Statistica Sinica **9**(1999), 119-135

WAVELET SHRINKAGE ESTIMATION OF CERTAIN POISSON INTENSITY SIGNALS USING CORRECTED THRESHOLDS

Eric D. Kolaczyk

Boston University

Abstract: Wavelet shrinkage estimation has been found to be a powerful tool for the non-parametric estimation of spatially variable phenomena. Most work in this area to date has concentrated primarily on the use of wavelet shrinkage techniques in contexts where the data are modeled as observations of a signal plus additive, Gaussian noise. In this paper, I introduce an approach to estimating intensity functions for a certain class of "burst-like" Poisson processes using wavelet shrinkage. The proposed method is based on the shrinkage of wavelet coefficients of the original count data, as opposed to the current approach of pre-processing the data using Anscombe's square root transform and working with the resulting data in a Gaussian framework. "Corrected" versions of the usual Gaussian-based shrinkage thresholds are used. The corrections explicitly account for effects of the first few cumulants of the Poisson distribution on the tails of the coefficient distributions. A large deviations argument is used to justify these corrections. The performance of the new method is examined, and compared to that of the pre-processing approach, in the context of an application to an astronomical gamma-ray burst signal.

Key words and phrases: Gamma-ray bursts, large deviations, Poisson processes, wavelets.

1. Introduction

Wavelet shrinkage techniques (e.g., Donoho, Johnstone, Kerkyacharian and Picard(1995)) have emerged recently as powerful methods for the non-parametric estimation of objects which may be characterized as 'spatially variable'. Examples of such objects include astronomical gamma-ray bursts, nuclear magnetic resonance spectra, and tomographic medical images. To date, most efforts by researchers have focused on problems in which the data are adequately modeled as observations of a signal plus additive, Gaussian noise. Deviations from this basic model often are dealt with through judicious use of transformations. For example, in the case of Poisson count data, a direct usage of Gaussian-based shrinkage thresholds usually is naive, given the effects of the Poisson distribution on the variance of the wavelet coefficients. Instead, the data typically are pre-processed using Anscombe's (1948) square root transformation, thereby normalizing the data and stabilizing the variance (e.g., Donoho (1993)). However,

in practice it is often found that this approach over-smooths the data, resulting in the elimination of sharp, brief structure in the underlying intensity function.

In this paper, I offer a new approach for adapting the wavelet shrinkage paradigm to Poisson data. Specifically, the original count data are analyzed, using an arbitrary orthonormal basis of periodic wavelets, but a set of "corrected" thresholds are applied to the corresponding wavelet coefficients. In deriving these thresholds, a large deviations probability calculation is used to approximate the tail probabilities of the wavelet coefficient distributions. The resulting thresholds explicitly correct the usual $(2\log(n))^{1/2}$ Gaussian-based thresholds for the effects of the Poisson distribution on these tails. In addition, I demonstrate that a particularly simple form of corrected thresholds may be derived for the special case of Haar wavelets, using a different set of approximations that are especially appropriate for situations involving very low levels of counts.

The rest of the paper is organized as follows. In Section 2, I define a certain class of "burst-like" Poisson processes and present certain distributional characteristics of the wavelet coefficients of such processes. The derivation of the corrected thresholds for these coefficients is given briefly in Section 3. Some issues regarding implementation of these results are discussed in Section 4, and the special case of Haar wavelets is presented. In Section 5, an application of the proposed methodology is presented in the context of estimating astronomical gamma-ray burst intensity functions. Finally, some discussion of additional related issues may be found in Section 6.

2. Burst-Like Poisson Processes

2.1. Background

Consider a Poisson process $N(0,t]$ on the interval $[0,1]$ with intensity function

$$\lambda(t) = \lambda_0 + \tilde{\lambda}(t),$$

where λ_0 is strictly positive and $\tilde{\lambda}(t) \geq 0$ may be described as being 'spatially inhomogeneous'. The process $N(0,t]$ is simply the sum of a homogeneous "background" Poisson process and a second, inhomogeneous Poisson process that tends to generate observations in "bursts". A perfect example of a phenomenon that may be well described by this model is the astronomical gamma-ray burst shown in Figure 1. This signal is one of the almost 4000 gamma-ray bursts collected since 1991 by the Burst and Transient Source Experiment (BATSE), on board NASA's Compton Gamma-Ray Observatory (Meegan (1992)). These bursts are observed on a fairly regular basis against the relatively constant background of gamma-ray emissions in outer space. However, despite having been first observed almost 30 years ago, much about the nature of the sources of these bursts remains a mystery. Additional details regarding these bursts are given in Section 5.

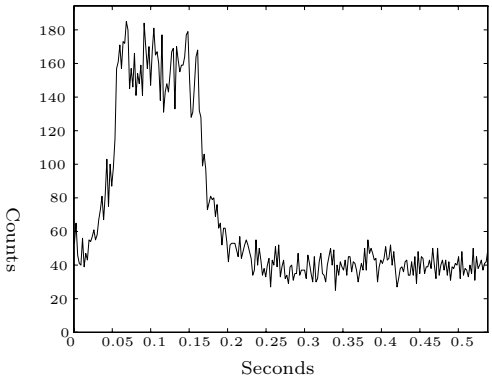

Figure 1. Gamma-Ray Burst (BATSE trigger 474). Original photon arrival times have been collected into $n = 512$ (roughly) two-millisecond bins. Only half of the burst data (about 0.54 seconds) that contain the actual pulse are displayed here.

The spatially variable nature of the data in Figure 1 suggests the use of wavelet shrinkage methods for estimating the underlying intensity function $\lambda(t)$. Such a methodology would be useful as a pre-processing step in large studies of many bursts (e.g., Norris et al. (1996)), as well as for more detailed analyses of the structure of individual bursts (e.g., Kolaczyk (1997)). The methodology in this paper will be based on the use of periodic wavelets on the interval $[0, 1]$. Specifically, functions

$$\psi^{\circ}_{j,k}(x) \equiv \sum_{l \in \mathbf{Z}} \psi_{j,k}(x + l),$$

where the $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ are dilations and translations of a single wavelet function $\psi \in L^2(\mathbb{R})$, defined so that the collection $\{\psi_{j,k}\}_{j,k \in \mathbf{Z}}$ forms an orthonormal basis of $L^2(\mathbb{R})$. The collection $\{\psi^{\circ}_{j,k}\}_{j \geq 0, k \in \mathbf{Z}}$ forms an orthonormal basis of the periodic functions in $L^2([0, 1])$. Wavelet functions typically are smooth, localized and oscillating. Additionally, wavelets are restricted to have zero integral, an important property in motivating the shrinkage of empirical wavelet coefficients in wavelet shrinkage methods. The reader is referred to Daubechies (1992) and Nason and Silverman (1994) for more background on wavelets.

2.2. Wavelet coefficients

Define the random variable $d_{j,k}$ to be the (j, k)th wavelet coefficient of the process $N(0, t]$ i.e.,

$$d_{j,k} \equiv \int_0^1 \psi^{\circ}_{j,k}(x)\, N(dx) = \sum_i \psi^{\circ}_{j,k}(x_i). \tag{1}$$

The summation in (1) expresses the wavelet coefficient in terms of the event arrival times $\{x_i\}$. By definition, the mean and variance of the wavelet coefficients take the form

$$E\left[d_{j,k}\right] = \int_0^1 \psi^{\circ}_{j,k}(x)\lambda(x)\,dx \quad \text{and} \quad \mathrm{Var}\left(d_{j,k}\right) = \int_0^1 \left[\psi^{\circ}_{j,k}(x)\right]^2 \lambda(x)\,dx.$$

Hence, the observed wavelet coefficients are unbiased estimates of the true coefficients of the intensity function $\lambda(t)$. Additionally, as might be expected, the variance of each coefficient is influenced by the value of $\lambda(t)$, primarily in the region of support of the corresponding wavelet function. Analogous expressions for higher order moments show a similar influence by $\lambda(t)$, indicating altogether that a simple Gaussian approximation to the coefficient distribution potentially may be quite unreasonable.

It is also instructive to consider the relationship between two distinct wavelet coefficients, say d_{j_1,k_1} and d_{j_2,k_2}. Their covariance takes the form

$$\mathrm{Cov}\left(d_{j_1,k_1}, d_{j_2,k_2}\right) = \int_0^1 \psi^{\circ}_{j_1,k_1}(x)\psi^{\circ}_{j_2,k_2}(x)\lambda(x)\,dx.$$

Hence, in general, the wavelet coefficients will be correlated. However, under $H_0 : \lambda(t) \equiv \lambda_0$, i.e., when the process $N(0,t]$ consists of simply a homogeneous background process, it follows from the above moment expressions and the orthonormality of the wavelet basis that the wavelet coefficients will be uncorrelated and possess identical marginal distributions, with zero mean and constant variance λ_0. It is under this particular null hypothesis that the thresholds in Section 3 are derived.

As a final point, it should be noted that the distributions of the wavelet coefficients $d_{j,k}$ are not necessarily independent, even under H_0. More specifically, a simple application of Campbell's theorem (e.g., see Kingman (1993), chapter 3) shows that the joint characteristic function of two wavelet coefficients is not equal to the product of the marginal characteristic functions unless the two wavelets have disjoint intervals of support. Therefore, when using wavelets of compact support, a short-term dependency among the coefficients is expected; with wavelets of infinite support, all coefficients are dependent to some degree, although the localized nature of the wavelet functions suggests that this dependency is effectively short-term as well.

2.3. Implications for wavelet shrinkage thresholds

The above results regarding the distributional properties of the $d_{j,k}$ have two key implications for the nature of the thresholds that are derived in this paper. First, thresholds designed for use in the Gaussian noise problem will be

inappropriate here; proper thresholds will need to incorporate some of the higher order distributional differences in the Poisson and Gaussian problems. Second, not only will the incorporation of these differences potentially change the form of the thresholds used, but also the number of thresholds used; the coefficient distribution in the Poisson problem commonly will not be symmetric, which suggests the need for a *pair* of thresholds.

In order to appreciate these issues in more detail, first recall the original formulation of the wavelet shrinkage methodology in the Gaussian noise problem. Donoho and Johnstone (1994) suggest that an n-length signal observed in additive, Gaussian noise may be "de-noised" through appropriate shrinkage towards zero of the empirical wavelet coefficients. Under the null hypothesis that the observations contain only noise, the wavelet coefficients, say $w_{j,k}$, will be independent and identically distributed Gaussian random variables, with zero mean and variance σ^2. The value $B_n^G = (2\log(n))^{1/2}\sigma$ is a probabilistic upper bound on the n coefficients, in the sense that

$$Pr(\max w_{j,k} \leq B_n^G) \quad \longrightarrow \quad 1 \qquad \text{as } n \to \infty. \tag{2}$$

See Leadbetter et al. (1983), page 14. Combined with the fact that the value $-B_n^G$ is in turn a probabilistic lower bound on the coefficients, this result suggests using the value B_n^G as a single threshold on the size (absolute value) of the wavelet coefficients. For data in which the underlying signal may be well-compressed by a wavelet transform, and for which there is a moderate to high signal-to-noise ratio, this threshold should serve to separate the "signal" and the "noise" into sets of "large" and "small" coefficients, respectively. Application of the appropriate inverse wavelet transform to the thresholded coefficients yields a de-noised estimate of the object underlying the data.

In moving from the Gaussian noise problem to that of burst-like Poisson processes, the analogous null hypothesis of "no signal" is $H_0 : \lambda(t) \equiv \lambda_0$. Under this condition, the wavelet coefficients $d_{j,k}$ have zero mean and common variance, as in the Gaussian setting, but they also have non-negligible skewness and kurtosis, not to mention some short-term dependency. These facts motivate the derivation in this paper of a new set of resolution level-dependent thresholds B_j, calibrated so that

$$Pr\Big(\max_{0\leq k\leq 2^j-1} d_{j,k} \leq B_j\Big) \tag{3}$$

approaches 1 at a rate similar to that in (2), as j increases. Additionally, not only will the thresholds B_j be different from (essentially, corrections of) the Gaussian thresholds, but the use of two thresholds $B_j^{max} \neq -B_j^{min}$ will be induced naturally within the derivation, in the case where the distribution of the $d_{j,k}$ is skewed.

Note that here and throughout the rest of this paper, probabilities involving the $d_{j,k}$ as in (3) will be written implicitly under the null hypothesis H_0.

3. Poisson Corrected Thresholds

A set of Poisson-corrected thresholds B_j are derived in this section for the wavelet coefficients $d_{j,k}$. The rate at which these thresholds allow (3) to tend to 1 is calibrated to that of the analogous B_j^G in the Gaussian case. Using suitable approximations for the probabilities in both the Poisson and Gaussian settings, this calibration takes the form of an equation which may be simplified to a defining equation for B_j. Details are given in Section 4 on how this latter equation may be used to solve for B_j numerically.

The primary task in the derivation of B_j is that of developing a usable approximation for the probability in (3). This task may be simplified somewhat by starting with the expression

$$Pr\Big(\max_{0\le j\le 2^j-1} d_{j,k} \le B_j\Big) \approx \exp\Big\{-2^j \frac{Pr\,(d_{j,k} > B_j)}{E[\mathcal{C}_{B_j}]}\Big\}. \tag{4}$$

Through (4), the original task is now reduced to that of finding a tractable expression for the tail probability of $d_{j,k}$. The approximation in (4) comes from application of Aldous' Poisson Clumping Heuristic (Aldous (1989)). The quantity $E[\mathcal{C}_{B_j}]$ is the expected "clump size", roughly the expected number of local exceedances of B_j, conditional on there being at least a single exceedance, which implies $\mathcal{C}_{B_j} \ge 1$. If it were the case that the $d_{j,k}$ were independent, we would have $E[\mathcal{C}_{B_j}] = 1$. In Section 4, it will be argued that while the coefficients themselves are not independent, the events $\{d_{j,k} > B_j\}$ are approximately independent for large B_j, which suggests that the expected 'clump size' is still approximately equal to 1.

The remainder of this section consists of two parts. In Section 3.1, an approximation for $P(d_{j,k} > B_j)$ is presented. Substituting this approximation into the right hand side of (4), the calibration referred to above may then be performed, resulting in the defining equation for B_j. Details of the calibration are given in Section 3.2.

3.1. Approximating the coefficient tail probability

Although we will see in Section 4 that a closed form expression for the coefficient tail probability $Pr(d_{j,k} > B_j)$ exists when using the Haar wavelet, this is not generally true for other choices of the wavelet function. Accordingly, we settle for an approximation to this tail probability. One might consider simply approximating the coefficient distribution by an appropriate normal distribution. However, the distributional characteristics derived in Section 2.2 suggest

that this approach will be unsatisfactory, especially as the probabilities under consideration are expected to be far out in the tails of the individual coefficient distributions.

This latter point suggests the use of a large deviations approximation. A method described in Feller (1971), Chapter 16.7, may be adapted to the present context for this purpose, yielding an expression that explicitly incorporates effects of just the first few cumulants on the far tail of the coefficient distribution.

Theorem 1. *Let* $\kappa_m = \lambda_0 \int_0^1 [\psi^\circ_{j,k}(x)]^m dx$ *denote the* m*-th cumulant of the distribution of* $d_{j,k}$*, and consider thresholds of the form* $B_j \equiv b_j \kappa_2^{1/2}$*. Then, under the assumption that* $b_j = o(\lambda_0^{3/10})$*,*

$$Pr\,(d_{j,k} > B_j) \sim \exp\{\gamma_1\, b_j^3 + \gamma_2\, b_j^4\}\,[1 - \Phi(b_j)]\,, \tag{5}$$

where

$$\gamma_1 = \frac{\kappa_3}{6\kappa_2^{3/2}} \quad \textit{and} \quad \gamma_2 = \frac{\kappa_2\kappa_4 - 2\kappa_3^2}{8\kappa_2^3}, \tag{6}$$

and Φ *is the standard normal cumulative distribution function.*

The proof of Theorem 1 relies on an exponential tilting of the distribution of $d_{j,k}$. This straightforward adaptation of Feller's (1971) method allows one, in principle, to work within a hierarchy of approximations between the two extremes of the Central Limit Theorem (simple, but less accurate) and, say, a standard saddlepoint approximation (accurate, but less practical from the standpoint of efficient implementation). Each successive approximation in this hierarchy incorporates cumulants of higher and higher orders, accompanied by the appropriate conditions on the growth of the thresholds b_j with that of λ_0. The expression in (5) may be viewed as a correction of the simple Central Limit Theorem approximation for the effects of skewness and kurtosis in the distribution of $d_{j,k}$. Note that the cumulants κ_m may be computed easily in the discrete setting. This point is critical because the coefficients γ_1 and γ_2 arise explicitly in the definition of the values b_j, as explained next. (The reader may notice that the coefficient γ_2 is slightly different from that obtained by Feller (1971) in the context of sums of independent random variables. In fact, the value for γ_2 given by Feller is incorrect, and may be shown to be equal to the quantity defined above.)

3.2. Calibrating the thresholds

A quick glance at the wavelet shrinkage literature to date is sufficient to demonstrate that the threshold B_n^G, or variants thereof, lies at the heart of a fair percentage of the methodology suggested for the Gaussian noise problem. Accordingly, I take this threshold to serve as a canonical choice with Gaussian data,

and calibrate a set of thresholds for Poisson data to have a similar probabilistic character. Specifically, as the probability in (2) tends to 1 asymptotically like

$$\exp\left\{-\left[4\pi\log(n)\right]^{-1/2}\right\}, \tag{7}$$

the Poisson thresholds B_j may be calibrated with B_j^G by setting the right hand side of (4) equal to (7), with $n = 2^j$, and substituting (5) appropriately. Taking logarithms twice on both sides of the resulting expression yields the equation

$$\gamma_2\, b_j^4 + \gamma_1\, b_j^3 + \log\left[1 - \Phi(b_j)\right] + \log(2^j) - \log(E[\mathcal{C}_{B_j}]) = -\frac{1}{2}\log[4\pi\log(2^j)].$$

Using the well-known approximation $1 - \Phi(x) \sim \phi(x)/x$, where $\phi(x)$ is the standard normal density function, and simplifying further, we arrive at the result

$$\gamma_2\, b_j^4 + \gamma_1\, b_j^3 - \frac{1}{2}\, b_j^2 - \log(b_j) + \log[2^j(2\log(2^j))^{1/2}] - \log(E[\mathcal{C}_{B_j}]) = 0. \tag{8}$$

Hence, the threshold $B_j \equiv b_j\kappa_2^{1/2}$ is proportional to the standard deviation of the coefficients of the background noise process, where the constant of proportionality is defined to be the solution of a particular equation. This equation involves b_j in polynomial and logarithmic terms, with coefficients that reflect the effects of the third and fourth order cumulants of the $d_{j,k}$. In addition, there is a constant term relating to the number of coefficients at resolution level j and, finally, the term capturing the effects of dependency among the coefficients through the expected clump size, $E[\mathcal{C}_{B_j}]$.

Details on the numerical implementation of (8) follow in the next section. However, before proceeding, note that technically the threshold just derived should be labeled B_j^{max}, denoting it as a probabilistic upper bound on the *maximum* of the wavelet coefficients. Recall from the discussion in Section 2.3 that, in general, a second threshold $B_j^{min} \neq -B_j^{max}$ is needed as well to bound the minimum of the coefficients. An argument analogous to that used above shows that $B_j^{min} = b_j^{min}\kappa_2^{1/2}$, where $-b_j^{min}$ is a solution to an equation identical in form to that in (8), but with the coefficients γ_1 and γ_2 replaced by

$$\tilde{\gamma}_1 = -\gamma_1 \quad \text{and} \quad \tilde{\gamma}_2 = \gamma_2.$$

4. Implementation

In this section, some issues relating to the computation of the threshold values B_j are discussed. An example is given as well, wherein these corrected Poisson thresholds are compared to the analogous Gaussian thresholds. Finally,

the context of data with very low levels of counts is discussed, where computation of the B_j may break down, and an alternative using Haar wavelets is suggested.

4.1. Computing the B_j

Although the results presented in Sections 2.2 and 3.1 were derived in the context of a continuous-time Poisson process $N(0,t]$, the adaptation of these results to discrete data is entirely straightforward. Therefore, from a computational point of view, the only non-trivial value involved in computing b_j from (8) is the logarithmic term involving the expected clump size, $E[\mathcal{C}_{B_j}]$. In general, for a discrete-time stochastic process, the expected clump size is bounded below by 1. When the components of the process are independent and identically distributed, we have $E[\mathcal{C}] = 1$ trivially. However, if there is dependency among these components, as is true in the case of the $d_{j,k}$, the value of $E[\mathcal{C}]$ may be greater than 1, and estimation of this value generally will be non-trivial, as it "... ultimately must involve the particular structure ..." of the underlying process (Aldous (1989)).

In the present context, an heuristic argument may be used to suggest that, as a rough estimate, we may take $E[\mathcal{C}_{B_j}] \approx 1$. Specifically, by definition of the expected clump size, the essential condition for this approximation to be valid is that, for fixed u,

$$Pr\Big(\frac{d_{j,k_0+u}}{\kappa_2^{1/2}} > b_j \Big| \frac{d_{j,k_0}}{\kappa_2^{1/2}} > b_j \Big) \longrightarrow 0 \quad \text{as} \quad b_j \longrightarrow \infty \tag{9}$$

(see Aldous (1989), pg. 50). By way of justification, begin by noting that the standardized wavelet coefficients in (9) are uncorrelated, have zero mean, unit variance, and cumulants κ_m that tend to zero like $\lambda_0^{-(m-2)/2}$, for $m \geq 3$. Hence, as $\lambda_0 \to \infty$, which is what drives b_j towards infinity, the probability in (9) tends towards a conditional probability between two uncorrelated, Gaussian random variables. Because uncorrelated Gaussian random variables are also independent, this probability looks asymptotically like the marginal tail probability of the standardized d_{j,k_0+u}. Under the condition $b_j = o(\lambda_0^{1/2})$, a modification of the arguments in Section 3.1 shows that this latter probability acts like $e^{-b_j^2/2}$, which tends to zero quickly as $b_j \to \infty$.

As a result of this argument, the term $\log(E[\mathcal{C}_{B_j}])$ in (8) essentially vanishes. Hence, the defining equation for b_j becomes

$$\gamma_2\, b_j^4 + \gamma_1\, b_j^3 - \frac{1}{2}\, b_j^2 - \log(b_j) + \log[2^j (2\log(2^j))^{1/2}] = 0, \tag{10}$$

which is just a fourth order polynomial, with the additional term $\log(b_j)$. The left hand side of (10) is quite smooth, and an estimate of b_j may be obtained easily,

for example using the golden section method (Press, Teukolsky, Vetterling and Flannery (1992)). Although more than one such zero will exist, the search may be restricted to a suitable region near $b_j^G \equiv (2\log(2^j))^{1/2}$, the analogue of b_j in the Gaussian setting. Note that b_j^G is a zero of (10) when $\gamma_1 = \gamma_2 = 0$, and recall that b_j essentially was derived as a correction of b_j^G. Because the coefficients γ_1 and γ_2 actually are quite small in practice, as compared to the coefficient -0.5 of the term b_j^2 in (10), these facts suggest that b_j be defined as the zero occurring close to b_j^G and not another (usually ridiculously large) value.

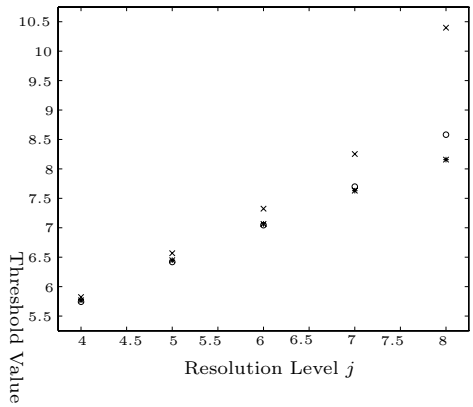

Figure 2. Comparison of Poisson corrected thresholds B_j^{max} ('x') and $-B_j^{min}$ ('o') with the analogous level-dependent Gaussian thresholds B_j^G ('$*$').

For the purpose of illustration, a comparison of the values B_j and B_j^G, for a particular case, are shown in Figure 2. In this example, the most nearly symmetric Daubechies wavelets of order 8 were used, with sample size $n = 512$ and $\lambda_0 = 6$. Thresholds were computed for resolution levels $j = 4, \ldots, 8$. These specific settings were chosen for this example because they correspond to those used in the analysis of the gamma-ray burst channel data in Section 5 below. Note that the relative discrepancy from B_j^G increases as resolution level increases, and is particularly evident at the highest resolution level. A simple calculation shows that, for all other factors fixed, the cumulants κ_m increase with increasing resolution level j like $2^{j(m-2)/2}$, for $m \geq 3$. Hence, the greater discrepancy at higher resolution levels reflects the fact that the distribution of coefficients at those levels is decidedly "less Gaussian".

4.2. Low-count data and the Haar wavelet

Experience has shown that for small values of λ_0 (depending on the choice of wavelet function), the value of γ_2 may be large enough to make the left hand side

of (10) strictly positive in the region around b_j^G. Hence, in situations involving very low levels of observed counts, a reasonable solution for b_j may not exist. For the choice of wavelet and sample size used in the example above, solutions at the highest resolution level ceased to exist for values of λ_0 less than about 5.5. This computational limitation perhaps is not entirely surprising, considering that the derivation of (8) is based on an argument asymptotic in λ_0. Below, I present an alternative set of thresholds, for the special case in which the Haar wavelet is used, that is particularly appropriate for applications with low levels of counts.

Assume that the observations from the process $N(0,t]$ have been binned to form the values $y_0, \ldots, y_{n-1}$, where $y_i \sim \text{Poisson}(\Lambda_i)$, $\Lambda_i = \int_{i/n}^{(i+1)/n} \lambda(u)du$, and $n = 2^J$ for some $J > 0$. The empirical wavelet coefficients in the Haar transform of the vector $\mathbf{y}$, are simply proportional to the differences of counts in pairs of adjacent bins, where the size of the bins is dyadic and determined by the particular resolution level j of a given coefficient. Recalling that the distribution of the difference of independent Poisson random variables may be expressed in terms of the distribution of a certain non-central chi-square random variable (Johnson (1959)), we find that the distribution of any given Haar coefficient is symmetric under $H_0 : \lambda(t) \equiv \lambda_0$, and write

$$\Pr\left(|d_{j,k}| < B_j\right) = 1 - 2\Pr\left[\chi^2_{(2m_j)}(\lambda_j) < \lambda_j\right], \tag{11}$$

where $m_j = 2^{(J-j)/2} B_j$ and $\lambda_j = \lambda_0/2^j$.

Inspired by the approach in Section 3, I use a set of simple approximations to the distribution of $d_{j,k}$, to facilitate a calibration with the canonical Gaussian threshold, B_j^G. Specifically, write

$$\Pr\left[\chi^2_{(2m_j)}(\lambda_j) < \lambda_j\right] \approx \Pr\left[\chi^2_{(f_j)} < \lambda_j/c_j\right] \approx \Pr\left[Z < \frac{(\lambda_j/c_j) - f_j}{\sqrt{f_j}}\right], \tag{12}$$

where $f_j = (2m_j + \lambda_j)^2/(2m_j + 2\lambda_j)$, $c_j = (2m_j + 2\lambda_j)/(2m_j + \lambda_j)$, and Z is a standard normal random variable. The first approximation is due to Patnaik (1948), and has been found to be especially accurate for what corresponds to low levels of λ_j (Johnson, Kotz and Balakrishnan (1995), page 464). The second approximation is a simple Central Limit Theorem approximation which, due to the reliance of f_j on the threshold B_j, can be expected to be quite reasonable in practical situations.

The desired calibration may be effected by equating the argument of the Gaussian tail probability in (12) with $z = (2\log(n))^{1/2}$. After simplifying, we find that the defining equation has a quadratic form i.e., $f_j^2 - 4\log(2^j)\, f_j -$

$8\log(2^j)\lambda_j = 0$. Taking the positive root of this equation and making the proper substitutions yields the threshold value

$$B_j = 2^{-(J-j)/2}\Big\{\log(2^j) + \Big[\log^2(2^j) + 2\log(2^j)\,\lambda_j\Big]^{1/2}\Big\}. \tag{13}$$

The relative simplicity of the above derivation, in contrast to that of the thresholds in Section 3, results in part from the availability of a closed-form expression for $\Pr(d_{j,k} > B_j)$ and from the independence of the Haar coefficients as well. The thresholds defined by (13) have been used successfully in estimating the intensity functions underlying a variety of astronomical gamma-ray burst signals like the one shown in Figure 1. Note that, due to the block-like structure of the Haar wavelets, implementation naturally is done within a cycle-spinning or translation-invariant framework (Donoho and Coifman (1995)). The reader is referred to Kolaczyk (1997) for details.

5. An Application

As an illustration of the primary methodology proposed in Section 3 of this paper, I provide an analysis of the astronomical gamma-ray burst (GRB) data presented earlier in Figure 1. The raw data were recorded as the arrival times of the high-energy photons (essentially light at the high end of the electro-magnetic spectrum) of which the burst is composed. Using specialized detectors equipped with charged-coupled devices (CCDs), arrivals initially are collected over four separate energy channels (i.e., at 25 - 58 keV, 58 - 115 keV, 115 - 320 keV, and > 320 keV). However, it is common to aggregate over all energies, and to collect the arrival times into adjacent, equi-length bins, yielding plots of total counts versus time, as shown in Figure 1. Astronomers have found that the counts may be well-modeled as having arrived from a Poisson process of unknown intensity (Scargle (1996)). GRBs such as this were observed first in the late 1960s by satellites designed to monitor violations of the nuclear test-ban treaty. Although thousands of these bursts have been observed since then, no two have been found to be alike, and their origin (and even much of their nature) remains one of the oldest mysteries in astronomy. The GRB in Figure 1 was observed over 1.08 seconds. Counts were placed into $n = 512$ bins. Note that the full data for this burst consist of the pulse-like event itself, followed by a period during which it is judged that the system has returned to the original background levels. Because the pulse occurred exclusively in the first half of the data, only this half (i.e., the first 256 bins) has been displayed in Figure 1 and the other figures below, so as to highlight the structure therein. However, the full set of data was used in all calculations.

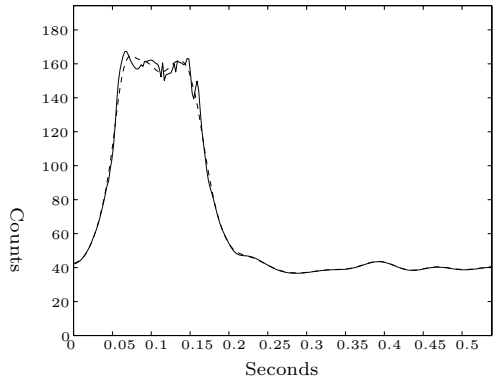

Figure 3. Estimates of the underlying intensity function for the Gamma-Ray Burst in Figure 1 using the Poisson corrected thresholds (solid), and the standard square-root based approach (dotted). The most nearly symmetric Daubechies wavelets of order 8 were used in both methods, with soft thresholding of the coefficients at resolution levels 5 through 8.

Figure 3 shows two estimates of the intensity function underlying the GRB. Both estimates were obtained using variations of the basic wavelet shrinkage methodology: wavelet transform, threshold coefficients, inverse wavelet transform. For the estimate drawn in a solid line, the methodology of this paper was used. Specifically, the wavelet transform of the 512 counts was computed, the level-dependent Poisson-corrected thresholds B_j were applied, and the corresponding inverse wavelet transform was computed. Alternatively for the estimate drawn in a dotted line, Anscombe's (1948) square root transformation was used first to pre-process the data, yielding observations $x_i \equiv 2(y_i+3/8)^{1/2}$ that are approximately normal in distribution, with roughly unit variance. The analogous level-dependent Gaussian thresholds $B_j^G = b_j^G$ (i.e., using unit variance) were applied in this case. Both methods were implemented within the translation-invariant de-noising framework of Donoho and Coifman (1995), which has been found to be useful in general for reducing pseudo-Gibbs phenomena.

Examining the estimates in Figure 3, it is clear that the Poisson-corrected thresholds have allowed for better preservation of the fine structure within the GRB. In particular, while the Poisson-corrected method has maintained the integrity of the three small bumps along the plateau, the square root method has smeared them into two larger bumps. Additionally, the former method has kept some evidence of the extremely sharp, brief spike at about 0.16 seconds; the latter method has eliminated any sign of this phenomenon. In the region from roughly 0.25 seconds onward, which contains the smooth, relatively constant background, both methods are almost identical in the estimated values they

yield. Thresholds in both methods were applied using soft thresholding (i.e., $\eta_B(x) = \mathrm{sgn}(x)(|x| - B)_+$), a fact which accounts for some of the attenuation in the estimates.

Of additional interest to astronomers is sometimes an analysis of the data by energy channel. Often the occurence of a phenomenon at more than one energy level is taken as additional evidence that it truly exists. When the photon arrival times are separated according to their original four channels, the level of counts involved typically is much lower than in the aggregate. This is particularly true in the fourth channel, which registers photons with energies in excess of 320 keV. Anscombe's square root transformation is less effective in settings involving very low levels of counts, and wavelet shrinkage estimates using data pre-processed with this transformation tend to exhibit noticeable over-smoothing of the data. Hence, any improvement in such settings is particularly welcome. Figure 4(a) shows the GRB of Figure 1 broken down by energy channel. The background level of counts corresponds to roughly 11 counts per bin in the first three channels, and just over 6 counts per bin in the fourth channel. Estimates of the underlying intensity functions are shown in Figure 4(b), deriving from the Poisson-corrected methods. The square root method was found to produce estimates with an unacceptably rounded appearance. On the other hand, the method using the Poisson-corrected thresholds has produced reasonably sharp estimates in areas where they seem quite plausible, such as the individual peaks in the third channel.

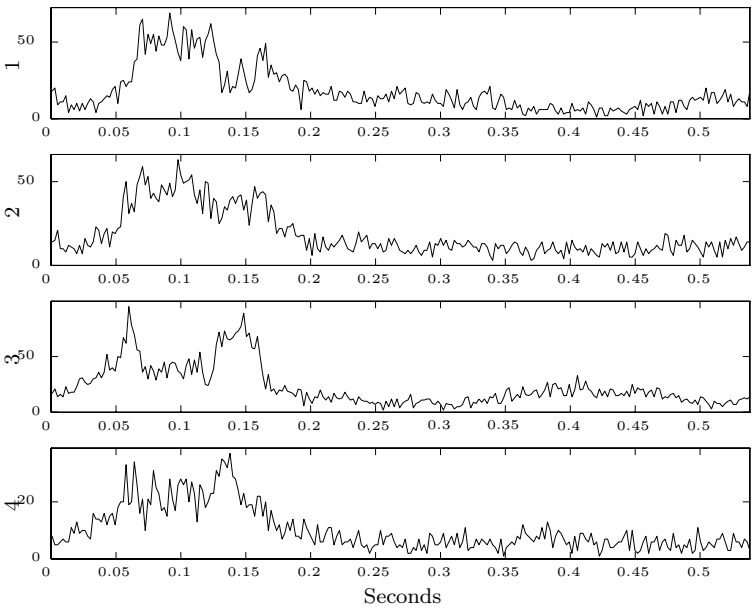

Figure 4(a)

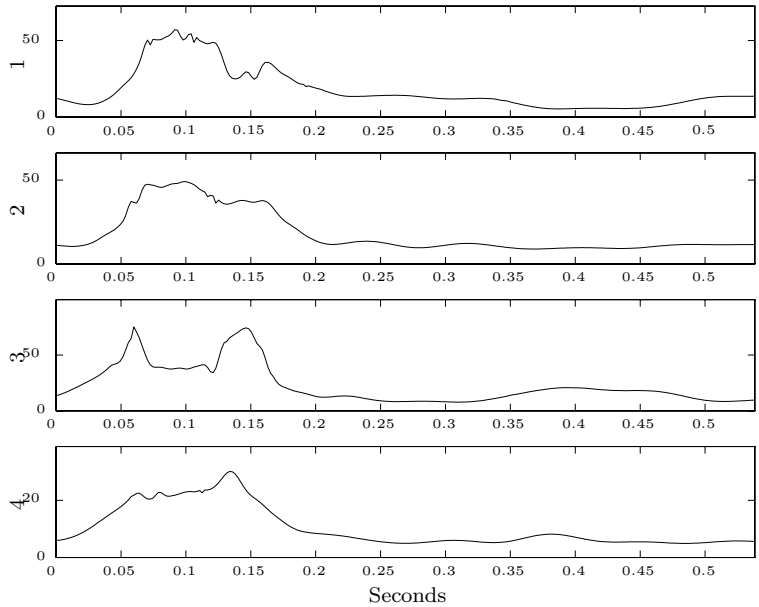

Figure 4(b)

Figure 4. (a): Decomposition of the Gamma-Ray Burst in Figure 1 into four energy channels. Reading from top to bottom, the channels are 25 - 58 keV, 58 - 115 keV, 115 - 320 keV, and > 320 keV. (b): Estimates of channel intensity functions are shown, based on Poisson corrected thresholds. Choice of wavelet and resolution levels thresholded is the same as in Figure 3.

6. Discussion

The distributional properties of the wavelet coefficients in the Poisson noise problem differ from those of the analogous coefficients in the Gaussian noise problem. In this paper I outline a method whereby these differences may be corrected for in the thresholding stage of the standard wavelet shrinkage methodology, instead of relying on the usual pre-processing tools. The corrections suggested here serve to account for effects of the first few cumulants of the Poisson distribution on the tails of the wavelet coefficient distributions. One may also account for the effect of short-term dependency among the coefficients. The brief argument used in Section 4.1 suggests that this effect may be treated as being negligible. However a more refined argument, perhaps leading to a more sensitive small-sample estimate of this effect, may be of interest. Additionally, although the defining equation presented here results from calibration with the Gaussian "universal" threshold i.e., $(2\log(n))^{1/2}$, the probabilistic behavior of other thresholds may be matched in an analogous manner. In a related direction, the large deviations approximation to the coefficient tail probability may be used in principle to modify

the False Discovery Rate (FDR) algorithm of Abramovich and Benjamini (1995) in an obvious manner. This algorithm results from considering the thresholding of wavelet coefficients from the perspective of multiple hypothesis testing, an approach also considered by Ogden and Parzen in (1996a) and (1996b).

From a broader perspective, there are further issues to consider. For example, a more careful study of the dependencies among the coefficients might yield measurable improvements via a joint thresholding rule, in place of the independent thresholding rule adopted here in analogy to the Gaussian case. Also, the replacement of simple thresholding rules by more flexible shrinkage rules merits attention. The adoption of various Bayesian paradigms (e.g., Abramovich, Sapatinas and Silverman (1998); Chipman, Kolaczyk and McCulloch (1997); and Clyde, Parmigiani and Vidakovic (1998)) has proven especially useful and natural for this purpose, in the Gaussian case. Finally, although the thresholds presented herein serve as a starting point for building up a body of knowledge with respect to wavelet shrinkage estimators from Poisson data, work in the Gaussian case is still far ahead as far as a study of general properties of the estimators is concerned. Work on this and some of the other points mentioned above is currently in progress by the author and colleagues.

Acknowledgements

The author would like to thank David Siegmund for a number of helpful discussions. The manuscript was prepared using computer facilities supported in part by National Science Foundation grant DMS 89-05292 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

References

Abramovich, F. and Benjamini, Y. (1995). Thresholding of wavelet coefficients as multiple Hypotheses testing procedure. In *Wavelets and Statistics* (Edited by A. Antoniadis and G. Oppenheim). Springer-Verlag, New York.

Abramovich, F., Sapatinas, T. and Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. Roy. Statist. Soc. Ser. B* **60**, 725-750.

Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic.* Springer-Verlag, New York.

Anscombe, F. (1948). The transformation of poisson, binomial and negative binomial data. *Biometrika* **35**, 246-254.

Chipman, H. A., Kolaczyk, E. A. and McCulloch, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *J. Amer. Statist. Assoc.* **92**, 1413-1421.

Clyde, M., Parmigiani, G. and Vidakovic, B. (1998). Multiple shrinkage and subset selection in wavelets. *Biometrika* **85**, 391-401.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia, Pennsylvania.

DeVore, R. A. and Lucier, B. J. (1992). Fast wavelet techniques for near-optimal image processing. IEEE-ICASSP-93, pgs. 48.3.1-48.3.7. IEEE Military Communications Conference, New York, NY.

Donoho, D. L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Proc. Sympos. Appl. Math.* Vol. 47 (Edited by Daubechies). AMS.

Donoho, D. L and Coifman, R. R. (1995). Translation-invariant De-Noising. In *Wavelets and Statistics* (Edited by A. Antoniadis and G. Oppenheim). Springer-Verlag, New York.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81**, 425-455.

Donoho, D. L, Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia ? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57**, 301-370.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2. Wiley, New York.

Johnson, N. L. (1959). On an extension of the connection between Poisson and χ^2-distributions. *Biometrika* **46**, 352-363.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Volume 2, Second Edition. Wiley, New York.

Kingman, J. F. C. (1993). *Poisson Processes.* Clarendon Press, Oxford.

Kolaczyk, E. D. (1997). Non-parametric estimation of Gamma-ray burst intensities using Haar wavelets. *The Astrophysical Journal* **483**, 340-349.

Leadbetter, M. A., Lindgrea, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

Meegan, C. A. et al. (1992). The spatial distribution of gamma ray bursts observed by BATSE. *Nature* **335**, 143-145.

Nason, G. P. and Silverman, B. W. (1994). The discrete wavelet transform in S. *J. Comput. Graphical Statist.* **3**, 163-191.

Norris, J. P. et al. (1996). Attributes of pulses in long bright Gamma-ray bursts. *The Astrophysical Journal* **459**, 393-412.

Ogden, T. and Parzen, E. (1996a). Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Comput. Statist. Data Anal.* **22**, 53-70.

Ogden, T. and Parzen, E. (1996b). Change-point approach to data analytic wavelet thresholding. *Statist. Comput.* **6**, 92-99.

Patnaik, P. B. (1949). The non-central χ^2- and F-distributions and their applications. *Biometrika* **36**, 202-232.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C.* Cambridge University Press.

Scargle, J. D. (1996). NASA ames research center. Personal communication.

Department of Mathematics and Statistics, Boston University, 111 Commington Street, Boston, MA 02215, U.S.A.

E-mial: kolaczyk@math.bu.edu

(Received March 1997; accepted April 1998)