Normalized kernel-weighted random measures

Jim Griffin

University of Kent

10 August 2007

Outline

Bayesian Density Regression

We observe data $(x_1, y_1), \ldots, (x_n, y_n)$
and we assume that $y_i \sim F_{x_i}$. We want to estimate F_x for $x \in \mathcal{X}$.

Bayesian Density Regression

We observe data $(x_1, y_1), \ldots, (x_n, y_n)$
and we assume that $y_i \sim F_{x_i}$. We want to estimate F_x for $x \in \mathcal{X}$.

We could build the hierarchical model

$$y_i \sim \upsilon(\psi_i, \phi)$$

$$\psi_i \sim G_{x_i}$$

$$G_{x_i} \stackrel{d}{=} \sum_{i=1}^{\infty} p_i(x_i)\delta_{\theta_i}$$

and $\theta_1, \theta_2, \theta_3, \cdots \stackrel{i.i.d.}{\sim} H$.

Bayesian Density Regression

We observe data $(x_1, y_1), \ldots, (x_n, y_n)$
and we assume that $y_i \sim F_{x_i}$. We want to estimate F_x for $x \in \mathcal{X}$.

We could build the hierarchical model

$$y_i \sim \upsilon(\psi_i, \phi)$$

$$\psi_i \sim G_{x_i}$$

$$G_{x_i} \stackrel{d}{=} \sum_{i=1}^{\infty} p_i(x_i)\delta_{\theta_i}$$

and $\theta_1, \theta_2, \theta_3, \cdots \stackrel{i.i.d.}{\sim} H$.

Then F_x can be estimated by $\mathrm{E}_{G,\phi|y}[\int k(y_i|\psi, \phi)\, dG_x(\psi)]$.

Bayesian Density Regression

- We would like G_x to be stationary.

Bayesian Density Regression

- We would like G_x to be stationary.
- ... and to have a way of controlling the dependence between G_x and G_y.

Possible approaches

Usually we generalize standard construction of priors for exchangeable sequences:

Possible approaches

Usually we generalize standard construction of priors for exchangeable sequences:

- Dirichlet process - DDP (MacEachern 1999)

Possible approaches

Usually we generalize standard construction of priors for exchangeable sequences:

- Dirichlet process - DDP (MacEachern 1999)
- Stick-breaking - πDDP (Griffin and Steel, 2006), Kernel-weighted stick-breaking (Dunson and Park, 2006)

Possible approaches

Usually we generalize standard construction of priors for exchangeable sequences:

- Dirichlet process - DDP (MacEachern 1999)
- Stick-breaking - πDDP (Griffin and Steel, 2006), Kernel-weighted stick-breaking (Dunson and Park, 2006)
- Pólya urn scheme (Caron et al, 2007)

Normalized random measures

We could extend the class of normalized random measures (Regazzini *et al* 2003, James *et al* 2005)).

Normalized random measures

We could extend the class of normalized random measures (Regazzini *et al* 2003, James *et al* 2005)).

Let (J, θ) follow an homogeneous Poisson process on $\mathbb{R}^+ \times \Theta$ with intensity $\kappa(J)h(\theta)$ and define

$$G = \frac{\sum_{i=1}^{\infty} J_i \delta_{\theta_i}}{\sum_{i=1}^{\infty} J_i}$$

Normalized random measures

We could extend the class of normalized random measures (Regazzini *et al* 2003, James *et al* 2005)).

Let (J,θ) follow an homogeneous Poisson process on $\mathbb{R}^+ \times \Theta$ with intensity $\kappa(J)h(\theta)$ and define

$$G = \frac{\sum_{i=1}^{\infty} J_i \delta_{\theta_i}}{\sum_{i=1}^{\infty} J_i}$$

then G follows a (homogeneous) NRM under suitable conditions for κ we have a random probability measure (infinite activity) and h is the density of the centring distribution.

Examples of NRMs

- Dirichlet process - Normalized gamma process

$$\kappa(J) = M \frac{\exp\{-J\}}{J}$$

- Normalized Generalized Gamma process

$$\kappa(J) = \frac{\gamma}{\Gamma(1-\gamma)} J^{-\gamma} \exp\{-rJ\|$$

Normalized kernel-weighted measures

Let (τ, J, θ) follow an homogeneous Poisson process on $\mathcal{X} \times \mathbb{R}^{+} \times \Theta$ with intensity $\kappa(J)h(\theta)$ and define

$$G_x = \frac{\sum_{i=1}^{\infty} k(x,\tau_i) J_i \delta_{\theta_i}}{\sum_{i=1}^{\infty} k(x,\tau_i) J_i}$$

for some kernel function $k(x, \tau_i)$ centred at τ_i

Normalized kernel-weighted measures

Let (τ, J, θ) follow an homogeneous Poisson process on $\mathcal{X} \times \mathbb{R}^{+} \times \Theta$ with intensity $\kappa(J)h(\theta)$ and define

$$G_x = \frac{\sum_{i=1}^{\infty} k(x, \tau_i) J_i \delta_{\theta_i}}{\sum_{i=1}^{\infty} k(x, \tau_i) J_i}$$

for some kernel function $k(x, \tau_i)$ centred at τ_i

For modelling, we wish to control

- Dependence between G_x and G_y. In these process, for a measureable set B, we can measure correlation through $\text{Corr}(G_x(B), G_y(B))$ which usually won't depend on B.
- The marginal prior of G_x for all x.

Normalized kernel-weighted measures

Dependence

The correlation of the unnormalized random measures is

$$\frac{\int k(x,\tau)k(y,\tau)\,d\tau}{\int k(x,\tau)^2\,d\tau}$$

This correlation will typically carry over to the normalized version unless we have a marginal processs that gives distributions with a few large jumps.

Normalized kernel-weighted measures

Dependence

The correlation of the unnormalized random measures is

$$\frac{\int k(x,\tau)k(y,\tau)\,d\tau}{\int k(x,\tau)^2\,d\tau}$$

This correlation will typically carry over to the normalized version unless we have a marginal processs that gives distributions with a few large jumps.

Stationarity

The form of κ can be derived to give particular marginal processes.

Ornstein-Uhlenbeck Dirichlet Process

With a 1D regressor, typically time, we fix the kernel function to be

$$k(x, \tau) = \exp\{-\lambda(x - \tau)\}\mathrm{I}(x > \tau).$$

and assume a marginal Dirichlet process.

Ornstein-Uhlenbeck Dirichlet Process

With a 1D regressor, typically time, we fix the kernel function to be

$$k(x,\tau) = \exp\{-\lambda(x-\tau)\}\mathrm{I}(x>\tau).$$

and assume a marginal Dirichlet process.

The unnormalized process must be a Gamma process.

Ornstein-Uhlenbeck Dirichlet Process

With a 1D regressor, typically time, we fix the kernel function to be

$$k(x, \tau) = \exp\{-\lambda(x - \tau)\}\mathrm{I}(x > \tau).$$

and assume a marginal Dirichlet process.

The unnormalized process must be a Gamma process.

The ideas of Barndorff-Nielsen and Shephard are useful to define this process. Let $\phi_1, \phi_2, \phi_3, \ldots$ are i.i.d. exponential (1) and $\tau_1, \tau_2, \tau_3, \ldots$ follow a Poisson process with intensity $M\lambda$ then

$$\gamma_t = \sum_{i=1}^{\infty} \mathrm{I}(\tau_i < t) \exp\{-\lambda_i \tau_i\} \phi_i$$

is Ga$(M, 1)$ distributed for all t.

Definition of OUDP

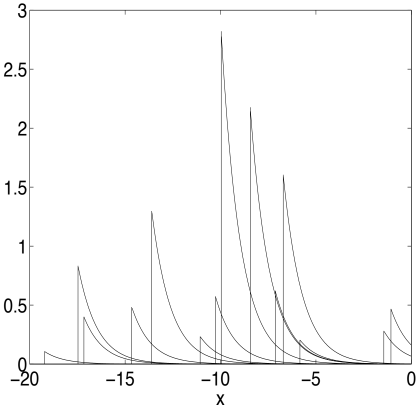
This is a construction when the covariate x is time. Define

$$G_x = \frac{\sum_{i=1}^{\infty} I(\tau_i < x)\exp\{-\lambda(x-\tau_i)\}J_i\delta_{\theta_i}}{\sum_{i=1}^{\infty} I(\tau_i < x)\exp\{-\lambda(x-\tau_i)\}J_i}$$

- τ follows a Poisson process with intensity λM.
- $J_1, J_2, J_3, \dots \overset{i.i.d.}{\sim} \mathrm{Ex}(1)$
- $\theta_1, \theta_2, \theta_3, \dots \overset{i.i.d.}{\sim} H$

or

(τ, J, θ) follows a Poisson process with intensity $\lambda M \exp\{-J\} h$

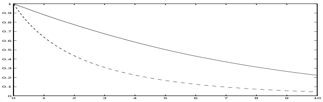


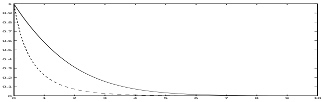
The autocorrelation at lag k is approximately

$$\exp\{-\lambda k\}\left[1+\frac{1}{M}(1-\exp\{-\lambda k\}\right]$$

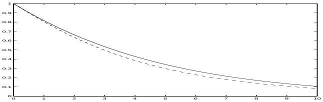
$\lambda = 0.25$ $\lambda = 1$

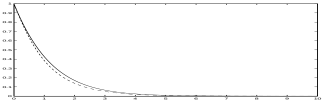
$M = 1$





$M = 4$





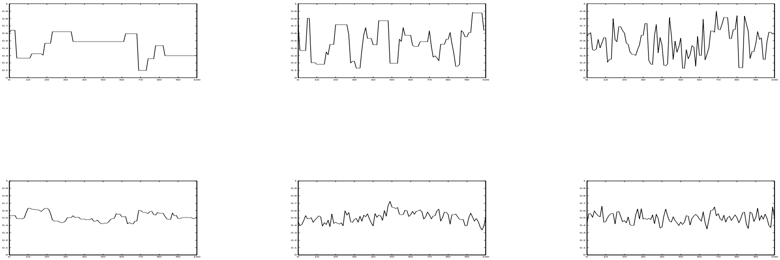
Dynamics of moments

The dynamics of the mean are $\mu_t = w_t \mu_{t-1} + (1 - w_t)\mu_G$

$\lambda = 0.125$ $\lambda = 0.5$ $\lambda = 2$

$M = 1$

$M = 16$



Computation

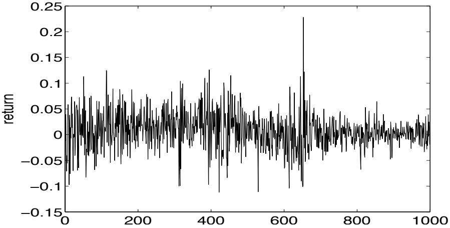
The stationarity of the process makes inference possible using fairly standard methods

$$G_t = \frac{\exp\{-\lambda t\}\gamma_0}{\exp\{-\lambda t\}\gamma_0 + \sum_{i=1}^m \exp\{-\lambda(t-\tau_i)\}J_i} G_0 + \frac{\sum_{i=1}^m \exp\{-\lambda(t-\tau_i)\}J_i}{\exp\{-\lambda t\}\gamma_0 + \sum_{i=1}^m \exp\{-\lambda(t-\tau_i)\}J_i}$$

where G_0 follows a Dirichlet process and γ_0 follows a gamma distribution with shape parameter M. Inference using:

- Gibbs sampling
- Particle filtering

Example - Brazilian stock index



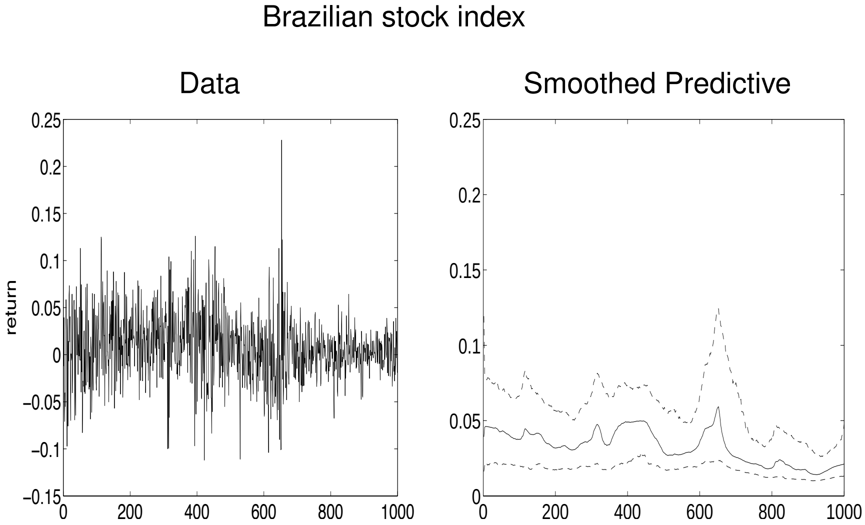
We observe $r_1, r_2, \ldots, r_T$ which are daily log returns and let

$$r_t | \sigma_t^2 \sim \mathrm{N}(0, \sigma_t^2)$$

$$\sigma_t^2 \sim F_t$$

where $\{F_t\}_0^T$ follows an OUDP, centred on an inverse Gaussian distribution, whose parameters are estimated from the marginal distribution of the data.

Example


Brazilian stock index
Data
Smoothed Predictive
return
0.25
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
0
200
400
600
800
1000

Generalizing to other marginal processes

For other marginal processes, let $w(a)$ be the Lévy density of the unnormalized marginal process.

Generalizing to other marginal processes

For other marginal processes, let $w(a)$ be the Lévy density of the unnormalized marginal process.

The intensity of the Poisson process of the unnormalized process with the kernel will be $w^\star(J)h(\theta)$ where $w^\star(J) = \lambda J w(J)$.

Generalizing to other marginal processes

For other marginal processes, let $w(a)$ be the Lévy density of the unnormalized marginal process.

The intensity of the Poisson process of the unnormalized process with the kernel will be $w^\star(J)h(\theta)$ where $w^\star(J) = \lambda J w(J)$.

A marginal NGG process arising from assuming the intensity $\kappa(J) = \frac{\gamma\lambda}{\Gamma(1-\gamma)} J^{1-\gamma} \exp\{-rJ\|$ which is a finite activity Poisson process.

In general, if we define a kernel $K(x, \tau)$ then the two measures are linked by the integral equation

$$\int_a^\infty w(J)\,dJ = \int_a^\infty w^\star(J)\nu(\{\tau | K(0,\tau) > a/J\})\,dJ$$

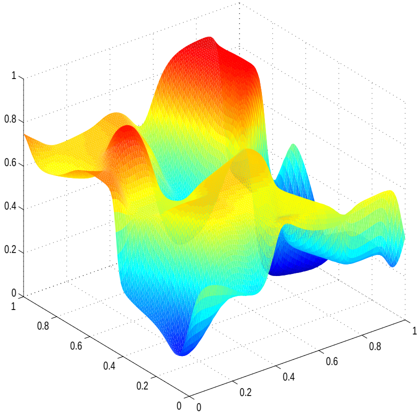
where ν is Lesbesgue measure.

This is a Volterra integral equation and can be solved using standard methods (in principle).

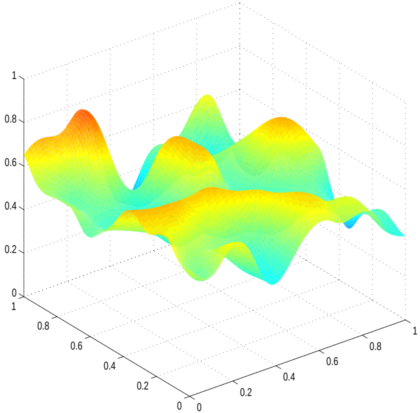
Generalizing to other kernels

In 2D, if the kernel $k(x, \tau) = \exp\{-\lambda|x - t|^2\}$ and we want a marginal Dirichlet process then the intensity function is $\frac{\lambda}{\pi}\exp\{-J\}h(\theta)$ (which is proportional to the intensity function for the OUDP)

$M = 1$



$M = 5$



Discussion

Normalized Kernel-Weighted Random Measures offer a way to model dependent nonparametric processes:

- Flexible kernels and marginal processes allow a large range of models to be defined.
- Computation is helped by representations through finite activity Poisson processes for some elements.
- and include continuous process on the space of measures.