_{2,1} & k_{2,2} & k_{2,3} & k_{2,4} & k_{2,5} \\ k_{3,1} & k_{3,2} & k_{3,3} & k_{3,4} & k_{3,5} \\ k_{4,1} & k_{4,2} & k_{4,3} & k_{4,4} & k_{4,5} \\ k_{5,1} & k_{5,2} & k_{5,3} & k_{5,4} & k_{5,5} \end{bmatrix} \right)$$

Marginalisation of Gaussians

The Key Ingredient for GPs

- Joint Gaussian $P(y_1, \ldots, y_N) = \mathcal{N}(\boldsymbol{K})$
- Marginal covariance is the submatrix.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} & k_{1,4} & k_{1,5} \\ k_{2,1} & k_{2,2} & k_{2,3} & k_{2,4} & k_{2,5} \\ k_{3,1} & k_{3,2} & k_{3,3} & k_{3,4} & k_{3,5} \\ k_{4,1} & k_{4,2} & k_{4,3} & k_{4,4} & k_{4,5} \\ k_{5,1} & k_{5,2} & k_{5,3} & k_{5,4} & k_{5,5} \end{bmatrix} \right)$$

Marginalisation of Gaussians

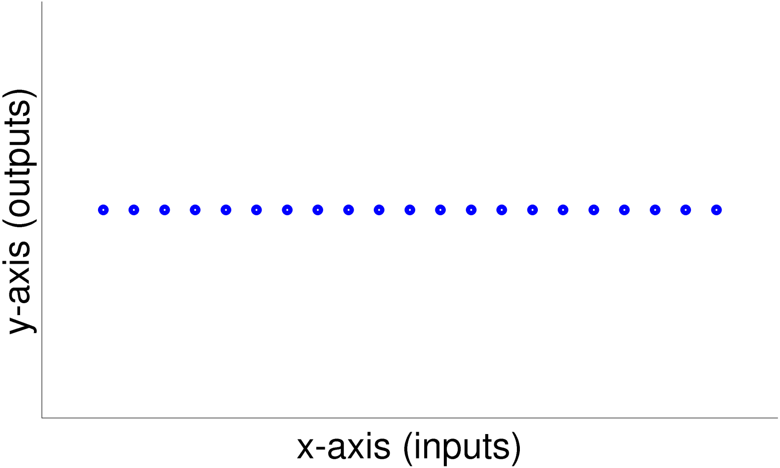
The Key Ingredient for GPs

- Joint Gaussian $P(y_1, \ldots, y_N) = \mathcal{N}(\boldsymbol{K})$
- Marginal covariance is the submatrix.
- Conditional is also Gaussian.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} & k_{1,4} & k_{1,5} \\ k_{2,1} & k_{2,2} & k_{2,3} & k_{2,4} & k_{2,5} \\ k_{3,1} & k_{3,2} & k_{3,3} & k_{3,4} & k_{3,5} \\ k_{4,1} & k_{4,2} & k_{4,3} & k_{4,4} & k_{4,5} \\ k_{5,1} & k_{5,2} & k_{5,3} & k_{5,4} & k_{5,5} \end{bmatrix} \right)$$

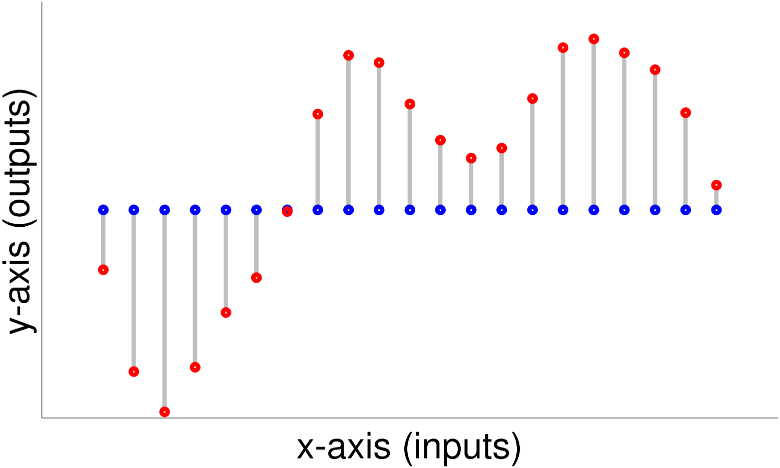
Why Gaussian Processes?

Nearby inputs have covarying outputs.



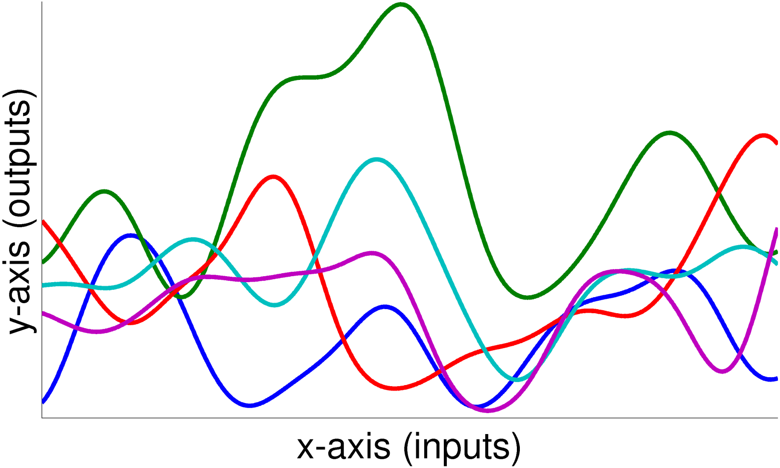
Why Gaussian Processes?

Nearby inputs have covarying outputs.



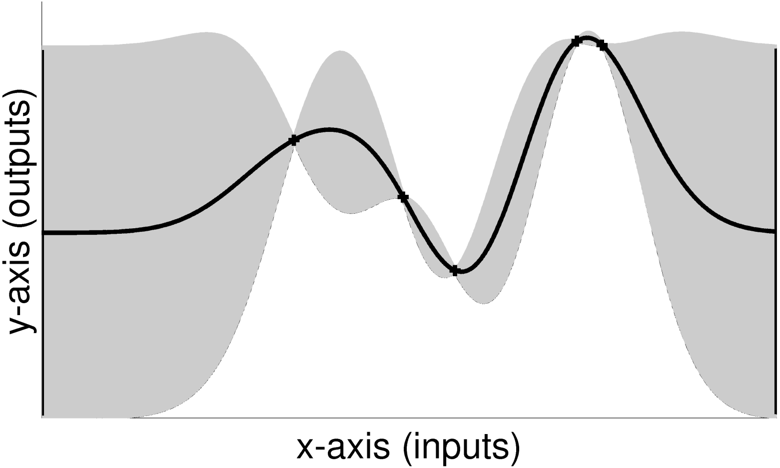
Why Gaussian Processes?

Nearby inputs have covarying outputs.

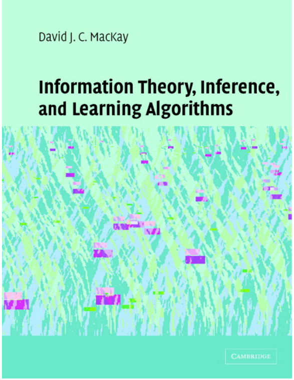


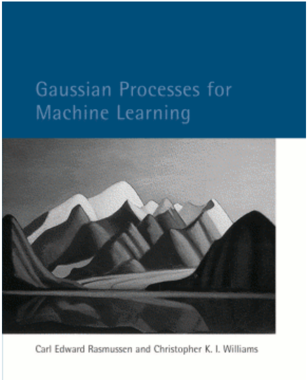
Why Gaussian Processes?

Nearby inputs have covarying outputs.



Recommended Reading





MacKay

Rasmussen and Williams

Both are free online!

The Log Gaussian Cox Process

- GPs give nonparametric priors on functions.
- Use them for priors on Poisson intensities!

Do the Natural Thing

- Exponentiate the draw from the GP:

$$g(t) \sim \mathcal{GP}(t, \theta)$$
$$\lambda(t) = \exp\{g(t)\}$$

- Rathbun and Cressie, 1994
- Jesper Møller and colleagues, 1998

The Log Gaussian Cox Process

Two Flies in the Ointment

1. $g(t)$ is infinite-dimensional.
2. The posterior is **doubly-intractable**.
    - Likelihood only known to within a constant.
    - Example: undirected graphical models

Likelihood of events $\{t_k\}_{k=1}^{K}$ between 0 and T:

$$p(\{t_k\}_{k=1}^{K} \,|\, g(t) = \boldsymbol{g}) = \exp\left\{ -\int_0^T \exp\{g(t)\}\,\mathrm{d}t + \sum_{k=1}^{K} g(t_k) \right\}$$

The Log Gaussian Cox Process

Two Flies in the Ointment

1. $g(t)$ is infinite-dimensional.
2. The posterior is **doubly-intractable**.
    - Likelihood only known to within a constant.
    - Example: undirected graphical models

Likelihood of events $\{t_k\}_{k=1}^K$ between 0 and T:

$$p(\{t_k\}_{k=1}^K \,|\, g(t) = \boldsymbol{g}) = \exp\left\{ -\int_0^T \exp\{g(t)\}\, \mathrm{d}t + \sum_{k=1}^K g(t_k) \right\}$$

Doubly-Intractable Inference

- Intractable **marginal likelihoods** are common in interesting Bayesian models.
- *This* difficulty depends on the parameters.
- Even Markov chain Monte Carlo is hard.

Good News

- Recent MCMC methods address inference in doubly-intractable models.
- Møller et al., 2004, Murray et al., 2006
- **The Catch:** You have to be able to generate *exact* data from the model.

Outline

The Poisson Process

The Gaussian Process

The Sigmoidal Gaussian Cox Process

Inference with the SGCP

Extensions

Summary

A Different Prior on Intensities

Like before, $g(t)$ is a draw from a GP. Transform:

$$\lambda(t) = \sigma(g(t))\,\bar{\lambda}(t)$$

The function $\sigma(\cdot)$ is a sigmoid, like:

$$\sigma(z) = \frac{1}{1 + \exp\{-z\}}$$

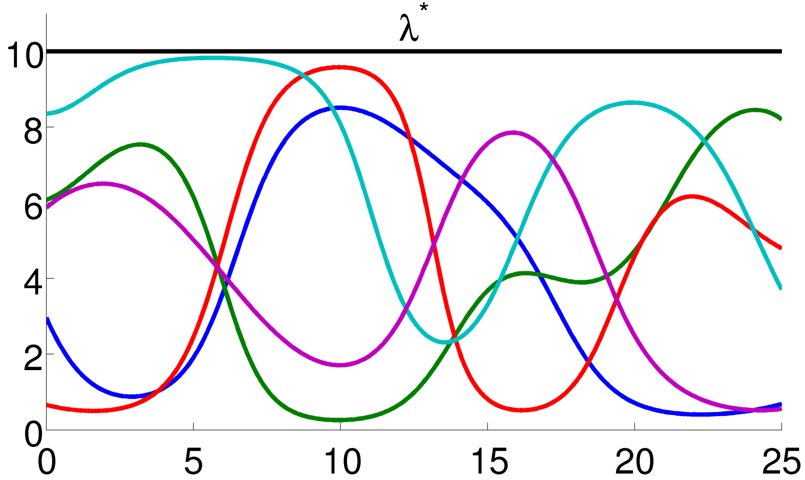
The *dominating intensity* $\bar{\lambda}(t)$ is something simple, like a constant function:

$$\bar{\lambda}(t) = \lambda^{\star}$$

We call this the **sigmoidal Gaussian Cox process** (SGCP).

Realizations from the SGCP Prior

Random positive functions beneath $\bar{\lambda}(t)$.



Generating Data from the SGCP

We can generate **exact Poisson data** from a random intensity drawn from this prior.

First, Two Useful Algorithms

- Simulating homogeneous Poisson data.
- How to perform **thinning** of Poisson data.

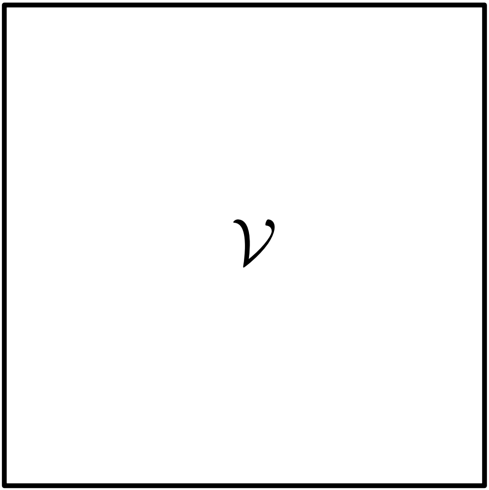
Simulate homogeneous Poisson data

Assuming intensity λ on region $\mathcal{V}$:

1. Find the measure of $\mathcal{V}$, i.e. $\mu(\mathcal{V})$.
2. Sample the number of events:
   $N(\mathcal{V}) \sim \mathcal{P}\mathcal{O}(\lambda\, \mu(\mathcal{V}))$
3. Distribute the $N(\mathcal{V})$ points independently and uniformly on $\mathcal{V}$.

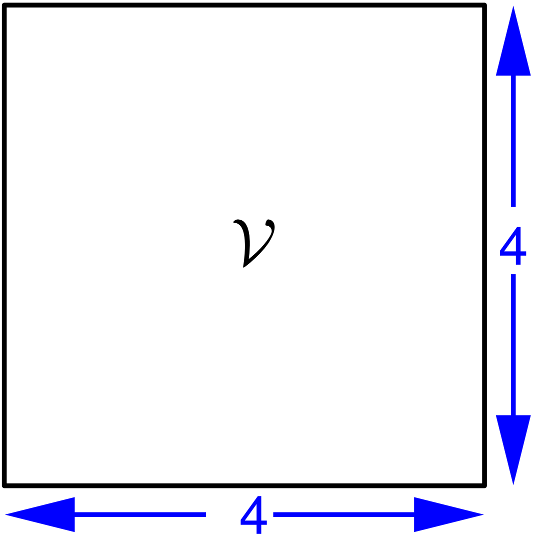
Simulate homogeneous Poisson data

Step 0: Region $\mathcal{V}$, constant intensity $\lambda^\star = 2$.



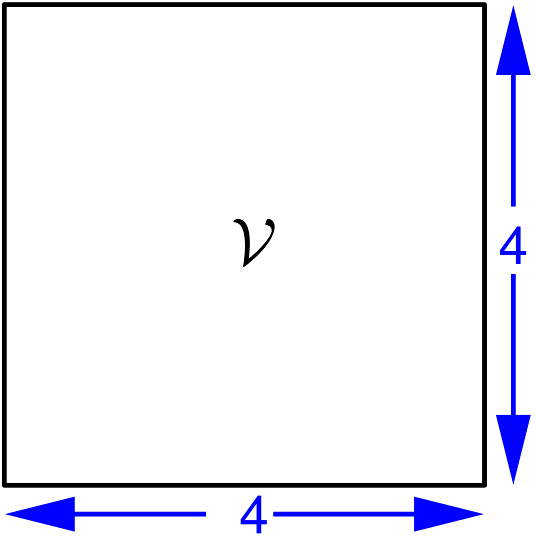
Simulate homogeneous Poisson data

Step 1: Get the volume of $\mathcal{V}$ . . .



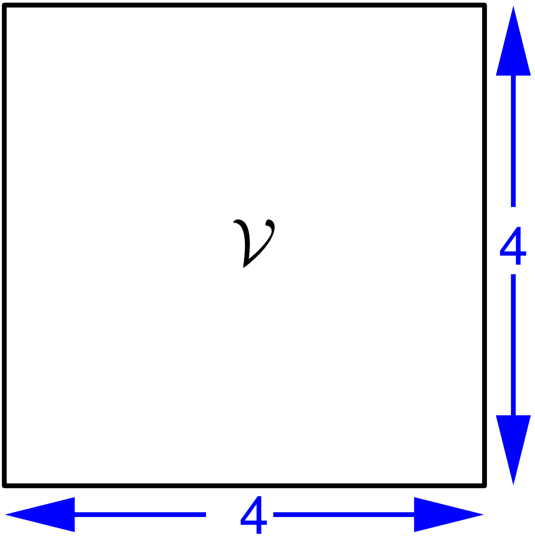
Simulate homogeneous Poisson data

Step 1: Get the volume of $\mathcal{V}$ ... $\mu(\mathcal{V}) = 4 \times 4 = 16$



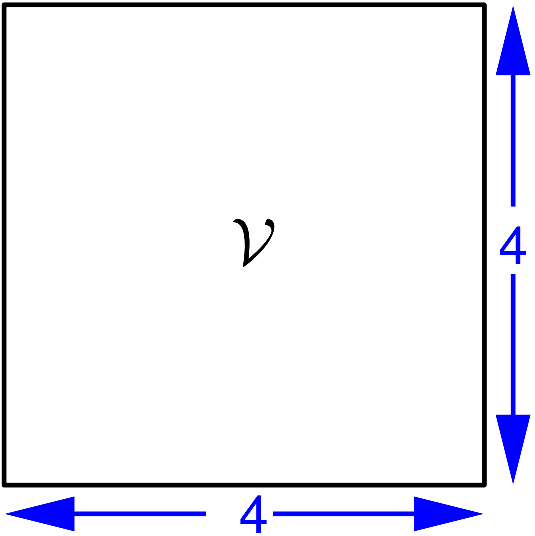
Simulate homogeneous Poisson data

Step 2: Draw the Poisson number of events K.



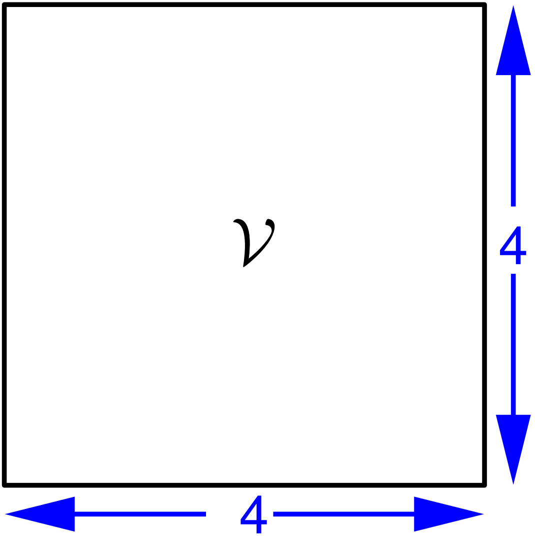
Simulate homogeneous Poisson data

Step 2: $K \sim \mathcal{P}o(\lambda^{\star}\mu(\mathcal{V}) = 2 \times 4 \times 4) \ldots$



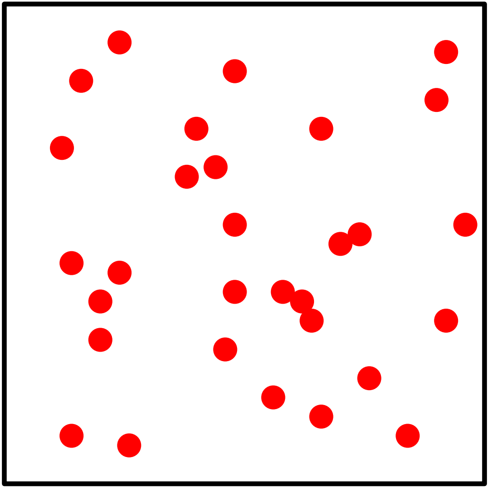
Simulate homogeneous Poisson data

Step 2: $K \sim \mathcal{P}o(\lambda^{\star}\mu(\mathcal{V}) = 2 \times 4 \times 4) \ldots$ 30



Simulate homogeneous Poisson data

Step 3: Distributed the K events uniformly in $\mathcal{V}$.



Independent Thinning

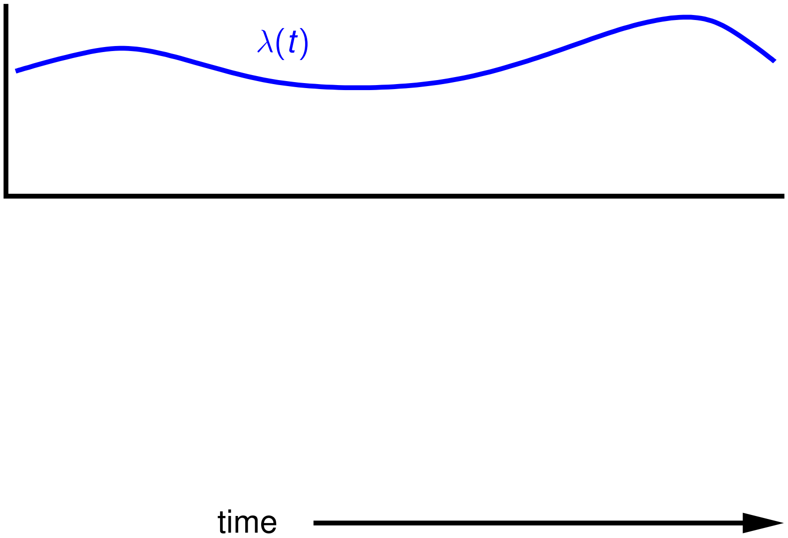
Due to Lewis and Shedler, 1979 – For some function $\phi : \mathbb{R}^D \rightarrow [0, 1]$:

1. Get some Poisson data $\{t_k\}_{k=1}^K$ from $\lambda(t)$.
2. Remove t_k with coin flip probability $1 - \phi(t_k)$.
3. The remaining events are Poisson with intensity $\phi(t)\lambda(t)$.

This is very similar to rejection sampling.

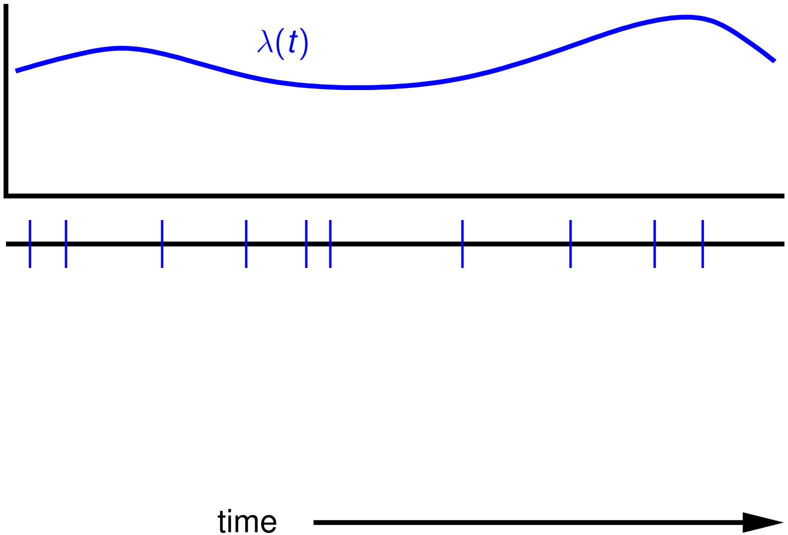
Independent Thinning

Step 1: Intensity function $\lambda(t)$.



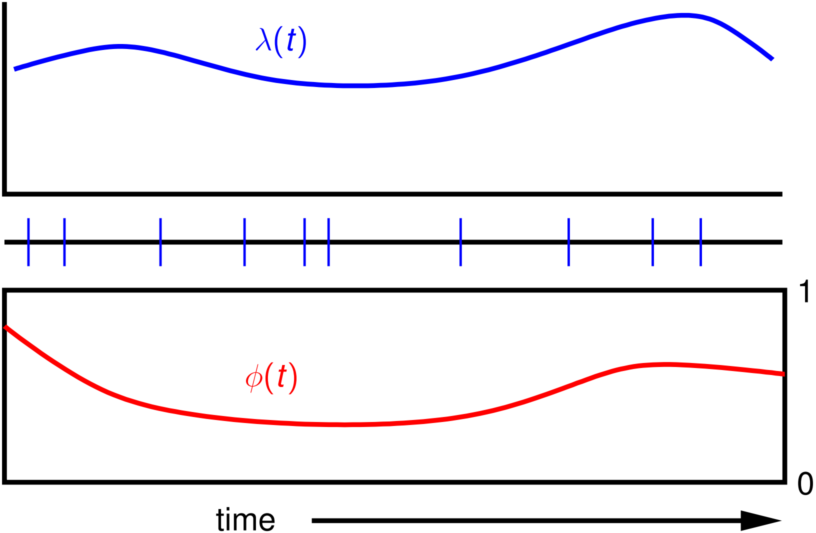
Independent Thinning

Step 2: Get some events from $\lambda(t)$.



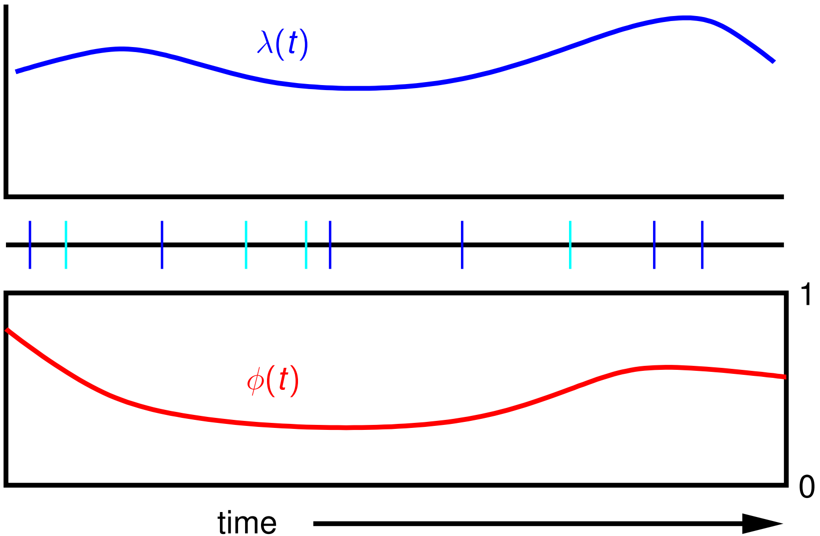
Independent Thinning

Step 3: Another function $\phi(t) : \mathcal{X} \rightarrow [0, 1]$.



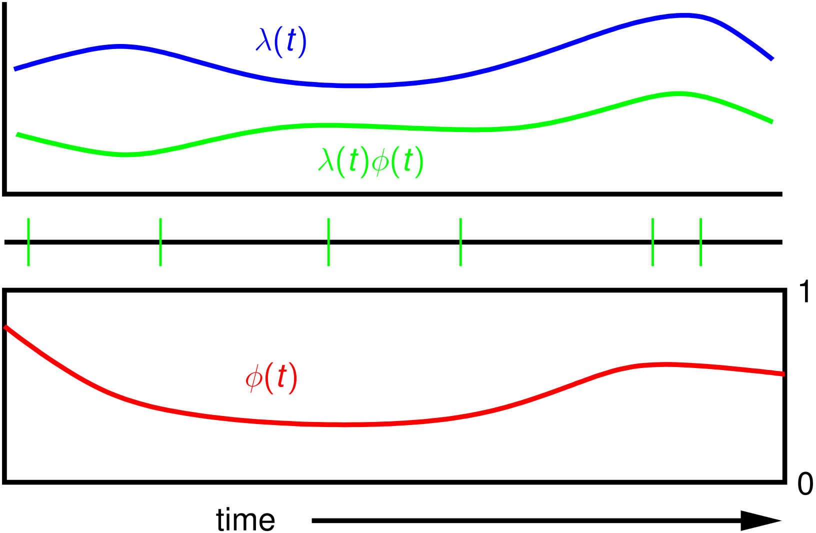
Independent Thinning

Step 4: Delete event t_k with probability $1 - \phi(t_k)$.



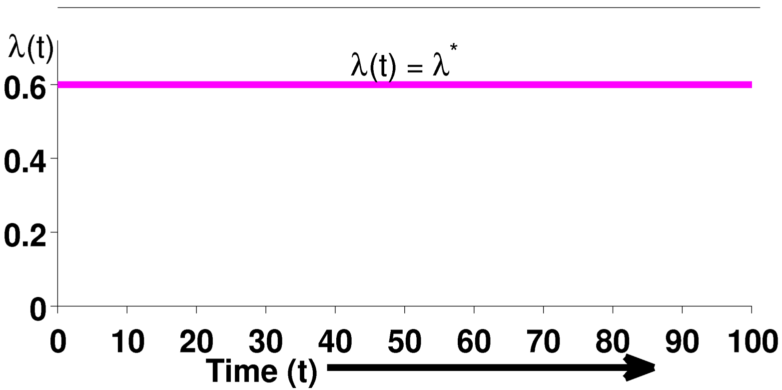
Independent Thinning

Step 5: Remaining events are from $\lambda(t)\phi(t)$.



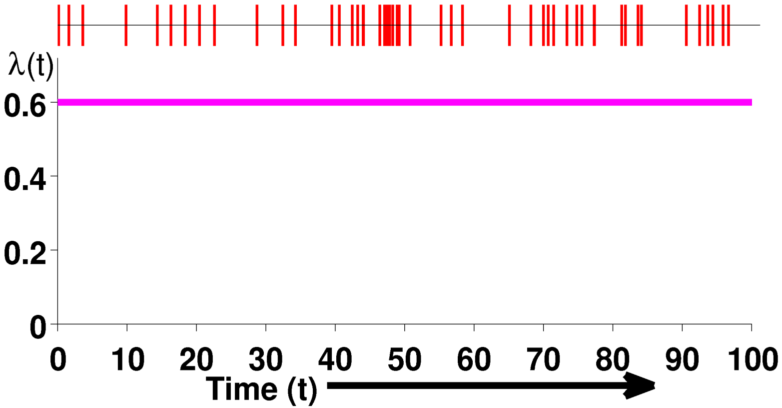
Back to the SGCP

1. Generate Poisson data from $\lambda(t) = \lambda^\star$.
2. Draw a sample from the GP at the events.
3. Thin events according to the GP draw.



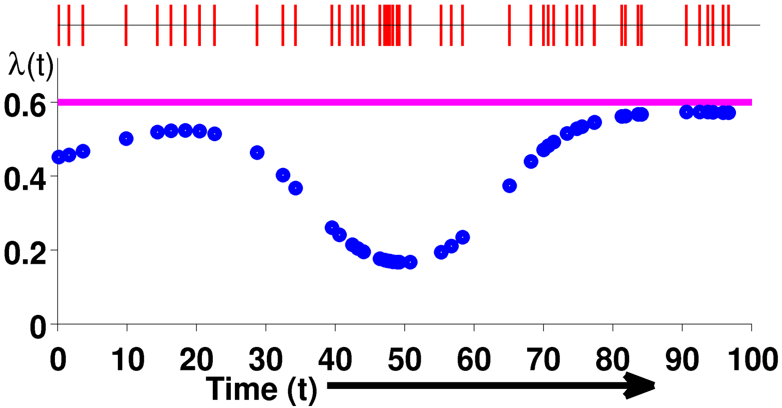
Back to the SGCP

1. Generate Poisson data from $\lambda(t) = \lambda^{\star}$.
2. Draw a sample from the GP at the events.
3. Thin events according to the GP draw.



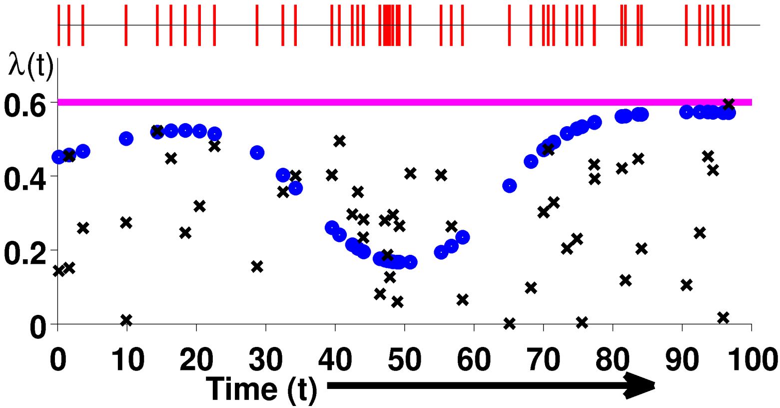
Back to the SGCP

1. Generate Poisson data from $\lambda(t) = \lambda^{\star}$.
2. Draw a sample from the GP at the events.
3. Thin events according to the GP draw.



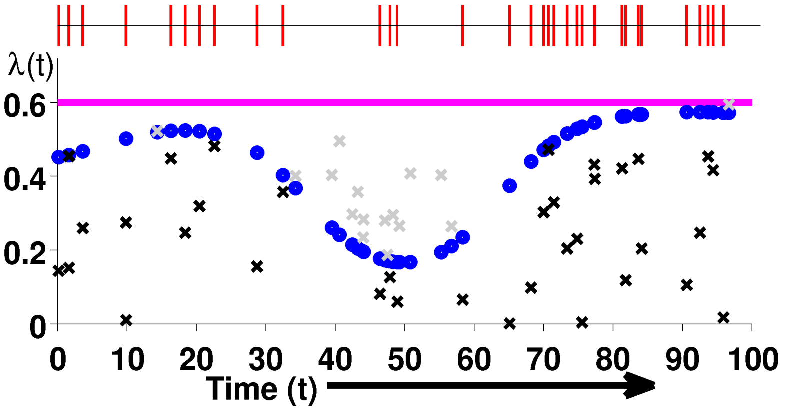
Back to the SGCP

1. Generate Poisson data from $\lambda(t) = \lambda^{\star}$.
2. Draw a sample from the GP at the events.
3. Thin events according to the GP draw.



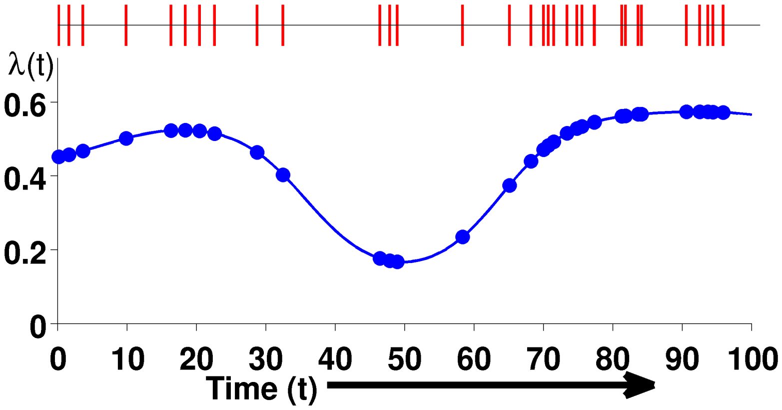
Back to the SGCP

1. Generate Poisson data from $\lambda(t) = \lambda^\star$.
2. Draw a sample from the GP at the events.
3. Thin events according to the GP draw.



Back to the SGCP

1. Generate Poisson data from $\lambda(t) = \lambda^{\star}$.
2. Draw a sample from the GP at the events.
3. Thin events according to the GP draw.



Properties of SGCP Generation

The data are **exactly drawn** from a Poisson process with a random intensity from the SGCP.

We did not have to discover the function at more than a **finite number** of locations.

We **did not have to integrate** the function.

Outline

The Poisson Process

The Gaussian Process

The Sigmoidal Gaussian Cox Process

Inference with the SGCP

Extensions

Summary

Inference with the SGCP

Given K events $\{t_k\}_{k=1}^K$ on $\mathcal{V}$, and the SGCP prior, what is the posterior distribution on $g(t)$?

Still doubly-intractable:

$$p(\{t_k\}_{k=1}^K \,|\, g(t), \lambda^\star) = \exp\left\{-\int_0^T \sigma(g(t))\, \lambda^\star \,\mathrm{d}t\right\} \prod_{k=1}^K \sigma(g(t_k))\, \bar{\lambda}(t_k)$$

Inference Via the Latent History

Augment the state with the “latent history” of the generative procedure. Assume there were M thinned events $\{s_m\}_{m=1}^{M}$ and write the joint distribution of everything:

$$p(\{t_k\}_{k=1}^{K}, \{s_m\}_{m=1}^{M}, \boldsymbol{g} \,|\, \lambda^\star, \theta) =$$

$$(\lambda^\star)^{K+M} \exp\{-\lambda^\star \mu(\mathcal{V})\} \prod_{k=1}^{K} \sigma(g(t_k)) \prod_{m=1}^{M} \sigma(-g(s_m))$$

$$\times\, \mathcal{GP}(\{g(t_k)\}, \{g(s_m)\} \,|\, \theta)$$

Ugly, but not intractable!

Inference Via the Latent History

- Homogeneous Poisson process

$$p(\{t_k\}_{k=1}^{K}, \{s_m\}_{m=1}^{M}, \boldsymbol{g} \mid \lambda^\star, \theta) =$$

$$(\lambda^\star)^{K+M} \exp\{-\lambda^\star \mu(\mathcal{V})\}$$

$$\times \prod_{k=1}^{K} \sigma(g(t_k))$$

$$\times \prod_{m=1}^{M} \sigma(-g(s_m))$$

$$\times \mathcal{GP}(\{g(t_k)\}, \{g(s_m)\} \mid \theta)$$

Inference Via the Latent History

- Homogeneous Poisson process
- Probability of unthinned events

$$p(\{t_k\}_{k=1}^{K}, \{s_m\}_{m=1}^{M}, \boldsymbol{g} \mid \lambda^\star, \theta) =$$

$$(\lambda^\star)^{K+M} \exp\{-\lambda^\star \mu(\mathcal{V})\}$$

$$\times \prod_{k=1}^{K} \sigma(g(t_k))$$

$$\times \prod_{m=1}^{M} \sigma(-g(s_m))$$

$$\times \mathcal{GP}(\{g(t_k)\}, \{g(s_m)\} \mid \theta)$$

Inference Via the Latent History

- Homogeneous Poisson process
- Probability of unthinned events
- Probability of thinned events

$$p(\{t_k\}_{k=1}^{K}, \{s_m\}_{m=1}^{M}, \boldsymbol{g} \mid \lambda^\star, \theta) =$$

$$(\lambda^\star)^{K+M} \exp\{-\lambda^\star \mu(\mathcal{V})\}$$

$$\times \prod_{k=1}^{K} \sigma(g(t_k))$$

$$\times \prod_{m=1}^{M} \sigma(-g(s_m))$$

$$\times \mathcal{GP}(\{g(t_k)\}, \{g(s_m)\} \mid \theta)$$

Inference Via the Latent History

- Homogeneous Poisson process
- Probability of unthinned events
- Probability of thinned events
- Gaussian process prior

$$p(\{t_k\}_{k=1}^{K}, \{s_m\}_{m=1}^{M}, \boldsymbol{g} \,|\, \lambda^\star, \theta) =$$

$$(\lambda^\star)^{K+M} \exp\{-\lambda^\star \mu(\mathcal{V})\}$$

$$\times \prod_{k=1}^{K} \sigma(g(t_k))$$

$$\times \prod_{m=1}^{M} \sigma(-g(s_m))$$

$$\times \mathcal{GP}(\{g(t_k)\}, \{g(s_m)\} \,|\, \theta)$$

Overview of the MCMC Sampler

We update each part of the latent state separately, conditioned on the others using a Gibbs-like procedure.

- Insert and remove latent thinned events via Metropolis–Hastings
- Move latent thinned events around via Metropolis–Hastings
- Sample the latent function via Hamiltonian Monte Carlo.

Also: hyperparameters of the GP, and $\lambda^\star$.

Inserting/Removing Latent Events

Birth Proposal

- Also propose a location $\hat{s}$ uniformly in $\mathcal{V}$.
- Draw $g(\hat{s})$ conditionally from the GP.

$$a_{ins} = \frac{\mu(\mathcal{V})\,\lambda^{\star}}{(M+1)}\sigma(-g(\hat{s}))$$

Death Proposal

- Pick one of the M at random.

$$a_{del} = \frac{M}{\mu(\mathcal{V})\,\lambda^{\star}}\sigma(-g(s_m))^{-1}$$

Moving Latent Events Around

- Iterate over each of the M events.
- Use a proposal distribution $q(\hat{s} \leftarrow s)$.
- Draw $g(\hat{s})$ conditionall from th GP.
- Accept with M-H ratio:

$$a_{loc} = \frac{q(s_m \leftarrow \hat{s}_m)\,\sigma(-g(\hat{s}_m))}{q(\hat{s}_m \leftarrow s_m)\,\sigma(-g(s_m))}$$

Updating the Latent Function

- The GP prior enforces a lot of structure.
- Use Hamiltonian Monte Carlo for efficiency.
- Uses gradients to reduce random walk behavior.

$$
\begin{aligned}
p(\boldsymbol{g} \,|\, \{t_k\}_{k=1}^{K}, \{s_m\}_{m=1}^{M}, \lambda^\star, \theta) \propto \\
\mathcal{GP}(\{g(t_k)\}, \{g(s_m)\} \,|\, \{t_k\}_{k=1}^{K}, \{s_m\}_{m=1}^{M}, \theta) \\
\times \prod_{k=1}^{K} \sigma(g(t_k)) \prod_{m=1}^{M} \sigma(-g(s_m))
\end{aligned}
$$

Updating Hyperparameters

GP Hyperparameters

Conditioned on the latent events and the latent function, just use the marginal likelihood.

Dominating Intensity Hyperparameters

Treat the union of observations and latent events as a parametric Poisson model. For the version with constant $\lambda^\star$, a gamma prior is conjugate:

$$\alpha = \alpha_0 + K + M$$
$$\beta = \beta_0 + \mu(\mathcal{V})$$

Empirical Evaluations

Synthetic Data

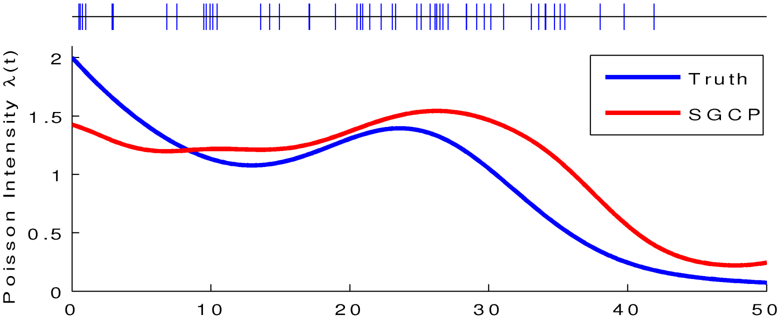
- Three known intensity functions.
- One training set, ten held-out test sets.
- Evaluated ℓ_2 norm and predictive logprob.
- Compared to kernel smoothing and LGCP.

Real-World Data

- Coal mining disasters in the UK, 1875-1962.
- Redwood forest data, scaled to unit square.

Synthetic Data Set 1

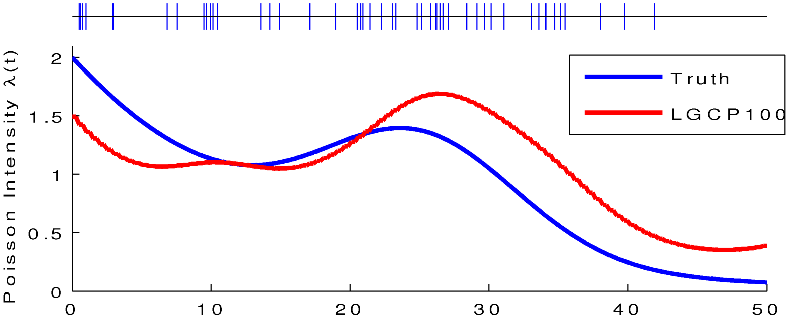
$$\lambda_1(s) = 2\exp\{-s/15\} + \exp\{-((s-25)/10)^2\} \text{ on } [0, 50]$$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_1(s)$	ℓ_2	**4.20**	6.65	5.96	6.12	5.44
	lp	**-45.11**	-46.41	-46.00	-46.80	-45.24

Synthetic Data Set 1

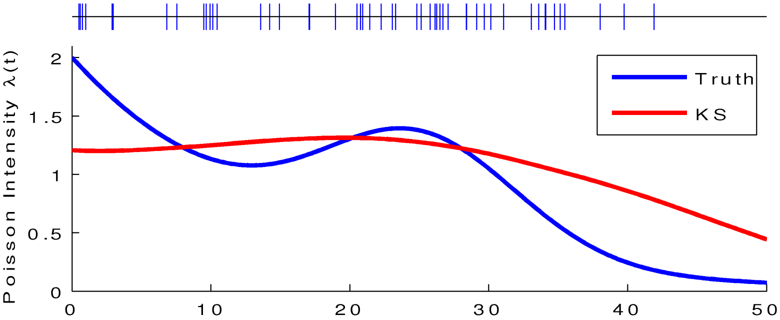
$$\lambda_1(s) = 2\exp\{-s/15\} + \exp\{-((s-25)/10)^2\} \text{ on } [0, 50]$$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_1(s)$	ℓ_2	**4.20**	6.65	5.96	6.12	5.44
	lp	**-45.11**	-46.41	-46.00	-46.80	-45.24

Synthetic Data Set 1

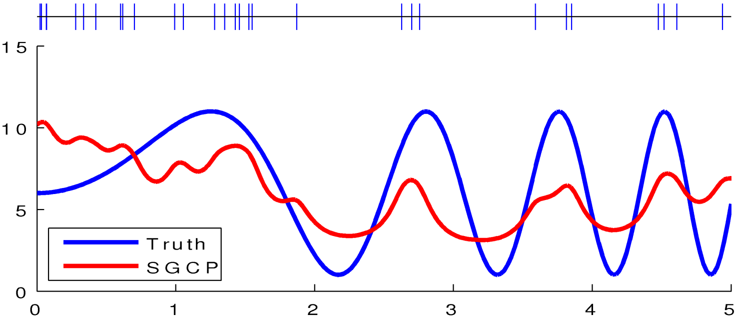
$$\lambda_1(s) = 2\exp\{-s/15\} + \exp\{-((s-25)/10)^2\} \text{ on } [0, 50]$$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_1(s)$	ℓ_2	**4.20**	6.65	5.96	6.12	5.44
	lp	**-45.11**	-46.41	-46.00	-46.80	-45.24

Synthetic Data Set 2

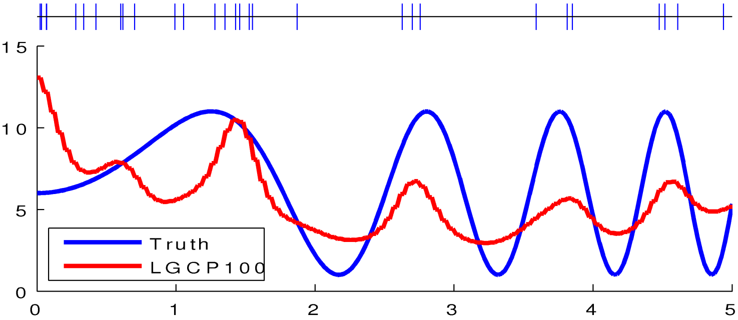
$$\lambda_2(s) = 5\sin(s^2)+6 \text{ on } [0,5]$$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_2(s)$	ℓ_2	**38.38**	73.71	70.34	53.27	43.51
	lp	24.45	**28.19**	23.36	22.89	25.29

Synthetic Data Set 2

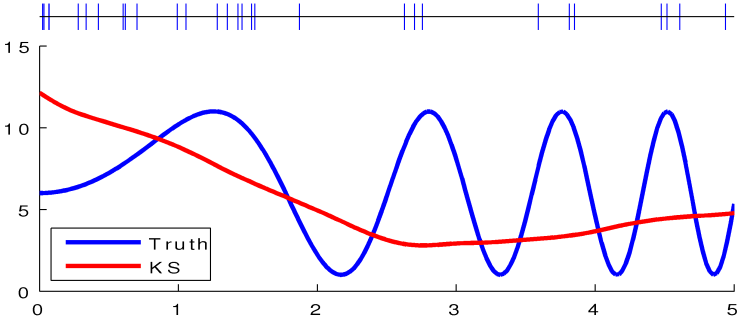
$$\lambda_2(s) = 5\sin(s^2)+6 \text{ on } [0,5]$$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_2(s)$	ℓ_2	**38.38**	73.71	70.34	53.27	43.51
	lp	24.45	**28.19**	23.36	22.89	25.29

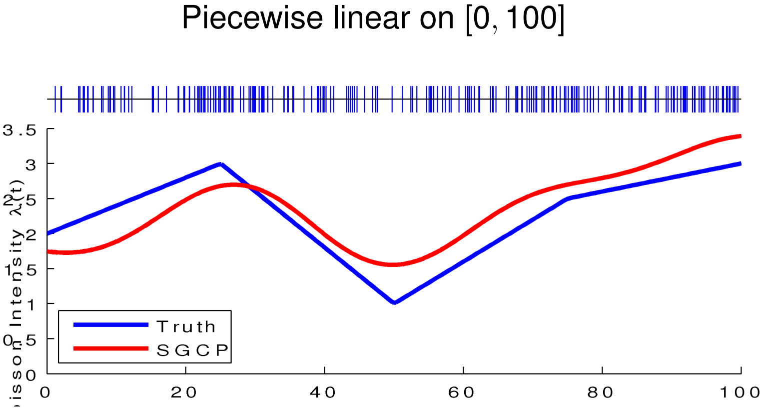
Synthetic Data Set 2

$$\lambda_2(s) = 5\sin(s^2) + 6 \text{ on } [0, 5]$$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_2(s)$	ℓ_2	**38.38**	73.71	70.34	53.27	43.51
	lp	24.45	**28.19**	23.36	22.89	25.29

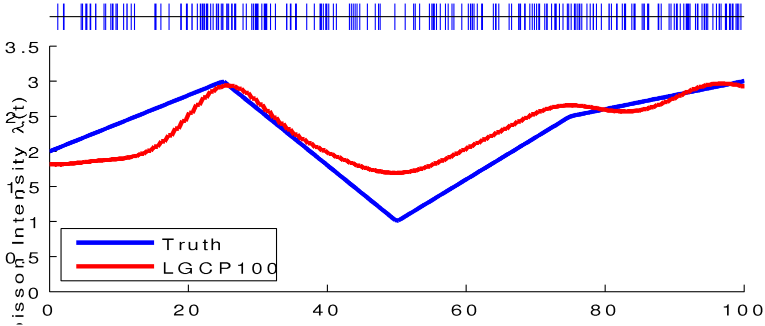
Synthetic Data Set 3



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_3(s)$	ℓ_2	11.41	30.56	90.76	22.14	**10.79**
	lp	**-43.39**	-46.47	-53.67	-52.31	-47.16

Synthetic Data Set 3

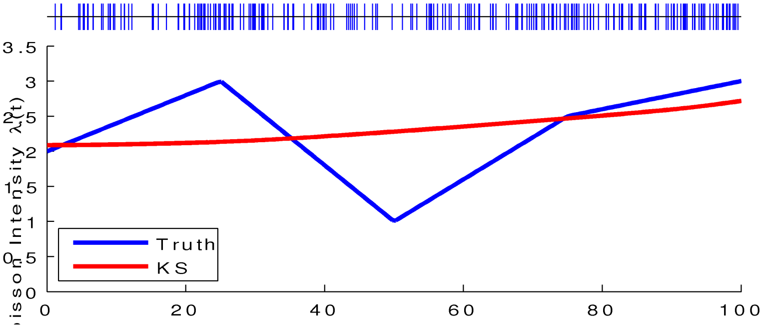
Piecewise linear on $[0, 100]$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_3(s)$	ℓ_2	11.41	30.56	90.76	22.14	**10.79**
	lp	**-43.39**	-46.47	-53.67	-52.31	-47.16

Synthetic Data Set 3

Piecewise linear on $[0, 100]$



		SGCP	KS	LGCP10	LGCP25	LGCP100
$\lambda_3(s)$	ℓ_2	11.41	30.56	90.76	22.14	**10.79**
	lp	**-43.39**	-46.47	-53.67	-52.31	-47.16

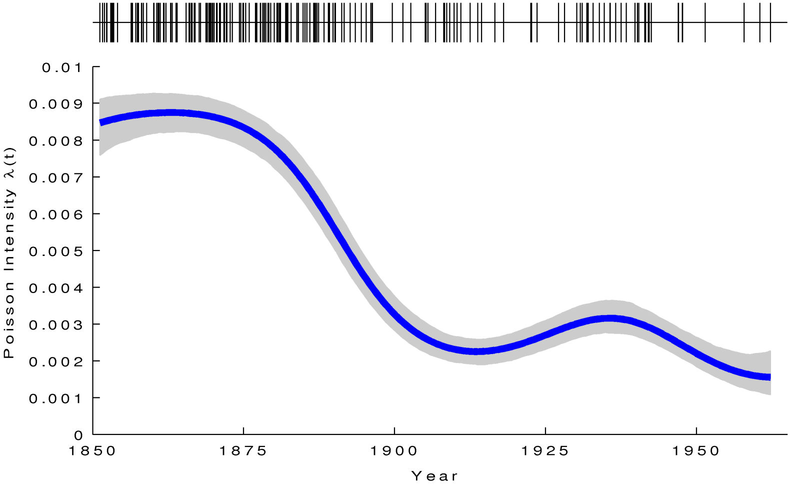
Coal Mining Disaster Data

- Coal mine disasters in UK between 1851 and 1962.
- 191 accidents.
- Commonly studied in changepoint models.
- Good example of a nonstationary Poisson process.

Only a “disaster” if ten or more people killed!

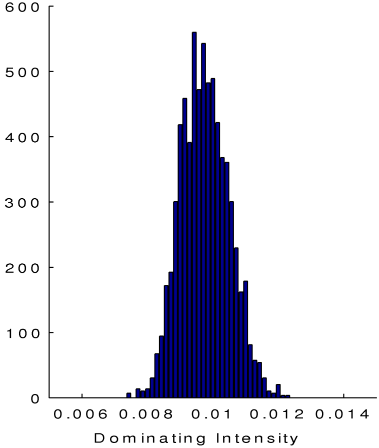
Coal Mining Disaster Data

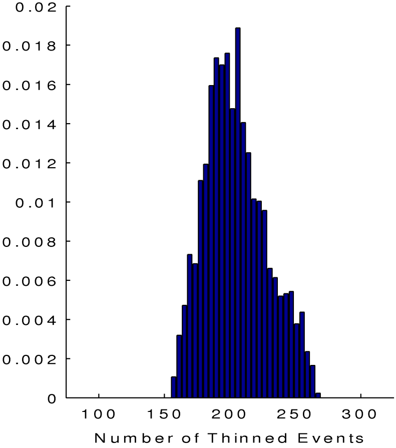
191 events between 15 March 1851 and 22 March 1962



Coal Mining Disaster Data

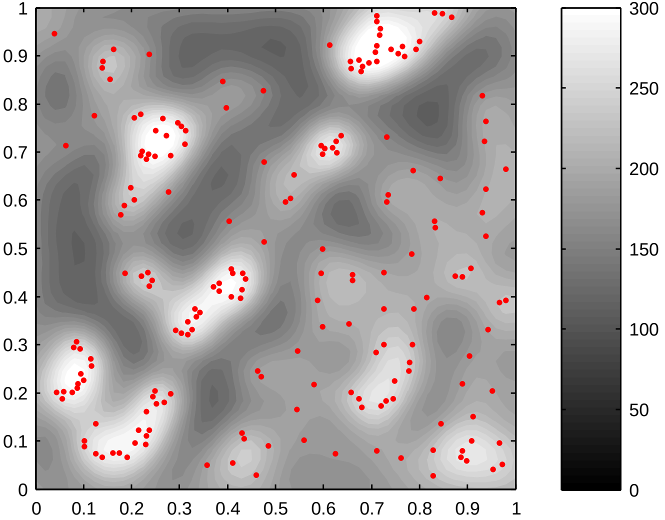
191 events between 15 March 1851 and 22 March 1962



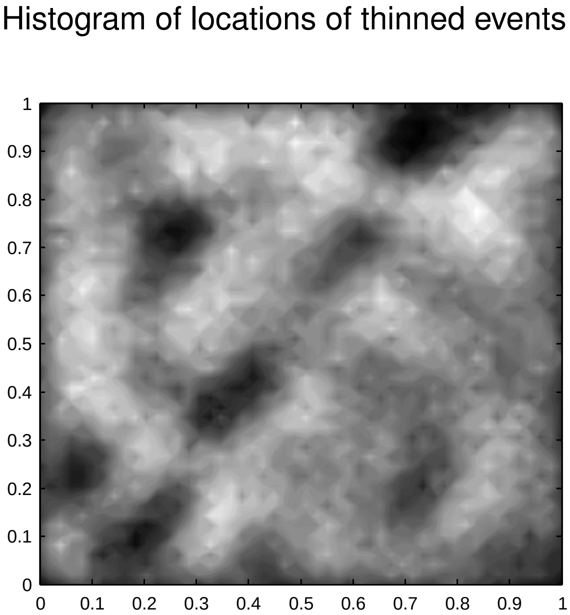


Redwoods

195 trees, scaled to the unit square



Redwoods


Histogram of locations of thinned events
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Outline

The Poisson Process

The Gaussian Process

The Sigmoidal Gaussian Cox Process

Inference with the SGCP

Extensions

Summary

Extended Point Processes

Marking

Additional random data associated with events.

- Example: random city locations each with random population

Interaction: Clustering

Points like to be closer together than Poisson.

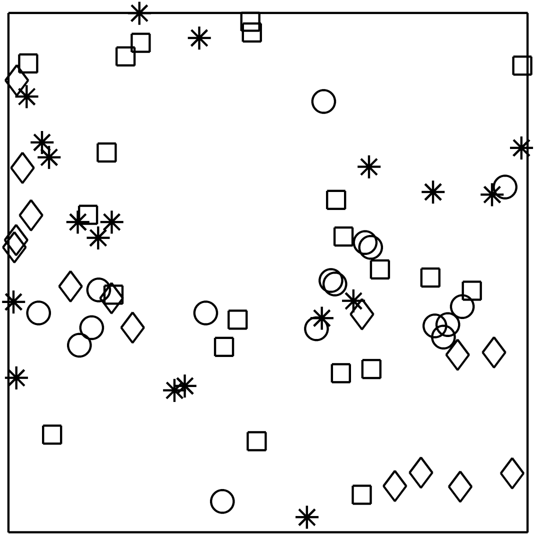
- Example: Plants germinate via seeds.

Interaction: Repulsion

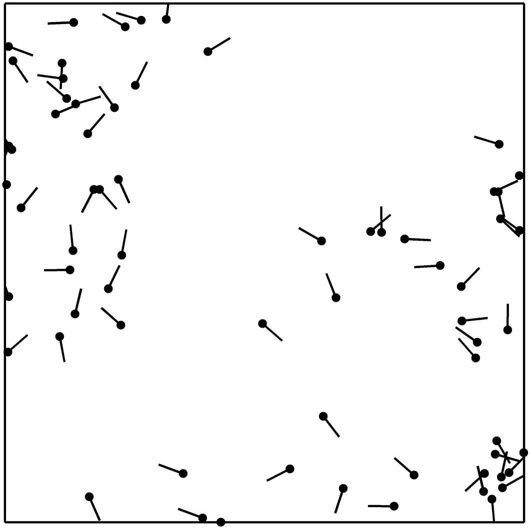
Points like to be farther apart than Poisson.

- Example: Neurons have refractory periods.

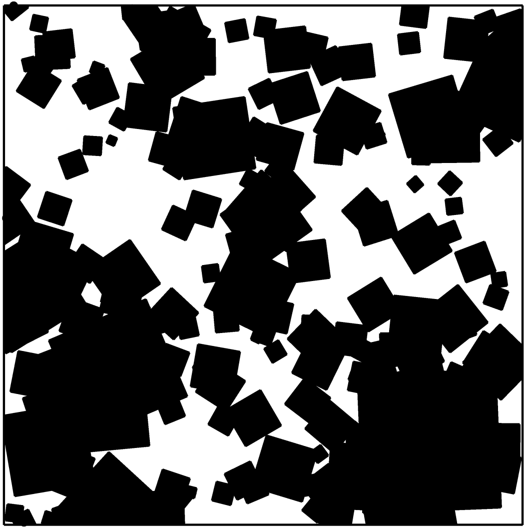
Marked Poisson Processes



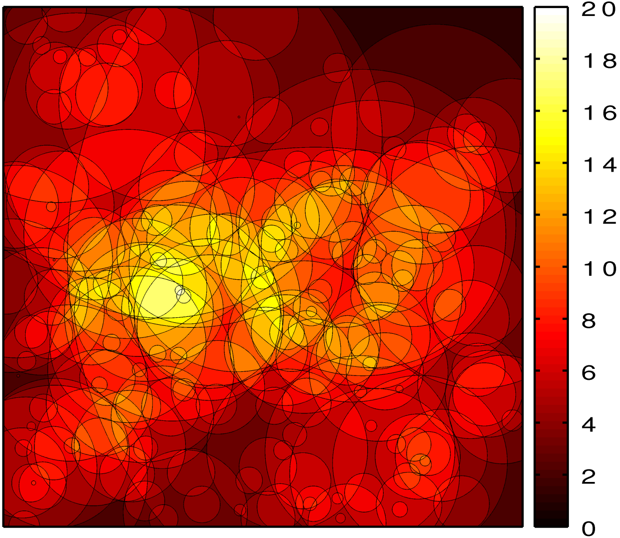
Marked Poisson Processes



The Boolean Model



The Boolean Model



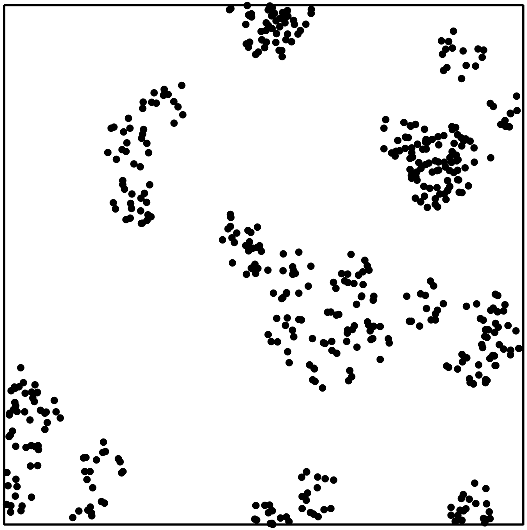
Interacting Point Processes

If the process is defined as a generative procedure, we can extend the SGCP directly to simulate data from it.

Contrast with general Gibbs/Markov point processes (e.g. Strauss process), where interaction is defined in terms of a potential function.

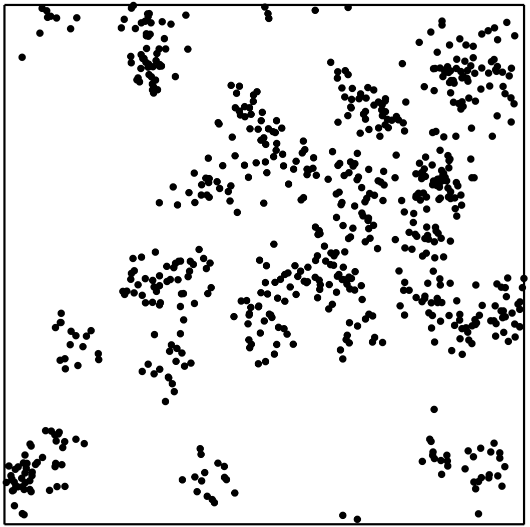
The Neyman–Scott Process

aka “The Matérn Cluster Process”

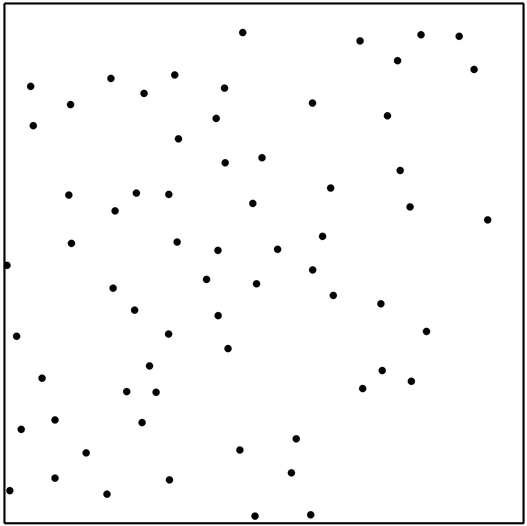


The Neyman–Scott Process

aka "The Thomas Process"



The Matérn Type III Process



Outline

The Poisson Process

The Gaussian Process

The Sigmoidal Gaussian Cox Process

Inference with the SGCP

Extensions

Summary

Summary

- The Poisson process is useful.
- NP Bayesian inference would be nice.
- GP priors on intensity functions have been intractable.
- Our construction uses a generative model to avoid intractability.
- The method is competitive in practice.
- The method can be extended to other point processes.
- **Bad News**: $O(N^3)$ scaling from GP.

Thanks

- Iain Murray (Toronto)
- David MacKay (Cambridge)
- Radford Neal (Toronto)
- Zoubin Ghahramani (Cambridge/CMU)
- Maneesh Sahani (Gatsby)

Funded by the Gates Cambridge Trust.