

**STATISTICAL TREATMENT OF GRAVITATIONAL
CLUSTERING ALGORITHM**

by

Yao Zhang

B.S. in Statistics, USTC, 1999

M.S in Mathematical Statistics, USTC, 2002

Submitted to the Graduate Faculty of
the Department of Statistics in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH
DEPARTMENT OF STATISTICS

This dissertation was presented

by

Yao Zhang

It was defended on

April 19, 2010

and approved by

Professor Satish Iyengar

Professor Henry Block

Assistant Professor Yu Cheng

Associate Professor Marek Druzdzel

Dissertation Director: Professor Satish Iyengar

ABSTRACT

**STATISTICAL TREATMENT OF GRAVITATIONAL CLUSTERING
ALGORITHM**

Yao Zhang, PhD

University of Pittsburgh, 2010

In neuroscience, simultaneously recorded spike trains from multiple neurons are increasingly common; however, the computational neuroscience problem of how to quantitatively analyze such data remains a challenge. Gerstein, et al. [5] proposed a gravitational clustering algorithm (GCA) for multiple spike trains to qualitatively study interactions, in particular excitation, among multiple neurons. This thesis is mainly focused on a probabilistic treatment of GCA and a statistical treatment of Gerstein’s interaction mode.

For a formal probabilistic treatment, we adopt homogeneous Poisson processes to generate the spike trains; define an interaction mode based on Gerstein’s formulation; analyze the asymptotic properties of its cluster index – GCA distances (GCAD). Under this framework, we show how the expectation of GCAD is related to a particular interaction mode, i.e., we prove that a time-adjusted-GCAD is a reasonable cluster index for large samples. We also indicate possible stronger results, such as central limit theorems and convergence to a Gaussian process.

In our statistical work, we construct a generalized mixture model to estimate Gerstein’s interaction mode. We notice two key features of Gerstein’s proposal: (1) each spike from each spike train was assumed to be triggered by either one previous spike from one other spike train or environment; (2) each spike train was transformed into a continuous longitudinal curve. Inspired by their work, we develop a Bayesian model to quantitatively estimate excitation effects in the network structure. Our approach generalizes the mixture model to

accommodate the network structure through a matrix Dirichlet distribution. The network structure in our model could either approximate the directed acyclic graph of a Bayesian network or be the directed graph in a dynamic Bayesian network. This model can be generally applied on high-dimensional longitudinal data to model its dynamics. Finally, we assess the sampling properties of this model and its application to multiple spike trains by simulation.

Keywords: Poisson process, generalized mixture model, matrix Dirichlet distribution, Bayes network, high-dimensional longitudinal data, multiple spike trains.

TABLE OF CONTENTS

1.0 INTRODUCTION	1
2.0 LITERATURE REVIEW	3
2.1 Dynamic Bayesian belief network	4
2.2 Gravitational Clustering Algorithm	6
2.3 Comparison of DBN and GCA	8
2.3.1 Interaction Modes	8
2.3.2 Mathematical aspects	9
2.3.3 Results	9
3.0 THEORETICAL ASPECTS OF GCA	11
3.0.4 Preliminaries	11
3.0.5 Moment properties of charge processes	13
3.0.6 Main results about GCAD ($K=2$)	14
3.0.7 Main results about GCAD ($K>2$)	15
4.0 GENERALIZED MIXTURE MODEL FOR HIGH-DIMENSIONAL LONGITUDINAL DATA	19
4.1 The problem	19
4.2 Statement of Generalized Mixture Model	20
4.3 Bayesian approach to computing the posterior	21
4.3.1 Preparation	22
4.3.2 Specification of prior distributions	24
4.3.3 Likelihood functions	26
4.3.4 Conditional Posterior distributions	28

4.3.5 Gibbs sampling procedure	33
5.0 SAMPLING AND SIMULATION STUDY	34
5.1 Some problems in sampling	34
5.2 Simulation study	38
5.3 Conclusion	40
APPENDIX. PROOFS OF RESULTS IN SECTION 3	56
INDEX	68
BIBLIOGRAPHY	70

LIST OF FIGURES

1	Compare samplers - CDF and PDF plots	41
2	Compare samplers - Empirical CDFs	42
3	Compare samplers - Tables for D_T and D_S	43
4	MD sample trace plot - MD sample trace plot for M	44
5	MD sample trace plot - MD sample trace plot for $M * 1000$	45
6	MD sample trace plot - MD sample trace plot for MM	46
7	MD sample trace plot - MD sample trace plot for $MM * 1000$	47
8	Simulation - Graph representation of copy matrix	48
9	Simulation - Posterior copy matrix table	49
10	Simulation - Raster plot for spike trains	50
11	Simulation - Plot for charge process	51
12	Simulation - Gravitational clustering plot	52
13	Simulation - Trace plot for posterior copy matrix	53
14	Simulation - QQ plot for p_{ij}	54
15	Simulation - Snow flake plots	55

1.0 INTRODUCTION

In computational neuroscience, analyzing simultaneously recorded multiple spike trains is a relatively new problem. A spike is one firing of a neuron and is usually considered to be a point event. A spike train is the sequence of spikes over time from one neuron and reflects the time-course of that neuron's activity. Since neurons interact with neighboring neurons, studying the nature of their interactions is of particular interest. A typical interaction is thought to be that firings of one neuron may be excited or inhibited by firings of other neurons. Simultaneously recorded multiple spike trains from multiple neurons reflect the time-course of the activity of a neuronal group which can be used to study the interactions among neurons.

For one neuron, statistical inference is well established by modeling a single spike train as a one-dimensional point process [3]. However, for multiple neurons, statistical inference becomes difficult if modeling multiple spike trains as a high-dimensional point process. Major difficulties are due to that the richness of interaction patterns among neurons, and the curse of dimensionality, or the typical paucity of data in the high-dimensional spaces involved.

Several simplifications for the interaction pattern have been proposed. Gerstein, et al. [5] used an interaction pattern similar to a directed acyclic graph (DAG) in a Bayesian belief network by only considering excitation effects. They applied a modified gravitational clustering algorithm on multiple spike trains. Using their method, neurons directly or indirectly connected by excitation were visually clumped into a single cluster. Rigat, et al. [10] used dynamic Bayesian network (DBN) on binned spike trains. Their method treated the neuronal interaction as a directed graph and could estimate the effect of excitation and/or inhibition.

In our probabilistic approach, to analyze GCA mathematically, we modeled spike trains

and their interactions using Poisson processes, and obtained large sample properties of GCA. And in our statistical approach, we adopted the interaction pattern from Gerstein, et al. [5] by creating a hidden neuron (we called it an environment) which accounts for all other factors beyond the excitation between recorded neurons. The DAG (directed cyclic graph) network like Bayesian belief network was difficult to model, so we proposed a semi-DAG network structure by enforcing an upper bound on the probability of forming (directed) rings. The excitation effect among neurons is parametrized by a copy matrix $(p_{ij})_{K \times K}$: p_{ij} is the probability that neuron i is triggered by (or copied with random disturbance from) neuron j . Thus, for each row of $(p_{ij})_{K \times K}$, a mixture model is applied. In order to link all these mixture models together and enforce a semi-DAG structure for the excitation effect, we generalized Dirichlet distribution to a matrix Dirichlet distribution. This generalization and further consideration of model identifiability led to several difficulties in implementing MCMC sampling. Thus, instead of the Metropolis-Hastings method, we designed a weighted sampler which controlled the sampling precision in a given finite number of steps for distributions with bounded support.

The remainder of this thesis is organized as follows. Chapter 2 reviews the literature on multiple spike train analysis, with a focus on GCA and DBN. Chapter 3 contains a theoretical study of GCA. Chapter 4 gives details of how to apply generalized mixture models to multiple spike trains. Chapter 5 discusses sampling problems and summarizes a simulation study.

2.0 LITERATURE REVIEW

The nervous system is very complex. Scientists are studying it from different angles. The quantitative analysis of the nervous system's (firing) activity is an important aspect of computational neuroscience. These studies are done at many levels of resolution. For the study of the dynamics of a single spike, see Koch [9] and Hodgkin, et al. [6]. For a single spike train, see Brown [3], for a review of one-dimensional point process models, for example, inhomogeneous Poisson and Gamma processes, and the inverse Gaussian process. For the study of two spike trains, Kass, et al. [12] used BARS (Bayesian Adaptive Regression Splines) to construct a test for the difference between their firing intensities. Kass, et al. [8] also considered the interaction between 2 spike trains by BARS. These methods focused on studying the firing intensities of neurons.

Brillinger [1] studied a multiple point process approach for a small group of neurons. Utikal [11] used empirical process to analyze the dependence between one spike train and other spike trains. Czanner et al. [4] developed a probabilistic model for the snowflake plot which is a powerful tool for depicting the joint activity of three neurons. Figure 15 shows snowflake plots on three of the neurons we used in our simulation studies. Most of these methods are intended to test the independence between one spike train and other spike trains.

As pointed out in Iyengar [7] and Brown, et al. [2], one problem in analyzing multiple spike trains was how to define an interaction pattern among neurons. Following this line, Rigat, et al. [10] proposed a dynamic Bayesian network (DBN) model to analyze interactions among multiple spike trains. There are two difficulties in applying this method: the process of binning spike trains into binary sequences can introduce noise; and identifiability of model parameters was not considered. However, the idea of applying Bayesian belief networks to

describe the interaction among neurons remains a very attractive approach.

Gerstein et al. [5] proposed an interaction pattern which only takes into account excitation effects, and applied a gravitational clustering algorithm (GCA) to qualitatively analyze this pattern. GCA is a numerical algorithm in pattern recognition whose aim in this context is to identify distinct independent neuronal sub-groups in a larger collection of simultaneously recorded neurons. This method requires the transformation of multiple spike trains into multiple longitudinal curves. Although our simulation studies show some oscillations of this method in the long run (i.e., there may be unstable clusterings), the transformed longitudinal curves do give more flexibility in statistical modeling. For example, it allows for the use of mixture models.

2.1 DYNAMIC BAYESIAN BELIEF NETWORK

We start with definitions related to the source data: a spike train is conveniently modeled as a point process for many purposes, but not for DBN or GCA, so we consider transformations by binning and by charge processes.

Data Type 1 (Spike Train[†]). $S = \{t_i > 0, t_{i+1} > t_i\}_{i \in \mathcal{N}}$ is said to be a spike train with i^{th} spike time t_i . It is a point process on \mathcal{R}^+ . We let $\{n_s(t)\}_{t \in \mathcal{R}^+}$ to be its corresponding counting process.

Generally there are many probabilistic models for S ; in this thesis, the basis of our simulations is the Poisson process with intensity r , $\mathcal{PP} \equiv \{s = \{t_i\}_{i \in \mathcal{N}}, \{n_s(t)\}_{t \in \mathcal{R}^+}, r\}$.

Data Type 2 (Binned Spike Train). A binary sequence $\Delta_s = \{\delta_1, \delta_2, \dots, \delta_M\}$ is a binned spike train of a spike train $S_{[0,T]} = \{t_1, t_2, \dots, t_N\}$ on $[0, T)$ if it is determined thus:

1. there is a finite partition $0 = \frac{0}{J}T < \frac{1}{J}T < \dots < \frac{J}{J}T = T$ of $[0, T)$: each time interval is considered to be a discrete time point;
2. δ_j is defined as $\sum_{i=1}^N I_{[\frac{j-1}{J}T, \frac{j}{J}T)}(t_i)$: $\delta_j = 1$ means a spike at j^{th} discrete time point and $\delta_j = 0$ means no spikes at j^{th} discrete time point.

† Figure 10 is a raster plot for 5 simulated spike trains.

To guarantee $\delta_j < 2$, usually $\frac{T}{j}$ is set to be 1 or 2 ms due to the typical neuron's refractory period. When there are several neurons, $\{\delta_{kj}\}_j$ denotes k^{th} neuron's binned spike train.

Definition 1 (DBN interaction pattern). *This interaction can be fully described by the following three components:*

1. an interaction matrix $B = (\beta_{kl})_{K \times K}$, where β_{kl} is the effect from l^{th} neuron to k^{th} neuron, with β_{kk} the self-refractory effect;
2. Baseline (or Spontaneous) activity vector $\Theta = (\theta_1, \theta_2, \dots, \theta_K)^T$: here θ_k describes the spontaneous activity of neuron k if it were isolated from the current neuron group and ignoring the self-refractory effect;
3. Linkage function: let $\pi_{kj} = \Pr\{\delta_{kj} = 1 | \text{history before time } j\}$ and $\tau_{kj} = \max\{l < j | \delta_{kl} = 1\}$: π_{kj} is k^{th} neuron's firing probability at time j conditioned on all neurons' histories before time j ; τ_{kj} is the last firing time of the k^{th} neuron before time j .

$$\log \left(\frac{\pi_{kj}}{1 - \pi_{kj}} \right) = \theta_k + \sum_{h=1}^K \beta_{kh} \left(\frac{\sum_{w=\tau_{kj}}^{j-1} \delta_{hw}}{j - \tau_{kj}} \right). \quad (2.1)$$

This definition is not yet a complete description of the interaction. To complete it, Rigat, et al. [10] assumed the independence of all such conditional probabilities to get

$$\Pr(\Delta | \Theta, B) = \prod_{k=1}^K \prod_{j=1}^M \pi_{kj}^{\delta_{kj}} (1 - \pi_{kj})^{1 - \delta_{kj}}. \quad (2.2)$$

Model Fitting Method 1 (DBN). *Bayesian method: $\text{Prior}(\Theta, B) \xrightarrow{\text{sampling}} \text{Posterior}(\Theta, B)$.*

This DBN model estimates all pairwise interactions by B , but it is different from the classic probabilistic network (Bayesian belief network, BBN) in an important way: if a directed graph is used to show interactions among neurons, rings are permitted in DBN but prohibited in BBN. Although Rigat, et al. [10] did not deal with the problem of parameter identifiability, they did recommend the use of interpretable parameters in terms of the biological mechanisms of neuronal firing and validated their model by simulation.

2.2 GRAVITATIONAL CLUSTERING ALGORITHM

In contrast to the DBN approach, GCA is not a network model. Rather, it is a clustering algorithm with an embedded dynamic system. A cluster is thought to be a neuronal sub-group with (direct or indirect) excitatory connections; thus, two clusters mean two unconnected sub-graphs. We now give the formal definitions.

Data Type 3 (Charge Process[‡]). A charge process $q_s(t)$ of a spike train $S_{[0,T]}$ is

$$q_s(t) = q_0 I_{\{n_s(t) > 0\}} \sum_{i=1}^{n_s(t)} e^{-\frac{t-t_i}{\tau}} - \bar{q}(t), \quad (2.3)$$

where q_0 is a positive constant, τ is the time constant which describes the exponential decay of a spike's excitation effect, and $\bar{q}(t)$ is a normalization function which makes $\mathbf{E}q_s(t) = 0$.

The normalizing function $\bar{q}(t)$ is usually constant, especially under stationary conditions. This charge process is a transformation of spike train: $q_s(t)$ is the active effect at time t . The sign of $q_s(t)$ in this context has the following interpretation: a neuron with spike train s excites other neurons if $q_s(t) > 0$ and is excited by other neurons if $q_s(t) < 0$.

Definition 2 (GCA interaction mode). This interaction mode also has three components:

1. a copy matrix[†] $P = (p_{kl})_{K \times K}$, $p_{kl} = \Pr(\text{spikes in } k^{\text{th}} \text{ spike train are copied from the } l^{\text{th}} \text{ spike train})$; this copying is done after a certain latency, ε , with a certain distribution, for example, a beta distribution on an interval $[a, b]$;
2. spontaneous spike trains $S^* = \{s_1^*, s_2^*, \dots, s_K^*\}$, each of whose spike trains model the activity if the corresponding neuron is isolated from its network;
3. a link which for neuron k , combine spikes copied from other neurons with the k^{th} neuron's spontaneous spike train, $s_k = \text{sort}\{s_k^*, \text{random copied spikes from } s_l \text{ with probability } p_{kl} \text{ and time lag } \varepsilon\}$

In this mode, we need some trigger neurons, for which $p_{.k} = 0$ or $s_k^* = s_k$. When we generate data, we first generate spikes for trigger neurons, then delete the trigger neurons and find new trigger neurons within remaining neuron subgroup; and repeat until no neurons

[‡] Figure 11 is an illustration of a charge process.

[†] Bottom table in Figure 9 is an example of copy matrix.

available. Note also that this interaction mode does not require charge processes, for it is about the firing rate or spikes. That is why we say that GCA is only a tool to analyze the interaction mode. The following GCA procedure is from Gerstein; it consists of three steps.

Model Fitting Method 2 (GCA). 1. *Trace calculation, which describes the neuron's trajectory in an embedded space:*

$$A(a, b) = \begin{cases} \frac{a-b}{|a-b|}, & |a - b| > c_0 \\ 0, & |a - b| \leq c_0 \end{cases} \quad (2.4)$$

$$f_k(t) = \sum_{l \neq k} q_l(t) A(x_l(t), x_k(t)) \quad (2.5)$$

$$\frac{dx_k(t)}{dt} = \sigma q_k(t) f_k(t) \quad (2.6)$$

Here, $x_k(t)$ is the location of the representation of neuron k in R^K at time t , $f_k(t)$ is the force field at the k^{th} neuron's position at time t , and $A(\cdot, \cdot)$ is an attenuation function. In fact, these equations define a dynamical system, the traces x_k of which are irregular, and the speed of those traces provide information about excitatory effects within a cluster.

2. *Distance measure[†] (Euclidean): the shorter of the distance between $x_k(t)$ and $x_l(t)$, the larger probability that k^{th} neuron and l^{th} neuron are within the same cluster.*

$$d_{kl}(t) = |x_k(t) - x_l(t)| \quad (2.7)$$

3. *Cluster recognition: The basic approach to the construction of distinct clusters is to view the traces in K -dim Euclidean space: if some traces get closer with time, we say that they are in one cluster; if they remain far apart, we say that they are in different clusters. Usually it is carried out by monitoring the collection of pairwise distances d_{kl} because of the difficulties of visualization in high dimensions.*

[†] Figure 12 is a plot of Distances by Time.

Gerstein showed that this procedure is powerful for detecting networks when individual firing rates are equal, but for networks with unequal firing rates, some undesirable oscillations occurred in their simulations. Our research on GCA is mainly focused on how to improve its sensitivity and how to visualize the traces.

2.3 COMPARISON OF DBN AND GCA

Strictly speaking, methods other than DBN and GCA mentioned in Chapter 2 are essentially designed for a small group of neurons. Currently only DBN and GCA can be applied to analyze multiple spike trains at the scale of tens or larger. However, they are very different in many aspects: their interaction mode, mathematical tools, and simulation results. However, both of them share the same limitations: they both require transformation of the original data, and their theoretical foundations are still not well studied. In this section we qualitatively study their differences.

2.3.1 Interaction Modes

GCA's interaction mode is suggested by Gerstein based on his experience in neuroscience. It can be viewed as a natural generalization of a widely-accepted method (crosscorrelogram, or its dynamic version, the joint peristimulus time histogram (JPSTH)) which are standard tools for analyzing two neurons; in contrast, DBN's interaction mode originates from BBN, which is more flexible than GCA. Other contrasts are the following:

1. Graph in DBN is Directed Cyclic Graph (DCG), in GCA is DAG;
2. In DBN, interaction includes excitation and inhibition; in Gerstein's implementation of GCA, only excitation is modeled;
3. In DBN, interaction takes effect by increasing or decreasing the temporal firing rate of triggered nodes; in GCA, interaction takes effect by directly copying a spike with a random time lag, so we can tell which spikes are from others, which are self generated;

4. In DBN, interactions between different pairs have different effective time ranges; in GCA, all interactions have the same latency distribution;
5. DBN's interaction mode is not as straightforward as GCA's interaction mode.

2.3.2 Mathematical aspects

Since GCA originated from pattern recognition and DBN originated from Bayesian methods (BBN originated from machine learning), they differ considerably in their mathematical aspects.

- DBN
- a. uses the Bernoulli distribution to construct likelihood function;
 - b. uses Bayesian methods to estimated parameters;
 - c. uses MCMC to obtain predictions;
 - d. uses proportion of fitting (and prediction) residuals that cross a threshold to evaluate the goodness-of-fit; and
 - e. uses artificial spike sorting errors to test the robustnes of this model;
- GCA
- a. uses concepts from physics to construct a dynamic system in high-dimensional Euclidean space;
 - b. uses numerical integration to obtain traces of this dynamic system;
 - c. uses a geometrical properties of those traces (Euclidean distances among them) to obtain clusters;
 - d. it makes no predictions because it does not try to fit data;
 - e. there areno goodness-of-fit tests because, once again, it does not try to fit data; and
 - f. there is no evaluation of robustness.

2.3.3 Results

1. GCA reads in data as time evolves; DBN reads in all data at the start;
2. GCA starts with a graph, and ends with clusters; DBN starts from graph, and ends with a graph;

3. GCA is not robust for networks with non unequal rates, according to my simulation; DBN is more robust, but has different performance for different rates, according to Rigat, et al. [10];
4. GCA's behavior is not well understand in theory, so much work left to be done; DBN has considerable theoretical support from Bayesian statistics; and
5. usually, GCA only requires record within one minute, DBN requires at least 10 minutes' record.

3.0 THEORETICAL ASPECTS OF GCA

GCA clusters spike trains using the following method: each neuron is represented by a virtual point in K -dimensional space (K is the number of neurons); initially, all pairs of points are equidistant; the products of paired charge processes will generate an uneven force field in this space; these virtual points' speeds are proportional to the force field; the final geometrical configuration of these points, along with cluster index determine the clustering. Mathematically, the traces of these virtual points are a K -dimensional non-autonomous dynamical system with random coefficients, so in general it is hard to draw a flow plot.

In this section, we mainly study $\mathbf{E}(x_k(t) - x_k(0))$: if $K = 2$, this is equivalent to $d_{12}(t)$; if $K > 2$, this will give us a draft picture of the geometrical pattern. There are two key aspects of this work: the first concerns the properties about charge processes and their products (in section 3.0.4 and 3.0.5); the second deals with how to calculate the integration of those products (in section 3.0.6 and 3.0.7).

3.0.4 Preliminaries

In this part, we expand on definition 2 and list some simple properties of related processes.

Definition 3 (Copy probability). $\mathcal{PP}_1 = \{s_1 = \{t_{1,i}\}_{i \in \mathcal{N}}, \{n_1(t)\}_{t \in \mathcal{R}^+}, r_1\}$ is said to be copied to \mathcal{MP}_2 with probability π and latency ε , if and only if

- (i) there exists a Poisson process $\mathcal{PP}_2^* = \{s_2^* = \{t_{2,i}^*\}_{i \in \mathcal{N}}, \{n_2^*(t)\}_{t \in \mathcal{R}^+}, r_2^*\}$ which is independent of \mathcal{PP}_1 ;
- (ii) there exists an i.i.d. sequence ε_i with distribution $U[0, \varepsilon_0]$, independent of all others; there exists another i.i.d. sequence ζ_i with distribution $\text{Bin}(1, \pi)$, independent of all others;

(iii) define a local copied process from \mathcal{PP}_1 ($t > \varepsilon_0$) thus:

$$\begin{aligned}
n_{\tilde{s}}(t) &\equiv \underbrace{I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \sum_{i=1}^{n_1(t - \varepsilon_0)} \zeta_i}_{n_{\tilde{s},1}(t)} \\
&\quad + \underbrace{I_{\mathcal{N}}((n_1(t)) - (n_1(t - \varepsilon_0))) \sum_{i=n_1(t - \varepsilon_0) + 1}^{n_1(t)} \zeta_i I_{[0,t]}(t_{1,i} + \varepsilon_i)}_{n_{\tilde{s},2}(t)} \quad (3.1)
\end{aligned}$$

$$\tilde{s}(t) \equiv \text{sort}\{t_{1,i} + \varepsilon_i \mid \zeta_i = 1, t_{1,i} + \varepsilon_i < t, i \in \mathcal{N}\} \quad (3.2)$$

$$\mathcal{LCP} = \{\tilde{s}, \{n_{\tilde{s}}(t)\}_{t \in \mathcal{R}^+}, r_s, U[0, \varepsilon_0]\} \quad (3.3)$$

here $\{n_{\tilde{s}}(t)\}_{t \in \mathcal{R}^+}$ is the counting process of point process \tilde{s} , and the intensity r_s is no longer a constant;

(iv) \mathcal{MP}_2 is the combination of \mathcal{PP}_2^* and \mathcal{LCP} :

$$\begin{aligned}
s_2(t) &\equiv \text{sort}\{s_2^*(t), \tilde{s}(t)\} \\
n_2(t) &\equiv n_2^*(t) + n_{\tilde{s}}(t) \\
r_2 &\equiv r_2(r_2^*, r_s, \varepsilon_0) \\
\mathcal{MP}_2 &= \{s_2 = \{t_{2,i}\}_{i \in \mathcal{N}}, \{n_2(t)\}_{t \in \mathcal{R}^+}, r_2\} \quad (3.4)
\end{aligned}$$

Lemma 1 (some properties of Poisson process). *For a standard Poisson process \mathcal{PP}_1 , we have the following properties*

- (i) $t_1, \dots, t_n \mid N(t) = n$ have the same distribution as order statistics of n iid $U(0, t)$ variates;
- (ii) Poisson process has stationary independent increments;

Lemma 2 (Properties of \mathcal{LCP}). *The process in formula (3.1) has the following properties:*

- (i) $n_{\tilde{s},1}$ is independent of $n_{\tilde{s},2}$;
- (ii) $n_{\tilde{s},1}$ is a Poisson process starting at time ε_0 and with intensity πr_1 ;
- (iii) $n_{\tilde{s},2}$ has distribution $\text{Poisson}(\frac{\pi}{2} r_1 \varepsilon_0)$.

Note that \mathcal{LCP} is not a Poisson process because it has dependent increments for $n_{\tilde{s}}$, with $r_s = r_s(t)$ in this case.

Lemma 3 (Charge process of \mathcal{LCP}). *First, the charge process of the point process in formula (3.1) to be (for $t > \varepsilon_0$) is the sum of the two components:*

$$q_{n_{\bar{s},1}}(t) \equiv I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \sum_{i=1}^{n_1(t - \varepsilon_0)} \zeta_i e^{-\frac{t - (t_{1,i} + \varepsilon_i)}{\tau}} - \pi \frac{\tau}{\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) r_1 \tau (1 - e^{-\frac{t - \varepsilon_0}{\tau}}) \quad (3.5)$$

$$\begin{aligned} q_{n_{\bar{s},2}}(t) &\equiv I_{\mathcal{N}}(n_1(t) - n_1(t - \varepsilon_0)) \sum_{i=n_1(t - \varepsilon_0) + 1}^{n_1(t)} \zeta_i I_{[0,t]}(t_{1,i} + \varepsilon_i) e^{-\frac{t - (t_{1,i} + \varepsilon_i)}{\tau}} \\ &- \pi r_1 \varepsilon_0 \left(\frac{\tau}{\varepsilon_0}\right)^2 \left(e^{-\frac{\varepsilon_0}{\tau}} - \left(1 - \frac{\varepsilon_0}{\tau}\right)\right) \\ &\stackrel{d}{=} I_{\mathcal{N}}(\tilde{n}_1(\varepsilon_0)) \sum_{i=1}^{\tilde{n}_1(\varepsilon_0)} \tilde{\zeta}_i I_{[0,\varepsilon_0]}(\tilde{t}_{1,i} + \tilde{\varepsilon}_i) e^{-\frac{\varepsilon_0 - (\tilde{t}_{1,i} + \tilde{\varepsilon}_i)}{\tau}} - \pi r_1 \varepsilon_0 \left(\frac{\tau}{\varepsilon_0}\right)^2 \left(e^{-\frac{\varepsilon_0}{\tau}} - \left(1 - \frac{\varepsilon_0}{\tau}\right)\right). \end{aligned}$$

In that case, we have

$$\begin{aligned} (i) \quad \mathbf{E}q_{\bar{s},1}(t) &= 0, \quad \mathbf{E}q_{\bar{s},1}^2(t) = \frac{\pi r_1 \tau^2}{4\varepsilon_0} (1 - e^{-\frac{2\varepsilon_0}{\tau}}) (1 - e^{-\frac{2(t - \varepsilon_0)}{\tau}}); \\ (ii) \quad \mathbf{E}q_{\bar{s},2}(t) &= 0. \end{aligned}$$

Finally, we introduce the following notation:

$$\begin{aligned} q_{\mathcal{LCP}}(t) &= q_{n_{\bar{s},1}}(t) + q_{n_{\bar{s},2}}(t) \\ q_{\mathcal{MP}}(t) &= q_{\mathcal{LCP}}(t) + \underbrace{I_{\mathcal{N}}(n_2^*(t)) \sum_{i=1}^{n_2^*(t)} e^{-\frac{t - t_{2,i}^*}{\tau}} - r_2^* \tau (1 - e^{-\frac{t}{\tau}})}_{q_2^*(t)} \end{aligned} \quad (3.6)$$

3.0.5 Moment properties of charge processes

This part is used to list moments of charge processes, which will be used in calculate GCAD.

Lemma 4 (Moments of charge process). *Let $s = \{t_i\}_{i \geq 1}$ be a Poisson process with intensity r and $\{n_s(t)\}_{t > 0}$ be its counting process, $q_s(t)$ its charge process as defined in Datatype 2.3:*

$$q_s(t) = q_0 I_{\{n_s(t) > 0\}} \sum_{i=1}^{n_s(t)} e^{-\frac{t - t_i}{\tau}} - \bar{q}(t);$$

for convenience, we set $q_0 = 1$ and $\bar{q}(t) = r\tau(1 - e^{-\frac{t}{\tau}})$, for $t, \delta > 0$. Then we have

$$\begin{aligned} (i) \quad \mathbf{E}q_s(t) &= 0. \\ (ii) \quad q_s(t + \delta) &\stackrel{d}{=} e^{-\frac{\delta}{\tau}} q_s(t) + \tilde{q}_s(\delta), \text{ where } \tilde{q}_s(t) \text{ is an i.i.d. copy of } q_s(t). \end{aligned}$$

(iii) $\mathbf{E}q_s^2(t) = \frac{r\tau}{2}(1 - e^{-\frac{2t}{\tau}})$, $c(t, \delta) \equiv \text{Cov}(q_s(t), q_s(t + \delta)) = e^{-\frac{\delta}{\tau}}\mathbf{E}q_s^2(t)$.

Lemma 5 (Moments of product of charge processes). *Let $q_1(t)$, $q_2(t)$ be two charge processes defined as in formula 2.3, and define $q_{1,2}(t) \equiv q_1(t)q_2(t)$,*

(i) *if $q_1(t)$ and $q_2(t)$ are independent and as defined in Lemma 4, then $\mathbf{E}q_{1,2}(t) = 0, \mathbf{E}q_{1,2}^2(t) = \mathbf{E}q_1^2(t)\mathbf{E}q_2^2(t)$;*

(ii) *if $q_1(t)$ and $q_2(t)$ are charge processes of \mathcal{PP}_1 and \mathcal{MP}_2 as in Definition 3, then $\mathbf{E}q_{1,2}(t) = \frac{r_1\pi\tau^2}{2\varepsilon_0}(1 - e^{-\frac{\varepsilon_0}{\tau}})(1 - e^{-\frac{2t-\varepsilon_0}{\tau}})$.*

3.0.6 Main results about GCAD (K=2)

In Theorem 1, we study the integration of charge process product; Theorem 2 is a simple application of Theorem 1. They state precisely the following facts for $K = 2$: if two processes are independent, the expectation of GCAD's deduction will be 0, but the variance of GCAD's deduction will increase over time; if two processes are dependent, the expectation of GCAD's deduction will increase with time. Proofs are in the Appendix.

Theorem 1 (2-dimension). *Let $q_1(t)$ and $q_2(t)$ be two charge processes defined as in formula 2.3, now define*

$$Q_{1,2,N}(t) \equiv \sum_{n=1}^N q_1\left(\frac{n}{N}t\right)q_2\left(\frac{n}{N}t\right)\frac{t}{N}$$

(i) *if $q_1(t)$ and $q_2(t)$ are defined as in Lemma 4 and independent (with intensity r_1 and r_2), then*

$$\mathbf{E}Q_{1,2,N}(t) = 0 \tag{3.7}$$

$$\lim_{N \rightarrow \infty} \mathbf{E}Q_{1,2,N}^2(t) \cong \frac{r_1r_2}{8}\tau^4\left(\frac{2t}{\tau} - 2.5\right) \tag{3.8}$$

(ii) *if $q_1(t)$ is copied to $q_2(t)$ with probability π and latency distribution $U[0, \varepsilon_0]$ (as defined in Lemma 5-(ii)), then*

$$\lim_{N \rightarrow \infty} \mathbf{E}Q_{1,2,N}(t) \cong \frac{\pi r_1\tau^2}{2\varepsilon_0}(1 - e^{-\frac{\varepsilon_0}{\tau}})\left(t - \varepsilon_0 - \frac{\tau}{2}e^{-\frac{\varepsilon_0}{\tau}}\right) \tag{3.9}$$

Theorem 2 (GCAD (K=2)). *For 2-dimensional GCA model defined by (2.3) and (2.4)-(2.6), let $h_1 \equiv \sup\{t \geq 0 | d_{12}(t) > c_0\}$, we have*

(i) if $n_1(t)$ and $n_2(t)$ are two independent Poisson process with intensity r_1 and r_2 , $d_{12}(0) \gg c_0$, then for $t < h_1$

$$\mathbf{E}d_{12}(t) = d_{12}(0)$$

$$\mathbf{Vard}_{12}(t) \cong r_1 r_2 \tau^3 t, \quad \text{if } t \gg \tau$$

(ii) if $n_1(t)$ and $n_2(t)$ as defined in Definition (3), $h_1 > t \gg \tau \gg \varepsilon_0$, then

$$\mathbf{E}d_{12}(t) \downarrow c_0 \quad (t \uparrow h_1)$$

From these theorems, we arrive at the following conclusions: the GCAD's expectation will be unchanged if two spike trains are independent; GCAD's expectation will decrease to c_0 if two spike trains are dependent. However, GCAD will become unreliable for large times because the variance of the distance increase with time; hence, the GCAD should be adjusted with time.

3.0.7 Main results about GCAD ($K > 2$)

We now turn to the analogous results for the case ($K > 2$). In order to do so, we need to introduce the notion of the depth of a copy matrix.

Definition 4 (Depth of copy matrix P).

$$D(P) \equiv \min\{n \geq 1 | P^n = 0\}$$

In order to use previous conclusions in section 3.0.5, we restrict our discussion to that each process is either a Poisson process or a mixture of Poisson process and several \mathcal{LCP} s, that is, $D(P) \leq 2$. In fact, in Theorem 2-(i), $D(P) = 1$; in Theorem 2-(ii), $D(P) = 2$. Next, we introduce additional notation and assumptions:

- (i) henceforth, assume that $|q_k(\cdot)| < M < +\infty$, $1 \leq k \leq K$;
- (ii) fix $t, \rho, \tau > 0$, $\forall \delta_0 \in (0, 1)$, and set $d_0 = \frac{2(K-1)M^2}{\delta_0} \rho \tau$, $N(\delta_0, t) = \frac{t}{\rho \tau}$;
- (iii) let $q_i(t)$ ($i = 1, 2, \dots, K$) be the charge process defined in (2.3), and let

$$\mathcal{F}_t \equiv \sigma\{q_i(s) | 1 \leq i \leq K, 0 \leq s \leq t\};$$

(iv) define the unit vector ($1 \leq i \neq j \leq K$), $\vec{1}_{ij}(t) \equiv \frac{x_j(t) - x_i(t)}{|x_j(t) - x_i(t)|}$;

(v) define a sum for $1 \leq m \leq N$ ($f_{ij,N}(0) = 0$), a generalization of $Q_{1,2,N}(t)$

$$f_{ij,N}\left(\frac{m}{N}t\right) \equiv \frac{t}{N} \sum_{n=1}^m q_{ij}\left(\frac{n-1}{N}t\right) \vec{1}_{ij}\left(\frac{n-1}{N}t\right);$$

(vi) for now, we focus only on $x_i(\cdot)$, which is a sum of finitely many ($N = \frac{t}{\rho\tau}$) terms:

$$x_i\left(\frac{m}{N}t\right) \equiv x_i(0) + \frac{t}{N} \sum_{n=1}^m \sum_{k \neq i} q_{ik}\left(\frac{n-1}{N}t\right) \vec{1}_{ik}\left(\frac{n-1}{N}t\right), \quad 1 \leq i \leq K \quad (3.10)$$

(vii) next, let

$$C_i\left(\frac{m}{N}t\right) \equiv \sum_{k \neq i} q_{ik}\left(\frac{m}{N}t\right) \vec{1}_{ik}\left(\frac{m}{N}t\right), \quad D_{ij}\left(\frac{m}{N}t\right) \equiv \frac{C_j\left(\frac{m}{N}t\right) - C_i\left(\frac{m}{N}t\right)}{|\vec{x}_i \vec{x}_j\left(\frac{m}{N}t\right)|};$$

(viii) finally, define a series for $n \in \mathcal{N}^*$ and $0 < u < 1$, $\rho_0 = 0, \rho_{n+1} = u(1 + \rho_n)$.

With above notations, we list several simple facts:

F1. $f_{ij}\left(\frac{m}{N}t\right), \vec{1}_{ij}\left(\frac{m}{N}t\right), x_i\left(\frac{m}{N}t\right) \in \mathcal{F}_{\frac{m-1}{N}t}$, $q_{ij}\left(\frac{m}{N}t\right), C_i\left(\frac{m}{N}t\right), D_{ij}\left(\frac{m}{N}t\right) \in \mathcal{F}_{\frac{m}{N}t}$;

F2. if $q_i(\cdot)$ is independent from $q_j(\cdot)$, then $\mathbf{E} \left[q_{ij}\left(\frac{m}{N}t\right) | \mathcal{F}_{\frac{m-1}{N}t} \right] = e^{-\frac{1}{\tau} \frac{2t}{N}} q_{ij}\left(\frac{m-1}{N}t\right)$;

F3. if $q_i(\cdot)$ and $q_j(\cdot)$ are defined as in Lemma 5-(ii), further assume $\varepsilon_0 = \frac{t}{N}$, then

$$\mathbf{E} \left[q_{ij}\left(\frac{m}{N}t\right) | \mathcal{F}_{\frac{m-1}{N}t} \right] = e^{-\frac{1}{\tau} \frac{2t}{N}} \left[q_{ij}\left(\frac{m-1}{N}t\right) + \overbrace{e^{\frac{1}{\tau} \frac{2t}{N}} \frac{r_1 \pi \tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}})^2}^{c_1(ij)} \right];$$

F4. for $n \in \mathcal{N}^*$, $\rho_n = \frac{u(1-u^n)}{1-u}$, $\sum_{i=0}^n \rho_i = \frac{u}{1-u} \left[n + 1 - \frac{1-u^{n+1}}{1-u} \right]$;

F5.

$$\vec{1}_{ij}\left(\frac{m}{N}t\right) = \frac{\vec{1}_{ij}\left(\frac{m-1}{N}t\right) + \frac{t}{N} D_{ij}\left(\frac{m-1}{N}t\right)}{|\vec{1}_{ij}\left(\frac{m-1}{N}t\right) + \frac{t}{N} D_{ij}\left(\frac{m-1}{N}t\right)|},$$

$$\tilde{O}^{(1)}\left(\frac{m-1}{N}t\right) \equiv \frac{1}{|\vec{1}_{ij}\left(\frac{m-1}{N}t\right) + \frac{t}{N} D_{ij}\left(\frac{m-1}{N}t\right)|} - 1;$$

F6. assuming that $|\vec{x}_i \vec{x}_j(\cdot)| \geq d_0$ for all following content, then we have

$$|D_{ij}(\cdot) \frac{t}{N}| \leq \delta_0, \quad |\tilde{O}^{(1)}(\cdot)| \leq \frac{\delta_0}{1 - \delta_0}. \quad (3.11)$$

Theorems 3-4 are the counterpart of Theorems 1-2, for higher dimensions. Theorem 4 is a simple application of Theorem 3. They show the link between clusters and gravitational traces: within each cluster, corresponding traces move toward each other; between clusters, cluster centers don't move. Compared to section 3.0.6, we have stronger assumptions but weaker conclusions.

Theorem 3 ($\mathbf{E}(f_{ij,N}(t)|\mathcal{F}_{\frac{m}{N}t})$). *If $|\overrightarrow{x_i x_j}(0)| > d_0$, let $h_{ij}(d_0) = \sup\{t \geq 0 | d_{ij}(t) > d_0\}$ and $u = e^{-\frac{1}{\tau} \frac{2t}{N}}$, then we have*

(i) *if $n_i(t)$ and $n_j(t)$ are two independent Poisson process with intensities r_i and r_j , respectively, then for $t < h_{ij}$ and $0 \leq m \leq N - 1$*

$$\begin{aligned} \mathbf{E}(f_{ij,N}(t)|\mathcal{F}_{\frac{m}{N}t}) &= f_{ij,N}(\frac{m}{N}t) + \frac{t}{N}q_{ij}(\frac{m}{N}t)\overrightarrow{1}_{ij}(\frac{m}{N}t)(1 + \rho_{N-1-m}) \\ &\quad + \mathbf{E}\left\{\sum_{n=m}^{N-1}\left[O^{(1)}(\frac{n}{N}t) + O^{(2)}(\frac{n}{N}t)\right]|\mathcal{F}_{\frac{m}{N}t}\right\} \end{aligned} \quad (3.12)$$

$$O^{(1)}(\frac{n}{N}t) = \frac{t}{N}q_{ij}(\frac{n}{N}t)\overrightarrow{1}_{ij}(\frac{n}{N}t)\rho_{N-1-n}\tilde{O}^{(1)}(\frac{n}{N}t) \quad (3.13)$$

$$O^{(2)}(\frac{n}{N}t) = \frac{t}{N}q_{ij}(\frac{n}{N}t)\rho_{N-1-n}\frac{\frac{t}{N}D_{ij}(\frac{n}{N}t)}{|\overrightarrow{1}_{ij}(\frac{n}{N}t) + \frac{t}{N}D_{ij}(\frac{n}{N}t)|} \quad (3.14)$$

(ii) *if $n_i(t)$ and $n_j(t)$ as defined in F.3, then for $t < h_{ij}$ and $0 \leq m \leq N - 1$*

$$\begin{aligned} \mathbf{E}(f_{ij,N}(t)|\mathcal{F}_{\frac{m}{N}t}) &= f_{ij,N}(\frac{m}{N}t) + \frac{t}{N}q_{ij}(\frac{m}{N}t)\overrightarrow{1}_{ij}(\frac{m}{N}t)(1 + \rho_{N-1-m}) \\ &\quad + \mathbf{E}\left\{\sum_{n=m}^{N-1}\left[O^{(1)}(\frac{n}{N}t) + O^{(2)}(\frac{n}{N}t)\right]|\mathcal{F}_{\frac{m}{N}t}\right\} \\ \frac{t}{N}c_1(ij)\overrightarrow{1}_{ij}(\frac{m}{N}t) \sum_{n=0}^{N-1-m} \rho_n &+ \mathbf{E}\left\{\sum_{n=m}^{N-1}\left[R^{(1)}(\frac{n}{N}t) + R^{(2)}(\frac{n}{N}t)\right]|\mathcal{F}_{\frac{m}{N}t}\right\} \end{aligned} \quad (3.15)$$

$$R^{(1)}(\frac{n}{N}t) = \frac{t}{N}\overrightarrow{1}_{ij}(\frac{n}{N}t)c_1(ij)\tilde{O}^{(1)}(\frac{n}{N}t)\sum_{l=0}^{N-1-n} \rho_l \quad (3.16)$$

$$R^{(2)}(\frac{n}{N}t) = \frac{t}{N}c_1(ij)\frac{\frac{t}{N}D_{ij}(\frac{n}{N}t)}{|\overrightarrow{1}_{ij}(\frac{n}{N}t) + \frac{t}{N}D_{ij}(\frac{n}{N}t)|}\sum_{l=0}^{N-1-n} \rho_l \quad (3.17)$$

$O^{(1)}(\cdot)$ and $O^{(2)}(\cdot)$ are defined as in (3.13) and (3.14).

Finally, note that $\mathbf{E}(f_{ij,N}(t)|\mathcal{F}_{\frac{0}{N}t}) = \mathbf{E}f_{ij,N}(t)$.

Theorem 4 (GCAD ($K > 2$), large inter-distance). Let $h_1(d_0) \equiv \bigwedge_{k \neq 1} h_{1k}(d_0)$, if $t < h_1(d_0)$ and fix $N = \frac{t}{\rho\tau}$, we have

(i) if $n_1(\cdot)$ is independent from $\{n_2(\cdot), \dots, n_K(\cdot)\}$, then

$$\lim_{d_0 \rightarrow +\infty} \mathbf{E} [x_1(t) - x_1(0)] = 0 \quad (3.18)$$

(ii) if $\{n_1(\cdot), \dots, n_{k_1}(\cdot)\}$ are independent from $\{n_{k_1+1}(\cdot), \dots, n_K(\cdot)\}$ and $\rho\tau = \varepsilon_0$, further assume $D(P) = 2$, then

$$\lim_{d_0 \rightarrow +\infty} \mathbf{E} [x_1(t) - x_1(0)] = \sum_{k=2}^{k_1} c_1(1k) \overrightarrow{1_{1k}}(0) \frac{u}{1-u} t \quad (3.19)$$

An important consequence of this theorem is immediate: if we put initial inter-neuron distances large enough, then in a particular time period, nodes within one cluster will move toward their center with steady velocities. However, the centers of each cluster will not move. However this result is only in the sense of its first moment; we have not yet studied in detail the 2^nd and higher order moments.

4.0 GENERALIZED MIXTURE MODEL FOR HIGH-DIMENSIONAL LONGITUDINAL DATA

In this chapter, we use a Bayesian model to quantitatively estimate the semi-DAG network of excitation pattern inspired by Gerstein, et al. [5]. This model is a generalization of mixture model from R^1 to R^K . It is also deeply involved in Bayesian belief network: by decreasing a control parameter[†] (a similar role of smoothing parameter in smoothing spline regression), this model will approximate a Bayesian belief network. In general, this model can be applied to study dynamics among high-dimensional longitudinal data.

4.1 THE PROBLEM

Our primary goal is to estimate the excitation pattern among multiple neurons. The pattern studied in this model is similar to Definition 2, but allow a small probability of forming rings. A copy matrix $(c_{kl})_{K \times K}$ is used to describe the excitation pattern as in Definition 2 and illustrated by a graph (see Figure 8). Nodes and (directed) edges correspond to neurons and excitation effect between neuron pairs. In his context, y_k denotes one neuron (or a node in graph) and y_{kt} is its activity at time t . Just as with binned spike trains, all y_{kt} 's are samples from corresponding charge processes at grid time points $t \in \{1, 2, \dots, T\}$. Using samples from the charge process curves gives more flexibility in modeling. A mixture model for y_k is used: with probability p_{kl} ($k \neq l$), y_{kt} is linearly predicted from $y_{l,t-1}$; with probability p_{kk} , y_{kt} is generated from environment. The semi-DAG network structure is described by enforcing a preset upper bound for all $p_{kl} \times p_{lk}$ s ($k \neq l$). Technically, the copy matrix

[†] This parameter will be specified in Definition 6.

$(p_{kl})_{K \times K}$ is given a matrix Dirichlet distribution (see Definition 6). Our model assumes Markov dependence for $y = (y_1, y_2, \dots, y_K)$.

4.2 STATEMENT OF GENERALIZED MIXTURE MODEL

In this section, we will give the details of our model. y_{kt} is the activity of neuron k at time t . $I_{\{\cdot\}}(\cdot)$ is the indicator function. $\lambda_{kt} = l (\neq k)$ means neuron k is excited by neuron l at time t ; $\lambda_{kt} = k$ means that neuron k is excited by environment at time t . Now, $\hat{y}_{kt} = \xi_{ki}(x_{it}^{(k)} - c_k) + \epsilon_{kt}$ is the linear model for predicting y_{kt} from $y_{i,t-1}$ (in our case, $x_{it}^{(k)} = y_{i,t-1}$). Next, $\sum_{j=1}^J I_{\{j\}}(d_{kt})e_{jt}$ is our model for the environment[§] which is a mixture of three normal distributions. Since the environment is defined to be the combination of random factors and unrecorded neurons which affect the activity of recorded neurons, it is assigned this mixture distribution. This model says that y_{kt} comes from two (mutually exclusive) sources: a simple linear regression (covariates can be added here) based on within-network excitation, or a component of environment. Let p_{ki} be the probability that y_k is excited by y_i , q_j be the weight of component in environment, and ϵ_{kt} be the observational error. Thus, $\{\xi_{ki}, c_k, \sigma^2, p_{ki}, q_j, m_j, s_j^2\}$ is our parameter set[‡], with the unobservable variables $\{\lambda_{kt}, d_{kt}\}$ \diamond .

This model clearly splits the activity of recorded neurons into two categories: within-network (recorded neurons) which is the interaction among recorded neurons; between-network-environment activity which is the interaction between recorded neurons and unrecorded neurons. By adjusting the complexity of environment, one can emphasize one category over another. Formally, the complete model is given by the following expressions:

$$y_{kt} = \sum_{i=1, i \neq k}^K I_{\{i\}}(\lambda_{kt}) \left[\xi_{ki}(x_{it}^{(k)} - c_k) + \epsilon_{kt} \right] + I_{\{k\}}(\lambda_{kt}) \sum_{j=1}^J I_{\{j\}}(d_{kt})e_{jt} \quad (4.1)$$

$$x_{it}^{(k)} \equiv y_{i(t-1)}; \quad k = 1, 2, \dots, K; \quad t \in \mathcal{T}_k \subseteq \mathcal{T} = \{1, 2, \dots, T\}.$$

§ This expression of environment creates a problem of unidentifiability under permutation.

‡ For model identifiability problem, $\xi_{ki} \geq 1$ and $c_k \geq 0$ are enforced in section 4.3.2.

◇ Their traces show the realization of random interactions at each time point.

$$y_{kt} | \lambda_{kt} = i \neq k, \xi_{ki}, x_{it}^{(k)}, c_k, \sigma^2 \propto (\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma^2} \frac{(y_{kt} - \xi_{ki}(x_{it}^{(k)} - c_k))^2}{2}} \quad (4.2)$$

$$y_{kt} | \lambda_{kt} = k, d_{kt} = j, m_j, s_j^2 \propto (s_j^2)^{-\frac{1}{2}} e^{-\frac{1}{s_j^2} \frac{(y_{kt} - m_j)^2}{2}} \quad (4.3)$$

$$\mathbf{P}(\lambda_{kt} = i | p_{k1}, p_{k2}, \dots, p_{kK}) = p_{ki} \quad (4.4)$$

$$\mathbf{P}(d_{kt} = j | q_1, \dots, q_J) = q_j \quad (4.5)$$

Model 4.1 has K mixture models which share a common observational error distribution and a common environment distribution. Since the studied excitation pattern is a semi-DAG, we need to enforce some relationships among those mixture models. This is realized by using a matrix Dirichlet distribution[†] for the copy matrix $P = (p_{ki})_{K \times K}$ instead of K separate distributions for P 's K rows. This distribution is the key point to model a semi-DAG network structure by parameters. However, this generalization leads to several technical difficulties, which we address in the next section. By adjusting the matrix Dirichlet distribution, one can get a network structure more or less similar to a DAG in Bayesian belief network.

4.3 BAYESIAN APPROACH TO COMPUTING THE POSTERIOR

In this section, we use the usual Bayesian method to fit our model: first, write out the likelihood; specify priors on the parameters; obtain all conditional posteriors; run a Markov chain Monte Carlo (MCMC) to get the posteriors; finally, use posterior mean and standard deviation to get parameter estimates. We now organize our Bayesian procedure as preparations, prior distributions, likelihood functions, conditional posteriors and Gibbs sampler procedure.

[†] In Definition 6, it is described as: $p_{ki}p_{ik} < c_0$ if $i \neq k$. c_0 is the control parameter which is used to approximate an ideal case $p_{ki}p_{ik} = 0$ (if $i \neq k$) by letting $c_0 \downarrow 0$.

4.3.1 Preparation

We start with notation used in later sections, and the define a matrix Dirichlet distribution.

Notations for the model structure

$\mathcal{K} = \{1, 2, \dots, K\}$, $\mathcal{K}_{-k} = \mathcal{K}/\{k\}$, $\mathcal{J} = \{1, 2, \dots, J\}$. K is the total number of neurons in observation, J is the total number of components assumed in environment.

$$\lambda_{kt}^* \equiv \begin{cases} \lambda_{kt}, & k \in \mathcal{K}, t \in \mathcal{T}_k; \\ *(undefined), & k \in \mathcal{K}, t \in \mathcal{T} \setminus \mathcal{T}_k. \end{cases}, \Lambda \equiv (\Lambda_1, \dots, \Lambda_K)^T \equiv (\lambda_{kt}^*)_{K \times T}.$$

λ_{kt} is defined in model 4.1, but some cannot be given a value since we only use part of observation for faster convergence and fitting stability.

$$\mathcal{T}_{ki} = \mathcal{T}_{ki}^\Lambda = \{t \in \mathcal{T}_k | \lambda_{kt} = i\}, \quad \mathcal{T}_{kkj} = \mathcal{T}_{kkj}^\Lambda = \{t \in \mathcal{T}_{kk} | d_{kt} = j\}.$$

These two sets are time sets partitioned by λ_{kt} and/or d_{kt} .

$$\lambda_{kt}^i \equiv \begin{cases} \lambda_{kt}, & k \in \mathcal{K}, t \in \mathcal{T}_{ki}; \\ *, & k \in \mathcal{K}, t \in \mathcal{T} \setminus \mathcal{T}_{ki}. \end{cases}, \quad \Lambda_{ki} = (\lambda_{k1}^i, \dots, \lambda_{kT}^i)^T.$$

$$Y_k = \{y_{kt} | t \in \mathcal{T}_k\}, \quad Y = \cup_{k \in \mathcal{K}} Y_k.$$

$$X_{ki} = \{x_{it}^{(k)} | t \in \mathcal{T}_{ki}\}, \quad X_k = \cup_{i \in \mathcal{K}_{-k}} X_{ki}, \quad X = \cup_{k \in \mathcal{K}} X_k.$$

$$d_{kt}^* \equiv \begin{cases} d_{kt}, & k \in \mathcal{K}, t \in \mathcal{T}_{kk}; \\ *, & k \in \mathcal{K}, t \in \mathcal{T} \setminus \mathcal{T}_{kk}. \end{cases}, \quad D_k = (d_{k1}^*, \dots, d_{kT}^*)^T,$$

$D = (D_1, \dots, D_K)^T$. It is similarly defined as λ_{kt}^* .

$$d_{kt}^j \equiv \begin{cases} d_{kt}, & k \in \mathcal{K}, t \in \mathcal{T}_{kkj}; \\ *, & k \in \mathcal{K}, t \in \mathcal{T} \setminus \mathcal{T}_{kkj}. \end{cases}, \quad D_{kj} = (d_{k1}^j, \dots, d_{kT}^j)^T.$$

$$\Theta_{1k} = \{\xi_{ki}, c_k, \sigma^2, m_j, s_j^2 | i \in \mathcal{K}_{-k}, j \in \mathcal{J}\}, \quad k \in \mathcal{K}; \quad P = (p_{ij})_{K \times K}, \quad P_k = (p_{k1}, \dots, p_{kJ})^T;$$

$Q = (q_1, \dots, q_J)^T$. These parameters are directly used in model 4.1.

$$\Theta_{2k} = \{\mu_{ki}, \delta^2, v_k, w^2, a, b, \mu_j, \tau^2, a_j, b_j | i \in \mathcal{K}_{-k}, j \in \mathcal{J}\}, \quad \text{for } k \in \mathcal{K}; \quad A = (\alpha_{ij})_{K \times K},$$

$$A_k = (\alpha_{k1}, \dots, \alpha_{kJ})^T, \quad B = (\beta_1, \dots, \beta_J)^T. \quad \text{These parameters are used to specify priors}$$

for previous parameters directly used in model 4.1.

$$\Theta_0 = \{\mu_{00}, \delta_0^2, m_0, n_0, v_0, w_0^2, m_1, n_1, \mu_0, \tau_0^2, m_2, n_2, a_{01}, a_{02}, b_{01}, b_{02}\} \text{ are initials for priors.}$$

$$\Theta_1 = \cup_{k \in \mathcal{K}} \Theta_{1k}, \quad \Theta_2 = \cup_{k \in \mathcal{K}} \Theta_{2k}.$$

Define matrix Dirichlet distribution[†] for $(p_{ki})_{K \times K}$ We define the matrix Dirichlet distribution

by three steps: first, do a transformation of Dirichlet distributions; second, define a bounded Dirichlet distribution; third, define matrix Dirichlet distribution. We

also list some simple properties for matrix Dirichlet distribution.

Let $(p_1, \dots, p_K) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$, $c_1, c_2, \dots, c_{K-1} \in (0, 1]$, $c_K = 1$. Consider the transformation ($k = 1, 2, \dots, K$)

$$z_k = \begin{cases} c_k p_k, & \text{if } c_k \in (0, 1) \\ c^* p_k, & \text{if } c_k = 1 \end{cases} \quad \text{or} \quad p_k = \begin{cases} \frac{z_k}{c_k}, & \text{if } c_k \in (0, 1) \\ \frac{z_k}{c^*}, & \text{if } c_k = 1, \end{cases} \quad (4.6)$$

where

$$c^* \equiv \frac{1 - \sum_{k=1, c_k < 1}^K z_k}{1 - \sum_{k=1, c_k < 1}^K \frac{z_k}{c_k}} = \frac{1 - \sum_{k=1, c_k < 1}^K c_k p_k}{1 - \sum_{k=1, c_k < 1}^K p_k} \geq 1$$

$$\begin{aligned} \therefore f(p_1, \dots, p_K) &\propto p_1^{\alpha_1-1} \dots p_K^{\alpha_K-1} \\ \therefore f(z_1, \dots, z_K) &\propto \prod_{k=1, c_k < 1}^K \left(\frac{z_k}{c_k}\right)^{\alpha_k-1} \times \prod_{k=1, c_k=1}^K \left(\frac{z_k}{c^*}\right)^{\alpha_k-1} \times \left| \frac{\partial(p_1, \dots, p_K)}{\partial(z_1, \dots, z_K)} \right| \\ &\propto \prod_{k=1}^K (z_k)^{\alpha_k-1} \times \left[\frac{1 - \sum_{k=1, c_k < 1}^K \frac{z_k}{c_k}}{1 - \sum_{k=1, c_k < 1}^K z_k} \right]^{\sum_{k=1, c_k=1}^K \alpha_k} \end{aligned}$$

This transformation will ensure that $z_k \leq c_k$ (for $k \leq K-1$). A slight modification leads to the following:

Definition 5. *Bounded Dirichlet distribution*

We say $(z_1, \dots, z_K) \sim \text{BD}(\alpha_1, \dots, \alpha_K; c_1, \dots, c_K; \gamma)$ iff

$$f(z_1, \dots, z_K) \propto \prod_{k=1}^K (z_k)^{\alpha_k-1} \times \left[\frac{1 - \sum_{k=1, c_k < 1}^K \frac{z_k}{c_k}}{1 - \sum_{k=1, c_k < 1}^K z_k} \right]^\gamma, \text{ here } \gamma > 0, c_1, \dots, c_{K-1} \in (0, 1], c_K = 1.$$

Its support is:

- $z_1 + \dots + z_K = 1$;
- $0 < z_k < c_k \leq 1$, $k = 1, \dots, K$;
- $0 < \sum_{k=1, c_k < 1}^K \frac{z_k}{c_k} < 1$.

† This distribution will give a stronger inter-relationship among submodels, see section 4.2. It can be viewed as a generalization of Dirichlet distribution.

The matrix Dirichlet distribution for $(p_{ij})_{K \times K}$ is defined as a series of conditional bounded Dirichlet distribution.

Definition 6. *Matrix Dirichlet distribution*

Let $P = (p_{ji})_{K \times K}$, $A = (a_{ij})_{K \times K}$, $c_0 \in (0, 1)$ (control parameter), $\Gamma = (\gamma_1, \dots, \gamma_K)$, $\gamma_k > 0$, $\alpha_{ij} > 0$,

we say $P \sim MD(A; \Gamma; c_0)$ iff P is defined by a series of conditional distributions:

- $(p_{11}, \dots, p_{1K}) | (A_1; \gamma_1; c_0) \sim BD(\alpha_{11}, \dots, \alpha_{1K}; c_{11} = \dots = c_{1K} = 1; \gamma_1)$;
- $(p_{i1}, \dots, p_{iK}) | (A_i; \gamma_i; c_0; p_{1i}, \dots, p_{i-1,i}) \sim BD(\alpha_{i1}, \dots, \alpha_{iK}; c_{i1} = 1 \wedge \frac{c_0}{p_{1i}}, \dots, c_{i,i-1} = 1 \wedge \frac{c_0}{p_{i-1,i}}, c_{ii} = \dots = c_{iK} = 1; \gamma_i)$, for $i = 2, 3, \dots, K$.

The matrix Dirichlet distribution has the following properties:

1. if $c_0 = 1$, rows of P are independent Dirichlet distributions; the smaller c_0 is, the tighter the submodels' interactions;
2. if $i \neq j$, then $p_{ij}p_{ji} < c_0$; these properties can be viewed as semi-DAG: if p_{ij} is relatively large, then p_{ji} can be forced to be close to 0. If we interpret p_{ij} to be the strength of excitation effect from neuron j to neuron i , we can use this property to approximate a one-way interaction mode by specifying a small c_0 ;
3. each row sum of P is 1; and
4. for a bounded Dirichlet distribution, $(z_k, c_k = 1 | z_k, c_k < 1) \sim (1 - \sum_{k=1, c_k < 1}^K z_k) \times \text{Dirichlet}(\alpha_k, c_k = 1)$.

4.3.2 Specification of prior distributions

This section gives details for all priors. The first-level priors for the parameters are directly used in model; second-level priors are used in specifying the first-level priors. Most priors are distributions frequently used in Bayesian approach; only few priors are special, for example, ξ_{ki} and c_k are bounded below.

First level prior distributions

$$\xi_{ki}^* | \lambda_{kt} = i \neq k, \mu_{ki}, \delta^2 \propto (\delta^2)^{-\frac{1}{2}} e^{-\frac{1}{\delta^2} \frac{(\xi_{ki}^* - \mu_{ki})^2}{2}}$$

$$\xi_{ki} = \dagger \xi_{ki}^* \vee 1 \quad (4.7)$$

$$c_k^* | \lambda_{kt} \neq k, v_k, w^2 \propto (w^2)^{-\frac{1}{2}} e^{-\frac{1}{w^2} \frac{(c_k^* - v_k)^2}{2}}$$

$$c_k = \dagger c_k^* \vee 0 \quad (4.8)$$

$$\sigma^2 | \lambda_{kt} \neq k, a, b \propto (\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}} \quad (4.9)$$

$$m_j | \lambda_{kt} = k, d_{kt} = j, \mu_j, \tau^2 \propto (\tau^2)^{-\frac{1}{2}} e^{-\frac{1}{\tau^2} \frac{(m_j - \mu_j)^2}{2}} \quad (4.10)$$

$$s_j^2 | \lambda_{kt} = k, d_{kt} = j, a_j, b_j \propto (s_j^2)^{-a_j-1} e^{-\frac{b_j}{s_j^2}} \quad (4.11)$$

$$\Gamma^\dagger(A, P, c_0) = (\gamma_1, \dots, \gamma_K)^T, \quad \gamma_k \equiv \sum_{l=1, c_0 \geq p_{lk}}^{k-1} \alpha_{kl} + \sum_{l=k}^K \alpha_{kl}$$

$$P|A; \Gamma(A, P, c_0); c_0 \sim \text{MD}(A; \Gamma; c_0) \quad (4.12)$$

$$q_1, \dots, q_J | \beta_1, \dots, \beta_J \sim \text{Dirichlet}(\beta_1, \dots, \beta_J) \propto q_1^{\beta_1-1} \dots q_J^{\beta_J-1} \quad (4.13)$$

Second level prior distributions

$$\mu_{ki} | \mu_{00}, \delta_0^2 \sim N(\mu_{00}, \delta_0^2) \propto e^{-\frac{1}{\delta_0^2} \frac{(\mu_{ki} - \mu_{00})^2}{2}} \quad (4.14)$$

$$\delta^2 | m_0, n_0 \sim IG(m_0, n_0) \propto (\delta^2)^{-m_0-1} e^{-\frac{n_0}{\delta^2}} \quad (4.15)$$

$$v_k | v_0, w_0^2 \sim N(v_0, w_0^2) \propto e^{-\frac{1}{w_0^2} \frac{(v_k - v_0)^2}{2}} \quad (4.16)$$

$$w^2 | m_1, n_1 \sim IG(m_1, n_1) \propto (w^2)^{-m_1-1} e^{-\frac{n_1}{w^2}} \quad (4.17)$$

$$\mu_j | \mu_0, \tau_0^2 \sim N(\mu_0, \tau_0^2) \propto e^{-\frac{1}{\tau_0^2} \frac{(\mu_j - \mu_0)^2}{2}} \quad (4.18)$$

$$\tau^2 | m_2, n_2 \sim IG(m_2, n_2) \propto (\tau^2)^{-m_2-1} e^{-\frac{n_2}{\tau^2}} \quad (4.19)$$

$$a, a_j | a_{01}, a_{02} \sim \text{Beta}(a_{01}, a_{02}) \propto a^{a_{01}-1} (U - a)^{a_{02}-1} I_{(0,U)}(a) \quad (4.20)$$

$$b, b_j | b_{01}, b_{02} \sim \text{Gamma}(b_{01}, b_{02}) \propto b^{b_{01}-1} e^{-b_{02}b} \quad (4.21)$$

$$\alpha_{ij} = \alpha_{00}, \beta_j = \beta_0. \quad (4.22)$$

‡ We make this requirement for two reasons: model identifiability and we only want to detect one-way interaction.

† Γ are not independent parameters.

4.3.3 Likelihood functions

In this section, we derive all conditional likelihood functions using the notation defined in section 4.3.1. A, c_0, B, Θ_0 are the initial values. Our goal is to calculate the likelihood $f(P, Q, \Theta_2 | Y, X; A, c_0, B, \Theta_0)$, Λ and D . In the distributional assumptions, we have assumed conditional independence in several places (and are implicit in the expressions below).

$$\begin{aligned}
f(Y|\Lambda, D, X, \Theta_1) &= \prod_{k \in \mathcal{K}} f(Y_k | \Lambda_k, D_k, X_k, \Theta_{1k}) \\
&= \prod_{k \in \mathcal{K}} \left\{ \prod_{i \in \mathcal{K}_{-k}} \left[\prod_{t \in \mathcal{T}_{ki}} f(y_{kt} | \lambda_{kt} = i, \xi_{ki}, x_{it}^{(k)}, c_k, \sigma^2) \right] \right. \\
&\quad \times \left. \prod_{j \in \mathcal{J}} \left[\prod_{t \in \mathcal{T}_{kj}} f(y_{kt} | \lambda_{kt} = k, d_{kt} = j, m_j, s_j^2) \right] \right\} \\
&\propto \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left\{ \prod_{i \in \mathcal{K}_{-k}} \left[(\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2} \frac{(y_{kt} - \xi_{ki}(x_{it}^{(k)} - c_k))^2}{2}} \right]^{I_{\{i\}}(\lambda_{kt})} \right. \\
&\quad \times \left. \prod_{j \in \mathcal{J}} \left[(s_j^2)^{-\frac{1}{2}} e^{-\frac{1}{2s_j^2} \frac{(y_{kt} - m_j)^2}{2}} \right]^{I_{\{k\}}(\lambda_{kt}) I_{\{j\}}(d_{kt})} \right\} \tag{4.23}
\end{aligned}$$

$$\begin{aligned}
f(\Lambda|P) &= \prod_{k \in \mathcal{K}} f(\Lambda_k | P_k) = \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} f(\lambda_{kt} | P_k) \\
&= \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \prod_{i \in \mathcal{K}} p_{ki}^{I_{\{i\}}(\lambda_{kt})} \tag{4.24}
\end{aligned}$$

$$\begin{aligned}
f(D|Q, \Lambda) &= \prod_{k \in \mathcal{K}} f(D_k | Q, \Lambda) \\
&= \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left[\prod_{j \in \mathcal{J}} q_j^{I_{\{j\}}(d_{kt})} \right]^{I_{\{k\}}(\lambda_{kt})} \tag{4.25}
\end{aligned}$$

$$\begin{aligned}
f(\Theta_1 | \Theta_2) &= \prod_{k \in \mathcal{K}} f(\Theta_{1k} | \Theta_{2k}) \\
&\propto \prod_{k \in \mathcal{K}} \left\{ \prod_{i \in \mathcal{K}_{-k}} \left\{ \left[(\delta^2)^{-\frac{1}{2}} e^{-\frac{1}{2\delta^2} \frac{(\xi_{ki} - \mu_{ki})^2}{2}} \right]^{I_{(1,+\infty)}(\xi_{ki})} \left[\Phi^\dagger \left(\frac{1 - \mu_{ki}}{\delta} \right) \right]^{I_{\{1\}}(\xi_{ki})} \right\} \right.
\end{aligned}$$

$$\begin{aligned}
& \times \left[(w^2)^{-\frac{1}{2}} e^{-\frac{1}{w^2} \frac{(c_k - v_k)^2}{2}} \right]^{I_{(0,+\infty)}(c_k)} \left[\Phi \left(-\frac{v_k}{w} \right) \right]^{I_{\{0\}}(c_k)} \Big\} \\
& \times (\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}} \times \prod_{j \in \mathcal{J}} \left[(\tau^2)^{-\frac{1}{2}} e^{-\frac{1}{\tau^2} \frac{(m_j - \mu_j)^2}{2}} (s_j^2)^{-a_j-1} e^{-\frac{b_j}{s_j^2}} \right] \quad (4.26)
\end{aligned}$$

$$\begin{aligned}
f(P|A; \Gamma; c_0) &= \text{MD}(P|A, \Gamma(A), c_0) = \prod_{k=1}^K f(P_k|A_k; P_1, \dots, P_{k-1}; c_0) \\
&= \prod_{k=1}^K \text{BD} \left(P_k|A_k; c_{k1} = 1 \wedge \frac{c_0}{p_{1k}}, \dots, c_{k,k-1} = 1 \wedge \frac{c_0}{p_{k-1,k}}, \right. \\
&\quad \left. c_{kk} = \dots = c_{kK} = 1; \gamma_k = \sum_{l=1, c_{kl}=1}^K \alpha_{kl} \right) \\
&\propto \prod_{k=1}^K \left\{ \left[\prod_{i=1}^K p_{ki}^{\alpha_{ki}-1} \right] \times \right. \\
&\quad \left. \left[\frac{1 - \sum_{l=1, c_0 < p_{lk}}^{k-1} p_{kl} \frac{p_{lk}}{c_0}}{1 - \sum_{l=1, c_0 < p_{lk}}^{k-1} p_{kl}} \right]^{l=1, \sum_{c_0 \geq p_{lk}}^{k-1} \alpha_{kl} + \sum_{l=k}^K \alpha_{kl}} \right\} \quad (4.27)
\end{aligned}$$

$$f(Q|B) \propto q_1^{\beta_1-1} q_2^{\beta_2-1} \dots q_J^{\beta_J-1} \quad (4.28)$$

$$\begin{aligned}
f(\Theta_2|\Theta_0) &= \prod_{k \in \mathcal{K}} f(\Theta_{2k}|\Theta_0) \\
&\propto \prod_{k \in \mathcal{K}} \left\{ \left[\prod_{i \in \mathcal{K}-k} e^{-\frac{1}{\delta_0^2} \frac{(\mu_{ki} - \mu_{00})^2}{2}} \right] \times e^{-\frac{1}{w_0^2} \frac{(v_k - v_0)^2}{2}} \right\} \\
&\quad \times (\delta^2)^{-m_0-1} e^{-\frac{n_0}{\delta^2}} \times (w^2)^{-m_1-1} e^{-\frac{n_1}{w^2}} \\
&\quad \times (\tau^2)^{-m_2-1} e^{-\frac{n_2}{\tau^2}} \times a^{a_{01}-1} (U - a)^{a_{02}-1} \times b^{b_{01}-1} e^{-b_{02}b} \\
&\quad \times \prod_{j \in \mathcal{J}} \left[e^{-\frac{1}{\tau_0^2} \frac{(\mu_j - \mu_0)^2}{2}} \times a_j^{a_{01}-1} (U - a_j)^{a_{02}-1} b_j^{b_{01}-1} e^{-b_{02}b_j} \right] \quad (4.29)
\end{aligned}$$

† $\Phi(\cdot)$ is cumulative density function of $N(0,1)$.

4.3.4 Conditional Posterior distributions

$$\begin{aligned}
P^{(n+1)}|\text{others} &\propto f(\Lambda^{(n)}|P)f(P|A; \Gamma(A); c_0) \\
&\propto \text{MD}(P|A; \Gamma(A); c_0) \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{K}} p_{ki}^{\sum_{t \in \mathcal{T}_k} I_{\{i\}}(\lambda_{kt}^{(n)})} \\
&\sim \text{MD}^\dagger(P|A + A^{(n)}; \Gamma(A + A^{(n)}); c_0) \\
A^{(n)} &= (\alpha_{ki}^{(n)})_{K \times K}, \quad \alpha_{ki}^{(n)} = \sum_{t \in \mathcal{T}_k} I_{\{i\}}(\lambda_{kt}^{(n)})
\end{aligned} \tag{4.30}$$

$$\begin{aligned}
Q^{(n+1)}|\text{others} &\propto f(D^{(n)}|Q, \Lambda^{(n)})f(Q|B) \\
&\propto q_1^{\beta_1-1} q_2^{\beta_2-1} \dots q_J^{\beta_J-1} \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left[\prod_{j \in \mathcal{J}} q_j^{I_{\{j\}}(d_{kt})} \right]^{I_{\{k\}}(\lambda_{kt})} \\
&\sim \text{Dirichlet}(B + B^{(n)}) \\
B^{(n)} &= (\beta_j^{(n)})_{J \times 1}, \quad \beta_j^{(n)} = \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}_k} I_{\{j\}}(d_{kt}) I_{\{k\}}(\lambda_{kt})
\end{aligned} \tag{4.31}$$

$$\begin{aligned}
\Theta_2^{(n+1)}|\text{others} &\propto f(\Theta_1^{(n)}|\Theta_2)f(\Theta_2|\Theta_0) \\
&\propto \prod_{k \in \mathcal{K}} \left\{ \left[\prod_{i \in \mathcal{K}_{-k}} e^{-\frac{1}{\delta_0^2} \frac{(\mu_{ki} - \mu_{00})^2}{2}} \right] \times e^{-\frac{1}{w_0^2} \frac{(v_k - v_0)^2}{2}} \right\} \\
&\times (\delta^2)^{-m_0-1} e^{-\frac{n_0}{\delta^2}} \times (w^2)^{-m_1-1} e^{-\frac{n_1}{w^2}} \\
&\times (\tau^2)^{-m_2-1} e^{-\frac{n_2}{\tau^2}} \times a^{a_{01}-1} (U - a)^{a_{02}-1} \times b^{b_{01}-1} e^{-b_{02}b} \\
&\times \prod_{j \in \mathcal{J}} \left[e^{-\frac{1}{\tau_0^2} \frac{(\mu_j - \mu_0)^2}{2}} \times a_j^{a_{01}-1} (U - a_j)^{a_{02}-1} \times b_j^{b_{01}-1} e^{-b_{02}b_j} \right] \\
&\times \prod_{k \in \mathcal{K}} \left\{ \prod_{i \in \mathcal{K}_{-k}} \left\{ \left[(\delta^2)^{-\frac{1}{2}} e^{-\frac{1}{\delta^2} \frac{(\xi_{ki}^{(n)} - \mu_{ki})^2}{2}} \right]^{I_{(1,+\infty)}(\xi_{ki}^{(n)})} \left[\Phi \left(\frac{1 - \mu_{ki}}{\delta} \right) \right]^{I_{\{1\}}(\xi_{ki}^{(n)})} \right\} \right\} \\
&\times \left[(w^2)^{-\frac{1}{2}} e^{-\frac{1}{w^2} \frac{(c_k^{(n)} - v_k)^2}{2}} \right]^{I_{(0,+\infty)}(c_k^{(n)})} \left[\Phi \left(-\frac{v_k}{w} \right) \right]^{I_{\{0\}}(c_k^{(n)})} \left\{ \right\} \\
&\times (\sigma_{(n)}^2)^{-a-1} e^{-\frac{b}{\sigma_{(n)}^2}} \times \prod_{j \in \mathcal{J}} \left[(\tau^2)^{-\frac{1}{2}} e^{-\frac{1}{\tau^2} \frac{(m_j^{(n)} - \mu_j)^2}{2}} (s_{j,(n)}^2)^{-a_j-1} e^{-\frac{b_j}{s_{j,(n)}^2}} \right]
\end{aligned} \tag{4.32}$$

† Sampling of MD will be discussed in next section.

$$\begin{aligned}
\mu_{ki}^{(n+1)}|_{\text{others}} &\propto e^{-\frac{1}{\delta_0^2} \frac{(\mu_{ki} - \mu_{00})^2}{2}} \times \left[e^{-\frac{1}{\delta_{(n)}^2} \frac{(\xi_{ki}^{(n)} - \mu_{ki})^2}{2}} \right]^{I_{(1,+\infty)}(\xi_{ki}^{(n)})} \left[\Phi \left(\frac{1 - \mu_{ki}}{\delta_{(n)}} \right) \right]^{I_{\{1\}}(\xi_{ki}^{(n)})} \\
&\sim^\dagger \begin{cases} e^{-\frac{1}{\delta_0^2} \frac{(\mu_{ki} - \mu_{00})^2}{2}} \Phi \left(\frac{1 - \mu_{ki}}{\delta_{(n)}} \right), & \text{if } \xi_{ki}^{(n)} = 1; \\ N \left(\frac{\frac{\mu_{00}}{\delta_0^2} + \frac{\xi_{ki}^{(n)}}{\delta_{(n)}^2}}{\frac{1}{\delta_0^2} + \frac{1}{\delta_{(n)}^2}}, \frac{1}{\frac{1}{\delta_0^2} + \frac{1}{\delta_{(n)}^2}} \right), & \text{if } \xi_{ki}^{(n)} > 1. \end{cases} \quad (4.33)
\end{aligned}$$

$$\begin{aligned}
v_k^{(n+1)}|_{\text{others}} &\propto e^{-\frac{1}{w_0^2} \frac{(v_k - v_0)^2}{2}} \times \left[e^{-\frac{1}{w_{(n)}^2} \frac{(c_k^{(n)} - v_k)^2}{2}} \right]^{I_{(0,+\infty)}(c_k^{(n)})} \left[\Phi \left(-\frac{v_k}{w_{(n)}} \right) \right]^{I_{\{0\}}(c_k^{(n)})} \\
&\sim^\dagger \begin{cases} e^{-\frac{1}{w_0^2} \frac{(v_k - v_0)^2}{2}} \Phi \left(-\frac{v_k}{w_{(n)}} \right), & \text{if } c_k^{(n)} = 0; \\ N \left(\frac{\frac{v_0}{w_0^2} + \frac{c_k^{(n)}}{w_{(n)}^2}}{\frac{1}{w_0^2} + \frac{1}{w_{(n)}^2}}, \frac{1}{\frac{1}{w_0^2} + \frac{1}{w_{(n)}^2}} \right), & \text{if } c_k^{(n)} > 0. \end{cases} \quad (4.34)
\end{aligned}$$

$$\begin{aligned}
\delta_{(n+1)}^2|_{\text{others}} &\propto (\delta^2)^{-m_0-1} e^{-\frac{n_0}{\delta^2}} \\
&\times \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{K}_{-k}} \left[(\delta^2)^{-\frac{1}{2}} e^{-\frac{1}{\delta^2} \frac{(\xi_{ki}^{(n)} - \mu_{ki}^{(n+1)})^2}{2}} \right]^{I_{(1,+\infty)}(\xi_{ki}^{(n)})} \left[\Phi \left(\frac{1 - \mu_{ki}^{(n+1)}}{\delta} \right) \right]^{I_{\{1\}}(\xi_{ki}^{(n)})}
\end{aligned}$$

$$\begin{aligned}
&\propto^\dagger (\delta^2)^{-m_0 - \frac{\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{K}_{-k}} I_{(1,+\infty)}(\xi_{ki}^{(n)})}{2} - 1} \times e^{-\frac{1}{\delta^2} \left\{ n_0 + \frac{1}{2} \left[\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{K}_{-k}} (\xi_{ki}^{(n)} - \mu_{ki}^{(n+1)})^2 I_{(1,+\infty)}(\xi_{ki}^{(n)}) \right] \right\}} \\
&\times \prod_{k \in \mathcal{K}} \prod_{i \in \mathcal{K}_{-k}} \left[\Phi \left(\frac{1 - \mu_{ki}^{(n+1)}}{\delta} \right) \right]^{I_{\{1\}}(\xi_{ki}^{(n)})} \quad (4.35)
\end{aligned}$$

† Sampling of this kind of distribution will be discussed in next section.

$$\begin{aligned}
w_{(n+1)}^2 | \text{others} &\propto (w^2)^{-m_1-1} e^{-\frac{n_1}{w^2}} \\
&\times \prod_{k \in \mathcal{K}} \left\{ \left[(w^2)^{-\frac{1}{2}} e^{-\frac{1}{w^2} \frac{(c_k^{(n)} - v_k^{(n+1)})^2}{2}} \right]^{I_{(0,+\infty)}(c_k^{(n)})} \left[\Phi \left(-\frac{v_k^{(n+1)}}{w} \right) \right]^{I_{\{0\}}(c_k^{(n)})} \right\} \\
&\propto^\dagger (w^2)^{-m_1 - \frac{1}{2} \sum_{k \in \mathcal{K}} I_{(0,+\infty)}(c_k^{(n)}) - 1} \times e^{-\frac{1}{w^2} \left[n_1 + \frac{1}{2} \sum_{k \in \mathcal{K}} (c_k^{(n)} - v_k^{(n+1)})^2 I_{(0,+\infty)}(c_k^{(n)}) \right]} \\
&\times \prod_{k \in \mathcal{K}} \left[\Phi \left(-\frac{v_k^{(n+1)}}{w} \right) \right]^{I_{\{0\}}(c_k^{(n)})} \tag{4.36}
\end{aligned}$$

$$\begin{aligned}
\tau_{(n+1)}^2 | \text{others} &\propto (\tau^2)^{-m_2 - \frac{J}{2} - 1} e^{-\frac{1}{\tau^2} \left[n_2 + \frac{1}{2} \sum_{j \in \mathcal{J}} (m_j^{(n)} - \mu_j^{(n)})^2 \right]} \\
&\sim IG(m_2 + \frac{J}{2}, n_2 + \frac{1}{2} \sum_{j \in \mathcal{J}} (m_j^{(n)} - \mu_j^{(n)})^2) \tag{4.37}
\end{aligned}$$

$$a^{(n+1)} | \text{others} \propto a^{a_{01}-1} (U - a)^{a_{02}-1} (\sigma_{(n)}^2)^{-a} I_{(0,U)}(a) \tag{4.38}$$

$$a_j^{(n+1)} | \text{others} \propto a_j^{a_{01}-1} (U - a_j)^{a_{02}-1} (s_{j,(n)}^2)^{-a_j} I_{(0,U)}(a_j) \tag{4.39}$$

$$b^{(n+1)} | \text{others} \propto b^{b_{01}-1} e^{-b(\frac{1}{\sigma_{(n)}^2} + b_{02})} \sim \text{Gamma}(b_{01}, \frac{1}{\sigma_{(n)}^2} + b_{02}) \tag{4.40}$$

$$b_j^{(n+1)} | \text{others} \propto b_j^{b_{01}-1} e^{-b_j(\frac{1}{s_{j,(n)}^2} + b_{02})} \sim \text{Gamma}(b_{01}, \frac{1}{s_{j,(n)}^2} + b_{02}) \tag{4.41}$$

$$\mu_j^{(n+1)} | \text{others} \sim N\left(\frac{\frac{\mu_0}{\tau_0^2} + \frac{m_j^{(n)}}{\tau_{(n+1)}^2}}{\frac{1}{\tau_0^2} + \frac{1}{\tau_{(n+1)}^2}}, \frac{1}{\frac{1}{\tau_0^2} + \frac{1}{\tau_{(n+1)}^2}}\right) \tag{4.42}$$

$$\begin{aligned}
\Lambda^{(n+1)} | \text{others} &\propto f(\Lambda | P^{(n+1)}) f(Y | \Lambda, D^{(n)}, X, \Theta_1^{(n)}) \\
&\propto \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left\{ \prod_{l \in \mathcal{K}} p_{kl,(n+1)}^{I_{\{l\}}(\lambda_{kt})} \times \prod_{i \in \mathcal{K}-k} \left[(\sigma_{(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma_{(n)}^2} \frac{(y_{kt} - \xi_{ki}^{(n)} (x_{it}^{(k)} - c_k^{(n)}))^2}{2}} \right]^{I_{\{i\}}(\lambda_{kt})} \right\}
\end{aligned}$$

† Sampling of this kind of distribution will be discussed in next section.

$$\times \prod_{j \in \mathcal{J}} \left[(s_{j,(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{s_{j,(n)}^2} \frac{(y_{kt} - m_j^{(n)})^2}{2}} \right]^{I_{\{k\}}(\lambda_{kt}) I_{\{j\}}(d_{kt}^{(n)})} \quad (4.43)$$

$$\begin{aligned} \mathbf{P}(\lambda_{kt}^{(n+1)} = l \neq k | \text{others}) &\propto p_{kl,(n+1)}^* \equiv p_{kl,(n+1)} (\sigma_{(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma_{(n)}^2} \frac{(y_{kt} - \xi_{kl}^{(n)}(x_{lt}^{(k)} - c_k^{(n)}))^2}{2}} \\ \mathbf{P}(\lambda_{kt}^{(n+1)} = k | \text{others}) &\propto p_{kk,(n+1)}^* \equiv p_{kk,(n+1)} \prod_{j \in \mathcal{J}} \left[(s_{j,(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{s_{j,(n)}^2} \frac{(y_{kt} - m_j^{(n)})^2}{2}} \right]^{I_{\{j\}}(d_{kt}^{(n)})} \\ \mathbf{P}(\lambda_{kt}^{(n+1)} = i | \text{others}) &= \frac{p_{ki,(n+1)}^*}{\sum_{l=1}^K p_{kl,(n+1)}^*}, \quad i \in \mathcal{K} \end{aligned} \quad (4.44)$$

$$\begin{aligned} D^{(n+1)} | \text{others} &\propto f(D | Q^{(n+1)}, \Lambda^{(n+1)}) f(Y | \Lambda^{(n+1)}, D, X, \Theta_1^{(n)}) \\ &\propto \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left[\prod_{j \in \mathcal{J}} q_{j,(n+1)}^{I_{\{j\}}(d_{kt})} \right]^{I_{\{k\}}(\lambda_{kt}^{(n+1)})} \\ &\quad \times \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \prod_{j \in \mathcal{J}} \left[(s_{j,(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{s_{j,(n)}^2} \frac{(y_{kt} - m_j^{(n)})^2}{2}} \right]^{I_{\{k\}}(\lambda_{kt}^{(n+1)}) I_{\{j\}}(d_{kt})} \\ &\propto \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left\{ \prod_{j \in \mathcal{J}} \left[q_{j,(n+1)} (s_{j,(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{s_{j,(n)}^2} \frac{(y_{kt} - m_j^{(n)})^2}{2}} \right]^{I_{\{j\}}(d_{kt})} \right\}^{I_{\{k\}}(\lambda_{kt}^{(n+1)})} \\ q_{j,(n+1)}^* &\equiv q_{j,(n+1)} (s_{j,(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{s_{j,(n)}^2} \frac{(y_{kt} - m_j^{(n)})^2}{2}} \end{aligned} \quad (4.45)$$

$$\mathbf{P}(d_{kt}^{(n+1)} = j | \lambda_{kt}^{(n+1)} \neq k) = \frac{1}{J} \quad (4.46)$$

$$\mathbf{P}(d_{kt}^{(n+1)} = j | \text{others}, \lambda_{kt}^{(n+1)} = k) = \frac{q_{j,(n+1)}^*}{\sum_{l=1}^J q_{l,(n+1)}^*}, \quad j \in \mathcal{J} \quad (4.47)$$

$$\begin{aligned} \Theta_1^{(n+1)} | \text{others} &\propto f(\Theta_1 | \Theta_2^{(n+1)}) f(Y | \Lambda^{(n+1)}, D^{(n+1)}, X, \Theta_1) \\ &\propto \prod_{k \in \mathcal{K}} \left\{ \prod_{i \in \mathcal{K}_{-k}} \left\{ \left[(\delta_{(n+1)}^2)^{-\frac{1}{2}} e^{-\frac{(\xi_{ki} - \mu_{ki}^{(n+1)})^2}{2\delta_{(n+1)}^2}} \right]^{I_{(1,+\infty)}(\xi_{ki})} \left[\Phi \left(\frac{1 - \mu_{ki}^{(n+1)}}{\delta_{(n+1)}} \right) \right]^{I_{\{1\}}(\xi_{ki})} \right\} \right. \\ &\quad \left. \times \left[(w_{(n+1)}^2)^{-\frac{1}{2}} e^{-\frac{1}{w_{(n+1)}^2} \frac{(c_k - v_k^{(n+1)})^2}{2}} \right]^{I_{(0,+\infty)}(c_k)} \left[\Phi \left(-\frac{v_k^{(n+1)}}{w_{(n+1)}} \right) \right]^{I_{\{0\}}(c_k)} \right\} \end{aligned}$$

$$\begin{aligned}
& \times (\sigma^2)^{-a^{(n+1)}-1} e^{-\frac{b^{(n+1)}}{\sigma^2}} \times \prod_{j \in \mathcal{J}} \left[(\tau_{(n+1)}^2)^{-\frac{1}{2}} e^{-\frac{1}{\tau_{(n+1)}^2} \frac{(m_j - \mu_j^{(n+1)})^2}{2}} (s_j^2)^{-a_j^{(n+1)}-1} e^{-\frac{b_j^{(n+1)}}{s_j^2}} \right] \\
& \times \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left\{ \prod_{i \in \mathcal{K}_{-k}} \left[(\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma^2} \frac{(y_{kt} - \xi_{ki}(x_{it}^{(k)} - c_k))^2}{2}} \right]^{I_{\{i\}}(\lambda_{kt}^{(n+1)})} \right. \\
& \quad \left. \times \prod_{j \in \mathcal{J}} \left[(s_j^2)^{-\frac{1}{2}} e^{-\frac{1}{s_j^2} \frac{(y_{kt} - m_j)^2}{2}} \right]^{I_{\{k\}}(\lambda_{kt}^{(n+1)}) I_{\{j\}}(d_{kt}^{(n+1)})} \right\} \tag{4.48}
\end{aligned}$$

$$\begin{aligned}
\xi_{ki}^{(n+1)} | \text{others} & \propto^\dagger \left[(\delta_{(n+1)}^2)^{-\frac{1}{2}} e^{-\frac{(\xi_{ki} - \mu_{ki}^{(n+1)})^2}{2\delta_{(n+1)}^2}} \right]^{I_{(1,+\infty)}(\xi_{ki})} \left[\Phi \left(\frac{1 - \mu_{ki}^{(n+1)}}{\delta_{(n+1)}} \right) \right]^{I_{\{1\}}(\xi_{ki})} \\
& \times \prod_{t \in \mathcal{T}_k} \left[(\sigma_{(n)}^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma_{(n)}^2} \frac{(y_{kt} - \xi_{ki}(x_{it}^{(k)} - c_k^{(n)}))^2}{2}} \right]^{I_{\{i\}}(\lambda_{kt}^{(n+1)})} \tag{4.49}
\end{aligned}$$

$$\begin{aligned}
c_k^{(n+1)} | \text{others} & \propto^\dagger \left[(w_{(n+1)}^2)^{-\frac{1}{2}} e^{-\frac{1}{w_{(n+1)}^2} \frac{(c_k - v_k^{(n+1)})^2}{2}} \right]^{I_{(0,+\infty)}(c_k)} \left[\Phi \left(-\frac{v_k^{(n+1)}}{w_{(n+1)}} \right) \right]^{I_{\{0\}}(c_k)} \\
& \times \prod_{t \in \mathcal{T}_k} \prod_{i \in \mathcal{K}_{-k}} \left[e^{-\frac{1}{\sigma_{(n)}^2} \frac{(y_{kt} - \xi_{ki}^{(n+1)}(x_{it}^{(k)} - c_k))^2}{2}} \right]^{I_{\{i\}}(\lambda_{kt}^{(n+1)})} \tag{4.50}
\end{aligned}$$

$$\sigma_{(n+1)}^2 | \text{others} \propto (\sigma^2)^{-a^{(n+1)}-1} e^{-\frac{b^{(n+1)}}{\sigma^2}} \times \tag{4.51}$$

$$\begin{aligned}
& \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \prod_{i \in \mathcal{K}_{-k}} \left[(\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma^2} \frac{(y_{kt} - \xi_{ki}^{(n+1)}(x_{it}^{(k)} - c_k^{(n+1)}))^2}{2}} \right]^{I_{\{i\}}(\lambda_{kt}^{(n+1)})} \\
& \sim IG(a^{(n+1)} + a^{(n+1)*}, b^{(n+1)} + b^{(n+1)*}) \tag{4.52} \\
a^{(n+1)*} & = \frac{1}{2} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}_k} \sum_{i \in \mathcal{K}_{-k}} I_{\{i\}}(\lambda_{kt}^{(n+1)}) \\
b^{(n+1)*} & = \frac{1}{2} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}_k} \sum_{i \in \mathcal{K}_{-k}} \left(y_{kt} - \xi_{ki}^{(n+1)}(x_{it}^{(k)} - c_k^{(n+1)}) \right)^2 I_{\{i\}}(\lambda_{kt}^{(n+1)})
\end{aligned}$$

† This distribution will be discussed in next section.

$$\begin{aligned}
m_j^{(n+1)} | \text{others} &\propto e^{-\frac{1}{\tau_{(n+1)}^2} \frac{(m_j - \mu_j^{(n+1)})^2}{2}} \times \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} e^{-\frac{1}{s_j^2} \frac{(y_{kt} - m_j)^2}{2}} I_{\{j\}}(d_{kt}^{(n+1)}) I_{\{k\}}(\lambda_{kt}^{(n+1)}) \\
&\sim N\left(\frac{\frac{\mu_j^{(n+1)}}{\tau_{(n+1)}^2} + \frac{\sum_{k \in \mathcal{K}, t \in \mathcal{T}_{kkj}} y_{kt}}{s_{j,(n)}^2}}{\frac{1}{\tau_{(n+1)}^2} + \frac{1}{s_{j,(n)}^2 / \sum_{k \in \mathcal{K}, t \in \mathcal{T}_{kkj}} 1}}, \frac{1}{\frac{1}{\tau_{(n+1)}^2} + \frac{1}{s_{j,(n)}^2 / \sum_{k \in \mathcal{K}, t \in \mathcal{T}_{kkj}} 1}}\right) \quad (4.53)
\end{aligned}$$

$$\begin{aligned}
s_{j,(n+1)}^2 | \text{others} &\propto (s_j^2)^{-a_j^{(n+1)} - 1} e^{-\frac{b_j^{(n+1)}}{s_j^2}} \prod_{k \in \mathcal{K}} \prod_{t \in \mathcal{T}_k} \left[(s_j^2)^{-\frac{1}{2}} e^{-\frac{1}{s_j^2} \frac{(y_{kt} - m_j)^2}{2}} \right]^{I_{\{k\}}(\lambda_{kt}^{(n+1)}) I_{\{j\}}(d_{kt}^{(n+1)})} \\
&\sim IG(a_j^{(n+1)} + a_j^{(n+1)*}, b_j^{(n+1)} + b_j^{(n+1)*}) \quad (4.54) \\
a_j^{(n+1)*} &= \frac{1}{2} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}_k} I_{\{j\}}(d_{kt}^{(n+1)}) I_{\{k\}}(\lambda_{kt}^{(n+1)}) \\
b_j^{(n+1)*} &= \frac{1}{2} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}_k} (y_{kt} - m_j^{(n+1)})^2 I_{\{j\}}(d_{kt}^{(n+1)}) I_{\{k\}}(\lambda_{kt}^{(n+1)})
\end{aligned}$$

4.3.5 Gibbs sampling procedure

With preparations above, we give the Gibbs sampling procedure for computing the posterior.

Step 0 Initial step.

$$\begin{aligned}
P^{(0)} &\sim P|A; \Gamma; c_0; \\
Q^{(0)} &\sim Q|B; \\
\Theta_2^{(0)} &\sim \Theta_2|\Theta_0; \\
\Lambda^{(0)} &\sim \Lambda|P^{(0)}; \\
D^{(0)} &\sim D|Q^{(0)}; \\
\Theta_1^{(0)} &\sim \Theta_1|\Theta_2^{(0)}.
\end{aligned}$$

Step 1 Iteration step. Assume n^{th} sample is finished, we get $(n+1)^{th}$ sample.

$$\begin{aligned}
P^{(n+1)} &\propto f(\Lambda^{(n)}|P) f(P|A; \Gamma(A); c_0); \\
Q^{(n+1)} &\propto f(D^{(n)}|Q, \Lambda^{(n)}) f(Q|B); \\
\Theta_2^{(n+1)} &\propto f(\Theta_1^{(n)}|\Theta_2) f(\Theta_2|\Theta_0); \\
\Lambda^{(n+1)} &\propto f(\Lambda|P^{(n+1)}) f(Y|\Lambda, D^{(n)}, X, \Theta_1^{(n)}); \\
D^{(n+1)} &\propto f(D|Q^{(n+1)}, \Lambda^{(n+1)}) f(Y|\Lambda^{(n+1)}, D, X, \Theta_1^{(n)}); \\
\Theta_1^{(n+1)} &\propto f(\Theta_1|\Theta_2^{(n+1)}) f(Y|\Lambda^{(n+1)}, D^{(n+1)}, X, \Theta_1).
\end{aligned}$$

5.0 SAMPLING AND SIMULATION STUDY

We now discuss several sampling problems we encountered in implementing the method from Section 4.3.4 and simulation study of applying Model 4.1 on transformed multiple spike trains.

5.1 SOME PROBLEMS IN SAMPLING

For some posterior distributions in Section 4.3.4, sampling is not straightforward. In this section, we discuss sampling problems for the matrix Dirichlet distribution and several complicated one-dimensional distributions. Since the Metropolis-Hastings algorithm did not work efficiently for matrix Dirichlet distributions and multimodal distributions, we designed a weighted sampler for our Gibbs iteration. Our weighted sampler is especially useful for a distribution with (1) bounded support, and (2) available uniform samplings on that support. In practice, for unbounded support, one can truncate the original distribution into one with bounded support which covers the bulk of probability mass of the original distribution. If the support of this distribution is (or can be covered by) a finite rectangular in R^n , one can always easily sample from the corresponding uniform distribution. In our case, the bounded Dirichlet distribution has an irregular bounded support but (of course) is covered by a rectangle. We also apply our weighted sampler to several complicated one-dimensional distributions in section 4.3.4.

Weighted sampler A weighted sampler is different from Metropolis-Hastings (MH) sampler in that it does not approximate the target distribution through iterations. Rather,

given a predetermined sampling precision, the weighted sampler can obtain one sample in a predetermined number of steps. The sampling precision can be evaluated both theoretically and by simulation. The weighted sampler also provides the possibility to control Bayesian iteration time. Generally, we write out this algorithm as follows.

Let X have pdf $f(x)$ with bounded support S ; $f(x)$ can be an arbitrary density function, and it is our target distribution. The weighted sampler procedure is:

1. get N samples: $\tilde{y} = (y_1, y_2, \dots, y_N)$, $y_n \text{ i.i.d. } \sim U[S]$ (uniform on S);
2. calculate $\tilde{w}(\tilde{y}) = (w_1 = \frac{f(y_1)}{\sum_{n=1}^N f(y_n)}, \dots, w_N = \frac{f(y_N)}{\sum_{n=1}^N f(y_n)})$;
3. get one sample $I(\tilde{y})$, such that $\mathbf{P}(I = n) = w_n$; and
4. set $\bar{x}(\tilde{y}) = y_{I(\tilde{y})} \in \{y_1, y_2, \dots, y_N\}$.

We now show by example that the empirical cdf $\hat{F}(x)$ of \hat{x} can approximate the true cdf $F(x)$ of $f(x)$ at any precision by increasing N . To evaluate the sampling precision, we introduce two measures: (1) $D_T(M) = \max_{\hat{x}_m, m=1, \dots, M} |\hat{F}(\hat{x}_m) - F(\hat{x}_m)|$ for the deviation of $\hat{F}(x)$ from $F(x)$; (2) let $\hat{x}_m, m = 1, \dots, M$ and $\hat{F}_x(x)$ be M samples and corresponding empirical cdf, $\hat{y}_m, m = 1, \dots, M$ and $\hat{F}_y(y)$ be another M samples and corresponding empirical cdf, $D_S(M) = \max_{z \in \{\hat{x}_m, \hat{y}_m, m=1, \dots, M\}} |\hat{F}_x(z) - \hat{F}_y(z)|$ for sampling stability. They are similar to the standard Kolmogorov-Smirnov statistics.

We now compare MH sampler and weighted sampler for a mixture distribution X : $P(X = N(-5, 1.5)) = 0.3$, $P(X = N(15, 1)) = 0.3$ and $P(X = \text{Exp}(0.8)) = 0.4$. Its pdf and cdf plot are given in Figure 1. In Figure 2 and Table 3, we show the empirical cdfs for four samplers: MH sampler with normal proposal density, MH sampler with uniform proposal density, weighted sampler and true sampler (generate samples according to its true distribution). We can see that the performance of the weighted sampler and true sampler is almost the same and much better than MH samplers by measure D_T . However, there is not much difference in D_S , which is expected according to Kolmogorov-Smirnov theorem.

Sample of P in (4.30) By definition 6 on P.24, we see that P can be sampled through a sequence of bounded Dirichlet distributions. Thus, we only need to sample a bounded Dirichlet distribution. This distribution has an irregularly shaped support and is possibly

multimodal. We do a simple comparison of weighted sampler and Metropolis-Hastings sampler. In a weighted sampler, the uniform distribution for the bounded Dirichlet distribution is Dirichlet distribution with all parameters set to 1. We use several values of N for our weighted sampler. The Metropolis-Hastings sampler to get a vector $(z_1, \dots, z_K) \sim \text{BD}(\alpha_1, \dots, \alpha_K; c_1, \dots, c_K; \gamma)$ is carried out thus:

- Initial step.
 $x^{(0)} \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$,
 $z^{(0)} = \text{transformation (4.6) of } x^{(0)}$;
- Sampling step. Assume $z^{(n)}$ obtained.
Let $x^{(n+1)} \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$, $U \sim U(0, 1)$,
 $z^{(n+1)*} = \text{transformation (4.6) of } x^{(n+1)}$,
 $a_z = 1 \wedge \frac{f_1(z^{(n+1)*})f_2(z^{(n)})}{f_1(z^{(n)})f_2(z^{(n+1)*})}$, where f_1 is the density of $\text{BD}(\alpha_1, \dots, \alpha_K; c_1, \dots, c_K; \gamma)$,
and f_2 is the true density of $z^{(n+1)*}$;
- Iteration step. If $a_z > U$, let $z^{(n+1)} = z^{(n+1)*}$ and $n + 1 \rightarrow n$; otherwise, make no change. Then go back to sampling step until convergence.

In practice, this procedure does not work well for BD because the acceptance rate is very small. So we do not present sampling results for MH procedure. By applying weighted sampler, however, we show several traceplots (Figure 4,5,6,7) to demonstrate that the weighted sampler does work. We sampled $P \sim MD(A, c_0)$: $c_0 = 0.1$; A takes value M , $M * 1000$, MM and $MM * 1000$. Here, the matrices M and MM are as:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0.7 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad MM = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

We can see that the weighted sampler successfully draws samples from a matrix Dirichlet distribution. For $A = MM$, p_{ij} and p_{ji} alternatively dominate each other; for $A = M$, some p_{ij} s always dominate p_{ji} .

Sample of μ in (4.33) and v in (4.34) The only problem is how to sample a density proportional to: $e^{-\frac{1}{\delta_0^2} \frac{(\mu_{ki} - \mu_{00})^2}{2}} \Phi\left(\frac{1 - \mu_{ki}}{\delta_{(n)}}. Obviously, we can use Metropolis-Hastings method again. But these distributions have up to three local modes, Metropolis-Hastings sampling method can hardly visit all modes, especially in some extreme cases, for example, if those modes are far apart. Therefore, we used a weighted sampler in this case.$

Sample of δ^2 in (4.35) and w^2 in (4.36) This is a serious problem, because these distributions are highly irregular and with lots of local modes. It is likely for Metropolis-Hastings sampling method to be trapped around a local mode. It is not easy to find a proposal distribution for efficiently implementing Metropolis-Hastings algorithm in this case due to the complicated form of target distributions. But it is a one-dimensional distribution and we can easily find an bounded interval which cover most of its probability, for which we use a weighted sampler.

Sample of a in (4.38) We can use either the MH or weighted sampler for a .

Sample of ξ in (4.49) and c in (4.50) Strictly speaking, this is not a sampling problem. Rather, it is a problem of posterior distribution which can be generally posed thus:

$$\begin{aligned} \theta &\sim \begin{cases} \mathbf{P}(\{\theta = c_0\}) = \Phi\left(\frac{c_0 - \mu_0}{\delta_0}\right), & \text{if } \theta = c_0; \\ \text{has pdf } \phi\left(\frac{\theta - \mu_0}{\delta_0}\right), & \text{if } \theta > c_0. \end{cases} \\ x|\theta &\sim \phi\left(\frac{x - \theta}{\sigma}\right). \end{aligned}$$

What is the distribution of $\theta|x$? The answer is simple.

$$\begin{aligned} \theta|x &\sim \begin{cases} \mathbf{P}(\{\theta = c_0\}) = \Phi\left(\frac{c_0 - \mu^*}{\delta_*}\right), & \text{if } \theta = c_0; \\ \text{has pdf } \phi\left(\frac{\theta - \mu^*}{\delta_*}\right), & \text{if } \theta > c_0. \end{cases} \\ \mu^* &= \frac{\frac{\mu_0}{\delta_0^2} + \frac{x}{\sigma^2}}{\frac{1}{\delta_0^2} + \frac{1}{\sigma^2}}, \quad \delta_*^2 = \frac{1}{\frac{1}{\delta_0^2} + \frac{1}{\sigma^2}} \end{aligned}$$

Based on this property, (4.49) and (4.50) can be rewritten as:

$$\xi_{ki}^{(n+1)} | \text{others} \sim \begin{cases} \mathbf{P}(\{\xi_{ki}^{(n+1)} = 1\}) = \Phi\left(\frac{1 - \mu_{ki}^{(n+1)*}}{\delta_{(n+1)*}}\right), & \text{if } \xi_{ki}^{(n+1)} = 1; \\ \text{has pdf } \phi\left(\frac{\xi_{ki}^{(n+1)} - \mu_{ki}^{(n+1)*}}{\delta_{(n+1)*}}\right), & \text{if } \xi_{ki}^{(n+1)} > 1. \end{cases} \quad (5.1)$$

$$\mu_{ki}^{(n+1)*} = \frac{\frac{\mu_{ki}^{(n+1)}}{\delta_{(n+1)}^2} + \frac{\sum_{t \in \mathcal{T}_k} I_{\{i\}}(\lambda_{kt}^{(n+1)})(x_{it}^{(k)} - c_k^{(n)})y_{kt}}{\sigma_{(n)}^2}}{\frac{1}{\delta_{(n+1)}^2} + \frac{\sum_{t \in \mathcal{T}_k} I_{\{i\}}(\lambda_{kt}^{(n+1)})(x_{it}^{(k)} - c_k^{(n)})^2}{\sigma_{(n)}^2}}$$

$$\delta_{(n+1)*}^2 = \frac{1}{\frac{1}{\delta_{(n+1)}^2} + \frac{\sum_{t \in \mathcal{T}_k} I_{\{i\}}(\lambda_{kt}^{(n+1)})(x_{it}^{(k)} - c_k^{(n)})^2}{\sigma_{(n)}^2}}$$

$$c_k^{(n+1)} | \text{others} \sim \begin{cases} \mathbf{P}(\{c_k^{(n+1)} = 0\}) = \Phi\left(-\frac{v_k^{(n+1)*}}{w_{(n+1)*}}\right), & \text{if } c_k^{(n+1)} = 0; \\ \text{has pdf } \phi\left(\frac{c_k^{(n+1)} - v_k^{(n+1)*}}{w_{(n+1)*}}\right), & \text{if } c_k^{(n+1)} > 0. \end{cases} \quad (5.2)$$

$$v_k^{(n+1)*} = \frac{\frac{v_k^{(n+1)}}{w_{(n+1)}^2} + \frac{\sum_{t \in \mathcal{T}_k} \sum_{i \in \mathcal{K}_{-k}} I_{\{i\}}(\lambda_{kt}^{(n+1)})(\xi_{ki}^{(n+1)})^2 (x_{it}^{(k)} - y_{kt} / \xi_{ki}^{(n+1)})}{\sigma_{(n)}^2}}{\frac{1}{w_{(n+1)}^2} + \frac{\sum_{t \in \mathcal{T}_k} \sum_{i \in \mathcal{K}_{-k}} I_{\{i\}}(\lambda_{kt}^{(n+1)})(\xi_{ki}^{(n+1)})^2}{\sigma_{(n)}^2}}$$

$$w_{(n+1)*}^2 = \frac{1}{\frac{1}{w_{(n+1)}^2} + \frac{\sum_{t \in \mathcal{T}_k} \sum_{i \in \mathcal{K}_{-k}} I_{\{i\}}(\lambda_{kt}^{(n+1)})(\xi_{ki}^{(n+1)})^2}{\sigma_{(n)}^2}}$$

5.2 SIMULATION STUDY

In Section 5.1, we studied sampling problems in fitting Model 4.1. In this section, we will show how Model 4.1 can be used to detect Gerstein's neuronal network interaction mode. Our simulation procedure is to start with a copy matrix, generate spike trains, apply gravitational clustering algorithm and then apply model 4.1 on the corresponding charge process.

Selection of network structure There are two practical problems in fitting the Bayesian Model 4.1: (1) the sampling of matrix Dirichlet distribution is not fast; (2) the convergence of MCMC update for Model 4.1 is not fast. Thus, we currently study a small network with only 5 nodes and 4 edges. This network is graphically represented in Figure 8. Corresponding raster plots, charge process, and gravitational time-distance plot are shown in Figure 10, 11, 12, respectively.

Selection of c_0 c_0 (see Definition 6) controls the upper bound of $p_{ij} \times p_{ji}$ ($i \neq j$). Ideally, to approximate a directed acyclic graph, we need specify a very small c_0 ; but small c_0 means slower in sampling matrix Dirichlet distribution. As a trade-off between CPU time and the degree of approximating a DAG, we set $c_0 = 0.1$.

Choose J in Model 4.1 J reflects the strength of background signal. After studies of various values, we set $J = 1$ because much larger values of J tend to erase all edges in the network.

Detect the convergence of MCMC update We set the minimum iteration in Gibbs sampling to be 6000 and then check trace plots for posteriors. If there is no visual deviation from convergence, we take it as convergence. In Figure 13, we give trace plots for p_{ij} (posterior copy matrix).

Rebuild network structure For a convergence MCMC iteration sequence, the first 4000 iterations are taken as data burn-in and after that take one record for every 20-iteration until the 9000th iteration. We then will have 250 estimated matrices P which directly reflect the network structure; however we cannot always get a clear estimation because all p_{ij} s are positive in theory. According to asymptotic theory, Bayesian posteriors approximate normal distributions when sample size is large. In our simulation study, we use data for 100 seconds and take one sample from transformed curve for every 0.002 seconds, so we have 50000 samples for Model 4.1. We can assume the posterior of p_{ij} to be normal and check its Normal QQ-plot (Figure 14).

Now we can say that if $mean(p_{ij}) - 2.576 \times sd(p_{ij}) > 0$, there is evidence that an edge exists from j to i . However, we can never obtain original copy matrix M . In Figure 9, we list $mean(p_{ij})$, $sd(p_{ij})$, $mean(p_{ij}) - 2.576 \times sd(p_{ij})$ and original copy matrix M . In Figure 8, we construct network plots with the same method for both estimated and true methods.

5.3 CONCLUSION

In this thesis, we developed a Poisson process based probability model to study the properties of the gravitational clustering algorithm (GCA) and showed that GCA should be adjusted by a time factor to balance the oscillation of GCA. The idea of the probability model can be generalized to study ordinary differential equations (ODE) in other contexts. We also developed a detailed mixture model to directly study the dependence among multiple spike trains. This model can be used to estimate a network structure by generalizing the Dirichlet distribution and possibly provides an alternative to a traditional Bayesian network and dynamic Bayesian network.

The future work can be extended in at least two directions: (1) refine the methods used in the probability model to obtain conclusions under much more general conditions; (2) propose a general Bayesian framework to make a statistical model with much more flexibility: statisticians shouldn't worry about the form of distributions and sampling techniques and should only focus on how to build and compare models. The second direction is particularly meaningful today because, for instance, the internet is creating giant data sets continuously; as a contrast, almost no classic statistical models can be directly applied on analyzing this kinds of data sets. Only a Bayesian model provides the possibility to handling such data sets; but current Bayesian methods are heavily affected by the form of probability distributions and sampling techniques. As I can see, a Bayesian framework based on discrete distributions can possibly provide a general solution (definitely with a heavy computational load).

Figure 1: CDF and PDF plot for X

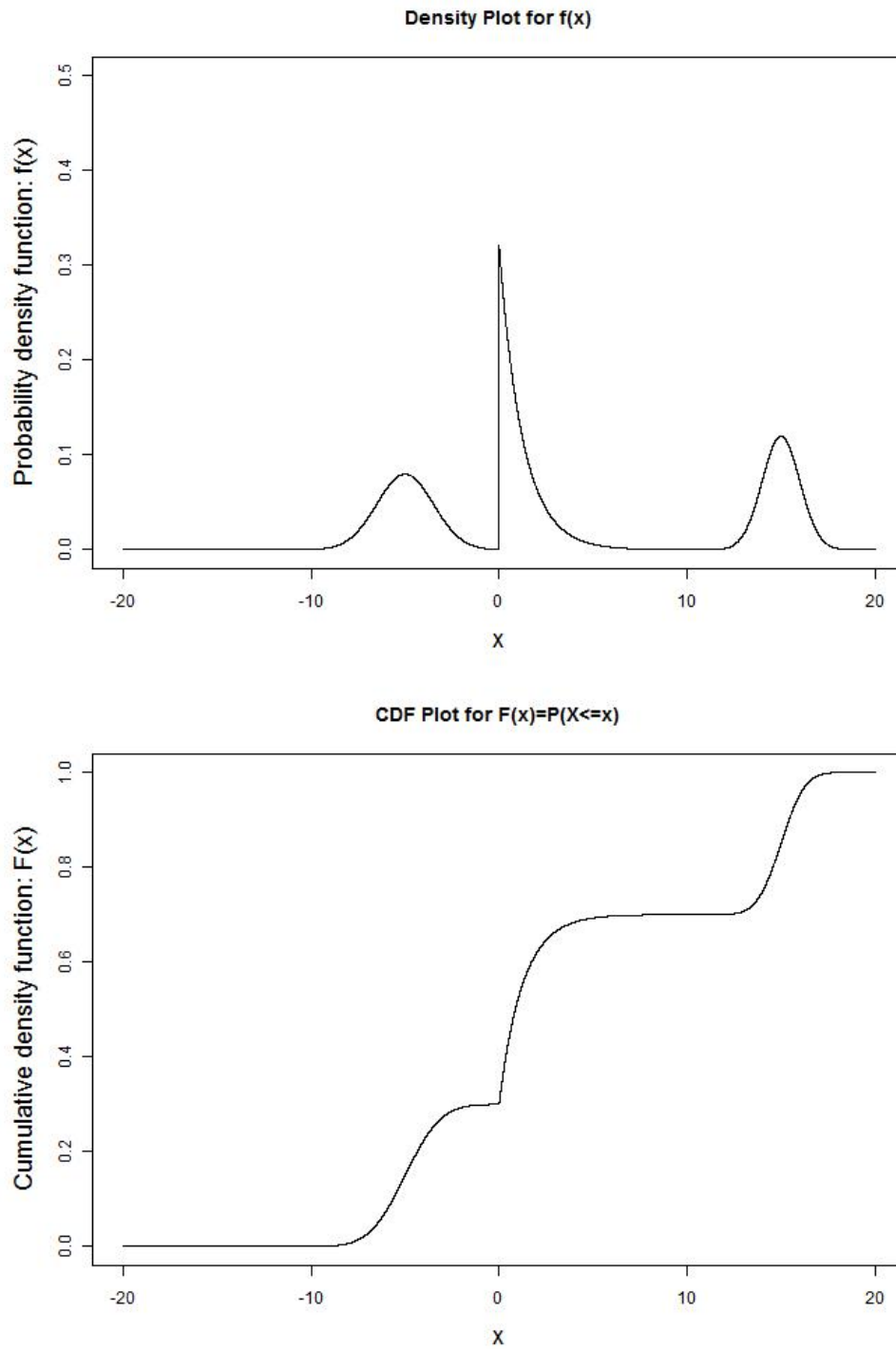


Figure 2: Empirical CDFs for 4 samplers

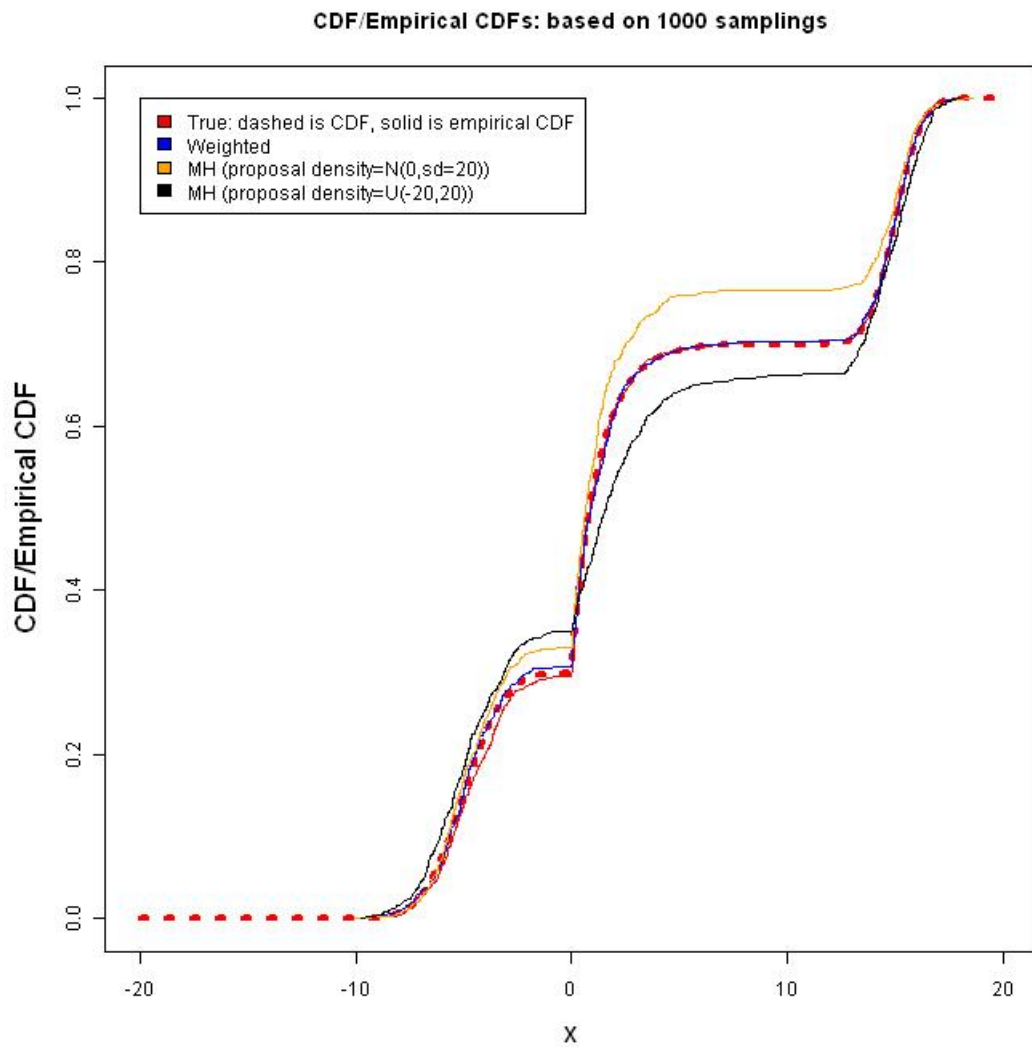


Figure 3: Tables for comparing 4 samplers: D_T and D_S

Compare 4 sampling methods		Sampling methods			
		True Sampler	Weighted Sampler	MH Sampler	MH Sampler
Sampling settings			Sample precision=100	proposal density=N(0,20)	proposal density=U(-20,20)
				max update steps=100	max update steps=100
D_T	mean	0.02664462	0.02753975	0.05800412	0.07958733
	sd	0.008270034	0.008488446	0.012601106	0.011547839
D_S	mean	0.038469	0.037707	0.038098	0.038851
	sd	0.01183889	0.01146467	0.01199868	0.0118039

Figure 4: MD sample trace plot: M

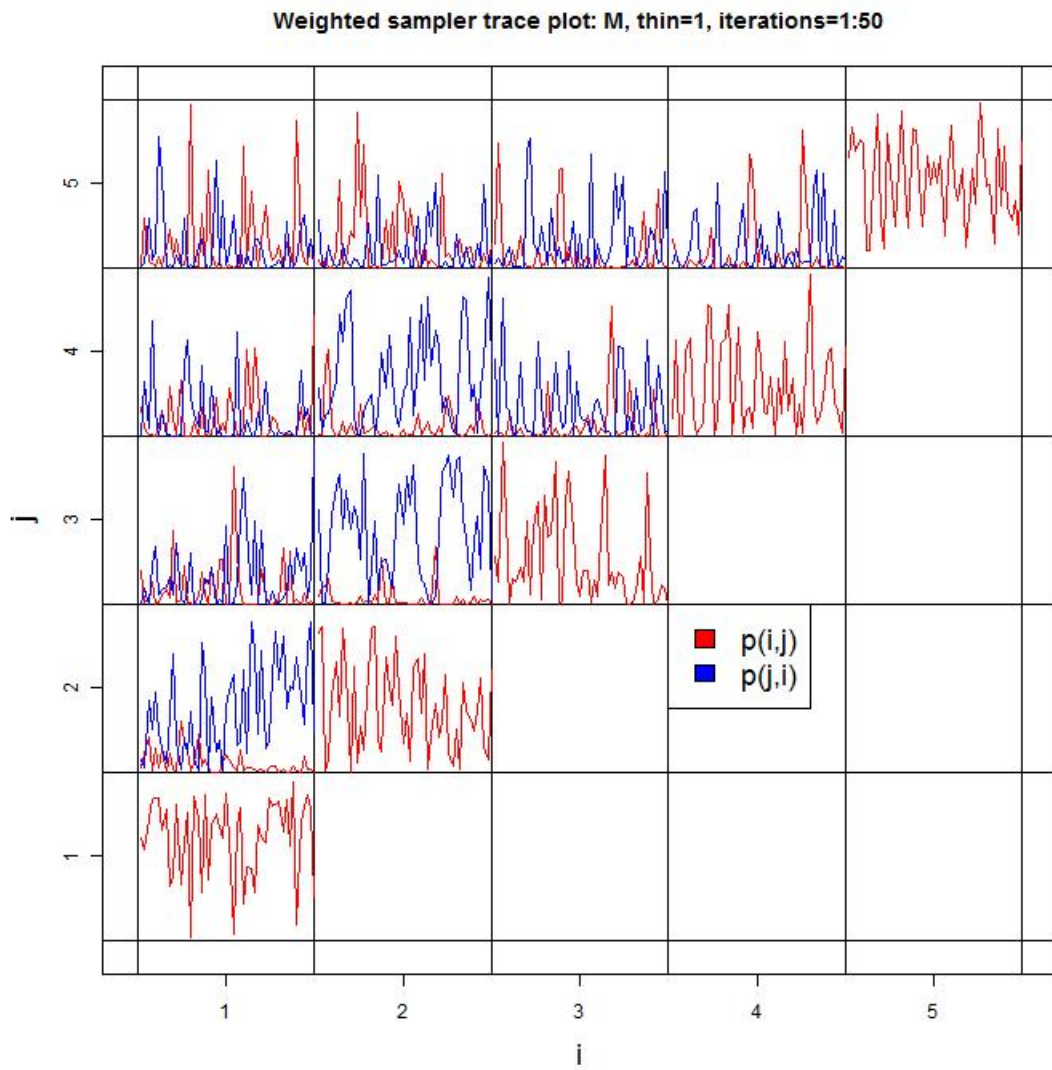


Figure 5: MD sample trace plot: $M * 1000$

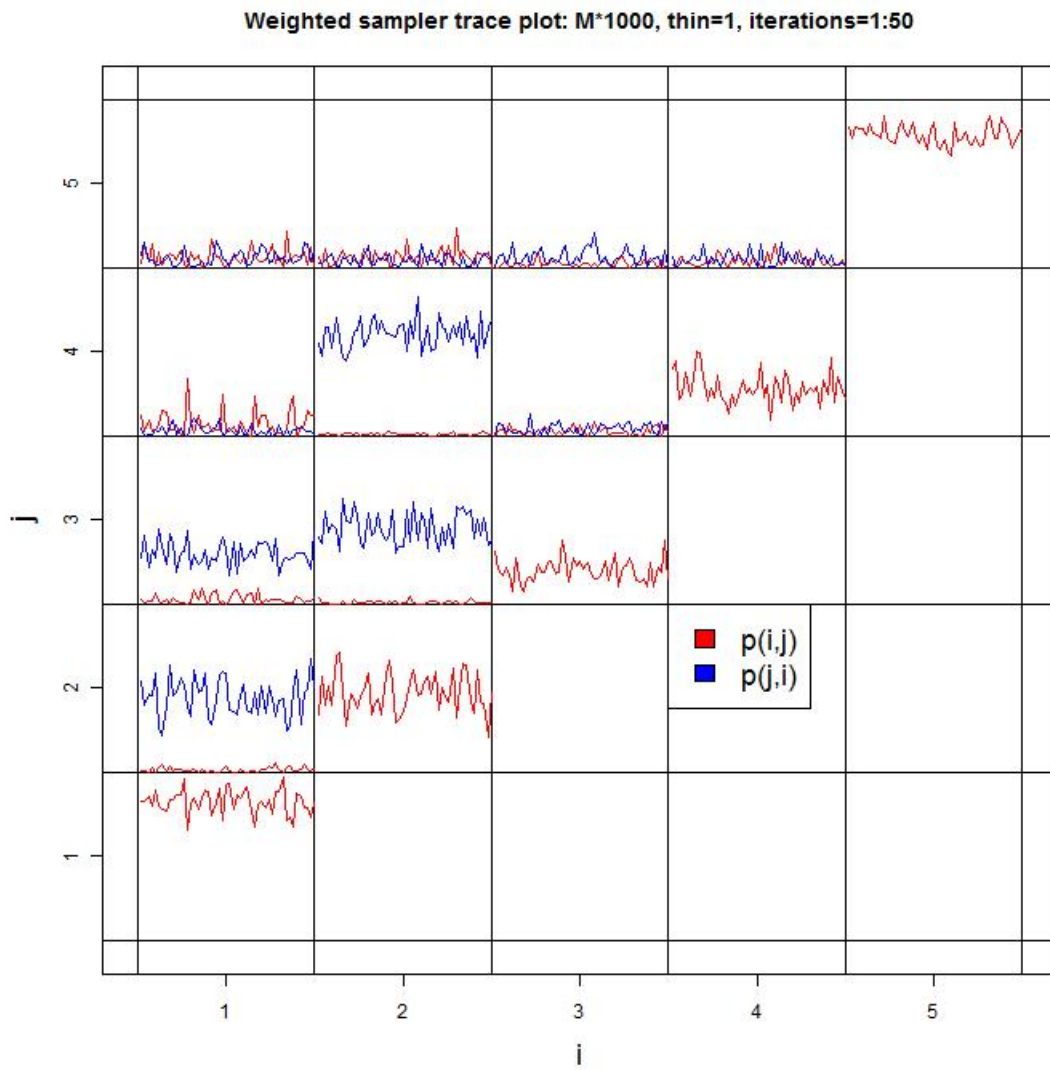


Figure 6: MD sample trace plot: MM

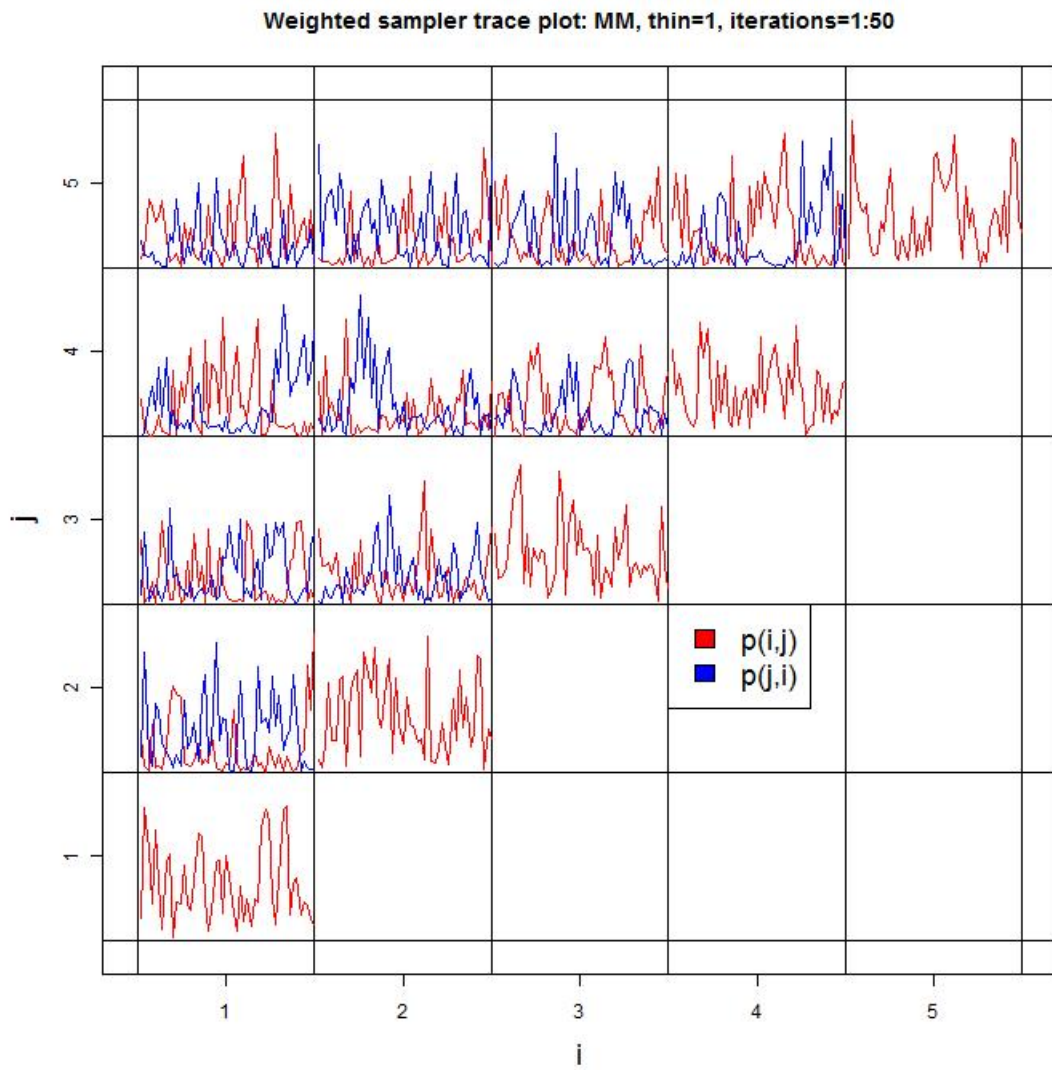


Figure 7: MD sample trace plot: $MM * 1000$

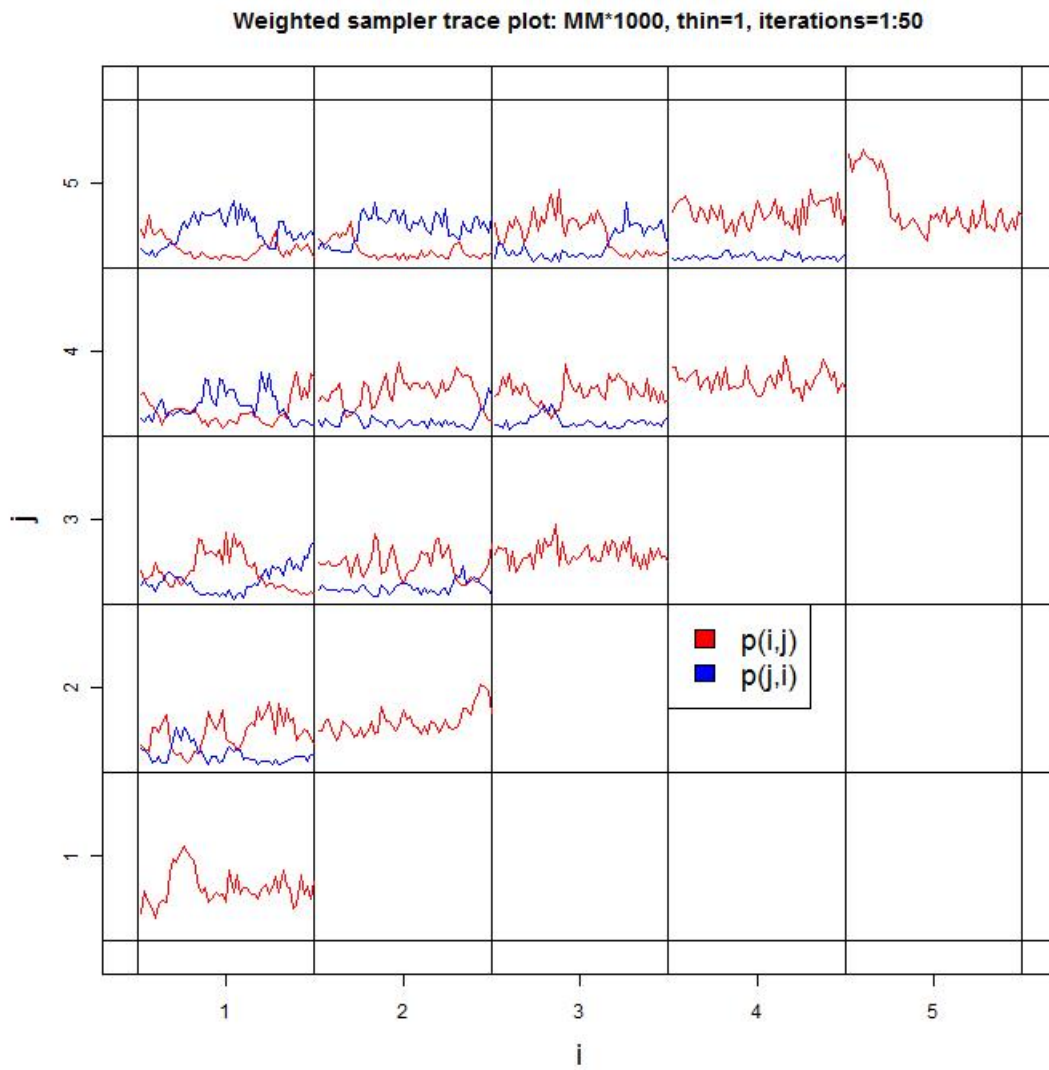


Figure 8: Graph for copy matrix

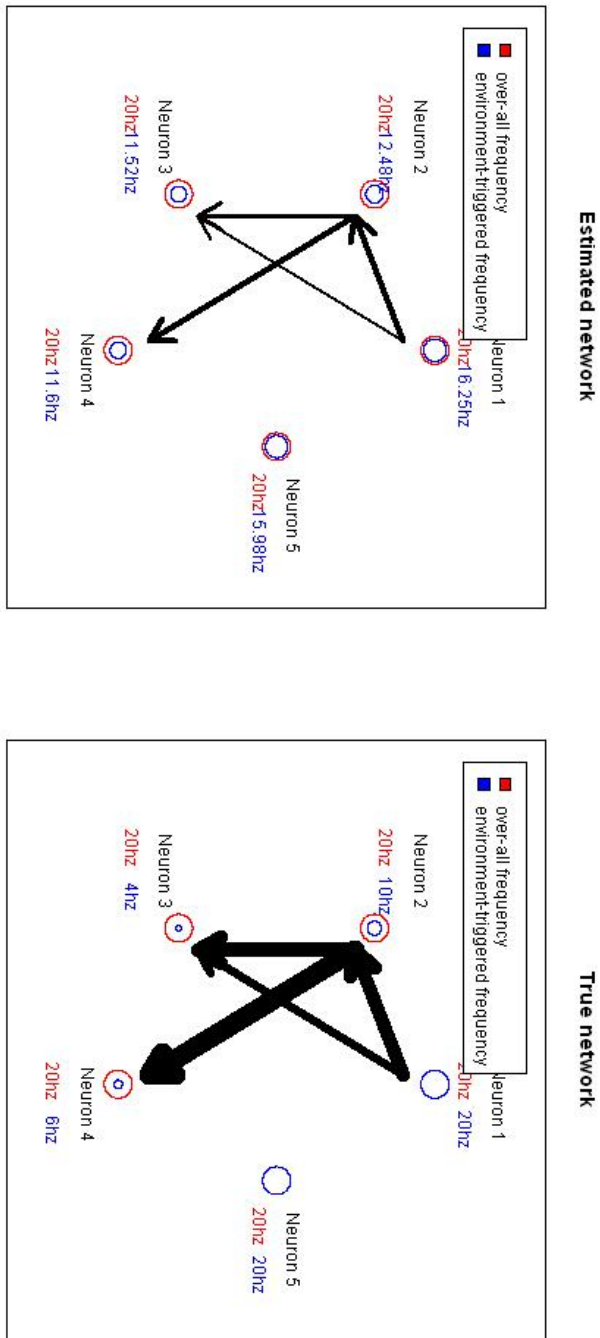


Figure 9: Table for posterior copy matrix

	Mean of P _{ij}				
	1	2	3	4	5
1	0.812604318	0.049281099	0.046723936	0.043537276	0.047853371
2	0.222523402	0.623905711	0.025415634	0.025264051	0.102891202
3	0.124536362	0.195906056	0.576226992	0.025515588	0.077815003
4	0.077663827	0.215542538	0.105654719	0.579861922	0.021276994
5	0.049632296	0.045835723	0.056897941	0.048850334	0.798783706
	SD of P _{ij}				
	1	2	3	4	5
1	0.057521356	0.038250833	0.038436534	0.036301695	0.039660036
2	0.05877459	0.06444056	0.019100875	0.0203469	0.046847146
3	0.045207719	0.054792335	0.061782102	0.018281538	0.0360265
4	0.040441707	0.064619477	0.049724749	0.067760652	0.016276356
5	0.041538848	0.039698026	0.045046029	0.040265059	0.055240348
	Mean(P _{ij})-2.576×SD(P _{ij})				
	1	2	3	4	5
1	0.664429306	-0.049253047	-0.052288574	-0.049975892	-0.054310881
2	0.071120058	0.45790683	-0.023788219	-0.027149563	-0.017787046
3	0.008081278	0.054761001	0.417076298	-0.021577655	-0.01498926
4	-0.02651401	0.049082766	-0.022436235	0.405310483	-0.020650899
5	-0.057371777	-0.056426393	-0.059140629	-0.054872459	0.65648457
	True copy matrix P				
	1	2	3	4	5
1	1	0	0	0	0
2	0.5	0.5	0	0	0
3	0.3	0.5	0.2	0	0
4	0	0.7	0	0.3	0
5	0	0	0	0	1

Figure 10: Raster plot

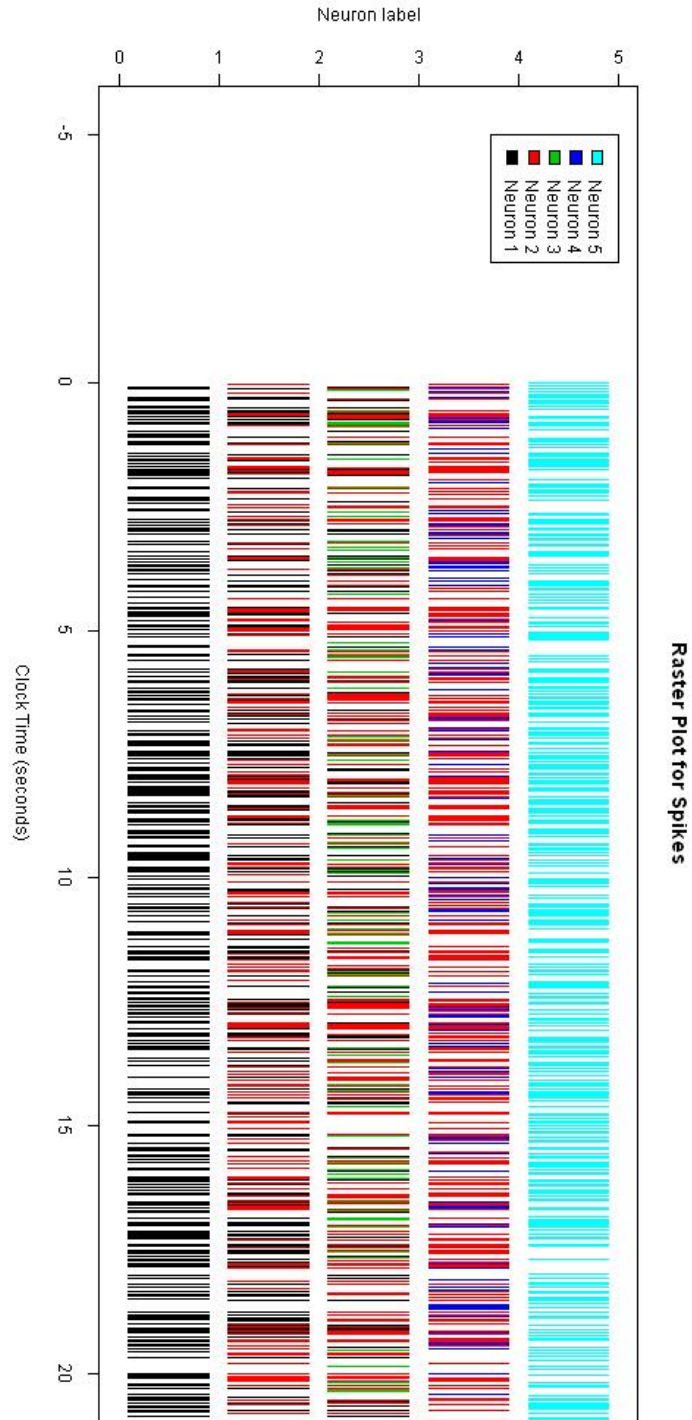


Figure 11: Plot for charge process

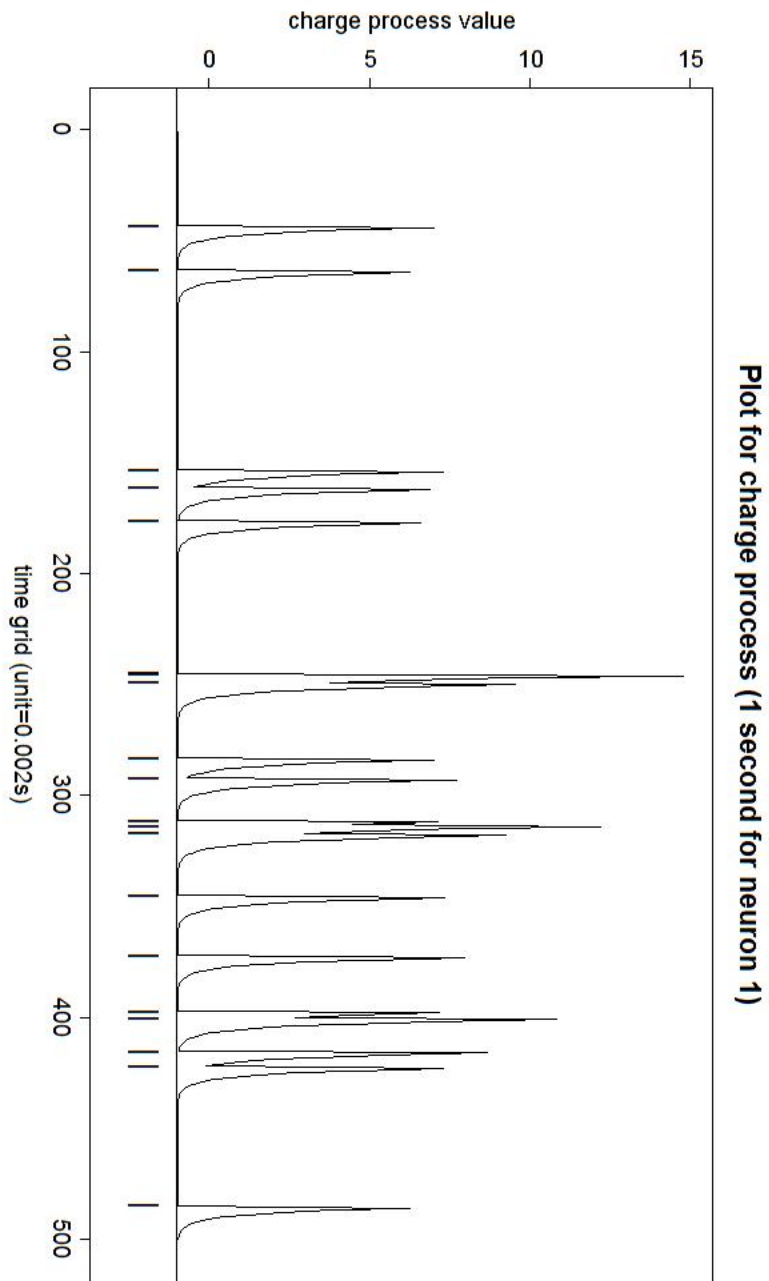


Figure 12: Gravitational clustering plot

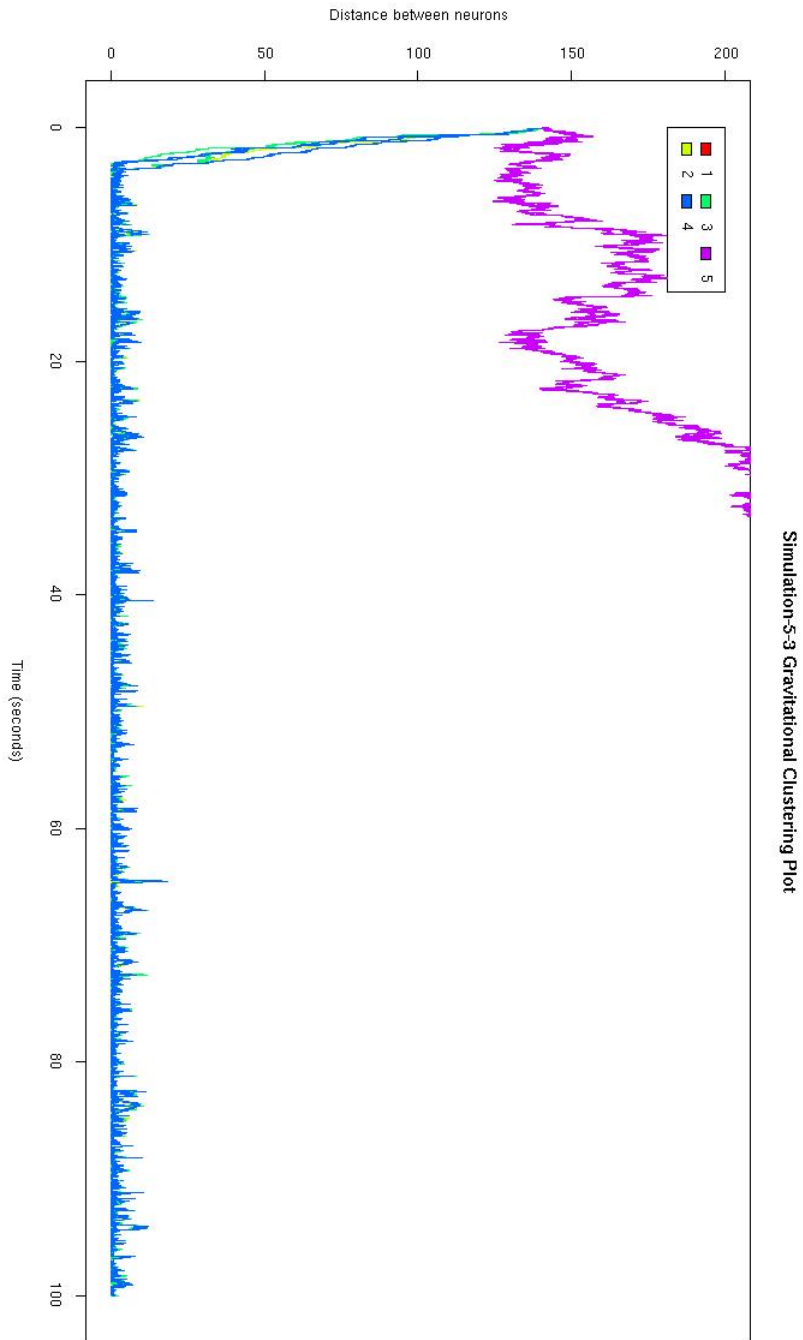


Figure 13: Trace plot for P

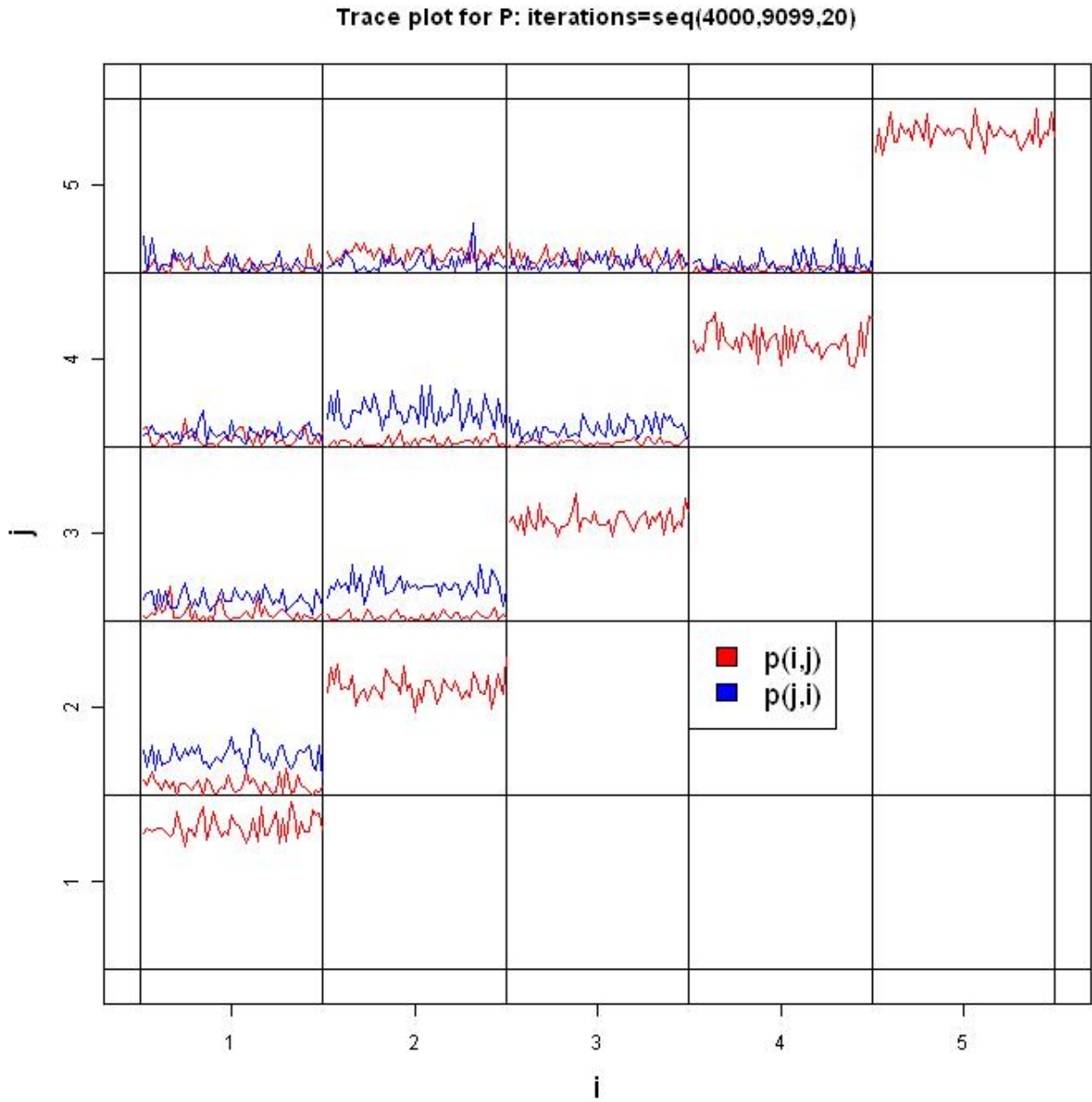


Figure 14: Normal QQ plot for p_{ij}

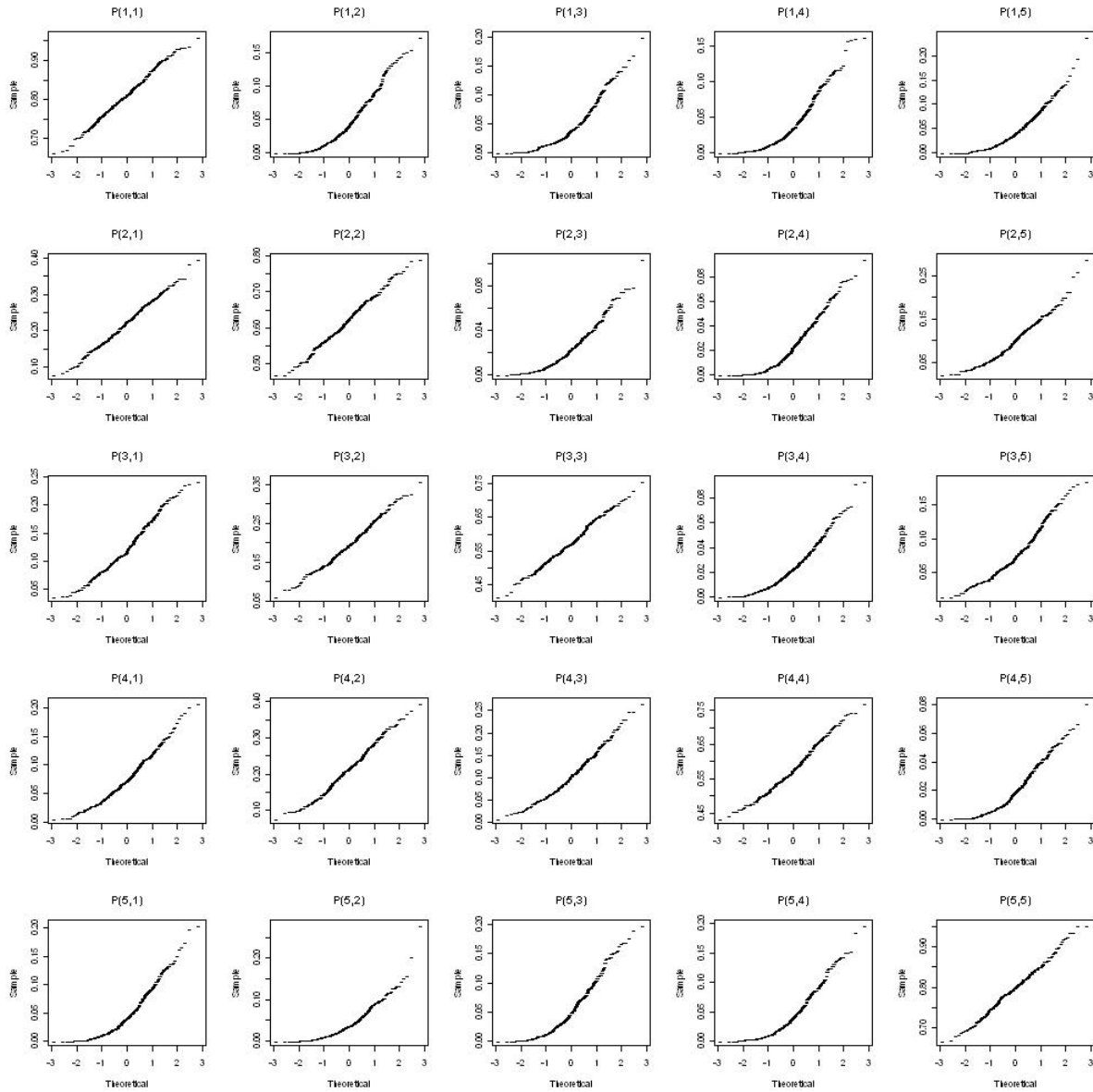
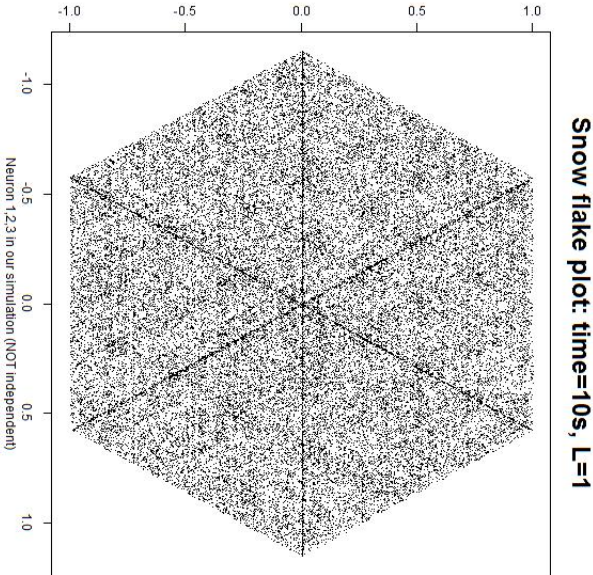
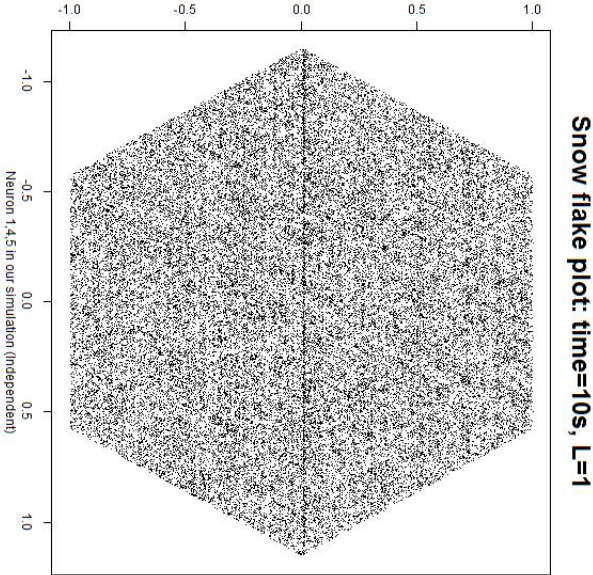


Figure 15: Snowflake plots



APPENDIX

PROOFS OF RESULTS IN SECTION 3

Proof of Lemma 2:

- (i) since $n_{\bar{s},1}(t)$ is measurable of $n_1(t - \varepsilon_0)$ and $n_{\bar{s},2}(t)$ is measurable of $n_1(t) - n_1(t - \varepsilon_0)$, noticing the independence of $n_1(t - \varepsilon_0)$ and $n_1(t) - n_1(t - \varepsilon_0)$, the independence of $n_{\bar{s},1}(t)$ and $n_{\bar{s},2}(t)$ follows;
- (ii) For $t > \varepsilon_0$ and $\delta > 0$, let $W \equiv n_{\bar{s},1}(t + \delta) - n_{\bar{s},1}(t)$ and $V \equiv n_{\bar{s},1}(t)$, let $\tilde{\zeta}_i$ be an i.i.d. copy of ζ_i , we only need to show
- a. $n_{\bar{s},1}(\varepsilon_0) = 0$ (obvious);
 - b. W is independent of V ;
 - c. $W \sim \text{Poisson}(\pi r_1 \delta)$;

By definition, we have

$$\begin{aligned}
 W &= I_{\mathcal{N}}(n_1(t - \varepsilon_0 + \delta) - n_1(t - \varepsilon_0)) \sum_{i=n_1(t - \varepsilon_0) + 1}^{n_1(t - \varepsilon_0 + \delta)} \zeta_i \\
 &= I_{\mathcal{N}}(n_1(t - \varepsilon_0 + \delta) - n_1(t - \varepsilon_0)) \sum_{i=1}^{n_1(t - \varepsilon_0 + \delta) - n_1(t - \varepsilon_0)} \tilde{\zeta}_i \\
 V &= I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \sum_{i=1}^{n_1(t - \varepsilon_0)} \zeta_i
 \end{aligned}$$

so, W is independent of V ; next is to obtain $\mathbf{P}(W)$, for $w \in \mathcal{N}^*$,

$$\begin{aligned} \mathbf{P}(W = w) &= \sum_{l=w}^{\infty} \mathbf{P}(n_1(t - \varepsilon_0 + \delta) - n_1(t - \varepsilon_0) = l) \binom{l}{w} \pi^w (1 - \pi)^{l-w} \\ &= \sum_{l=w}^{\infty} \frac{(r_1 \delta (1 - \pi))^{l-w}}{(l-w)!} \times \frac{(\pi r_1 \delta)^w}{w!} e^{-r_1 \delta} \\ &\sim \text{Poisson}(\pi r_1 \delta). \end{aligned}$$

(iii) let $\tilde{\mathcal{P}}\mathcal{P}_1$ and $\tilde{\varepsilon}_i$ be the i.i.d. copy of $\mathcal{P}\mathcal{P}_1$ and ε_i , we have

$$\begin{aligned} n_{\tilde{s},2} &\sim I_{\mathcal{N}}(\tilde{n}_1(\varepsilon_0)) \sum_{i=1}^{\tilde{n}_1(\varepsilon_0)} \tilde{\zeta}_i I_{[0,\varepsilon_0)}(\tilde{t}_{1,i} + \tilde{\varepsilon}_i) \\ &\sim I_{\mathcal{N}}(\tilde{n}_1(\varepsilon_0)) \sum_{i=1}^{\tilde{n}_1(\varepsilon_0)} \tilde{\zeta}_i I_{[0,\varepsilon_0)}(u_i + \tilde{\varepsilon}_i), \end{aligned}$$

where u_i is an i.i.d. sequence with distribution $U[0, \varepsilon_0]$, so it shares the same form of $n_{\tilde{s},1}$ but with a different $\tilde{\zeta}_i$; a simple calculation shows that

$$\mathbf{P}(\tilde{\zeta}_i I_{[0,\varepsilon_0)}(u_i + \tilde{\varepsilon}_i) = 1) = \pi/2$$

so result follows. ■

Proof of Lemma 3:

(i)

$$\begin{aligned} \mathbf{E}I_{\mathcal{N}}(n_1(t - \varepsilon_0)) &\sum_{i=1}^{n_1(t-\varepsilon_0)} \zeta_i e^{-\frac{t-(t_{1,i}+\varepsilon_i)}{\tau}} \\ &= \mathbf{E}I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \sum_{i=1}^{n_1(t-\varepsilon_0)} \zeta_i e^{-\frac{t-\varepsilon_0-t_{1,i}-(\varepsilon_i-\varepsilon_0)}{\tau}} \\ &= \mathbf{E}\zeta_1 \mathbf{E}e^{\frac{\varepsilon_1-\varepsilon_0}{\tau}} \mathbf{E}I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \sum_{i=1}^{n_1(t-\varepsilon_0)} e^{-\frac{t-\varepsilon_0-t_{1,i}}{\tau}} \\ &= \pi \times \frac{\tau}{\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) \times r_1 \tau (1 - e^{-\frac{t-\varepsilon_0}{\tau}}) \end{aligned}$$

$$\begin{aligned}
& \mathbf{E}I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \left[\sum_{i=1}^{n_1(t-\varepsilon_0)} \zeta_i e^{-\frac{t-(t_{1,i}+\varepsilon_i)}{\tau}} \right]^2 \\
&= \mathbf{E} \sum_{i=1}^{n_1(t-\varepsilon_0)} \zeta_i e^{-\frac{2[(t-\varepsilon_0-t_{1,i})+(\varepsilon_0-\varepsilon_i)]}{\tau}} \\
&+ \mathbf{E} \sum_{i \neq l} \zeta_i \zeta_l e^{-\frac{\varepsilon_0-\varepsilon_i}{\tau}-\frac{\varepsilon_0-\varepsilon_l}{\tau}} e^{-\frac{(t-\varepsilon_0)-t_{1,i}}{\tau}-\frac{(t-\varepsilon_0)-t_{1,l}}{\tau}} \\
&= \pi \mathbf{E} e^{-\frac{2(\varepsilon_0-\varepsilon_1)}{\tau}} \mathbf{E} n_1(t - \varepsilon_0) \mathbf{E} e^{-\frac{2(t-\varepsilon_0-u)}{\tau}} \\
&+ (\mathbf{E} \zeta_1)^2 \mathbf{E} (n_1^2(t - \varepsilon_0) - n_1(t - \varepsilon_0)) (\mathbf{E} e^{-\frac{\varepsilon_0-\varepsilon_i}{\tau}})^2 (\mathbf{E} e^{-\frac{t-\varepsilon_0-u}{\tau}})^2 \\
&= \pi \times \frac{\tau}{2\varepsilon_0} \left[1 - e^{-\frac{2\varepsilon_0}{\tau}} \right] \times r_1 \frac{\tau}{2} \left[1 - e^{-\frac{2(t-\varepsilon_0)}{\tau}} \right] \\
&\quad + \left[\pi \frac{\tau}{\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) r_1 \tau (1 - e^{-\frac{t-\varepsilon_0}{\tau}}) \right]^2
\end{aligned}$$

(ii)

$$\begin{aligned}
& \mathbf{E}I_{\mathcal{N}}(n_1(t) - n_1(t - \varepsilon_0)) \sum_{i=n_1(t-\varepsilon_0)+1}^{n_1(t)} \zeta_i I_{[0,t)}(t_{1,i} + \varepsilon_i) e^{-\frac{t-(t_{1,i}+\varepsilon_i)}{\tau}} \\
&= \pi \mathbf{E} \sum_{i=1}^{n_1(\varepsilon_0)} I_{[0,\varepsilon_0)}(t_{1,i} + \varepsilon_i) e^{-\frac{\varepsilon_0-(t_{1,i}+\varepsilon_i)}{\tau}} \\
&= \pi \mathbf{E} n_1(\varepsilon_0) \mathbf{E} \left(I_{[0,\varepsilon_0)}(u + \varepsilon_1) e^{-\frac{\varepsilon_0-(u+\varepsilon_1)}{\tau}} \right) \\
&= \pi \times r_1 \varepsilon_0 \times \left(\frac{\tau}{\varepsilon_0} \right)^2 (e^{-\frac{\varepsilon_0}{\tau}} - (1 - \frac{\varepsilon_0}{\tau}))
\end{aligned}$$

■

Proof of Lemma 4:

(i) That $\mathbf{E}q_s(t) = 0$ is an easy verification;

$$\begin{aligned}
q_s(t + \delta) &= I_{\mathcal{N}}(n_s(t + \delta)) \sum_{i=1}^{n_s(t+\delta)} e^{-\frac{t+\delta-t_i}{\tau}} - r\tau(1 - e^{-\frac{t+\delta}{\tau}}) \\
&= I_{\mathcal{N}}(n_s(t)) \sum_{i=1}^{n_s(t)} e^{-\frac{t+\delta-t_i}{\tau}} - r\tau(1 - e^{-\frac{t}{\tau}})e^{-\frac{\delta}{\tau}} \\
&+ I_{\mathcal{N}}(n_s(t + \delta) - n_s(t)) \sum_{i=n_s(t)+1}^{n_s(t+\delta)} e^{-\frac{t+\delta-t_i}{\tau}} - r\tau(1 - e^{-\frac{t+\delta}{\tau}} - (1 - e^{-\frac{t}{\tau}})e^{-\frac{\delta}{\tau}}) \\
&\stackrel{d}{=} e^{-\frac{\delta}{\tau}}q_s(t) + \tilde{q}_s(\delta) \quad (\text{by Lemma 2-(i)})
\end{aligned}$$

(ii) by Lemma 3-(i), let $\varepsilon_0 \downarrow 0$ and $\pi = 1$, we have $\mathbf{E}q_s^2(t) = \frac{r\tau}{2}(1 - e^{-\frac{2t}{\tau}})$;

by (ii), we have $c(t, \delta) \equiv Cov(q_s(t), q_s(t + \delta)) = e^{-\frac{\delta}{\tau}}\mathbf{E}q_s^2(t)$.

(iii)

■

Proof of Lemma 5:

(i) easy to check;

(ii) by Lemma 4(ii) and Formula 3.6, we have

$$q_1(t) \stackrel{d}{=} e^{-\frac{\varepsilon_0}{\tau}}q_1(t - \varepsilon_0) + \tilde{q}_1(\varepsilon_0)$$

$$q_2(t) = q_2^*(t) + q_{n_{\bar{s}},1}(t) + q_{n_{\bar{s}},2}(t)$$

so, $\mathbf{E}q_{1,2}(t) = e^{-\frac{\varepsilon_0}{\tau}} \mathbf{E}q_1(t - \varepsilon_0)q_{n_{\bar{s}},1} + \mathbf{E}\tilde{q}_1(\varepsilon_0)q_{n_{\bar{s}},2}$.

$$\begin{aligned}
\therefore \mathbf{E}I_{\mathcal{N}}(n_1(t - \varepsilon_0)) &= \sum_{i,l=1}^{n_1(t-\varepsilon_0)} \zeta_l e^{-\frac{\varepsilon_0 - \varepsilon_l}{\tau}} e^{-\frac{2(t-\varepsilon_0) - \varepsilon_{1,i} - \varepsilon_{1,l}}{\tau}} \\
&= \mathbf{E}I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \sum_{i=1}^{n_1(t-\varepsilon_0)} \zeta_i e^{-\frac{\varepsilon_0 - \varepsilon_i}{\tau}} e^{-\frac{2(t-\varepsilon_0) - \varepsilon_{1,i}}{\tau}} \\
&+ \mathbf{E}I_{\mathcal{N}}(n_1(t - \varepsilon_0)) \sum_{i \neq l} \zeta_l e^{-\frac{\varepsilon_0 - \varepsilon_l}{\tau}} e^{-\frac{2(t-\varepsilon_0) - \varepsilon_{1,i} - \varepsilon_{1,l}}{\tau}} \\
&= r_1 \pi \frac{\tau}{\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) \frac{\tau}{2} (1 - e^{-\frac{2(t-\varepsilon_0)}{\tau}}) + (r_1 \tau (1 - e^{-\frac{t-\varepsilon_0}{\tau}}))^2 \pi \frac{\tau}{\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) \\
&\therefore \mathbf{E}q_1(t - \varepsilon_0)q_{n_{\bar{s}},1} = r_1 \pi \frac{\tau}{\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) \frac{\tau}{2} (1 - e^{-\frac{2(t-\varepsilon_0)}{\tau}})
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}I_{\mathcal{N}}(\tilde{n}_1(\varepsilon_0)) &= \sum_{i=1}^{\tilde{n}_1(\varepsilon_0)} \tilde{\zeta}_i I_{[0,\varepsilon_0]}(\tilde{t}_{1,i} + \tilde{\varepsilon}_i) e^{-\frac{2\varepsilon_0 - (2\tilde{t}_{1,i} + \tilde{\varepsilon}_i)}{\tau}} \\
&= \mathbf{E}\tilde{n}_1(\varepsilon_0) \times \mathbf{E}\tilde{\zeta}_1 \times \iint_{\substack{0 < u + \varepsilon < \varepsilon_0 \\ 0 < u, \varepsilon < \varepsilon_0}} \frac{1}{\varepsilon_0^2} e^{-\frac{2\varepsilon_0 - (2u + \varepsilon)}{\tau}} du d\varepsilon \\
&= r_1 \varepsilon_0 \pi \frac{\tau^2}{\varepsilon_0^2} \left[1 - e^{-\frac{\varepsilon_0}{\tau}} - \frac{1}{2} (1 - e^{-\frac{2\varepsilon_0}{\tau}}) \right]
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}I_{\mathcal{N}}(\tilde{n}_1(\varepsilon_0)) &= \sum_{i \neq l}^{\tilde{n}_1(\varepsilon_0)} \tilde{\zeta}_i I_{[0,\varepsilon_0]}(\tilde{t}_{1,i} + \tilde{\varepsilon}_i) e^{-\frac{2\varepsilon_0 - (2\tilde{t}_{1,i} + \tilde{\varepsilon}_i)}{\tau}} \\
&= \mathbf{E}(\tilde{n}_1^2(\varepsilon_0) - \tilde{n}_1(\varepsilon_0)) \times \mathbf{E}\tilde{\zeta}_1 I_{[0,\varepsilon_0]}(u + \tilde{\varepsilon}_1) e^{-\frac{\varepsilon_0 - (u + \tilde{\varepsilon}_1)}{\tau}} \times \mathbf{E}e^{-\frac{\varepsilon_0 - u}{\tau}} \\
&= (r_1 \varepsilon_0)^2 \times \pi \frac{\tau^2}{\varepsilon_0^2} \left(\frac{\varepsilon_0}{\tau} - 1 + e^{-\frac{\varepsilon_0}{\tau}} \right) \times \frac{\tau}{\varepsilon_0} \left[1 - e^{-\frac{\varepsilon_0}{\tau}} \right]
\end{aligned}$$

$$\begin{aligned}
& \because \mathbf{E}\tilde{q}_1(\varepsilon_0)q_{n_{\tilde{s}},2} \\
& \quad = \mathbf{E}I_{\mathcal{N}}(\tilde{n}_1(\varepsilon_0)) \sum_{i,l=1}^{\tilde{n}_1(\varepsilon_0)} \tilde{\zeta}_i I_{[0,\varepsilon_0]}(\tilde{t}_{1,i} + \tilde{\varepsilon}_i) e^{-\frac{2\varepsilon_0 - (2\tilde{t}_{1,i} + \tilde{\varepsilon}_i)}{\tau}} \\
& \quad - r_1\tau(1 - e^{-\frac{\varepsilon_0}{\tau}}) \times \pi r_1\varepsilon_0 \frac{\tau^2}{\varepsilon_0^2} \left(e^{-\frac{\varepsilon_0}{\tau}} - \left(1 - \frac{\varepsilon_0}{\tau}\right) \right) \\
& \quad \therefore \mathbf{E}\tilde{q}_1(\varepsilon_0)q_{n_{\tilde{s}},2} = r_1\varepsilon_0\pi \frac{\tau^2}{\varepsilon_0^2} \left[1 - e^{-\frac{\varepsilon_0}{\tau}} - \frac{1}{2}(1 - e^{-\frac{2\varepsilon_0}{\tau}}) \right]
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\mathbf{E}q_{1,2}(t) &= e^{-\frac{\varepsilon_0}{\tau}} \frac{r_1\pi\tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) (1 - e^{-\frac{2(t-\varepsilon_0)}{\tau}}) + \frac{r_1\pi\tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}})^2 \\
&= \frac{r_1\pi\tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) (1 - e^{-\frac{2t-\varepsilon_0}{\tau}})
\end{aligned}$$

■

Proof of Theorem 1:

(i)

$$\begin{aligned}
Q_{1,2,N}^2 &= \left[\sum_{n=1}^N q_1\left(\frac{n}{N}t\right) q_2\left(\frac{n}{N}t\right) \frac{t}{N} \right]^2 \\
&= \underbrace{\sum_{i=1}^N q_1^2\left(\frac{i}{N}t\right) q_2^2\left(\frac{i}{N}t\right) \frac{t^2}{N^2}}_{\text{Part1}} \\
&\quad + 2 \underbrace{\sum_{i=1}^{N-1} \sum_{l=1}^{N-i} q_1\left(\frac{i}{N}t\right) q_2\left(\frac{i}{N}t\right) q_1\left(\frac{i+l}{N}t\right) q_2\left(\frac{i+l}{N}t\right) \frac{t^2}{N^2}}_{\text{Part2}}
\end{aligned}$$

$\because q_1(\cdot)$ and $q_2(\cdot)$ are bounded

$$\therefore \lim_{N \rightarrow \infty} \mathbf{E}\text{Part1} = 0$$

$$\begin{aligned} \mathbf{E}\text{Part2} &= 2 \sum_{i=1}^{N-1} \sum_{l=1}^{N-i} e^{-\frac{2}{\tau} \frac{l}{N} t} \mathbf{E}q_1^2\left(\frac{i}{N}t\right) \mathbf{E}q_2^2\left(\frac{i}{N}t\right) \frac{t^2}{N^2} \quad (\text{by Lemma 5-(i)}) \\ &= 2 \sum_{i=1}^{N-1} \frac{r_1 r_2 \tau^2}{4} (1 - e^{-\frac{2}{\tau} \frac{i}{N} t})^2 \sum_{l=1}^{N-i} e^{-\frac{2lt}{\tau N}} \frac{t^2}{N^2} \\ &= 2 \sum_{i=1}^{N-1} \frac{r_1 r_2 \tau^2}{4} (1 - e^{-\frac{2}{\tau} \frac{i}{N} t})^2 \frac{e^{-\frac{2t}{\tau N}} \left[1 - e^{-\frac{2t}{\tau N}(N-i)}\right]}{1 - e^{-\frac{2t}{\tau N}}} \frac{t^2}{N^2} \\ &\stackrel{N\theta_N \equiv \frac{2t}{\tau}}{=} \frac{r_1 r_2 \tau^4}{8} \sum_{i=1}^{N-1} \theta_N^2 (1 - e^{-i\theta_N})^2 \frac{e^{-\theta_N} (1 - e^{-(N-i)\theta_N})}{1 - e^{-\theta_N}} \\ &= \frac{r_1 r_2 \tau^4}{8} \sum_{i=1}^{N-1} (1 - e^{-i\theta_N})^2 (1 - e^{-(N-i)\theta_N}) \theta_N \times \frac{\theta_N e^{-\theta_N}}{1 - e^{-\theta_N}} \end{aligned}$$

Notice $\lim_{N \rightarrow \infty} \theta_N \downarrow 0$, we have

$$\lim_{N \rightarrow \infty} \frac{\theta_N e^{-\theta_N}}{1 - e^{-\theta_N}} = 1$$

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{i=1}^{N-1} (1 - e^{-i\theta_N})^2 (1 - e^{-(N-i)\theta_N}) \theta_N &= \int_0^{\frac{2t}{\tau}} (1 - e^{-s})^2 (1 - e^{-(\frac{2t}{\tau} - s)}) ds \\ &= s - 2.5 + 2(1 + s)e^{-s} + \frac{1}{2}e^{-2s} \Big|_0^{\frac{2t}{\tau}} \\ &\cong \frac{2t}{\tau} - 2.5 \quad (\because t \gg \tau) \end{aligned}$$

So, Equation (3.8) is obtained when $t \gg \tau$. This leads us to conjecture that $\lim_{N \rightarrow \infty} Q_{1,2,N}(t)$ may well be a Brownian Motion (for t sufficiently large).

(ii) by Lemma 5-(ii), we have(for $t > \varepsilon_0$)

$$\begin{aligned}
\mathbf{E}Q_{1,2,N}(t) &= \sum_{n=1}^N \mathbf{E}q_{1,2}\left(\frac{n}{N}t\right) \frac{t}{N} \\
&= \sum_{n=\lceil \frac{N\varepsilon_0}{t} \rceil + 1}^N \frac{t}{N} \frac{r_1 \pi \tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) (1 - e^{-\frac{2n}{N}t - \frac{\varepsilon_0}{\tau}}) \\
\therefore \lim_{N \rightarrow \infty} \mathbf{E}Q_{1,2,N}(t) &= \frac{r_1 \pi \tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) \int_{\varepsilon_0}^t (1 - e^{-\frac{2s - \varepsilon_0}{\tau}}) ds \\
&= \frac{r_1 \pi \tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) \left(s + \frac{\tau}{2} e^{\frac{\varepsilon_0}{\tau}} e^{-\frac{2s}{\tau}} \right) \Bigg|_{\varepsilon_0}^t \\
&\cong \frac{\pi r_1 \tau^2}{2\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}}) (t - \varepsilon_0 - \frac{\tau}{2} e^{-\frac{\varepsilon_0}{\tau}}) \quad (\because t \gg \tau)
\end{aligned}$$

■

Proof of Theorem 2:

Without loss of generality, we can set $-x_1(0) = x_2(0) = d_0 \gg c_0$; then we have

$$\frac{d(x_1(t) - x_2(t))}{dt} = 2q_1(t)q_2(t) \quad (\because d_{12}(t) > c_0 \Rightarrow A(x_2, x_1) = 1)$$

which (when $t < h_1$) is the same as

$$\begin{aligned}
\overline{x_2 x_1}^\rightarrow(t) &= \overline{x_2 x_1}^\rightarrow(0) + 2 \int_0^t q_{1,2}(s) ds \\
\therefore \overline{x_2 x_1}^\rightarrow(t) &= -2d_0 + 2 \lim_{N \rightarrow \infty} Q_{1,2,N}(t)
\end{aligned}$$

(i) by Theorem 1-(i),

$$\begin{aligned}
\mathbf{E}\overline{x_2 x_1}^\rightarrow(t) &= -2d_0 + 2\mathbf{E} \lim_{N \rightarrow \infty} Q_{1,2,N}(t) \\
&= -2d_0 + 2 \lim_{N \rightarrow \infty} \mathbf{E}Q_{1,2,N}(t) \\
&= -2d_0 \\
\therefore \mathbf{E}d_{12}(t) &= \mathbf{E}|\overline{x_2 x_1}^\rightarrow(t)| = d_{12}(0)
\end{aligned}$$

$$\begin{aligned}
\mathbf{Var}\overrightarrow{x_2x_1}(t) &= \mathbf{Var}(-2d_0 + 2 \lim_{N \rightarrow \infty} Q_{1,2,N}(t)) \\
&= 4\mathbf{Var}(\lim_{N \rightarrow \infty} Q_{1,2,N}(t)) \\
&= 4 \lim_{N \rightarrow \infty} \mathbf{Var}Q_{1,2,N}(t) \\
\therefore \mathbf{Var}d_{12}(t) &\cong r_1r_2\tau^3t
\end{aligned}$$

(ii) by Theorem 1-(ii)

$$\begin{aligned}
\mathbf{E}\overrightarrow{x_2x_1}(t) &= -2d_0 + 2\mathbf{E} \lim_{N \rightarrow \infty} Q_{1,2,N}(t) \\
&= -2d_0 + 2 \lim_{N \rightarrow \infty} \mathbf{E}Q_{1,2,N}(t) \\
&\cong -2d_0 + \frac{\pi r_1 \tau^2}{\varepsilon_0} (1 - e^{-\frac{\varepsilon_0}{\tau}})(t - \varepsilon_0 - \frac{\tau}{2} e^{-\frac{\varepsilon_0}{\tau}}) \\
&\cong -2d_0 + \pi r_1 \tau t \\
\therefore \mathbf{E}d_{12}(t) &= \mathbf{E}|\overrightarrow{x_2x_1}(t)| \downarrow c_0 \quad (t \uparrow h_1)
\end{aligned}$$

■

Proof of Theorem 3:

(i) for $m = N - 1$, conclusion holds; assume for $1 \leq m \leq N - 1$ it holds, we prove it holds for $m - 1$ (math induction).

$$\begin{aligned}
&\mathbf{E} \left(f_{ij,N}(t) | \mathcal{F}_{\frac{m-1}{N}t} \right) = \mathbf{E} \left\{ \left[\mathbf{E} \left(f_{ij,N}(t) | \mathcal{F}_{\frac{m}{N}t} \right) \right] | \mathcal{F}_{\frac{m-1}{N}t} \right\} \\
&= \mathbf{E} \left\{ \left[f_{ij,N} \left(\frac{m}{N}t \right) + \frac{t}{N} q_{ij} \left(\frac{m}{N}t \right) \overrightarrow{1}_{ij} \left(\frac{m}{N}t \right) (1 + \rho_{N-1-m}) \right] | \mathcal{F}_{\frac{m-1}{N}t} \right\} \\
&\quad + \mathbf{E} \left\{ \left[\mathbf{E} \left\{ \sum_{n=m}^{N-1} \left[O^{(1)} \left(\frac{n}{N}t \right) + O^{(2)} \left(\frac{n}{N}t \right) \right] | \mathcal{F}_{\frac{n}{N}t} \right\} \right] | \mathcal{F}_{\frac{m-1}{N}t} \right\}
\end{aligned}$$

using F.1, rewrite the expectation above

$$\begin{aligned}
&= f_{ij,N} \left(\frac{m}{N}t \right) + \frac{t}{N} \mathbf{E} \left[q_{ij} \left(\frac{m}{N}t \right) | \mathcal{F}_{\frac{m-1}{N}t} \right] \overrightarrow{1}_{ij} \left(\frac{m}{N}t \right) (1 + \rho_{N-1-m}) \\
&\quad + \mathbf{E} \left\{ \sum_{n=m}^{N-1} \left[O^{(1)} \left(\frac{n}{N}t \right) + O^{(2)} \left(\frac{n}{N}t \right) \right] | \mathcal{F}_{\frac{m-1}{N}t} \right\}
\end{aligned}$$

now using F.2 and decomposition F.5, we have

$$\begin{aligned}
& \frac{t}{N} \mathbf{E} \left[q_{ij} \left(\frac{m}{N} t \right) | \mathcal{F}_{\frac{m-1}{N} t} \right] \overrightarrow{1}_{ij} \left(\frac{m}{N} t \right) (1 + \rho_{N-1-m}) \\
&= \frac{t}{N} e^{-\frac{1}{\tau} \frac{2t}{N}} q_{ij} \left(\frac{m-1}{N} t \right) (1 + \rho_{N-1-m}) \times \\
& \left\{ \overrightarrow{1}_{ij} \left(\frac{m-1}{N} t \right) \left[1 + \tilde{O}^{(1)} \left(\frac{m-1}{N} t \right) \right] + \frac{\frac{t}{N} D_{ij} \left(\frac{m-1}{N} t \right)}{\left| \overrightarrow{1}_{ij} \left(\frac{m-1}{N} t \right) + \frac{t}{N} D_{ij} \left(\frac{m-1}{N} t \right) \right|} \right\} \\
&= \frac{t}{N} q_{ij} \left(\frac{m-1}{N} t \right) \rho_{N-1-(m-1)} \overrightarrow{1}_{ij} \left(\frac{m-1}{N} t \right) \\
&+ \frac{t}{N} q_{ij} \left(\frac{m-1}{N} t \right) \rho_{N-1-(m-1)} \overrightarrow{1}_{ij} \left(\frac{m-1}{N} t \right) \tilde{O}^{(1)} \left(\frac{m-1}{N} t \right) \\
&+ \frac{t}{N} q_{ij} \left(\frac{m-1}{N} t \right) \rho_{N-1-(m-1)} \frac{\frac{t}{N} D_{ij} \left(\frac{m-1}{N} t \right)}{\left| \overrightarrow{1}_{ij} \left(\frac{m-1}{N} t \right) + \frac{t}{N} D_{ij} \left(\frac{m-1}{N} t \right) \right|}
\end{aligned}$$

by noticing that

$$f_{ij,N} \left(\frac{m}{N} t \right) = f_{ij,N} \left(\frac{m-1}{N} t \right) + \frac{t}{N} q_{ij} \left(\frac{m-1}{N} t \right) \overrightarrow{1}_{ij} \left(\frac{m-1}{N} t \right)$$

and rearranging terms, we see that the result holds for $m-1$.

(ii) the same method as in (i), just notice that F.3 replaces F.2 in this procedure. ■

Proof of Theorem 4:

Notice that $\mathbf{E} [x_1(t) - x_1(0)] = \sum_{k=2}^K \mathbf{E} f_{1k,N}(t)$, we only need to estimate $\mathbf{E} f_{1k,N}(t)$ by Theorem (3).

(i) by Theorem (3)-(i), we have

$$\begin{aligned}
|\mathbf{E} \left(f_{1k,N}(t) | \mathcal{F}_{\frac{0}{N} t} \right)| &= \underbrace{\left| f_{1k,N} \left(\frac{0}{N} t \right) + \frac{t}{N} q_{1k} \left(\frac{0}{N} t \right) \overrightarrow{1}_{1k} \left(\frac{0}{N} t \right) (1 + \rho_{N-1-0}) \right|}_{=0} \\
&+ \mathbf{E} \left\{ \left| \sum_{n=0}^{N-1} \left[O^{(1)} \left(\frac{n}{N} t \right) + O^{(2)} \left(\frac{n}{N} t \right) \right] | \mathcal{F}_{\frac{0}{N} t} \right| \right\} \\
&\leq \sum_{n=0}^{N-1} \frac{t}{N} 2M^2 \rho_{N-1-n} \frac{\delta_0}{1 - \delta_0} \quad (\because F.6) \\
&\leq \frac{2M^2}{1 - \delta_0} \frac{ut}{1 - u} \delta_0 \quad (\because F.4)
\end{aligned}$$

notice $d_0 \rightarrow +\infty \iff \delta_0 \rightarrow 0$, we see

$$\lim_{d_0 \rightarrow +\infty} |\mathbf{E}[x_1(t) - x_1(0)]| \leq \lim_{\delta_0 \rightarrow +\infty} \frac{2(K-1)M^2}{1-\delta_0} \frac{ut}{1-u} \delta_0 = 0$$

(ii) from above (Theorem 4-(i)), we have

$$\lim_{d_0 \rightarrow +\infty} \mathbf{E}[x_1(t) - x_1(0)] = \lim_{d_0 \rightarrow \infty} \sum_{k=2}^{k_1} \mathbf{E}f_{1k,N}(t)$$

by Theorem (3)-(ii), we have (for $2 \leq k \leq k_1$)

$$\begin{aligned} \mathbf{E}\left(f_{1k,N}(t) \middle| \mathcal{F}_{\frac{0}{N}t}\right) &= \underbrace{f_{1k,N}\left(\frac{0}{N}t\right) + \frac{t}{N}q_{1k}\left(\frac{0}{N}t\right)\overrightarrow{1}_{1k}\left(\frac{0}{N}t\right)(1 + \rho_{N-1-0})}_{P1} \\ &\quad + \mathbf{E}\left\{\underbrace{\sum_{n=0}^{N-1} \left[O^{(1)}\left(\frac{n}{N}t\right) + O^{(2)}\left(\frac{n}{N}t\right)\right]}_{P2} \middle| \mathcal{F}_{\frac{0}{N}t}\right\} \\ &\quad + \underbrace{\frac{t}{N}c_1(1k)\overrightarrow{1}_{1k}\left(\frac{0}{N}t\right) \sum_{n=0}^{N-1-0} \rho_n}_{P3} + \mathbf{E}\left\{\underbrace{\sum_{n=0}^{N-1} \left[R^{(1)}\left(\frac{n}{N}t\right) + R^{(2)}\left(\frac{n}{N}t\right)\right]}_{P4} \middle| \mathcal{F}_{\frac{0}{N}t}\right\} \end{aligned}$$

by same method as in (i), we see $\lim_{d_0 \rightarrow +\infty} |P1 + P2| = 0$; using F.4 and F.6

$$\begin{aligned} P3 &= \frac{t}{N}c_1(1k)\overrightarrow{1}_{1k}(0) \frac{u}{1-u} \left[N - \frac{1-u^N}{1-u} \right] \\ &= c_1(1k)\overrightarrow{1}_{1k}(0) \frac{u}{1-u} \left[1 - \frac{1}{N} \frac{1-u^N}{1-u} \right] t \\ \lim_{d_0 \rightarrow +\infty} |P4| &\leq \sum_{n=0}^{N-1} \frac{t}{N}c_1(1k) \frac{2\delta_0}{1-\delta_0} \frac{u}{1-u} \left[n - \frac{1-u^n}{1-u} \right] \\ &= c_1(1k) \frac{2\delta_0}{1-\delta_0} \frac{u}{1-u} \left[\frac{N-1}{2} - \frac{1}{1-u} + \frac{1}{N} \frac{1-u^N}{(1-u)^2} \right] t \end{aligned}$$

notice in the usual setting $N \gg 1$, we have

$$\lim_{d_0 \rightarrow \infty} \frac{|P3|}{|P4|} \cong \lim_{d_0 \rightarrow \infty} \frac{1}{\delta_0 N} = \lim_{\delta_0 \rightarrow 0} \frac{\rho\tau}{t} \frac{1}{\delta_0} = +\infty$$

so finally, we obtain

$$\begin{aligned}
\lim_{d_0 \rightarrow +\infty} \mathbf{E} [x_1(t) - x_1(0)] &= \lim_{d_0 \rightarrow \infty} \sum_{k=2}^{k_1} \mathbf{E} f_{1k,N}(t) \\
&= \lim_{d_0 \rightarrow \infty} \sum_{k=2}^{k_1} c_1(1k) \overrightarrow{1}_{1k}(0) \frac{u}{1-u} t (1 + O(\delta_0 \frac{t}{\rho\tau})) \\
&= \sum_{k=2}^{k_1} c_1(1k) \overrightarrow{1}_{1k}(0) \frac{u}{1-u} t
\end{aligned}$$

INDEX

- LCP*, 12, 13
- MP*, 12

- active effect, 6
- attenuation function, 7

- BARS
 - Bayesian Adaptive Regression Splines, 3
- Bayesian belief network, 3, 19
 - BBN, 5, 8, 9
 - DBN, 3–5, 8, 9
 - DBN interaction mode, 5
 - dynamic Bayesian network, 1
 - semi-Bayesian belief network, 19
- Bernoulli distribution, 9
- beta distribution, 6

- charge process, 6, 7, 14
- cluster algorithm, 6
- cluster recognition, 7
- computational neuroscience, 1, 3
- control parameter, 19, 21, 24
- copy matrix, 6
- copy probability, 11
- crosscorrelogram, 8

- DAG, 1, 8
- DCG, 8
 - directed cyclic graph, 8
- depth of copy matrix, 15
- directed acyclic graph, 1
- Dirichlet distribution, 2, 23
 - Bounded Dirichlet distribution, 23, 24, 35
 - Matrix Dirichlet distribution, 24
- distance, 7
- dynamic system, 6, 9, 11
- dynamical system, 7

- effective time range, 9

- equal firing rates network, 8
- Euclidean space, 9
- excitation, 8

- firing activity, 3
- firing intensity, 3
- firing rate, 7
- flow plot, 11
- force field, 7

- GCA, 4, 6–8
 - GCA Distance, 7
 - GCA interaction mode, 6
 - gravitational clustering algorithm, 1, 4, 6
- GCA procedure, 7
- GCA's distances, iii
- GCAD, iii, 13–15
- generalized mixture model, 2, 20
- Gibbs sampling procedure, 33
- goodness-of-fit, 9

- indicator function, 20
- inhibition, 8
- intensity, 4
- interaction, 3, 4
- interaction mode, 7, 8

- joint peristimulus time histogram, 8
- JPSTH, 8

- latency distribution, 6, 9
- likelihood function, 9

- machine learning, 9
- mathematical tool, 8
- matrix, 2, 21, 24
- MCMC, 2, 9
- mixture model, 2, 4, 21
- model identifiability, 25

- moment, 13, 14
- nervous system, 3
- network model, 6
- normalization function, 6
- numerical integration, 9
- oscillation, 8
- partition, 4
- pattern recognition, 4, 9
- physics, 9
- Poisson process, iii, 11, 12
- probabilistic network, 5
- process
 - empirical process, 3
 - Gamma process, 3
 - inverse Gaussian process, 3
 - point process, 1, 3, 4
 - Poisson process, 3, 4
- product, 14
- random time lag, 8
- Sampler
 - Metropolis-Hastings sampler, 2
 - weighted sampler, 2
- sampler
 - Metropolis-Hastings sampler, 36
- self-refractory effect, 5
- sensitivity, 8
- snowflake plot, 3
 - probabilistic model for snowflake plot, 3
- spike, 1, 3
 - multiple spike trains, 1, 2, 4, 8
 - spike train, 1, 3, 4
 - spontaneous spike train, 6
- spike sorting, 9
- stationary independent increment, 12
- temporal firing rate, 8
- time decaying constant, 6
- trigger neuron, 6
- unequal firing rates network, 8
- unequal rates network, 10
- unidentifiability, 20
- unit vector, 16

BIBLIOGRAPHY

- [1] D. R. Brillinger.
Nerve cell spike train data analysis: A progression of technique.
Journal of American Statistical Association, pages 260–271, 1992.
- [2] E. Brown, R. Kass, and P. Mitra.
Multiple neural spike train data analysis: state-of-the-art and future challenges.
Nature: Neuroscience, 7:456–461, 2004.
- [3] E. N. Brown.
Theory of point process for neural systems.
In C.C.Chow, B.Gutkin, D.Hansel, C.Meunier, and J.Dalibard, editors, *Methods and Models in Neurophysics*, chapter 14, pages 691–726. Paris, Elsevier, 2005.
- [4] G. Czanner, S. Grun, and S. Iyengar.
Theory of the snowflake plot and its relations to higher-order analysis methods.
Neural Computation, 17(7):1456–1479, 2005.
- [5] G. L. Gerstein and A. M. H. J. Aertsen.
Representation of cooperative firing activity among simultaneously recorded neurons.
Journal of Neurophysiology, 54(6):1513–1528, 1985.
- [6] A. L. Hodgkin and A. F. Huxley.
A quantitative description of membrane potential on sodium conductance in the giant axon of loligo.
Journal of Physiology, 117(4):500–544, 1952.
- [7] S. Iyengar.
The analysis of multiple neural spike train analysis.
In N. Bolakrishnan, editor, *Advances in Methodological and Applied Aspects of Probability and Statistics*, pages 507–524. Gordon and Breack, 2001.
- [8] R. E. Kass, V. Ventura, and C. Cai.
Statistical smoothing of neuronal data.
Network: Computation in Neural Systems, 14:5–15, 2003.
- [9] C. Koch.
Biophysics of Computation: Information Processing in Single Neurons.
Oxford University Press, 1999.
- [10] F. Rigat, M. d. Gunst, and J. v. Pelt.
Bayesian modeling and analysis of spatio-temporal neuronal networks.
Bayesian Analysis, in press.
- [11] K. J. Utikal.
Nonparametric inference for markovian interval processes.
Stochastic processes and their applications, pages 1–23, 1997.
- [12] V. Ventura, R. Carta, R. E. Kass, S. N. Gettner, and C. R. Olson.
Statistical analysis of temporal evolution in single-neuron firing rates.
Biostatistics, 1:1–20, 2002.