WAVELET-DOMAIN MODELING AND ESTIMATION OF POISSON PROCESSES

Klaus E. Timmermann and Robert D. Nowak

Michigan State University
East Lansing, MI 48824
timmerm4@egr.msu.edu nowak@egr.msu.edu

ABSTRACT

This paper develops a new wavelet-domain Bayesian framework for modeling and estimating the intensity of a Poisson process directly from count observations. A new multiscale, multiplicative innovations model is developed as a prior for the underlying intensity function. The new prior model leads to a simple and efficient closed-form estimator that requires $O(N)$ computations, where N is the dimension of the intensity function. We compare the new method with previously proposed wavelet-based approaches to this problem.

1. INTRODUCTION

This paper considers the problem of estimating the intensity of a Poisson process from a single observation of the process. We observe counts

$$\boldsymbol{c} \sim \text{Poisson}(\boldsymbol{\lambda}) \tag{1}$$

where $\boldsymbol{\lambda}$ is a 1-d or 2-d intensity function. The intensity function is discretized and to simplify the presentation we assume that it is a 1-d signal vector, although all results are easily extended to images. Furthermore, we assume that both $\boldsymbol{c}$ and $\boldsymbol{\lambda}$ are $N \times 1$ vectors. The contribution of this paper is a novel modeling and estimation framework for Poisson processes based in the wavelet-domain. The Poisson processes are encountered in many applications including medical and astronomical imaging.

There are several reasons for adopting a wavelet-domain framework:

- Real-world intensity functions often display self-similarity across scales — wavelet transforms provide efficient representations
- The Poisson distribution is self-reproducing across scale — sum of Poisson variates is Poisson
- Coarse-scale estimators of intensities are very reliable (high SNR) — reliable information can be passed to finer scales
- Useful Bayesian priors are easily specified in the wavelet-domain

We will discuss some of these motivations in more detail later.

In this paper we describe a new probability model for intensity functions called the *multiscale multiplicative innovations* (MMI) model. The MMI model leads to a very simple and powerful Bayesian procedure for estimating the intensity of a Poisson process. Other wavelet-based estimators for Poisson intensity estimation have been proposed in the literature. A simple wavelet-based approach to this problem is to take the square-root of the counts (a variance stabilizing transformation that makes the data approximately Gaussian) and then apply standard wavelet thresholding techniques for noise removal. A more sophisticated wavelet thresholding approach is taken in [1] for the estimation of burst-like Poisson processes. A wavelet-based method for the estimation of more general Poisson intensities is developed in [2]. In this approach, the PRESS-optimal estimator developed in [3] employs the method of cross-validation to design wavelet-domain filters for intensity estimation. These methods can provide satisfactory results in certain situations. However, none of these methods adopts a Bayesian perspective and hence do not explicitly make use of prior information that may be available. We will show that Bayesian estimation procedure developed in this paper can significantly outperform existing methods.

2. THE HAAR WAVELET TRANSFORM

This paper uses the following wavelet representation and notation throughout. Let $\mathbf{c}_0$ designate the data sequence of counts $\boldsymbol{c}$ of length $N = 2^J$, and let $c_{0,k}$ be its k^{th} element. The subscript 0 denotes the finest scale (resolution) of analysis. A multiscale analysis of $\mathbf{c}_0$ can be obtained by iterating

$$\begin{aligned} c_{j,k} &= c_{j-1,2k} + c_{j-1,2k+1} \\ d_{j,k} &= c_{j-1,2k} - c_{j-1,2k+1} \end{aligned} \tag{2}$$

for $j = 1, \ldots, J$ and $k = 0, \ldots, N/2^j - 1$, and $J = \log_2(N)$. Here, J denotes the coarsest scale of analysis.

$c_{j,k}$ and $d_{j,k}$ are termed the scaling and wavelet coefficients of the data, respectively, at scale j and position k. These coefficients are simply the unnormalized Haar transform coefficients. The coefficients are not normalized so that the Poisson nature of the data is preserved at all scales. Furthermore, since the data are obtained by counting the number of events occuring in disjoint regions of space of equal size (intervals in 1-d, pixels in 2-d), the Haar transform is ideally suited for this problem. The scaling coefficients $\mathbf{c}_j = \{c_{j,k}\}_{k=0}^{N/2^j-1}$ represent a lower resolution representation of the data $\mathbf{c}_{j-1}$. The "detail" information in $\mathbf{c}_{j-1}$, which is absent in $\mathbf{c}_j$, is conveyed by the sequence of wavelet coefficients $\mathbf{d}_j = \{d_{j,k}\}_{k=0}^{N/2^j-1}$. Note that $\mathbf{c}_{j-1}$ can be perfectly reconstructed from $\mathbf{c}_j$ and $\mathbf{d}_j$.

In a similar fashion, define the scaling coefficients $\lambda_{j,k}$ and the wavelet coefficients $\theta_{j,k}$ of the intensity function $\boldsymbol{\lambda}$. Note that

This work was supported by the National Science Foundation, grant no. MIP–9701692

$\lambda_{j,k} = E[c_{j,k}]$ and $\theta_{j,k} = E[d_{j,k}]$, where $E[\cdot]$ denotes the expectation operator. Hence, the intensity estimation problem is equivalent to estimating the means of the scaling and wavelet coefficients of the data and then inverting the transformation.

3. A NEW PROBABILITY MODEL FOR INTENSITY FUNCTIONS

To formulate a Bayesian estimator for this problem, we assume a prior probability model for the unknown intensity. The observed data $\boldsymbol{c}$ is the realization of a random sequence $\boldsymbol{C}$ ($\sim$ Poisson($\boldsymbol{\lambda}$)), and $\boldsymbol{\lambda}$ is regarded as an unknown realization of a random sequence $\boldsymbol{\Lambda}$ with prior density $f_{\boldsymbol{\Lambda}}$. Given this prior, we seek the minimum mean-square error (mmse) estimator

$$\begin{aligned}\hat{\boldsymbol{\lambda}} &= E[\boldsymbol{\Lambda}\,|\boldsymbol{C} = \boldsymbol{c}] \\ &= \int_{\boldsymbol{\lambda}} \boldsymbol{\lambda}\, f(\boldsymbol{\lambda}|\boldsymbol{c})\, d\boldsymbol{\lambda} \end{aligned} \qquad (3)$$

where $f(\boldsymbol{\lambda}|\boldsymbol{c})$ is the conditional probability density function $f_{\boldsymbol{\Lambda}|\boldsymbol{C}}(\boldsymbol{\lambda}|\boldsymbol{c})$. We will often follow this simplifying convention of implicitly specifying pdf's and probability mass functions by their arguments.

A Bayesian approach facilitates the solution of (3) by expressing it in terms of the prior density $f_{\boldsymbol{\Lambda}}$. Applying Bayes' theorem to (3)

$$\hat{\boldsymbol{\lambda}} = \frac{\int_{\boldsymbol{\lambda}} \boldsymbol{\lambda}\, P(\boldsymbol{c}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}}{\int_{\boldsymbol{\lambda}} P(\boldsymbol{c}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}} \qquad (4)$$

where $P(\boldsymbol{c}|\boldsymbol{\lambda}) = \prod_l \lambda_l^{c_l} e^{-\lambda_l}$. The Bayes' estimator (4) poses two interrelated problems. First, the specification of a meaningful and useful prior $f(\boldsymbol{\lambda})$. Second, the numerical computation of the estimator. The remainder of this section describes a new prior probability model for intensities.

There are several important reasons for adopting a multiscale approach to this problem.

- The multiscale decomposition is fractal in nature, so the self-similar nature of real-world intensities is easily modeled.
- The multiscale decomposition preserves the Poisson characteristic of the data at each scale due to the reproducing property of the Poisson distribution, i.e., $c_i \sim$ Poisson(λ_i), c_i independent $\Rightarrow \sum c_i \sim$ Poisson($\sum \lambda_i$).
- Prior models that are mathematically tractable, computationally practical, and empirically supported, can be specified very naturally.
- Poisson data is much more reliable at coarse scales than at fine resolutions (higher counts $\Rightarrow$ higher signal-to-noise ratio). Therefore, more reliable coarse-scale estimators can be leveraged to better improve higher resolution estimators.

Together, the two first points enable a similar treatment of the data at all scales. This, in turn, leads to a simple algorithm and facilitates the specification of the prior distribution; thus, point three. The fourth point motivates an estimation process that evolves from coarse to fine scales. The estimate of the intensity at scale j can be used to efficiently compute the subsequent estimation step at the next finer scale $j-1$. These assertions will become evident in the next section.

We are now in a position to postulate a novel, wavelet-domain probabilistic model for the intensity. Let $\Lambda_{j,k}$ and $\Theta_{j,k}$ denote the random variables corresponding to the j,k-th scaling and wavelet coefficient of the intensity, respectively. At the coarsest scale $j = J$, the single scaling coefficient $\Lambda_{J,0}$ has a density with support on $\mathbb{R}^+$. For example, the gamma density is conjugate to the Poisson and is especially useful. Next, introduce independent perturbations $\{\Delta_{j,k}\}$ and model the wavelet coefficients by

$$\Theta_{j,k} = \Lambda_{j,k}\, \Delta_{j,k} \qquad (5)$$

Each wavelet coefficient is modeled as independent perturbation of its corresponding scaling coefficient. Furthermore, the perturbations at all scales and positions $\{\Delta_{j,k}\}$ are assumed to be mutually independent. Applying the recursions (3) to these coefficients, we find that $\Lambda_{j-1,k} = \frac{1}{2}(\Lambda_{j,[k/2]} + (-1)^k\, \Theta_{j,[k/2]})$, where $[\cdot]$ stands for the largest integer no greater than its argument. Then, using (5)

$$\begin{aligned}\Lambda_{j-1,2k} &= \Lambda_{j,k}\, \frac{(1+\Delta_{j,k})}{2} \\ \Lambda_{j-1,2k+1} &= \Lambda_{j,k}\, \frac{(1-\Delta_{j,k})}{2}\end{aligned} \qquad (6)$$

We can interpret the refinement in (7) as a multiscale innovations structure, with the innovation $\frac{(1+\Delta_{j,k})}{2}$ entering as a multiplicative, rather than additive, perturbation. We call this model a *multiscale multiplicative innovations* (MMI) model. The model is graphically depicted in Figure 1. The multiplicative innovation structure is well-suited to the Poisson nature of the problem, as demonstrated in the next section.

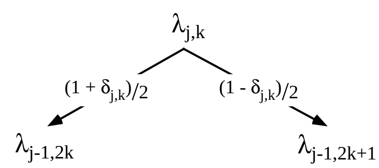

Figure 1: *Multiscale multiplicative innovations model. A coarse-scale probability model of the intensity is refined via multiplicative perturbations.*

The key properties of the MMI model are:

- The model gives rise to signals of a fractal nature, as are typical in real-world imagery. A fractal structure is guaranteed as long as the priors for the perturbations $\{\Delta_{j,k}\}$ are chosen to be self-similar themselves across scales.
- The model is essentially invariant to the length of the observation time. Since we are most interested in temporally homogeneous processes, it would be undesirable if the prior depended on the observation time interval in a complicated manner. Due to the multiplicative innovations structure, only the coarsest scale of the prior is dependent on the observation time interval.
- The model provides a mathematically tractable match to the Poisson nature of the data, as will become apparent in Section 3.

In the MMI model, the prior density f_Δ for $\Delta_{j,k}$ can be chosen as identical for all j,k, or may be defined to be distributed differently at each scale for added flexibility. Location dependence can also be introduced, but we have not pursued this at present. Desired properties for f_Δ include support on the $[-1,1]$ interval, symmetry about the origin, unimodality, and concentration around zero. The first property is due to the fact that the range of $\Theta_{j,k}$ is $[-\Lambda_{j,k}, \Lambda_{j,k}]$. The second arises from the assumption that there is no reason *a priori* to favor $\Lambda_{j-1,2k}$ over $\Lambda_{j-1,2k+1}$, or vice-versa. The third and fourth properties are based on the characteristics of observed wavelet coefficient distributions resulting from natural signals [4].

One very general class of probability density functions that possesses the desired characteristics are beta-mixture densities of the form

$$f(\delta) \;=\; \sum_{i=1}^{M} p_i \frac{(1-\delta^2)^{s_i-1}}{B(s_i, s_i)\; 2^{2s_i-1}} \qquad (7)$$

for $-1 \leq \delta \leq 1$, where B is the Euler beta function, $0 \leq p_i \leq 1$ is the weight of the i-th beta density $\frac{(1-\delta^2)^{s_i-1}}{B(s_i,s_i)\; 2^{2s_i-1}}$ with parameter s_i, and $\sum_{i=1}^{M} p_i = 1$. Figure 2 depicts a mixture of two beta densities.

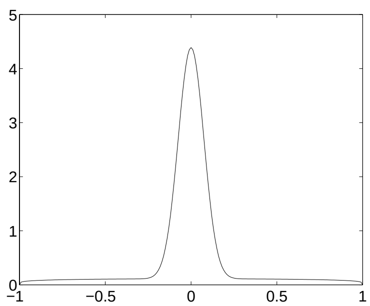

Figure 2: *Two component Beta-mixture distribution with* $p_1 = 1 - p_2 = .2$*,* $s_1 = 1.2$*, and* $s_2 = 90$*.*

4. BAYESIAN ESTIMATION

4.1. Wavelet Coefficient Estimation

Now using the MMI prior model, we can derive a simple wavelet-domain Bayesian estimator of the intensity. Recalling the definition $\boldsymbol{c}_{j-1}$ is the $j-1$ scale representation of the data, the best estimate of the wavelet coefficient $\theta_{j,k}$, given $\boldsymbol{c}_{j-1}$, is formulated as follows.

$$\begin{aligned} \widehat{\theta}_{j,k} = E\left[\Theta_{j,k} | \boldsymbol{c}_{j-1}\right] &= E\left[\Lambda_{j,k}\Delta_{j,k} | \boldsymbol{c}_{j-1}\right] \\ &= E\left[\Lambda_{j,k} | \boldsymbol{c}_{j-1}\right]\; E\left[\Delta_{j,k} | \boldsymbol{c}_{j-1}\right] \quad (8) \end{aligned}$$

where we have exploited the independence between $\Lambda_{j,k}$ and $\Delta_{j,k}$.

Using (7), it can be shown that

$$\begin{aligned} \widehat{\delta}_{j,k} &= E\left[\Delta_{j,k} | \boldsymbol{c}_{j-1}\right] \\ &= d_{j,k}\, \frac{\sum_i p_i \frac{B(s_i + c_{j-1,2k},\, s_i + c_{j-1,2k+1})}{B(s_i,s_i)\;(2s_i + c_{j,k})}}{\sum_i p_i \frac{B(s_i + c_{j-1,2k},\, s_i + c_{j-1,2k+1})}{B(s_i,s_i)}} \quad (9) \end{aligned}$$

Due to page limitations we do not give all the steps here, but the interested reader can refer to [5].

Now, the only remaining issue is the computation of $\widehat{\lambda}_{j,k}$. Here is where we exploit our multiscale structure. Recall that at the coarsest scale J, we have a single scaling coefficient $\lambda_{J,0}$. In this case, it can be shown [5] that the Bayes' estimate $\widehat{\lambda}_{J,0} = E\left[\Lambda_{J,0} | \boldsymbol{c}_{J-1}\right] = E\left[\Lambda_{J,0} | c_{J,0}\right]$. However since $c_{J,0}$ is typically large, the estimate $\widehat{\lambda}_{j,k} = c_{J,0}$ is a very reasonable in practice. With the coarsest scaling coefficient in hand, we can iterate the estimation process by exploiting the linearity of the conditional expectation operator

$$\begin{aligned} \widehat{\lambda}_{j-1,k} &= E\left[\frac{1}{2}(\Lambda_{j,[k/2]} + (-1)^k\,\Theta_{j,[k/2]}) \,\middle|\, \boldsymbol{c}_{j-2}\right] \\ &= \frac{1}{2}\left(\widehat{\lambda}_{j,[k/2]} + (-1)^k\,\widehat{\theta}_{j,[k/2]}\right) \quad (10) \end{aligned}$$

to obtain a higher resolution estimate of the intensity at scale $j-1$. *Note that the resulting intensity estimates are non-negative.* That is, $\widehat{\lambda}_{j,k} \geq 0$ for $j = 0, \ldots, J$ and all k. The overall algorithm is described below.

Wavelet-Domain Bayesian Intensity Estimation

1. Estimate coarsest scale coefficients
   $\widehat{\lambda}_{J,k} = c_{J,k}$
2. For $j = J$ down to 1
   Compute $\widehat{\delta}_{j,k}$ according to (9)
   Compute $\widehat{\theta}_{j,k} = \widehat{\lambda}_{j,k}\widehat{\delta}_{j,k}$
   Combine $\widehat{\lambda}_{j-1,k} = \frac{1}{2}\left(\widehat{\lambda}_{j,[k/2]} + (-1)^k\,\widehat{\theta}_{j,[k/2]}\right)$

In general, the complexity of the proposed estimation filter, implemented by means of the fast wavelet transform, is $\mathcal{O}(N)$, the same order as the fast wavelet transform itself. Thus, the proposed filter is computationally efficient.

4.2. Estimation of Prior Parameters

The distribution of real-world data wavelet coefficients often fit a profile which resembles that of Figure 2, as previously discussed. This makes parametric representation of the distributions very fitting and, therefore, facilitates estimation of the distribution. Here we give a very simple approach based on moment matching.

Let $\mathrm{P}_{j,k}$ be the random variable given by $\mathrm{P}_{j,k} = \frac{1}{2}(1 + \Delta_{j,k})$, and assume that $\Delta_{j,k}$ has an M-component beta-mixture density with parameters $\{p_i, s_i\}_{i=1}^{M}$. Using (6) we obtain

$$\Lambda_{j-1,2k} = \Lambda_{j,k}\mathrm{P}_{j,k} \qquad (11)$$

Since $\Delta_{j,k}$, and therefore $\mathrm{P}_{j,k}$, are independent of Λ, $E\left[\Lambda^n_{j-1,2k}\right] = E\left[\Lambda^n_{j,k}\right]\; E\left[\mathrm{P}^n_{j,k}\right]$ and

$$E\left[\mathrm{P}^n_{j,k}\right] \;=\; \frac{E\left[\Lambda^n_{j-1,2k}\right]}{E\left[\Lambda^n_{j,k}\right]} \qquad (12)$$

The moments $E\left[\Lambda^n_{j-1,2k}\right]$ and $E\left[\Lambda^n_{j,k}\right]$ are easily estimated from the data. For example, $E\left[\Lambda_{j,k}\right] \approx \frac{1}{N/2^j - 1}\sum_k c_{j,k}$ and $E\left[\Lambda^2_{j,k}\right] \approx \frac{1}{N/2^j - 1}\sum_k \left(c^2_{j,k} - c_{j,k}\right)$. Substituting these estimates for various n into (12) produces a set of equations that can be solved for the parameters $\{p_i, s_i\}_{i=1}^{M}$.

5. COMPARISON OF WAVELET-BASED INTENSITY ESTIMATORS

Here we compare the performance of the new Bayesian estimation algorithm with several existing methods. To assess the performance of each method four test intensity functions were used. These functions were the "Doppler," "Blocks," "HeaviSine," and "Bumps" test signals proposed in [6]. Since the intensity functions must be non-negative, the test functions were shifted and scaled to obtain an intensities with a desired peak value and a minimum value of $\frac{1}{\text{peak value}}$. We compare the performance of the PRESS-optimal estimator (PRESS) [2], the new Bayesian estimator (BAYES) described in this paper with a two component beta-mixture model for the innovations with parameters[1] $s_1 = 2$ and $s_2 = 10000$., the square-root estimation methods using the Haar wavelet transform (D2), and the square-root estimation method using the Daubechies-8 (D8) wavelet[2]. The square-root method first computes the square-root of the counts, then treats the square-root data as though it were Gaussian and applies a soft-threshold nonlinearity to "denoise" the data. After denoising the square-root data the result is squared to obtain an intensity estimate. For both square-root methods the universal threshold proposed in [6] was used, where N is the length of the data vector. Table 1 gives the mean-square errors (MSEs) of the various methods for a peak intensity of 8. Table 2 gives the MSEs of each method for a peak intensity of 128. All MSEs are normalized by the norm of the underlying intensity function. Table 1 shows that at low intensities (low signal-to-noise) the Bayesian estimator dramatically outperforms all other methods. Table 2 shows that at higher intensities the performance of the other methods is closer to the Bayes' estimator's performance, but that the Bayes' estimator is still the best overall choice.

Table 1: *MSE results for various test intensities and estimation algorithms. Peak intensity in each case is* 8.

Intensity	*PRESS*	*BAYES*	*D2*	*D8*
Doppler	0.062	0.030	0.056	0.049
Blocks	0.072	0.027	0.071	0.086
HeaviSine	0.064	0.017	0.036	0.032
Bumps	0.219	0.164	0.470	0.477

Table 2: *MSE results for various test intensities and estimation algorithms. Peak intensity in each case is* 128.

Intensity	*PRESS*	*BAYES*	*D2*	*D8*
Doppler	0.006	0.007	0.014	0.008
Blocks	0.005	0.002	0.010	0.014
HeaviSine	0.005	0.003	0.006	0.003
Bumps	0.020	0.025	0.113	0.097

[1] The mixing parameter $p_1 = 1 - p_2$ is determined using the data-adaptive moment matching method given in Section 4.2

[2] The method proposed in [1] is not compared since it is derived under a "burst-like" process model which is not appropriate for these test functions with the exception of the Bumps function

6. CONCLUDING REMARKS

We have introduced a Bayesian approach for Poisson intensity estimation. We argued that wavelet-domain analysis is the right framework for carrying the Bayesian estimation. We introduced a novel MMI prior model for intensity functions based on a *multiplicative* innovations structure. The MMI captures many of the key features of real-world intensity functions and provides an excellent match to the Poisson distribution. The MMI model facilitates a wavelet-domain Bayesian estimation procedure that proceeds in a natural fashion from coarse-to-fine resolutions. The estimator has a simple closed expression and can be implemented in $O(N)$ operations, where N is the dimension of the finest resolution of the discretized intensity.

We also note that the Haar transform is not shift-invariant, and therefore the prior model and estimator developed in this paper are implicitly shift-dependent. This type of shift-dependence can lead to blocking artifacts in the reconstruction. To remedy this potential shortcoming, we may regard the prior and estimator developed above as being conditioned on a particular shift. Then, in the spirit of Bayesian estimation, we can place a prior distribution on all possible shifts (*e.g.*, uniform), and compute the shift-unconditional estimator by averaging over the shift prior. This procedure can be regarded as a Bayesian interpretation of the translation-invariant wavelet de-noising scheme for Gaussian signals introduced in [7].

Finally we have compared the performance of the wavelet-domain Bayesian estimator with other existing methods using a suite of test intensity functions. This comparison shows the new method significantly outperforms other wavelet-based methods and that the two component beta-mixture model for the innovations provides a good prior for the many signals of diverse characteristics. In future work, we plan to investigate extensions of this framework to related point processes.

7. REFERENCES

[1] E. D. Kolaczyk, "Estimation of intensities of burst-like Poisson processes using Haar wavelets," *preprint*, 1997.

[2] R. D. Nowak and R. G. Baraniuk, "Wavelet-based filtering for photon imaging systems," *IEEE Trans. Image Processing*, submitted April 1997.

[3] R. D. Nowak, "Optimal signal estimation using cross-validation," *IEEE Signal Processing Letters*, vol. 4, no. 1, pp. 23–25, 1997.

[4] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, to appear in Special Issue on Theory and Applications of Filter Banks and Wavelets, 1998.

[5] K. E. Timmermann and R. D. Nowak, "Multiscale Bayesian analysis of Poisson data," *Michigan State University, Dept. Elec. Eng. Tech. Rep.*, October, 1997.

[6] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, Dec. 1995.

[7] R. Coifman and D. Donoho, "Translation invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics*, vol. New York: Springer-Verlag, pp. 125–150, 1995.