Bayesian Deconvolution of Cyclostationary Processes Based on Point Processes

†Christophe ANDRIEU - †Patrick DUVAUT - †‡Arnaud DOUCET
†ENSEA - ETIS Groupe Signal - 95014 Cergy Pontoise France
‡LETI-CEA Technologies Avancées 91191 Gif sur Yvette France
andrieu@ensea.fr - duvaut@ensea.fr - douceta@ensea.fr

ABSTRACT

In this paper we address the problem of the Bayesian deconvolution of a widely spread class of processes, filtered point processes, whose underlying point process is a self-excited point process. In order to achieve this deconvolution, we perform powerful stochastic algorithm, the Markov chains Monte Carlo (MCMC), which despite their power have not been yet widely used in signal processing.

1. Introduction

In this paper we recall a general modelization of filtered point processes, which is useful in practice. We then use it to model a particular class of filtered point processes, whose statistics are periodic in time. We then present an algorithm based on Markov chain Monte Carlo, which estimates all the conditional densities of the parameters of the process. We then present the results of a simulation.

2. General modelization

Point processes are suitable tools that allow physicists to model many situations of interest: indeed they are a natural way for modeling highly localized events occuring randomly at given times or points of the state space. A natural way of mathematicaly modeling such processes is to use the following notion of counting process. Let us consider $N(t)$ a stochastic process such that:

$$\begin{aligned} dN(t) &= 1 \text{ in case of an event} \\ dN(t) &= 0 \text{ otherwise} \end{aligned} \tag{1}$$

$N(t)$ is thus the number of events that occured until time t.

Furthermore, each of the event may have another characteristic value $\underline{\theta}_t$ (apart from its own time of occurence), which may also be stochastic. The process $(N(t), \underline{\theta}_t)_{t\in\mathbb{R}}$ is named a marked point process.

Nevertheless in most situations, the underlying point process does not appear directly as an observation. The point process may be filtered (due to the width of the bandpass of the detector, the reponse of a mechanical structure to pulses...), and thus the observed process is modelled as:

$$x(t) = \sum_i h\left(t - t_i, \underline{\theta}_{t_i}\right) \tag{2}$$

which may be rewritten using a stochastic integral in the following useful manner:

$$x(t) = \int_{\mathbb{R}} h\left(t-u, \underline{\theta}_u\right) dN(u) \tag{3}$$

Then the statistics of the process may be defined. We suppose the times and the marks to be independent, the marks to be iid. Concerning the time of arrival, we use the formalism of the so-called self-excited point processes, that is we suppose the point process to be defined with the following conditional probability densities (considering that no events happened before $t = 0$):

$$\begin{aligned} & P\left(t \le t_{n+1} \le t + dt / N(t_n) = n, t_{1\to n}\right) \\ = \;& \lambda\left(t / N(t), t_{1\to n}\right) dt + o(dt) \end{aligned} \tag{4}$$

(where $t_{1\to n} = \{t_1, \ldots, t_n\}$) which we will note in order to alleviate notations and when no confusion is possible $\lambda(t)$. This definition leads to the interesting following formula:

$$P\left(t_{n+1} \le t / t_{1\to n}\right) = 1 - \exp\left(-\int_{t_n}^{t} \lambda(t)\, dt\right) \tag{5}$$

This formula is useful when simulating such point processes [4]. The statistics of the process x can be calculated using the multicoincidence densities [2] defined as follows:

$$\begin{aligned} & \mathbb{E}\left(dN(t_1) \times \ldots \times dN(t_n)\right) \\ = \;& f_n(t_1, \ldots, t_n) = \mathbb{E}\left(\lambda(t_1)\right) \prod_{i=2}^{n} \mathbb{E}\left(\lambda(t_i) / t_{1\to i-1}\right) \end{aligned} \tag{6}$$

where the expectation is to be taken over the possible realizations of $N(.)$.

One can obtain for example the first order and second order moments:

$$\begin{aligned} \mathbb{E}(x(t)) &= \int_{\mathbb{R}} \mathbb{E}\left(h\left(t-u, \underline{\theta}_u\right)\right) f_1(u)\, du \\ \mathbb{E}(x(t_1)\, x(t_2)) &= \int_{\mathbb{R}} \left[\mathbb{E}\left(h\left(t_1 - u_1, \underline{\theta}_{u_1}\right) h\left(t_2 - u_2, \underline{\theta}_{u_2}\right)\right)\right. \\ & \left. \left(\delta(u_1 - u_2)\, f_1(u_1) + f_2(u_1, u_2)\right) du_1 du_2\right] \end{aligned} \tag{7}$$

3. Application to cyclostationary processes

A stochastic process is said to be cyclostationary up to order 2 when the corresponding statistics depend periodically on time [1], that is there exists $T \in \mathbb{R}$ such that:

$$\begin{aligned} \mathbb{E}\left(x\left(t+T\right)\right) &= \mathbb{E}\left(x\left(t\right)\right) \\ \mathbb{E}\left(x\left(t_1+T\right)x\left(t_2+T\right)\right) &= \mathbb{E}\left(x\left(t_1\right)x\left(t_2\right)\right) \end{aligned} \tag{8}$$

Assuming $h\left(.,.\right)$ to have a finite support and $\left(\underline{\theta}_t\right)_{t\in\mathbb{R}}$ to be stationary, one easily deduces that the process appearing in the preceeding section is cyclostationary up to order 2 if and only if:

$$\begin{aligned} f_1\left(u+T\right) &= f_1\left(u\right) \\ f_2\left(u+T,v+T\right) &= f_2\left(u,v\right) \end{aligned} \tag{9}$$

Furthermore, we show that the bifrequency representation of x will be these of

$$x_0\left(t\right) = \sum_{k=0}^{+\infty} h\left(t-kT,\underline{\theta}_{kT}\right) \tag{10}$$

but weighted by the Fourier transform of $\delta\left(u_1-u_2\right)f_1\left(u_1\right) + f_2\left(u_1,u_2\right)$, which shows that the more x is "time shifted", the more its Fourier transform is concentrated on the first diagonal of the bifrequency plane, and thus weakens the DCS (degree of cyclostationarity)[1] and the performance of algorithms relying only on correlation property. It is this stationarizing effect which motivated our work.

4. Bayesian solution to the problem

We model the observed signal in the following manner:

$$y\left(t\right) = \int_{\mathbb{R}} h\left(t-u,\underline{\theta}_u\right)dN\left(u\right) + n\left(t\right) \tag{11}$$

where $n\left(t\right)$ is the observation noise. This leads to the discrete observations $y_{1\to n}$.

The Bayesian solution to the problem consists in evaluating the following conditional probability density:

$$p\left(t_{1\to p},\underline{\theta}_{1\to p},\underline{\gamma}_n,\underline{\gamma}_t,\underline{\gamma}_{\underline{\theta}} \middle/ y_{1\to n}\right) \tag{12}$$

where $\underline{\gamma}_n,\underline{\gamma}_t,\underline{\gamma}_{\underline{\theta}}$ are respectively the parameters for the observation noise, the intensity of the point process and the parameters of the mark. The knowledge of this conditional density will then allow us to calculate all the statistics of interest, including MAP, confidence intervals, expectations. It is worth noticing that we consider the parameters $\underline{\gamma}_*$ to be stochastic, which may not seem natural (on the contrary of the missing data $t_{1\to p}$ and $\underline{\theta}_{1\to p}$) to people not acquainted with Bayesian techniques, but is common in the field: this may be seen as a regularization constraint.

In order to make this evaluation, one will perform a stochastic algorithm, a Markov chain Monte Carlo. This class of algorithm allows one to construct a Markov chain of samples whose equilibrium density is $p\left(t_{1\to p},\underline{\theta}_{1\to p},\underline{\gamma}_n,\underline{\gamma}_t,\underline{\gamma}_{\underline{\theta}} \middle/ y_{1\to n}\right)$. In other words, one will construct the density with a chain of samples which will represent it. We recall below the two most popular MCMC methods, the Metropolis-Hastings (MH) algorithm, and the Gibbs sampler.

4.1. The Metropolis-Hastings algorithm

Suppose that one wants to sample from the following multivariate probability density:

$$\pi\left(x\right) \tag{13}$$

1. Set $k=0$ and arbitrarily draw $x^{(0)}$.
2. Propose a candidate sample according to proposal density $q\left(x^{(k)},x^{(k+1)}\right)$
3. Accept this candidate with acceptance probability:

$$\alpha\left(x^{(k)},x^{(k+1)}\right) = \begin{cases} \min\left\{1,\frac{\pi\left(x^{(k+1)}\right)q\left(x^{(k)},x^{(k+1)}\right)}{\pi\left(x^{(k)}\right)q\left(x^{(k+1)},x^{(k)}\right)}\right\} \\ 1 \text{ if } \pi\left(x^{(k)}\right)q\left(x^{(k+1)},x^{(k)}\right)=0 \end{cases}$$

4. Set $k=k+1$ and go to 2.

Under relatively weak conditions, as $k\to+\infty$, $x^{(k)}$ tends to be drawn according to $\pi\left(x\right)$.

4.2. The Gibbs sampler

Suppose that one wants to sample from the following probability density:

$$\pi\left(x\right) = \pi\left(x_{1\to n}\right) \tag{14}$$

where $x_{1\to n}$ is a partition of x

1. Set $k=0$ and arbitrarily draw $x_{1\to n}^{(0)}$
2. (a) Draw $x_1^{(k+1)}$ according to $p\left(x_1 \middle/ x_{2\to n}^{(k)}\right)$
   (b) Draw $x_i^{(k+1)}$ according to $p\left(x_i \middle/ x_{1\to i-1}^{(k+1)},x_{i+1\to n}^{(k)}\right)$ for $i=2,\ldots,n-1$.
   (c) Draw $x_n^{(k+1)}$ according to $p\left(x_i \middle/ x_{1\to n-1}^{(k+1)}\right)$
3. Set $k=k+1$ and go to 2.

Then, under mild conditions on the conditional densities, as $k\to+\infty$ $x_{1\to n}^{(k)}$ will be distributed according to $\pi\left(x\right)$.

One sees that although the Gibbs sampler is a particular case of chained MH algorithms, it is - intuitively - more valuable than a general MH algorithm insofar as this later algorithm proceeds in two steps which one could call "more or less crude proposition" and "refinement of the proposition" on the contrary of the Gibbs sampler which makes "clevered" propositions.

5. Application

5.1. Sampling scheme

In our case, when performing the Gibbs sampler, one would need to draw from the 5 following densities:

$$
\begin{array}{l}
p\left(t_{1\to p}\Big/ y_{1\to n}, \underline{\theta}_{1\to p}, \underline{\gamma}_n, \underline{\gamma}_t, \underline{\gamma}_{\underline{\theta}}\right) \\
p\left(\underline{\theta}_{1\to p}\Big/ y_{1\to n}, t_{1\to p}, \underline{\gamma}_n, \underline{\gamma}_t, \underline{\gamma}_{\underline{\theta}}\right) \\
p\left(\underline{\gamma}_\alpha\Big/ y_{1\to n}, \underline{\theta}_{1\to p}, t_{1\to p}, \underline{\gamma}_\beta, \underline{\gamma}_\gamma\right)
\end{array}
\tag{15}
$$

where α, β, γ are to be taken in $\{n, t, \underline{\theta}\}$. Then using the Bayes rule it yields, noting that $\propto$ means "is proportional to" and $\mathcal{S} = \left\{t_{1\to p}, \underline{\theta}_{1\to p}, \underline{\gamma}_n, \underline{\gamma}_t, \underline{\gamma}_{\underline{\theta}}, y_{1\to n}\right\}$:

$$
\begin{array}{rcl}
p\left(t_{1\to p}/\mathcal{S}\backslash\{t_{1\to p}\}\right) & \propto & \mathcal{L}\left(y_{1\to n}\right) p\left(t_{1\to p}/\,\underline{\gamma}_t\right) \\
p\left(\underline{\theta}_{1\to p}/\mathcal{S}\backslash\left\{\underline{\theta}_{1\to p}\right\}\right) & \propto & \mathcal{L}\left(y_{1\to n}\right) p\left(\underline{\theta}_{1\to p}/\,\underline{\gamma}_{\underline{\theta}}\right) \\
p\left(\underline{\gamma}_\alpha/\mathcal{S}\backslash\left\{\underline{\gamma}_\alpha\right\}\right) & \propto & p\left(\alpha_{1\to n_\alpha}/\,\underline{\gamma}_\alpha\right) p\left(\underline{\gamma}_\alpha\right)
\end{array}
\tag{16}
$$

with

$$
\mathcal{L}\left(y_{1\to n}\right) = p\left(y_{1\to n}\Big/ t_{1\to p}, \underline{\theta}_{1\to p}, \underline{\gamma}_n, \underline{\gamma}_t, \underline{\gamma}_{\underline{\theta}}\right) \tag{17}
$$

The choice of this sampling scheme is motivated by the fact that we group the correlated random variables, in order to improve the rate of convergence of the algorithm. At this point one could think that applying a simple resampling procedure would be accurate. This would consist for the five cases in drawing according to the prior of the parameter, and then accept or reject according to the likelihood of the data conditionaly to the parameters [3]. Nevertheless, for the first two cases, due to the shape of the likelihood $\mathcal{L}$ and the fact that computers have finite precision representation of numbers - and thus rounding small values of the likelihood to zero - new parameters will be rarely accepted, if ever...

These two steps of the sampling scheme will thus require the use of a MH algorithm, which allows one to circumvent this problem. Furthermore, we are in the general case confronted with a tradeoff between the speed of convergence of the algorithm, and numerical limits, which prevent us from drawing the whole $t_{1\to p}$ or $\underline{\theta}_{1\to p}$ at a time. In the case when the $t_{1\to p}$ or $\underline{\theta}_{1\to p}$ are correlated, one notes that:

$$
\begin{array}{rcl}
p\left(t_{i\to j}/\,\underline{\gamma}_t, t_{1\to i-1}, t_{j+1\to p}\right) & \propto & p\left(t_{i\to j}/\,\underline{\gamma}_t, t_{1\to i-1}\right) \\
 & & p\left(t_{j+1\to p}/\,\underline{\gamma}_t, t_{1\to i-1}\right)
\end{array}
$$

which with the help of (5) allows one to sample $t_{i\to j}$ conditionaly to other times.

5.2. Simulation

We treat here the simple case when the time t_n depends only on the number of points that occured in the past. It is a zero memory self-excited point process. The priors we take are as follows:

$$
\begin{array}{rcll}
\lambda\left(u/\,N\left(u\right), T\right) & = & \mathcal{N}\left(\left(N\left(u\right)+1\right)T, \sigma_t\right) & \\
T & \sim & \mathcal{U}\left(0, M\right) & (18) \\
\underline{\theta} & \sim & \mathcal{N}\left(m_{\underline{\theta}}, \sigma_{\underline{\theta}}\right) & \\
n_k & \sim & \mathcal{N}\left(0, \sigma_n\right) & (19)
\end{array}
$$

The noise is assumed to be white, and the shape to be a decaying exponential.

The priors on the parameters are taken to be

$$
\begin{array}{rcll}
m_\theta & \sim & \mathcal{U}\left(0, M\right) & \\
\sigma_\alpha & \sim & Inv-\chi^2\left(\upsilon, \mu\right) & (20)
\end{array}
$$

so that

$$
p\left(\sigma_\alpha/\mathcal{S}\backslash\left\{\sigma_\alpha\right\}\right) = Inv-\chi^2\left(\upsilon+n_\alpha, \frac{\upsilon\mu+n_\alpha\tau}{\upsilon+n_\alpha}\right) \tag{21}
$$

where

$$
\tau = \frac{1}{n_\alpha}\sum_{j=1}^{n_\alpha}\left(\alpha_i - m_\alpha\right)^2
$$

We have taken the following values: $T = 30.0, \sigma_t = 5.0, m_{\underline{\theta}} = 1.0, \sigma_{\underline{\theta}} = .1, \upsilon = 2$ (this ensures infinite variance for the prior) $\mu = .3$. We took the following realization of the process, giving the 14 following values for the 7 impulses:

Time	35.8	63.1	90.4	121.7	146.5	188.4	210.3
Amp.	1.17	1.02	1.08	.855	.929	1.12	.93

We then performed the algorithm for $\sigma_n = .3$. We present the estimates of several conditional densities of the parameters. Due to the lack of space, we only give the conditional expectations of the times and amplitudes. Finaly we give the burn-in period of the estimates of the chains for the times, for $\sigma_n = .3$ and $\sigma_n = .4$: the algorithm converges very fast in the first case and requires more iterations in the second case. One also notices that the variance naturally increases. The results are quite good insofar as we estimate all the parameters and the SNR is low. Of course excellent results have been obtained when we have performed the algorithm with $\sigma_n = .1$ and $\sigma_n = .2$.

Time	36.4	63.5	90.5	123.9	148.6	188.5	210.3
Amp.	.88	.92	.85	.55	.77	1.12	.76

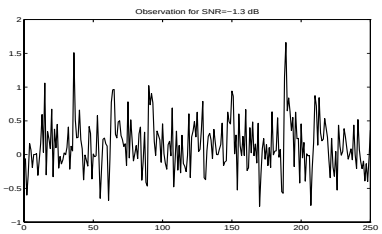

Figure 1: Observations at $-1.3dB$

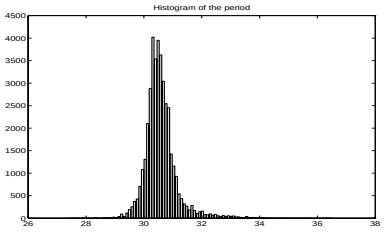

Figure 2: Estimation of $p\left(T/y_{1\to n}\right)$

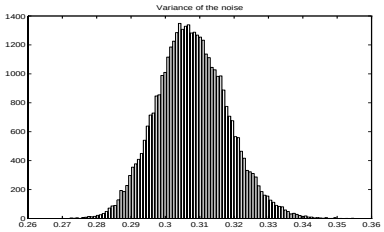

Figure 3: Estimation of $p\left(\sigma_n/y_{1\to n}\right)$

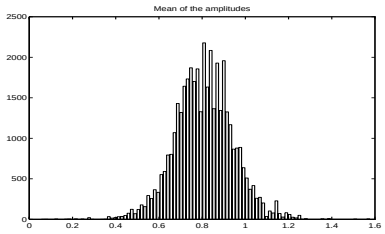

Figure 4: Estimation of $p\left(m_{\underline{\theta}}/y_{1\to n}\right)$

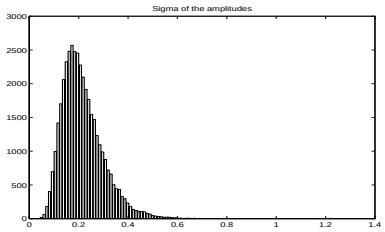

Figure 5: Estimation of $p\left(\sigma_{\underline{\theta}}/y_{1\to n}\right)$

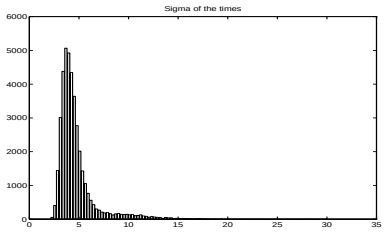

Figure 6: Estimation of $p\left(\sigma_t/y_{1\to n}\right)$

6. Conclusion

We present here an application of a modelization of a large class of filtered point processes to cyclostationary processes which may be very useful in many applications. One of the advantage of the modelization is that it uses all the prior information which is often in practice available, but would be ignored using only correlation properties. Indeed when highly time disturbed we have shown that such process may be weakly cyclostationary. In order to estimate the parameters of the given model, we propose an original stochastic algorithm, based on MCMC, which have not become yet popular in the signal processing community, despite their numerous applications. Furthermore, such an approach could be extended to the deconvolution of superimposed filtered point processes, and the case of the unknown a priori number of impulses will be treated in a later paper.

References

[1] W.A. Gardner, Statistical Spectral Analysis: A Non Probabilistic Theory. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[2] O. Macchi, "The Coincidence Approach to Stochastic Point Processes", Adv. Appl. Prob. 7, 83-122 (1975).

[3] A.F.M. Smith G.O. Roberts 'Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods' J.R. Statist. Soc. B (1993) 55, No. 1, pp. 3-23.

[4] D. L. Snyder, M. I. Miller "Random Point Processes in Time and Space", Second Edition Springer-Verlag 1991.

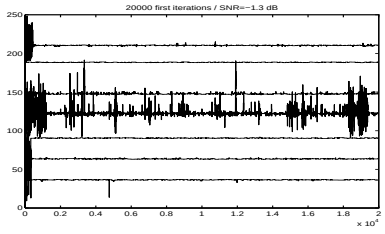

Figure 7: Burn-in period of times at $-1.3dB$

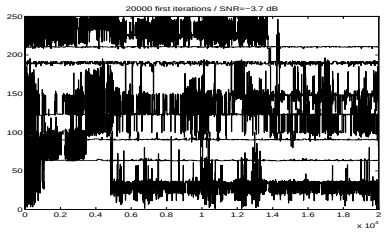

Figure 8: Burn-in period of times at $-3.7dB$