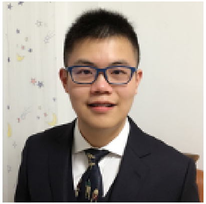
Bayesian Detection and Prediction of Disruptive Events using Twitter

Theo Kypraios

School of Mathematical Sciences @ University of Nottingham

University of Crete – March, 2024

Acknowledgements



Jiale Tao
University of Nottingham

Karthik Bharath
University of Nottingham

University of Nottingham UK | CHINA | MALAYSIA

Table of contents

Motivation

Knowledge Transfer Partnerships (KTPs)

- A KTP is a three-way project between a company, a university, and a recent graduate, where the business is able to utilise academic expertise in order to develop a new technology or improve products and processes.
- Great opportunity for all three parties involved.
- KTP with Delta Rail (now Resonate) back in 2016 [Ian Dryden, David Hodge].
- Delta Rail / Resonate is a technology company specialising in rail and connected transport solutions. Offices in Derby (about 20 km from UoN campus).

Company's Brief

Resonate's brief

Explore a "link" between **disruption** reported by customers and that measured by the railway .... In other words, can we **identify** a major **disruption** based on what people are saying on Twitter?

Available Data

Any tweet that mentioned/tagged a railway company from 01/01/2015 to 31/12/2015:

Exploratory Data Analysis

Raw Data: 2.2M Tweets

cleaned_tweets2015

	A	B
1	21/01/2015 17:26	@witheta @SouthernRailUK Not a damn chance. I have annual season from ELD and can't see them offering diddly squat.
2	21/01/2015 17:26	Just asked if my early train would definitely depart before the later one and was told 'It's a gamble' @SouthernRailUK #LondonBridgeTrains
3	21/01/2015 17:27	RT @FJGSolicitors: .@VirginTrains to the rescue - the power of social http://t.co/5MvBt5wgmg http://t.co/Vqtc4eVvSI
4	21/01/2015 17:27	@eastcoastuk Yes they were empathetic (chilly) but it's out of their hands... An ongoing issue? Someone said staff wearing coats yesterday!
5	21/01/2015 17:27	@TfLTravelAlerts @districtline @DistrictLame 2/2 at Victoria station
6	21/01/2015 17:27	@greateranglia @Sasbo23 GK is basically saying they have no spares as they sent back 36 trains to save cash when they arrived
7	21/01/2015 17:27	@LondonMidland thanksss ,úåÔ∏è
8	21/01/2015 17:27	RT @jodrell: @richard_je @greateranglia old and unpleasant sounds about right :-(
9	21/01/2015 17:28	Twitter would I get in trouble if I dialled 999 and reported @SW_Trains for crimes of extortionate prices? It'll be worth it #swtrains
10	21/01/2015 17:29	RT @sussmes88: @greateranglia 17.11 Brentwood to Shenfield - black hat found and passed to staff at Shenfield.
11	21/01/2015 17:29	@TLRailUK @SouthernRailUK just but explaining why that train always crawls along and is always late. Never any announcements on board. Ta!
12	21/01/2015 17:29	@MWilliamsWBA @LondonMidland my M.P replied and will raise their appalling service with the minister for transport on my behalf
13	21/01/2015 17:29	@Dombelina @SouthernRailUK at least they were honest!
14	21/01/2015 17:29	Are @Se_Railway planning on running trains to charring cross and Waterloo at the weekends or is this the new norm #blackheath #WeekendTravel
15	21/01/2015 17:29	Tube from Paddington to Kings Cross is a joke. Like a third world service. Reflects so badly on London from a prime station @TfL
16	21/01/2015 17:29	@SW_Trains do you ever plan on making trains more frequent on your very overcrowded lines? Like the reading line? It's getting tedious.
17	21/01/2015 17:29	@SouthernRailUK @panthrosighalot very clever 'I believe' not 'it will be'. Get out clause released!
18	21/01/2015 17:29	@greateranglia complete robbery !! you add on an extra ¬£100 to season toys knowing you won't be running trains for every wkd in 2 months
19	21/01/2015 17:29	Standing on the train out of London is always good fun @SouthernRailUK #LondonBridgeTrains
20	21/01/2015 17:29	@SouthernRailUK @AlexBakerface has been seen on one of your trains. Expect delays...don't ask me why!
21	21/01/2015 17:29	Thanks Grant @fgw. Looks like we're stuck with crappy old tugs for 2+ years. Then probably more delays when electrification is installed.
22	21/01/2015 17:29	@demiii_ @c2c_rail2 @c2c_Rail I don't even know what that means!
23	21/01/2015 17:29	@greateranglia can you turn the heating on please - 1730 to Gidea Park from Liverpool St carriage number 64568
24	21/01/2015 17:30	RT @sussmes88: @greateranglia 17.11 Brentwood to Shenfield - black hat found and passed to staff at Shenfield.

Exploratory Data Analysis

Not all tweets refer to disruption (in a broad sense) and/or tag railway company:

- `Hi Sophie. We are experiencing delays of up to an hour on this route due to an accident on the A259. @SouthernRailUK`
- `@TLRailUK I'm pretty sure you didn't. Selfish train company. Only out to rip me off and make me late either to work or from work.`
- `Why would it make it go any faster? Just passing Maidenhead 4 mins before due in Reading. Not an excessive delay yet.`
- `Excellent journeys with @eastcoastuk on Monday. Great service good food very pleasant. Thank you.`

Text $\rightarrow$ Counts

Data pre-processing:

1. Identify the tweets that refer to "delay" using related words; [off-the-self sentiment analysis / classification];
2. divide the day (24 hours) in fixed intervals (e.g. 15-minutes);
3. count the number of "delay tweets" per interval.

There are many ways to pre-process the data (especially step 1 above) – this is **not** the focus of this talk.

Our starting point is the temporal count data obtained at the end of step 3 and we treat these as our observed data ($\mathbf{x}$).

Curating the Data

We split the 24 hours of a calendar day into 96 15-minutes intervals:

Tweet	Day	Time Interval	Company	"Delay"
selfish ... make me late ...	Mon 03/05	08:30-08:45	Thames Link	1
... experiencing delays up to ...	Tue 08/12	16:30-16:45	Southern Rail	1
... great service ...	Fri 17/02	18:30-18:45	East Coast	0
⋮	⋮	⋮	⋮	⋮

Notation

Denote by $x^j_{d,i}$ the number of "delay" tweets in the i_{th} interval on a given day d and which refer to / tag company j, where $i = 1, \dots, 96$, $d \in \{01/01, \dots, 31/12\}$ and $j = 1, \dots, m$.

Modelling

Start Simple: Aggregate over companies

- We aggregate over the different companies and consider all the tweets which refer to a given day of the week, e.g. Monday.
- We end up with a discrete-value time series of length $T = 52 \times 96$:

$$\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{52})$$

  where $\boldsymbol{x}_k = (x_{k,1}, \ldots, \boldsymbol{x}_{k,96})$ and $k = 1, \ldots, 52$.
- We denote by $x_{k,i}$ the number of “delay” tweets in the i_{th} 15-minute interval (irrespective of which railway company is tagged) which was tweeted on the k_{th} Monday in the dataset.

A Bayesian Hierarchical Model

We assume that the number of "delay" tweets at a given time-interval tell us something about the (unobserved) state of the railway.

We wish to build a hierarchical model to learn the patterns in the observed data ...

... both when there is a major disruption in the railway and when there is not.

The model should reflect the nature of the data: a combination of the normal pattern and occasional additional counts caused by disruptive events.

Our Focus

Predicting the state of the railway at a given interval having observed the number of "delay" tweets.

It is Wednesday 08:30.

Between 08:15-08:30 there have been 60 tweets mentioning "delay".

Is 60 an unusually large number of tweets for that time on a Wednesday or something to be expected?

Modelling the State of the Railway

Assumption:

The state of the railway during a fixed time interval is either "normal" or "disrupted"; the latter corresponds to intervals where there is a major disruptive event and the former where there is not.

We model this behaviour using a Markov chain in discrete time.

Let $\{Z_t,\ t = 0, 1, \dots\}$ be an irreducible homogeneous Markov chain on the state space $\mathcal{S} = \{0, 1\}$ with transition probability matrix $M_Z^\star$; that is

$$m_{ij}^\star = P(Z_t = j \mid Z_{t-1} = i), \quad \text{for} \quad i = 0, 1,$$
$$\text{where} \quad 0 = \text{“normal”} \quad \text{and} \quad 1 = \text{“disrupted”}.$$

Modelling the Volume of Tweets

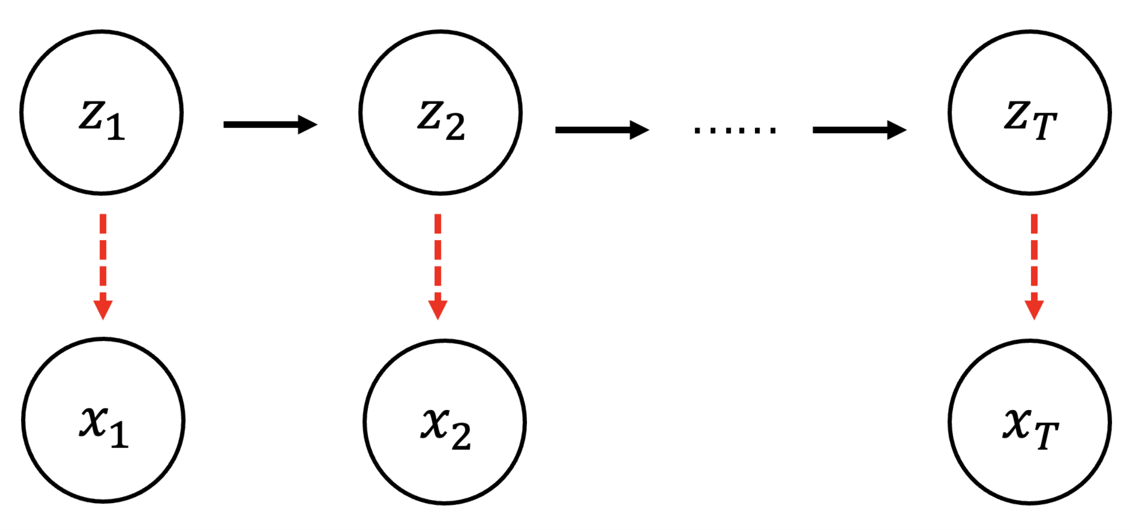
Let the non-negative integer-valued random process $\{X_t,\ t = 0, 1, \ldots\}$ represent the number of tweets posted at time t which are decomposed as follows:

$$X_t = \begin{cases} Y_t, & \text{if} \quad Z_t = 0; \\ Y_t + W, & \text{if} \quad Z_t = 1; \end{cases} \qquad t = 1, 2, \ldots, T. \tag{1}$$

We further assume that

- $Y_t | Z_t \overset{\text{iid}}{\sim} Po(\lambda(t))$, i.e. conditional on $\{Z_t,\ t = 1, \ldots, T\}$, $Y_1, \ldots, Y_T$ are assumed to be mutually independent.
- $W \sim Po(c)$ and independent of $\{Y_t,\ t = 1, 2, \ldots, T\}$.

Graphical Model



MMPP / HMM

The model can be either viewed a discretised version of a Markov Modulated Poisson Process (MMPP) or a Hidden Markov Model (HMM).

- the number of "delay" tweets we expect to see at time t (Y_t) follows a Poisson distribution with a time-dependent mean $\lambda(t)$ if the state of the railway is "normal";
- if there is a major disruptive event at time t then there is an additional number of "delay" tweets (W) that follows a Poisson distribution with mean c.

How to Model $\lambda(t)$?

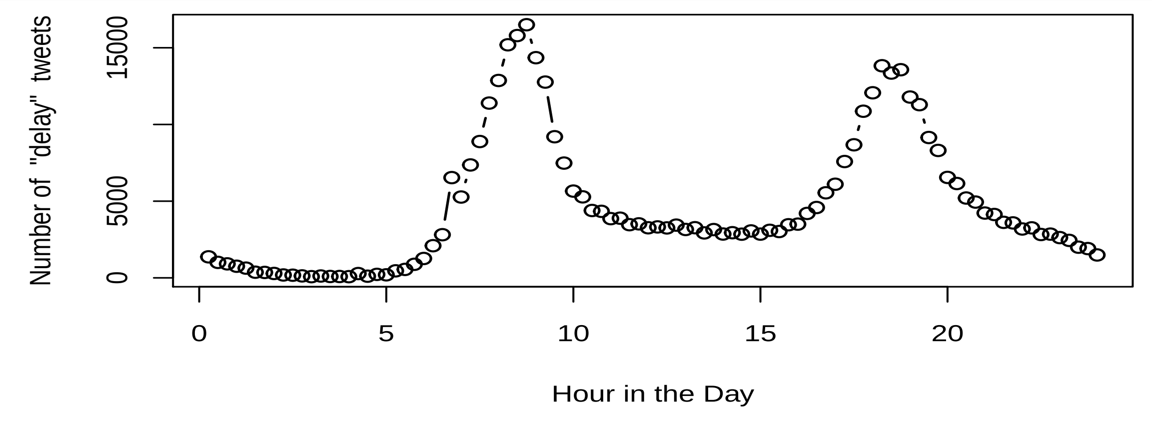


Figure 1: Frequency of "delay" tweets throughout the day

Modelling the Intensity

Figure 1 suggests that the average number of "delay" tweets, $\lambda(t)$, can be modelled with the following parametric function:

$$\lambda(t) = c_1 \cdot \exp\left\{-\frac{1}{2\sigma_1^2}(t-\mu_1)^2\right\} + c_2 \cdot \exp\left\{-\frac{1}{2\sigma_2^2}(t-\mu_2)^2\right\} + c_3, \tag{2}$$

The parameters μ_1 and μ_2 represent the times during the morning and afternoon/evening rush hours at which the number of tweets is maximum;

The rest parameters of the function $\lambda(t)$, c_1, c_2, and c_3 are scaling constants.

Bayesian Learning

Bayesian Inference

Set-up:

- observed data: $\boldsymbol{x} = (x_1, \dots, x_T)$;
- model parameters: $\boldsymbol{\theta} = (\mu_1, \mu_2, c_1, c_2, c_3, c)$;
- observed-data likelihood: $\pi(x_1, \dots, x_T | \boldsymbol{\theta})$

Observed-data likelihood, $\pi(x_1, \dots, x_T | \boldsymbol{\theta})$, not readily available because $\boldsymbol{z} = (z_1, \dots, z_T)$ is unobserved and hence will have to be integrated out:

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{z}} (\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\theta}) \,.$$

The sum becomes computationally infeasible for large T.

Bayesian Inference: Data Augmentation

We adopt a Bayesian framework and treat the unobserved states of the Markov chain $\{Z_t\}$ additional parameters to be inferred alongside the model parameters $\boldsymbol{\theta}$.

$$
\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{x}) &\propto \pi(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \\
&\propto \pi(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\theta})\,\pi(\boldsymbol{z}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}) \\
&\propto \pi(\mu_1, \mu_2)\,\pi(m^\star_{00}, m^\star_{01})\,\pi(m^\star_{10}, m^\star_{11})\,\pi(c)\,\pi(c_1)\,\pi(c_2)\,\pi(c_3) \\
&\times \prod_{t=1}^{T} P(Z_t = z_t|Z_{t-1} = z_{t-1}, \boldsymbol{m}^\star) \times \prod_{t=1}^{T} P(X_t = x_t|Z_t = z_t, \boldsymbol{c}, \boldsymbol{\mu}).
\end{aligned}
$$

Bayesian Inference: Augmented Likelihood

Having augmented the observed data $\boldsymbol{x}$ with the unobserved state of the railway ($\boldsymbol{z}$), the augmented likelihood

$$\prod_{t=1}^{T} P(X_t = x_t | Z_t = z_t, \boldsymbol{c}, \boldsymbol{\mu})$$

is straightforward to compute.

Each term in the product is the probability mass function from a Poisson distribution. If the state of the railway is “normal” that is, a Poisson($\lambda(t)$); if “disrupted” then it would be a Poisson($\lambda(t) + c$).

Next step: Prior specification for $\pi(\boldsymbol{\theta})$.

Bayesian Inference: Prior on parameters governing $M^\star$

four parameters of the transition probability matrix $M_z^\star$, namely $m_{00}^\star, m_{01}^\star, m_{10}^\star$ and $m_{00}^\star$,

$$M^\star = \begin{bmatrix} m_{00}^\star & m_{01}^\star \\ m_{10}^\star & m_{11}^\star \end{bmatrix}.$$

We assign Dirichlet distributions to these probabilities preserving the constraints that $m_{01}^\star + m_{00}^\star = 1$ and $m_{10}^\star + m_{11}^\star = 1$.

$$(m_{00}^\star, m_{01}^\star) \sim \mathrm{Dir}(\alpha_1, \alpha_2)$$
$$(m_{10}^\star, m_{11}^\star) \sim \mathrm{Dir}(\alpha_3, \alpha_4)$$

where $\alpha_1, \ldots, \alpha_4$ are further hyper-parameters.

Bayesian Inference: Prior on parameters governing $\lambda(t)$

Based on the interpretation of the parameters μ_1 and μ_2, we assume that a-priori are distributed as the ordered statistics of two *independent* Uniform random variables in (0,24):

$$\pi(\mu_1, \mu_2) = \frac{1}{288} \mathbb{1}\{0 \leq \mu_1 \leq 24\} \mathbb{1}\{0 \leq \mu_2 \leq 24\} \mathbb{1}\{\mu_1 < \mu_2\},$$

The first two terms come from the Uniform distribution and the last one comes from the order statistics since we require $\mu_1 < \mu_2$.

We further assign that $c \sim \text{Exp}(\beta)$ and independent Uniform priors on c_1, c_2 and c_3: $c_1, c_2 \sim U(0, 150)$ and $c_3 \sim U(0, 50)$.

Sampling from the Posterior distribution $\pi(\boldsymbol{\theta}, z|x)$

We develop a bespoke MCMC algorithm to sample from $\pi(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{x})$:

1. `initialisation; repeat steps 2-7 until convergence:`
2. `Sample from` $\pi(m^\star_{00}, m^\star_{01}|\boldsymbol{z})$ `directly;`
3. `Sample from` $\pi(m^\star_{10}, m^\star_{11}|\boldsymbol{z})$ `directly;`
4. `Sample from` $\pi(c|\boldsymbol{x}, c_1, c_2, c_3)$ `using Metropolis-Hastings;`
5. `Sample from` $\pi(\mu_1, \mu_2|\boldsymbol{x}, \boldsymbol{z}, c_1, c_2, c_3)$ `using Metropolis-Hastings;`
6. `Sample from` $\pi(c_1, c_2, c_3|\boldsymbol{x}, \boldsymbol{z}, \mu_1, \mu_2)$ `using Metropolis-Hastings;`
7. `Sample from` $\pi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\theta})$ `using the Forward-Backward algorithm.`

MCMC

A lot of work has gone into the development of the MCMC algorithm.

Focus on Prediction

Our main focus is to predict the state of the railway at a particular time-interval t_k when we have observed data *up to and including* time interval t_k; that is, deriving the posterior distribution $\pi(Z_k = s|\boldsymbol{x}_{1:k})$:

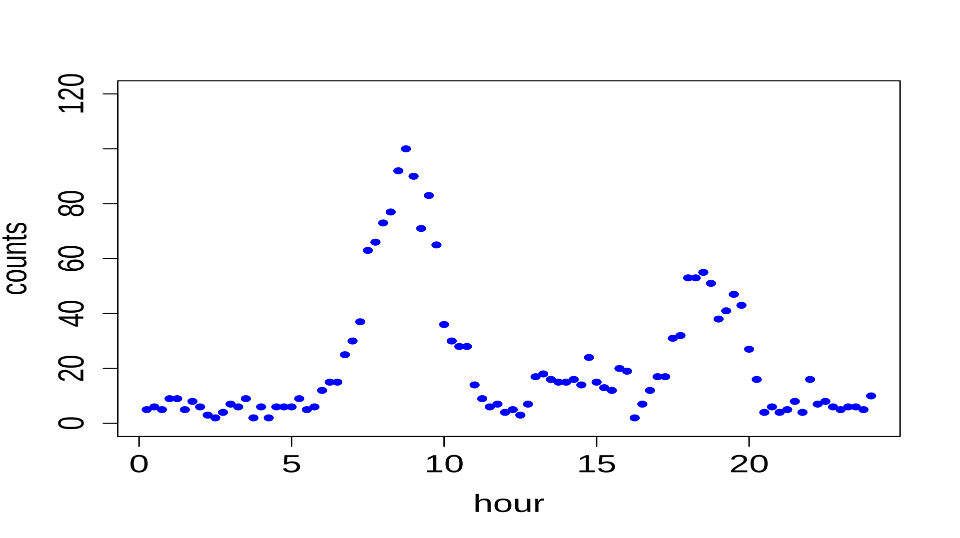
$$\pi(Z_k = s|\boldsymbol{x}_{1:k}) = \frac{\pi(x_k|Z_k = s)\,\pi(Z_k = s|\boldsymbol{x}_{1:(k-1)})}{\pi(x_k|\boldsymbol{x}_{1:(k-1)})} \propto \pi(x_k|Z_k = s)\,\pi\left(Z_k = s|\boldsymbol{x}_{1:(k-1)}\right)$$

Following some manipulation of the formulae we obtain the following recursion which enables us predict Z_k when the new data x_k comes if the model has been fitted up to time x_{k-1}

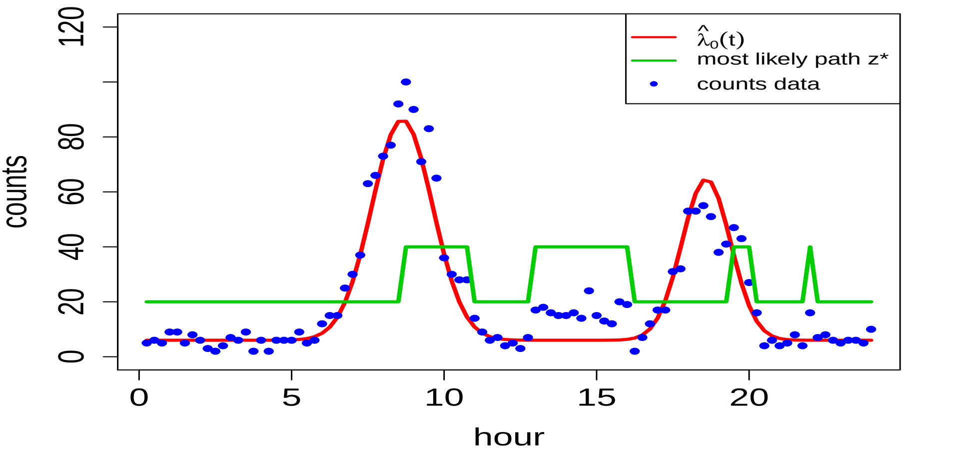
$$p(Z_k = s|\boldsymbol{x}_{1:k}) \propto p(x_k|Z_k = s) \sum_{l \in \{1,2\}} \pi(Z_k = s|z_{k-1} = l)\,\pi(Z_{k-1} = l|\boldsymbol{x}_{1:(k-1)}) \tag{3}$$

Does this really work?

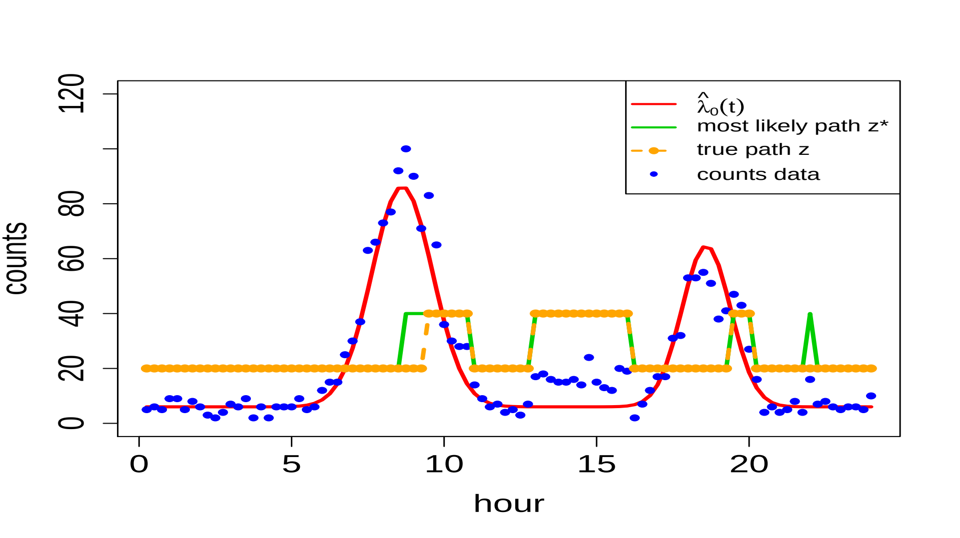
Synthetic Data



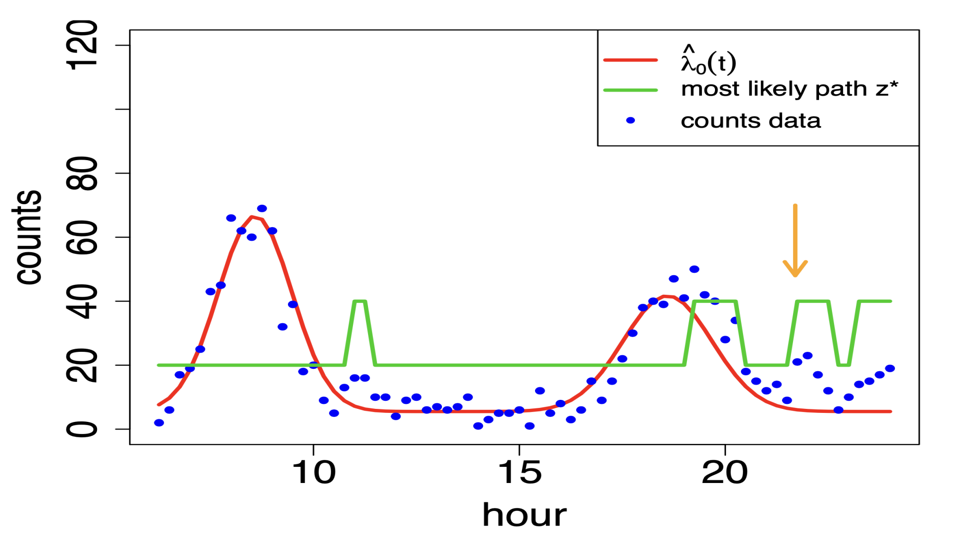
Synthetic Data



Synthetic Data



Incident on the 17th February 2015 – Great Western Railway



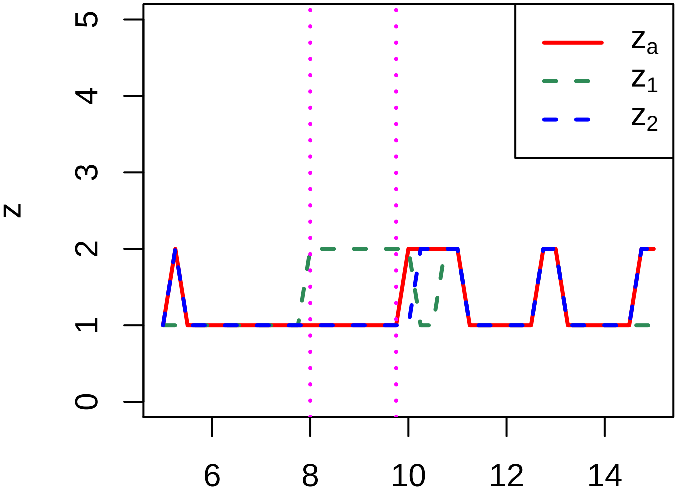
Reports of person being hit by a train around 20:58.

Model Extension

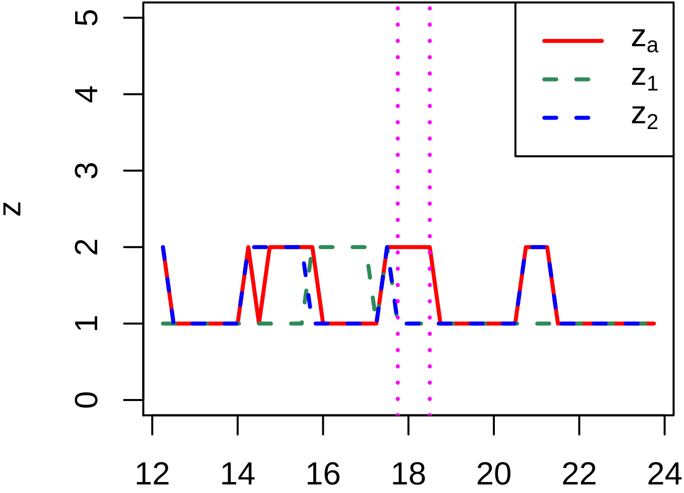
Capitalising on the Data's Rich Structure

- We wish to capitalise on the rich structure that the data has.
- One approach is to consider data (tweets/counts) from different companies separately.
- We can then fit/train our model to data from each company and make predictions.
- By doing so, we ignore the dependence between companies.

False Negative



False Positive



Capturing the Dependence Structure

So far we have considered the “state of the railway” in a broad sense without making any distinction between different companies.

However, there could be a major disruptive event that affects only one railway company whilst the rest are operating as normal.

Alternatively, there could be a major disruptive that affects more than one railway company.

We are now interested in answering the following questions:

- Can we predict the status of different railway companies *simultaneously*?
- Can we measure the extend to which one company’s state affects the state of another?

Bivariate MMPP / Coupled HMM

We want to model the state of the railway jointly for these two companies and consider a bivariate Markov Chain in discrete-time:

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}^{(1)} \\ \boldsymbol{Z}^{(2)} \end{pmatrix} = \begin{pmatrix} Z_1^{(1)}, Z_2^{(1)}, \ldots, Z_T^{(1)} \\ Z_1^{(2)}, Z_2^{(2)}, \ldots, Z_T^{(2)} \end{pmatrix}$$

The observed data are also considered jointly

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}^{(1)} \\ \boldsymbol{x}^{(2)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)}, x_2^{(1)}, \ldots, x_T^{(1)} \\ x_1^{(2)}, x_2^{(2)}, \ldots, x_T^{(2)} \end{pmatrix}$$

Model Assumptions

- The future states only on the current states

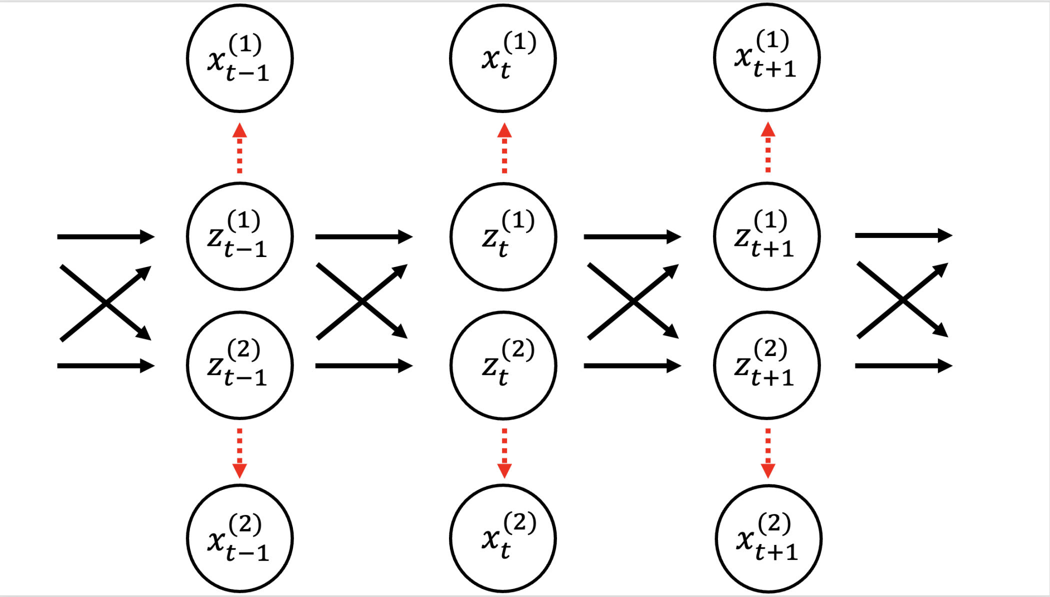
$$\boldsymbol{Z}_{t+1} \perp\!\!\!\perp \boldsymbol{Z}_{1:(t-1)} \mid \boldsymbol{Z}_{\boldsymbol{t}}.$$

- Conditional independence: $\pi\left(X_{1,t}, X_{2,t} \mid Z_{1,t}, Z_{2,t}\right) = \pi\left(X_{1,t} \mid Z_{1,t}\right) \times \pi\left(X_{2,t} \mid Z_{2,t}\right)$
- The model on $\pi\left(X_{j,i} \mid Z_{j,i}\right)$ for $j = 1, 2$ is the same as in the univariate case:

$$X_{j,t} = \begin{cases} Y_{j,t}, & \text{if} \quad Z_{j,t} = 1; \\ Y_{j,t} + W, & \text{if} \quad Z_{j,t} = 2; \end{cases} \qquad t = 1, 2, \ldots, T \quad \text{and} \quad j = 1, 2.$$

where $Yj, t \sim \text{Poisson}(\lambda(t))$ and $W \sim \text{Poisson}(c)$.

Bivariate MMPP / Coupled HMM



Modelling the Bivariate Markov Chain

- Denote the discrete state space of $\boldsymbol{Z}^{(1)}$ and $\boldsymbol{Z}^{(2)}$ to be $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively, i.e. $\mathcal{S}_1 = \mathcal{S}_2 = \{\text{normal}, \text{disrupted}\}$
- Denote that the cardinality of $\mathcal{S}_1$ and $\mathcal{S}_2$ are d_1 and d_2 $(d_1, d_2 \in \mathbb{N})$.
- One can construct an auxiliary univariate Markov chain $\{Z_i^\star\}_{t=1}^{T}$ with state space $\mathcal{S}^\star$ with cardinality $d_1 \times d_2$. Since $Z_t^\star$ and $\boldsymbol{Z}_t$ share the same cardinality, one can find a bijective mapping between them and have a single Markov Chain in the extended state space:

$$\mathcal{S}^\star = \{\{\text{normal}, \text{disrupted}\}, \{\text{normal}, \text{normal}\}, \\ \{\text{disruptive}, \text{normal}\}, \{\text{disrupted}, \text{disrupted}\}\}$$

Modelling the Bivariate Markov Chain

The transition probability matrix $M^\star$ is of dimension 4×4 and has $4 \times 3 = 12$ unknown parameters to be estimated from the data.

In general, if we have m companies, then $M^\star$ will be of dimension $2^m \times 2^m$ with $2^m \times (2^m - 1)$ unknown parameters to be estimated.

For instance, if $m = 4$ then we have 240 parameters to be estimated.

If $m = 8$ we have 65,280 parameters.

In addition to the parameter growing rapidly, we can't say anything about the dependence between companies.

Bayesian Inference for the Bivariate MMPP / Coupled HMM

- Given the observed data $\boldsymbol{x} = (x_1, x_2)$ we can make inference (and prediction) in a similar fashion to the univariate case.

- Can construct an MCMC algorithm to sample from the target density:

$$\pi(\boldsymbol{\theta}, Z^\star | \boldsymbol{x}).$$

- For such an algorithm to be efficient, it should involve sampling efficiently from the unobserved (bivariate) state:

$$\pi(Z^\star | \boldsymbol{\theta}, \boldsymbol{x}).$$

  A Forward-Backward algorithm becomes infeasible as the number of companies/chains increases.

- Gibbs sampler is also possible, i.e. alternate between $\pi(\boldsymbol{Z}^{(1)} | Z^{(2)}, \boldsymbol{\theta}, \boldsymbol{x})$. and $\pi(\boldsymbol{Z}^{(2)} | Z^{(1)}, \boldsymbol{\theta}, \boldsymbol{x})$.

Modelling the Bivariate Markov Chain

Additional assumption:

$$Z_t^{(1)} \perp\!\!\!\perp Z_t^{(2)} \mid Z_{1:t-1}^{(1)}, Z_{1:t-1}^{(2)}.$$

Instead of parameterising $\left\{Z^{(1)}, Z^{(2)}\right\}$ with $M^{\star}$, we introduce parameters, $\boldsymbol{\eta}$:

$$\eta_{abd}^{(i)} = P\left(Z_t^{(i)} = d \mid Z_{t-1}^{(1)} = a, Z_{t-1}^{(2)} = b\right), \quad \text{where } i = 1, 2 \quad \text{and } a, b, d \in \{0, 1\}.$$

Such a parameterisation imposes some constraints since $\eta_{ab1}^{(i)} + \eta_{ab0}^{(i)} = 1$. That means we can't recover all elements of $M^{\star}$; **but:**

- If interested in prediction, then we don't need the full $M^{\star}$; knowing $\boldsymbol{\eta}$ is sufficient;
- We are going to model η to understand how does one chain (company) affects the other.

Modelling the Marginal Probabilities η

If we consider data from two companies, then we have 8 independent η parameters.

Denote by

$$\eta^{(i)}_{z_1 z_2 1} = P\left(Z^{(i)}_t = 1 \,|\, Z^{(1)}_{t-1} = z_1, Z^{(2)}_{t-1} = z_2\right).$$

We model $\operatorname{logit} \eta^{(i)}_{z_1 z_2 1}$ as follows:

$$\begin{aligned} \operatorname{logit}\left(\eta^{(i)}_{z_1\, z_2 1}\right) &= \beta^{(i)}_0 + \beta^{(i)}_1 \mathbb{1}\left(Z^{(1)}_{t-1} = 1\right) + \beta^{(i)}_2 \mathbb{1}\left(Z^{(2)}_{t-1} = 1\right) \\ &+ \beta^{(i)}_3 \mathbb{1}\left(Z^{(1)}_{t-1} = 1 \text{ and } Z^{(2)}_{t-1} = 1\right) \end{aligned}$$

Parameters β_1, β_2 and β_3 provide information about how does the state of one company affect the state of the other.

Putting Everything Together

- Introducing the η parametrisation not only allows us to investigate the dependence between the different railway companies but also enable us efficient sampling of the unobserved process $\boldsymbol{Z} = \left(\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}\right)$ given observed data $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.
- We have a developed an efficient algorithm to sample from $\pi(\boldsymbol{Z}|\boldsymbol{x}, \boldsymbol{\theta})$ that scales well with the number of companies.
- We have also utilised a Polya-Gamma representation to sample efficiently from $\pi(\boldsymbol{\beta}\,|\,\boldsymbol{x}, \boldsymbol{Z})$.

Conclusions

Epilogue

- We have developed a probabilistic modelling framework for analysing multivariate discrete time-series data and an efficient Markov Chain Monte Carlo algorithm to allow fitting the model within a Bayesian framework.
- Key to our work is a reparameterisation which is sufficient for prediction (which is our focus).
- There has also been a significant amount of work with regards to the theoretical properties of our model.
- There are numerous ways to move this work forward, e.g.
    - Classifying the tweets better;
    - Estimate $\lambda(t)$ non-parametrically;
    - Different models for η.