BSTERGM: Bayesian Separable-Temporal Exponential family Graph Models

Choi Seokjun

16 Oct. 2020

Outline

ERGM: Exponential family Random Graphs Models

Definition (ERGM: Exponential family Random Graphs Models)

Let $\mathcal{Y} \subset \{0,1\}^{n^2}$ 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 are called the ERGM.

STERGM: Separable-Temporal Exponential family Random Graphs Models

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

Let $Y_{t-1} = y_{t-1} \in \mathcal{Y}$ be given and $Y_t \in \mathcal{Y}$. Let $\mathcal{Y}^+ \subset \mathcal{Y}$ be a set of all graphs having equal or additional edges comparing to y_{t-1}. Likewise, define $\mathcal{Y}^- \subset \mathcal{Y}$ having equal or sparser edges. For $Y_t^+ \in \mathcal{Y}^+$, $Y_t^- \in \mathcal{Y}^-$, set Formation model:

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

Dissolution model:

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

for $\theta^+, \theta^- \in \mathbb{R}^p$, where $s(y_t^+, y_{t-1}), s(y_t^-, y_{t-1}) \in \mathbb{R}^n$ are (functions 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 $\mathcal{Y}$ by the boolean algebra edgewisely, 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 are 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 rule as the 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, $s'(y_t^{\cdot}) - s'(y_{t-1})$. For example, some candidates of s' are:

- the number of edges
- k-stars
- triangles

Note that we can set s' by a function of sufficient network 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^-?

Fitting the BSTERGM

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 $y_{ex} \in \mathcal{Y}$ (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.

$$log\pi = (\theta^+_* - \theta^+_{m-1})(s'(y^+_t) - s'(y^+_{ex})) + (\theta^-_* - \theta^-_{m-1})(s'(y^-_t) - s'(y^-_{ex})) + log\frac{P(\theta^+_*, \theta^-_*)}{P(\theta^+_{m-1}, \theta^-_{m-1})}$$

Moreover, it is noteworthy that for the second step, an auxiliary MCMC chain which generate an exchange sample is used.

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 follow the common inference procedure for the Bayes models.

- An 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 of θ^+, θ^-, we can use the auxiliary chain algorithm once again.

Algorithm (GOF procedure)

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

1. sample θ_s^+, θ_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 a box-plot of $g(y_s)$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.