Online Monitoring and Anomaly Detection in Multivariate Stochastic Processes

Alex Bolton, Nick Heard a.bolton12@imperial.ac.uk

Imperial College London

Objective

The aim of this project is to develop methods for monitoring patterns of behaviour in dynamic networks.

Applications

An active computer virus may change the behaviour of some computers in a network. If this can be detected then the virus may be combatted. A wide variety of other examples are of interest, including:

- Analysing how patterns of drug seizures change over time.
- Classifying software as malign code or malware.

Change point methodology

Interactions between nodes in a network form a stochastic process, and change point detection methods can detect behaviour changes.

Suppose we observe N processes operating simultaneously in a network, giving data D. Each univariate random process may be generated from a different distribution. We have change points $\tau = (\tau_1, \ldots, \tau_k)$. Each τ_i has an associated 'marked' vector $I_i \in \{0,1\}^N$ noting which univariate pocesses change at τ_i. The posterior density $p(I, \tau, k \mid D)$ is estimated using a reversible jump MCMC (RJMCMC) scheme, which proposes change points and marked vectors. Change point detection for multivariate processes has not received much attention in literature.

The likelihood of the data is calculated as a product over processes and over intervals between change points:

$$L(D \mid I, \tau, k) = \prod_{j=1}^{N} \prod_{l:I_{l,j}=1} L(D_{l,j}).$$

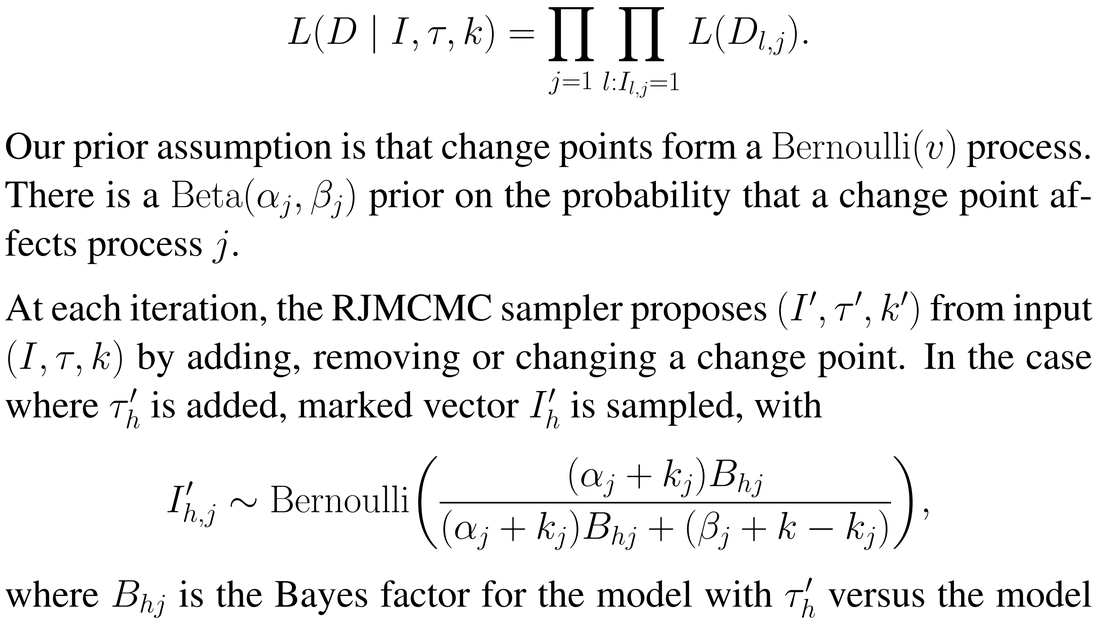
Our prior assumption is that change points form a $\mathrm{Bernoulli}(v)$ process. There is a $\mathrm{Beta}(\alpha_j, \beta_j)$ prior on the probability that a change point affects process j.

At each iteration, the RJMCMC sampler proposes (I', τ', k') from input (I, τ, k) by adding, removing or changing a change point. In the case where τ'_h is added, marked vector I'_h is sampled, with

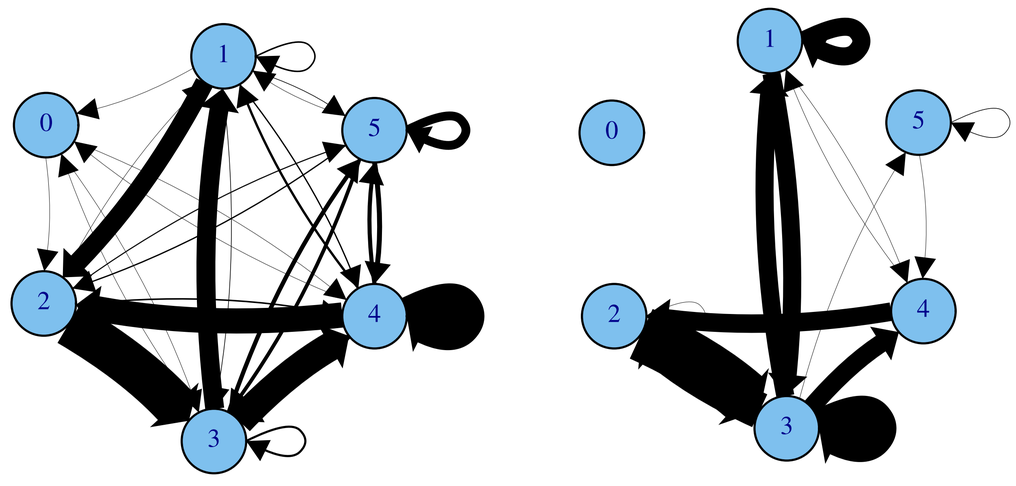
$$I'_{h,j} \sim \mathrm{Bernoulli}\left(\frac{(\alpha_j + k_j)B_{hj}}{(\alpha_j + k_j)B_{hj} + (\beta_j + k - k_j)}\right),$$

where B_{hj} is the Bayes factor for the model with τ'_h versus the model without. The change point is accepted with probability

$$\min\left\{1, \frac{v}{1-v}\frac{d_{k+1}(n-1-k)}{b_k(k+1)}\prod_{j=1}^{N}\frac{\beta_j + k - k_j + (\alpha_j + k_j)B_{hj}}{\alpha_j + \beta_j + k}\right\}.$$



Daily drug seizure probabilities for 5 different towns. Created using United Nations Office on Drugs and Crime data.



Transitions in a sequential trace of malware, the left graph shows the first 2^{10} transitions, the right graph shows the final 2^{10}.

Regime switching methodology

The data may be generated from a fixed (but unknown) number of regimes. Processes subject to the same regime have similar behaviour.

Change points have marked vectors which note regimes. The regime switching pattern is modelled as a Markov chain. So for any univariate process j, the probability that regime m is adopted at τ_i only depends on the regime at τ_{i-1}. A $\mathrm{Dirichlet}$ prior is used for the probability vector $(w_{j,u,1}, \ldots, w_{j,u,r_j})$, where $w_{j,u,m}$ is the probability that process j will switch from regime u to regime m at a change point. The prior distribution for the positions of change points is a $\mathrm{Bernoulli}(v)$ process.

The likelihood is given by a product over processes and over regimes:

$$L(D \mid I, \tau, k) = \prod_{j=1}^{N} L(D_j) = \prod_{j=1}^{N}\prod_{i=1}^{r_j} L(D_{i,j}).$$

The posterior $p(I, \tau, k \mid D)$ can be calculated with an RJMCMC sampler. At each iteration the sampler proposes (I', τ', k') from (I, τ, k). If we have added a change point τ'_h then we sample $I'_{h,j}$ as follows:

$$\mathbb{P}(I'_{h,j} = u) = \left(\frac{q^u_{h,j}B^u_{h,j}}{\sum_{m=1}^{r_j+1} q^m_{h,j}B^m_{h,j}}\right), \text{ for } u = 1, \ldots, r_j + 1.$$

Let $B_{h,j}$ be the Bayes factor for the model where τ'_h affects process j against the model where it does not. We also set $q_{h,j}$ so that

$$\prod_{j=1}^{N} q_{h,j} = \frac{\pi(I' \mid k')}{\pi(I \mid k)}.$$

The acceptance probability for this new change point is then

$$\min\left\{1, \frac{v}{1-v}\frac{d_{k+1}(n-1-k)}{b_k(k+1)}\prod_{j=1}^{N}\sum_{u=1}^{r_j+1} q^u_{h,j}B^u_{h,j}\right\}.$$

Obstacles

- Missing data are a common problem, as are seasonal patterns in behaviour, making inference difficult.
- Our model assumes that behaviour changes occur instantly at change points. Changes occuring over an interval may be missed.
- To calculate likelihoods we need to know how each univariate process is generated. The likelihood $L(D_{l,j}) = \int L(D_{l,j} \mid \theta_{l,j})\,\pi(\theta_{l,j})d\theta_{l,j}$ can be estimated using MCMC methods, but the sampler will be much more efficient if we can calculate the integral explicitly.

Summary

RJMCMC samplers will be used to estimate the posterior density of change points and regimes in network data. This work has a variety of applications. For example, modelling a sequential trace of software as a Markov chain with changing transition matrices may help us to classify the software as malign code or malware.

EPSRC Pioneering research and skills