Nonparametric Bayesian decompounding

Frank van der Meulen
Delft University of Technology

joint work with Shota Gugushvili (Leiden University)
and Peter Spreij (University of Amsterdam)

EMS 2015, Amsterdam

Compound Poisson process

- $N = (N_t,\, t \geq 0)$ is a Poisson process with a constant intensity $\lambda > 0$.
- $Y_1, Y_2, Y_3 \ldots$ is a sequence of independently distributed random variables, independent of N and having a common density f.
- A *compound Poisson process (CPP)* $X = (X_t,\, t \geq 0)$ is defined as

$$X_t = \sum_{j=1}^{N_t} Y_j.$$

Statistical problem

- The 'true' parameters: $\lambda = \lambda_0$ and $f = f_0$.
- Observations: a discrete time sample $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ is available from the CPP X, where $\Delta > 0$ is
  - fixed, or
  - tends to zero (high frequency observations)

  We will assume $\Delta = \Delta_n = n^{-\alpha}$, $\alpha \in [0, 1)$.
- Problem: (nonparametric) estimation of λ_0 and f_0.

Statistical problem

- The 'true' parameters: $\lambda = \lambda_0$ and $f = f_0$.
- Observations: a discrete time sample $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ is available from the CPP X, where $\Delta > 0$ is
  - fixed, or
  - tends to zero (high frequency observations)

  We will assume $\Delta = \Delta_n = n^{-\alpha}$, $\alpha \in [0, 1)$.
- Problem: (nonparametric) estimation of λ_0 and f_0.

Brief outline of this talk:

1. Equivalent problem, prior and likelihood.
2. Bayesian rate of contraction.
3. Numerical issues and a few simulation results.

Equivalent problem

The random variables

$$Z_i^\Delta = X_{i\Delta} - X_{(i-1)\Delta}, \qquad 1 \le i \le n$$

are IID. Each Z_i^Δ is distributed as

$$Z^\Delta = \sum_{j=1}^{T} Y_j,$$

where T is independent of the sequence $Y_1, Y_2, \ldots$ and has a Poisson distribution with parameter $\Delta\lambda$.

Equivalent problem

The random variables

$$Z_i^\Delta = X_{i\Delta} - X_{(i-1)\Delta}, \qquad 1 \leq i \leq n$$

are IID. Each Z_i^Δ is distributed as

$$Z^\Delta = \sum_{j=1}^{T} Y_j,$$

where T is independent of the sequence $Y_1, Y_2, \ldots$ and has a Poisson distribution with parameter $\Delta\lambda$.

- *Equivalent problem:* Estimate λ_0 and f_0 based on the sample $\mathcal{Z}_n^\Delta = (Z_1^\Delta, Z_2^\Delta, \ldots, Z_n^\Delta)$.

Recovering λ_0 and f_0 from $\mathcal{Z}_n^\Delta$ is called *decompounding*.

History

- Frequentist approaches:
  - Early contributions [Buchmann and Grübel(2003)] and [Buchmann and Grübel(2004)]: estimation of the distribution function F_0,
  - later contributions [Comte et al.(2014)], [Duval(2013)] and [van Es et al.(2007)] concentrated on estimation of the density f_0 instead.

History

- Frequentist approaches:
  - Early contributions [Buchmann and Grübel(2003)] and [Buchmann and Grübel(2004)]: estimation of the distribution function F_0,
  - later contributions [Comte et al.(2014)], [Duval(2013)] and [van Es et al.(2007)] concentrated on estimation of the density f_0 instead.
- Bayesian approaches:
  - In [Ghosal and Tang(2006)] and [Tang and Ghosal(2007)] nonparametric Bayesian estimation of the transition density of a discretely observed Markov processes is studied.
  - A parametric Bayesian approach to inference for compound Poisson processes is studied in [Insua et al.(2012)].

Class of densities

The class $\mathcal{F}$ of densities f is that of location mixtures of normal densities. So

$$f(x) = f_{H,\sigma}(x) = \int \phi_\sigma(x - z)\mathrm{d}H(z),$$

where ϕ_σ denotes the density of the $N(0, \sigma^2)$ distribution and H is a mixing measure.

Prior distributions

$$f(x) = f_{H,\sigma}(x) = \int \phi_\sigma(x - z)\mathrm{d}H(z) \tag{1}$$

- Independent priors on λ (prior Π_1) and f (prior Π_2).

Prior distributions

$$f(x) = f_{H,\sigma}(x) = \int \phi_\sigma(x - z)\mathrm{d}H(z) \tag{1}$$

- Independent priors on λ (prior Π_1) and f (prior Π_2).
- The prior for f is defined as the law of the function $f_{H,\sigma}$ as in (1), with H assumed to follow a Dirichlet process prior D_α with base measure α and σ a-priori independent with distribution Π_3.

Prior distributions

$$f(x) = f_{H,\sigma}(x) = \int \phi_\sigma(x - z)\mathrm{d}H(z) \tag{1}$$

- Independent priors on λ (prior Π_1) and f (prior Π_2).
- The prior for f is defined as the law of the function $f_{H,\sigma}$ as in (1), with H assumed to follow a Dirichlet process prior D_α with base measure α and σ a-priori independent with distribution Π_3.

Dirichlet mixture of normal densities as a prior Π_2, introduced in the context of Bayesian density estimation by [Ferguson(1983)] and [Lo(1984)].

Likelihood for Z^Δ

Notation

$\mathbb{Q}^{\Delta}_{\lambda,f}$	law of Z^Δ (law of $(X_{i\Delta} - X_{(i-1)\Delta})$
$\mathbb{Q}^{n,\Delta}_{\lambda,f}$	law of $\mathcal{Z}^\Delta_n$, $\mathcal{Z}^\Delta_n = (Z_\Delta, \dots, Z_{n\Delta})$
$\mathbb{R}^{\Delta}_{\lambda,f}$	law of $(X_t,\, t \in [0,\Delta])$ (the CPP on $[0,\Delta]$)

Likelihood for Z^{Δ}

For $\lambda, \tilde{\lambda} > 0$ and $f, \tilde{f}$ positive and continuous, we have for $\Delta > 0$

$$\frac{\mathrm{d}\mathbb{R}^{\Delta}_{\lambda,f}}{\mathrm{d}\mathbb{R}^{\Delta}_{\widetilde{\lambda},\widetilde{f}}} = \exp\left(\int_0^{\Delta} \int_{\mathbb{R}} \log\left(\frac{\lambda f(x)}{\widetilde{\lambda}\widetilde{f}(x)} \right) \mu(dt, dx) - (\lambda - \widetilde{\lambda}) \right),$$

and

Likelihood for Z^Δ

For $\lambda, \tilde{\lambda} > 0$ and $f, \tilde{f}$ positive and continuous, we have for $\Delta > 0$

$$\frac{\mathrm{d}\mathbb{R}^\Delta_{\lambda,f}}{\mathrm{d}\mathbb{R}^\Delta_{\widetilde{\lambda},\widetilde{f}}} = \exp\left(\int_0^\Delta \int_{\mathbb{R}} \log\left(\frac{\lambda f(x)}{\widetilde{\lambda}\widetilde{f}(x)}\right) \mu(dt, dx) - (\lambda - \widetilde{\lambda})\right),$$

and

$$k^\Delta_{\lambda,f}(Z^\Delta) = k^\Delta_{\lambda,f}(X_\Delta) := \frac{\mathrm{d}\mathbb{Q}^\Delta_{\lambda,f}}{\mathrm{d}\mathbb{Q}^\Delta_{\widetilde{\lambda},\widetilde{f}}}(X_\Delta) = \mathrm{E}_{\,\mathbb{R}^\Delta_{\widetilde{\lambda},\widetilde{f}}}\left(\frac{\mathrm{d}\mathbb{R}^\Delta_{\lambda,f}}{\mathrm{d}\mathbb{R}^\Delta_{\widetilde{\lambda},\widetilde{f}}} \,\middle|\, X_\Delta\right).$$

Likelihood of $\mathcal{Z}_n^\Delta$ and posterior measure

The likelihood given the observations $\mathcal{Z}_n^\Delta$ is

$$L_n^\Delta(\lambda, f) = \prod_{i=1}^{n} k_{\lambda,f}^\Delta(Z_i^\Delta).$$

By Bayes' theorem, the posterior measure of any measurable set $A \subset (0, \infty) \times \mathcal{F}$ is given by

$$\Pi(A|\mathcal{Z}_n^\Delta) = \frac{\iint_A L_n^\Delta(\lambda, f)\, \mathrm{d}\Pi_1(\lambda)\mathrm{d}\Pi_2(f)}{\iint L_n^\Delta(\lambda, f)\, \mathrm{d}\Pi_1(\lambda)\mathrm{d}\Pi_2(f)}.$$

Aims and results

Rescaled Hellinger distance

$$h^{\Delta}(\mathbb{Q}_{\lambda_0,f_0},\mathbb{Q}_{\lambda,f}) = \frac{1}{\sqrt{\Delta}} h(\mathbb{Q}_{\lambda_0,f_0},\mathbb{Q}_{\lambda,f}).$$

As $\Delta \to 0$, this converges to $h(\lambda_0 f_0, \lambda f)$.

Aims and results

Rescaled Hellinger distance

$$h^{\Delta}(\mathbb{Q}_{\lambda_0,f_0},\mathbb{Q}_{\lambda,f}) = \frac{1}{\sqrt{\Delta}} h(\mathbb{Q}_{\lambda_0,f_0},\mathbb{Q}_{\lambda,f}).$$

As $\Delta \to 0$, this converges to $h(\lambda_0 f_0, \lambda f)$.

Define

$$A(\varepsilon_n, M) = \{(\lambda, f) : h^{\Delta}(\mathbb{Q}^{\Delta}_{\lambda_0,f_0},\mathbb{Q}^{\Delta}_{\lambda,f}) > M\varepsilon_n\}.$$

We shall say that ε_n is a posterior contraction rate, if there exists a constant $M > 0$, such that

$$\Pi(A(\varepsilon_n, M)|\mathcal{Z}^{\Delta}_n) \to 0 \qquad (2)$$

in $\mathbb{Q}^{\Delta,n}_{\lambda_0,f_0}$-probability as $n \to \infty$.

Assumptions on data generating process

(i) λ_0 is in a compact set $[\underline{\lambda}, \overline{\lambda}] \subset (0, \infty)$;

(ii) The true density f_0 is a location mixture of normal densities, i.e.

$$f_0(x) = f_{H_0,\sigma_0}(x) = \int \phi_{\sigma_0}(x - z)\mathrm{d}H_0(z)$$

for some fixed distribution H_0 and a constant $\sigma_0 \in [\underline{\sigma}, \overline{\sigma}] \subset (0, \infty)$. Furthermore, for some $0 < \kappa_0 < \infty$, $H_0[-\kappa_0, \kappa_0] = 1$, i.e. H_0 has compact support.

Assumptions on the prior

(i) The prior on λ, has a density π_1 such that

$$0 < \underline{\pi}_1 \leq \pi_1(\lambda) \leq \overline{\pi}_1 < \infty, \quad \lambda \in [\underline{\lambda}, \overline{\lambda}].$$

(ii) The prior on σ has a density π_3 such that

$$0 < \underline{\pi}_3 \leq \pi_3(\sigma) \leq \overline{\pi}_3 < \infty, \quad \sigma \in [\underline{\sigma}, \overline{\sigma}].$$

Assumptions on the prior

(i) The prior on λ, has a density π_1 such that

$$0 < \underline{\pi}_1 \leq \pi_1(\lambda) \leq \overline{\pi}_1 < \infty, \quad \lambda \in [\underline{\lambda}, \overline{\lambda}].$$

(ii) The prior on σ has a density π_3 such that

$$0 < \underline{\pi}_3 \leq \pi_3(\sigma) \leq \overline{\pi}_3 < \infty, \quad \sigma \in [\underline{\sigma}, \overline{\sigma}].$$

(iii) The base measure α of the Dirichlet process prior D_α has a continuous density on an interval $[-\kappa_0 - \zeta, \kappa_0 + \zeta]$, for some $\zeta > 0$, is bounded away from zero there, and for all $t > 0$ satisfies the tail condition

$$\alpha(|z| > t) \lesssim e^{-b|t|^\delta} \tag{3}$$

with some constants $b > 0$ and $\delta > 0$;

Main result

Theorem

Under the assumptions of the previous 2 slides, provided $n\Delta \to \infty$, there exists a constant $M > 0$, such that for

$$\varepsilon_n = \frac{\log^{\kappa}(n\Delta)}{\sqrt{n\Delta}}, \quad \kappa = \max\left(\frac{2}{\delta}, \frac{1}{2}\right) + \frac{1}{2},$$

we have

$$\Pi\left(A\left(\varepsilon_n, M\right) \middle| \mathcal{Z}_n^{\Delta}\right) \to 0$$

in $\mathbb{Q}_{\lambda_0, f_0}^{\Delta, n}$-probability as $n \to \infty$.

Short discussion of the main result

- The (frequentist) minimax convergence rate for estimation of $k_{\lambda,f}$ is unknown, but by analogy to [Ibragimov and Khas'minskiĭ(1982)] and the lower bound derived in [Duval(2013)], is expected to be of order $\log^{1/4}(n)/\sqrt{n}$ (cf. [Ghosal and van der Vaart(2001)]).

Short discussion of the main result

- The (frequentist) minimax convergence rate for estimation of $k_{\lambda,f}$ is unknown, but by analogy to [Ibragimov and Khas'minskiĭ(1982)] and the lower bound derived in [Duval(2013)], is expected to be of order $\log^{1/4}(n)/\sqrt{n}$ (cf. [Ghosal and van der Vaart(2001)]).
- Our posterior contraction rate is the same as for Bayesian density estimation in [Ghosal and van der Vaart(2001)].

Short discussion of the main result

- The (frequentist) minimax convergence rate for estimation of $k_{\lambda,f}$ is unknown, but by analogy to [Ibragimov and Khas'minskiĭ(1982)] and the lower bound derived in [Duval(2013)], is expected to be of order $\log^{1/4}(n)/\sqrt{n}$ (cf. [Ghosal and van der Vaart(2001)]).
- Our posterior contraction rate is the same as for Bayesian density estimation in [Ghosal and van der Vaart(2001)].
- Extensions to *multivariate* setting possible with a similar result, slightly weaker, but *more complicated* technical assumptions like in [Shen et al. (2013)].

About the proof

- General results in Bayesian nonparametric statistics, such as Theorem 2.1 in [Ghosal et al.(2000)] and Theorem 2.1 in [Ghosal and van der Vaart(2001)] are not easily applicable.

About the proof

- General results in Bayesian nonparametric statistics, such as Theorem 2.1 in [Ghosal et al.(2000)] and Theorem 2.1 in [Ghosal and van der Vaart(2001)] are not easily applicable.
- The proof mimmicks the main steps of the proof of Theorem 2.1 in [Ghosal et al.(2000)], while also employing some results on the Dirichlet location mixtures of normal densities from [Ghosal and van der Vaart(2001)].

About the proof

- General results in Bayesian nonparametric statistics, such as Theorem 2.1 in [Ghosal et al.(2000)] and Theorem 2.1 in [Ghosal and van der Vaart(2001)] are not easily applicable.
- The proof mimmicks the main steps of the proof of Theorem 2.1 in [Ghosal et al.(2000)], while also employing some results on the Dirichlet location mixtures of normal densities from [Ghosal and van der Vaart(2001)].
- Significant part of technicalities are characteristic of the decompounding problem only.

Computational difficulties

The density of a nonzero increment z on a time interval of length Δ is given by

$$p(z \mid \lambda, f) = \frac{e^{-\lambda\Delta}}{1 - e^{-\lambda\Delta}} \sum_{k=1}^{\infty} \frac{(\lambda\Delta)^k}{k!} f^{(*k)}(z),$$

Computational difficulties

The density of a nonzero increment z on a time interval of length Δ is given by

$$p(z \mid \lambda, f) = \frac{e^{-\lambda\Delta}}{1 - e^{-\lambda\Delta}} \sum_{k=1}^{\infty} \frac{(\lambda\Delta)^k}{k!} f^{(*k)}(z),$$

We assume the jump size distribution is a mixture of $J \geq 1$ Gaussians:

$$f(\cdot) = \sum_{j=1}^{J} \rho_j \phi(\cdot; \mu_j, 1/\tau), \qquad \sum_{j=1}^{J} \rho_j = 1$$

Parameters: $\mu = (\mu_1, \ldots, \mu_J)$ and τ.

Computational difficulties

The density of a nonzero increment z on a time interval of length Δ is given by

$$p(z \mid \lambda, f) = \frac{e^{-\lambda\Delta}}{1 - e^{-\lambda\Delta}} \sum_{k=1}^{\infty} \frac{(\lambda\Delta)^k}{k!} f^{(*k)}(z),$$

We assume the jump size distribution is a mixture of $J \geq 1$ Gaussians:

$$f(\cdot) = \sum_{j=1}^{J} \rho_j \phi(\cdot; \mu_j, 1/\tau), \qquad \sum_{j=1}^{J} \rho_j = 1$$

Parameters: $\mu = (\mu_1, \ldots, \mu_J)$ and τ.

Infeasible to generate independent realisations of the posterior distribution of (λ, f).

Reparametrisation and prior specification

Instead of parametrising with $(\lambda, \rho_1, \ldots, \rho_J)$ we define

$$\psi_j = \lambda \rho_j, \qquad j = 1, \ldots, J.$$

Then

$$\lambda = \sum_{j=1}^{J} \psi_j \quad \text{and} \quad \rho_j = \frac{\psi_j}{\sum_{j=1}^{J} \psi_j}.$$

Define $\theta = (\psi, \mu, \tau)$ and $\psi = (\psi_1, \ldots, \psi_J)$.

Reparametrisation and prior specification

Instead of parametrising with $(\lambda, \rho_1, \ldots, \rho_J)$ we define

$$\psi_j = \lambda \rho_j, \qquad j = 1, \ldots, J.$$

Then

$$\lambda = \sum_{j=1}^{J} \psi_j \quad \text{and} \quad \rho_j = \frac{\psi_j}{\sum_{j=1}^{J} \psi_j}.$$

Define $\theta = (\psi, \mu, \tau)$ and $\psi = (\psi_1, \ldots, \psi_J)$.

Take partially conjugate priors

$$\begin{aligned} \psi_1, \ldots, \psi_J &\overset{\text{iid}}{\sim} \mathcal{G}(\alpha_0, \beta_0) \\ \mu \mid \tau &\sim \mathcal{N}([\xi_1, \ldots, \xi_J]', I_{J \times J}(\tau\kappa)^{-1}) \\ \tau &\sim \mathcal{G}(\alpha_1, \beta_1) \end{aligned}$$

with positive hyperparameters $(\alpha_0, \beta_0, \alpha_1, \beta_1, \kappa)$ fixed.

Introducing auxiliary variables

Let

$$\mathcal{I} = \{i \in \{1, \ldots, n\} : \ z_i \neq 0\}$$

denote the set of observations with nonzero jump sizes

Introducing auxiliary variables

Let

$$\mathcal{I} = \{i \in \{1, \ldots, n\} : \; z_i \neq 0\}$$

denote the set of observations with nonzero jump sizes

- For $i \in \mathcal{I}$ and $j \in \{1, \ldots, J\}$, let n_{ij} denote the number of jumps of type j on segment i.
- Denote the set of all auxiliary variables by

  $$\mathbf{a} = \{a_i, \; i \in I\},$$

  where

  $$a_i = (n_{i1}, n_{i2}, \ldots, n_{iJ}).$$

Data augmentation algorithm

Construct a Metropolis-Hastings algorithm to draw from

$$p(\theta, \mathbf{a} \mid z) = \frac{p(\theta, z, \mathbf{a})}{p(z)},$$

where

$$\theta = (\psi, \mu, \tau), \qquad \psi = (\psi_1, \ldots, \psi_J), \qquad \mu = (\mu_1, \ldots, \mu_J).$$

Data augmentation algorithm

Construct a Metropolis-Hastings algorithm to draw from

$$p(\theta, \mathbf{a} \mid z) = \frac{p(\theta, z, \mathbf{a})}{p(z)},$$

where

$$\theta = (\psi, \mu, \tau), \qquad \psi = (\psi_1, \dots, \psi_J), \qquad \mu = (\mu_1, \dots, \mu_J).$$

The two main steps of the algorithm are:

1. *Update segments:* for each segment $i \in \mathcal{I}$, draw a_i conditional on $(\theta, z, \mathbf{a}_{-i})$;
2. *Update parameters:* draw θ conditional on $(z, \mathbf{a})$.

Some numerical results

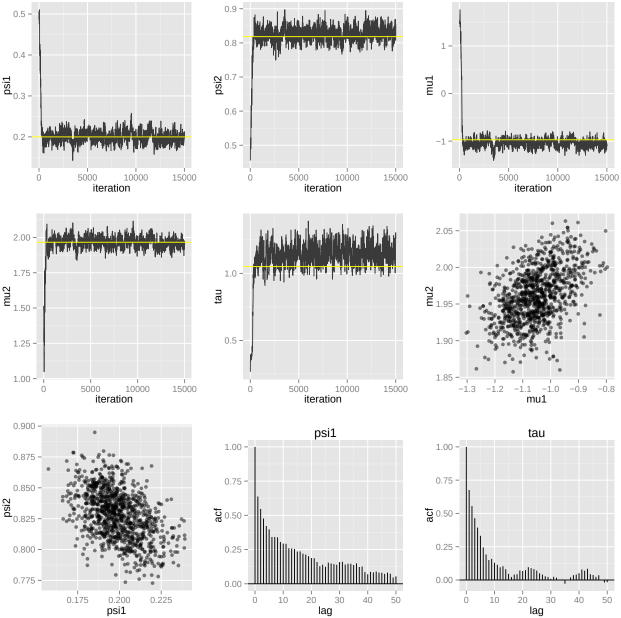
- We simulated $n = 5000$ segments with $\Delta = 1$, $\mu_1 = 2$, $\mu_2 = -1$, $\tau = 1$, $\psi_1 = 0.8\lambda$ and $\psi_2 = 0.2\lambda$.
- Prior-hyperparameters: $\mathcal{E}(1)$ priors on all ψ_j and τ. Furthermore: $\mu_j \mid \tau \sim \mathcal{N}(0, \tau^{-1})$.

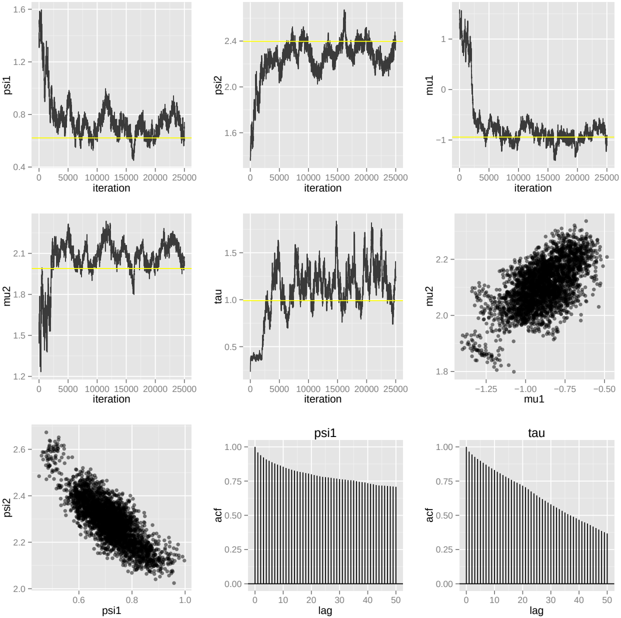
Some numerical results

- We simulated $n = 5000$ segments with $\Delta = 1$, $\mu_1 = 2$, $\mu_2 = -1$, $\tau = 1$, $\psi_1 = 0.8\lambda$ and $\psi_2 = 0.2\lambda$.
- Prior-hyperparameters: $\mathcal{E}(1)$ priors on all ψ_j and τ. Furthermore: $\mu_j \mid \tau \sim \mathcal{N}(0, \tau^{-1})$.

We show results for

- $\lambda = 1$: $\psi_1 = 0.8$ and $\psi_2 = 0.2$
- $\lambda = 3$: $\psi_1 = 2.4$ and $\psi_2 = 0.6$


psi1
0.5
0.4
0.3
0.2
0
5000
10000
15000
iteration
psi2
0.9
0.8
0.7
0.6
0.5
0
5000
10000
15000
iteration
mu1
1
0
−1
0
5000
10000
15000
iteration
mu2
2.00
1.75
1.50
1.25
1.00
0
5000
10000
15000
iteration
tau
1.0
0.5
0
5000
10000
15000
iteration
mu2
2.05
2.00
1.95
1.90
1.85
−1.3
−1.2
−1.1
−1.0
−0.9
−0.8
mu1
psi2
0.900
0.875
0.850
0.825
0.800
0.775
0.175
0.200
0.225
psi1
psi1
acf
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
lag
tau
acf
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
lag


psi1
iteration
psi2
iteration
mu1
iteration
mu2
iteration
tau
iteration
mu2
mu1
psi2
psi1
psi1
acf
lag
tau
acf
lag

References I

- B. Buchmann and R. Grübel. Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.*, 31:1054–1074, 2003.
- B. Buchmann and R. Grübel. Decompounding Poisson random sums: recursively truncated estimates in the discrete case. *Ann. Inst. Statist. Math.*, 56:743–756, 2004.
- F. Comte, C. Duval and V. Genon-Catalot. Nonparametric density estimation in compound Poisson process using convolution power estimators. *Metrika*, 77:163–183, 2014.
- C. Duval. Density estimation for compound Poisson processes from discrete data. *Stoch. Proc. Appl.*, 123:3963–3986, 2013.
- B. van Es, S. Gugushvili and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13:672–694, 2007.
- T.S. Ferguson. Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics*, 287–302, Academic Press, New York, 1983.

References II

- S. Ghosal, J.K. Ghosh and A.W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28:500–531, 2000.
- S. Ghosal and Y. Tang. Bayesian consistency for Markov processes. *Sankhyā*, 68:227–239, 2006.
- S. Ghosal and A.W. van der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.*, 29:1233–1263, 2001.
- S. Ghosal and A.W. van der Vaart. Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.*, 35:697–723, 2007.
- I.A. Ibragimov and R.Z. Khas'minskiĭ. An estimate of the density of a distribution belonging to a class of entire functions (Russian). *Teor. Veroyatnost. i Primenen.*, 27:514–524, 1982.
- D.R. Insua, F. Ruggeri and M.P. Wiper. *Bayesian Analysis of Stochastic Process Models*. John Wiley & Sons, 2012.

References III

- A.Y. Lo. On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.*, 12:351–357, 1984.
- Y. Tang and S. Ghosal. Posterior consistency of Dirichlet mixtures for estimating a transition density. *J. Statist. Plann. Inference*, 137:1711–1726, 2007.
- W. Shen, S.T. Tokdar and S. Ghosal. Adaptive multivariate density estimation with Dirichlet mixtures. *Biometrika*, 100: 632–640, 2013.