WAVELETS IN BAYESIAN CHANGE-POINT ANALYSIS

R. Todd Ogden, University of South Carolina
Department of Statistics, University of South Carolina, Columbia, SC 29208

KEY WORDS: empirical coefficients, discrete wavelet transform, Haar wavelet, posterior density, projection

A Bayesian analysis based on the empirical wavelet coefficients is developed for the standard change-point problem. This analysis is considered first for the piecewise constant Haar wavelet basis, then extended to using smooth and orthogonal boundary wavelets. The analysis, though developed for use in the standard change-point model, can be applied to estimating a discontinuity in an otherwise smooth function by using only higher level coefficients in the computation of the posterior distribution, thereby effectively removing a "smooth" of the function. The procedure is illustrated with simulated data examples.

1 Introduction

In the general change-point problem, one observes independent ordered data $Y_1, \ldots, Y_n$, for which

$$Y_i \sim \mathrm{N}(f(i/n), \sigma^2),$$

for some function f defined on $[0,1]$. In the standard model, f has the form

$$f(u) = \begin{cases} \mu, & 0 \le u < \tau, \\ \mu + \Delta, & \tau \le u \le 1, \end{cases}$$

for some *change-point* $\tau \in [0,1)$.

There are two stages to classical change-point inference: testing for the existence of a change-point, and, if it is determined that a change-point exists, estimating the parameters τ, μ, Δ, and σ^2. Note that in the second stage, the estimation of the change-point τ is key, since given $\hat{\tau}$, estimation of the other three parameters is quite straightforward. The methods developed in this paper are aimed primarily at the Bayesian estimation of τ and can also be used as graphical diagnostics for determining whether a change-point is present.

The standard change-point problem considered above can be generalized by considering other possibilities for the mean-generating function f. In particular, let

$$f(u) = \begin{cases} g(u), & 0 \le u < \tau, \\ h(u), & \tau \le u \le 1, \end{cases}$$

for a change-point τ, and define $\Delta = h(\tau) - g(\tau)$.

A traditional nonparametric approach to the standard change-point problem involves the cumulative sum (CUSUM) process (see, e.g., Csorgő and Horváth (1997)). Define the *sample Brownian bridge process* on $[0,1]$ to be

$$\tilde{B}_n(u) = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{[nu]} (Y_i - \bar{Y})}{\hat{\sigma}}$$

for some consistent estimator $\hat{\sigma}$ of σ. Under the null hypothesis, this process converges in distribution to a standard Brownian bridge stochastic process. Various classical test statistics for the change-point problem and goodness-of-fit (Kolmogorov-Smirnov, Anderson-Darling, Cramér-von Mises, etc.) are based on this process. Plotting the sample Brownian bridge process provides a useful diagnostic.

Define the "raw" estimator of $f(u)$ to be the piecewise constant function

$$\tilde{f}(u) = \begin{cases} Y_i, & \frac{i-1}{n} \le u < \frac{i}{n},\ i = 1, \ldots, n \\ 0, & u = 1. \end{cases}$$

2 Wavelets

Wavelets have recently become very popular in statistics, as well as in a number of other applications. Since a basic understanding of wavelets is key to this paper, a brief introduction follows. Good mathematical treatments are given in Daubechies (1992) and Kaiser (1994), among other sources. An introduction with a lower level of mathematical sophistication focusing on statistical application is provided in Ogden (1996).

The simplest of all wavelet bases is that due to Haar (1910). The *mother wavelet* for the Haar system is defined as

$$\psi(x) = \begin{cases} 1, & 0 \le x < 1/2 \\ -1, & 1/2 \le x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

An entire family of wavelets is constructed by dilating and translating the mother wavelet:

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

Thus translated and dilated, the set of Haar functions $\{\psi_{j,k},\ j,k \in \mathbb{Z}\}$ form an orthonormal basis for $L^2(\mathbb{R})$.

Associated with the Haar mother wavelet is a *father wavelet* or *scaling function* ϕ, defined (in the Haar case) to be

$$\phi(x) = \begin{cases} 1, & 0 \le x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

An entire set of scaling functions are constructed from this function according to

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k).$$

Defining the subspace V_j to be the span of $\{\phi_{j,k},\ k \in \mathbb{Z}\}$ gives a nested sequence of approximation spaces in $L^2(\mathbb{R})$. The subspace V_j consists of all functions in $L^2(\mathbb{R})$ which are piecewise constant over intervals of length 2^{-j}. Thus, any L^2 function can be approximated arbitrarily well by projecting it onto a V_j subspace with j chosen to be large enough.

The set of wavelet functions with dilation index j forms an orthonormal basis for the "detail space" between successive approximations V_j and W_j, i.e.,

$$V_{j+1} = V_j \bigoplus W_j.$$

This is the basis for the *multiresolution analysis* principle of wavelets.

The piecewise constant multiresolution analysis for the Haar wavelet basis is extended by taking smoother mother and father wavelets, such as the compactly supported orthonormal families of Daubechies wavelets (see Daubechies (1992)). The same principles apply as those for the Haar basis, with approximation spaces and details spaces which are as smooth as the functions which generate them.

Since the unit interval $[0,1]$ is of interest in the change-point problem as defined earlier, an orthonormal basis for $L^2[0,1]$ can be formed by restricting the wavelets to live on the unit interval. This is easily enough done for the Haar wavelets, since $\text{supp}\psi_{j,k} = [2^{-j}k, 2^{-j}(k+1))$, but it is not as straightforward for smoother wavelets. One common method for adapting smooth wavelets to the interval is "periodic boundary handling," in which the function f is considered to be periodic with period 1. Cohen, Daubechies, and Vial (1993) develop a method to modify the usual Daubechies wavelets to live on the interval — the resulting wavelet functions are not all simply dilations and translations of a single mother wavelet however. When wavelets (and their corresponding scaling functions) are appropriately modified, they form a multiresolution analysis of $L^2[0,1]$, with nested approximation spaces and with wavelets forming bases for the detail spaces in between.

If the true function f were known, *wavelet coefficients* could be computed via

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = \int f(u)\psi_{j,k}(u)\,du. \qquad (1)$$

Since f is never known, *empirical wavelet coefficients* can be calculated by substituting $\tilde{f}$ in place of f in (1):

$$w_{j,k} = \langle \tilde{f}, \psi_{j,k} \rangle = \int \tilde{f}(u)\psi_{j,k}(u)\,du. \qquad (2)$$

Although the expression in (2) is useful for thinking about empirical wavelet coefficients, it should be noted that they are not actually computed that way, but by a fast wavelet algorithm.

For the development of this paper, it is assumed that the sample size n is a power of two. This will result in $n-1$ wavelet coefficients. When the sample size is not a power of two, various practical remedies are possible — see Ogden (1997) for discussion and comparison.

3 Bayesian analysis

Richwine (1996) applied the Haar wavelet basis to the change-point problem.

Since the empirical coefficients are independent,

$$p(\underset{\sim}{w}\,|\tau, \Delta, \sigma^2) = \prod_j \prod_k p(w_{j,k}|\tau, \Delta, \sigma^2),$$

where $\underset{\sim}{w}$ represents the vector of coefficients. The posterior distribution of the parameters is thus

$$p(\tau, \Delta, \sigma^2|\underset{\sim}{w}) \propto \prod_j \prod_k p(w_{j,k}|\tau, \Delta, \sigma^2) \cdot \pi(\tau, \Delta, \sigma^2) \qquad (3)$$

for a joint prior distribution π on τ, Δ, and σ^2.

The distribution of a single Haar wavelet coefficient defined as in (2) is

$$w_{j,k}|\tau, \Delta, \sigma^2 \sim \mathrm{N}\left(\Delta q_{j,k}(\tau), \sigma^2/n\right) \qquad (4)$$

where the mean function $q_{j,k}(\tau)$ is defined as

$$q_{j,k}(\tau) = \frac{2^{j/2}}{n} \times$$

$$
\begin{cases} 2^{-j}k-[n\tau], & \\ \qquad\qquad 2^{-j}k \le \tau < 2^{-j}(k+\frac{1}{2}) & \\ -(n2^{-j}(k+1)-[n\tau]), & \\ \qquad\qquad 2^{-j}(k+\frac{1}{2}) \le \tau < 2^{-j}(k+1). & \end{cases}
$$

for $\tau \in [0,1]$. This $q_{j,k}$ function is continuous and piecewise linear, shaped like an inverted "witch's hat" — it is zero outside of the support of the corresponding wavelet and goes to a peak at the midpoint of the support of $\psi_{j,k}$. This function gives some insight into how a single empirical Haar wavelet coefficient is at detecting a jump point: If a change occurs outside the support of $\psi_{j,k}$, it has no power to detect it; the power for detecting change is highest when the change occurs near the midpoint of the support of $\psi_{j,k}$. This agrees well with intuition.

Applying the above distributional result to the expression (3) for the prior distribution of τ, Δ, and σ^2 gives

$$
\begin{aligned} p(\tau,\Delta,\sigma^2|\underset{\sim}{w}) \quad &\propto \quad \sigma^{-(n-1)/2}\pi(\tau,\Delta,\sigma^2)\times \\ &\exp\left(-\sum_j\sum_k(w_{j,k}-\Delta q_{j,k}(\tau))^2/(2\sigma^2)\right) \end{aligned}
$$

Note that if Δ and σ^2 were known, then the mode of the posterior distribution (5) would come for the τ which minimizes the squared differences between the observations and their expectations. Figure 1, taken from Richwine (1996) illustrates this point for a simulated data set in which σ^2 and Δ are known to be one. The mode of the posterior distribution occurs for the value of τ for which the values of $q_{j,k}(\tau)$ and $w_{j,k}$ line up most closely. For this illustration, the change-point actually occurred at $\tau = 0.72$, so the values line up much more closely than for $\tau = 0.20$.

The analysis discussed thus far can be generalized by allowing for a smoother wavelet basis. In particular, the Daubechies wavelets adapted to the interval (see Cohen, Daubechies, and Vial (1993)) are used. These wavelets also provide an orthonormal basis for $L^2[0,1]$, so the expression (4) for the conditional distribution of the empirical coefficient $w_{j,k}$ given τ, Δ, and σ holds, with $q_{j,k}(\tau)$ defined to be

$$
q_{j,k}(\tau) = \int_\tau^1 \psi_{j,k}(t)\,dt.
$$

In practice, it is not necessary to compute this integral — the value of the $w_{j,k}(\tau)$'s for all j and k can be computed by applying the discrete wavelet transform algorithm to the vector $(0,\ldots,0,1,\ldots,1)'$ ($[n\tau]$ zeroes, $n-[n\tau]$ ones).

The analysis just described is appropriate for the standard change-point model, in which the function f is assumed to be piecewise constant with a single jump. One attractive feature of such an analysis is that it can be applied also to more general change-point problems, in particular, when the underlying function f is smooth except for a point of abrupt change.

To apply the analysis to this situation, recall that each empirical wavelet coefficient contains information about how the function is changing in a localized region (which is just the support of the corresponding wavelet function). Since a change-point is a localized phenomenon, it is natural to use only the more localized (higher level) coefficients in the computation of the posterior density, leaving the lower-level coefficients to describe the large-scale features of the data.

To be more precise, the raw function $f^{\sim}$ is first projected onto the approximation space V_{j_0} for some suitably chosen integer j_0. The posterior density is computed using only the wavelet coefficients with dilation index $j \ge j_0$. Thus, the projection gives a good smooth estimate of f, and the "residual" wavelet coefficients are analyzed to find the probable location of the change-point.

The Jeffrey's noninformative joint prior for τ, Δ, and σ is given by

$$
\pi(\tau,\Delta,\sigma^2) \propto \frac{1}{\sigma},
$$

so the joint posterior density is given by

$$
\begin{aligned} p(\tau,\Delta,\sigma^2|\underset{\sim}{w}) &\propto \sigma^{-(M+2)/2}\times \\ &\exp\{\frac{\sum_{j\ge j_0}\sum_k(w_{j,k}-\Delta q_{j,k}(\tau))^2}{2\sigma^2}\}, \end{aligned}
$$

where $M < n$ is the number of wavelet coefficients considered.

Integrating out Δ and σ gives the marginal posterior of the change-point τ to be

$$
p(\tau|\underset{\sim}{w}) \propto C^{-1/2}\left(A-\frac{B^2}{C}\right)^{-(M-2)/4},
$$

where $A = \sum_{j\ge j_0}\sum_k w_{j,k}^2$, $B = \sum_{j\ge j_0}\sum_k w_{j,k}q_{j,k}(\tau)$, and $C = \sum_{j\ge j_0}\sum_k q_{j,k}^2(\tau)$.

Figure 2 and Figure 3 show this method in action with simulated data. The upper left-hand plot in each figure is the raw time series; the upper right-hand plot is the true mean function, a sine function with a jump. In the bottom left-hand plot is the "removed trend" — the projection of f^wig onto the approximation space V_{j_0}. The marginal posterior density of τ (using only the higher level

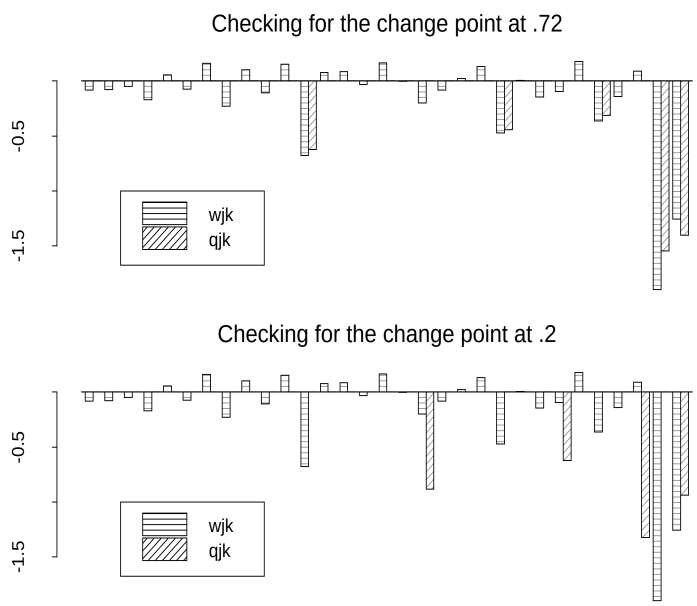

Figure 1: Comparing $w_{j,k}$'s with their expectations for two values of τ

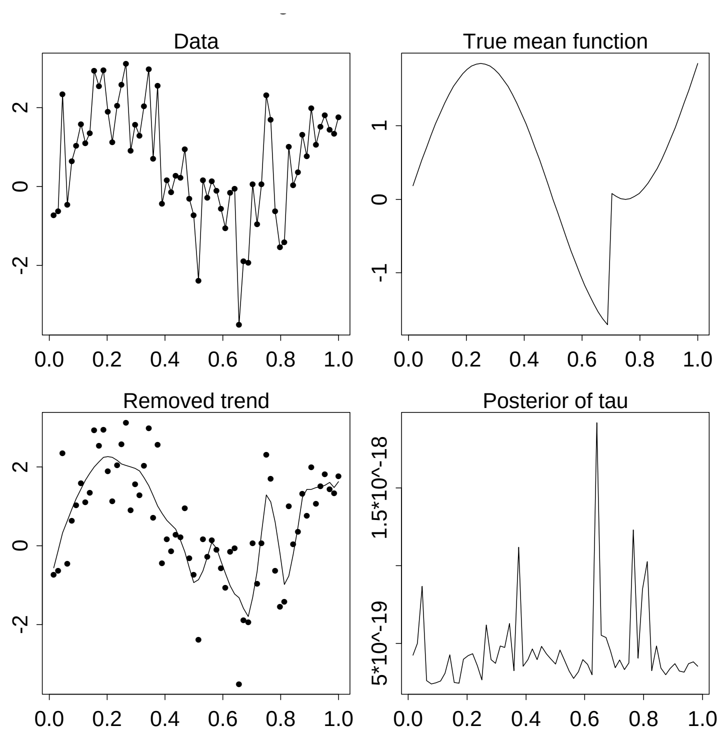

Figure 2: Simulated data with sine function with jump, signal-to-noise ratio 1

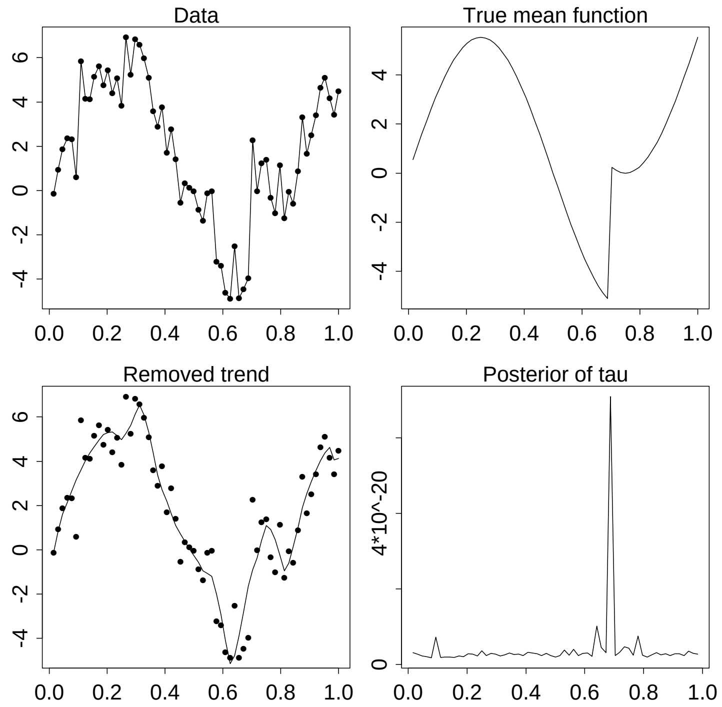

Figure 3: Simulated data with sine function with jump, signal-to-noise ratio 3

wavelet coefficients) is plotted in the lower right-hand plot of each figure. For these examples, only coefficients at the highest two levels of resolution were considered in computation of the posterior density ($16 + 32 = 48$ coefficients used; the remaining coefficients were applied to calculating the "removed trend"). Signal-to-noise ratio is defined here as the ratio of the standard deviation of the sequence $f(1/n), \ldots, f((n-1)/n), f(1)$ to the noise level σ.

From these simulated data examples, it is readily seen that the posterior marginal distribution can provide a valuable diagnostic tool for change-point analysis. Given a loss function, Bayes estimators can also be computed directly.

4 Conclusions

The greatest criticism for wavelet-based methods is their lack of *translation invariance* (see Coifman and Donoho (1995)). It is easy to see, for example, that using the Haar basis, it will be much easier to locate a change at, for instance $\tau = 1/2$, than it will be for a non-dyadic point such as $\tau = 1/3$. Work is currently underway by the author to adapt the analysis in this paper to get around this problem by applying the stationary wavelet transform as discussed by Nason and Silverman (1995).

Another practical issue is the choice of the integer j_0. Although the examples included in the text were constrained somewhat by software limitations, the problem of choosing this parameter needs to be addressed. Some data-based choice based on the total energy of empirical coefficients at each level might be appropriate.

References

Cohen, A., Daubechies, I., and Vial, P. (1993). Multiresolution analysis, wavelets and fast algorithms on an interval. *Applied Computational Harmonic Analysis* **1**: 54–81.

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*. Antoniadis, A., and Oppenheim, G. (eds.). New York: Springer-Verlag. pp. 125–150.

Csorgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, New York.

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

Haar, A. (1910). Zur Theorie der orthoganalen Funktionen-Systeme. *Annals of Mathematics*

69: 331–371.

Kaiser, G. (1994). *A Friendly Guide to Wavelets.* Birkhäuser, Boston.

Nason, G. P. and Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics.* Antoniadis, A., and Oppenheim, G. (eds.). New York: Springer-Verlag. pp. 125–150.

Ogden, R. T. (1996). *Essential Wavelets for Statistical Applications and Data Analysis.* Birkhäuser, Boston.

Ogden, R. T. (1997). On preconditioning the data for the wavelet transform when the sample size is not a power of two, to appear, *Communications in Statistics B: Simulation and Computation.*

Richwine, J. E. (1996). Bayesian estimation of change-points using Haar wavelets. Master's thesis, University of South Carolina Department of Statistics, Columbia, South Carolina.