Stochastic Actor-Oriented Models for Social Networks

## 0.1 Terminology & Notation

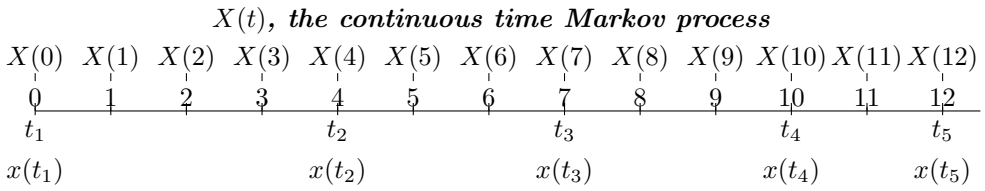
- **Network** - A collection of actors and ties between them. A whole network (all of the nodes and the ties between them) is denoted $x$. In SAOMs, the networks are observed in waves for time points $m = 1, \ldots, M$. Sometimes, the networks observed are denoted as $x(t_m)$.

- **Node** or **Actor** - The entities that form ties in a network. The number of nodes in a network is denoted $n$. The set of nodes, $\{i : i = 1, \ldots n\}$ is denoted $\mathcal{N}$.

- **Edge** or **Tie** - The connections between actors in a network. A tie from actor $i$ to actor $j$ is denoted by $x_{ij}$. In SAOMs, ties are almost always directed, so $x_{ij}$ does not necessarily equal $x_{ji}$. $x_{ij} = 1$ if there is a tie from actor $i$ to actor $j$ and $x_{ij} = 0$ if there is not a tie from actor $i$ to actor $j$. Ties from an actor to itself, $x_{ii}$, are structurally zero, meaning that self-ties are not allowed. In a network of $n$ nodes, there are $n(n-1)$ possible ties.

- **Adjacency matrix** - The adjacency matrix of a network is a numerical representation of it. The adjaceny matrix of $x$ is denoted $\mathcal{A}(x)$, is of dimension $n \times n$ and has $ij^{th}$ entry equal to the value of $x_{ij}$ for all $i \neq j \in \mathcal{N}$. The diagonal entries of $\mathcal{A}(x)$, $ii$ for $i = 1, \ldots, n$, are all structurally 0.

- $\mathbf{X(t)}$ - A continuous time Markov process *in which* the observed panel of $M$ networks of size $n$ is embedded. For each $t = 1, 2, 3, \ldots$, there is one tie change. In other words, the network $X(t+1)$ only differs from $X(t)$ by one tie, $x_{ij}$. Call this tie $x_{ij}(t)$ in network $X(t)$ and $x_{ij}(t+1)$ in network $X(t+1)$. Then, $x_{ij}(t+1) = 1 - x_{ij}(t)$. Note that $t$ is different from $t_m$. Here is an illustrated example:




  Thus, each network observation, $x(t_m)$ represents the concatenation of (usually) multiple consecutive steps in the Markov chain in SAOMs. The state space for the the Markov chain is $\mathcal{S}$, the large but finite set of all possible adjacency matrices for $n$ nodes with zeroes on the diagonal. Note that $|\mathcal{S}| = 2^{n(n-1)}$.[1] Also note that the time points $t$ are not strictly 1 unit of time apart. The space between them is dictated by the distribution of the waiting times in the Markov process.

- **Rate of Change** - One of the defining elements of SAOMs is the function for the rate of change of a network and of the actors in that network. For actor $i$, the rate function is commonly denoted $\lambda_i(\alpha, x)$, which dictates how quickly actor $i$ gets opportunities to change one of its ties, $x_{ij}$. In this function, $\alpha$ is a parameter and $x$ is the current network state. The rate function for the whole network is $\lambda(\alpha, x) = \sum_i \lambda_i(\alpha, x)$. For any time point, $t$, where $X(t) \equiv x$, the waiting time to the next change opportunity by *any* actor has distribution $Exponential(\lambda(\alpha, x))$. This is because the chain of network evolutions is Markovian. There are many possibilities for the rate function. The simplest is that it is constant between observations $x(t_m)$. The rate function can also depend on covariate values of the actors or structural network elements such as outdegree. For instance, $\lambda_i(\alpha, x) = \exp\{\sum_k \alpha_k r_{ik}(x)\}$, where $r_{ik}(x)$ is a function of the current network state and the node $i$ (one of the aforementioned possible dependencies), and $k = 1, \ldots K$ could be any number of functions as chosen by the researcher from previous subject matter knowledge or otherwise.

- **Probability of Change (nodes)** - Given that a change, call it $X(t) \equiv x \to X(t+1)$ occurs, each node has its own probability of changing. Recall that from time $t$ to time $t+1$, only one tie is changing (call it $x_{ij}$), so that only one node, $i$, is given the power to change one of its $n-1$ available ties from $x_i j$ to $1 - x_{ij}$. The probability that node $i$ will be the one to change a tie, given $x \to X(t+1)$ is $\pi_i(\alpha, x) \equiv \pi_i(\alpha, x | x \to X(t+1)) = \frac{\lambda_i(\alpha, x)}{\lambda(\alpha, x)}$.

- **Permitted Changes** - First note a slight abuse of notation. The current state of the network, which was previously denoted $X(t)$ and $x$, will now be denoted $x^0$. Then, let $x$ instead represent the new, as yet undetermined network $X(t+1)$. So, given that a change, $x^0 \to x$ will occur, and that node $i$ has been selected to change a tie, there is a set of permitted changes, $\mathcal{A}_i(x^0) = \{\mathcal{A}(x^0) \cup \mathcal{A}_i^r(x^0)\}$, where $\mathcal{A}(x^0)$ is

[1] See http://oeis.org/A053763.

the current network's adjacency matrix, and $\mathcal{A}_i^r(x^0) = \{x | x_{ij} = 1 - x_{ij}^0$ for only one $j \in \mathcal{N}, j \neq i, \}$. So, the set of permitted changes includes the set of all adjacency matrices with all entries equivalent to the current adjacency matrix with the exception of one of the non-fixed, $n-1$ entries of the $i^{th}$ row and the current adjacency matrix, meaning that $|\mathcal{A}_i(x^0)| = n$. Additionally, define the complete set of possible changes from $x^0 \to x$ as $\mathcal{A}_{1:n}(x^0) = \{\mathcal{A}(x^0) \cup \mathcal{A}_1(x^0) \cup \mathcal{A}_2(x^0) \cup \cdots \cup \mathcal{A}_n(x^0)\}$. This is the set of all networks that differ from $x^0$ by only one tie and has size $|\mathcal{A}_{1:n}(x^0)| = n(n-1)+1$.

- **Objective function** - So, once a change in node $i$ is given, *how* does that node determine which of the $n$ permitted changes it will make? Its choice is determined in large part by its objective function, $f_i(\beta, x^0, x)$, where $\beta$ is a parameter. This function can be interpreted as the "relative attractiveness" for node $i$ associated with making a change from the current network state, $x^0$ to a new network state, $x$, where $x \in \mathcal{A}_i(x^0)$. It is assumed that actors want to maximize this function when given an opportunity for change. For example, if changing from $x^0$ to $x$ will decrease the value of $f_i$ for all $x \in \mathcal{A}_i(x^0)$, the node will be very likely to *not* change any ties. Many models don't include the current state, $x^0$, in the objective function, but instead only consider $x$, the future state, and just attempt to get closer to their optimum state regardless of where they currently are. The objective function is usually defined linearly as $f_i(\beta, x^0, x) = \sum_k \beta_k s_{ik}(x^0, x)$, where the functions $s_{ik}$ are determined from subject matter knowledge. These, like the $r_{ik}$ in the rate function, can be actor covariates or structural variables like outdegree or reciprocated ties.

- **Probability of Change (tie given node)** - Given the change of a tie from node $i$, the probability of the change $x^0 \to x$, where $x \in \mathcal{A}_i(x^0)$ is $p_i(\beta, x^0, x) \equiv p_i(\beta, x^0, x | x^0 \to x, x \in \mathcal{A}_i(x^0)) = \frac{\exp\{f_i(\beta, x^0, x)\}}{\sum_{\tilde{x}} \exp\{f_i(\beta, x^0, \tilde{x})\}}$, where $\tilde{x} \in \mathcal{A}_i(x^0)$. For a specific tie, $x_{ij}$, with $j \neq i$, this expression can be re-written as $p_{ij}(\beta, x^0) = \frac{\exp\{f_i(\beta, x^0, x^j)\}}{f_i(\beta, x^0, x^0) + \sum_{j \neq i} \exp\{f_i(\beta, x^0, x^j)\}}$, where $x^j$ has ties $x_{ih}^j = x_{ih}^0$ for all $h \neq j$ and $x_{ih}^j = 1 - x_{ih}^0$ for $h = j$. Using this new expression, we can write the probability of no change for node $i$ ($p_i(\beta, x^0, x^0)$ in the original notation) as $p_{i0} \equiv p_i(\beta, x^0, x^0) = 1 - \sum_{j \neq i} p_{ij}(\beta, x^0)$.

- **Intensity Matrix** - The intensity matrix, $Q$, of a continuous time Markov process describes the rate of change between states of the chain. For networks, there are a very large number of possible states: $2^{n(n-1)}$. But, due to the properties of continuous-time Markov processes, there are only $n$ possible states given the current state, $n-1$ of which are uniquely determined by the node $i$ that is given the opportunity to change. Thus, the intensity matrix $Q$ is a very sparse $2^{n(n-1)} \times 2^{n(n-1)}$ matrix, with only $n(n-1)+1$ non-zero entries in each row. Note that $n(n-1)$ of these represent the possible states that are one edge different from a given state, and the additional non-zero entry is for the state to remain the same. All other entries in a row are zero because those column states cannot be reached from the row state by just one change. The entries of $Q$ are defined as follows: let $y \neq z \in \{1, 2, \ldots, 2^{n(n-1)}\}$ be indices of two different possible states of the network, $x^y, x^z \in \mathcal{S}$. For a timepoint $t$ define $x^y = X(t)$ and let $x^z \in \mathcal{A}_i(x^y)$ for each $i \in \mathcal{N}$. Then the $yz^{th}$ entry of $Q$ is:

$$q(x^y, x^z) = \begin{cases} \lambda_i(\alpha, x^y) p_i(\beta, x^y, x^z) & \text{if } x^z \in \mathcal{A}_i(x^y), \forall i \in \mathcal{N} \\ 0 & \text{if } x^z \notin \mathcal{A}_i(x^y) \text{ for any i } \in \mathcal{N} \end{cases}$$

  Properties of intensity matrices dictate that the diagonal entry, where the state doesn't change, is defined as $q(x^y, x^y) = -\sum_{z \neq y} q(x^y, x^z)$ so that the row sum is 0. Thus, the rate of change between any two states that differ by only one tie from $i \to j$ is the product of the rate at which actor $i$ gets to change a tie and the probability that the tie that will change is the tie to node $j$.[2] Furthermore, the theory of continuous time Markov chains gives that the matrix of transition probabilities between observation times $t_{m-1}$ and $t_m$ is dependent only on the difference between timepoints, $t_m - t_{m-1}$ and is equal to $e^{(t_m - t_{m-1})Q}$, where $Q$ is the matrix defined above and $e^X$ for a real or complex square matrix $X$ is equal to $\sum_{k=0}^{\infty} \frac{1}{k!} X^k$. There is a package in `R` to do this matrix exponential operation, but I can't imagine doing this for networks for any $n > 5$.

- **Moving from $x(t_{m-1})$ to $x(t_m)$** - First define $x(t_1)$ as the first network observed and $x(t_M)$ as the last network observed. Networks in between are written $x(t_m)$ for $m = 2, \ldots, M-1$. The observations $x(t_1), x(t_2), \ldots, x(t_M)$ are embedded in the continuous-time Markov chain $\{X(t) | t_1 \leq t \leq t_M\}$. Moving from the observation $x(t_{m-1})$ to the observation $x(t_m)$ requires $R$ changes or steps in the Markov process $X(t)$ where $R = \sum_{i,j \in \mathcal{N}} |x_{ij}(t_m) - x_{ij}(t_{m-1})|$. Note that the observed network change from $x(t_{m-1})$ to $x_(t_m)$ is conditionally independent of all other network observations given $x(t_{m-1})$. Denote the $R$ timepoints in the the Markov process that correspond to the unobserved changes in the network as $T_1, T_2, \ldots, T_R$ and define $T_0 = t_{m-1}$ and $T_R \leq t_m < T_{R+1}$. For each timepoint $T_r$, for $r = 1, \ldots R$, there is a single actor,

[2] Just to be clear, the change is from $x_{ij}^y$ to $x_{ij}^z = 1 - x_{ij}^y$.

denoted $I_r$ that gets an opportunity to change at this timepoint. This actor changes one of its ties to one node, $J_r$. If there is no change, $I_r = J_r$. Formally, the pair $(I_r, J_r)$ is the only $(i, j)$ pair for which $x_{ij}(T_{r-1}) \neq x_{ij}(T_r)$ if such a pair exists, and $(I_r, J_r) = (I_r, I_r)$ otherwise. The triplet $(T_r, I_r, J_r)$ forms a stochastic process for $r = 1, \ldots, R$ that, given the prior observation, $x(t_{m-1})$ for $m = 2, \ldots, M$, completely determines $x(t)$ for $t_{m-1} < t < t_m$.

- **Augmenting Data** (see section on ML estimation) - The augmenting data consist of $R$ and the pairs $(I_r, J_r)$ for $r = 1, \ldots, R$. (Note the timepoints $T_r$ are left out of the augmenting data.) The possible outcomes of the augmenting data are determined by the condition that if for a given tie, $x_{ij}(t_{m-1})$, the number of timepoints $1 \leq r \leq R$ at which $(i_r, j_r) = (i, j)$ is even if $x_{ij}(t_m) = x_{ij}(t_{m-1})$ and is odd if $x_{ij}(t_m) \neq x_{ij}(t_{m-1})$.

- **Sample Path** - The sample path is the stochastic process $V = ((I_r, J_r) : r = 1, \ldots, R)$. Note that the elements of $V$ where $I_r = J_r$ are redundant. However, they are kept in the process because they help with computation of the likelihood. For notational convenience, define $x^{(r)} = x(t_r)$. Using this notation, write $x(t_{m-1}) = x^(0)$ for the independent cases where $m = 2, \ldots, M$. Then, the networks $x^{(r)}$ and $x^{(r-1)}$ only differ in the element $(I_r, J_r) \in V$, provided that $I_r \neq J_r$.

- **PDF of Sample Path** - The sample path has probability distribution function

$$p_{sp}(V|\alpha, \beta, x^{(0)}) = Pr(T_R \leq t_m < T_{R+1}|x^{(0)}, V, \alpha, \beta) \times \prod_{r=1}^{R} \pi_{i_r}(\alpha, x^{(r-1)}) p_{i_r, j_r}(\beta, x^{(r-1)}).$$

The first component, $Pr(T_R \leq t_m < T_{R+1}|x^{(0)}, V, \alpha, \beta)$ is the probability that your next network observation at time point $t_m$ comes *before* the next change in the continuous time Markov chain, $X(T_{R+1})$. From a discussion that follows, $Pr(T_R \leq t_m < T_{R+1}|x^{(0)}, V, \alpha, \beta)$, the probability that your next network observation at time point $t_m$ comes *before* the next change in the continuous time Markov chain, $X(T_{R+1})$, can be approximated by

$$\kappa(\alpha, x(t_{m-1}), V) \approx \int_{t_{m-1}}^{t_m} p_{T_R}(s) \Pr(\{T_{R+1} - T_R > t_m - T_R\}|T_R = s) ds$$

$T_{R+1}$ has distribution Exponential($\lambda(\alpha, x^{(R)})$). So, $\Pr(\{T_{R+1} - T_R > t_m - T_R\}) = \Pr(T_{R+1} > t_m) = 1 - \Pr(T_{R+1} < t_m) = 1 - (1 - \exp\{-\lambda(\alpha, x^{(R)}) t_m\}) = \exp\{-\lambda(\alpha, x^{(R)}) t_m\}$. So, the above integral can be written as

$$\kappa(\alpha, x(t_{m-1}), V) \approx \int_{t_{m-1}}^{t_m} p_{T_R}(s) \exp\{-\lambda(\alpha, x^{(R)})(t_m - s)\}.ds$$

This quantity is approximately equal to $\frac{p_{T_R}(t_m)}{\lambda(\alpha, x^{(R)})}$. The second expression, $\prod_{r=1}^{R} \pi_{i_r}(\alpha, x^{(r-1)}) p_{i_r, j_r}(\beta, x^{(r-1)})$, is the PDF of the stochastic process $V$. So, the probability of the sample path can be approximated by:

$$p_{sp}(V|\alpha, \beta, x^{(0)}) \approx \frac{p_{T_R}(t_m)}{\lambda(\alpha, x^{(R)})} \times \prod_{r=1}^{R} \pi_{i_r}(\alpha, x^{(r-1)}) p_{i_r, j_r}(\beta, x^{(r-1)}).$$

So, for observed data, augmented by the sample path, "the likelihood conditional on $x(t_{m-1})$ can be expressed directly, either exactly [(for constant $\lambda_i$)] or in good approximation."

- **Distribution of waiting times between $x(t_{m-1})$ and $x(t_m)$** - Let $T_{r+1} - T_r$ be the waiting time between one of the $r = 1, \ldots, R$ changes between $x(t_{m-1})$ and $x(t_m)$. After conditioning on $x(t_{m-1}) \equiv x^{(0)}$ and $V$, these waiting times have distribution $T_{r+1} - T_r \sim$ Exponential($\lambda(\alpha, x^{(r)})$) for $r = 0, 1, \ldots, R-1$. Following from this, the distribution of $T_R - t_{m-1}$, the time it takes for the network to move from state $x(t_{m-1})$ to state $x(t_m)$ is the convolutions of the Exponential($\lambda(\alpha, x^{(r)})$) distributions for $r = 0, \ldots, R-1$. The pdf of $T_R - t_{m-1}$ is[3]:

$$h(T_R - t_1|\alpha, x(t_{m-1}), V) = \sum_{r=0}^{R-1} \frac{\lambda(\alpha, x^{(0)}) \cdot \lambda(\alpha, x^{(1)}) \cdot \lambda(\alpha, x^{(R-1)})}{\prod_{q \neq r, q=0}^{R-1} (\lambda(\alpha, x^{(q)}) - \lambda(\alpha, x^{(r)}))} \cdot \exp\{-(T_R - t_{m-1}) \cdot \lambda(\alpha, x^{(r)})\}$$

In the special case that the $\lambda_i(\alpha, x)$ rates are constant, call it $\alpha_1$, then the number of steps in the Markov chain to move from $x(t_{m-1})$ to $x(t_m)$, $R$ has distribution $R \sim$ Poisson($n\alpha_1(t_m - t_{m-1})$). In this case, $Pr(T_R \leq t_m < T_{R+1}|x^{(0)}, V, \alpha, \beta) = \kappa(\alpha_1, x^{(0)}, V) = \exp\{-n\alpha_1(t_m - t_{m-1})\} \frac{(n\alpha_1(t_m - t_{m-1}))^R}{R!}$. However, when the individual rates are different for each actor, this probability must be approximated.

[3]Source: `http://www.ccms.or.kr/data/pdfpaper/jcms21_4/21_4_501.pdf`

- **Distribution of $T_R$** - $T_R$ is the timepoint in the Markov chain at which the state $x(t_m)$ is acheived. Since $T_R - t_{m-1}$ is a convolution of exponentials with rate parameters $\lambda(\alpha, x^{(r)})$ for $r = 0, R-1$, by the Lyapounov Central Limit Theorem, the distribution of $T_R$ is approximately normal with mean $\mu_\alpha$ and variance $\sigma^2_\alpha$ where

$$\mu_\alpha = \sum_{r=0}^{R-1} \frac{1}{\lambda(\alpha, x^{(r)})} \quad \text{and} \quad \sigma^2_\alpha = \sum_{r=0}^{R-1} \frac{1}{\lambda(\alpha, x^{(r)})^2}$$

  The PDF of $T_R$ is thus $p_{T_R} \approx \frac{1}{\sqrt{2\pi\sigma^2_\alpha}} \exp\left\{\frac{-((t-t_{m-1})-\mu_\alpha)^2}{2\sigma^2_\alpha}\right\}$.

## 0.2 Important Facts about SAOMs

- In SAOMs, the observed networks in a longitudinal study, call them $x_{t_1}, \ldots, x_{t_M}$, are discrete observations embedded in a continuous time process, $X(t)$. This means that the probability model is defined *independently* of the observational design. Thus, the time between observations can be irregular and still be analyzed under this model.

- The continuous-time Markov process $\{X(t)|t_1 \leq t \leq t_M\}$ has a distribution that consitutes an exponential family, which are of the form $f(y|\theta) = \exp\{\sum_{j=1}^{s} q_j(\theta)T_j(y)\}c(\theta)h(y)$ where $y$ is the data vector and $\theta$ is the scalar or vector parameter. This means the SAOMs can be viewed as "incompletely observed exponential family" models.

- Inference is computer-intensive and time-consuming because of "elaborate MCMC procedures" that must be implemented.

- The definition of SAOMs means that data can be simulated directly from the probability distribution given the initial network state $x_{t_1}$.

- However, SAOMs are "too complicated for the calculation of likelihoods or estimators in closed form." The distribution, however, can be easily simulated from, so that estimation can proceed by MCMC simulation.

- ML estimation of SAOMs is based on the missing data principle. The "missing data" here are the paths from $x(t_{m-1})$ to $x(t_m)$ for all $m = 2, \ldots, M$. This set can be denoted for each $m$ by $V_m = \{(I_{m1}, J_{m1}), \ldots, (I_{mR_m}, J_{mR_m})\}$. So, the complete set of augmenting data is $V = (V_2, \ldots V_M)$.

## 0.3 Maximum Likelihood Estimation

Sketch of Algorithm:

1. Use the first network observation, $x(t_1)$, as a starting value. All analysis proceeds conditioning on $x(t_1)$.

2. For $m = 2, 3, \ldots, M$ where $M$ is the total number of observed networks, augment the observed data with random draws from the sequence of intermediate steps in the Markov process, $x(t)$, that could have led from $x(t_{m-1})$ to $x(t_m)$. i.e. draw from the set

$$\{(x(t_1^*), x(t_2^*), \ldots x(t_{C-1}^*))' | x(t_1^*) \in \mathcal{A}_{1:n}(x(t_{m-1})) \& x(t_2^*) \in \mathcal{A}_{1:n}(x(t_1^*)) \& x(t_3^*) \in \mathcal{A}_{1:n}(x(t_2^*)) \& \cdots \& x(t_m) \in \mathcal{A}_{1:n}(x(t_{C-1}^*))\}$$

   where $C = \sum_{i,j \in \mathcal{N}} |x_{ij}(t_m) - x_{ij}(t_{m-1})|$. Note that $C$ in the number of single edge changes needed to get from $x(t_{m-1})$ to $x(t_m)$, so $C - 1$ is the number of intermediate networks. These draws can be simulated using a Metropolis-Hastings algorithm.

3. Use the draws from 2 as updates in a Robbins-Monro algorithm (?) to find the solution of the likelihood equation.

Full Iterative Algorithm: For each iteration $N = 1, 2, \ldots$

1. For each $m = 2, \ldots, M$, make a large number of Metropolis-Hastings steps of the following form. Let $\underline{v} = ((i_1, j_1), \ldots, (i_R, i_R))$ be a given path from $x(t_{m-1})$ to $x(t_m)$ in the whole possible set of paths $\mathcal{V}$. Then propose a new path, $\tilde{\underline{v}}$ from the proposal distribution that consists of all of the possible small changes:

   (a) "Paired Deletions" - Of all pairs of indices $r_1, r_2$ such that $(i_{r_1}, j_{r_1}) = (i_{r_2}, j_{r_2})$, $i_{r_1} \neq j_{r_1}$, randomly select a pair $(r_1, r_2)$ and delete both from the current path.

   (b) "Paired Insertions" - Randomly select an edge $(i, j) \in \mathcal{N}^2$ with $i \neq j$. Randomly choose 2 indices, $r_1, r_2$. Insert $(i, j)$ immediately before $r_1$ and $r_2$.

   (c) "Single Insertions" - At a random place in the current path, insert $(i, i)$ for a random $i \in \mathcal{N}$.

(d) "Single Deletions" - Of all elements in the current path that satisfy $i_r = j_r$, randomly delete one of them.

(e) "Permutations" - For randomly chosen $r_1 < r_2$, where $r_2 - r_1$ is bounded from above by some small-ish number to avoid lengthy computation (tuning parameter?), the segment of consecutive elements $(i_r, j_r)$, $r = r_1, \ldots, r_2$ is randomly permuted.

Denote the proposal probabilities $u(\underline{\tilde{v}}|\underline{v})$ and the target probabilities $p(\underline{v})$. The target distribution is $p(\underline{v}) \propto p_{SP}(\underline{v}|\alpha, \beta, x(t_{m-1}))$. Then the acceptance probability for a proposal path $\underline{\tilde{v}}$ and a current path $\underline{v}$ is: $\min\left\{1, \frac{p(\underline{\tilde{v}})u(\underline{v}|\underline{\tilde{v}})}{p(\underline{v})u(\underline{\tilde{v}}|\underline{v})}\right\}$. The series of M-H steps will result in a new set of augmenting data, $(v^{(N)}) = (v_2^{(N)}, \ldots, v_M^{(N)})$.

2. Compute

$$S_{XV}(\theta^{\hat{(N)}}; x, v^{(N)}) = \sum_{m=2}^{M} S_m(\theta^{\hat{(N)}}; x(t_{m-1}), v_m^{(N)})$$

$S_{XV}(\theta^{\hat{(N)}}; x, v^{(N)})$ is the complete data score function. $S_m(\theta^{\hat{(N)}}; x(t_{m-1}), v_m^{(N)}) = \frac{\partial \log p_m(v_m;\theta|x(t_{m-1}))}{d\theta}$ is the total data score function at step $m$. $p_m$ is $p_{SP}(v_m; \theta = (\alpha, \beta)'|x(t_{m-1}))$, the probability of the sample path $v_m$.

3. Update

$$\theta^{(N+1)} = (\alpha^{(N+1)}, \beta^{(N+1)}) = \theta^{(N)} + a_N D^{-1} S_{XV}(\theta^{\hat{(N)}}; x, v^{(N)})$$

where $D$ is a MC estimate of the complete data observed Fisher information matrix, $D_{XV}(\theta) = -\frac{\partial S_{XV}(\theta:x,V)}{\partial \theta}$ estimated for $\theta = \theta^{(1)}$. (i.e. evaluated at the starting value of $\theta$.) The starting value is the method of moments estimator, $\hat{\theta}^{(1)}$. (See Snijders 2001 for how to get the MoM estimate). The sequence $a_N$ is just a sequence that approaches 0 as $N \to \infty$. (I'm not sure exactly what? Maybe just $\frac{1}{N}$ would work? There's no specifics in Snijders 2001, and I'm looking through the RSiena manual for clues...).