A wavelet-Fisz algorithm for Poisson intensity estimation

Piotr Fryźlewicz *         Guy P. Nason †

June 24, 2002

Abstract

This article introduces a new method for the estimation of the intensity of an inhomogeneous one-dimensional Poisson process. The wavelet-Fisz transformation transforms a vector of binned Poisson counts to approximate normality with variance one. Theory shows that, asymptotically, the transformed vector is normal and the elements uncorrelated. Hence we can use any suitable Gaussian wavelet shrinkage method to estimate the Poisson intensity. Since the wavelet-Fisz operator does not commute with the shift operator we can dramatically improve accuracy by always cycle spinning before the wavelet-Fisz transform as well as optionally after.

Extensive simulations show that our approach was never worse than any of the state-of-the-art competitors and usually significantly outperformed them. Our method is fast, simple, automatic and easy to code. The wavelet-Fisz principle extends to many dimensions and would also be able to handle several types of noise including binomial, negative binomial and Gamma.

Our technique is applied to the estimation of the intensity of earthquakes in northern California. We show that our technique gives visually similar results to the current state-of-the-art.

Keywords: Poisson process, variance stabilizing transform, transform to Gaussian, cycle spinning, denoising

---

*Piotr Fryźlewicz is Graduate Student, Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK (email: p.z.fryzlewicz@bristol.ac.uk).
†Guy P. Nason is Professor of Statistics, Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK (email: g.p.nason@bristol.ac.uk).
1 Introduction

Wavelet methods have now become a useful tool in the area of curve estimation, including in particular regression problems where noise is Gaussian, as well as density estimation (a general overview of wavelet methods in statistics can be found for example in Vidakovic (1999), see Daubechies (1992) for a mathematical introduction to wavelets). Some authors have also considered the problem of estimating the intensity of a Poisson process using a wavelet-based technique. The usual (regression) setting is as follows: the possibly inhomogeneous one-dimensional Poisson process is observed on the interval $[0,T]$, and discretized into a vector $v = (v_0, v_1, \ldots, v_{N-1})$, where $v_n$ is the number of events falling into the interval $[nT/N, (n + 1)T/N)$, and $N = 2^J$ is an integral power of two. Each $v_n$ can be thought of as coming from a Poisson distribution with an unknown parameter $\lambda_n$, which needs to be estimated. The approach proposed by Donoho (1993) consists in first preprocessing the data using Anscombe’s (1948) square-root transformation, so that the noise becomes approximately Gaussian. Then the analysis proceeds as if the noise was indeed Gaussian, yielding (after applying the inverse square-root transformation) an estimate of the intensity of the process.

The current state-of-the-art methods are based on Bayesian techniques described in Kolaczyk (1999a) and Timmermann and Nowak (1997, 1999). Kolaczyk (1999a) introduces a Bayesian multiscale algorithm to estimate the discretized intensity. However, rather than transforming the data using a wavelet transform, he considers recursive dyadic partitions, and places prior distributions at the nodes of the binary trees associated with these partitions. The Bayesian methods outperform earlier techniques in Kolaczyk (1997, 1999b), Nowak and Baraniuk (1999) and also the recent technique of Antoniadis and Sapatinas (2001) (since the latter is equivalent to Nowak and Baraniuk (1999) for Poisson data). The very recent article by Sardy et al. (2002) describes a computationally intensive $l_1$-penalized likelihood method that can be used for estimating Poisson intensities.

Other recent contributions to the field of wavelet-based intensity estimation include Patil and Wood (2000), who concentrate on the theoretical MISE properties of wavelet intensity estimators, where the intensity is a random process rather than a deterministic function (or, after discretization, a deterministic vector). Brillinger (1998) gives a brief overview of wavelet-based methodology in the analysis of point process data, and obtains the estimate of the autointensity function of the well-known California earthquake data.

In this article, we propose an alternative wavelet-based algorithm for estimating the deterministic discretized intensity function of an inhomogeneous one-dimensional Poisson process. Our method is based on the asymptotic normality of a certain function of the Haar wavelet and scaling coefficients of the vector $v$, the property first observed and proved by Fisz (1955), but (obviously) not set in the wavelet context at that time. In his paper, Fisz uses this property to test the hypothesis that two Poisson variables have equal means, and the hypothesis that their means are both equal to a given number.

The idea behind our algorithm is the following: we first preprocess the (Poisson) vector $v$ using a nonlinear wavelet-based transformation, which we call the wavelet-Fisz transformation, and then treat the preprocessed vector as if it were i.i.d. Gaussian of unit variance. In other words, we provide a new variance stabilizing transform, which operates in the wavelet domain, and not in the
time domain, like the standard square-root transformation.

The main advantages of our method are the following.

1. Its performance is extremely good, see below.

2. It is of computational order $N$ (or $NM$ if $M$ cycle-spins are used), in practice the software is itself very fast.

3. It is extremely simple and easy to code.

4. It is fully automatic (up to any parameters that the Gaussian denoiser requires).

5. It can make use of any signal+Gaussian noise denoising technology, an area where a vast amount of research effort has been and is being expended. Hence our method can only get better as we take advantage of newer Gaussian denoisers.

6. It can be easily extended to more than one dimension.

7. In principle it can handle many types of noise (Poisson, Gamma, binomial and negative binomial).

The papers by Kolaczyk (1999a), Timmermann and Nowak (1999) and the review paper Besbeas et al. (2002) all conclude that the Bayesian methods proposed by Kolaczyk (1999a) and Timmermann and Nowak (1999) are the best currently available. Sections 3.2 and 3.3 demonstrate that our algorithm was never worse than the above Bayesian methods and usually significantly outperformed them. Combined with the advantages above we have no hesitation in strongly recommending it.

Section 2 introduces the wavelet-Fisz transform and describes its theoretical and empirical properties. Section 3 introduces the wavelet-Fisz algorithm for estimating Poisson intensities and performs a thorough simulation study on its performance. Section 4 exhibits our algorithm and that of Kolaczyk (1999a) on a data set derived from the well-known Northern Californian Earthquake database studied by Brillinger (1998). Finally, Section 5 provides some conclusions and ideas for future exploration.

2 The wavelet-Fisz transformation

2.1 The Algorithm

Given the vector $\mathbf{v} = (v_0, v_1, \ldots, v_{N-1})$ of Poisson counts, we perform the wavelet-Fisz transformation as follows:

1. We apply one step of the Haar discrete wavelet transform to $\mathbf{v}$ (in other words, we operate on the finest scale only); however, we use the non-normalized filters $\{1/2, -1/2\}$ and $\{1/2, 1/2\}$, rather than the usual $\{1/\sqrt{2}, -1/\sqrt{2}\}$ and $\{1/\sqrt{2}, 1/\sqrt{2}\}$. We place the smooth in vector $s_n := (v_{2n} + v_{2n+1})/2$, and the detail in vector $d_n := (v_{2n} - v_{2n+1})/2$, for $n = 0, 1, \ldots, N/2 - 1$. 

3
2. We now set

\[
    f_n = \begin{cases} 
        0 & \text{if } s_n = 0, \\
        \frac{d_n}{\sqrt{s_n}} & \text{otherwise,}
    \end{cases}
\]

for \( n = 0, 1, \ldots, N/2 - 1 \). Given that each \( v_k \) is a realization of the variable \( V_k \sim \text{Pois}(\lambda_k) \), we have that

\[
    f_n \text{ is a realization of } Z_n := \begin{cases} 
        0 & \text{if } V_{2n} - V_{2n+1} = 0, \\
        \frac{V_{2n} - V_{2n+1}}{\sqrt{2V_{2n} + V_{2n+1}}} & \text{otherwise.}
    \end{cases}
\]

Since the sequence of variables \((V_0, V_1, \ldots, V_{N-1})\) is assumed to be independent, the theorem by Fisz (1955) implies that

\[
    Z_n \overset{d}{\to} N(0, 1/2) \quad \text{as} \quad (\lambda_{2n}, \lambda_{2n+1}) \to (\infty, \infty) \quad \text{and} \quad \lambda_{2n}/\lambda_{2n+1} \to 1.
\]

Therefore, we can expect \( f_n \) to come from a distribution which is close to \( N(0, 1/2) \) provided that \( \lambda_{2n} \) and \( \lambda_{2n+1} \) are “large and close enough”. Motivated by this observation, we replace the mother wavelet coefficients as follows:

\[
    d_n := f_n.
\]

The fluctuations in \( v \) which are mainly due to noise will carry over to approximately normally distributed (with mean 0 and variance 1/2) fluctuations in the modified detail vector \( d \). On the other hand, “significant” (i.e. caused by changes in the underlying intensity) fluctuations in \( v \) will result in the detail coefficients being substantially deviated from normal.

3. Following the above procedure, we modify the detail coefficients at all scales, leaving all the scaling coefficients intact. The detail coefficients (corresponding to Poisson noise) at all scales are now approximately Gaussian.

4. We now treat the modified detail coefficients as if they were in fact unmodified detail coefficients resulting from the plain Haar decomposition of some vector \( u \). We reconstruct the vector \( u \) using the usual inverse Haar transform, bearing in mind that non-normalized filters have to be used again (the filters used to decompose vector \( v \) were \( \{1/2, -1/2\} \) and \( \{1/2, 1/2\} \) instead of \( \{1/\sqrt{2}, -1/\sqrt{2}\} \) and \( \{1/\sqrt{2}, 1/\sqrt{2}\} \)). There is a one-to-one correspondence between vectors \( v \) and \( u \), and we will denote

\[
    u = \mathcal{F}v.
\]

The nonlinear operator \( \mathcal{F} \) defines the \textit{wavelet-Fisz transformation}.

Note that our wavelet-Fisz transform is a mapping \( \mathbb{R}^N \to \mathbb{R}^N \) whereas the Fisz (1955) transform is \( \mathbb{R}^2 \to \mathbb{R} \). As was argued in this section, if \( v \) is the vector of Poisson counts, then the Haar detail coefficients of \( u = \mathcal{F}v \) are approximately Gaussian. In the next section, we will argue that \( u \) itself is also approximately normal with variance one, so that it can be viewed as a “Gaussianized” version of the original vector \( v \). The computational complexity of the wavelet-Fisz transformation is the same as that of the discrete Haar transform, which is \( O(N) \).
2.2 An Example — 8 Data Points

Let \(v = (v_0, v_1, \ldots, v_7)\), and \(v_i > 0\) for all \(i\). The wavelet-Fisz transform \(u = \mathcal{F}v\) is given by

\[
\begin{align*}
  u_0 &= \frac{\sum_{i=0}^{7} v_i}{8} + \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} + \frac{v_0 + v_1 - (v_2 + v_3)}{2 \sqrt{3}} + \frac{v_0 - v_1}{\sqrt{2} v_0 + v_1}, \\
  u_1 &= \frac{\sum_{i=0}^{7} v_i}{8} + \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} - \frac{v_0 + v_1 - (v_2 + v_3)}{2 \sqrt{3}} - \frac{v_0 - v_1}{\sqrt{2} v_0 + v_1}, \\
  u_2 &= \frac{\sum_{i=0}^{7} v_i}{8} + \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} - \frac{v_0 + v_1 - (v_2 + v_3)}{2 \sqrt{3}} + \frac{v_2 - v_3}{\sqrt{2} v_2 + v_3}, \\
  u_3 &= \frac{\sum_{i=0}^{7} v_i}{8} - \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} + \frac{v_0 + v_1 - (v_2 + v_3)}{2 \sqrt{3}} - \frac{v_2 - v_3}{\sqrt{2} v_2 + v_3}, \\
  u_4 &= \frac{\sum_{i=0}^{7} v_i}{8} - \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} + \frac{v_4 + v_5 - (v_6 + v_7)}{2 \sqrt{3}} + \frac{v_4 - v_5}{\sqrt{2} v_4 + v_5}, \\
  u_5 &= \frac{\sum_{i=0}^{7} v_i}{8} - \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} + \frac{v_4 + v_5 - (v_6 + v_7)}{2 \sqrt{3}} - \frac{v_4 - v_5}{\sqrt{2} v_4 + v_5}, \\
  u_6 &= \frac{\sum_{i=0}^{7} v_i}{8} - \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} - \frac{v_4 + v_5 - (v_6 + v_7)}{2 \sqrt{3}} + \frac{v_6 - v_7}{\sqrt{2} v_6 + v_7}, \\
  u_7 &= \frac{\sum_{i=0}^{7} v_i}{8} - \frac{\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i}{2 \sqrt{2}} - \frac{v_4 + v_5 - (v_6 + v_7)}{2 \sqrt{3}} - \frac{v_6 - v_7}{\sqrt{2} v_6 + v_7}.
\end{align*}
\]

The structure of the underlying Haar inverse DWT can clearly be seen in the formula. From the above example, it is immediate that

1. The means of a vector and its wavelet-Fisz transform are equal: \(\sum_{i=0}^{n} v_i = \sum_{i=0}^{n} u_i\).
2. If \(c\) is a vector whose components are all equal, then \(\mathcal{F}c = c\).

2.3 The Formula

We will now introduce an explicit general formula for the operator \(\mathcal{F}\). Let \(v = (v_0, v_1, \ldots, v_{N-1})\) be the vector of Poisson counts, and let \(u = (u_0, u_1, \ldots, u_{N-1})\) be the wavelet-Fisz transform of \(v\): \(u = \mathcal{F}v\). Bearing in mind that \(N\) is an integral power of two, we denote \(J = \log_2(N)\). We introduce the family of Haar wavelet filters \(\{\psi^{j,k}\}\), where \(j = 0, 1, \ldots, J - 1\) is the scale parameter, and \(k = l2^{J-j}, l = 0, 1, \ldots, j\), is the location parameter. The components of \(\psi^{j,k}\) will be denoted...
by $\psi_n^{j, k}$, for $n = 0, 1, \ldots, N - 1$. We define

$$
\psi_n^{j, k} = \begin{cases} 
0 & \text{for } n < k \\
1 & \text{for } k \leq n < k + 2^{j-1} \\
-1 & \text{for } k + 2^{j-1} - 1 \leq n < k + 2^j \\
0 & \text{for } k + 2^j \leq n.
\end{cases}
$$

(6)

Similarly, we introduce the family of Haar scaling filters $\{\phi_n^{j, k}\}$, whose components will be denoted by $\phi_n^{j, k}$ (the range of $j, k, n$ remains unchanged). We define

$$
\phi_n^{j, k} = \begin{cases} 
0 & \text{for } n < k \\
1 & \text{for } k \leq n < k + 2^j \\
0 & \text{for } k + 2^j \leq n.
\end{cases}
$$

(7)

Our definition of discrete Haar wavelets is similar to that of Nason et al. (2000), Section 2. The difference is that we “pad” the wavelet filter vectors with zeros on both sides so that they have all length $N$, and we do not normalize them.

Further, let $\langle \cdot, \cdot \rangle$ denote the inner product of two vectors, and let $b^j(n) = (b_0^j(n), b_1^j(n), \ldots, b_{2^j-1}^j(n))$ be the binary representation of an integer $n$, where $n < 2^j$.

The formula for the $n$th element of $u = \mathcal{F}v$ is

$$
\begin{align*}
    u_n &= \frac{\langle \phi_0^{0,0}, v \rangle}{N} + \sum_{j=0}^{J-1} (-1)^{b_j^j(n)} 2^j \frac{2^j}{N} c_{j, n}(v),
\end{align*}
$$

(8)

where

$$
c_{j, n} = \begin{cases} 
\frac{\langle \phi_{j, n/2^{j-1}}^{j-1}, v \rangle}{\langle \phi_{j, n/2^{j-1}}^{j-1}, v \rangle} & \text{if } \langle \phi_{j, n/2^{j-1}}^{j-1}, v \rangle > 0 \\
0 & \text{otherwise}.
\end{cases}
$$

(9)

We now quote two important properties of the operator $\mathcal{F}$, the proofs are in the appendix. Proposition 2.1 says that the coefficients of the wavelet-Fisz transformed vector of Poisson counts (where the underlying intensity is constant $\lambda_n = \lambda$ for all $n = 1, \ldots, N$) are asymptotically uncorrelated, and Proposition 2.2 says that these coefficients are also asymptotically normal with variance one. These two properties will be useful later, as they enable us to treat the preprocessed vector as Gaussian with uncorrelated noise, so that the well-established wavelet denoising techniques can be applied.

**Proposition 2.1** Let $V = (V_0, V_1, \ldots, V_{N-1})$ be a vector of i.i.d. Poisson variables with mean $\lambda$, and let $N$ be an integral power of two. Let $U = \mathcal{F}V$. For $m \neq n$, we have

$$
\text{cor}(U_m, U_n) \to 0 \quad \text{as} \quad \lambda \to \infty \quad \text{and} \quad \lambda/N \to 0.
$$

(10)
Proposition 2.2 Let $\mathbf{V} = (V_0, V_1, \ldots, V_{N-1})$ be a vector of i.i.d. Poisson variables with mean $\lambda$, and let $N$ be an integral power of two. Let $\mathbf{U} = \mathcal{F}\mathbf{V}$. For all $n = 0, 1, \ldots, N - 1$, we have
\begin{equation}
U_n - \lambda = M + X_n,
\end{equation}
where
\begin{align*}
M & \xrightarrow{d} 0 \quad \text{as} \quad \lambda/N \rightarrow 0 \\
X_n & \xrightarrow{d} N(0, 1) \quad \text{as} \quad (\lambda, N) \rightarrow (\infty, \infty).
\end{align*}

Simulations suggest that the wavelet-Fisz transformation works well (i.e., transforms vectors of independent Poisson variables to vectors of nearly uncorrelated, approximately normal variables with variance one) in the case of non-constant intensities, too, provided that:

- the sample size $N$ and the minimum of the intensity vector $\min_n \lambda_n$ are “large”;
- the mean of the intensity vector $\bar{\Lambda}$ is small compared to the sample size $N$.

As an example, let us consider the intensity as in the top plot of Figure 1 (a rescaled version of Donoho and Johnstone (1994) bumps function). This intensity vector will be denoted by $\Lambda$, and $\mathbf{v}$ will denote a sample path generated from it. We define the Anscombe (1948) transformation by
\begin{equation}
\mathcal{A}\mathbf{v} = 2\sqrt{\mathbf{v} + 3/8}.
\end{equation}

The Anscombe transformation was proposed by Donoho (1993) as a variance stabilizing transform appropriate for converting Poisson signals to “nearly-Gaussian” signals. Figure 1 compares the Q-Q plots of $\mathbf{v} - \Lambda$, $\mathcal{A}\mathbf{v} - \mathcal{A}\Lambda$ and $\mathcal{F}\mathbf{v} - \mathcal{F}\Lambda$. Clearly, the wavelet-Fisz transformation does a better job in “Gaussianizing” $\mathbf{v}$: the vector $\mathcal{F}\mathbf{v} - \mathcal{F}\Lambda$ is “the most nearly normal” vector among the above. In particular, the wavelet-Fisz transformed data is less “stepped” and looks more like variates from a continuous distribution than a discrete one. The Anscombe-transformed data appears more “stepped” for lower quantiles than for higher ones. Further, the tails for wavelet-Fisz are more normal than for Anscombe which in turn is more normal than the count data. Figure 2 shows the autocorrelation function of $\mathbf{v} - \Lambda$, $\mathcal{A}\mathbf{v} - \mathcal{A}\Lambda$, and $\mathcal{F}\mathbf{v} - \mathcal{F}\Lambda$. Note that the structure of the acf of all the three vectors is similar, which may suggest that the wavelet-Fisz transformation does not introduce a significantly large amount of extra correlation, even if the intensity function is non-constant. The lack of correlation enables us to use any of the popular methods for wavelet denoising of signals contaminated by independent Gaussian noise.
Figure 1: From top to bottom: intensity vector $\Lambda$ of Donoho and Johnstone (1994) \textbf{bumps} function and the sample path $\mathbf{v}$, the Q-Q plots of vectors $\mathbf{v} - \Lambda$, $\mathbf{A} \mathbf{v} - \mathcal{A} \Lambda$, and $\mathcal{F} \mathbf{v} - \mathcal{F} \Lambda$. See text for further discussion.
Figure 2: From top to bottom: the acf of $v - \Lambda$, $Av - A\Lambda$, and $Fv - F\Lambda$. 
3 Poisson intensity estimation

We propose the following core algorithm for estimating the intensity $\Lambda$ of a Poisson process:

[A1] Given the vector $\mathbf{v}$ of Poisson observations, preprocess it using the wavelet-Fisz transformation to obtain $\mathcal{F}\mathbf{v}$.

[A2] Denoise $\mathcal{F}\mathbf{v}$ using any suitable ordinary wavelet denoising technique, appropriate for Gaussian noise (i.e. DWT — thresholding — inverse DWT). Denote the smoothed version of $\mathcal{F}\mathbf{v}$ by $\widehat{\mathcal{F}\mathbf{A}}$. We can optionally exploit the fact that the asymptotic variance of the noise is equal to one.

[A3] Perform the inverse wavelet-Fisz transform to obtain $\mathcal{F}^{-1}(\widehat{\mathcal{F}\mathbf{A}})$ and take it to be the estimate of the intensity.

The next sections evaluate the performance of our Poisson intensity estimation algorithm and compare it to existing techniques.

3.1 Methods for Poisson intensity estimation

Existing methods. As mentioned in the introduction the Bayesian methods due to Kolaczyk (1999a) and Timmermann and Nowak (1997, 1999) are currently state-of-the-art. For more details of this sort of comparison, see Besbeas et al. (2002).

Hence our simulation study directly compared our technique with the Bayesian methods only. To compare our technique with Kolaczyk (1999a) we used Eric Kolaczyk’s BMSMShrink MATLAB software. As we did not have access to Timmermann and Nowak’s software we exactly reproduced the simulation setup as in Timmermann and Nowak (1999) and compared our results to their Tables I and II. (Incidentally, the methods in Kolaczyk (1999a) and Timmermann and Nowak (1999) are very similar: the underlying Bayesian model is exactly the same, although the hyperparameter estimation is slightly different (Kolaczyk (2001)).

Our method. The following describes the common features for our Poisson intensity estimation.

1. All our techniques always involve the wavelet-Fisz transform, [A1], of the data, and the inverse wavelet-Fisz transform, [A3].

2. In step [A2] of our algorithm the wavelet denoising technique may be of a translation invariant (TI) transform type, see Donoho and Coifman (1995). We refer to TI-denoising at this stage as “internal” cycle spinning (CS).

3. In step [A2] we could use any one of a number of wavelet families (e.g. multiwavelet, see Downie and Silverman (1998), complex-valued, see Lina (1997) etc.) for the denoising. In our simulations below we use wavelets of different degrees of smoothness from the Daubechies’ family, see Daubechies (1992).
4. Let $\mathcal{S}$ be the shift-by-one operator from Nason and Silverman (1995). The wavelet-Fisz transform is not translation-equivariant since $\mathcal{F}\mathcal{S} \neq \mathcal{S}\mathcal{F}$. This non-commutativity implies that it is beneficial to apply CS to the whole algorithm $[\text{A1}]-[\text{A3}]$ even if $[\text{A2}]$ uses a TI technique. We call this “external” CS.

Due to the particular type of nonlinearity of the wavelet-Fisz transform there is no fast $O(N \log N)$ algorithm for the external CS. Therefore, we implement external CS by actually shifting the data before $[\text{A1}]$, shifting back the estimate after $[\text{A3}]$, and averaging over the estimates obtained through several different shifts.

For a data set of length $N$ there are $N$ possible shifts. However, we have found that good results can be obtained using only approximately $N/10$ shifts (or even $N/100$ for $N$ suitably large).

Note that there is no point in doing external CS with the Anscombe transformation, $\mathcal{A}$, provided one has carried out internal CS, since Anscombe’s transformation commutes with the shift operator: $\mathcal{A}\mathcal{S} = \mathcal{S}\mathcal{A}$.

The following list labels and describes the wavelet denoising methods that we choose to use in $[\text{A2}]$. In each case $\mathcal{F}^{\text{label}}$ denotes the use of the wavelet-Fisz transform and its inverse.

- **$\mathcal{F}^{\text{hard}}$:** Universal hard thresholding from Donoho and Johnstone (1994) as implemented in WaveThresh (Nason (1998)) with default parameters (e.g. uses MAD variance estimation on all coefficients). Only uses external CS (100 shifts).

- **$\mathcal{F}^{\text{CV}}$:** Cross-validation method from Nason (1996) as implemented in WaveThresh using default parameters but hard thresholding. Only uses external CS (100 shifts).

- **$\mathcal{F}^{\text{BT}}$:** A variant of the greedy tree algorithm from Baraniuk (1999). Only uses external CS (10 shifts).

- **$\mathcal{F}^{\text{eBayes}}$:** The eBayes procedure as described by Johnstone and Silverman (2001) using their software kindly supplied by Bernard Silverman. Here we set the variance parameter of the normalized coefficients to be equal to 1, see (12). Full internal CS ($O(N \log N)$), part of the eBayes software) and 10 shifts for external CS.

- **$\mathcal{F}^{\text{TIU}}$:** Universal hard threshold corresponding with variance=1, see again (12). Full internal CS and 10 shifts for external CS.

**Hybrids.** We also looked at the performance of certain hybrid methods. These estimate the intensity by averaging the results of two of the above wavelet-Fisz methods.

- **$\mathcal{H}_1$:** average of $\mathcal{F}^{\text{CV}}$ and $\mathcal{F}^{\text{TIU}}$.
- **$\mathcal{H}_2$:** average of $\mathcal{F}^{\text{CV}}$ and $\mathcal{F}^{\text{BT}}$.
- **$\mathcal{H}_3$:** average of $\mathcal{F}^{\text{CV}}$ and $\mathcal{F}^{\text{eBayes}}$. 

11
Figure 3: Left: Scaled and shifted Donoho and Johnstone (1994) blocks function, and its clipped version: clipped blocks. Right: The true intensity function (dashed) and an estimate computed using our algorithm using hybrid method H3. This was the best reconstruction (in terms of MISE) out of 50.

3.2 Empirical performance — clipped blocks.

For our first set of simulations the “true” intensity is the clipped blocks function shown on the left hand side of Figure 3. The clipped blocks signal is of length $N = 1024$ and obtained from the blocks function of Donoho and Johnstone (1994) by setting all negative values to zero, scaling it so that the maximum intensity is 15.6 and then adding 3. We also examine the same signal but scaled by factors of $1/6$, $1/3$ and $10/3$. These scalings give us a range of low and high intensity settings with large spreads of low intensity. The minimum and maximum intensities are, for each of these scalings: $3$–$18.6$, $0.5$–$3.1$, $1.0$–$6.2$ and $10$–$62$.

The simulation results reported in Table 1 are the MISE per bin: that is we compute the sum of the squared errors between our estimate and the true intensity, then divide by the number of bins (1024) and then take the mean over all simulations.

It is clear from Table 1 that the hybrid methods do extremely well especially at medium and high intensity levels. Hybrids can easily be formulated due to the large number of methods available for denoising Gaussian contaminated signals. The right hand figure in Figure 3 shows a particular sample reconstruction using the hybrid method H3.

Out of all the non-hybrid methods BMSMSShrink does better at scale 1, is comparable at scale $1/3$ but beaten at the other two scales. The method $F \rightarrow CV$ does well at low and high intensities: maybe this reflects the asymptotics of $\lambda / N \rightarrow 0$ and $\lambda \rightarrow \infty$ respectively (Propositions 2.1 and 2.2)?
Table 1: MISE per bin (×100 and rounded) for clipped block intensity estimation using several procedures as denoted in the text for a variety of intensity scalings. The best result(s) in each column is (are) indicated by a box. The results are based on 25 simulations except for BMSMShrink which is based on 40.

<table>
<thead>
<tr>
<th>Method</th>
<th>Scaling</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/6</td>
<td>1/3</td>
<td>1</td>
<td>10/3</td>
</tr>
<tr>
<td><strong>BMSMShrink</strong></td>
<td>9</td>
<td>19</td>
<td>61</td>
<td>192</td>
</tr>
<tr>
<td>(F_{\text{CV}})</td>
<td>8</td>
<td>21</td>
<td>66</td>
<td>190</td>
</tr>
<tr>
<td>(F_{\text{BT}})</td>
<td>11</td>
<td>25</td>
<td>74</td>
<td>189</td>
</tr>
<tr>
<td>(F_{\text{Bayes}})</td>
<td>8</td>
<td>19</td>
<td>67</td>
<td>214</td>
</tr>
<tr>
<td>(H_1)</td>
<td>9</td>
<td>18</td>
<td>62</td>
<td>178</td>
</tr>
<tr>
<td>(H_2)</td>
<td>9</td>
<td>19</td>
<td>53</td>
<td>165</td>
</tr>
<tr>
<td>(H_3)</td>
<td>8</td>
<td>18</td>
<td>59</td>
<td>183</td>
</tr>
</tbody>
</table>

Table 2: Normalized MISE values (×10000) for existing Bayesian techniques and our \(F_{\text{Bayes}}\) method using Daubechies’ least asymmetric wavelets with 10 vanishing moments. The best results are indicated by a box.

<table>
<thead>
<tr>
<th>Intensity</th>
<th>Peak intensity=8</th>
<th>Peak intensity=128</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F_{\text{Bayes}})</td>
<td>(BMSMShrink)</td>
<td>(F_{\text{Bayes}})</td>
</tr>
<tr>
<td>Doppler Blocks</td>
<td>154</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>178</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>1475</td>
<td>1792</td>
</tr>
</tbody>
</table>

3.3 Empirical performance — standard test functions

The simulation setup in this section is the same as that described in Timmermann and Nowak (1999) and the results here can be directly compared. The results cited here as \(BAYES\) are taken from and refer to the Bayesian method developed in Timmermann and Nowak (1999).

The study in Timmermann and Nowak (1999) obtains two sets of intensity functions of length \(N = 1024\) from the test functions from Donoho and Johnstone (1994). Each set is obtained by shifting and scaling to achieve \((\text{min}, \text{max})\) intensities of \((1/8, 8)\) and \((1/128, 128)\). The true intensity functions for the \((1/8, 8)\) case are shown as dashed lines in Figure 4. All the results in this section are based on 25 independent simulations.

The results reported in Table 2 are the MISE normalized by the squared \(l_2\) norm of the true intensity vector, multiplied by 10000 and then rounded for clarity of presentation (this is exactly the same performance measure as in Timmermann and Nowak (1999) which is useful for comparability). The results show that our \(F_{\text{Bayes}}\) method dramatically outperforms the existing state-of-the-art methods especially for the lower intensity, except for the blocks function. We should also emphasize that
Figure 4: Selected estimates for various true intensity functions (dashed, described in text). All estimates were the best out of 50 simulations. The estimation method in each case was $F \triangleright U$ with Daubechies least-asymmetric wavelets with 10 vanishing moments except for blocks which used $H3$ with Haar wavelets.
Table 3: Normalized MISE values ($\times 10000$) for existing Bayesian techniques and three wavelet-Fisz estimation methods using Haar wavelets. The best results are indicated by a box.

<table>
<thead>
<tr>
<th>Peak Intensity</th>
<th>BAYES</th>
<th>BMSMShrink</th>
<th>$\mathbf{F} \bowtie \mathbf{U}$</th>
<th>$\mathbf{H2}$</th>
<th>$\mathbf{H3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>178</td>
<td>126</td>
<td>175</td>
<td>120</td>
<td>116</td>
</tr>
<tr>
<td>128</td>
<td>27</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

our $\mathbf{F} \bowtie \mathbf{U}$ method is far simpler and quicker (as a very rough guide $\text{BMSMShrink}$ took about 40 minutes for 25 simulations in MATLAB; our $\mathbf{F} \bowtie \mathbf{U}$ took 8 minutes in SPlus. However, the wavelet transforms in $\mathbf{F} \bowtie \mathbf{U}$ are implemented in C, but not the wavelet-Fisz transform nor the external cycle-spinning.)

In Figure 4 the small spike in the heavisine function is not picked up well at intensity 8 but is almost always clearly estimated at intensity 128 (not shown). However, it should be said that the spike is almost completely obscured by noise in all realizations at intensity 8 so it would be extremely difficult for any method to detect it. We are impressed with the quality of the estimates using the new wavelet-Fisz method, particularly with bumps and doppler. Also, the reconstruction of blocks is very accurate. Overall, it must be remembered that the reconstructions are usually going be less impressive than the equivalent problem where the test functions are contaminated with Gaussian noise.

The main reason why our $\mathbf{F} \bowtie \mathbf{U}$ method in Table 2 does not work well for blocks is that a smooth wavelet is used in the Gaussian denoising step [A2]. We performed additional simulations using Haar wavelets and the results are summarised in Table 3. It can be seen that the hybrid methods do extremely well, even $\mathbf{F} \bowtie \mathbf{U}$ does well for high intensities. The BAYES method of Timmermann and Nowak (1999) is not competitive in this context. We should also mention that the wavelet-Fisz methods not based on simple universal thresholding take a bit longer than $\mathbf{F} \bowtie \mathbf{U}$; indeed their speed of execution is roughly equivalent to $\text{BMSMShrink}$.

We also performed an indirect comparison with the computationally intensive $l_1$-penalized likelihood method proposed by Sardy et al. (2002) using results presented therein. In simulated comparisons on blocks and bumps, at best, they quote improvements in median observed mean-squared errors by a factor of 2.4 (comparing their method with Donoho-Anscombe). Our simple transform offers further improvements: the corresponding factors in our case range from 3.8 to 11.3.

**Summary.** For all test signals, except blocks, our $\mathbf{F} \bowtie \mathbf{U}$ method gives better performance (1–1.9 improvement in MISE) and is much quicker. For blocks the hybrid methods give better performance (1.1 improvement) but are comparable in speed to $\text{BMSMShrink}$.

Our recommendation is that if one suspects the intensity is piecewise constant then use Haar wavelets and a hybrid method such as H2 or H3 otherwise we strongly recommend the use of $\mathbf{F} \bowtie \mathbf{U}$ with a smooth wavelet.
Figure 5: The number of earthquakes of magnitude $\geq 3.0$ which occurred in Northern California in 1024 consecutive weeks, the last week being 29 Nov – 5 Dec 2000.

4 Application to earthquake data

In this section, we analyse Northern Californian earthquake data, available from: http://quake.geo.berkeley.edu. We analyse the time series $N_k$, $k = 1, \ldots, 1024$, where $N_k$ is the number of earthquakes of magnitude 3.0 or more which occurred in the $k$th week, the last week under consideration being 29 November – 5 December 2000. The time series, created in S-Plus, is plotted in Figure 5.

Our aim is to extract the intensity which underlies the realization of this process. For the purposes of this example we shall use the $BMSMShrink$ methodology of Kolaczyk (1999a) and our hybrid $H_2$ method with Haar wavelets. The rationale for using $H_2$ is that:

- it appears that the true earthquake intensity is highly non-regular and $H_2$ with Haar wavelets worked the best on the blocks and clipped blocks simulation examples;

- the earthquake data exhibits medium to high intensities and $H_2$ was better than the other hybrids for this situation;

see Tables 1,3.

Figures 6 and 7 show the estimates obtained using $BMSMShrink$ and $H_2$. The estimates are very similar in Figure 6 when plotted on the 0-250 vertical scale to include the peak at 274 weeks. The real differences in the estimates can be seen in Figure 7 which encompasses a smaller vertical range.
Figure 6: Intensity estimates for earthquake data for weeks 200 to 299. Dotted line is $BMSMShrink$ estimate and solid is $H2$ estimate.

as there are no large peaks during the period 300-399 weeks. However, the estimates again are visually not too different: the $H2$ estimate is a little less variable (although we tried $H3$ which was slightly more variable). Although with this real data there is clearly no right or wrong answer it is reassuring that they do give such similar visual results even though $BMSMShrink$, $H2$ and $H3$ are based on completely different philosophies.

5 Conclusions

In this paper, we have described a new wavelet-based technique for bringing vectors of Poisson counts to normality with variance one. The technique, named the wavelet-Fisz transformation, was applied to estimating the intensity of an inhomogeneous Poisson process, yielding a method whose performance was better than that of the current state-of-the-art in terms of both accuracy and speed.

For Poisson intensity estimation our methodology requires two components. The first, the wavelet-Fisz transform, is very simple and easy to code. The second component can be any suitable Gaussian denoising procedure: we have used and compared a variety of wavelet methods ranging from the fast universal thresholding to more complicated techniques such as cross-validation, Baraniuk trees and empirical Bayes. Since any Gaussian denoiser can be used the wavelet-Fisz algorithm can only improve as the field develops. We suspect the majority of multiscale denoising techniques
are developed for Gaussian noise and the wavelet-Fisz transform can exploit these for estimating Poisson intensities.

We believe that one of the reasons why the performance of the wavelet-Fisz algorithm is so good is due to the non-commutativity of the Fisz and shift operators: hence enabling meaningful cycle spinning. Also, the Fisz transform itself is a more effective normalizer than Anscombe.

Our recommendation is that if one suspects the intensity is piecewise constant then one should use Haar wavelets and a hybrid method such as H2 or H3 otherwise we strongly recommend the use of F▷U with a smooth wavelet.

Future ideas. An important feature of our algorithm is that it can be easily extended to more than one dimension: in particular, we are keen to exploit this idea for images.

Also, the wavelet-Fisz transform has the potential to “Gaussianize” a wide range of other noise distributions such as binomial, Gamma and negative binomial. As $\chi^2$ is a special case of the Gamma distribution we believe that the wavelet-Fisz algorithm could be a potentially useful variance-stabilizer for a putative (wavelet) periodogram smoothing technique for (locally stationary) time series analysis. See Priestley (1981) for details on periodogram smoothing for stationary time series and Nason et al. (2000) for smoothing of wavelet periodograms of locally stationary wavelet processes.

Clearly, it would be interesting to explore the use of the wavelet-Fisz transform in conjunction with new and exciting denoisers, not necessarily wavelet based.
Software. The S-Plus routines written and used by us can be downloaded from the associated web page: http://www.stats.bris.ac.uk/~mapzf/Poisson/Poisson.html

Acknowledgements

We would like to thank Eric Kolaczyk for kindly supplying his BMSMShrink software. Thanks also to Anestis Antoniadis, Eric, Rob Nowak, Theofanis Sapatinas and Kamila Żychaluk for helpful discussions. Piotr Fryźlewicz was supported by a University of Bristol Research Scholarship, an ORS award and Unilever Research. Guy Nason gratefully acknowledges support from EPSRC grants GR/M10229/01 and GR/A01664/01.

Appendix: Proofs

Proof of 2.1

Proof. We will first calculate the correlation between the modified detail coefficients at two different scales. The detail coefficient at any given scale has the form

$$D_f = (X_0 - X_1)f(X_0 + X_1)$$

where $X_0$ and $X_1$ are some independent, identically distributed Poisson variables, and $f(x) = x^{-1/2}$ with $f(0) = 0$. The detail coefficient at any coarser scale depends on $X_0, X_1$ through their sum only, i.e. we have

$$D_c = g(X_0 + X_1),$$

where $g$ also depends on some other Poisson variables $X_i, i \neq 0, 1$. Since $X_0, X_1$ are identically distributed, we obviously have

$$\mathbb{E}(D_f) = 0 \text{ and } \mathbb{E}(D_f D_c) = 0,$$

and so $\text{cov}(D_f, D_c) = 0$. We can show in a similar way that the smooth coefficient $\langle \phi^{0.0}, \nu \rangle / N$ is uncorrelated with any of the detail coefficients.

We are now in a position to calculate $\text{cov}(U_m, U_n)$. From formula (8) it is clear that the variables will share the “smooth” term $\langle \phi^{0.0}, \nu \rangle / N$, which we will denote by $M$ to simplify the notation, Since the integer $[n / 2^{j-j}] 2^{j-j}$ (see formula (8)) depends only on the first $j$ bits in the binary expansion of $n$, the variables $U_m$ and $U_n$ will also share the term

$$X := \sum_{j=0}^{J-1} (-1)^{b_j(n)} 2^{i_j} c_{j,J_n}(\nu),$$

(15)
where \( J^* = \text{min}\{j : b_j^T(n) \neq b_j^T(m)\} \). Using the definition in formula (9), it can be proved that
\[
(-1)^{b_j^T(m)} 2^{\frac{j-J}{2}} c_{j,J,m}(\mathbf{V}) = -(-1)^{b_j^T(n)} 2^{\frac{j-J}{2}} c_{j,J,n}(\mathbf{V}).
\]
(16)

The term on the lhs of equation (16) will be denoted by \( Y \). We also denote
\[
Z_1 = \sum_{j=J^*+1}^{J-1} (-1)^{b_j^T(n)} 2^{\frac{j-J}{2}} c_{j,J,m}(\mathbf{V})
\]
\[
Z_2 = \sum_{j=J^*+1}^{J-1} (-1)^{b_j^T(n)} 2^{\frac{j-J}{2}} c_{j,J,n}(\mathbf{V}).
\]

It takes a closer look at formula (8) to see that \( Z_1 \) and \( Z_2 \) are independent (they are functions of different components of vector \( \mathbf{V} \)). Using the formulae in (14), we now write
\[
\text{cov}(U_m, U_n) = \text{cov}(M + X - Y + Z_1, M + X + Y + Z_2) = \text{Var}(M) + \text{Var}(X) - \text{Var}(Y).
\]

For \( \lambda \) large enough, as \( X \) and \( Y \) become approximately normal (see Fisz (1955)), we have
\[
\text{Var}(X) \leq \sum_{j=0}^{J^*-1} (1 + \epsilon)2^{j-J} = (1 + \epsilon)(2^{J^*-J} - 2^{-J})
\]
\[
\text{Var}(Y) \geq (1 - \epsilon)2^{J^*-J}
\]

Moreover, we have \( \text{Var}(M) = \lambda/N \to 0 \) by assumption. Since \( N = 2^J \to \infty \), we have
\[
\text{cov}(U_m, U_n) \leq \lambda/N + (1 + \epsilon)(2^{J^*-J} - 2^{-J}) - (1 - \epsilon)2^{J^*-J} \to 0
\]
as \( \epsilon \to 0 \) (note that \( 2^{J^*-J} \) is constant), which completes the proof.

**Proof of 2.2**

**Proof.** Without loss of generality, let us concentrate on \( U_0 \). Let \( J = \log_2(N) \), and let us denote
\[
W_j(\lambda) = \begin{cases}
\frac{\sum_{i=0}^{2^j} V_i - \sum_{i=0}^{2^j-1} V_i}{\sqrt{\sum_{i=0}^{2^j-1} V_i}} & \text{if } \sum_{i=0}^{2^j-1} V_i > 0 \\
0 & \text{otherwise}
\end{cases}
\]
(17)
to emphasize the dependence of \( W_j \) on \( \lambda \). The following equality holds (see the example in section 2.2 and formulas (8) and (9))
\[
U_0 = N^{-1} \sum_{i=0}^{N-1} V_i + \sum_{j=0}^{J-1} 2^{\frac{j-J}{2}} W_j(\lambda).
\]
(18)
Set \[ M = N^{-1} \sum_{i=0}^{N-1} V_i - \lambda \text{ and } X_0 = \sum_{j=0}^{J-1} 2^{-j-1} W_j(\lambda). \]

We will first show that \( X_0 \overset{d}{\to} N(0, 1) \) as \((\lambda, J) \to (\infty, \infty)\). Let us fix \( \epsilon_1 > 0 \). By Fisz theorem, if \( \lambda \) or \( j \) are large enough, then we have
\[
\text{Var}(W_j(\lambda)) = (1 + \epsilon_j^2) \leq 1 + \epsilon_1,
\]
where \( |\epsilon_j^2| < \epsilon_1 \). Also, for all \( \lambda \), the variables \( W_j(\lambda) \) are uncorrelated (see the proof of Proposition 2.1).

Using the symmetry of \( W_j(\lambda) \), the Chebyshev inequality, the orthogonality of \( W_j(\lambda) \), and the formula (19), for large \( \lambda \) and \( J \) we have
\[
P \left( \sum_{j=J}^{M-1} 2^{\frac{j-1}{2}} W_j(\lambda) < -\epsilon \right) = P \left( \sum_{j=J}^{M-1} 2^{\frac{j-1}{2}} W_j(\lambda) > \epsilon \right)
\leq \epsilon^{-2} \text{Var} \left( \sum_{j=J}^{M-1} 2^{\frac{j-1}{2}} W_j(\lambda) \right)
\leq \epsilon^{-2} \sum_{j=J}^{\infty} 2^{\frac{j-1}{2}} \text{Var} (W_j(\lambda))
\leq \epsilon^{-2} (1 + \epsilon_1) 2^{-J}.
\]

Clearly, we have that
\[
\forall \epsilon \exists J_0 \quad \forall J \geq J_0 \quad \epsilon^{-2} (1 + \epsilon_1) 2^{-J} \leq \epsilon.
\]

Observe now that
\[
\forall J \quad \sum_{j=0}^{J-1} 2^{\frac{j-1}{2}} W_j(\lambda) \overset{d}{\to} N(0, 1 - 2^{-J}) \quad \text{as} \quad \lambda \to \infty.
\]

Here we have a finite linear combination of orthogonal variables, each of which converges in distribution to \( N(0, 1) \) by Fisz theorem. The finite linear combination will therefore converge to the finite linear combination of orthogonal (= independent) normal variables, whose variances sum up to \( 1 - 2^{-J} \). Denote by \( S_{\lambda^2}(t) \) the survival function of a normal variable with mean zero and variance \( \sigma^2 \). Note two properties of the family \( \{S_{1-2^{-j}}(\lambda)\}_{j=1}^\infty \): \( |S_{1-2^{-j}}(\cdot) - S_1(\cdot)|_\infty \to 0 \) as \( J \to \infty \); \( \{S_{1-2^{-j}}(t)\}_{j=1}^\infty \) is uniformly Lipschitz continuous with Lipschitz constant \( L = 1/\sqrt{\pi} \).

Now fix \( \epsilon > 0 \) and choose the corresponding \( J_0 \) in (20). For an arbitrary fixed \( t \), examine the
difference
\[ D_1 = \left| P \left( \sum_{j=0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > t \right) - S_1(t) \right|. \] (22)

We have

\[ P \left( \sum_{j=0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > t \right) =
\]
\[ P \left( \sum_{j=0}^{J_0-1} 2^{-\frac{j}{2}} W_j(\lambda) + \sum_{j=J_0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > t \right) \leq
\]
\[ P \left( \bigg\{ \sum_{j=0}^{J_0-1} 2^{-\frac{j}{2}} W_j(\lambda) > t - \epsilon \bigg\} \lor \bigg\{ \sum_{j=J_0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > \epsilon \bigg\} \right) \leq
\]
\[ P \left( \sum_{j=0}^{J_0-1} 2^{-\frac{j}{2}} W_j(\lambda) > t - \epsilon \right) + P \left( \sum_{j=J_0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > \epsilon \right) \leq
\]
\[ (S_{1-2^{-J_0}}(t - \epsilon) + \epsilon) + \epsilon \leq S_{1-2^{-J_0}}(t) + \epsilon/\sqrt{\pi} + 2\epsilon \leq S_1(t) + \epsilon + (1/\sqrt{\pi} + 2) \epsilon \leq S_1(t) + 4\epsilon. \] (23)

On the other hand, we have

\[ P \left( \sum_{j=0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > t \right) \geq
\]
\[ P \left( \bigg\{ \sum_{j=0}^{J_0-1} 2^{-\frac{j}{2}} W_j(\lambda) > t + \epsilon \bigg\} \land \bigg\{ \sum_{j=J_0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > -\epsilon \bigg\} \right) =
\]
\[ P \left( \sum_{j=0}^{J_0-1} 2^{-\frac{j}{2}} W_j(\lambda) > t + \epsilon \right) + P \left( \sum_{j=J_0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > -\epsilon \right) -
\]
\[ P \left( \bigg\{ \sum_{j=0}^{J_0-1} 2^{-\frac{j}{2}} W_j(\lambda) > t + \epsilon \bigg\} \lor \bigg\{ \sum_{j=J_0}^{J-1} 2^{-\frac{j}{2}} W_j(\lambda) > -\epsilon \bigg\} \right) \geq
\]
\[ (S_{1-2^{-J_0}}(t + \epsilon) - \epsilon) + (1 - \epsilon) - 1 \geq S_{1-2^{-J_0}}(t) - \frac{1}{\sqrt{\pi}} \epsilon - 2\epsilon \geq
\]
\[ S_1(t) - \epsilon - \left( \frac{1}{\sqrt{\pi}} + 2 \right) \epsilon \geq S_1(t) - 4\epsilon. \] (24)

Inequalities (23) and (24) together prove that the difference \( D_1 \) of formula (22) is arbitrarily small for \( \lambda \) and \( J \) large enough, which proves the convergence.
We will now show that $M \xrightarrow{d} 0$ as $\lambda/N \to 0$. We denote by $S^0(t)$ the survival function of the constant variable 0. Consider the difference

$$D_2 = \left| P \left( N^{-1} \sum_{i=0}^{N-1} (V_i - \lambda) > t \right) - S^0(t) \right|. \quad (25)$$

For $t > 0$, we have

$$\begin{align*}
D_2 &= P \left( N^{-1} \sum_{i=0}^{N-1} (V_i - \lambda) > t \right) \leq N^{-2} t^{-2} \mathbb{E} \left( \sum_{i=0}^{N-1} (V_i - \lambda) \right)^2 \\
&= N^{-2} t^{-2} \sum_{i=0}^{N-1} \text{Var}(V_i) = N^{-1} t^{-2} \lambda \to 0 \quad \text{as} \quad \lambda/N \to 0. \quad (26)
\end{align*}$$

For $t < 0$, we have

$$\begin{align*}
D_2 &= \left| P \left( N^{-1} \sum_{i=0}^{N-1} (V_i - \lambda) > -|t| \right) - 1 \right| = P \left( N^{-1} \sum_{i=0}^{N-1} (V_i - \lambda) \leq -|t| \right) \\
&= P \left( -N^{-1} \sum_{i=0}^{N-1} V_i + \lambda \geq |t| \right) \leq N^{-1} t^{-2} \lambda \to 0 \quad \text{as} \quad \lambda/N \to 0. \quad (27)
\end{align*}$$

Inequalities (26) and (27) show that $M \xrightarrow{d} 0$ as $\lambda/N \to 0$. The proof of Proposition 2.2 is completed.

References


Nason, G. P. (1998), WaveThresh3 Software. Department of Mathematics, University of Bristol, Bristol, UK.


