BAYESIAN ESTIMATION AND DETECTION OF SHOT NOISE PROCESSES USING REVERSIBLE JUMPS

Christophe ANDRIEU - Patrick DUVAUT
ENSEA-ETIS Groupe Signal / 6, avenue du Ponceau 95014 Cergy Cedex France.
andrieu@ensea.fr-duvaut@ensea.fr

ABSTRACT

In this paper we propose an original algorithm for the Bayesian joint estimation and detection of shot noise processes. The solution we propose relies on Markov chain Monte Carlo methods and provides the *a posteriori* probability density of the unknown parameters conditionally to the observations. The solution we propose provides many degrees of freedom for the inclusion of any *a priori* knowledge.

1. INTRODUCTION

Shot noise processes are very important in many fields of physics and applied physics, since they model a surprisingly large amount of phenomenons [5], for which part of the information is characterized by the localization of events in a continuum. Achieving the detection and the estimation of such processes may not be easy due to observation noise and overlapping. What we propose in this paper is a Bayesian solution to this problem using powerful stochastic algorithms, which allow to jointly solve the detection/estimation problem, the Markov chain Monte Carlo (MCMC). Our procedure provides an estimation of the *a posteriori* probability density of all the parameters conditionally to the observations. Until recently the detection, that is the determination of the number of events, could not be simply treated using a classical MCMC [1], since they did not allow for dimension change in the set of parameters. Recently, the revolutionary paper [4] has provided a general framework to treat this problem, allowing the different dimensions to cooperate and share information.

2. MODELIZATION/NOTATIONS/GOAL

The observed continuous time process can be written in the following way,

$$y(t) = \sum_{k=1}^{K(L)} h(t - \tau_k, \boldsymbol{\theta}_k) + n(t)$$

for $t \in [0, L]$, where $K(L)$ is the (random) number of events which occured during the observation time, with arrival times $\tau_{1\to K(L)}$ $(\tau_1 < \tau_2 < \ldots < \tau_{K(L)})$ and other characteristic random parameters $\boldsymbol{\theta}_{1\to K(L)}$ of the response of the shot-noise, $h(.,.)$. $n(t)$ is the observation noise, whose parameter vector is $\boldsymbol{\Upsilon}$. Here and in the whole paper we use the following useful notation $\mathbf{X}_{1\to N} \triangleq \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$. The statistics of $\tau_{1\to K(L)}$, which we model as a generalized Poisson process, are provided by the intensity of the process, $\lambda(., \boldsymbol{\Lambda})$, which may depend on both time and past of the process. $\boldsymbol{\Lambda}$ is the set of the possibly unknown parameters of the intensity.

In practice the process is sampled at frequency $\nu_s = 1/T_s$ and we observe for $n = 1, \ldots, N$:

$$y_n \triangleq \sum_{k=1}^{K(L)} h(nT_s - \tau_k, \boldsymbol{\theta}_k) + n(nT_s)$$

The aim of the paper is the estimation of the following probability density:

$$\pi(K(L), \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\Upsilon} / y_{1\to N}) \tag{1}$$

with $\boldsymbol{\tau} = \tau_{1\to K(L)}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_{1\to K(L)}$. This density probability gathers all the information concerning the unknown parameters provided by the observations and the, possibly non informative, *prior* on the parameters.

3. SOLUTION USING MCMC

The estimation of (1) requires the estimation of untractable high-dimensional integrals. To circumvent this problem we use a Markov chain Monte Carlo algorithm, which consists in building a Markov chain whose equilibrium density is (1). The algorithm we propose is a hybrid MCMC since it consists in combining several Metropolis-Hastings-Green (MHG) steps whose principle is briefly recalled below.

3.1. The MHG move

Suppose one wants to sample from $\pi(d\mathbf{X})$ where $\mathbf{X}$ may change of nature or size ($\pi(d\mathbf{X})$ is a probability distribution, not a density $\pi(\mathbf{X})$!).

1. Initialize the chain with $\mathbf{X}^0$ and set $k = 1$
2. Propose $\mathbf{X}^*$ from $q(\mathbf{X}^{k-1}, .)$
3. Evaluate $\alpha = \frac{\pi(d\mathbf{X}^*)q(\mathbf{X}^*, d\mathbf{X}^{k-1})}{\pi(d\mathbf{X}^{k-1})q(\mathbf{X}^{k-1}, d\mathbf{X}^*)}$
4. $\mathbf{X}^k = \mathbf{X}^*$ with probability $\min\{1, \alpha\}$, else $\mathbf{X}^k = \mathbf{X}^{k-1}$
5. $k \leftarrow k + 1$, go to 2.

In the case when there is no movement between spaces of different dimensions the probabilities are replaced by their densities, and this is just a classical Metropolis-Hastings (MH) step. When not the case, that is for example when one jumps from space $\mathcal{C}_1$ to space $\mathcal{C}_2$, the proposal densities corresponding to $q(.,.)$ do not take the same form depending on the starting space. Furthermore, those moves can not be arbitrary so as to achieve convergence to the required density, and must be defined by pairs. A typical way to perform a move between spaces will be as follows:

Copyright 1997 IEEE

optionally draw a random vector $\mathbf{U}_i$ (i corresponds to the starting space), which will generally achieve the dimension matching between spaces, and propose $\mathbf{X}_k^* = g_i(\mathbf{X}_{k-1}, \mathbf{U}_i)$, where $g_1(.,.) = g_2^{-1}(.,.)$ is an invertible deterministic transformation. The choice between those two moves is made randomly with probability b_k.

In what follows, we note $MHG^1(\pi, q)$ a step of this algorithm or a combination of such steps which update the components of the parameter one after the other, leading to a so called "MH one at a time". Suppose for example that one wants to sample from $\pi(\mathbf{X}_{1\to n})$ then proceed as follows:

1. Initialize the chain with $\mathbf{X}_{1\to n}^0$ and set $k = 1$
2. For $i = 1, \dots, n$
   (a) Propose $\mathbf{X}_i^*$ from $q(\mathbf{X}_{1\to i-1}^k \mathbf{X}_{i\to n}^{k-1}, .)$
   (b) Evaluate $\alpha = \frac{\pi(d\mathbf{X}_{1\to i-1}^k d\mathbf{X}_i^* d\mathbf{X}_{i+1\to n}^{k-1})}{\pi(\mathbf{X}_{1\to i-1}^k \mathbf{X}_{i\to n}^{k-1})} \times \frac{q(\mathbf{X}_{1\to i-1}^k \mathbf{X}_i^* \mathbf{X}_{i+1\to n}^{k-1}, d\mathbf{X}_i^{k-1}, .)}{q(\mathbf{X}_{1\to i-1}^k \mathbf{X}_{i\to n}^{k-1}, d\mathbf{X}_i^*, .)}$
   (c) $\mathbf{X}_i^k = \mathbf{X}_i^*$ with probability $\min\{1, \alpha\}$, else $\mathbf{X}_i^k = \mathbf{X}_i^{k-1}$
3. $k \leftarrow k + 1$, go to 2.

3.2. The hybrid sampler

Then the whole algorithm, the hybrid sampler, can be written as follows:

1. Initialize $K^0(L), \boldsymbol{\tau}^0, \boldsymbol{\theta}^0, \boldsymbol{\Lambda}^0, \boldsymbol{\Upsilon}^0$ and $k = 1$.
2. Choose with probability b_k among the two following movements (add or delete an impulse):
   (a) $MHG^1(\pi(K(L), \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\Upsilon}/y_{1\to N}), q_b)$
   (b) $MHG^1(\pi(K(L), \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\Upsilon}/y_{1\to N}), q_d)$
3. $MHG^1\left(\pi\left(\tau_{1\to K^k(L)}^k / \overline{\tau_{1\to K^k(L)}^k}\right), q_\tau\right)$
4. $MHG^1\left(\pi\left(\boldsymbol{\theta}_{1\to K^k(L)}^k / \overline{\boldsymbol{\theta}_{1\to K^k(L)}^k}\right), q_\theta\right)$
5. $MHG^1(\pi(\boldsymbol{\Lambda}/\overline{\boldsymbol{\Lambda}}), q_\Lambda)$
6. $MHG^1(\pi(\boldsymbol{\Upsilon}/\overline{\boldsymbol{\Upsilon}}), q_\Upsilon)$
7. $k \leftarrow k + 1$ and go to 2.

where $\overline{\mathbf{X}} \triangleq \{K(L), \boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{\Lambda}, \boldsymbol{\Upsilon}, y_{1\to N}\} \setminus \{\mathbf{X}\}$ at the current step of the algorithm. This algorithm converges to the required density under mild conditions [6] we do not detail here.

4. APPLICATION

We have treated the following application: $\lambda(kT_s) = .025\exp(-\cos(2\pi.1k))$, $h(t - \tau_i, \alpha_i, a_i, b_i, \nu_i) = \mathbb{I}_{t \geq \tau_i} e^{-\alpha_i(t-\tau_i)}[a_i \cos(2\pi\nu_i(t - \tau_i)) + b_i \sin(2\pi\nu_i(t - \tau_i))]$ and the noise is white Gaussian $n_k \sim \mathcal{N}(0, \sigma_n)$. In the application $\sigma_n = .5$.

4.1. Prior and instrumental density

The priors we have chosen are as follows: $[a_i, b_i]^T \sim \mathcal{N}(\mu_{ab}, \sigma_{ab})$, $\nu_i \sim \mathcal{U}_{[0,.5]}$ (the $(a_i, b_i, \nu_i)_{i=1,\dots,K(L)}$ are iid) and $\sigma_n \propto 1/\sigma_n$ which is a conjuguate non-informative prior. In what follows we do not identify the $\alpha_{1\to K(L)}$ we suppose it set to 25. If they were unknown one would take for example $\exp(-\alpha_i) \sim \mathcal{U}_{[0,1]}$. In order to estimate $\lambda(t, \boldsymbol{\Lambda})$, we propose to model this intensity by a piecewise constant periodic function, with an unknown number $P(T)$ of intervals $I_{1\to P(T)}$ characterized by the positions of their middles $\xi_{1\to P(T)}$ ($\xi_1 < \xi_2 < \dots < \xi_{P(T)}$), modeled as a homogeneous Poisson process with intensity λ_ξ and their log-levels $\eta_{1\to P(T)}$. Thus $\lambda(t, \boldsymbol{\Lambda}) = \sum_{k=1}^{P(T)} \mathbb{I}_{I_k}(t \bmod T)\exp(\eta_k)$ (we suppose here that T is known). We borrow the idea of [2] for the *prior* of the intensity, that is $\eta_{1\to P(T)}/\xi_{1\to P(T)} \sim \mathcal{N}(\boldsymbol{\mu}_\eta, \Sigma_\eta)$ with $\Sigma_\eta$which penalizes high jumps between neighbour large steps: it provides smoothness to the solution (see 4.4. and [2] for more details). The relative simplicity and the generality of this approach have motivated our choice.

The instrumental densities are $q_{ab}(a, b) \propto \pi(a, b/\overline{a_i, b_i})\,\pi_{ab}(a, b)$, $q_\nu(\nu) \propto \sum_{k=1}^{N_\nu} \pi\left(\nu(k)/\overline{\nu(k)}\right)\mathbb{I}_{[\nu_k - \eta_\nu, \nu_k + \eta_\nu]}(\nu)$, that is we provide as proposal density in the last case an approximation of the conditional densities of these parameters by a mixture of uniform laws weighted by some discrete values of the real conditional densities. This method provides in practice a good mixing of the Markov chain [3], and preserves the "MCMC spirit", circumventing any cumbersome maximization necessary in other proposed methods. In our case $\nu(k) = .5\frac{k-1}{N/2}$ ($k = 1, \dots, N/2$) which allows the use of a classical FFT. When one wants to estimate $\alpha_{1\to K(L)}$ then one can propose $q_\alpha(\alpha) \propto \sum_{k=1}^{N_\alpha} \pi\left(\alpha(k)/\overline{\alpha(k)}\right)\mathbb{I}_{[\alpha_k - \eta_\alpha, \alpha_k + \eta_\alpha]}(\alpha)$. The first density can be evaluated directly in our case. We note $\Sigma_i^{-1} = \frac{\sum_{k=1}^N \mathbf{H}_i \mathbf{H}_i^T}{\sigma_n^2}$, $\boldsymbol{\mu}_i = \Sigma_i \sum_{k=1}^N \mathbf{H}_i y_k$ with $\mathbf{H}_i^T = [\,h(0T_s - \tau_i)\ \dots\ h((n-1)T_s - \tau_i)\,]$. Then with $S_i^{-1} = \Sigma_i^{-1} + \Sigma_0^{-1}$ and $\mathbf{m}_i = S_i(\Sigma_0^{-1}\boldsymbol{\mu}_0 + \Sigma_i^{-1}\boldsymbol{\mu}_i)$, $q_{ab}(a_i, b_i) \propto \exp\left(-\frac{1}{2}\left(\begin{bmatrix} a_i \\ b_i \end{bmatrix} - \mathbf{m}_i\right)^T S_i^{-1}\left(\begin{bmatrix} a_i \\ b_i \end{bmatrix} - \mathbf{m}_i\right)\right)$. Obviously this provides a "Gibbs like" step, since in that later case the acceptance rate is 1. Due to the choice on the *prior* of σ_n one can choose $q(\sigma_n/\overline{\sigma_n})$ such that $\sigma_n^* \sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{2N}\sum_{i=1}^N\left(y_i - \sum_{k=1}^{K(L)} h(iT_s - \tau_k, \boldsymbol{\theta}_k)\right)^2\right)$ which is also a "Gibbs like" step. We propose a random walk for the update of the $\tau_{1\to K(L)}$ that is $\tau_i^* \sim \mathcal{N}(\tau_i^{k-1}, \sigma_\tau)$. In this last case the variance is chosen quite small, 0.01.

4.2. "Birth-Death" move

We provide here more details on the propositions of the dimension change. One chooses with probability b_k (we have taken $b_k = .5$) between the following moves:

- "birth move" is done in the following manner:
  1. Draw $\tau_{p+1} \sim \frac{\lambda(.,\boldsymbol{\Lambda})}{\int_0^L \lambda(u,\boldsymbol{\Lambda})du}$
  2. Draw $\nu_{p+1} \sim \mathcal{U}_{[0,.5]}(.)$
  3. Draw $(a_{p+1}, b_{p+1})^T \sim \pi\left(a_{p+1}, b_{p+1}/\overline{a_{p+1}, b_{p+1}}\right)$

Copyright 1997 IEEE

Thus, $q_b(a_{p+1}, b_{p+1}, \nu_{p+1}, \tau_{p+1}) = b_k \frac{\lambda(\tau_{p+1},\Lambda)}{\int_0^L \lambda(u,\Lambda)du} \times 2\mathbb{I}_{[0,.5]}(\nu_{p+1}) \times \pi\left(a_{p+1}, b_{p+1}/\overline{a_{p+1}, b_{p+1}}\right)$ (The last density can be evaluated directly and easily from $q_{ab}(a_i, b_i)$).

- "death move", which is the reverse move, is as follows:
  1. Draw a point among the current $p+1$ points.
  2. Suppress it from the current estimation.

Thus $q_d(p+1) = (1-b_k) \times 1/(p+1)$

4.3. Non informative prior

The choice of the *prior* is very important in our case. Indeed when there is no dimension change, one can choose $\mu_{ab}=0, \sigma_{ab}^{-1} \to 0$ which provides non informative *prior* on the amplitudes. Nevertheless this is not easily done when there are dimension changes since such *prior* provide information on the dimension, thus leading to arbitrary model choice. This lead us to add the following constraint: we have set the following ratio $\frac{(\sigma_{ab}^{p+1})^{p+1}}{(\sigma_{ab}^{p})^{p}}.5$ to 1.0 (σ_{ab}^k corresponds to dimension k and .5 corresponds to the probability of the frequency) which cancels the influence of the *prior* on the Bayes factors, that is on the choice of the dimension. Then one can take the limit above which removes information of the *prior* within a dimension. The ratio of the *prior* on the amplitudes and the frequency for a birth is thus $1/4\pi$. Of course other *prior* could be proposed, (depending on knowledge about the problem) providing another penalization of the likelihood and thus another Bayesian model selection rule.

Thus $\alpha_{birth} = Likelihood\_Ratio \times \lambda(\tau_{p+1}, \Lambda) \times 1/4\pi \times q_d(p+1)/q_b(a_{p+1}, b_{p+1}, \nu_{p+1}, \tau_{p+1})$. Of course $\alpha_{death} = 1/\alpha_{birth}$.

4.4. The intensity

Here we provide some elements on the way the intensity is estimated. For more details the reader is invited to refer to [2]. $[\Sigma_\eta]_{(k,k)} = \sigma^2 \frac{\xi_{k+1}-\xi_{k-1}}{2}$ and $[\Sigma_\eta]_{(k,k+1)} = -\sigma^2 \beta \frac{\xi_{k+1}-\xi_k}{2}$ (we do not detail the obvious cases of the first and last steps) where β is near to 1 so as to cancel the effect of μ_η ($\mu_\eta = -3$ is our choice), and σ controls the smoothness of the intensity ($\sigma = .5$ is our choice). The addition of a move is made in the same spirit as for an impulse, but with constraints: for the birth of a new interval between ξ_k and ξ_{k+1}, the new intensity is $\eta_{new} = \frac{\xi_{new}-\xi_k}{\xi_{k+1}-\xi_k}\eta_k + \frac{\xi_{k+1}-\xi_{new}}{\xi_{k+1}-\xi_k}\eta_{k+1} + \varepsilon$, where ε is a random variable, drawn from an instrumental density (that is almost free, see [2]). The death move is made with the reverse transformation. In our application $\lambda_\xi = 50$.

5. RESULTS

The process we have generated has the following original parameters:

i	1	2	3	4	5	6
τ	.020	.172	.219	.376	.726	1.164
a	-1.16	-4.58	-.73	-1.76	-1.76	-.63
b	-1.22	.86	4.48	1.74	1.89	-3.15
ν	.37	.12	.44	.12	.14	.16

The figures we present are the following:

- The observation and the original process.
- The estimation of $p(K(L)/y_{1\to N})$.
- The superimposed estimations of $p(\tau_i/y_{1\to N}, K(L))$ for $K(L) = 4, 5, 6$ and $i = 1, \ldots, (L)$.
- The restoration of the process for $K(L) = 6$ with the marginal MAP values (read on the graphs).
- The estimation of $p(P(L)/y_{1\to N})$
- The estimation of the intensity with $P(L) = 6$ with the marginal MAP values (read on the graph).
- We do not present results on the noise, since it is always one of the easiest parameter to identify.

Those results are interesting, since they reflect the uncertainty of a decision in the presence of noise: small impulses and noise can be mistaken, and overlapping can cause "identification like" problems. The first impulse seems to be ignored, and another pulse is added between pulse 2 and 3.

6. CONCLUSION

We propose in this paper a Bayesian solution to the problem of joint detection/estimation of shot-noise process using MCMC. Those algorithms allow the estimation of the *a posteriori* probability density of the parameters conditionally to the observations, which concentrates all the probabilistic information concerning those parameters. Results are satisfactory and the method provides many degrees of freedom for regularization, thus allowing the natural inclusion of any *prior* knowledge. Furthermore this algorithm requires only to run a single Markov chain which makes the best use of the common information of the different models, which is usually impossible using other methods.

REFERENCES

[1] C. Andrieu, A. Doucet and P. Duvaut, " Bayesian Deconvolution of Cyclostionary Processes based on Point Processes ", *Eusipco 1996 Trieste, Italy*

[2] E. Arjas and J. Heikkinen, "An Algorithm for nonparametric Bayesian Estimation of a Poisson intensity", *preprint from Dpt. of Statistics Univ. of Jyväskylä* 1996.

[3] A. Doucet, C. Andrieu and P. Duvaut, "Joint Bayesian Detection and Estimation of Sinusoids in Noise", *submitted to IEEE SP*, 1996.

[4] P.J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, 82, 711-732, 1995.

[5] D.L. Snyder and M.I. Miller, *Random Point Processes in Time and Space*, 2^{nd} edition, Springer-Verlag 1991.

[6] L. Tierney, "Markov Chains for Exploring Posterior Distributions", *The Annals of Statistics*, Vol. 22, No. 4, 1701-1762, 1994.

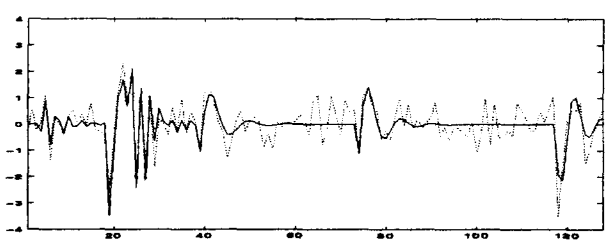

Figure 1. Observation (dotted lines) - Original

Copyright 1997 IEEE

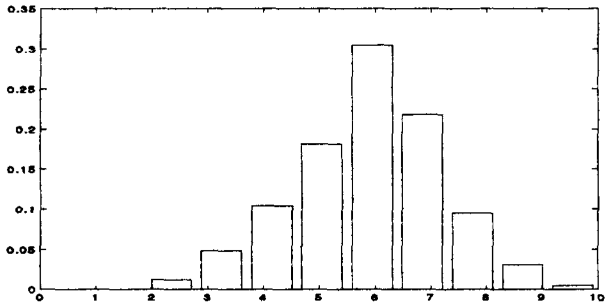

Figure 2. Estimations of $p\,(K\,(L)/\,y_{1\rightarrow N})$

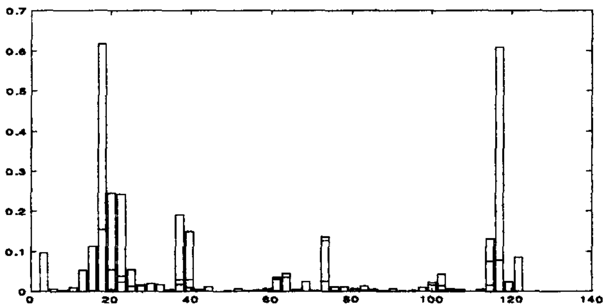

Figure 3. Estimations of the $p\,(\tau_i/\,y_{1\rightarrow N}, K\,(L) = 4)$

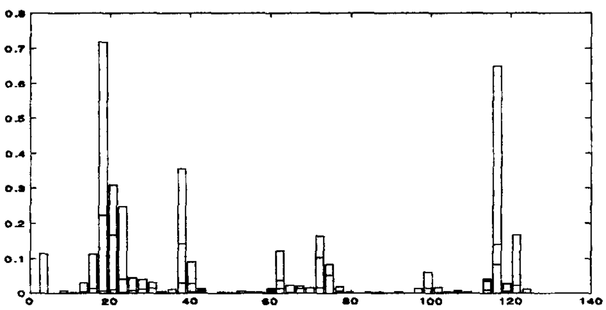

Figure 4. Estimation of the $p\,(\tau_i/\,y_{1\rightarrow N}, K\,(L) = 5)$

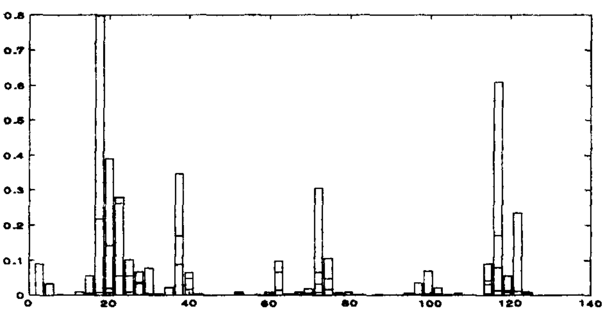

Figure 5. Estimations of the $p\,(\tau_i/\,y_{1\rightarrow N}, K\,(L) = 6)$

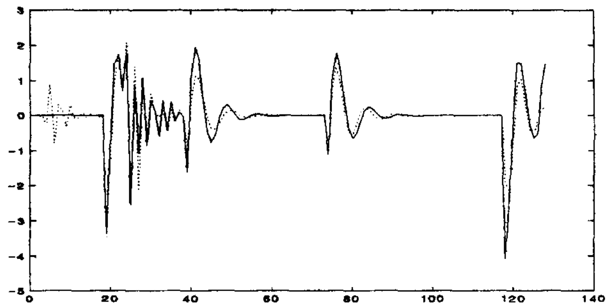

Figure 6. Restoration of the process for $K\,(L) = 6$

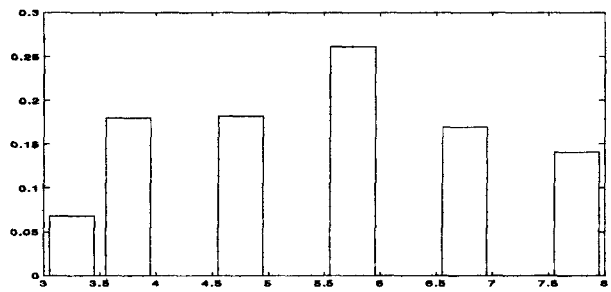

Figure 7. $p(\,P\,(L)/\,y_{1\rightarrow N})$

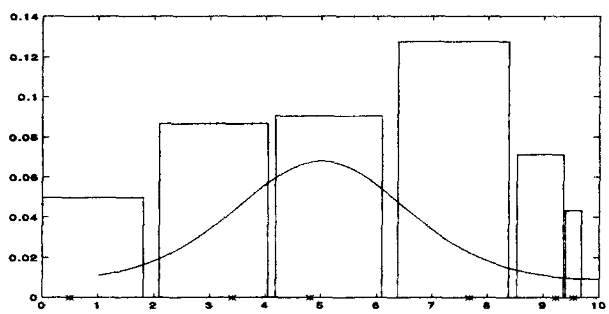

Figure 8. Restoration of the intensity for $P\,(L) = 6$

Copyright 1997 IEEE