Code for "Compound random measures and their use in Bayesian nonparametrics"

Jim E. Griffin and Fabrizio Leisen*
University of Kent

Please note that this code is designed to illustrate the methods in the paper. The code has been optimised and so should not be used for the comparison of algorithms.

1 Introduction

Code is provided for both a univariate and a multivariate NCoRM mixture model with $\Gamma(\phi, 1)$ score distribution. There is a version of the NCoRM mixture model with a Dirichlet process constructed by normalising a CoRM with gamma process marginals. The Lévy intensity of the gamma process is parameterised in the following way in the code

$$\rho(x) = Mx^{-1}\exp\{-x\}$$

and α is a probability measure (the centring measure). An NCoRM process with Normalised generalised gamma marginals can be constructed by normalising a CoRM with generalised gamma process marginals. The Lévy intensity of the gamma process is parameterised in the following way in the code

$$\rho(x) = \frac{M\sigma}{\Gamma(1-\sigma)}x^{-1-\sigma}\exp\{-x\}$$

and α is a probability measure (the centring measure).

The univariate NCoRM mixture model assumes that

$$y_i \sim \mathrm{N}(y_i|\mu_i, \sigma_i^2)$$

* *Corresponding author*: Jim E. Griffin, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury CT2 7NF, U.K. *Email*: jeg28@kent.ac.uk

and the centring measure is $\alpha(mu, \sigma^2) = \mathrm{N}()\Gamma()$. The multivariate NCoRM mixture model assumes that

$$y_i \sim \mathrm{N}(y_i|\mu_i, \Sigma_i)$$

and centring measure is $\alpha(\mu, \Sigma)$. The second model is used in the example in the paper. The Matlab functions for each model are listed below.

Dirichlet process marginals

```
[output] = norm_mult_DP_MVN_PU(data, group, burnin, numbofits, every, pred)
[output] = norm_mult_DP_PU(data, group, burnin, numbofits, every, pred)
```

Inputs

- `data` – An $(n \times p)$-dimensional matrix of data values (where $p = 1$ for `norm_mult_DP_PU`)
- `group` – An $(n \times 1)$-dimensional vector indicating the group of each data point. The groups should be numbered $1, \ldots, K$ (for K groups).
- `burnin` – The length of the burn-in period for the MCMC sampler.
- `numbofits` – The total number of samples to be collected for each parameter.
- `every` – The level of thinning
- `pred` – A $(N \times p)$-dimensional matrix of values at which the posterior mean density will be evaluated. The posterior mean marginal density of the i-th dimension of the data will be evaluated at the point in the i-th column.

Outputs

- `output` – This a `struct` structure. The fields are:
  - `pred` – The inputted value of pred
  - `probpred` – A $(N \times p)$-dimensional matrix of evaluations of the posterior mean density of each dimension.
  - `phi` – The posterior sample of ϕ.
  - `M` – The posterior sample of M, a parameter of the Lévy intensity.
  - `s` – The posterior sample of allocation variables of the mixture model.

Normalised generalised gamma marginals

`[output] = norm_mult_DP_MVN_PU(data, group, burnin, numbofits, every, pred)`
`[output] = norm_mult_DP_PU(data, group, burnin, numbofits, every, pred)`

Inputs

- `data` – An $(n \times p)$-dimensional matrix of data values (where $p = 1$ for `norm_mult_DP_PU`)
- `group` – An $(n \times 1)$-dimensional vector indicating the group of each data point. The groups should be numbered $1, \ldots, K$ (for K groups).
- `burnin` – The length of the burn-in period for the MCMC sampler.
- `numbofits` – The total number of samples to be collected for each parameter.
- `every` – The level of thinning
- `pred` – A $(N \times p)$-dimensional matrix of values at which the posterior mean density will be evaluated. The posterior mean marginal density of the i-th dimension of the data will be evaluated at the point in the i-th column.

Outputs

- `output` – This a `struct` structure. The fields are:
  - `pred` – The inputted value of pred
  - `probpred` – A $(N \times p)$-dimensional matrix of evaluations of the posterior mean density of each dimension.
  - `phi` – The posterior sample of ϕ.
  - `NGGsigma` – The posterior sample of σ, a parameter of the Lévy intensity.
  - `M` – The posterior sample of M, a parameter of the Lévy intensity.
  - `s` – The posterior sample of allocation variables of the mixture model.