Practical and Consistent Nonparametric Inference in Gaussian Cox Processes

Ryan Prescott Adams

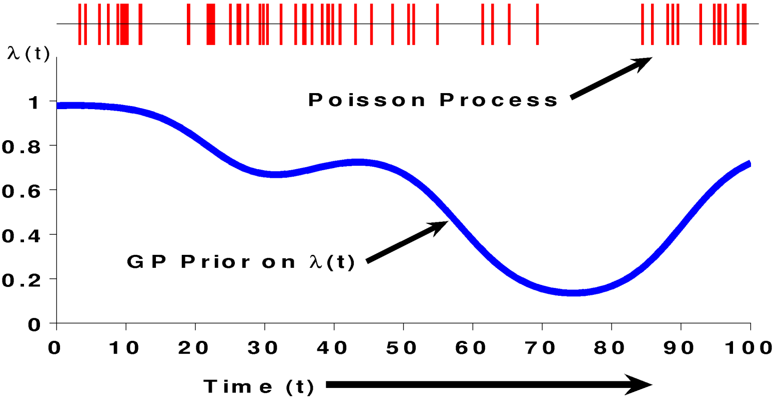
Cavendish Laboratory
University of Cambridge
http://www.inference.phy.cam.ac.uk/rpa23/

CIfAR Summer School

Joint work with Iain Murray and David MacKay

The Main Idea of the Talk

- Observe a sequence of events:
  $0 < t_1 < t_2 < t_3 < \cdots$
- Model as Poisson with rate $\lambda(t)$.
- Use a Gaussian process prior on $\lambda(t)$.



Outline

Outline

Modeling with Poisson Processes

GPs for Inhomogeneous Poisson Processes

- Gaussian Processes
- The Log-Gaussian Cox Process
- A Generative Cox Process
- Latent History Inference
- Coal Mine Disaster Data

Future Directions

Summary

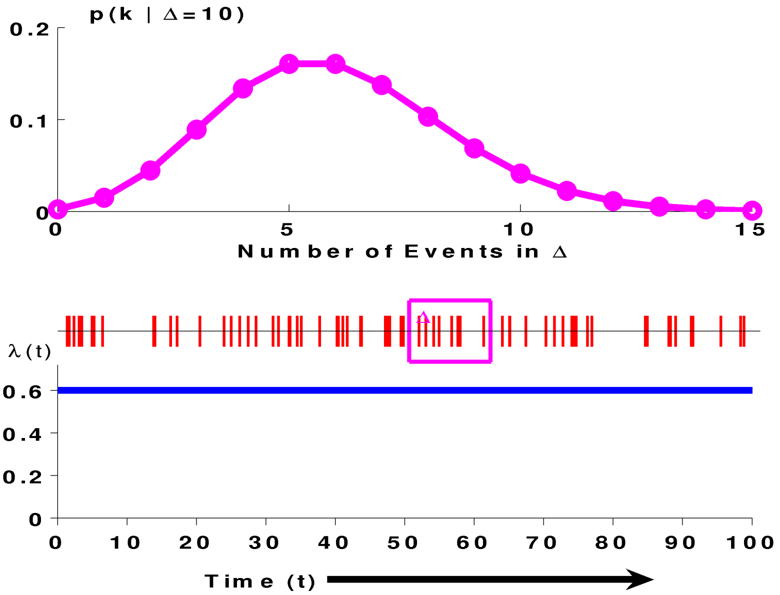
The Basic Poisson Process

- A **point process** of events in time.
- Has a **rate** (or intensity) $\lambda > 0$.
  - Expected number of events per unit time.
- An interval of length Δ_t has a Poisson-distributed number of events N_{Δ_t}:

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

- Disjoint intervals are independent.
- The time between arrivals is exponentially-distributed with parameter λ.

The Basic Poisson Process

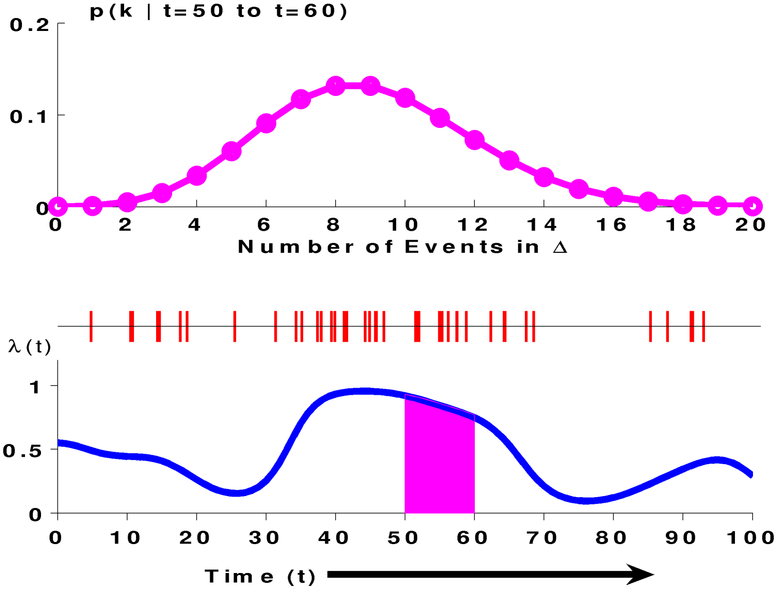


Inhomogeneous Poisson Process

- The rate is a function of time: $\lambda(t) \geq 0$.
- Disjoint intervals are still independent.
- The number of events between t_1 and t_2 is Poisson-distributed as:

$$p(k \mid \lambda(t)) = \frac{1}{k!}\left[\int_{t_1}^{t_2} dt\ \lambda(t)\right]^k \exp\left\{-\int_{t_1}^{t_2} dt\ \lambda(t)\right\}$$

Inhomogeneous Poisson Process



Inhomogeneous Poisson Process

Varying λ is much more interesting!

How can we make $\lambda(t)$ nonparametric?

How can we do inference with a NP $\lambda(t)$?

Modeling the Rate Function

The Inference Problem

Given some events $t_1, t_2, \ldots$, what is $\lambda(t)$?

The Prior on $\lambda(t)$

- We might not know much about it.
- We don't want to choose a functional form.
- We think the function is not too bumpy.
- Use a transformed Gaussian process!

Outline

Modeling with Poisson Processes

GPs for Inhomogeneous Poisson Processes

- Gaussian Processes
- The Log-Gaussian Cox Process
- A Generative Cox Process
- Latent History Inference
- Coal Mine Disaster Data

Future Directions

Summary

Why Gaussian Processes?

The GP is a nonparametric prior on functions.

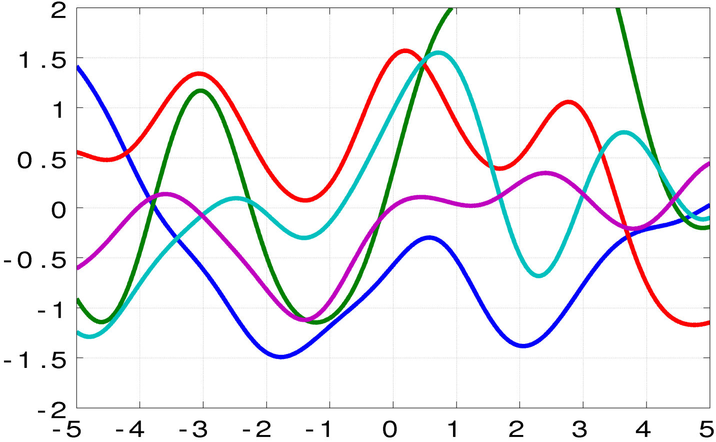
GP Components:

- Input space $\mathcal{X}$ (e.g. $[0, \infty)$ for time)
- Output space $\mathcal{Y} = \mathbb{R}$
- Covariance function
  $K(x, x'; \theta) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.
- Mean function $m(x; \theta) : \mathcal{X} \to \mathbb{R}$.

The predictive distribution and the marginal likelihood are easy.

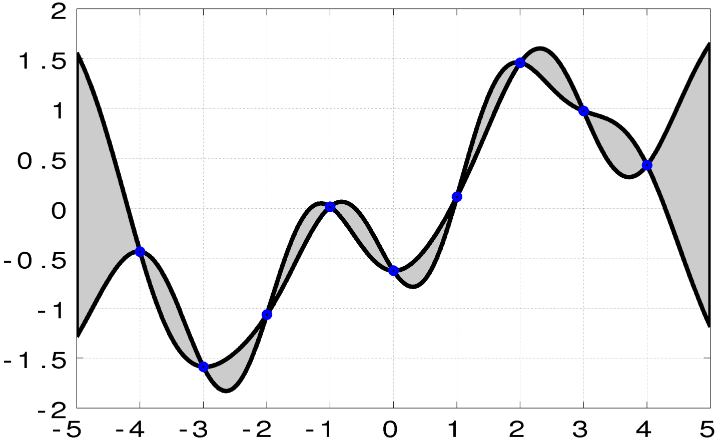
Gaussian Processes

"Nearby inputs have covarying outputs"



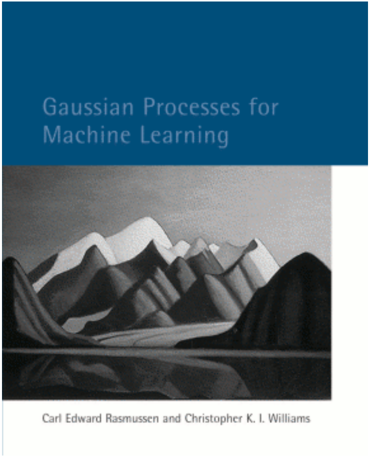
Gaussian Processes

"Nearby inputs have covarying outputs"



Gaussian Processes

Chris Williams will tell all on Saturday.



Read Chris and Carl's book.

Back to Poisson Inference

If we use a stochastic process to describe $\lambda(t)$, then the whole thing is called a **Cox Process**.

The Log-Gaussian Cox Process

Exponentiate the draw from the GP:

$$\lambda(t) = \exp\{g(t)\}, \text{ where } g(t) \sim \mathcal{GP}.$$

Nonparametric Inference is Intractable

We can't integrate random infinite-dimensional functions, but we need $\int dt \; \exp\{g(t)\}$ for the likelihood.

Our Solution

A Generative Prior

If we can generate exact data from the prior, we can almost always perform inference.

Rather than exponentiating $g(t) \sim \mathcal{GP}$, put it through a sigmoid, e.g. a scaled logistic:

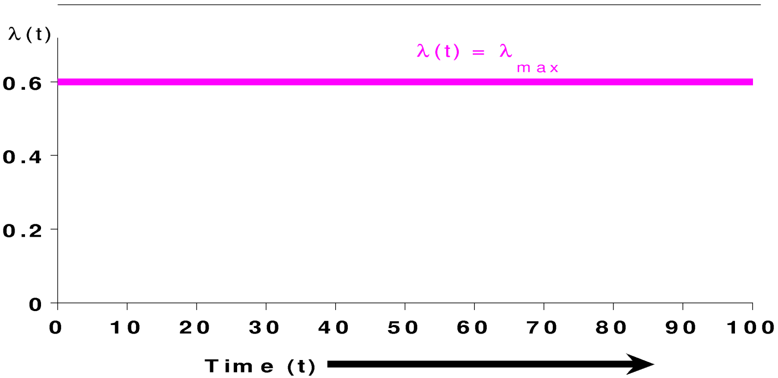
$$\lambda(t) = \frac{\lambda_{\text{max}}}{1 + \exp\{-g(t)\}}$$

where λ_{max} is a maximum rate.

We can generate exact data from this prior.

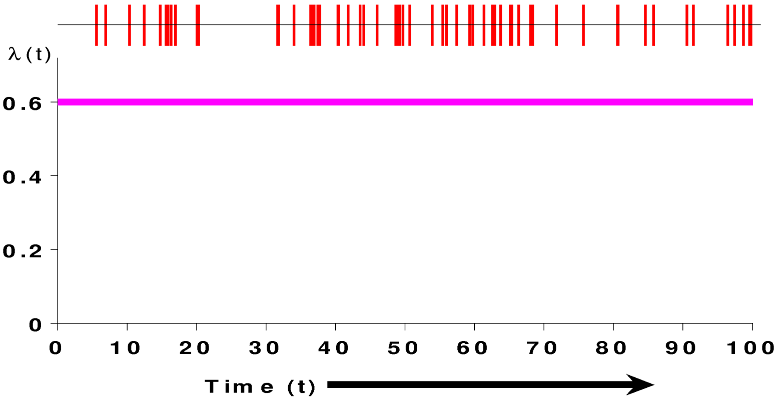
Thinning (Rejection Sampling)

1. Generate Poisson data from $\lambda(t) = \lambda_{\mathrm{max}}$.
2. Draw a sample from the GP at the events.
3. Reject some of the events.



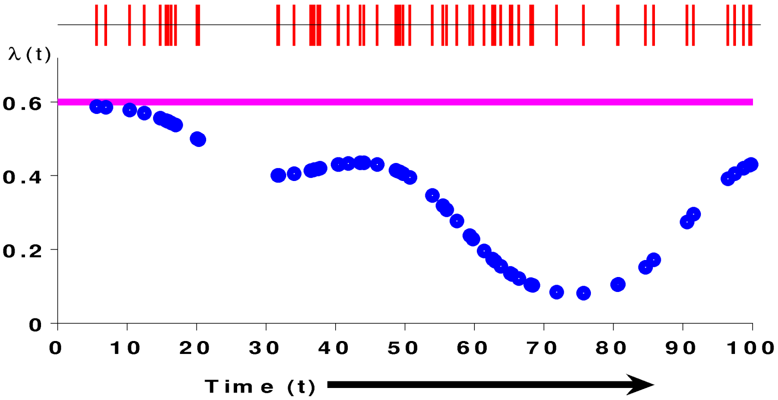
Thinning (Rejection Sampling)

1. Generate Poisson data from $\lambda(t) = \lambda_{\text{max}}$.
2. Draw a sample from the GP at the events.
3. Reject some of the events.



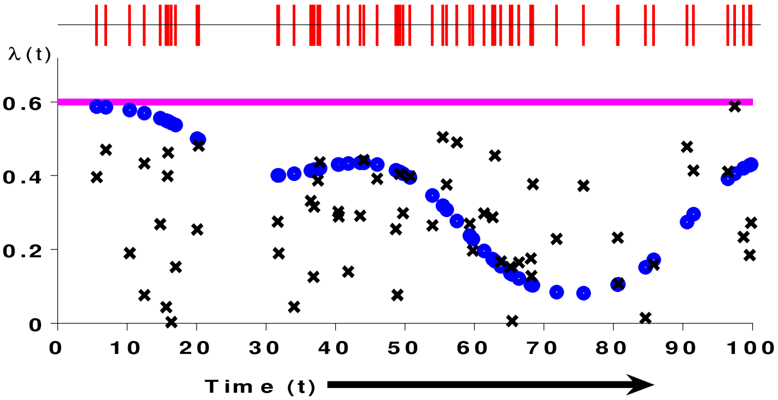
Thinning (Rejection Sampling)

1. Generate Poisson data from $\lambda(t) = \lambda_{\text{max}}$.
2. Draw a sample from the GP at the events.
3. Reject some of the events.



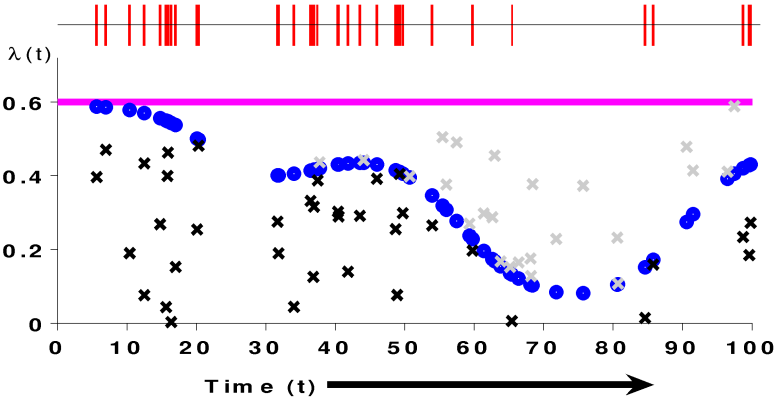
Thinning (Rejection Sampling)

1. Generate Poisson data from $\lambda(t) = \lambda_{\text{max}}$.
2. Draw a sample from the GP at the events.
3. Reject some of the events.



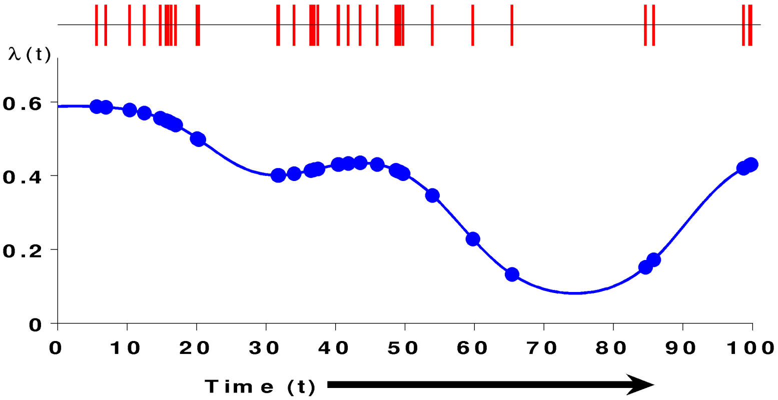
Thinning (Rejection Sampling)

1. Generate Poisson data from $\lambda(t) = \lambda_{\text{max}}$.
2. Draw a sample from the GP at the events.
3. Reject some of the events.



Thinning (Rejection Sampling)

1. Generate Poisson data from $\lambda(t) = \lambda_{\text{max}}$.
2. Draw a sample from the GP at the events.
3. Reject some of the events.



Features of the Generative Prior

The procedure is **exact**.

We did not need to know the function everywhere.

We did not have to calculate $\int dt\ \lambda(t)$.

Inference Via the Latent History

If we use this prior, we are saying that the data were generated via the thinning procedure.

Most of the state of that procedure is unknown, but can be treated as latent variables.

We have a well-defined probabilistic model on which to do inference.

We infer the **latent history** of the thinning process.

Inference Via the Latent History

We Don't Know:

- How many events were removed?
  - Denote this as M.
- When did the removed events happen?
  - Denote these as $\{\hat{t}_m\}_{m=1}^{M}$.
- What was the latent function?
  - Denote these as $\{g_n\}_{n=1}^{N}$ and $\{\hat{g}_m\}_{m=1}^{M}$.

$$p(M, \{\hat{t}_m\}, \{g_n\}, \{\hat{g}_m\} \,|\, \{t_n\}, T, \lambda_{\text{max}}) =$$

$$\frac{\lambda_{\text{max}}^{M}}{M!} \exp\{-T\lambda_{\text{max}}\}\, \mathcal{GP}(\{g_n\}, \{\hat{g}_m\} \,|\, \{t_n\}, \{\hat{t}_m\})$$

$$\times \left[\prod_{n=1}^{N} \frac{1}{1+\exp\{-g_n\}}\right] \left[\prod_{m=1}^{M} \frac{1}{1+\exp\{\hat{g}_m\}}\right]$$

Infer the Latent History

We use MCMC to sample from the posterior.
Two types of step:

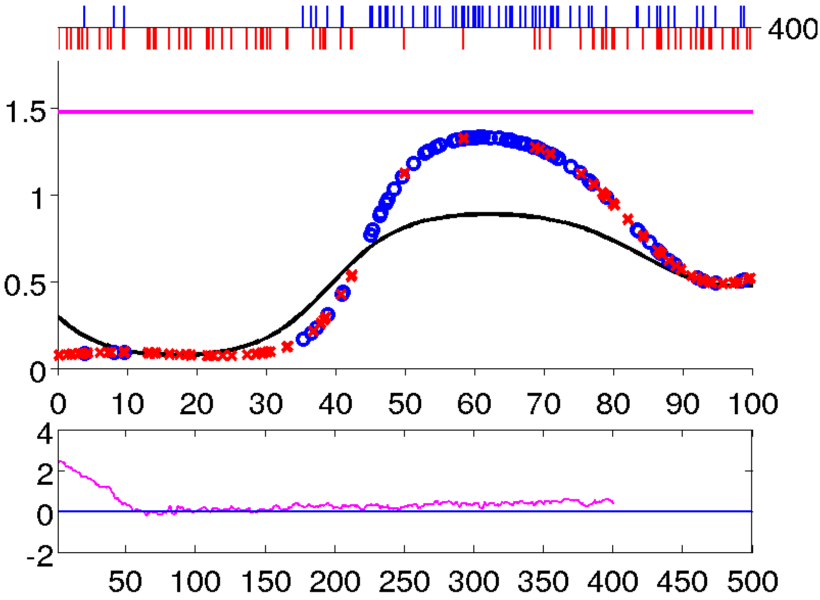
- Modify the number of latent rejections.
- Adjust the locations of the rejections and tweak the latent function.

We can also sample from GP hyperparameters:

- After each Markov step in the latent history, take a step in the GP hyperparameters.

We can also infer the value of λ_{max}.

Infer the Latent History



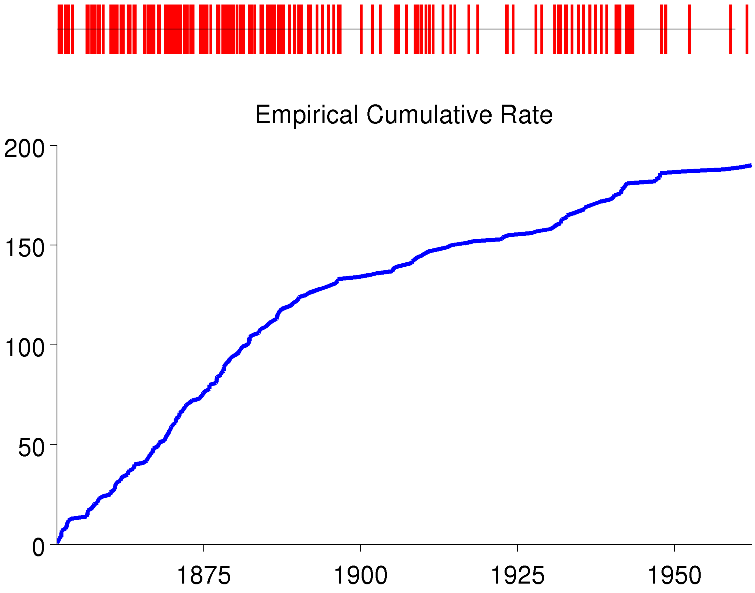
Infer the Latent History

Everybody likes movies!

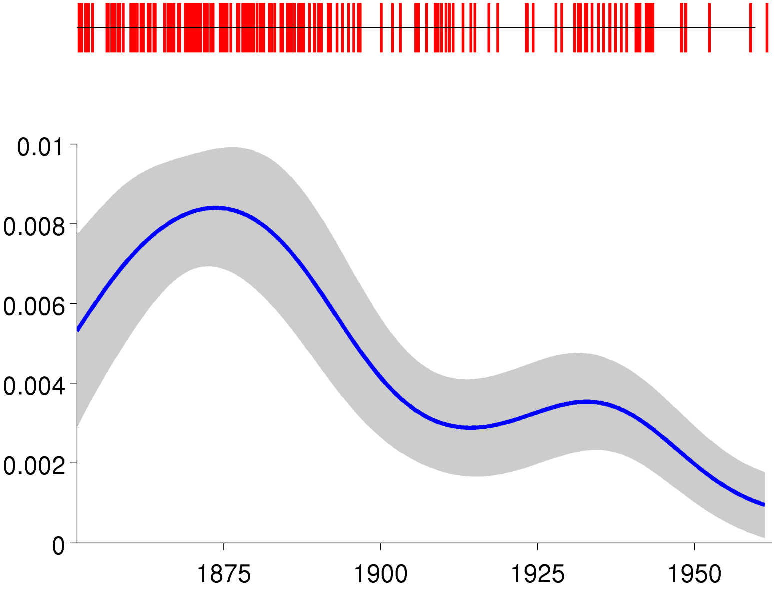
Coal Mine Disaster Data

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

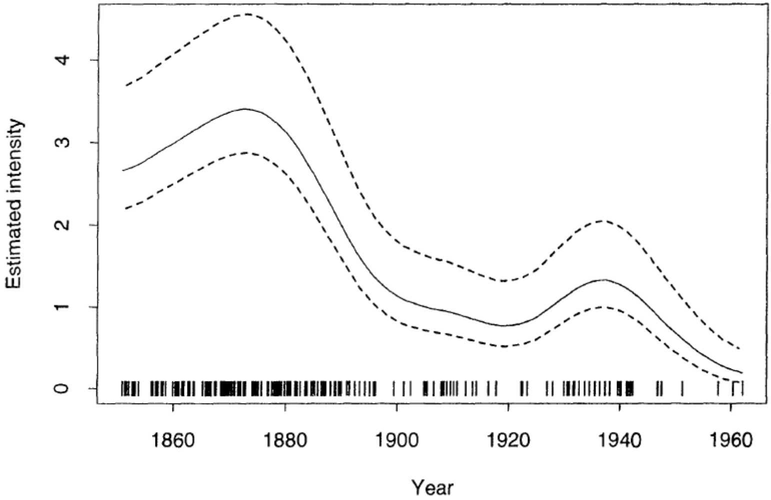
Coal Mine Disaster Data



Coal Mine Disaster Data



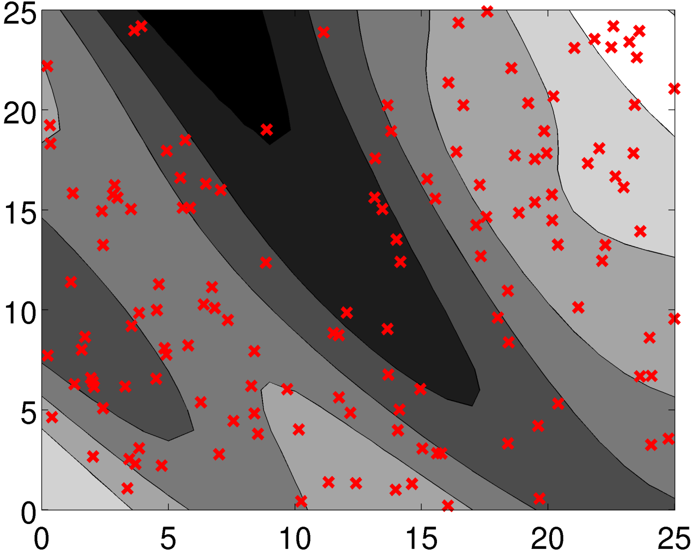
Coal Mine Disaster Data



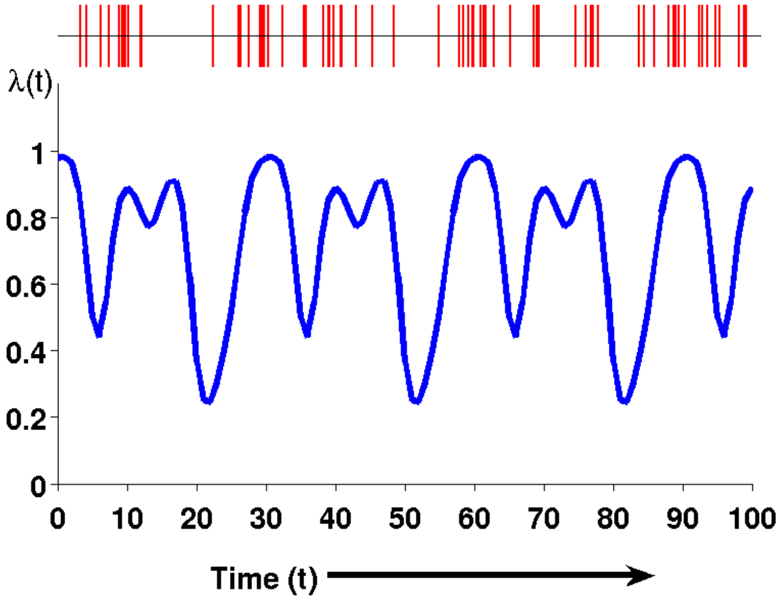
Cowling, Hall & Phillips, JASA 1996

Outline

Spatial Data



Nonstationary Covariance Functions



Sparse GP Approximations

GPs are expensive: $O(n^3)$ in the data.

We are $O((N+M)^3)$ per MCMC step.

One Possibility:

Introduce sparsity by “clumping” data together.

Coupled Poisson Processes

Examine structure **across** point processes:

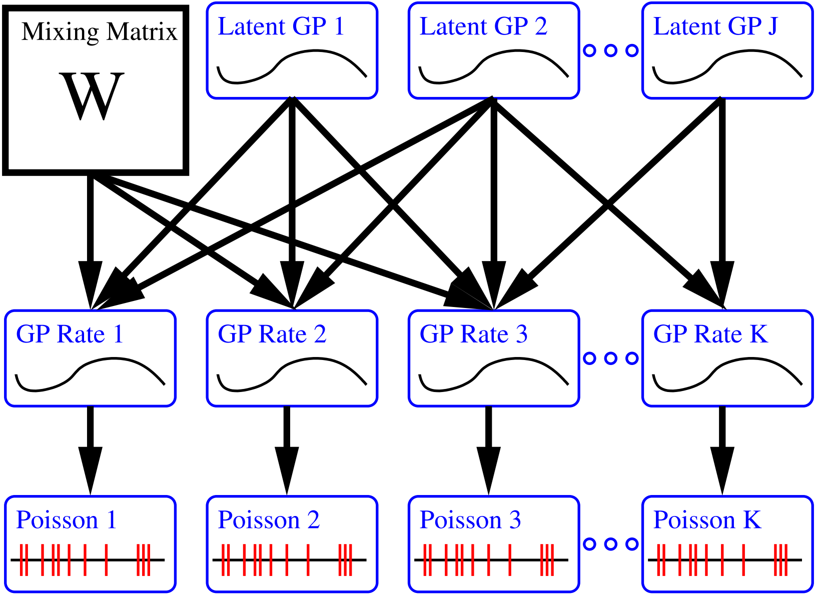
- Neural populations
- Arrivals at hospitals in a city
- Bank failures by country

Model them as Poisson processes that are independent, given the intensity functions.

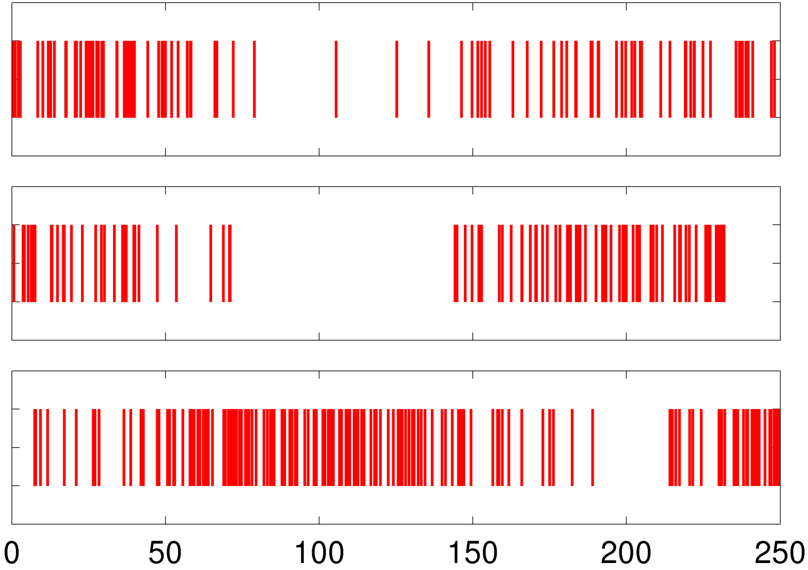
Use a Semiparametric Latent Factor Model (Teh, et al. 2004) to couple the rate functions.

Could also introduce latent input spaces.

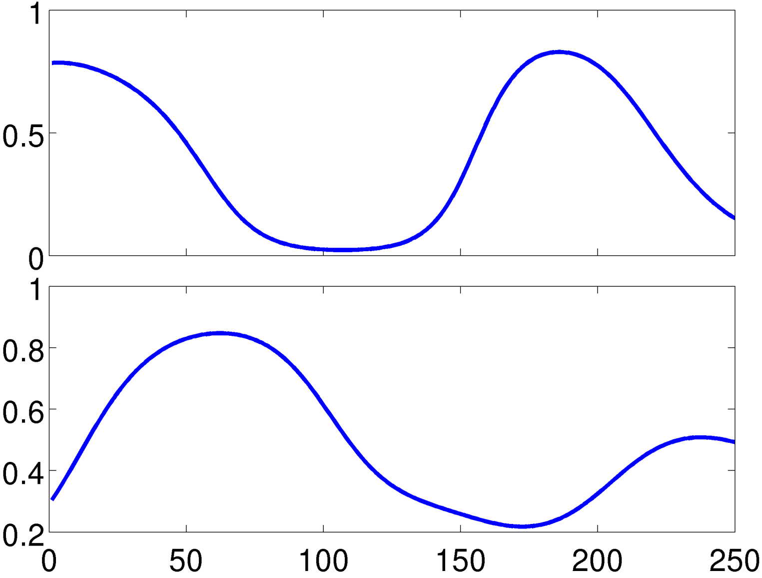
Coupled Poisson Processes



Coupled Poisson Processes



Coupled Poisson Processes



Outline

Summary

- Inhomogeneous Poisson process are useful models.
- Generally, using a GP rate function is hard.
- We can solve this with a bounded rate function.
- We can do inference of the function, hyperparameters and bound.
- There are neat directions to go with this.

Thanks

- Iain Murray
- David MacKay
- Radford Neal
- Zoubin Ghahramani

RPA is funded by the Gates Cambridge Trust and IAM is funded by the government of Canada.