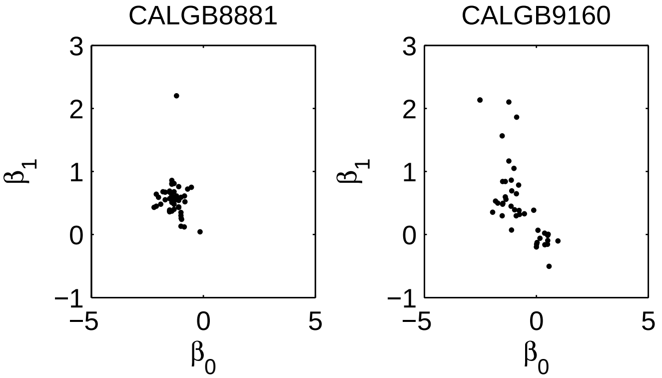
Compound Random Measures

Jim Griffin
(joint work with Fabrizio Leisen)

University of Kent

Introduction: Two clinical studies



Infinite mixture models

This data could be analysed using two infinite mixture models

$$\text{CALGB8881: } f_1(y) = \sum_{j=1}^{\infty} w_j^{(1)} k(y|\theta_j)$$

and

$$\text{CALGB9160: } f_2(y) = \sum_{j=1}^{\infty} w_j^{(2)} k(y|\theta_j)$$

where

- $w_j^{(k)} > 0$ for $j = 1, 2, \ldots$ and $\sum_{j=1}^{\infty} w_j^{(k)} = 1$ for $k = 1, 2$.
- $k(y|\theta)$ is a p.d.f. for y with parameter θ.

We need to put a prior on the $w^{(1)}$, $w^{(2)}$ and θ (random probability measure).

Some dependent random probability measures: stick-breaking

θ are i.i.d. and

$$w_j^{(k)} = V_j^{(k)} \prod_{i<j} \left(1 - V_i^{(k)}\right)$$

- Hierarchical Dirichlet Process (Teh *et al*, 2006): $V_j^{(k)} \sim \text{Be}\left(\alpha_0 \beta_j, \alpha_0 \left(1 - \sum_{l=1}^{j} \beta_l\right)\right)$,

$$\beta_j' \sim \text{Be}(1, \gamma), \qquad \beta_j = \beta_j' \prod_{l=1}^{j-1} (1 - \beta_l'),$$

- Probit stick-breaking processes, etc.: $\left(V_j^{(1)}, V_j^{(2)}\right)$ are correlated and independent of $\left(V_i^{(1)}, V_i^{(2)}\right)$ for $i \neq j$.

University of Kent

Completely random measures

$\tilde{\mu}$ is a completely random measure (CRM) on Θ if, for any disjoint subsets $A_1, \ldots, A_n$, $\tilde{\mu}(A_1) \ldots, \tilde{\mu}(A_n)$ are mutually independent.

We concentrate on completely random measures (CRM's) which can be represented in terms of jump sizes J_i and jump locations θ_i as

$$\tilde{\mu} = \sum_{i=1}^{\infty} J_i \delta_{\theta_i}$$

where δ is Dirac's delta function and have Lévy-Khintchine representation

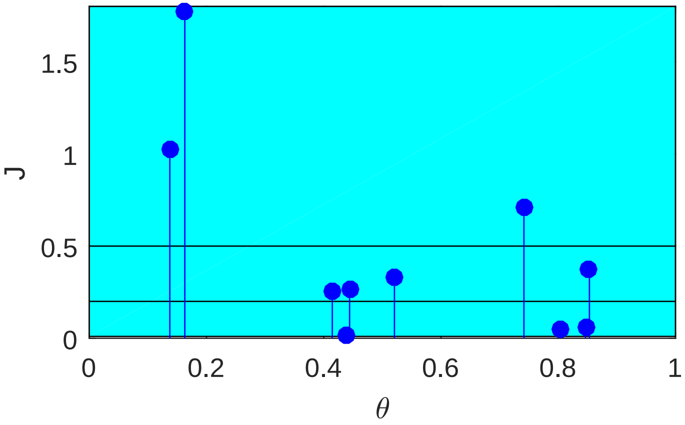
$$\mathbb{E}\left[e^{-\int f(\theta)\tilde{\mu}(d\theta)}\right] = e^{-\int_0^\infty \int \left[1-e^{-sf(\theta)}\right]\alpha(d\theta)\rho(ds)}$$

where α and ρ are measures for which $\int \alpha(d\theta) < \infty$.

University of Kent

Completely random measures

Poisson process with intensity $\alpha(d\theta)\rho(ds)$.



Examples of CRM's

Many processes that we use in Bayesian nonparametrics are CRM's

- Gamma process - $\rho(ds) = s^{-1} \exp\{-s\}\, ds$.
- Beta process - $\rho(ds) = \beta s^{-1}(1-s)^{\beta-1}\, ds$.

or can be derived from CRM's

- Normalizing a Gamma process, *i.e.* taking $\tilde{p} = \tilde{\mu}/\tilde{\mu}(\Theta)$, leads to a Dirichlet process.
- A beta process prior for $p_1, p_2, \ldots$ can be used to define an Indian buffet process.

University of Kent

Vectors of CRMs

It is useful to define d related CRM's.

Suppose that $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are CRM's on Θ with marginal Lévy intensities $\bar{\nu}_j(ds, d\theta) = \nu_j(ds)\alpha(d\theta)$

Then $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are a vector of CRM's if there is a Lévy-Khintchine representation of the form

$$\mathbb{E}\left[e^{-\tilde{\mu}(f_1)-\cdots-\tilde{\mu}_d(f_d)}\right] = e^{-\psi^{\star}_{\rho,d}(f_1,\ldots,f_d)}$$

where

$$\psi^{\star}_{\rho,d}(f_1, \ldots, f_d) = \int_{(\mathbb{R}^+)^d} \int \left[1 - e^{-s_1 f_1(\theta)-\cdots-s_d f_d(\theta)}\right] \alpha(d\theta)\rho_d(ds_1, \ldots ds_d)$$

and

$$\nu_j(ds) = \int \rho_d(ds_1, \ldots, ds_d).$$

Compound Random Measures: Definition

A compound random measure (CoRM) is a vector of CRM's with intensity

$$\rho_d(ds_1,\ldots,ds_d)=\int z^{-d}h\left(\frac{s_1}{z},\ldots,\frac{s_d}{z}\right)\,ds_1\ldots ds_d\,\nu^\star(dz)$$

where

- $s_1,\ldots,s_d$ are called scores.
- H is a score distribution with density h.
- $\nu^\star$ is the Lévy intensity of a directing Lévy process.

which satisfies the condition $\int \min(1,\| s \|)z^{-d}h\left(\frac{s_1}{z},\ldots,\frac{s_d}{z}\right)\nu^\star(dz)<\infty$ where $\| s \|$ is the Euclidean norm of the vector $\mathbf{s}=(s_1,\ldots,s_d)$.

University of Kent

A representation of a CoRM

Realizations of a CoRM can be expressed as

$$\tilde{\mu}_j = \sum_{i=1}^{\infty} m_{j,i} \, J_i \, \delta_{\theta_i}$$

where

- $m_{1,i}, \ldots, m_{d,i} \overset{i.i.d.}{\sim} H$
- $\tilde{\eta} = \sum_{i=1}^{\infty} J_i \, \delta_{\theta_i}$ is a CRM with Lévy intensity $\nu^{\star}(ds) \, \alpha(d\theta)$.

CoRMs with independent gamma scores

We will concentrate on the class of CoRMs for which

$$h(s_1/z, \ldots, s_d/z) = \prod_{j=1}^{d} f(s_j/z)$$

where f is the p.d.f. of a gamma distribution with shape ϕ, $f(x) = \frac{1}{\Gamma(\phi)} x^{\phi-1} \exp\{-x\}$.

University of Kent

Properties of CoRMs with independent score distributions

- The Lévy copula can be expressed as a univariate integral.
- Let $M_z^f(t) = \int e^{ts} z^{-1} f(s/z) ds$ be the moment generating function of $z^{-1} f(s/z)$ then

$$\psi_{\rho,d}(\lambda_1, \dots, \lambda_d) = \int_{(\mathbb{R}^+)^d} \int \left[1 - e^{-s_1\lambda_1 - \dots - s_d\lambda_d}\right] \rho_d(ds_1, \dots ds_d)$$

$$= \psi_{\rho,d}(\lambda_1, \dots, \lambda_d) = \int \left(1 - \prod_{j=1}^{d} M_z^f(-\lambda_j)\right) \nu^\star(z) dz$$

- This expression can be used to calculate quantities such as $\text{Corr}(\tilde{\mu}_k(A), \tilde{\mu}_m(A))$.

University of Kent

CoRMs with gamma distributed scores

Consider a CoRM process with independent Ga$(\phi, 1)$ distributed scores. If the CoRM process has gamma process marginals then

$$\rho_d(s_1, \ldots, s_d) = \frac{\left(\prod_{j=1}^{d} s_j\right)^{\phi-1}}{[\Gamma(\phi)]^{d-1}} |\boldsymbol{s}|^{-\frac{d\phi+1}{2}} e^{-\frac{|\boldsymbol{s}|}{2}} W_{\frac{(d-2)\phi+1}{2}, -\frac{d\phi}{2}}(|\boldsymbol{s}|) \tag{1}$$

where $|\boldsymbol{s}| = s_1 + \cdots + s_d$ and W is the Whittaker function. If the CoRM process has σ-stable process marginals then

$$\rho_d(s_1, \ldots, s_d) = \frac{\left(\prod_{j=1}^{d} s_j\right)^{\phi-1}}{[\Gamma(\phi)]^{d-1}} \frac{\sigma\Gamma(\sigma + d\phi)}{\Gamma(\sigma+\phi)\Gamma(1-\sigma)} |\boldsymbol{s}|^{-\sigma-d\phi}. \tag{2}$$

University of Kent

CoRMs with exponentially distributed scores

Consider a CoRM process with independent exponentially distributed scores. If the CoRM has gamma process marginals we recover the multivariate Lévy intensity of Leisen *et al* (2013),

$$\rho_d(s_1, \ldots, s_d) = \sum_{j=0}^{d-1} \frac{(d-1)!}{(d-1-j)!} |\boldsymbol{s}|^{-j-1} e^{-|\boldsymbol{s}|}.$$

Otherwise, if σ-stable marginals are considered then we recover the multivariate vector introduced in Leisen and Lijoi (2011) and Zhu and Leisen (2014),

$$\rho_d(s_1, \ldots, s_d) = \frac{(\sigma)_d}{\Gamma(1-\sigma)} |\boldsymbol{s}|^{-\sigma-d}.$$

CoRMs with independent gamma scores: specific marginals

The Lévy intensity of $\tilde{\mu}_j$

$$\nu_j(ds) = \int z^{-1} f(s/z) ds \, \nu^\star(dz) = \nu(ds).$$

If we have independent gamma scores, the directing Lévy intensity $\nu^\star$ is linked to the marginal Lévy intensity by

$$\nu^\star\left(\frac{1}{t}\right) = t^{2-\phi} \mathcal{L}^{-1}\left(\frac{\Gamma(\phi)}{s^{\phi-1}} \nu(s)\right)(t)$$

where $\mathcal{L}^{-1}$ is the inverse Laplace transform.

University of Kent

CoRMs with independent gamma scores: specific marginals

The intensity of the directing Lévy process is

$$\nu^\star(z) = z^{-1}(1-z)^{\phi-1}, \qquad 0 < z < 1$$

leads to a marginal gamma process for which

$$\nu(s) = s^{-1}\exp\{-s\}, \qquad s > 0$$

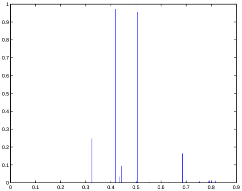
Remarks

- $\nu^\star$ is the the Lévy intensity of a beta process.
- If $\nu^\star$ is the Lévy intensity of a Stable-Beta process (Teh and Görür, 2009), the marginal process is a generalized gamma process.

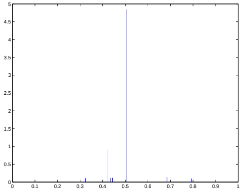
University of Kent

NCoRM: Gamma marginal, $\phi = 1$

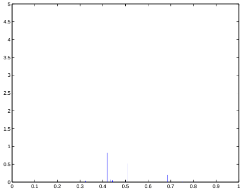
DLP



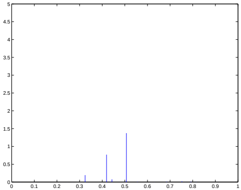
Dim. 1



Dim. 2

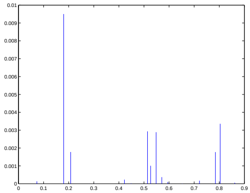


Dim. 3

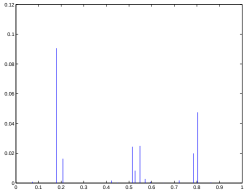


NCoRM: Gamma marginal, $\phi = 10$

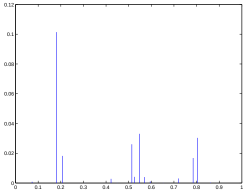
DLP



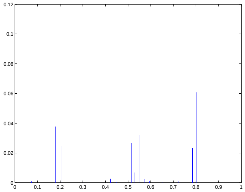
Dim. 1



Dim. 2



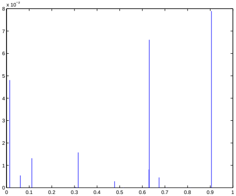
Dim. 3



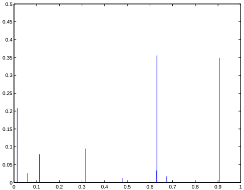
University of Kent

NCoRM: Gamma marginal, $\phi = 50$

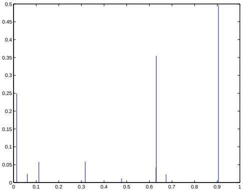
DLP



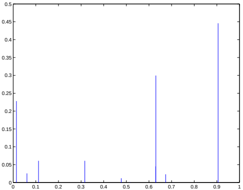
Dim. 1



Dim. 2



Dim. 3



University of Kent

Marginal beta process

A CoRM with beta process marginals ($\nu(s) = \beta s^{-1}(1-s)^{\beta-1}$) can be constructed using

- A beta score distribution with parameters α and 1
- A directing Lévy intensity

$$\nu^\star(z) = \beta z^{-1}(1-z)^{\beta-1} + \frac{\beta(\beta-1)}{\alpha}(1-z)^{\beta-2}$$

i.e. a superposition of a beta process and a compound Poisson process with beta jump distribution.

University of Kent

Links to other processes

Other processes can be expressed as CoRM's:

- Superpositions/Thinning: *e.g.* Griffin *et al* (2013), Chen *et al* (2014), Lijoi and Nipoti (2014), Lijoi *et al* (2014a, b) using mixture score distributions

$$h(s) = \pi\delta_{s=0} + (1-\pi)h^{\star}(s).$$

- Lévy copulae: *e.g.* Leisen and Lijoi (2011), Leisen *et al* (2013), Zhu and Leisen (2014).

Normalized Compound Random Measures (NCoRM)

A vector of random probability measures can be defined by normalizing each dimension of the CoRM so that

$$p_k = \frac{\tilde{\mu}_k}{\tilde{\mu}_k(\Theta)} = \sum_{j=1}^{\infty} w_j^{(k)} \delta_{\theta_j}.$$

CoRMs on more general spaces

For a more general space $\mathbb{X}$, we define $\tilde{\mu}(\cdot; x)$ to be a completely random measure for $x \in \mathbb{X}$.

The collection $\{\tilde{\mu}(\cdot; x) | x \in \mathbb{X}\}$ can be given a CoRM prior with

$$\tilde{\mu}(\cdot; x) = \sum_{j=1}^{\infty} m_j(x) \, J_j \, \delta_{\theta_j}$$

where $m_k(x)$ is a realisation of a random process on $\mathbb{X}$.

Example
$\mathbb{X} = \mathbb{R}^p$, $m_k(x) = \exp\{r_k(x)\}$ where $r_k(x)$ is given a zero-mean Gaussian process prior (see Ranganath and Blei, 2015).

University of Kent

Inference with NCoRM's

We assume that the data are $(x_1, y_1), \ldots, (x_n, y_n)$ and are modelled as

$$y_i|\zeta_i \overset{ind.}{\sim} k(y_i|\zeta_i), \quad \zeta_i \sim p(\cdot; x_i) = \frac{\tilde{\mu}(\cdot; x_i)}{\tilde{\mu}(\Theta; x_i)}, \quad i = 1, 2, \ldots, n$$

where $k(y|\theta)$ is a probability density function for y with parameter θ and $\{p(\cdot; x)|x \in \mathbb{X}\}$ is given an NCoRM prior.

University of Kent

MCMC inference for infinite mixture models

Introducing allocation variables $c_1, \ldots, c_n$, the posterior is proportional

$$p(y, c|m, J, \theta) = \left[\prod_{i=1}^{n} k\left(y_i|\theta_{c_i}\right) \frac{J_{c_i}\, m_{c_i}(x_i)}{\sum_{l=1}^{\infty} J_l\, m_l(x_i)}\right].$$

This form is not tractable due to the infinite sum in the denominator of each term. This can be addressed using the identity

$$\frac{1}{\sum_{l=1}^{\infty} J_l\, m_l(x_i)} = \int_0^{\infty} \exp\left\{-v_i \sum_{l=1}^{\infty} J_l\, m_l(x_i)\right\} dv_i$$

University of Kent

MCMC inference for infinite mixture models

Introducing latent variables v_i leads to a suitable form of augmented posterior for MCMC

$$
\begin{aligned}
& p(y, c, v | m, J, \theta) \\
&= \prod_{i=1}^{n} \left[k\left(y_i|\theta_{c_i}\right) J_{c_i}\, m_{c_i}(x_i) \exp\left\{ -v_i \sum_{l=1}^{\infty} J_l\, m_l(x_i) \right\} \right] \\
&= \prod_{j=1}^{K} \left[\prod_{\{i|c_i=j\}} k\left(y_i|\theta_j\right) J_j^{a_j} \prod_{\{i|c_i=j\}} m_j(x_i) \right] \exp\left\{ -\sum_{l=1}^{\infty} J_l \sum_{i=1}^{n} v_i\, m_l(x_i) \right\}
\end{aligned}
$$

where there are K distinct values of c_i and $a_j = \sum_{i=1}^{n} \mathrm{I}(c_i = j)$.

University of Kent

MCMC inference for infinite mixture models: Finite $\mathbb{X}$, independent scores

In this case, we can define a marginal sampler (*e.g.* Favaro and Teh, 2013) by integrating over J and m.

- $\int J^a \, \nu^\star(J) \, dJ$ is typical for marginal samplers of normalized random measure mixtures.
- Integrals of $\prod_{\{i|c_i=j\}} m_j(x_i)$ will be a product of moments of the scored distribution.
- $\mathrm{E}[\exp\left\{-\sum_{l=1}^{\infty} J_l \sum_{i=1}^{n} v_i \, m_l(x_i)\right\}]$ can be evaluated either exactly or as a univariate integral.

University of Kent

MCMC inference for infinite mixture models: General $\mathbb{X}$

Pseudo-marginal methods (Andrieu and Roberts, 2009) are useful for a target density of the form

$$\pi(\theta) \propto f(\theta)\, g(\theta)$$

where $g(\theta)$ cannot be directly evaluated.

Samples from the target density

$$\hat{\pi}(\theta) \propto f(\theta)\, \hat{g}(\theta)$$

where $\mathrm{E}[\hat{g}(\theta)] = g(\theta)$ will have the distribution π.

In our target, the problem is evaluating
$\mathrm{E}[\exp\left\{-\sum_{l=1}^{\infty} J_l \sum_{i=1}^{n} v_i\, m_l(x_i)\right\}] = \exp\{-\psi(v)\}$

Unbiased estimation of the Laplace transform

The Poisson estimator (see Papaspiliopoulos, 2011) of $L_\phi = \exp\left\{-\int_D \phi(x)\,dx\right\}$ is

$$\hat{L}_\phi = \prod_{i=1}^{K}\left(1 - \frac{\phi(x_i)}{a\,C\,\kappa(x_i)}\right)$$

where κ is a p.d.f. on D, $C > \frac{\phi(x)}{\kappa(x)}$ for $x \in D$, $a > 1$, $K \sim \text{Pn}(a\,C)$ and $x_i \stackrel{i.i.d.}{\sim} \kappa$. Then,

$$\text{E}[\hat{L}_\phi] = \exp\left\{-\int_D \phi(x)\,dx\right\}$$

and

$$\text{V}[\hat{L}_\phi] = L_\phi^2\left(\exp\left\{\frac{1}{a\,C}\int_D \frac{\phi(x)^2}{\kappa(x)}\,dx\right\} - 1\right) < \infty.$$

University of Kent

Unbiased estimation of $\exp\{-\psi(v)\}$

Assuming that $x_1, x_2, \ldots, x_n$ are distinct, $m_i^\star = m(x_i)$ and $m^\star = (m_1^\star, \ldots, m_n^\star)$, $\exp\{-\psi_{\rho,d}(v)\}$ can be re-expressed as

$$\exp\left\{-\int_{(\mathbb{R}^+)^n}\int_0^\infty \left(1-\exp\left\{-z\sum_{i=1}^n v_i\, m_i^\star\right\}\right) h(m^\star)\,\nu^\star(z)\,dz\,dm^\star\right\} = \prod_{k=1}^n L_k$$

where

$$L_k = \exp\left\{-\int_{(\mathbb{R}^+)^n}\int_0^\infty v_k\, m_k^\star\, h(m^\star)\exp\left\{-t\sum_{i=1}^n v_i\, m_i^\star\right\} T_{\nu^\star}(t)\,dt\,dm^\star\right\}$$

and $T_{\nu^\star}(t) = \int_t^\infty \nu^\star(z)\,dz$ (tail mass function).

Unbiased estimation of the Laplace transform

L_k can be estimated using the Poisson estimator with $x = (z, m_k^\star)$, $D = (0, \infty) \times (\mathbb{R}^+)^n$ and

$$\phi(z, m_k^\star) = v_k\, m_k^\star\, h(m_k) \exp\left\{ -t \sum_{i=1}^{n} v_i\, m_k^\star \right\} T_{\nu^\star}(t) < \infty.$$

A suitable approximating density is

$$\kappa(z, m_k^\star) = \kappa_{\tilde{\nu}}(z) \frac{m_k^\star\, h(m_k^\star)}{\mathrm{E}[m_k^\star]}$$

where $\kappa_\nu(z) > T_\nu(z)$ for all $z \in \mathbb{R}^+$.

University of Kent

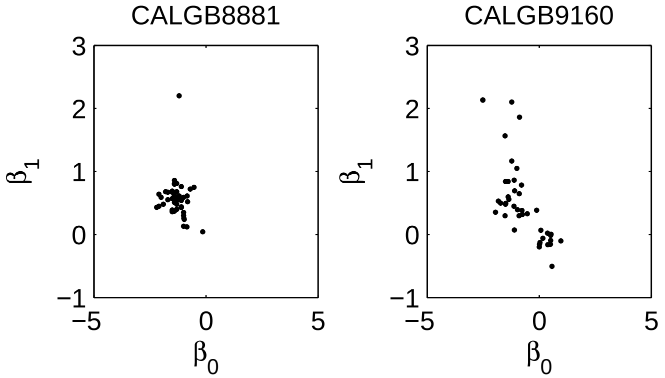
A sampler for more general processes

A pseudo-marginal sampler is used with

- $\exp\{-\psi_{\rho,d}(v)\}$ estimated by the Poisson estimator.
- The jumps are not integrated out and values for empty clusters are proposed from $h(m, J) \propto h(m_1/z, \ldots, m_K/z) z \exp\{-vz\}\nu^\star(z)$.
- An interweaving scheme for m and z (Yu and Meng, 2011).

University of Kent

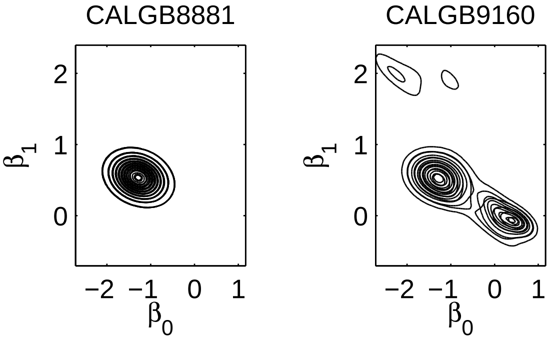
Two clinical studies



University of Kent

Two clinical studies: Posterior mean densities

Results using a CoRM with independent gamma scores.



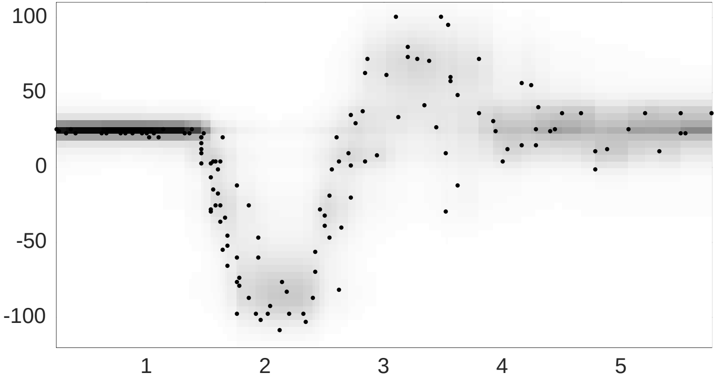
Example: Nonparametric regression

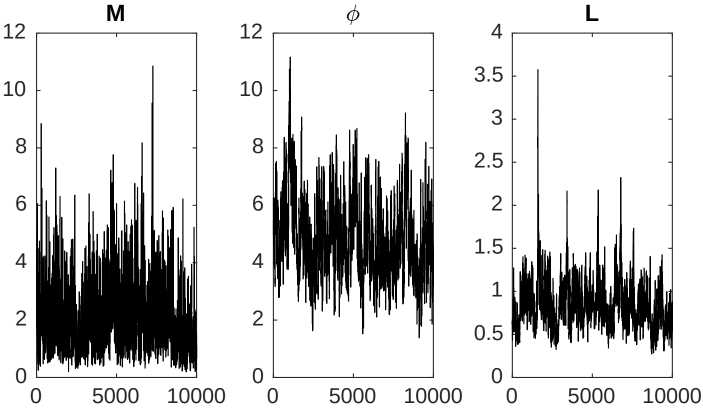
We consider the classic motorcycle data which records head acceleration at different times after impact.

$$f(y) = \sum_{j=1}^{\infty} w_j(x) \mathrm{N}(y|\mu_j, \sigma_j^2)$$

where

- $w_j = \frac{\exp\{r_k(x)\} J_k}{\sum_{m=1}^{\infty} \exp\{r_m(x)\} J_m}$
- $r_m(x)$ are given independent Gaussian process prior with squared exponential covariance function.
- $J_1, J_2, \ldots$ follow a Gamma process with Lévy intensity $M\,x^{-1} \exp\{-x\}$.


100
50
0
-50
-100
1
2
3
4
5



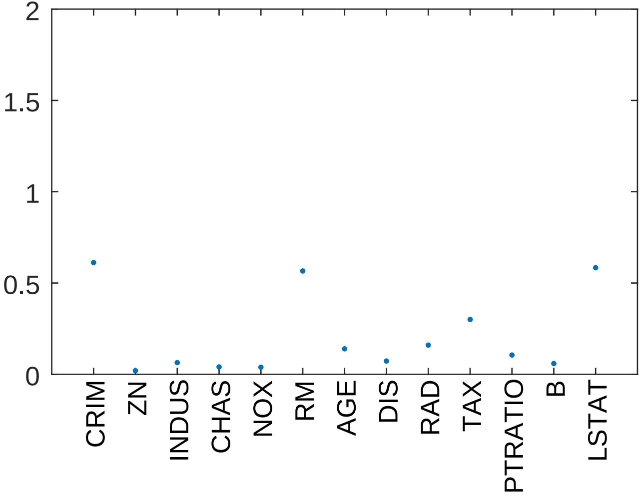
University of Kent

Example: Nonparametric variable selection

The classic Boston housing data record the median value of owner-occupied homes in 506 areas of Boston and the values of 14 attributes that are thought to effect house prices.

The covariance function $k(x, x') = \exp\{-\sum_{i=1}^{p} w_j (x_j - x_j')^2\}$ and $p(w_j) \propto (1 + w_j)^{-1}$.

Posterior median and 95% credible intervals for w_j



Summary

- CoRM processes are a unifying framework for a wide-range of proposed vectors of CRMs.
- CoRM process are vectors of CRM's which are constructed in terms of a (univariate) CRM and a distribution (which defines the dependence).
- Several MCMC methods for NCoRM mixture models are developed. These include methods which depend on the availability of analytical forms for some integrals with respect to the score distribution and methods which do not.
- Modelling dependence through distributions allows a wide-range of dependent nonparametric models to be developed (*e.g.* regression, time series, etc.).

University of Kent

References

C. Andrieu and G. O. Roberts (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, **37**, 697–725.

S. Favaro and Y. W. Teh (2013). MCMC for Normalized Random Measure Mixture Models. *Statistical Science*, **28**, 335–359.

J. E. Griffin, M. Kolossiatis and M. F. J. Steel (2013). Comparing Distributions By Using Dependent Normalized Random-Measure Mixtures. *Journal of the Royal Statistical Society, Series B*, **75**, 499–529.

F. Leisen and A. Lijoi (2011). Vectors of Poisson-Dirichlet processes. *J. Multivariate Anal.*, **102**, 482–495.

F. Leisen, A. Lijoi and D. Spano (2013). A Vector of Dirichlet processes. *Electronic Journal of Statistics* **7**, 62–90.

A. Lijoi, and B. Nipoti (2014), A class of hazard rate mixtures for combining survival data from different experiments. *Journal of the American Statistical Association*, **109**, 802–814.

A. Lijoi, B. Nipoti and I. Prünster (2014a), Bayesian inference with dependent normalized completely random measures, *Bernoulli*, **20**, 1260–1291.

A. Lijoi, B. Nipoti and I. Prünster (2014b), Dependent mixture models: clustering and borrowing information, *Computational Statistics and Data Analysis*, **71**, 417–433.

University of Kent

References

O. Papaspiliopoulos (2011). A methodological framework for Monte Carlo probabilistic inference for diffusion processes. In *Bayesian Time Series Models* (D. Barber, A. Taylan Cemgil and S. Chippia, Eds), Cambridge University Press.

R. Ranganath and D. M. Blei (2015). Correlated Random Measures. *arXiv:1507.00720*

Y. W. Teh and D. Görür (2009). Indian Buffet Proceses with Power-law Behavior. In *Advances in Neural Information Processing Systems 22* (Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta, Eds.), 1838–1846.

Y. W. Teh, M. I. Jordan, M. J. Beal and D. M. Blei (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, **101**, 1566–1581.

Y. Yu and X.-L. Meng (2011). To Center or Not to Center: That is Not the Question – An Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency. *Journal of Computational and Graphical Statistics*, **20**, 531–570.

W. Zhu and F. Leisen (2014). A multivariate extension of a vector of Poisson-Dirichlet processes. To appear in the *Journal of Nonparametric Statistics*.