RIKEN BSI BSIS Technical Report: No05-1.

Comparison of single spike train descriptive models by information geometric measure

Hiroyuki Nakahara[1,3,*], Shun-ichi Amari[1], Barry J. Richmond[2]

[1] Lab. for Mathematical Neuroscience, RIKEN Brain Science Institute

2-1 Hirosawa, Wako, Saitama, 351-0198 Japan

[2] Lab. of Neuropsychology, National Institute of Mental Health

National Institutes of Health, Bethesda, MD 20892, U.S.A.

[3] Correspondence to *hiro@brain.riken.jp*

Abstract

In examining spike trains, different groups propose different models for describing the structure of these trains. The different models often seem quite similar, but because their formalism is different it is often not obvious how predictions made would differ. We use the information geometric measure, an orthogonal coordinate representation of point processes, to express different models of stochastic point processes in

*This technical report is an extended version of a journal paper, "A comparison of descriptive models of a single spike train by information geometric measure", including some examples and figures.

a common coordinate system. Within such a framework, it becomes straightforward to visualize the higher order correlations of different models and thereby assess the differences between models. Here we apply the information geometric measure to compare two similar, but not identical, models of neuronal spike trains, namely the inhomogeneous Markov and the mixture of Poisson models. It shows that they differ in the second and higher order interaction terms. For the mixture of Poisson model the higher order interactions (2nd and higher) are of comparable magnitude within each order, whereas for the inhomogeneous Markov model the 2nd and higher-order interactions have a structure of alternating signs over different orders. This provides guidance about what measurements of data that would separate the efficacy of the two models. As newer models are proposed they also can be compared to these models using information geometry.

1 Introduction

Over the past two decades studies of the information carrying properties of neuronal spike trains have become more intense and sophisticated. Many early studies of neuronal spike trains concentrated mainly on using general methods to reduce the dimensionality of the description. Recently, however, specific models have been developed to take into account findings from both experimental and theoretical biophysical data, (Dean, 1981; Richmond and Optican, 1987; Abeles, 1991; Bialek et al., 1991; Reid et al., 1992; Softky and Koch, 1993; Shadlen and Newsome, 1994; Victor and Purpura, 1997; Stevens and Zador, 1998; Oram et al., 1999; Shinomoto et al., 1999; Meister

and Berry, 1999; Baker and Lemon, 2000; Kass and Ventura, 2001; Reich et al., 2001; Brown et al., 2002; Wiener and Richmond, 2003; Beggs and Plenz, 2003; Fellous et al., 2004) (references therein). These newer approaches guess at specific structures that give rise to the spike trains in experimental data. Because these models have specific structures, fitting these models is translated into estimating the parameters of the model, rather than using general approaches to dimensionality reduction. There are several benefits of having these more descriptive models. First, all of the approaches describe data in a succinct manner. Second, the more principled models make their assumptions explicit, i.e., they declare which properties in data are considered important. Third, parametric models have 'practical' value for data analysis because the parameter values of a model can often be reasonably well estimated even with the limited number of samples that can be obtained in experiments.

When considering different models it is natural to ask which model is really good, or perhaps, more properly, what do we learn about the system from the different models, and in what ways are the models equivalent, or different? If the differences can be seen explicitly, experiments can be designed to evaluate features that distinguish the models. A powerful approach for distinguishing them is to project them into a single coordinate frame, especially an orthogonal coordinate one. Information geometric measure (IG) (Amari, 2001; Nakahara and Amari, 2002b) provides a orthogonal coordinate system to make such projections for models of point processes. Using the IG measure, we consider a probability space, where each 'point'

in the space corresponds to a probability distribution. Estimating the probability distribution of the spike train from experimental data corresponds to identifying the location and spread of a point in this space. In this context, different assumptions underlying the different models of the spike train correspond to different constraints of the search (Fig 1). Once different models are re-represented by the common coordinate system, it is possible to compare the nature of different models. Thus, we can compare the regions of the geometric expression that can be reached by the different models, and the regions that overlap and the regions that are unique can be identified.

Here we use the IG measure to compare two different stochastic models of spike trains, the inhomogeneous Markov (IM) (Kass and Ventura, 2001) and the mixture of Poisson (MP) models (Wiener and Richmond, 2003). Experimentally recorded spike trains generally depart from Poisson statistics (see Section 2.) The variance-to-mean relation is seldom one, the interval distribution is often not truly exponential, and distributions of counts from a counting interval are not Poisson. Recently, both IM and MP models have been proposed to deal with these deviations. Both treat the spike trains as point processes, but each emphasizes different properties. The intersection of the two models is an inhomogeneous Poisson process, which is a special and perhaps uninteresting case. The IM model model emphasizes the non-Poisson nature of interval distributions, whereas the MP model emphasizes the non-Poisson nature of the spike count distribution (Fig 2; Section 3). To compare the two models, we re-represent the two models by the IG measure as a common coordinate system (Section 4). It shows their predictions about

second and higher-order statistics (Section 5). The 'first-cut' coordinates of the IG representation shows how to use higher order interactions among spikes to distinguish between these models.

2 Preliminaries

Consider an inhomogeneous Poisson process. For a spike train of a single neuron, consider a time period of N bins, where each bin is so short that it can have at most a single spike. Each neuronal spike train is then represented by a binary random vector variable. Let $X^N = (X_1, \cdots, X_N)$ be N binary random variables and let $p(\boldsymbol{x}^N) = P\left[X^N = \boldsymbol{x}^N\right], \boldsymbol{x}^N = (x_1, \cdots, x_n), x_i = 0, 1$, be its probability, where $p(\boldsymbol{x}^N) > 0$ is assumed for all $\boldsymbol{x}$. Each X_i indicates a spike in the i-th bin, by $X_i = 1$, or no spike, by $X_i = 0$. With this notation, the inhomogeneous Poisson process is given by

$$p(\boldsymbol{x}^N) = \prod_i^N \eta_i^{x_i} \left(1 - \eta_i\right)^{1 - x_i}$$

where $\eta_i = E\left[x_i\right]$. The probability of a spike occurrence in a bin is independent from those of the other bins, i.e.

$$p(\boldsymbol{x}^N) = \prod_i^N p(x_i), \text{ where } p(x_i) = \eta_i^{x_i} \left(1 - \eta_i\right)^{1 - x_i}.$$

Then, $(\eta_1, \cdots, \eta_N)$, or the PSTH (peri-stimulus histogram), obtained from experimental data is sufficient to estimate parameter values of this model. In this condition experimental data analysis is simple, and spike generation from this model easy. This independence property leads to well known facts: the count statistics obey the Poisson distribution and the interval statistics

obey the exponential distribution. The simplicity makes the Poisson model popular as a choice for the descriptive model. Experimental findings often suggest that the empirical probability distributions of spike counts and intervals are close to the Poisson distribution and the exponential distribution, respectively, but, so frequently, they depart from these as well (Dean, 1981; Tolhurst et al., 1983; Gawne and Richmond, 1993; Gershon et al., 1998; Lee et al., 1998; Stevens and Zador, 1998; Maynard et al., 1999; Oram et al., 1999). These findings provoked studies considering a larger class of models, including the IM and MP models (see Section 3.)

The Poisson process occupies only a small subspace compared with a full space, or the original class of probability distributions $p(\boldsymbol{x}^N)$. The number of all possible spike pattern is 2^N, since $X^N \in \{0,1\}^N$. Therefore, each $p(\boldsymbol{x}^N)$ is given by 2^n probabilities

$$p_{i_1 \cdots i_N} = \text{Prob}\left\{X_1 = i_1, \cdots, X_N = i_N\right\}, \quad i_k = 0, 1, \text{ subject to } \sum_{i_1,\cdots,i_N} p_{i_1 \cdots i_N} = 1.$$

The set of all the possible probability distributions $\{p(\boldsymbol{x})\}$ forms a $\left(2^N - 1\right)$-dimensional manifold $\boldsymbol{S}_N$. To represent a 'point' in $\boldsymbol{S}_N$, i.e. a probability distribution $p(\boldsymbol{x}^N)$, one simple coordinate system, called the P-coordinate system is given by $\{p_{i_1 \cdots i_N}\}$ above, where $\{p_{i_1 \cdots i_N}\}$ corresponds to 2^N probabilities. Every set of values has 2^N probabilities, with each corresponding to a specific probability distribution, $p(\boldsymbol{x}^N)$. Since $\{p_{i_1 \cdots i_N}\}$ sums to one, the effective coordinate dimension is $2^N - 1$ (instead of 2^N).

For the information geometric measure (IG) we use two other coordinate systems. The first is the θ-coordinate system that is explained in Section 4,

and the second coordinate system is the η-coordinate systems, given by the expectation parameters,

$$\begin{aligned}
\eta_i &= E[x_i] = \text{Prob}\{x_i = 1\}, \quad i = 1, \cdots, n \\
\eta_{ij} &= E[x_i x_j] = \text{Prob}\{x_i = x_j = 1\}, \quad i < j \\
\eta_{i_1 i_2 \cdots i_l} &= E[x_{i_1} \cdots x_{i_l}] = \text{Prob}\{x_{i_1} = x_{i_2} = \cdots = x_{i_l} = 1\} \quad i_1 < i_2 < ... < i_l.
\end{aligned}$$

All η_{ijk}, etc., together have $2^N - 1$ components, that is,

$$\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_N) = (\eta_i, \eta_{ij}, \eta_{ijk}, \cdots, \eta_{12...N})$$

has $2^N - 1$ components, where we write $\boldsymbol{\eta}_1 = (\eta_i)$, $\boldsymbol{\eta}_2 = (\eta_{ij})$ and so on, forming the η-coordinate system in $\boldsymbol{S}_N$.

Any probability distribution of X^N can be completely represented by P-coordinates or η-coordinates, if and only if all of the coordinates are used. If any model of the probability distribution of X^N has fewer parameters than $2^N - 1$ (this is usually the case), the probability space in which the model lies is restricted. Since the Poisson process uses $\boldsymbol{\eta}_1$ as its coordinates, it lies in a much smaller subspace than the full space $\boldsymbol{S}_N$.

The components of the η-coordinates, $\boldsymbol{\eta}_1 = (\eta_i) = (\eta_1, \eta_2, ..., \eta_N)$, $(i = 1, ..., N)$ are familiar, since they would correspond to the PSTH in experimental data analysis. It is taken to represent the expectation of the time course of the neural firing, expressed as the probability of a spike at each time or as the firing frequency. Below we freely exchange the PSTH and $\boldsymbol{\eta}_1$ for simplicity. To be careful, however, we mention the difference between $\boldsymbol{\eta}_1$ and the probability density of firing, since the PSTH is often regarded

as the latter as well. The probability density will be re-calculated if the bin size changes so that it is invariant under the change of the bin size and has the unit as the number of firings per unit (infinitesimal) time. The firing frequency would become the probability density as the time resolution approaches zero. In data analysis, the firing frequency is then taken to be the empirically measured density. In contrast, each of $\boldsymbol{\eta}_1$, i.e. η_i, is a probability of a bin, $P[X_i = 1]$, not a probability density. For η_i, it is assumed that each bin can contain at most only one spike. In practical data analysis, it thus can be regarded as the density, as long as the bin size is small enough. In general, though, we must be aware of this difference and be careful in translating between them. How large to make the bins is an important question but beyond the scope of the present study. Some may concern what if an event is found not to occurr in data, i.e. if zero probability must be assigned to the event. In such a case, we can, in principle, re-construct a probability model from which the events of zero probabilities are omitted, or impose some assumptions on those zero probabilities, which seems more useful in practice (see (Nakahara and Amari, 2002b).)

For presentation simplicity, we sometimes write $X_i = 0$ and $X_i = 1$ as x_i^0 and x_i^1, respectively, e.g., $x^N = (X_1 = 0, X_2 = 0, X_3 = 1 \cdots X_N = 1) = (x_1^0, x_2^0, x_3^1 v, \cdots x_N^1)$. The notation of $p_{(ijk)}$ etc is used to define

$$p_{(i_1 i_2 \ldots i_k)} = P\left[x_1^0, \cdots, x_{i_1}^1, \cdots x_{i_2}^1, \cdots x_{i_k}^1 \cdots x_N^0\right].$$

We also use $p_{(0)} = P\left[x_1^0, \cdots\cdots x_N^0\right]$. The cardinality of a spike train, i.e.

the number of spikes or the spike count, is

$$Y \equiv \left| X^N \right| = \#\left\{ X_i = 1 \right\}.$$

Given a specific spike train x^N, n is reserved to indicate $n = \left| x^N \right|$ and

$$s(1), s(2), \cdots, s(n)$$

are used to denote specific timings of spikes that are the set of indices having $x_i = 1$. For example, with this notation, we write

$$x^N = (x_1^0, \cdots x_{s(1)-1}^0, x_{s(1)}^1, x_{s(1)+1}^0 \cdots x_{s(2)}^1, \cdots x_{s(3)}^1, x_{s(3)+1}^0, \cdots x_{s(n)}^1, x_{s(n)+1}^0, ... x_N^0).$$

3 Two Parametric Models for Single Spike Train

Here we present the original formulations of the inhomogeneous Markov (IM) (Kass and Ventura, 2001) and the mixture of Poisson (MP) models (Wiener and Richmond, 2003) .

3.1 Inhomogeneous Markov model

The IM model was developed as a tractable class of spike interval probability distributions to account for the observation that the spike firing over bins is not completely independent and thus departs from the Poisson process (Kass and Ventura, 2001; Ventura et al., 2002).

The inhomogeneous Markov assumption is the key of the IM model, assuming the following equality: for any spike event, $x_{s(l)}^1$ $(l \leq n)$,

$$P\left[x_{s(l)}^1 \middle| x_1, \cdots x_{s(l)-1} \right] = P\left[x_{s(l)}^1 \middle| x_{s(l-1)}^1, x_{s(l-1)+1}^0, \cdots x_{s(l)-1}^0 \right].$$

The probability of firing at a time given its past, which is the left side of the equation, depends only on the time of the last spike, and the time since it occurred. Denote the right hand side by $\tilde{K}_{s(l-1),s(l)}$, as

$$\tilde{K}_{s(l-1),s(l)} = P\left[x^1_{s(l)} \middle| x^1_{s(l-1)}, x^0_{s(l-1)+1}, \cdots x^0_{s(l)-1}\right],$$

where $l = 2, ..., n$ (potentially n is up to N).

If $\tilde{K}_{s(l-1),s(l)} = \eta_{s(l)} = P\left[x^1_{s(l)}\right]$, this is an inhomogeneous Poisson process. By explicitly including the parameter, $\{\tilde{K}_{s(l-1),s(l)}\}$, the IM model enlarges the class of probabilities beyond the Poisson process, while reducing the space in which it lies to be smaller than a "full" probability space, $\boldsymbol{S}_N$. The original parameters of the IM model are given by $\{\eta_i, \tilde{K}_{i,j}\}$ $(i, j = 1, , , , N,\ i < j)$. Once we define another quantity, $K_{s(l-1),s(l)}$, such that

$$\tilde{K}_{s(l-1),s(l)} = \eta_{s(l)} K_{s(l-1),s(l)}$$

(provided $\eta_i > 0$), they are given by

$$\{\eta_i, K_{i,j}\} \ \ (i, j = 1, , , , N,\ i < j).$$

After some calculations, we obtain the following:

Proposition 1

Given the original parameters $\{\eta_i, \tilde{K}_{i,j}\}$ $(i, j = 1, , , , N,\ i < j)$, probability of any spike train under the IM model is given by

$$\begin{aligned} &P_{IM}(\boldsymbol{x}^N) \\ &= \left\{ \prod_{l=1}^{s(1)-1} (1 - \eta_l) \right\} \eta_{s(1)} \left\{ \prod_{l=2}^{n} \tilde{K}_{s(l-1),s(l)} \right\} \prod_{l=1}^{n} \prod_{k1}^{s(l+1)-s(l)-1} (1 - \tilde{K}_{s(l),s(l)+k}), \end{aligned} \tag{1}$$

where and hereafter P_{IM} is used to denote a probability distribution of the IM model and the convention $s(n+1) = N+1$ is introduced without loss of generality. (See Appendix A for the proof.)

Equation (1) indicates that the probability of any spike train $p(\boldsymbol{x}^N)$ depends only on η_i and $\tilde{K}_{i,j}$ together under the IM model. Thus, $\{\eta_i, \tilde{K}_{i,j}\}$ $(i,j = 1,,,N,\ i < j)$ is one coordinate system for the IM model and $\{\eta_i, K_{i,j}\}$ is another. The number of parameters, or the dimensionality of the IM model, is $N + \frac{N(N-1)}{2}$.

A subclass of IM model, called multiplicative inhomogeneous Markov (MIM) model, was also proposed (Kass and Ventura, 2001). In addition to the IM assumption, they assumed another constraint on $K_{i,j}$, given by

$$a_{j-i} \equiv K_{i,j}. \tag{2}$$

This assumption is not equivalent to assuming $\tilde{K}_{i,j}\tilde{K}_{i',j'}$, where $j-i = j'-i'$. The assumption further constrains the probability space to search (Fig 2.) The dimensionality of the MIM model is $2N-1$. The MIM model is easily expressed by substituting $\tilde{K}_{i,j} = \eta_j a_{j-i}$ in Eq 1.

3.2 Mixture of Poisson model

The MP model is primarily motivated by the desire to make the model more account for the spike count statistics over the interval of interest(Wiener and Richmond, 2003). In many neurophysiological recordings of single neurons, the spike count data, or the probability distribution of spike counts, is easy to obtain, and arguably the most robust measure to estimate. It is also of

interest for biophysical reasons. The MP model *does not* refer to a mixture of Poisson distribution but refers to a specific spike descriptive model, as shown below.

The MP model begins by fitting the spike count distribution using one or more Poisson distributions, i.e., a mixture of Poisson distributions. The mixture of Poisson distribution is a larger model of the spike count distribution than the Poisson distribution, as desired. The mixture of Poisson distribution itself cannot determine a spike train generation without further assumptions. In the original work (Wiener and Richmond, 2003) each trial was drawn from one of the Poisson distributions; each k-th component of the Poisson process is chosen with a probability π_k in each trial of experiment, generating a spike train of the trial. Thus the MP model enjoys the simplicity of the Poisson process for generating spike trains in each trial.

Let us write each k-th (inhomogeneous) Poisson process, P_k,

$$P_k\left[X^N = \boldsymbol{x}^N\right] = \prod_i^N \eta_{i,k}^{x_i} \left(1 - \eta_{i,k}\right)^{1-x_i}$$

where we define $\eta_{i,k} = E_k\left[x_i\right]$, and E_k denotes the expectation with respect to the probability P_k. The trial-by-trial mixture of the Poisson processes is given by

$$P\left[X^N = \boldsymbol{x}^N\right] = \sum_k^K \pi_k P_k\left[X^N = \boldsymbol{x}^N\right],$$

where $\{\pi_k\}$ are mixing probabilities with $\sum_{k=1}^{K} \pi_k = 1$. The corresponding spike count distribution is the mixture of Poisson distribution, given by $P\left[Y = y\right] = \sum_k^K \pi_k P_k\left[Y = y\right]$. Here the Poisson distribution of each k-th

component is given by

$$P_k\left[Y=y\right]=\frac{y^{\lambda_k}}{y!}e^{-\lambda_k},\quad \text{where we have}\quad \lambda_k=\sum_i^N \eta_{i,k}.$$

An important issue is how we estimate a spike generation of each k-th component, i.e. $\{\eta_{i,k}\}$ $(i=1,...,N)$. Consider first a single Poisson process, for which we pretend to have only a single component in the above formulation. Can we recover $\{\eta_{i,1}\}$ from λ_1? We can get $\eta=\lambda_1/N$, if the process is homogeneous. If inhomogeneous, the solution is not unique: various sets of $\{\eta_{i,1}\}$ may match a value of λ_1. In practice we get $\{\eta_{i,1}\}$ by looking at the PSTH from the same experimental data. If there is more than one component, the PSTH tells us only the left hand side of the equation below

$$\eta_i=E\left[X_i\right]=\sum_k^K \pi_k\eta_{i,k}\ (i=1,...N).$$

In this general case, to obtain $\{\eta_{i,k}\}$ of each k-th component, the approach taken by the MP model is to assume that the overall shape of PSTH is the same among all components (Wiener and Richmond, 2003). This assumption implies that there exists, for each k, a constant α_k such that $\eta_{i,k}=\alpha_k\eta_i$ for any $i=1,\cdots,N$. By taking the sum with respect to i, the value of α_k is given by

$$\alpha_k=\frac{\lambda_k}{c_1},\quad \text{where we defined}\quad c_1\equiv\sum_i^N\eta_i.$$

The MP model, as a generative model of spike trains, is the trial-by-trial mixture of Poisson process with this assumption. We summarize as follows.

Proposition 2

Given the original parameters $\{\pi_k, \lambda_k, \eta_i\}$ $(k = 1, ..., K;\ \ i = 1, ..., N)$, the probability distribution of any spike pattern of the MP model is given by

$$P_{MP}(\boldsymbol{x}^N) \quad = \quad \sum_k^K \pi_k P_k(\boldsymbol{x}^N) \tag{3}$$

where and hereafter P_{MP} denotes the probability distribution of the MP model and P_k denotes the probability distribution of the k-th component,

$$P_k(\boldsymbol{x}^N) = \prod_i^N \eta_{i,k}^{x_i} \left(1 - \eta_{i,k}\right)^{1-x_i}, \tag{4}$$

where $\eta_{i,k}$ is defined by

$$\eta_{i,k} \quad = \quad \frac{\lambda_k}{c_1} \eta_i \left(i = 1, \cdots, N\right). \tag{5}$$

Here, c_1 is a constraint of the model parameters, given by

$$c_1 \quad = \quad \sum_i^N \eta_i = \sum_i^N \sum_k^K \pi_k \eta_{i,k} = \sum_k^K \pi_k \lambda_k. \tag{6}$$

Another constraint of the model parameter is $\sum_k^K \pi_k = 1$.

Thus, the dimensionality of the MP model is $2K + N - 2$. There is one issue worth mentioning on the constraint on c_1 (Eq 6). c_1 is intrinsic, or a part of, the MP model. It is then natural to estimate all of the parameters together $\{\pi_k, \lambda_k, \eta_i\}$, for example, using a maximum likelihood estimation on both empirical distributions (the PSTH and spike counts) together with the two constraints above. Or, if the estimation with each distribution is done separately first, we then need to project this estimated 'point' onto the space restricted by Eq 6, i.e. further re-estimate the parameter values with

this constraint. The procedure in the original work (Wiener and Richmond, 2003), where the estimation is done separately, seems sufficient in practice but a mathematical rigorous examination remains for future work. At least, examination of the constraint on c_1 would work as a sanity check.

4 Representation of Two Models by Information Geometric Measure

4.1 Information Geometric Measure

Having established the two models, we re-write them in information geometric (IG) measure (Amari, 2001; Nakahara and Amari, 2002b). Here we first give a brief description of the IG measure. The usefulness of IG measure was studied earlier for spike data analysis (Nakahara and Amari, 2002a; Nakahara et al., 2002; Nakahara and Amari, 2002b; Amari et al., 2003) and for DNA microarray data (Nakahara et al., 2003). Although these studies emphasized neural population firing and interactions among neurons (or gene expressions), almost all the earlier results can be directly applied in analyzing single neuron spike trains because the mathematical formulation is general in the sense that it can be applied to any binary random vector. We mention that the IG measure has some broad roots for data analysis in a log-linear model (Bishop et al., 1975; Whittaker, 1990).

Let us first introduce the θ-coordinate system, defined by,

$$\log P\left[X^N = \boldsymbol{x}^N\right] = \sum \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j + \sum_{i<j<k} \theta_{ijk} x_i x_j x_k \cdots + \theta_{1\cdots N} x_1 \cdots x_N - \psi,$$

where the indices of θ_{ijk}, etc., satisfy $i < j < k$ and ψ is a normalization term, corresponding to $-\log p(x_1 = x_2 = ... = x_N = 0)$. This log expansion is not an approximation, but is exact. All θ_{ijk}, etc., together have $2^N - 1$ components, that is,

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N) = (\theta_i, \theta_{ij}, \theta_{ijk}, \cdots, \theta_{12...N}),$$

and forms the θ-coordinate system in $\boldsymbol{S}_N$, also called θ-coordinates. This θ-coordinates can represent any probability distribution in $\boldsymbol{S}_N$. It is straightforward to write down any components of the θ-coordinates in relation to the P-coordinate system. Here, we list a few first terms,

$$\psi = -\log p_{(0)}, \theta_i = \log \frac{p_{(i)}}{p_{(0)}}, \theta_{ij} = \log \frac{p_{(ij)}p_{(0)}}{p_{(i)}p_{(j)}}, \theta_{ijk} = \log \frac{p_{(ijk)}p_{(i)}p_{(j)}p_{(k)}}{p_{(0)}p_{(ij)}p_{(jk)}p_{(ik)}}.$$

For the later use, let us explicitly write the components of both η- and θ-coordinates with respect to P-coordinates.

Theorem 3

$$\theta_{i_1 i_2 ... i_l} = \sum_{m=0}^{l} \sum_{A \in \Omega_m(X^{l*})} \log p_{(A)}^{(-1)^{l-m}} \tag{7}$$

$$\eta_{i_1 i_2 ... i_l} = \sum_{m=0}^{N-l} \sum_{A \in \Omega_m(\bar{X}^{l*})} p_{(\{i_1, i_2, ... i_l\} \cup A)}, \tag{8}$$

where some conventional notation is introduced; X^{l*} indicates the specific l-tuple among X^N, namely $\{X_{i_1}, X_{i_2}, ...X_{i_l}\}$. $\Omega_m(X^{l*})$ indicates the set of all possible m-tuple of X^{l*}. In the first equation, given $A \in \Omega_m(X^{l*})$, A indicates each element of $\Omega_m(X^{l*})$ and the summation is taken over all the elements of $\Omega_m(X^{l*})$. The same convention applies to the second equation except that the summation is taken over $\Omega_m(\bar{X}^{l*})$ and $\bar{X}^{l*}$ is defined such

that $X^N = X^{l*} \cup \bar{X}^{l*}$ and $X^{l*} \cap \bar{X}^{l*} = \phi$. Given $A\{j_1, ..., j_m\}$, $p_{(A)}$ is used as the same notation as $p_{(j_1 \ldots j_m)}$ in the first equation and $p_{(\{i_1, i_2, \ldots i_l\} \cup A)}$ as $p_{(i_1 i_2 \ldots i_l j_1 \ldots j_m)}$ in the second equation.

The proof can be easily derived by using Rota's method, i.e. the principle of inclusion-exclusion, and is omitted. As shown previously (Amari, 2001; Nakahara and Amari, 2002b), the IG measure, the η- and θ-coordinates together, effectively uses the dually flat structure of the η-coordinates and θ-coordinates in $\boldsymbol{S}_N$, which is the property of being e-flat and m-flat in a more general term (Amari and Nagaoka, 2000). The notion of e-flat and m-flat is proved to underlie various useful properties of probability distributions, especially the exponential family of probability distribution. The IG coordinates allow examination of the different order interactions of neural spike firing and then allows calculation the information conveyed by these different orders (Nakahara and Amari, 2002b). In particular, the mixed coordinates of different orders, constructed from η-coordinates and θ-coordinates, is useful for examining different order interactions (Amari, 2001; Nakahara and Amari, 2002b). This occurs because under the mixed coordinates, the components of η-coordinates (or θ-ones) can be treated independently from those of the other, i.e. θ-coordinates (or η-ones), due to the dual orthogonality of η- and θ-coordinates. This makes comparison of each order interaction between any probability distributions transparent.

Among the coordinates, the 1st order (or 1st-cut) mixed coordinate $(\boldsymbol{\eta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N)$ is useful for dissociating the mean firing rate $\boldsymbol{\eta}_1$ from all the

second and higher-order interactions under null hypothesis of any correlated firing because all the components of $\boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N$ can be treated independently from $\boldsymbol{\eta}_1$. This 1st-cut mixed coordinates is most suitable for comparing the IM and MP models. The original parameters of the IM model are $\left\{\eta_i, \tilde{K}_{i,j}\right\}$ and those of the MP model are $\{\pi_k, \lambda_k, \eta_i\}$. These parameters are regarded as observable and thereby used in fitting data (Fig 2). Since both models share the parameters $\boldsymbol{\eta}_1$, the difference between them lies with other components in the full probability space. Then, the 1-cut mixed coordinates can characterize the difference, discussing the interaction terms $\boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N$, separately from $\boldsymbol{\eta}_1$.

We now make two remarks useful in the following sections. First concerns the notion of restricted probability space. Recall that the Poisson process is the intersection of the IM and MP models, which is given by $p(\boldsymbol{x}^N) = \prod_{i=1}^{N} \eta_i^{x_i} (1-\eta_i)^{1-x_i}$. This model corresponds to

$$\log p(\boldsymbol{x}^N) = \sum_i \theta_i x_i - \psi, \tag{9}$$

where $\theta_i = \log \frac{\eta_i}{1-\eta_i}$. Therefore, we have $(\theta_i) \Leftrightarrow (\eta_i)$. In other words, the Poisson process lies in a restricted subspace smaller than the full space, characterized by setting $(\boldsymbol{\theta}_2, ...\boldsymbol{\theta}_N) = (0, ..., 0)$. Under the maximum entropy principle, the Poisson process is the distribution that can be determined completely by the first-order statistics $\boldsymbol{\eta}_1$. The IM and MP models expand this restricted space differently. Consider the following,

$$\log p\left(\mathbf{x}^N\right) = \sum \theta_i x_i + \sum \theta_{ij} x_{(ij)} - \psi,$$

where $\theta_i = \log \frac{p_{(i)}}{p_{(0)}}$ and $\theta_{ij} = \log \frac{p_{(ij)} p_{(0)}}{p_{(i)} p_{(j)}}$. Now, $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ becomes the coor-

dinate system and $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ is another coordinate system, i.e. η-coordinates. The number of parameters is $N + \frac{N(N-1)}{2}$. This model has been extensively studied in various fields, e.g. the spin glasses and/or Boltzmann machine. This model is characterized by setting $(\boldsymbol{\theta}_3, ..., \boldsymbol{\theta}_N) = (0, ..., 0)$ and thus, the model that can be determined completely by the first and second-order statistics $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ under the maximum entropy principle. We call this model the 2nd-order IG model.

Second concerns the different representations of the η- and θ-coordinates. Their dual orthogonality is broader than the specific representations of the two coordinates described above. Specifically, there can be different representations of the two coordinates, due to the degrees of freedom in affine coordinate system. This notion will be easily understood in the following example. Hereafter, when the distinction is needed, we refer to the original dual coordinates, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N)$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_N)$ as the 'standard' IG, or θ- and η-, coordinates.

Example: Consider a simple example, $X^2 = (X_1, X_2)$. We have

$$\log p(\boldsymbol{x}^2) = \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_{(12)} - \psi.$$

The standard θ- and η-coordinates are given by $\theta_1 = \log \frac{p_{(1)}}{p_{(0)}}$, $\theta_2 = \log \frac{p_{(2)}}{p_{(0)}}$ and $\theta_{12} = \log \frac{p_{(12)} p_{(0)}}{p_{(1)} p_{(2)}}$ and by $\eta_1 = E[X_1], \eta_2 = E[X_2]$, and $\eta_{12} = E[X_1 X_2]$, respectively. Now, for example, consider:

$$\log p(\boldsymbol{x}^2) = \theta_1' x_1 + \theta_2' x_2 + \theta_{12}' (1 - x_1) x_2 - \psi'.$$

Here $(\theta_1', \theta_2', \theta_{12}')$ is a new θ-coordinates of X^2. The new η-coordinates is given by $\eta_1' = E[X_1] = \eta_1$, $\eta_2' = E[X_2] = \eta_2$, and $\eta_{12}' = E[(1 - X_1)(X_2)] =$

$\eta_2 - \eta_{12}$. Each of the two new coordinates is linearly related to each standard coordinates and thus, the new coordinates systems are dually orthogonal to each other, forming a new set of η- and θ-coordinates.

4.2 IM model

We aim to represent the IM model by standard θ-coordinates. We introduce another representation of θ- and η-coordinates for convenience. Let us define, for $i < j$,

$$\tilde{X}_{i,j} \equiv X_i X_j \prod_{l=i+1}^{j-1} (1 - X_l). \tag{10}$$

$\tilde{X}_{i,j}$ becomes 1 only when there are spikes at the i-th and j-th bins and no spikes between them. This is a natural quantity to be dealt with the IM model. We have

$$\tilde{\eta}_{i,j} \equiv E\left[\tilde{X}_{i,j}\right] = \eta_i \tilde{K}_{i,j} \prod_{l=i}^{j-i-1} \left(1 - \tilde{K}_{i,i+l}\right). \tag{11}$$

Since $\left\{\eta_i, \tilde{K}_{i,j}\right\}$ is a coordinate system of the IM model, Eq 11 implies that $\{\eta_i, \tilde{\eta}_{i,j}\}$ is also another coordinate system, which is another representation (but not the standard one) of η-coordinates of the IM model. In correspondence to $\{\tilde{\eta}_{i,j}\}$, we introduce $\{\tilde{\theta}_{i,j}\}$ by

$$\log P_{IM}(\boldsymbol{x}^N) \equiv \sum \theta_i X_i + \sum_{i<j} \tilde{\theta}_{i,j} \tilde{X}_{i,j} - \psi. \tag{12}$$

We see that $\{\theta_i, \tilde{\theta}_{i,j}\}$ is another representation of θ-coordinates of the IM model, corresponding to the η-coordinates, $\{\eta_i, \tilde{\eta}_{i,j}\}$. We have

$$\theta_i = \log \frac{p_{(i)}}{p_{(0)}}, \ \tilde{\theta}_{i,j} = \log \frac{p_{(ij)} p_{(0)}}{p_{(i)} p_{(j)}}. \tag{13}$$

The dimensionality of the coordinates is $N + \frac{N(N-1)}{2}$ and is equal to that of the 2nd-order model of the standard coordinates. However, the restricted probability spaces of IM and 2nd-order models are different.

To see how the IM model is embedded in the full space, let us write the IM model in the standard θ-coordinates. By expanding Eq 12 with the definition of $\tilde{X}_{i,j}$, we get $\theta_{ij} = \tilde{\theta}_{i,j}, \theta_{ijk} = -\tilde{\theta}_{i,k}, \theta_{ijkl} = \tilde{\theta}_{i,l}$, or in general,

$$\theta_{i_1 i_2 \cdots i_k} = (\ -1)^k \, \tilde{\theta}_{i_1, i_k} \ \ (k \geq 2). \tag{14}$$

This equation indicates that the restricted probability space of the IM model imposes a specific alternating structure in the higher-order interaction term. It also indicates that the IM model is not distinguishable from the 2nd-order model of the standard θ-coordinates in terms of the second-order interaction, since we have $\theta_{ij} = \tilde{\theta}_{i,j}$. The relation of $\{\theta_i, \tilde{\theta}_{i,j}\}$ with the original parameters $\{\eta_i, K_{i,j}\}$ is given as follows.

Theorem 4

The IM model is represented by the standard θ-coordinates in relation to its original parameters $\{\eta_i, K_{i,j}\}$ $(i < j, \ \ ; i, j = 1, ..., N)$ as

$$\theta_i = \log \frac{\eta_i}{1-\eta_i} + \sum_{l=i+1}^{N} \log \left(1 + \eta_l \frac{1 - K_{i,l}}{1 - \eta_l}\right) \tag{15}$$

$$\theta_{i_1 i_2 \cdots i_k} = (\ -1)^k \, \tilde{\theta}_{i_1, i_k} \ \ (k \geq 2) \tag{16}$$

$$\tilde{\theta}_{i,j} = \log K_{i,j} - \sum_{l=j}^{N} \log \left(1 + \eta_l \frac{1 - K_{i,l}}{1 - \eta_l}\right). \tag{17}$$

See Appendix B. First, note that the 1st-order component θ_i deviates from that of the Poisson process (which would have the form of $\log \frac{\eta_i}{1-\eta_i}$ in

the right hand side). Second, $\tilde{\theta}_{i,j}$ becomes zero if any $K_{i,j}$ is equal to one, which corresponds to the case that the IM model reduces to the Poisson process. Thus, in the above of $\tilde{\theta}_{i,j}$, the discrepancy of the IM model from the Poisson process is indicated by the first term $\log K_{i,j}$ (in reference to $\log 10$) and also $1 - K_{i,l}$ $(l = j, ..., N)$ in the second term. We see that $\tilde{\theta}_{i,j}$ depends not only on $K_{i,j}$ and η_j but also on $K_{i,l}$ and η_l for $l = j + 1, ..., N$.

We now approximate $\tilde{\theta}_{i,j}$ to grasp its nature. First, $\eta_i \ll 1$ holds in most data. Second, the original work (Kass and Ventura, 2001) suggests that $K_{i,j}$ is roughly within a range of $[0.4, 1.6]$, implying that the probability of a spike occurring is not strongly dependent on the time of the previous spike (refer to their paper for a question of relation to the refractory period). In any case, since this estimation is done only with one type of data, further examination is required to take it as a general phenomenon. Yet, it seems unlikely that $K_{i,j}$ takes a different order of magnitude. Then, let us assume

$$\eta_l \frac{1 - K_{i,l}}{1 - \eta_l} \ll 1, \tag{18}$$

and also notice that in many situations, we have $\eta_l \ll K_{i,j}$. In such a case, we approximately have

$$\tilde{\theta}_{i,j} \simeq \log K_{i,j} - \sum_{l=j}^{N} \eta_l \frac{1 - K_{i,l}}{1 - \eta_l}. \tag{19}$$

We observe that $K_{i,j}$ is the dominant term in $\tilde{\theta}_{i,j}$ and that at the same time, the terms $(1 - K_{i,l})$ $(l = j + 1, ...N)$ also contribute to $\tilde{\theta}_{i,j}$. As far as the order of η_l does not differ between each other, for further simplification, let

us put $\bar{\eta}_j = \frac{1}{N-j+1}\sum_{l=j}^{N}\eta_l$, and then we have

$$\tilde{\theta}_{i,j} \simeq \log K_{i,j} - \frac{\bar{\eta}_j}{1-\bar{\eta}_j}\sum_{l=j}^{N}(1-K_{i,l}). \tag{20}$$

For the multiplicative IM (MIM) model, we can replace $K_{i,j}$ by a_{j-i} in Theorem 4 to obtain the exact expression. In case of using the above approximation, we get

$$\tilde{\theta}_{i,j} \simeq \log a_{j-i} - \frac{\bar{\eta}_j}{1-\bar{\eta}_j}\sum_{l=j}^{N}(1-a_{l-i}). \tag{21}$$

Note that even with the MIM model, $\tilde{\theta}_{i,j}$ cannot be represented by the terms that only use a_{j-i}. In other words, even for a fixed $k = j - i$, $\tilde{\theta}_{i,j}$ takes different values and the range of its values is determined by the summation in the second term (see Example below). The summation goes from a_{j-i} to a_{N-i}. Thus, N, the number of bins, affects the value of $\tilde{\theta}_{i,j}$. This is because $\tilde{\theta}_{i,j}$ is defined with respect to a given period, whereas a_k is not. In this sense, the second term reflects a boundary effect.

Example In Fig 3 (A-C), we show an example of MIM model and manipulated only $\{\eta_i\}$ among the original parameters from A to C. The overall shape of $\tilde{\theta}_{i,j}$ follows the shape of a_{j-i} $(= K_{i,j})$, while $\tilde{\theta}_{i,j}$ varies even with a fixed k. The range obtained by the above approximation is indicated by thick lines. We note that in this example of a_{j-i} (which very roughly imitates the curve estimated in the original work), the minimum is almost always equal to set $i = 1$ for each $k = j - i$. Thus, the lower thick line is almost equal to $\log a_k$.

4.3 MP model

Let us first write the MP model in terms of the log expansion, based on its original definition (Proposition 2) as

$$\log P_{MP}(\boldsymbol{x}^N) \;\; = \;\; \log \sum_k^K \pi_k P_k(\boldsymbol{x}^N). \tag{22}$$

To represent the MP model in the standard θ-coordinates using relation to the original parameters, i.e. $\{\pi_k, \lambda_k, \eta_i\}$ $(k = 1, \cdots, K\ , i = 1, \cdots, N)$, note that the probabilities of the MP model, denoted by p^{MP}, are represented by the original parameters (Proposition 2),

$$p^{MP}_{(i_1 \cdots i_l)} = \sum_k^K \pi_k p^k_{(i_1 \cdots i_l)}, \tag{23}$$

where p^k, for each k-th component inhomogeneous Poisson process, is given by

$$p^k_{(0)} = \prod_{j=1}^N (1 - \eta_{j,k}), \;\; p^k_{(i)} = p^k_{(0)} \frac{\eta_{i,k}}{1 - \eta_{i,k}}, \;\; \cdots, p^k_{(i_1 \cdots i_l)} = p^k_{(0)} \prod_{j=1}^l \frac{\eta_{i_j,k}}{1 - \eta_{i_j,k}}.$$

Also recall that we have $\eta_{i,k} = \frac{\lambda_k}{c_1}\eta_i$ where $c_1 = \sum_i^N \eta_i$. Then we have

Theorem 5

The probabilities of the MP model are given by

$$p^{MP}_{(i_1 \cdots i_l)} = \frac{\prod_{j=1}^l \eta_{i_j}}{c_1^l} \sum_k^K \pi_k \lambda_k^l p^k_{(0)} \prod_{j=1}^l \left(1 - \frac{\lambda_k}{c_1}\eta_{i_j}\right)^{-1}, \tag{24}$$

where p^k indicates the probabilities due to the k-th component and $p^k_{(0)}$ is given by $p^k_{(0)} = \prod_{j=1}^N (1 - \eta_{j,k}) = \prod_{j=1}^N \left(1 - \frac{\lambda_k}{c_1}\eta_i\right)$.

By using Theorems 3 and 5 together, any component of the standard θ-coordinates of the MP model is represented by the original parameters of the MP model. For example, the first and second order components of the standard θ-coordinates is given by

$$\theta_i = \log \frac{\eta_i}{1-\eta_i} + \log \frac{\sum_k^K \pi_k p_{(0)}^k \left(1 - \frac{c_1 - \lambda_k}{c_1 - \lambda_k \eta_i}\right)}{\sum_k^K \pi_k p_{(0)}^k} \tag{25}$$

and

$$\theta_{ij} = \log \frac{\left(\sum_k^K \pi_k p_{(0)}^k\right)\left(\sum_k^K \pi_k p_{(0)}^k \frac{1}{1-\frac{\lambda_k}{c_1}\eta_i} \frac{1}{1-\frac{\lambda_k}{c_1}\eta_j}\right)}{\left(\sum_k^K \pi_k p_{(0)}^k \frac{1}{1-\frac{\lambda_k}{c_1}\eta_i}\right)\left(\sum_k^K \pi_k p_{(0)}^k \frac{1}{1-\frac{\lambda_k}{c_1}\eta_j}\right)}. \tag{26}$$

The higher-order interaction terms, such as $\theta_{ijk}, \theta_{ijkl}$ (and so on), can be derived in a similar manner. In the above equations, first, note that the 1st order term θ_i indicates that the MP model deviates somewhat from a Poisson process. Second, the MP model induces the 2nd-order interactions as evident in the above expression. More generally, the MP model induces the higher-order interaction, despite the fact that each component of the MP model (i.e. each Poisson process) itself does not have any interaction term. This is because the summation terms in the p^{MP} appears as a ratio (and is not cancelled out) in computing the θ-coordinates Third, the term $\theta_{i_1,...,i_l}$ of the MP model is permutation-free over its given indices, $\{i_1, ..., i_l\}$, or equivalently that the value of $\theta_{i_1,...,i_l}$ depend upon the choice of the indices only through the magnitudes of $\{\eta_{i_1}, ..., \eta_{i_l}\}$. Mathematically, this is clear because the term $\theta_{i_1,...,i_l}$ depends on the bin indices only through the term $\prod_{j=1}^{l} \left(1 - \frac{\lambda_k}{c_1}\eta_{i_j}\right)^{-1}$. With this property, for example, $\theta_{ij} = \theta_{i'j'}$, if $(\eta_i, \eta_j) =$

$(\eta_{i'}, \eta_{j'})$ or $= (\eta_{j'}, \eta_{i'})$. Such a property does not exist in the IM model.

Finally, we observe that the MP model tends to produce components of comparable magnitude in each order interaction when the order (i.e. l) is not too high. To see this, first note that it is reasonable (in most cases at least) to assume $1 \gg \eta_{i,k} = \frac{\lambda_k}{c_1}\eta_i$ for any i, k. Then we have approximations, $p^k_{(0)} \simeq \exp\left(-\frac{\lambda_k}{c_1}\sum_i \eta_i\right) = e^{-\lambda_k}$ and $\prod_{j=1}^{l}\left(1 - \frac{\lambda_k}{c_1}\eta_{i_j}\right)^{-1} \simeq 1 + \frac{\lambda_k}{c_1}\sum_{j=1}^{l}\eta_{i_j}$. With these approximations, the summation term of p^{MP} is

$$\sum_k^K \pi_k \lambda_k^l p^k_{(0)} \prod_{j=1}^{l}\left(1 - \frac{\lambda_k}{c_1}\eta_{i_j}\right)^{-1} \simeq \sum_k^K \pi_k \lambda_k^l e^{-\lambda_k}\left(1 + \frac{\lambda_k}{c_1}\sum_{j=1}^{l}\eta_{i_j}\right). \quad (27)$$

The l-th order term $\theta_{i_1,\ldots,i_l}$ is represented by a ratio of these summation terms and in each summation term, the bin indices appear only as $\sum_{j=1}^{l}\eta_{i_j}$, as evident in the above approximation. Because $1 \gg \eta_i$ can usually be assumed and the difference among η_i is of secondary order, the magnitude of the same order interaction is expected to be similar, as far as the order l is not so large. This implies that in such a case, parameters of $\{\pi_k, \lambda_k\}$ becomes dominant factors in the interaction terms.

Example; We illustrate our observations on the MP model, using some examples (Fig 4). In Fig 4 A is shown the PSTH and spike count distribution of the MP model and the corresponding θ-coordinates (up to the 6-th order; i.e., from θ_i to $\theta_{i_1,\ldots,i_6}$). Since the PSTH is uniform, there is a single value at each order. As mentioned above, we see the 2nd-order and higher-order interaction exist. At the same time, their magnitudes are relatively small (e.g. compare with those shown for the MIM model in Fig 3.) In Fig 4 B-D, only a part of the original parameters ($\{\eta_i, \pi_k, \lambda_k\}$) is changed from those

of Fig 4 A; namely $\{\eta_i\}$, $\{\lambda_k\}$ and $\{\pi_k\}$ are changed in Fig 4 B, C and D, respectively (see figure legend for details). A variation of $\{\eta_i\}$ induces only a small variation in θ-coordinates, i.e, comparable magnitude in each order (Fig 4 B). Compared with the change in $\{\eta_i\}$, the change in $\{\lambda_k\}$ or $\{\pi_k\}$ induces a relatively larger interaction term (Fig 4 C, D.)

5 Discussion

Different generative point process models are expected to give results with different high order statistical moments. Because the information geometric (IG) measure provides a complete space for representing point processes, any point process model can be mapped into this space, allowing the subspaces occupied by different models to be located, described and compared. The two models examined in this manuscript, the inhomogeneous Markov (IM) and mixture of Poisson (MP) models, were constructed to account for the interval distribution and count distribution, respectively. In their native forms it is difficult to quantify the ways in which the high order moments differ.

However, the differences between the IM and MP models can be seen well by using the first cut mixed coordinates, $(\boldsymbol{\eta}_1; \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N)$, and inspecting its 2nd and higher components, $\boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N$. The two models differ only in the 2nd and higher-order terms. For θ_{ij} (or equivalently $\tilde{\theta}_{i,j}$) the 3rd and higher-order interaction terms have a structure of alternating signs in the IM model. The components are permutation-free, and of comparable magnitude for the

MP model, at least for the first few components (e.g, Eqns 26, 27), whereas the components represented by successive information geometric terms may vary considerably for the MIM model (cf Eqns 20, 21).

The results from this analysis make it possible to consider what would be required to distinguish between these models for experimental data. Consider the IG 2nd-order model (which is natural as 2nd-order model under maximum entropy principle), using $\log p(\boldsymbol{x}^N) = \sum_i \theta_i x_i + \sum_{i,j} \theta_{ij} x_i x_j - \psi$. In this reduced form, the two models are different in $\{\theta_{ij}\}$. We approximate the MP model, i.e., components of comparable magnitude, by letting $\theta_{ij} = \theta$. The MIM model is similar to letting $\{\theta_{ij}\}$ be a specific sequence (a function of $\{a_k\}$, cf Eq 21.)

Experimental datasets are often (perhaps even always) too small to compare higher-order interactions above about three-way. However, it might require less data to recognize that $\{\theta_{ij}\}$ are approximately equal or that they seem to follow the prediction of the $\{a_k\}$ than to use the models, as originally formulated. The two models might be distinguished by inspecting whether the 2nd and 3rd order interaction terms have alternating signs, i.e., $\theta_{i_1 i_2 \cdots i_k} = (-1)^k \tilde{\theta}_{i_1, i_k}$ $(k \geq 2)$, so the third-order interaction is negative if the second-order interaction is positive, and vice versa. If there is a positive correlation of spikes between two bins (e.g. at i_1 and i_k bins), and a positive correlation among the spikes for three bins (e.g. at i_1, i_k, and i_l bins, where l is one of $\{2, ..., k-1\}$), the data can not have come from a process described by the IM model. If the correlations are of opposite signs, then the data could have come from either model (in the MP models, the third-

order interaction can be positive or negative, regardless of the sign of the second-order interaction) and it would be necessary to examine higher order interactions. There has been to our knowledge no experimental study suggesting alternating sign interactions over the 2nd and 3rd-oder interactions that is predicted by the IM model. Previous studies found only positive (not negative) third-order interactions for single and multiple neuron spike trains at least in a few cases (Abeles et al., 1993; Riehle et al., 1997; Prut et al., 1998; Oram et al., 1999; Lestienne and Tuckwell, 1998; Baker and Lemon, 2000). However, one needs to be cautious because it is possible that the investigators were mostly interested in finding positive interactions, whether second, third or higher-order.

There are two caveats to consider when comparing the two models with data. First, other sources of experimental variability may add difficulties. For example, the latency might vary across trials. This source of variability is not considered in the models but it surely affects the interaction terms. Second, in fitting experimental data, the PSTH ($\{\eta_i\}$) was smoothed in both models. Smoothing reduces the effective dimensionality of the models. The MIM model had additional smoothing to estimate $\{\tilde{\eta}_{i,j}\}$, reducing the model dimensionality even further. Smoothing affects, or blurs in general, all interaction terms and if too severe, will make it difficult to distinguish the two. Either of these circumstances could mask real differences in the data. Despite the caveats, however, the results here suggest that experimental data might distinguish between the models by comparing the two and three way relations among spikes. Finally, our approach is applicable to other

descriptive models of single neuronal spike trains and can be combined with analysis of activity in neural population (Nakahara and Amari, 2002b).

Acknowledgments

HN is supported by Grant-in-Aid on Priority Areas (C) and the Fund from the US-Japan Brain Research Cooperative Program of MEXT, Japan. HN thanks M. Arisaka for technical assistance and O. Hikosaka for support.

References

M. Abeles. 1991. *Corticonics: neural circuits of the cerebral cortex*. Cambridge University Press, Cambridge: UK.

M. Abeles, H. Bergman, E. Margalit, and E. Vaadia. 1993. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J Neurophysiol*, 70:1629-1638.

S. Amari and H. Nagaoka. 2000. *Methods of Information Geometry*. AMS and Oxford University Press.

S. Amari, H. Nakahara, S. Wu, and Y. Sakai. 2003. Synfiring and higher-order interactions in neuron pool. *Neural Computation*, 15:127–142.

S. Amari. 2001. Information geometry on hierarchical decomposition of stochastic interactions. *IEEE Trans on IT*, pages 1701–1711.

S. N. Baker and R. N. Lemon. 2000. Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *Journal of Neurophysiology*, 84:1770–1780.

J. M. Beggs and D. Plenz. 2003. Neuronal avalanches in neocortical circuits. *J Neurosci*, 23(35):11167–77.

W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck, and D. Warland. 1991. Reading a neural code. *Science*, 252(5014):1854–7.

Y. M. M Bishop, S. E. Fienberg, and P. W. Holland. 1975. *Discrete Multivariate Analysis*. MIT Press, Cambridge, USA.

E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank. 2002. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2):325–346.

A. F. Dean. 1981. The variability of discharge of simple cells in the cat striate cortex. *Experimental Brain Research*, 44:437–440.

J. M. Fellous, P. H. Tiesinga, P. J. Thomas, and T. J. Sejnowski. 2004. Dis-

covering spike patterns in neuronal responses. *J Neurosci*, 24(12):2989–3001.

T. J. Gawne and B. J. Richmond. 1993. How independent are the messages carried by adjacent inferior temporal cortical neurons. *Journal of Neuroscience*, 13(7):2758–2771.

E. D. Gershon, M. C. Wiener, P. E. Latham, and B. J. Richmond. 1998. Coding strategies in monkey v1 and inferior temporal cortices. *Journal of Neurophysiology*, 79:1135–1144.

R. E. Kass and V. Ventura. 2001. A spike-train probability model. *Neural Computation*, 13(8):1713–20.

D. Lee, N. L. Port, W. Kruse, and A. P. Georgopoulos. 1998. Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *Journal of Neuroscience*, 18(3):1161–1170.

R. Lestienne and H. C. Tuckwell. 1998. The significance of precisely replicating patterns in mammalian CNS spike trains. *Neuroscience*, 82:315-336.

E. M. Maynard, N. G. Hatsopoulos, C. L. Ojakangas, B. D. Acuna, J. N. Sanes, R. A. Normann, and J. P. Donoghue. 1999. Neuronal interactions

improve cortical population coding of movement direction. *J Neurosci*, 19(18):8083–93.

M. Meister and M.J. Berry. 1999. The neural code of the retina. *Neuron*, 22(3):435–50.

H. Nakahara and S. Amari. 2002a. Information-geometric decomposition in spike analysis. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *NIPS*, volume 14, pages 253–260. MIT Press: Cambridge.

H. Nakahara and S. Amari. 2002b. Information geometric measure for neural spikes. *Neural Computation*, pages 2269–2316.

H. Nakahara, S. Amari, M. Tatsuno, S. Kang, and K. Kobayashi. 2002. Examples of applications of information geometric measure to neural data. RIKEN BSI BSIS Tech Report No02-1 (www.brain.riken.go.jp/labs/mns/nakahara/papers/TR_IGspike.pdf).

H. Nakahara, S. Nishimura, M. Inoue, G. Hori, and S. Amari. 2003. Gene interaction in dna microarray data is decomposed by information geometric measure. *Bioinformatics*, 19:1124–1131.

M. W. Oram, M. C. Wiener, R. Lestienne, and B. J. Richmond. 1999.

Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *Journal of Neurophysiology*, 81(6):3021–33.

Y. Prut, E. Vaadia, H. Bergman, I. Haalman, H. Slovin, and M. Abeles. 1998. Spatiotemporal structure of cortical activity: properties and behavioral relevance. *Journal of Neurophysiology*, 79:2857-2874.

D. S. Reich, F. Mechler, and J. D. Victor. 2001. Temporal coding of contrast in primary visual cortex: when, what, and why. *J Neurophysiol*, 85(3):1039–50.

R. C. Reid, J. D. Victor, and R. M. Shapley. 1992. Broadband temporal stimuli decrease the integration time of neurons in cat striate cortex. *Visual Neuroscience*, 9:39–45.

B. J. Richmond and L. M. Optican. 1987. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. ii. quantification of response waveform. *J Neurophysiol*, 57(1):147–61.

A. Riehle, S. Grun, M. Diesmann, A. Aertsen. 1997. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, 278:1950-1953.

M. N. Shadlen and W. T. Newsome. 1994. Noise, neural codes and cortical organization. *Curr Opin Neurobiol*, 4(4):569–79.

S. Shinomoto, Y. Sakai, and S. Funahashi. 1999. The ornstein-uhlenbeck process does not reproduce spiking statistics of neurons in prefrontal cortex. *Neural Comput*, 11(4):935–51.

W. R. Softky and C. Koch. 1993. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *J Neurosci*, 13(1):334–50.

C. F. Stevens and A. M. Zador. 1998. Input synchrony and the irregular firing of cortical neurons. *Nat Neurosci*, 1(3):210–7.

D. J. Tolhurst, J. A. Movshon, and A. F. Dean. 1983. The statisical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, 23(8):775–785.

V. Ventura, R. Carta, R. E. Kass, S. N. Gettner, and C. R. Olson. 2002. Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics*, 3(1):1–20.

J. D. Victor and K. P. Purpura. 1997. Metric-space analysis of spike trains: theory, algorithms and application. *Network*, 8:127–164.

J. Whittaker. 1990. *Graphical models in applied multivariate statistics*. John wiley & sons, Chichester, England.

M. C. Wiener and B. J. Richmond. 2003. Decoding spike trains instant by instant using order statistics and the mixture-of-poissons model. *J Neurosci*, 23(6):2394–406.

Appendices

Appendix A

The IM assumption is given by, for any spike event, $x^1_{s(l)}$ $(l \leq n)$,

$$P\left[x^1_{s(l)}\middle| x_1, \cdots x_{s(l)-1}\right] = P\left[x^1_{s(l)}\middle| x^1_{s(l-1)}, x^0_{s(l-1)+1}, \cdots x^0_{s(l)-1}\right] = \tilde{K}_{s(l-1),s(l)}$$

where $l = 2, ..., n$ (potentially n is up to N). Under the IM model, the probability of any spike train is written by two quantities, namely η_i and $\tilde{K}_{i,j}$. First note that we have

$$P\left[X^N = x^N\right] = P\left[x^0_1, \cdots x^1_{s(1)}, \cdots x^0_{s(2)-1}\right] P\left[x^1_{s(2)}, \cdots x_N\middle| x^0_1 \cdots x^1_{s(1)}, \cdots x^0_{s(2)-1}\right].$$

Second, under the IM assumption, we can write

$$P\left[x^1_{s(2)}, \cdots x_N\middle| x^0_1 \cdots x^1_{s(1)}, \cdots x^0_{s(2)-1}\right] = \prod_{l=2}^{n} \tilde{q}_{s(l),s(l+1)} \tilde{K}_{s(l-1),s(l)}, \quad (28)$$

where we define

$$\tilde{q}_{s(l-1),s(l)} \equiv P\left[x^0_{s(l-1)+1}, x^0_{s(l-1)+2}, \cdots x^0_{s(l)-1}\middle| x^1_{s(l-1)}\right], \quad (29)$$

for $l = 2, ..., n$, with the convention $s(n+1) = N+1$. To see that $\{\tilde{q}_{i,j}\}$ is determined solely by $\{\tilde{K}_{i,j}\}$, note that we have $\tilde{K}_{i,j} = 1 - \frac{\tilde{q}_{i,j+1}}{\tilde{q}_{i,j}}$. Given this identify, we have $\tilde{q}_{i,j} = \prod_{l2}^{j-i-1}(1 - \tilde{K}_{i,i+l})\tilde{q}_{i,i+2}$ and $\tilde{q}_{i,j+2} = 1 - \tilde{K}_{i,i+1}$. Therefore we obtain

$$\tilde{q}_{i,j} \equiv P\left[x^0_{j-1}, x^0_{j-2}, \cdots x^0_{i+1}\middle| x^1_i\right] = \prod_{l1}^{j-i-1}(1 - \tilde{K}_{i,i+l}). \quad (30)$$

Third, by defining $P_{init} = P\left[x^0_1, \cdots x^1_{s(1)}, \cdots x^0_{s(2)-1}\right]$, we have, under the IM assumption,

$$P_{init} = P\left[x^0_1, \cdots x^1_{s(1)}\right] P\left[x^0_{s(2)-1}, \cdots x^0_{s(1)+1}\middle| x^1_{s(1)}\right] = P\left[x^0_1, \cdots x^1_{s(1)}\right] \tilde{q}_{s(1),s(2)}.$$

In the original work (Kass and Ventura, 2001), there was no explicit mention of how to treat the quantity $P\left[x_1^0, \cdots x_{s(1)}^1\right]$, but, with our understanding of its spirit and our preference of simplicity, we define this quantity as

$$P\left[x_1^0, \cdots x_{s(1)}^1\right] = \left\{\prod_{l=1}^{s(1)-1}(1-\eta_l)\right\}\eta_{s(1)}.$$

Then, we obtain

$$P_{init} = \left\{\prod_{l=1}^{s(1)-1}(1-\eta_l)\right\}\eta_{s(1)}\tilde{q}_{s(1),s(2)}. \tag{31}$$

Taken all together, Proposition 1 is proved.

Appendix B

It suffices to obtain θ_i and $\tilde{\theta}_{i,j}$ in order to get the values of the standard θ-coordinates of the IM model, since we have

$$\theta_{i_1 i_2 \cdots i_k} = (\ -1)^k\, \tilde{\theta}_{i_1, i_k} \ \ (k \geq 2).$$

To compute θ_i and $\tilde{\theta}_{i,j}$, it suffices to obtain the expression of $p_{(0)}, p_{(i)}$ and $p_{(ij)}$, due to Eq 13. Using the notation $\tilde{q}_{i,j}$ (see Appendix A), we have $p_{(0)} = \prod_{i=1}^{N}(1-\eta_i)$ and

$$p_{(i)} = \left\{\prod_{l=1}^{i-1}(1-\eta_l)\right\}\eta_i\tilde{q}_{i,N+1}, \ \ p_{(ij)} = \left\{\prod_{l=1}^{i-1}(1-\eta_l)\right\}\eta_i\tilde{q}_{i,j}\tilde{K}_{i,j}\tilde{q}_{j,N+1}.$$

Therefore, by using the identity $\tilde{q}_{i,j} = \prod_{l=1}^{j-i-1}\left(1-\tilde{K}_{i,i+l}\right)$, we get

$$\theta_i = \log\frac{\eta_i}{1-\eta_i} + \log\prod_{l=i+1}^{N}\frac{1-\tilde{K}_{i,l}}{1-\eta_l}, \ \ \tilde{\theta}_{i,j} = \log\frac{\tilde{K}_{i,j}}{\eta_j} - \log\prod_{l=j}^{N}\frac{1-\tilde{K}_{i,l}}{1-\eta_l}.$$

By using $\tilde{K}_{i,j} = \eta_j K_{i,j}$, we get the expression in the theorem.

Figure legends

Figure 1: Schematic drawing to show that each parametric model for spike train description is embedded as a subspace in a full probability space. Given a parametric model, estimating the parameter values (or the probability distribution) from experimental data corresponds to identifying a 'point' in the subspace.

Figure 2: Schematic drawing to indicate how raw experimental data is converted to estimation of parameter values of the two different models. The raw data, or raster data (top), can be converted to different formats of data, namely $K_{i,j}$, PSTH, and spike count histogram (middle). For the MIM model, $K_{i,j} = a_{j-i}$ and PSTH data is used, whereas for the MP model, PSTH and spike count histogram is used. The case of the MIM model, not of the IM model is shown here only for presentation simplicity.

Figure 3: Example for illustration of MIM model. In each row (A-C), shown is the relationship of the original parameters ($\{\eta_i\}$ i.e. PSTH, and $K_{i,j} = a_k$, where $k = j-i$, in the case of MIM model) in two left subfigures with the ISI (inter spike interval) distribution and the θ-coordinates, ($\tilde{\theta}_{i,j}$) in right two subfigures. Only $\{\eta_i\}$ is changed in (A)-(C); uniform in (A), with the ratio of $1:2$ in (B),with the ratio of $1:3$ in (C). Dashed line in the ISI figure is the distribution if we assume the inhomogeneous Poisson distribution with the same form of PSTH. The shaded region in $\tilde{\theta}_{i,j}$ subfigure indicates the range of the values that it takes, given a fixed $k = j - i$. The thick upper

and lower lines indicate the maximum and minimum values obtained by the approximation (in the main text.)

Figure 4: Example for illustration of MP model. (A) Shown is the relationship of the original parameters ($\{\eta_i\}$ i.e. PSTH, and $\{\pi_k, \lambda_k\}$), in two left subfigures, with the different order of θ-coordinates $\theta_{i_1,...,i_l}$, where l is up to 6, in right subfigure. Here $\{\eta_i\}$ is uniform, given by $\eta_i = 0.03$ (30 Hz); $\pi_k = 0.5$ $(k = 1, 2)$; $\lambda_1 = 4$ and $\lambda_2 = 8$. Dashed lines in the spike count figure indicates the distribution of each component (without weighing by π_k) in the mixture of Poisson distribution. This model is used as a reference. In the following figures (B)-(D), only some of the original parameters are changed, which is shown, and the adjacent figure is the corresponding $\theta_{i_1,...,i_l}$. Note that in changing some of the original parameters, the constraint $c_1 = \sum_i^N \eta_i = \sum_k^K \pi_k \lambda_k$ should hold. (B) Only $\{\eta_i\}$ is changed; the ratio of $\{\eta_i\}$ is changed as $1 : 2$ and $1 : 9$ in the left and right, respectively. In each order, there can be different values, due to the combination of the indices of $\{i_1, ..., i_l\}$ and the number of variation in each order results in $l + 1$. To show the variation, the plotted values in each order are slightly displaced horizontally. (C) Only λ_k is changed; (λ_1, λ_2) is $(2, 10)$ and $(1, 11)$ in the left and right, respectively. We do not show the case such as $(0, 12)$ or $(6, 6)$, because they correspond to a single Poisson distribution. (D) Only π_k is changed; (π_1, π_2) is $(0.2, 0.8)$ and $(0.8, 0.2)$ in the top and bottom, respectively. Due to the constraint of c_1, the summation of η_i changes in this case and therefore, the corresponding PSTH is also shown.

Figure 1

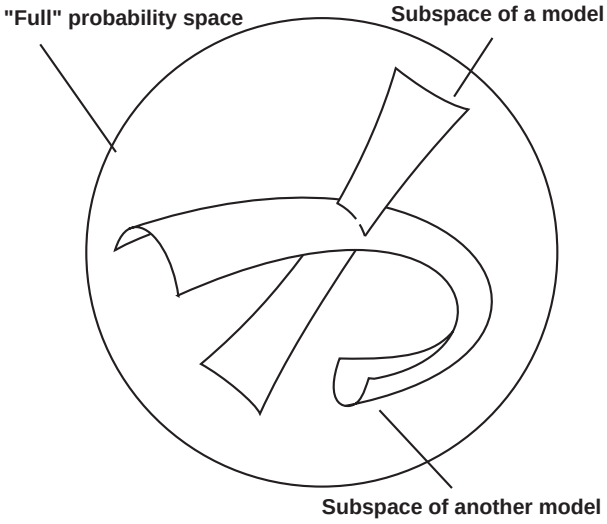

Figure 1/Nakahara

Figure 1:

Figure 2

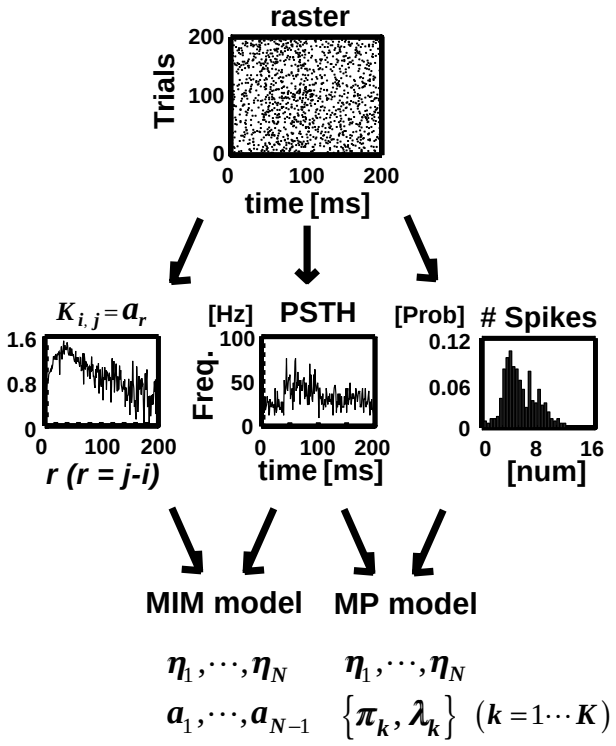

Figure 2/Nakahara

Figure 2:

Figure 3

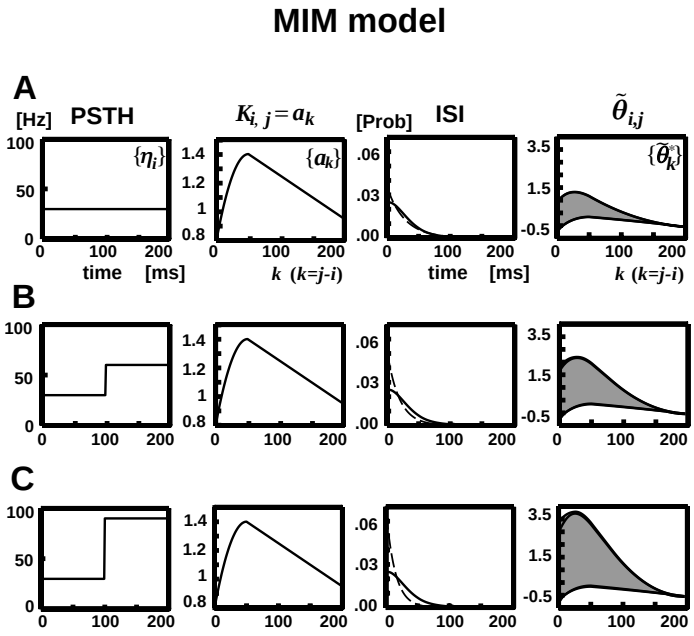

Figure 3/Nakahara

Figure 3:

Figure 4

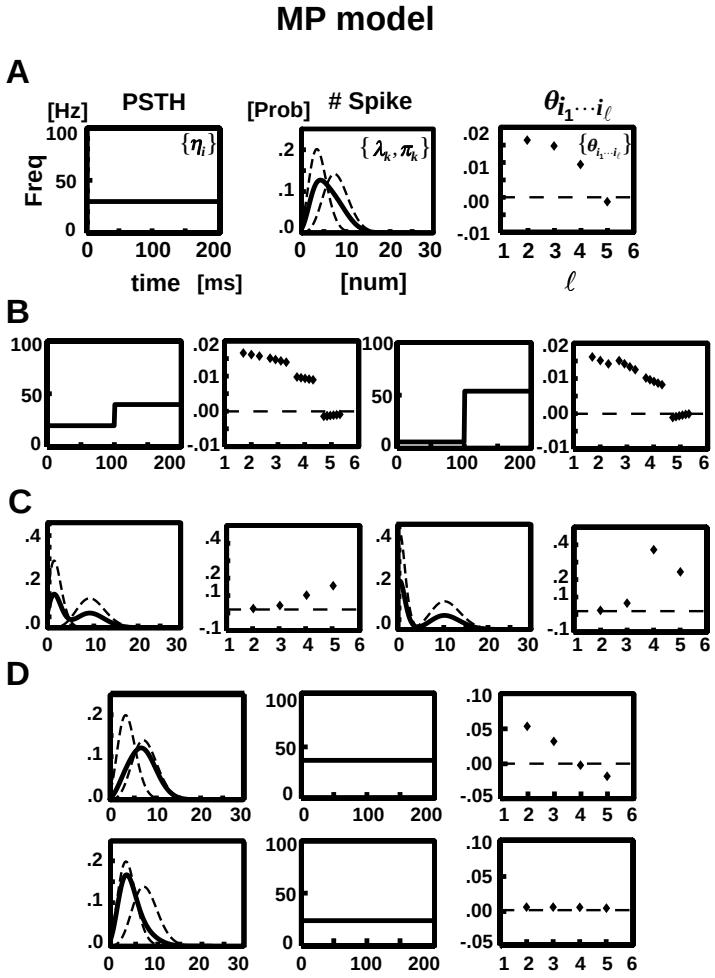

Figure 4/Nakahara

Figure 4: