Point Processes, Discrete Random Probability Measures, and their use in Bayesian Statistics

Katja Ickstadt

SAMSI, October 2009
OVERVIEW

1. Introduction to Point Processes and Random Probability Measures

2. Point Processes as Priors in Bayesian Statistics
   2.1 Prior for Poisson process intensity based on the gamma process, with application in spatial statistics
   2.2 Prior for continuous monotonic function based on variants of the Dirichlet process, with an application in dose-response and growth curve analysis

3. Conclusions
1. Introduction to Point Processes and Random Probability Measures
A point process is a stochastic model for the locations of events \( \{ s_j \} \) in some set \( \mathcal{X} \).

If \( \mathcal{X} \subset \mathbb{R}^d, d = 2, 3 \) it is called a spatial point process.

If each \( s \) has an associated observation \( z \in \mathcal{F} \), the \( (s, z) \in \mathcal{X} \times \mathcal{F} \) are realizations of a marked point process.
Definition: Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\Phi$ a collection of locally finite counting measures on $\mathcal{X}$. Let $\sigma(\Phi)$ be the smallest $\sigma$-field that is generated by the sets $\{\varphi \in \Phi : \varphi(A) = n\}$, $\forall A \in \mathcal{B}(\mathcal{X})$. A point process $N$ on $\mathcal{X}$ is a measurable mapping $(\Omega, \mathcal{A}) \rightarrow (\Phi, \sigma(\Phi))$.

A point process induces a probability measure $P(N \in B)$ on $(\Phi, \sigma(\Phi))$, $B \in \sigma(\Phi)$. 
**Poisson Process**

$N$ is a *(inhomogeneous) Poisson process* on $\mathcal{X}$ with nonnegative $\sigma$-finite *intensity measure* $\lambda(dx)$, if

1. for all $A \in \mathcal{B}(\mathcal{X})$ with $\lambda(A) < \infty$ and $k \in \mathbb{N}_0$

$$P(N(A) = k) = \frac{\lambda(A)^k e^{-\lambda(A)}}{k!}$$

holds and

2. the random variables $N(A_1), \ldots, N(A_k)$ are independent for disjoint sets $A_1, \ldots, A_k \in \mathcal{B}(\mathcal{X})$ with

$$\forall j = 1, \ldots, k : \lambda(A_j) < \infty.$$

It is $\lambda(A) = \int_A \lambda(s) \omega(ds)$ with density $\lambda(s)$ with respect to $\omega(ds)$. 
Poisson Process

• If $\lambda(A) = \lambda \omega(A)$ for all $A \in \mathcal{B}(X)$ for $\lambda > 0$ constant, then the Poisson process is homogeneous.

  ($= \lambda |A|$ if $\omega$ is the Lebesgue-measure)

• For a homogeneous Poisson process the locations $\{ s_j \}$ are uniformly distributed on $A$ given $N(A)$, the number of points in $A$.

• In spatial statistics a homogeneous Poisson process is also referred to as complete spatial randomness.
SIMULATION

To simulate a homogeneous Poisson process on $A$:

1. Draw $N(A)$ from a Poisson distribution.
2. For $j = 1, \ldots, N(A)$ draw $s_j$ from a uniform distribution on $A$.

To simulate an inhomogeneous Poisson process:

1. Draw points from a homogeneous Poisson process with intensity $\lambda^* = \sup_{s \in A} \lambda(s)$
2. Keep each point with probability $\frac{\lambda(s_j)}{\lambda^*}$
Implemented in R package spatstat (function rpoispp)
Homogeneous Poisson Process

on Unit Square with Lambda=100

Homogeneous Poisson Process

on Unit Square with Lambda=500
Inhomogeneous Poisson Process

on Unit Square
EXTENSIONS OF THE POISSON PROCESS

- **Cox process**: If $\Lambda(s)$ is a non-negative valued stochastic process and conditional on $\Lambda(s)$ the point process is a inhomogeneous Poisson process with intensity $\Lambda(s)$.

- **Inhibitory processes**: Processes where points tend to reject one another, e.g. Strauss processes where no two observations can be closer than $\delta$.

- **Cluster processes**: Processes where points tend to occur in clusters, e.g. Neyman-Scott processes.
**NEYMAN-SCOTT PROCESS**

**Definition:** A (homogeneous) Neyman-Scott process is a point process $N$ on $\mathcal{X}$ for which

1. The locations of parents (cluster centers) follow a Poisson process.

2. The number of offspring for each parent are independent, identically distributed.

3. The distributions of offspring around their parents are independently drawn for each offspring.

4. The process consists only of the offspring.
Homogeneous Neyman–Scott Process

5 Offspring Per Parent

Homogeneous Neyman–Scott Process

Poisson Distributed Offspring
APPLICATION: SIMULATION OF RAS-PROTEINS ON CELL MEMBRANES

- Ras-Proteins play an important role in the signal transduction on cell membranes.
- Some of the proteins lie on the membrane by themselves, some in clusters.
- Percentage of proteins that are clustered and size of clusters are of interest to biologists.
APPLICATION: SIMULATION OF RAS-PROTEINS ON CELL MEMBRANES

- Images of proteins are taken via fluorescence microscopy.
- Only a small part of the proteins can be made visible in each image.
- Algorithms to detect proteins on the image are available, but not much is known about their performance.
- Simulate pictures to test performance.
- Protein patterns are simulated as realizations of Neyman-Scott processes.
Fluorescence microscopy image of Ras-proteins (bad quality)
Simulated image of Ras-proteins
Simulated image of Ras-proteins with added noise
**Discrete Random Probability Measures**

Let $G^*$ be a class of probability measures on $(\Omega, \mathcal{A})$. The *Dirichlet process* (Ferguson (1973)) is a distribution $\mathcal{P}(G)$ on the distributions $G \in G^*$, a *random probability measure*.

**Definition:** A random probability measure $G$ follows a *Dirichlet process* with mass (concentration) parameter $\alpha$ and base measure $G_0$, i.e. $G \sim DP(\alpha, G_0)$, if it holds $\forall$ partitions $(A_1, \ldots, A_k)$ of $\Omega$ that:

\[
(G(A_1), \ldots, G(A_k)) \sim Dirichlet(\alpha \cdot G_0(A_1), \ldots, \alpha \cdot G_0(A_k)).
\]
**Dirichlet Process**

Consider the partition \((A, A^C)\), then for all \(A \in \mathcal{A}\):

\[
G(A) \sim Beta(\alpha \cdot G_0(A), \alpha \cdot (1 - G_0(A)))
\]

It follows that

\[
E(G(A)) = G_0(A)
\]
\[
Var(G(A)) = \frac{G_0(A)(1 - G_0(A))}{\alpha + 1}
\]

\(G_0\) is the expectation of the random measure \(G\) and \(\alpha\) determines the variability around \(G_0\).
Properties of the Dirichlet Process

i) Conjugacy:

If $S_1, \ldots, S_n \mid G \overset{i.i.d.}{\sim} G$ and $G \sim DP(\alpha, G_0)$, then

$$G \mid s_1, \ldots, s_n \sim DP(\alpha + n, G_0^*)$$

with

$$G_0^* = \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{j=1}^{n} \delta(s_j) ,$$

$\delta(s_j)$ being the point measure at $s_j$.

The distribution $G_0^*$ is thus a mixture of $G_0$ and the empirical distribution of the $s_j$. 
Properties of the Dirichlet Process

The posterior predictive distribution of $S_{n+1}$ is given by $G_0^*$:

$$S_{n+1}|s_1, \ldots, s_n, \alpha, G_0 \sim \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{j=1}^{n} \delta(s_j)$$

Because of the $\delta(s_j)$ there is positive probability that the next observation is equal to one of the previous.

The distribution of $S_{n+1}$ is that of a Pólya urn: If a ball of a certain color is drawn two balls of the color are placed back in the urn and with probability $\alpha/(\alpha + n)$ a ball of a new color is drawn, where the color is determined by a draw of $G_0$. 
Properties of the Dirichlet Process

ii) Discreteness:

Sethuraman (1994) showed that $G$ is discrete with probability 1 even if $G_0$ is continuous and can be written as

$$G = \sum_{j=1}^{\infty} w_j \delta(s_j),$$

where $s_j \overset{i.i.d.}{\sim} G_0$ and

$$w_j = U_j \prod_{h=1}^{j-1} (1 - U_h) \quad \text{and} \quad U_j \overset{i.i.d.}{\sim} Beta(1, \alpha).$$

This representation is referred to as infinite mixture model.
Properties of the Dirichlet Process

The $w_j$ have a "stick-breaking" construction. Proportions $U_j \sim Beta(1, \alpha)$ are successively broken of the remaining "stick"
**Gamma Process**

**Definition:**  
$\Gamma$ is a *Gamma process* on $\Omega$ with nonnegative $\sigma$-finite shape measure $\alpha(ds)$ and inverse scale $\beta > 0$,  

$$
\Gamma \sim Ga(\alpha(ds), \beta^{-1})
$$

if
**Gamma Process**

i) for all measurable sets $A \in \mathcal{A}$ with $\alpha(A) < \infty$ the random variable $\Gamma(A)$ has a density

$$f_{\Gamma(A)}(t) = \frac{\beta^{\alpha(A)} t^{\alpha(A) - 1} e^{-\beta t}}{\Gamma(\alpha(A))}, \quad t > 0,$$

and

ii) the random variables $\Gamma(A_1), \ldots, \Gamma(A_k)$ are independent for all measurable, and disjoint $A_1, \ldots, A_k \in \mathcal{A}$ with $\alpha(A_j) < \infty, \ j = 1, \ldots, k$. 

**Gamma Process**

- For all $A \in \mathcal{A}$ the random variable $\Gamma(A)$ thus follows a $Ga(\alpha(A), \beta^{-1})$ distribution with

  $$\alpha(A) = \int_A \alpha(ds) = \int_A \alpha(s)r(ds)$$

  the density $\alpha(s)$ being with respect to a reference measure $r(ds)$.

- The inverse scale parameter can be dependent on $s$,
  $\Gamma \sim Ga(\alpha(ds), \beta(s)^{-1})$ is then an inhomogeneous Gamma process.
Properties of the Gamma Process

• As the Poisson process and the Dirichlet process the Gamma process is almost surely discrete. Its realizations consist of countably $\infty$ many jumps at locations $s_j$ with heights $u_j > 0$, so that

$$\Gamma = \sum_j u_j \delta(s_j).$$

• Heights $u_j$ and locations $s_j$ can be generated with the Inverse Lévy Measure Algorithm (Wolpert und Ickstadt (1998)). This has similar algorithmic advantages as the "stick-breaking": The largest point masses can be generated first.
**Inverse Lévy Measure Algorithm**

1. Fix a large $M \in \mathbb{N}$ and choose any convenient distribution $\Pi(ds)$ on $\Omega$ from which samples can be drawn and such that the shape measure $\alpha(ds)$ has a density $\alpha(s) \equiv \alpha(ds)/\Pi(ds)$.

2. For $j = 1, \ldots, M$ generate $s_j \overset{i.i.d.}{\sim} \Pi(ds)$.

3. Construct the first $M$ event times $\tau_j$, $j = 1, \ldots, M$ of a standard Poisson process, e.g. by the cumulative sum of independent exponential random variables $T_i$, i.e., $\tau_j = \sum_{i \leq j} T_i$. 

**Inverse Lévy Measure Algorithm**

4. Set $u_j = E_1^{-1}(\tau_j / \alpha(s_j))\beta^{-1}$, where

$$E_1(t) = \int_{t}^{\infty} e^{-u} u^{-1} du$$

is the exponential integral function.

5. Set $\Gamma_M(ds) \equiv \sum_{j \leq M} u_j \delta(s_j) \approx \Gamma(ds)$.

The algorithm works analogously for other Lévy processes; just replace step 4. by setting $u_j$ equal to the inverse Lévy measure $l(\tau_j, s_j)^{-1}$. 
Properties of the Gamma Process

- The Gamma process has independent increments.
- There is an important relation between Gamma and Dirichlet process: The normalized Gamma process $\Gamma/\Gamma(\Omega)$ is a Dirichlet process.


2. Point Processes in Bayesian Nonparametrics
POINT PROCESSES IN BAYESIAN NONPARAMETRICS

• What is Bayesian Nonparametrics (BNP)?

• BNP uses prior distribution on infinite dimensional “parameters”, i.e. functions instead of vectors → priors supported on
  – Space of continuous probability meas. supported on \( \mathbb{R} \)
  – Space of monotonic continuous functions on \([0, 1]\)

• Prior distributions → stochastic processes instead of finite dim. distributions

• O’Hagan and Forster (2004, chapter 13) and Müller and Quintana (2004) provide reviews of BNP methods

• Point processes play an important role as priors in BNP
MIXTURES IN BAYESIAN NONPARAMETRICS

• One approach to build a (nonparametric) prior distribution for a function $f(x)$ is by using a discrete mixture:

  Pre-specify a function $k(x, \theta)$ and model $f(x)$ as

  $$f(x) = \sum_j u_j k(x, \theta_j) = \int k(x, \theta) P(d\theta)$$

• Examples:
  
  – $f(x)$ probability density:
    $k(x, \theta)$ density, $P(d\theta)$ discrete probability measure
  
  – $f(x)$ intensity:
    $k(x, \theta)$ e.g. density, $P(d\theta)$ discrete measure
  
  – $f(x)$ general function:
    $k(x, \theta)$ e.g. density, $P(d\theta)$ discrete signed measure
MIXTURES IN BAYESIAN NONPARAMETRICS

- What prior to use for $P(d\theta) = \sum_j u_j \delta(\theta_j)$?

- Use a point process, random measure, random probability measure,...!

- Rest of this talk illustrates two nonparametric priors based on point processes

  2.1 Prior for Poisson process intensity based on the gamma process

  2.2 Prior for continuous monotonic function based on variants of the Dirichlet process
2.1 Point Process Priors in Spatial Statistics
OVERVIEW

• Introduction

• Example: Leukemia data set

• Poisson/gamma models

• Spatial analysis of Leukemia data set
INTRODUCTION

- Spatial epidemiology, i.e., data and (possibly) covariates have a spatial component

- Problem: Data are given with different spatial resolution:
  - Administrative: population, disease counts, unemployment, . . .
  - Regular grid: environmental data, sea levels, . . .
  - Irregular grid: wind speeds, weather conditions, . . .

- Aims: Disease mapping and ecologic regression without aggregating all data to a coarse, common spatial resolution
Leukemia data set
EXAMPLE: LEUKEMIA DATA

- Counts of Leukemia cases for children on ward level in London
- Obtained from Office for National Statistics and Thames Cancer register
- 295 registered cases from 1985 to 1996
EXAMPLE: EXPECTED LEUKEMIA CASES

- Population count $N_{ist}$ is stratified according to age (0-4, 5-9 and 10-14 years) and gender on ward level from the census of 1981 and 1991.

- Calculation of expected cases $E_i$ in region $i$:

$$E_i = \sum_{st} r_{st} N_{ist}$$
Example: Standardized Mortality Ratios

- Standardized Mortality Ratio in ward $i$:
  \[ \text{SMR}_i = \frac{\text{observed cases}}{\text{expected cases}} \]
  \[ \frac{O_i}{E_i} \]

- $\text{SMR}_i$: MLE for a non-spatial Poisson model

- Aim: model-based smoothing of the SMRs
**Example: Benzene Emission**

- Covariate: Benzene emission on 1 km $\times$ 1 km grid in tons/year
- Taken from *atmospheric emissions inventory* of London
- Different spatial resolution
Poisson/gamma models
**Standard Approaches**

- **Aggregate** data and covariates with the aim of a joint spatial resolution (e.g. ward level)

- Disadvantage: **Ecological bias**

- **Analysis**: Poisson–log normal model or Poisson–log normal model with Markov Random Field structure to describe spatial random effects (Besag et al., 1991, Best et al., 2001)
ALTERNATIVE APPROACH

- Random field generalization of conjugate Poisson/gamma-models (Wolpert and Ickstadt, 1998, Best et al., 2000), which allows data and covariates on the measured spatial scale

+ Covariates can be incorporated as excess or relative risk factors (Breslow and Day, 1980) \(\leadsto\) different interpretations

  - Excess risk factor: “competing risk”: \(\Lambda = \beta_0 + \rho(x)\)
  - Relative risk factor: “modifying factor”: \(\Lambda = \beta_0 \times \theta(x)\)
POISSON/GAMMA MODELS: LATENT TERM

- Latent spatial term: discrete mixture
- Gamma process $\Gamma(ds)$: Realizations consist of infinitely many jumps (latent sources) $\mu_S = (\mu_X, \mu_Y)$ with jump heights $\gamma_s$
- $k(y, s)$: kernel function, e.g. Gaussian kernel, sensible choice for unobserved spatial covariates
- Of interest: $k(y, s)$, i.e. the effect of jump $s$ of the gamma process at location $y$

$$k(y, s) = \exp \left\{ -\frac{1}{2} \left( \frac{(y_X - \mu_X)^2}{\sigma_1^2} + \frac{(y_Y - \mu_Y)^2}{\sigma_2^2} \right) \right\}$$

according to a bivariate Gaussian kernel with random expectation $(\mu_X, \mu_Y)$ evaluated at $(y_X, y_Y)$
Poisson/Gamma Models: Latent Term

- Realization of a latent source $k(y, s) \gamma_s$:

$$\Rightarrow \text{latent term: } \int_S k(y, s) \Gamma(ds) \beta_* = \sum_s k(y, s) \gamma_s \beta_*.$$
Poisson/gamma Models

Observe locations $y$ and attributes $a \Rightarrow$ marked points $x = (y, a)$. Hence:

$$N(dy \, da) \sim \text{Pois}\left( \Lambda(y, a) w_Y(dy) w_A(da) \right)$$

Normalize: $w_A(da) \equiv 1$, $w_Y(dy) \equiv$ Population structure

$$\Lambda(y, a) = \left( \sum_{j \in J_A} a_j \beta_j \right) \times \exp \left( \sum_{j \in J_M} a_j \beta_j \right)$$
**Poisson/gamma Models**

Observe locations $y$ and attributes $a \Rightarrow$ marked points $x = (y, a)$.

Hence:

\[ N(dy \, da) \sim \text{Pois}(\Lambda(y, a) w_Y(dy) w_A(da)) \]

Normalize: $w_A(da) \equiv 1$, $w_Y(dy) \hat{=} \text{Population structure}$

\[
\Lambda(y, a) = \left( \sum_{j \in J_A} a_j \beta_j + \int_S k(y, s) \Gamma(ds) \beta_* \right) \times \exp \left( \sum_{j \in J_M} a_j \beta_j \right)
\]
Poisson/Gamma Models

Observe locations $y$ and attributes $a \Rightarrow$ marked points $x = (y, a)$.
Hence:

$$N(dy \, da) \sim \text{Pois}(\Lambda(y, a)w_Y(dy) \, w_A(da))$$

Normalize: $w_A(da) \equiv 1$, $w_Y(dy) \equiv$ Population structure

$$\Lambda(y, a) = \left( \sum_{j \in J_A} a_j \beta_j \right) \times \exp \left( \sum_{j \in J_M} a_j \beta_j \right)$$

$$\Lambda(y, a) = \left( \sum_{j \in J_A} a_j \beta_j + \int_S k(y, s) \Gamma(ds) \beta_* \right) \times \exp \left( \sum_{j \in J_M} a_j \beta_j \right)$$

$$\Gamma(ds) \sim \text{Gamma}(\alpha^\beta(ds), \tau^\beta(ds))$$

$$\beta \sim \pi(\beta)d(\beta)$$
Spatial analysis of Leukemia data set
Spatial Analysis of Leukemia Data

- Expected counts on ward level (reference)
- Observed covariate: benzene on grid level
- Latent covariates as random field surface

a) Poisson/gamma model without benzene
\[ \Lambda^a(y) = \beta_0 + \beta_k \sum_s k(y, s) \gamma_s \]

b) Poisson/gamma model with additive benzene effect
\[ \Lambda^b(y, B) = \beta_0 + \beta_{benz}^A B(y) + \beta_k \sum_s k(y, s) \gamma_s \]

c) Poisson/gamma model with multiplicative benzene effect
\[ \Lambda^c(y, B) = (\beta_0 + \beta_k \sum_s k(y, s) \gamma_s) \times \exp(\beta_{benz}^M B(y)) \]
**Prior Distributions**

\( \beta \)-Parameter: prior expectations: \[
\frac{\text{#observations}}{\text{# risk factor} \times \text{# expected cases}}
\]

\[
\beta_0 \sim \text{Gamma}(0.575, \tau_0)
\]

\[
\beta_{\text{benz}} \sim \text{Gamma}(0.575, \tau_{\text{benz}})
\]

\[
\beta_* \sim \text{Gamma}(0.575, \tau_{\text{latent}})
\]

**Gamma process:**

\[
\gamma_m \sim \text{Gamma}(\alpha_{\gamma}, \tau_{\gamma})
\]

\[
\alpha_{\gamma} = \text{area} \times \tau_{\gamma}
\]

\[
\tau_{\gamma} = \frac{1}{\text{# latent risk factors}}
\]
**Prior Distribution**

\[
\begin{align*}
\mu_X & \sim \text{Unif(bounding box)} \\
\sigma_X & \sim \log\text{Normal}(0, p_X) \\
p_X & = \begin{cases} 
2 & \text{if } z > 0.5 \\
6 & \text{if } z \leq 0.5
\end{cases} \\
z & \sim \text{Unif}(0, 1)
\end{align*}
\]

similarly for \(\mu_Y\) and \(\sigma_Y\)

model choice: via Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002):

\[
\text{DIC} = \text{"model fit" (mean deviance) + "complexity" (number of effective parameters)}
\]
**RESULTS**

a) Poisson/gamma model without influence of benzene

Posterior Means of estimated $\hat{SMR}_i$ (left) and random field locations (9 latent sources) (right)
RESULTS

b) Poisson/gamma model with additive effect of benzene

Posterior Means of estimated $\hat{SMR}_i$ (left) and random field locations (1 latent source) (right)
c) Poisson/gamma model with multiplicative effect of benzene

Posterior Means of estimated $\hat{SMR}_i$ (left) and random field locations (9 latent sources) (right)
## Model Comparison: DIC

<table>
<thead>
<tr>
<th># latent sources</th>
<th>Model b)</th>
<th>Model c)</th>
<th>Model a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>417.0</td>
<td>415.5</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>390.8</td>
<td>391.0</td>
<td>390.8</td>
</tr>
<tr>
<td>2</td>
<td>394.6</td>
<td>397.3</td>
<td>396.6</td>
</tr>
<tr>
<td>3</td>
<td>417.1</td>
<td>391.8</td>
<td>393.3</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>388.9</td>
<td>390.5</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>386.9</td>
<td>388.2</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>385.2</td>
<td>386.4</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>384.1</td>
<td>384.8</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>383.4</td>
<td>383.8</td>
</tr>
<tr>
<td>9</td>
<td>—</td>
<td>382.8</td>
<td>383.2</td>
</tr>
<tr>
<td>10</td>
<td>—</td>
<td>382.5</td>
<td>383.0</td>
</tr>
</tbody>
</table>
CONCLUSION

• Discrete mixture (here based on the gamma process) allows for building a flexible and adequate BNP model for the latent spatial term

• The overall semi-parametric Poisson/gamma random field model allows for ecologic regression without ecological bias

• Covariates may enter multiplicatively or additively; in Leukemia example: Multiplicative modelling of benzene is preferred

• Properties of the model class have been checked and the class has been compared to standard approaches in an extensive simulation study (Sturtz, 2007)

• Model fitting can be done in WinBUGS


2.2 Point Process Priors for Monotone Regression
OVERVIEW

- Monotone Regression
  - Nonparametric Prior for Monotone Functions
  - Interpretation
  - Implementation

- Application to Dose-Response Analysis

- Application to Growth Curves
Monotone Regression
**Motivation Monotonic Regression**

- **Clinical trials**: Statistical methodology determined before collecting data, *e.g.* functional form of monotonic dose-response curve in dose-finding trial $\rightarrow$ uncertainty

- **Way out**: a) Use class of param. dose-response functions + model selection/averaging  
  *Here*: b) Directly use nonparametric model

- **Problem**: Nonparametric models are often difficult to interpret  
  $\rightarrow$ incorporation of prior knowledge difficult

- **Aim of this work**: Provide flexible nonparametric models that allow incorporation of prior information on interpretable aspects of the underlying function
PARAMETRIC DOSE RESPONSE MODELS

- Sigmoid Emax model: \( \mu(x) = \beta_0 + \beta_1 \frac{x^\delta}{x^\delta + ED_{50}^\delta} \)

- Logistic model: \( \mu(x) = \beta_0 + \beta_1 \frac{1}{1 + \exp((ED_{50} - x)/\delta)} \)

- General structure

\[ \mu(x) = \beta_0 + \beta_1 \mu^0(x), \]

where \( \mu^0(x) \) is a probability distribution function

\( \rightarrow \) \( \beta_0, \beta_1 \) interpretable as baseline and maximum effect

- In the following: Instead of assuming a parametric model for \( \mu^0(\cdot) \), model it nonparametrically
MONOTONE REGRESSION

• Decompose monotone function $\mu(.)$ on $[0, 1]$ as

$$\mu(x) = \beta_0 + \beta_1 \mu^0(x),$$

where $\mu^0(x)$ is a probability distribution function on $[0, 1]$

• **Advantage:** $\beta_0, \beta_1$ often interpretable in application context
  (e.g. baseline and maximum effect in dose-response analysis)

• Which priors should be used?
  – For $\beta_0, \beta_1$ conjugate priors can be used
  – Assume **nonparametric** prior for $\mu^0(.)$
MONOTONE REGRESSION

• Model $\mu^0(.)$ as discrete mixture of parametric probability distribution functions $F(x, \xi)$ with $\xi \in \Xi$

$$
\mu^0(x) = \sum_{j=1}^{J} w_j F(x, \xi_j)
$$

$$
= \int_{\Xi} F(x, \xi) P(d\xi),
$$

where $P(d\xi)$ is discrete probability distribution, with support $\xi_1, \ldots, \xi_J$ and $\sum_{j=1}^{J} w_j = 1$

• $P(d\xi)$ is unknown and needs to be ‘learned’ from the data

• What prior to use for $P(d\xi)$?
MONOTONE REGRESSION: PRIOR

• Ongaro and Cattaneo (2004) propose the following stochastic process as prior for discrete probability measures

\[ P(d\xi) = \sum_{j=1}^{J} w_j \delta_{\xi_j}(d\xi), \]

where \( J \) is a positive, integer-valued random variable,
\( \xi_j \overset{iid}{\sim} P_0, \) \( P_0 \) continuous probability distr. on \( \Xi \)
\( w_1, \ldots, w_J \sim Q_J, \) \( Q_J \) prob. distr. on \( J - 1 \) dim. simplex,
\( \xi_j \) independent of \( w_j \) and \( J \)
**MONOTONE REGRESSION: PRIOR**

Contains many popular discrete random measures used in BNP as special case

- Dirichlet process $P \sim DP(\alpha P_0)$
  
  $J = \infty$ and $\mathbf{w} = (w_1, \ldots, w_j, \ldots)' \sim GEM(\alpha)$
  
  (Sethuraman, 1994)

- Truncated stick breaking process
  
  $J = N$, $\mathbf{w} \sim GDir_{N-1}(a, b)$ (Ishwaran, James, 2001)

- In our situation:
  
  Reasonable to assume $J < \infty$. One possible choice:

  $J \sim TPois(\lambda)$, $\mathbf{w}|J \sim Dir_{J-1}(\delta \mathbf{1})$, $\forall J > 0$, $\lambda, \delta > 0$

  $\leftrightarrow$ This ensures full support (see Ongaro and Cattaneo, 2004)
MONOTONE REGRESSION: PRIOR

• Summary: \( \mu^0(x) \) is modelled as
  \[
  \sum_{j=1}^{J} w_j F(x, \xi_j) = \int_{\Xi} F(x, \xi) P(d\xi)
  \]
  and a general random probability measure assumed for \( P(d\xi) \)

• Main advantage over fixed basis functions for monotone regression (e.g. monotonic splines)
  – Functional basis itself is learned from data (\( \xi_j \) and \( J \) treated as unknown, not only the basis coefficients)
  – No particular structure for locating the \( w_j, \xi_j \) in parameter space \( \rightarrow \) high flexibility

\[\longleftrightarrow \text{Sparsity: Very small } J \text{ sufficient for fitting smooth functions.}\]

• See Clyde and Wolpert (2007) for more on flexibility/sparsity
MONOTONE REGRESSION: INTERPRETATION

• Prior mean and covariance for $\mu^0(x)$:

$$E(\mu^0(x)) = \int_\Xi F(x, \xi) dP_0$$

$$\text{Cov}(\mu^0(x_1), \mu^0(x_2)) = k_0 \left\{ \int_\Xi F(x_1, \xi) F(x_2, \xi) dP_0 - \int_\Xi F(x_1, \xi) dP_0 \int_\Xi F(x_2, \xi) dP_0 \right\}$$

where $k_0 = E \left( \sum_{j=1}^{J} w_j^2 \right) \in [0, 1]$

• $P_0$ determines prior mean and prior correlation

• $k_0$ is main factor controlling variability
MONOTONE REGRESSION: INTERPRETATION

• If $(w_1, \ldots, w_J)' \sim \text{Dir}_{J-1}(\delta 1), \; \delta > 0$

\[ E(\sum_{j=1}^{J} w_j^2 | J) = \frac{\delta + 1}{J \delta + 1} \]

High variability $\rightarrow$ small $\delta$ or small $J$

• In summary: (i) Parameters have a meaning in the model
  (ii) Priors can be chosen to match information and uncertainty

• In applications: Useful to simulate prior distribution and look at additional summaries of $\mu^0(.)$ (e.g. prior quantiles, maximum increase, smoothness, ...), to calibrate priors.
MONOTONE REGRESSION: IMPLEMENTATION

- Which parametric base cdf \( F(., .) \) should be used?
- \( F(., .) \) should be
  - (i) **Flexible** enough to approximate any continuous cdf
    \[ \text{necessary for full support property} \]
  - (ii) Computationally simple \( \rightarrow \) fast to evaluate
- Quite common in the literature: **beta cdf**
  \[ \rightarrow \text{is known to have approximation property (i)} \]
- Problem: \( F(., .) \) is then regularized **incomplete beta function**, \( i.e. \) not available in closed form \( \rightarrow \) computationally expensive
**MONOTONE REGRESSION: IMPLEMENTATION**

Use “two-sided power” (TSP) distribution (van Dorp and Kotz, 2002)

\[
F(x | \xi) = \begin{cases} 
    m \left( \frac{x}{m} \right)^{\nu} & 0 \leq x \leq m \\
    1 - (1 - m) \left( \frac{1-x}{1-m} \right)^{\nu} & m \leq x \leq 1 
\end{cases}
\]
**MONOTONE REGRESSION: IMPLEMENTATION**

ad (i) For a given continuous cdf $G(.)$ on $[0, 1]$ there exist $J, m_j, \nu_j$ such that

$$\sup_{x \in [0,1]}(|\mu^0(x) - G(x)|) \leq \kappa(J)(1 + 2e^{-1}),$$

where $\mu^0(x) = \sum_{j=1}^{J} w_j F(x, m_j, \nu_j)$ and

$$\kappa(J) := \sup\{0, \ldots, J-1\} \{G\left(\frac{k+1}{J}\right) - G\left(\frac{k}{J}\right)\}$$

ad (ii) **Advantage** over beta cdf: One evaluation (in C++) is around 10 to 15 times faster than the beta cdf implemented in the GSL library.
MONOTONE REGRESSION: IMPLEMENTATION

- RJ-MCMC (for example embedded in a Gibbs sampler) can be used to obtain an approximate sample of the varying dimensional parameter \((J, w_1, \ldots, w_J, \xi_1, \ldots, \xi_J)\)
  → Obtain sample of posterior for \(\mu^0(.)\)

- In the applications we tested and in a simulation study for various test functions, the algorithm works quite well (because of factorization into location and scale parameter?)
Dose Response Analysis
Dose Response analysis

• Dose ranging trial for the treatment of the irritable bowel syndrome (Biesheuvel and Hothorn, 2002), with four active doses (1,2,3,4) and placebo

• Primary continuous endpoint: baseline adjusted abdominal pain score → larger values correspond to better treatment effect

• 369 patients completed study with almost balanced allocations across the dose groups

• Estimate

\[ MED_\Delta = \min_{x \in (0,4)} \{ x : \mu(x) > \mu(0) + \Delta \} \]
**Dose Response analysis**

- Model: \( y_j \sim N(\mu(x_j), \sigma^2), \ x_j \in [0, 1] \) where
  \[
  \mu(x) = \beta_0 + \beta_1 \mu^0(x),
  \]
  with \( \mu^0(x) = \int_{\Xi} F(x, \xi) P(d\xi) \)

- We consider two scenarios:
  
  A **weakly informative** and an **informative** selection of priors

- For informative setting: The variability of the prior for \( \beta_0 \) is set to the information obtained by approximately 30 patients and for \( \beta_1 \) to the information obtained from 3 patients
**Dose Response analysis: Priors**

<table>
<thead>
<tr>
<th></th>
<th>weakly informative</th>
<th>informative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta, \sigma^2$</td>
<td>$\propto \sigma^{-2}$</td>
<td>*</td>
</tr>
<tr>
<td>$J$</td>
<td>$TPoi(0.5)$</td>
<td>$TPoi(0.5)$</td>
</tr>
<tr>
<td>$w</td>
<td>J$</td>
<td>$Dir_{J-1}(1)$</td>
</tr>
<tr>
<td>$P_0$</td>
<td>$U(0,4) \times U(1,70)$ (linear increas. prior mean)</td>
<td>$Beta(0,4,1,2) \times U(1,70)$ (concavely incr. prior mean)</td>
</tr>
</tbody>
</table>

$* \beta_0 \sim t_4(0.21, 0.036)$

$\beta_1 \sim t_4(0.55, 0.36)$

$\sigma^2 \sim IG(3.6, 4)$

Prior specification implies $k_0 \approx 0.91$
POSTERIOR DISTRIBUTION OF $\mu(.)$
**Posterior distribution of** $MED_{0.25}$

<table>
<thead>
<tr>
<th></th>
<th>0.025-Q.</th>
<th>0.25-Q.</th>
<th>Median</th>
<th>0.75-Q.</th>
<th>0.975-Q.</th>
<th>Mean</th>
<th>Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>WI</td>
<td>0.16</td>
<td>0.48</td>
<td>0.76</td>
<td>1.00</td>
<td>3.04</td>
<td>0.90</td>
<td>0.44</td>
</tr>
<tr>
<td>I</td>
<td>0.16</td>
<td>0.44</td>
<td>0.72</td>
<td>0.92</td>
<td>1.96</td>
<td>0.74</td>
<td>0.19</td>
</tr>
</tbody>
</table>

- **Point estimate:** 0.76 resp. 0.72 (median of the posterior)
- For the same data set, the hyperbolic Emax model results in a dose estimate of 0.74 (ML-estimate)
Growth Curves
**GROWTH CURVES**

- Classical example for monotone regression: Growth curve of a boy’s height at 83 days over one school year (Ramsay, 1998)

- Priors: Non-informative for $\beta$, $\sigma^2$,

$$P_0 = (U(0, 320) \times U(1, 70))$$

$$J \sim TPois(1), \mathbf{w} | J \sim Dir_{J-1}(\delta \mathbf{1})$$

- Posterior for $J$: Posterior Mean 4.6, Posterior Variance 1.3
POSTERIOR DISTRIBUTION OF $\mu(.)$
CONCLUSIONS

- Here, a generalization of the Dirichlet process allows for building a flexible, adequate, and sparse BNP model for monotone functions.
- The model allows for incorporation of knowledge on important aspects of the curve (baseline, maximum effect, shape) → very relevant in clinical trials.
- The method performs very reasonable for real data sets and also quite competitive in a simulation study (compared to recent classical approaches).
- Various possibilities to embed this model into more complex hierarchical models (e.g. other endpoints (binary or count data), covariates, ...).
• Extension to multivariate monotonicity (relevant for drug combination trials, epidemiology,...)

• Extensions to convex/concave or unimodal shape constraints (relevant in economics, dose-response analysis, ...)


OVERALL CONCLUSIONS

- Point processes and discrete random probability measures are fairly general mathematical objects

- They may be used directly as models for the data (describing or simulating, for example, point patterns as illustrated in part 1 of the talk)

- They may be used as priors in Bayesian nonparametric inference (as illustrated in part 2 of the talk). Mixtures based on point processes can then be employed as sparse and flexible models for all sorts of functions (densities, intensities, functions possibly with shape constraints) in various dimensions.
ACKNOWLEDGMENTS

• Sibylle Sturtz, IQWiG Institute for Quality and Efficiency in Health Care, Cologne

• Björn Bornkamp & Arno Fritsch, Department of Statistics, TU Dortmund
Winbugs Code for 2.1
model
{
  #priors
  #======
  beta.0 ~ dgamma(a.0, tau.0)
  beta.latent ~ dgamma(a.latent, tau.latent)
  a.0 <- 0.575
  tau.0 <- a.0 * 2 * expect
  a.latent <- 0.575
  tau.latent <- a.latent * 2 * expect

  for (s in 1:Source)
  {
    delta[s] ~ dgamma(a.delta, tau.delta)
  }
  a.delta <- area * tau.delta
  tau.delta <- 1/Source
```r
#Sprungstellen
#================
for(i in 1:Source)
{
  moveX[i] ~ dunif(dist1[i], dist2[i])
  moveY[i] ~ dunif(dist3[i], dist4[i])
  Sx.sourceMove[i] <- Sx.source[i] + moveX[i]
  Sy.sourceMove[i] <- Sy.source[i] + moveY[i]
}
for (sx in 1:Source)
{
  for (i in 1:I)
  {
    distanceX[sx, i] <- abs(wardXcenter[i]-Sx.sourceMove[sx])
    distanceY[sx,i] <- abs(wardYcenter[i]-Sy.sourceMove[sx])
    kernel[sx,i] <- exp(-(pow(distanceX[sx,i]/(2*rhoX[sx]),2)+
                        pow(distanceY[sx,i]/(2*rhoY[sx]), 2)))
  }
  rhoY[sx] ~ dlnorm(0,wert[kY])
  rhoX[sx] ~ dlnorm(0,wert[kX])
}
kX <- step(PX - 0.5) + 1
PX ~ dunif(0,1)
kY <- step(PY-0.5) + 1
PY ~ dunif(0,1)
```
# Intensitäten

for (i in 1:I) {
    count[i] ~ dpois(lambda[i])
    lambda[i] <- p[i]*pop[i]
    latent[i] <- inprod(delta[,], kernel[,i])
    latent.term[i] <- beta.latent*latent[i]
    p[i] <- (beta.0 + latent.term[i])
}
Simulation Study for 2.2
**Simulation Study**

- Compare proposed method with recent classical approaches to nonparametric monotone regression
- **Local linear** regression and monotonization (*monreg* R-package, Dette et al., 2006)
- Constrained optimization with a spline basis (*mgcv* R-package, Wood, 2007)
- Our approach with non-informative prior for $\beta$, $\sigma^2$ and reasonable “default” prior for $\mu^0(.)$ (linear increasing prior mean, with $J \sim TPoi(1)$ and $w \sim Dir(1) \rightarrow k_0 \approx 0.84$)
Simulation Study: Test Functions

- Simulate normal variates with mean $N(\mu_j(i/49), \sigma^2)$, $i = 0, \ldots, 49$, $j = 1, 2, 3$.
- Use $\sigma = 0.05$ and $\sigma = 0.2$
- Repeat every scenario 2,000 times and calculate

$$MAE = \frac{1}{11} \sum_{i=0}^{10} |\mu(i/10) - \hat{\mu}(i/10)|,$$
### Simulation Study: Results

<table>
<thead>
<tr>
<th>Testfct.</th>
<th>$\sigma$</th>
<th>LocLin/Mon</th>
<th>MonRS</th>
<th>MonBayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1(.)$</td>
<td>0.05</td>
<td>0.0216</td>
<td>0.0149</td>
<td>0.0117</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.0628</td>
<td>0.0527</td>
<td>0.0475</td>
</tr>
<tr>
<td>$\mu_2(.)$</td>
<td>0.05</td>
<td>0.0274</td>
<td>0.0172</td>
<td>0.0188</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.0632</td>
<td>0.0582</td>
<td>0.0557</td>
</tr>
<tr>
<td>$\mu_3(.)$</td>
<td>0.05</td>
<td>0.0176</td>
<td>0.0161</td>
<td>0.0212</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.0520</td>
<td>0.0502</td>
<td>0.0561</td>
</tr>
</tbody>
</table>

Table 1: Mean absolute estimation error for compared methods.