BSTERGM: Bayesian Separable-Temporal Exponential family Graph Models

Choi Seokjun

16 Oct. 2020

Outline

Random graphs

Definition (Random graphs and related terminologies)

Let $\mathcal{B} = \{0, 1\}$. For given $n \in \mathbb{N}$,

- The set $\mathcal{Y} \subset \mathcal{B}^{n^2}$ is a set of graphs of n nodes (without weights for each nodes and edges.)
- Let Ω be an event set. We say $Y : \Omega \to \mathcal{Y}$ is a random variable for a graph, or a random graph.
- For a random graph $Y \in \mathcal{Y}$, denote the edge between i-th node and j-th node by Y_{ij} for $i, j = 1, 2, ..., n$, satisfying $Y_{ij} = 1$ if the edge is connected. Otherwise, $Y_{ij} = 0$.
- If edges of Y have directions, then Y is called a directed graph. Otherwise, Y is called a undirected graph.

Let me notate a realization of random graph by y and its edges by y_{ij} for $i, j = 1, 2, ..., n$.

Here is a remark. These are obvious that, for given $n \in \mathbb{N}$,

- $|\mathcal{Y}| = 2^{n(n-1)}$ if $\mathcal{Y}$ is the set of all directed graphs not permitting self-connecting edges.
- $|\mathcal{Y}| = 2^{n(n-1)/2}$ if $\mathcal{Y}$ is the set of all graphs of undirected one.

;thus, the size of $\mathcal{Y}$ grows exponentially when n increases.

ERGM: Exponential family Random Graphs Models

Definition (ERGM: Exponential family Random Graphs Models)

Let $\mathcal{Y}$ be a set of graphs with n nodes and Y be a random graph (on $\mathcal{Y}$.) Set a distribution on $\mathcal{Y}$ to

$$P(Y = y; \theta) = \frac{exp(\theta^T s(y))}{c(\theta)}$$

for some $\theta \in \mathbb{R}^p$, where $s(y) \in \mathbb{R}^p$ is a vector which is part of y's sufficient network statistics, and $c(\theta) \in \mathbb{R}$ is a normalizing constant satisfying $c(\theta) = \sum_{y \in \mathcal{Y}} exp(\theta^T s(y))$.
Models which have a such form called the ERGM.

Definition (TERGM: Temporal Exponential family Random Graphs Models)

Let $\mathcal{Y}$ be a set of graphs with n nodes. Let $Y_1 = y_1 \in \mathcal{Y}$ be given and $Y_2, ..., Y_T (T \in \mathbb{N})$ be random graphs (on $\mathcal{Y}$). Set a distribution on $\mathcal{Y} \times ... \times \mathcal{Y}$ ($T - 1$ folds) to

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}; \theta) = \frac{exp(\theta^T s(y_t, y_{t-1}))}{c(\theta, y_{t-1})}$$

for $\theta \in \mathbb{R}^p$ and with the first-order Markov assumption $P(Y_2, Y_3, ..., Y_T | Y_1) = P(Y_2|Y_1)P(Y_3|Y_2)...P(Y_T|Y_{T-1})$, where $s(y_t, y_{t-1}) \in \mathbb{R}^n$ is a part of sufficient network statistics, and $c(\theta, y_{t-1}) = \sum_{y \in \mathcal{Y}} exp(\theta^T s(y, y_{t-1}))$ is a normalizing constant. Models which have a such form called the TERGM.

STERGM: Separable-Temporal Exponential family Random Graphs Models

Definition (STERGM: Separable-Temporal Exponential family Random Graphs Models)

Let $\mathcal{Y}$ be a set of graphs with n nodes. Let $Y_1 = y_1 \in \mathcal{Y}$ be given and $Y_2, ..., Y_T (T \in \mathbb{N})$ be random graphs (on $\mathcal{Y}$). Set a distribution on $\mathcal{Y} \times ... \times \mathcal{Y}$ ($T-1$ folds) by following way:
Let $\mathcal{Y}^+|_t$ be a subset of $\mathcal{Y}$ consisting all graphs which have equal or additional edges comparing to y_{t-1}.
Likewise, let $\mathcal{Y}^-|_t$ be a subset of $\mathcal{Y}$ consisting all graphs which have equal or sparse edges comparing to y_{t-1}.
Next, for $Y_t^+ : \Omega \to \mathcal{Y}^+|_t$, $Y_t^- : \Omega \to \mathcal{Y}^-|_t$, $y_t^+ \in \mathcal{Y}^+|_t$ and $y_t^- \in \mathcal{Y}^-|_t$, set

$$P(Y_t^+ = y_t^+ | Y_{t-1} = y_{t-1}; \theta^+) = \frac{exp((\theta^+)^T s(y_t^+, y_{t-1}))}{c(\theta^+, y_{t-1})}, P(Y_t^- = y_t^- | Y_{t-1} = y_{t-1}; \theta^-) = \frac{exp((\theta^-)^T s(y_t^-, y_{t-1}))}{c(\theta^-, y_{t-1})}$$

for some $\theta^+, \theta^- \in \mathbb{R}^p$, $s(y_t^+, y_{t-1}), s(y_t^-, y_{t-1}) \in \mathbb{R}^n$, which are parts of sufficient network statistics, and normalizers $c(\theta^+, y_{t-1}) = \sum_{y^+ \in \mathcal{Y}^+} exp((\theta^+)^T s(y^+, y_{t-1})), c(\theta^-, y_{t-1}) = \sum_{y^- \in \mathcal{Y}^-} exp((\theta^-)^T s(y^-, y_{t-1}))$.
Then, defining operations $+, -$ on $\mathcal{Y}$ following the boolean algebra edgewise-ly, set y_t to

$$y_t = y_t^+ - (y_{t-1} - y_t^-) = y_t^- + (y_t^+ - y_{t-1})$$

Additionally, assume that

- The first-order Markov assumption: $P(Y_2, ..., Y_T | Y_1) = P(Y_2|Y_1)...P(Y_T|Y_{T-1})$
- The separability: the conditional independence between Y_t^+ and Y_t^- for all $t = 2, ..., T$; thus, $P(Y_t = y_t | Y_{t-1} = y_{t-1}; \theta^+, \theta^-) = P(Y_t^+ = y_t^+ | Y_{t-1} = y_{t-1}; \theta^+) P(Y_t^- = y_t^- | Y_{t-1} = y_{t-1}; \theta^-)$

Models which have a such form called the STERGM.

BSTERGM: Bayesian STERGM

To convert the STERGM to the Bayesian setting, put priors $p(\theta^+), p(\theta^-)$ over θ^+, θ^- and take the Bayes theorem as a inference rule. Then, the posterior of θ^+, θ^- becomes

$$P(\theta^+, \theta^- | y_t, y_{t-1}) = \frac{P(Y_t^+ = y_t^+ | y_{t-1}, \theta^+) P(Y_t^- = y_t^- | y_{t-1}, \theta^-) P(\theta^+), P(\theta^-)}{c(\theta^+, y_{t-1}) c(\theta^-, y_{t-1})}$$

where $y_t = y_t^+ - (y_{t-1} - y_t^-) = y_t^- + (y_t^+ - y_{t-1})$.

Remarks:

- We cannot compute the normalizing constants $c(\theta^+, y_{t-1})$, $c(\theta^-, y_{t-1})$ practically because we need to sum up too many terms.
- The constants are doubly intractable: they depend on θ^+,θ^-, the parameters of a model. Thus, we cannot use an ordinary MCMC algorithm to get the posterior sample.

Network Statistics for BSTERGM

As a usual, $s(y_{t}^{\cdot}, y_{t-1})$ would be chosen by a difference of network statistics, $s'(y_{t}^{\cdot}) - s'(y_{t-1})$. A common candidates of s' are:

- the number of edges
- node degree distribution (each order)
- edgewise shared partner distribution (each order)
- dyadwise shared partner distribution (each order)
- k-star distribution
- triangle distribution

In fact, a bundle of all order of node degree distribution and all order of edgewise shared partner distribution form sufficient statistics of a graph. Also, we can set s' by a function of these statistics.

Fitting the BSTERGM

Suppose that a sequence of graph samples $y_1, ..., y_T$ is observed
(then, $y_2^+, ..., y_T^+$ and $y_2^-, ..., y_T^-$ are uniquely determined,)
and we want to fit the observation using BSTERGM.

How do we find the posterior distribution of θ^+, θ^-?

Note that an ordinary MCMC algorithm does not work in our situation, because the constant part remains when calculating the MCMC ratio. To solve this doubly-intractable constant problem, we should use more special MCMC technique, the exchange MCMC algorithm.

Fitting the BSTERGM

Here is the algorithm of our main chain for the exchange algorithm.

Algorithm (the main chain)

Let $y_1, ..., y_T$ be given. For $m = 1, ..., M$,

1. Propose candidates θ_*^+, θ_*^- from $\epsilon(.|\theta_{m-1}^+, \theta_{m-1}^-)$.
2. Select a lag $(t-1, t)$ randomly on $2 \leq t \leq T$.
3. Generate an exchange graph $y_{ex} \in \mathcal{Y}|_t$ (with y_{ex}^+, y_{ex}^-) at the θ_*^+, θ_*^-.
4. Calculate the exchange MCMC ratio π at the lag,

$$\pi = \frac{P(y_t^+|y_{t-1}, \theta_*^+)P(y_t^-|y_{t-1}, \theta_*^-)p(\theta_*^+, \theta_*^-)}{P(y_t^+|y_{t-1}, \theta_{m-1}^+)P(y_t^-|y_{t-1}, \theta_{m-1}^-)p(\theta_{m-1}^+, \theta_{m-1}^-)} \frac{P(y_{ex}^+|y_{t-1}, \theta_{m-1}^+)P(y_{ex}^-|y_{t-1}, \theta_{m-1}^-)}{P(y_{ex}^+|y_{t-1}, \theta_*^+)P(y_{ex}^-|y_{t-1}, \theta_*^-)}$$

5. With probability $min(\pi, 1)$, accept the proposal and put $(\theta_m^+, \theta_m^-) = (\theta_*^+, \theta_*^-)$.
   Otherwise, reject the proposal and put $(\theta_m^+, \theta_m^-) = (\theta_{m-1}^+, \theta_{m-1}^-)$.

Observe that, the π has no normalizing constant terms because they are canceled out by added exchange terms.

$$log\pi = (\theta_*^+ - \theta_{m-1}^+)(s(y_t^+, y_{t-1}) - s(y_{ex}^+, y_{t-1})) + (\theta_*^- - \theta_{m-1}^-)(s(y_t^-, y_{t-1}) - s(y_{ex}^-, y_{t-1})) + log\frac{P(\theta_*^+, \theta_*^-)}{P(\theta_{m-1}^+, \theta_{m-1}^-)}$$

$$= (\theta_*^+ - \theta_{m-1}^+)(s'(y_t^+) - s'(y_{ex}^+)) + (\theta_*^- - \theta_{m-1}^-)(s'(y_t^-) - s'(y_{ex}^-)) + log\frac{P(\theta_*^+, \theta_*^-)}{P(\theta_{m-1}^+, \theta_{m-1}^-)}$$

Fitting the BSTERGM

We need one more thing. To generate the exchange graph y_{ex} for time t (of the second part of main chain), we should have a generative algorithm at a given parameter points. However, we still do not know the normalizing constant. Thus, I use one more MCMC chain.

Algorithm (the auxiliary chain)

Let $\theta^+, \theta^-, y_{t-1}$ be given. For $k = 1, ..., K$,

1. Select one edge randomly from the y_{k-1}, say, $y_{k-1;ij}$.
2. Propose a new graph y_* with switching the y_{ij} value from y_{k-1}:
   If $y_{k-1;ij} = 1$, then set $y_{*;ij} = 0$(dissolution case.) Oterwise, If $y_{k-1;ij} = 0$, then set $y_{*;ij} = 1$(formation case.)
   (If sample graphs are undirected, switch y_{ji} simultaneously.)
3. Take θ as θ^+ or θ^- according to the case. Calculate the MCMC ratio ϕ,

$$\phi = \frac{P(Y_t = y_*|y_{t-1}, \theta)}{P(Y_t = y_{k-1}|y_{t-1}, \theta)} = \frac{exp(\theta^T s(y_*, y_{t-1}))}{exp(\theta^T s(y_{k-1}, y_{t-1}))} = exp(\theta^T (s'(y*) - s'(y_{k-1})))$$

4. With probability $min(\phi, 1)$, accept the proposal and put $y_k = y_*$.
   Otherwise, reject the proposal and put $y_k = y_{k-1}$.

After K iteration, use the last network as the exchange sample in the second part of the main algorithm.

MCMC Diagnosis

Since we run two kinds of MCMC chains, we should proceed two diagnostic task.

The main chain produces the posterior samples of parameter, so the procedure is same as ordinary MCMC case.

- Cut burn-in period. Do thinning if it is needed.
- Depict traceplots of each parameter chain to check the convergence and the mixing.
- Depict autocorrelation plot. Calculate ESS if it is needed.

Next, checking all auxiliary chains is practically irritating one. In general, it is suffice to check the auxiliary chain of the last iteration (of the main chain) with statistics included in the model.

- Calculate network statistics of all graphs produced by the last auxiliary chain.
- Depict traceplots of each statistics to check the convergence and the mixing.

Inference and Prediction

Since we have posterior samples by running the main chain as many as we want, we can do basic inference procedure.

- A outlining shape of the posterior of $\theta^+, \theta^- | y_1, y_2, ..., y_T$ by histogram.
- An approximated summary statistics: mean, mode, variance, ...
- An approximated quantile and probability interval

Moreover, we already have a generative algorithm at the specific parameter point, we can predict the form of network at $T+1, T+2, ...$ using posterior sample following standard Bayesian method. For example, for predicting $T+1$,

- Run K iteration using auxiliary chain at y_T with each posterior sample points.
- Take the each network as a predicted result.

If you need, calculate some network statistics for the results and get a summary statistics of them.

Goodness of Fit

To evaluate the goodness of fit for the posterior θ^{+}, θ^{-}, we can use the auxiliary chain algorithm once again.

Algorithm (GOF procedure)

For t=2,...,T
For s=1,...,S

1. sample $\theta_s^{+}, \theta_s^{-}$ from the estimate of posterior.
2. simulate y_s using the auxiliary chain under y_{t-1}.
3. calculate $g(y_s)$, some higher degree statistics (eg. Node-degree dist & Edgewise Shared Partner dist)

Draw the box-plot of $g(y_s)$ and compare with $g(y_t)$.

Supplements

You can find the C++ implementation (using Armadilo: see http://arma.sourceforge.net/)
of BSTERGM fitting, diagnostic, and GOF algorithms
at my Github page: https://github.com/letsjdosth/BayesianSTERGM.