MCMC INFERENCE OF THE SHAPE AND VARIABILITY OF TIME-RESPONSE SIGNALS

Dmitriy A. Katz-Rogozhnikov, Kush R. Varshney, Aleksandra Mojsilović, and Moninder Singh

Business Analytics and Mathematical Sciences Department, IBM Thomas J. Watson Research Center
1101 Kitchawan Rd., Route 134, Yorktown Heights, NY 10598, USA

ABSTRACT

Signals in response to time-localized events of a common phenomenon tend to exhibit a common shape, but with variable time scale, amplitude, and delay across trials in many domains. We develop a new formulation to learn the common shape and variables from noisy signal samples with a Bayesian signal model and a Markov chain Monte Carlo inference scheme involving Gibbs sampling and independent Metropolis-Hastings. Our experiments with generated and real-world data show that the algorithm is robust to missing data, outperforms the existing approaches and produces easily interpretable outputs.

Index Terms— time-response signal, multiple alignment, Markov chain Monte Carlo, outsourcing

1. INTRODUCTION

Responses over time to singular events are studied in many domains, for example, the electric potential in the brain of a subject who has been presented with a visual stimulus [1], the financial performance of a company that has agreed to an outsourcing engagement [2], and the vital signs of a patient who has been administered a medication.

As discussed in Sec. 1, the general signal model that we consider is given by (1). The particular noise we consider is zero-mean white Gaussian noise with variance $\sigma^2$. The whiteness of the noise in the model introduces statistical independencies among times which we take advantage of in inference. The particular form of the common shape $f(t)$ we consider is a piecewise polynomial spline with a fixed number of knots $m$. Let the times of the knots be $t_1, t_2, \ldots, t_m$, and the values of the common shape function at those times be $f_1, f_2, \ldots, f_m$. The value $f(t)$ at a time point $t$ other than the knots may be interpolated given $f_1, \ldots, f_m$. The fixed parameterization for $f(t)$ in terms of $f_1, \ldots, f_m$ permits us to develop a tractable inference procedure. The signal $r_i(t)$ is measured at $t_i$ (not necessarily uniformly spaced) and is denoted $r^{(i)}_1, \ldots, r^{(i)}_{t^{(i)}}$. The number of measured points and their times may be different for different trials, and different from the common shape function.

There are $3n + m$ random variables in the model we would like to infer, namely $(A_i, b_i, d_i)$, $i = 1, \ldots, n$ and $f_j, j = 1, \ldots, m$. Uninformative priors are taken for the $f_j$. The priors of the scale and shift variables, $p(A_i; \theta_A), p(b_i; \theta_B)$ and $p(d_i; \theta_d)$ with hyperpa-

2. HIERARCHICAL BAYESIAN SIGNAL MODEL

As discussed in Sec. 1, the general signal model that we consider is given by (1). The particular noise we consider is zero-mean white Gaussian noise with variance $\sigma^2$. The whiteness of the noise in the model introduces statistical independencies among times which we take advantage of in inference. The particular form of the common shape $f(t)$ we consider is a piecewise polynomial spline with a fixed number of knots $m$. Let the times of the knots be $t_1, t_2, \ldots, t_m$, and the values of the common shape function at those times be $f_1, f_2, \ldots, f_m$. The value $f(t)$ at a time point $t$ other than the knots may be interpolated given $f_1, \ldots, f_m$. The fixed parameterization for $f(t)$ in terms of $f_1, \ldots, f_m$ permits us to develop a tractable inference procedure. The signal $r_i(t)$ is measured at $t_i$ (not necessarily uniformly spaced) time points $t^{(i)}_1, \ldots, t^{(i)}_{t^{(i)}}$ and is denoted $r^{(i)}_1, \ldots, r^{(i)}_{t^{(i)}}$. The number of measured points and their times may be different for different trials, and different from the common shape function.

There are $3n + m$ random variables in the model we would like to infer, namely $(A_i, b_i, d_i)$, $i = 1, \ldots, n$ and $f_j, j = 1, \ldots, m$. Uninformative priors are taken for the $f_j$. The priors of the scale and shift variables, $p(A_i; \theta_A), p(b_i; \theta_B)$ and $p(d_i; \theta_d)$ with hyperpa-
parameters $\theta_A$, $\theta_b$ and $\theta_d$, are the same for all $i$ and are given later in the paper. A graphical model representation of the signal model is provided in Fig. 1. Plate notation is used in the figure to indicate repetition of the variables for each trial, $i = 1, \ldots, n$. Let $A$ denote all the variables $A_1, \ldots, A_n$, $b$ denote all the variables $b_1, \ldots, b_n$, and $d$ denote all the variables $d_1, \ldots, d_n$.

Due to the additive Gaussian noise, the conditional pdf of $f_j$, 

$$p\left(f_j | A, b, d, f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_m, r_1^{(1)}, \ldots, r_n^{(n)}\right),$$

is Gaussian. The mean and variance of this Gaussian may be derived based on spline interpolation formulas. For example with a piecewise linear $f(t)$, letting

$$\tau_{ij} = \{k \mid t_{j-1} < b_k + d_i < t_j\},$$
$$\tau_{i} = \{k \mid t < b_k + d_i \leq t+1\},$$
$$\tau_{ij}^+ = \{k \mid t < b_k + d_i < t+1\},$$
$$\sum_{i=1}^{n} (A_j - A_i f_{j-1} + \tau_{j} - A_j f_{j+1} + \tau_{j+1}),$$

and the mean is

$$\sum_{i=1}^{n} (\tau_{ij} A_j - A_i f_{j-1} + \tau_{ij} A_j - A_i f_{j+1} + \tau_{ij+1}),$$

and the variance is

$$\sum_{i=1}^{n} (\tau_{ij} A_j - A_i f_{j-1} + \tau_{ij} A_j - A_i f_{j+1} + \tau_{ij+1}).$$

With a linear spline, $f_j$ depends on $f_{j-1}$ and $f_{j+1}$, but not the other values of the shape function; this may or may not be true with other interpolating splines.

$A_j$ depends on the measurements, $b_i$ and $d_i$ of trial $i$ as well as the common $f_j$ values, but not on the measurements, $b_i$ and $d_i$ of any other trial $i' \neq i$. Similarly for $b_i$ and $d_i$. The conditional densities of the scale and shift variables, due to the white Gaussian noise, are:

$$p\left(A_j | b_i, d_i, f_1, \ldots, f_m, r_1^{(1)}, \ldots, r_n^{(n)}\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{n} (r_k^{(i)} - A_i f(b_k^{(i)} + d_i))^2\right).$$

In (4)–(6), the $f(b_t^{(i)} + d_i)$ terms involve spline interpolation.

### 3. MCMC Sampling

The new hierarchical Bayesian signal model set forth in the previous section factors in such a way (illustrated in Fig. 1) that lends itself to inference via Gibbs sampling [6]. Each of the $3n + m$ random variables is sampled in turn, conditioned on all of the other variables and the measured time-response signals.

The $f_j$ are sampled within a Gibbs iteration according to

$$p\left(f_j | A, b, d, f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_m, r_1^{(1)}, \ldots, r_n^{(n)}\right),$$

which are Gaussians with means and variances discussed in the previous section.

The $A_i, b_i$ and $d_i$ are difficult to sample directly due to the spline interpolation. For these variables, we use independent Metropolis-Hastings sampling [7, 8]. That is, the variable is drawn independently from its prior distribution and then accepted or rejected according to the conditional densities (4)–(6). For example, at the current Gibbs iteration, a candidate amplitude $A_i^{(\text{candidate})}$ is drawn from $p(A_i | \theta_A)$, $A_i^{(\text{candidate})}$ is accepted as $A_i^{(\text{current})}$ with probability $\min\{R, 1\}$, and $A_i^{(\text{previous})}$ is taken as $A_i^{(\text{current})}$ with probability $\max\{1 - R, 0\}$, where $R$ is the Hastings ratio

$$\exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} (r_k^{(i)} - A_i^{(\text{candidate})} f(b_k^{(i)} + d_i))^2\right).$$

Since only the ratio of two conditional densities is needed, the normalizing constant of the conditional pdf is not required. Independent Metropolis-Hastings is less commonly used than perturbation-based and other flavors of Metropolis-Hastings, but in the overall problem at hand, we observe that the independent flavor works well, with short mixing times and without getting ‘stuck.’

### 4. Results on Generated Data

In this section, we illustrate the proposed signal processing technique on data generated from the model (1). Specifically, we use $A_i$ that are uniform over the interval $[0.5, 1.5]$, $b_i$ that are uniform over the interval $[0.5, 1]$, and $d_i$ that are uniform over the interval $[-3, 3]$. The common shape is $f(t) = \left(t^3/10 - t^2/2 - t/2\right)e^{-t/4}$, the zero-mean additive white Gaussian noise has variance $\sigma^2 = 1$, and measurements are taken for all $n$ trials at times $t = 1, 2, \ldots, 16$. Note that in generating the data, $f(t)$ is not a spline; however, we use a piecewise linear $f(t)$ in the MCMC inference. One realization of the data for $n = 100$ is shown in Fig. 2(a).
Fig. 2. Generated data (a) response signals, (b) aligned with [4], (c) aligned with proposed MCMC inference, and (d) the averages of (a) in green, (b) in cyan, and (c) in red along with the true $f(t)$ in blue.

Table 1. Mean-squared error between true $f$ and estimate for different $n$ averaged over 50 instances of data with standard deviation over the instances in parentheses.

<table>
<thead>
<tr>
<th>$n$</th>
<th>simple avg</th>
<th>Listgarten</th>
<th>HB MCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>2.11 (0.63)</td>
<td>1.67 (0.60)</td>
<td>1.24 (1.28)</td>
</tr>
<tr>
<td>100</td>
<td>2.05 (0.42)</td>
<td>1.58 (0.44)</td>
<td>0.51 (0.41)</td>
</tr>
<tr>
<td>200</td>
<td>2.00 (0.32)</td>
<td>1.51 (0.38)</td>
<td>0.37 (0.24)</td>
</tr>
<tr>
<td>400</td>
<td>1.98 (0.27)</td>
<td>1.53 (0.33)</td>
<td>0.24 (0.14)</td>
</tr>
</tbody>
</table>

Table 2. Mean-squared error between true $f$ and estimate for different missing data percentages at $n = 100$ averaged over 50 instances of data with standard deviation over the instances in parentheses.

<table>
<thead>
<tr>
<th>$f$</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>simple avg</td>
<td>2.05 (0.42)</td>
<td>2.06 (0.43)</td>
<td>2.08 (0.42)</td>
<td>2.07 (0.44)</td>
</tr>
<tr>
<td>Listgarten</td>
<td>1.58 (0.44)</td>
<td>1.53 (0.41)</td>
<td>1.54 (0.45)</td>
<td>0.61 (0.52)</td>
</tr>
<tr>
<td>HB MCMC</td>
<td>0.51 (0.41)</td>
<td>0.53 (0.41)</td>
<td>0.54 (0.45)</td>
<td>0.61 (0.52)</td>
</tr>
</tbody>
</table>

5. RESULTS ON REAL-WORLD DATA

In this section we consider the use of our method in a business analytics application. Corporate leaders are often interested in quantifying an impact of a major initiative to company performance. One such initiative is outsourcing, where an external vendor manages a portion of the company’s operations, in order to reduce expenses and increase earnings. Therefore, financial performance metric time signals, such as selling, general and administrative expenses (SG&A) growth rate, and earnings before tax (EBT) growth rate, are expected to respond to a company entering into an outsourcing engagement.

We use the proposed signal model and inference methodology to understand the impact of outsourcing on the future performance of companies.

We expect outsourcing to change the growth rate of SG&A and EBT for a period of time, with the impact taking a common shape across companies. We also expect the amplitude of the impact to be different and for the impact to occur on different time scales for different companies. In addition, the data available for outsourcing events is the date at which the outsourcing engagement deal was signed, but not when the outsourcing was actually rolled out. It is thus important to include time delay parameters in the signal model. Performance metric measurements are quite noisy; the white Gaussian model is not inappropriate in this setting and is applied because of its mathematical convenience.

We use a uniform distribution over the interval $[0, 1]$ as $p(A_i; \theta_A)$ and a uniform distribution over the interval $[1/2, 1]$ as $p(b_i; \theta_b)$. The delay prior is the following:

$$p(d_i; \theta_d) = \begin{cases} 1/2, & 0 < d_i \leq 1 \\ -d_i/4 + 3/4, & 1 < d_i \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

where time is measured in quarters of years. The $p(d_i; \theta_d)$ distribution is easily sampled using Smirnov transformation of uniform random numbers.

We show results of MCMC inference for SG&A on $n = 249$ companies in Fig. 3 and for EBT on $n = 216$ companies in Fig. 4. Companies whose available response signals contain missing values are discarded to allow comparison with [4], which is why $n$ is different for the two financial metrics. Subfigure (a) shows the noisy signal samples and (b) shows the result of taking a simple average over the $n$ signals at each time. Subfigure (c) shows the latent trace obtained using [4], whereas (d) shows the median values of 500 MCMC samples of $f$ from iterations 2501 to 3000.

For SG&A, we see that the simple average, the latent trace, and the median common shape $f(t)$ are all similar looking, notably decreasing for about four quarters after $t = 0$. Expenses do decrease as a result of outsourcing. Earnings are also impacted by outsourcing, but less directly and with some delay. The simple average produces a fairly noisy signal for EBT while the latent trace and median common shape are smoother. In all three plots, we see that after $t = 0$, the EBT growth rate initially decreases for a couple of quarters and then increases for a few quarters. The simple average and the median common shape are more in line with each other than the latent trace of [4], which is a result of [4] incorporating nonuniform scalings of time in order to align signals. In the latent trace, the initial decrease in EBT growth rate after $t = 0$ is stretched out in comparison with other parts of the signal. In understanding the business impact of outsourcing, such nonuniformities make interpretation difficult.
Another aspect of the hierarchical Bayes model is that we have access to the highly interpretable \(A, b, d\), and \(\theta\) variables. The figures show histograms over the \(n\) companies of median \(A, b, d\), and \(\theta\), illustrating the variability of these parameters. Interestingly, the distribution of the delay parameter \(d\) is approximately the same for both SG&A and EBT, centered around one quarter after the signing of the outsourcing deal; this indicates that the delay parameter truly is capturing the time between the signing of the deal and the roll out. The amplitude distributions are different, which is to be expected because outsourcing has different relative effects on SG&A and EBT for different companies.

6. CONCLUSION

In this paper, we have developed a Bayesian inference methodology for determining the common structure and variability of a collection of time-response signals. It is shown to work well on generated and real-world data, to be robust to missing data, and to produce easily interpretable outputs including uniform time scalings.

That we see a common delay distribution between SG&A and EBT is capturing the time between the signing of the deal and the roll out. The amplitude distributions are different, which is to be expected because outsourcing has different relative effects on SG&A and EBT for different companies.

of the model includes adding one further layer of hierarchy, by not taking \(\theta A, \theta b,\) and \(\theta d\) to be fixed hyperparameters, but to be random variables that must also be inferred. In such an extended model, we could understand the variability of \(A, b,\) and \(d\) in a more rigorous fashion than examining histograms such as those in Fig. 3 and Fig. 4.

7. REFERENCES