Conditional Mixture Model for Correlated Neuronal Spikes

Shun-ichi Amari
amari@brain.riken.jp
Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, Wako-shi,
Hirosawa 2-1, Saitama 351-0198, Japan

Analysis of correlated spike trains is a hot topic of research in computational neuroscience. A general model of probability distributions for spikes includes too many parameters to be of use in analyzing real data. Instead, we need a simple but powerful generative model for correlated spikes. We developed a class of conditional mixture models that includes a number of existing models and analyzed its capabilities and limitations. We apply the model to dynamical aspects of neuron pools. When Hebbian cell assemblies coexist in a pool of neurons, the condition is specified by these assemblies such that the probability distribution of spikes is a mixture of those of the component assemblies. The probabilities of activation of the Hebbian assemblies change dynamically. We used this model as a basis for a competitive model governing the states of assemblies.

1 Introduction

Neurons emit spikes for transmitting information, and the spike timing is often stochastic. Spike trains may therefore be regarded as the outcome of a stochastic process (see, e.g., Kass, Ventura, & Brown, 2005). The simplest case is a Poisson process, which consists of a sequence of independent events with a fixed probability of occurrence. The behavior of a real neuron is not necessarily Poissonian, and for this reason, detailed considerations have been given to its description in terms of a renewal or Markovian process. The process is usually not stationary, and the firing rate generally varies over time.

When we consider a pool of neurons, the spikes from these neurons are generally not independent but correlated. Synchronous firing is an interesting area of research (Abeles, 1991; Gerstein & Perkel, 1969; Diesmann, Gewaltig, & Aertsen, 1999; Dayan & Abbot, 2005), in which correlations, particularly higher-order correlations, play a fundamental role (Martignon, von Hasseln, Grün, Aertsen, & Palm, 1995; Amari, Nakahara, Wu, & Sakai, 2003). We need to analyze correlated multineuron spiking processes with temporal structures (Salinas & Sejnowski, 2001). Here, we consider mostly the probability distributions of correlated spikes of a pool of neurons in

one time bin of 5 msec to 20 msec, and the firing rates (firing probabilities) of neurons may change over a longer time. We shall focus on the spatial correlations represented by pairwise covariances (e.g., de la Rocha, Doiron, Shea-Brown, Josic, & Reyes, 2007; Berger, Warren, Normann, Arieli, & Grün, 2007; Tatsuno, Fellous, & Amari, 2009), and touch on the higher-order moments.

The family of all joint probability distributions of spikes from n neurons has $2^n - 1$ degrees of freedom. Such a family has been studied by using information geometry (Amari & Nagaoka, 2000; Amari, 2001; Nakahara & Amari, 2002), where the firing rates and higher-order correlations are decomposed orthogonally (Amari, 2009). However, the full model is too complex and too detailed to be useful for explaining the stochastic nature of observed data. We need simpler tractable models. Such models include additive interaction, elimination, and replacement models (Cox & Isham, 1999; Kuhn, Aertsen, & Rotter, 2003; Kuhn, Rotter, & Aertsen, 2002; Niebur, 2007, see also Feng & Brown, 2000; Destexhe & Pare, 1999; Gutig, Aharonov, Rotter, & Sompolinsky, 2003). Recently Brette (2009) proposed a stochastic additive mixture model.

A mixture model is popular in statistics (Titterington, Smith, & Makov, 1985; McLachlan & Peel, 2000). We propose a conditional mixture model, which is a mixture of component distributions, where the mixing coefficients are determined by the state of the neuron pool. Typically when Hebbian cell assemblies coexist in a pool of neurons, the mixing coefficients are determined from the states of the assemblies. We first demonstrate that the additive, elimination, and replacement models can be regarded as special cases of our model. The mixture model is tractable because it does not have many degrees of freedom. We clarify its capabilities and limitations in terms of moments and show that our model is much simpler yet less restrictive than existing models.

After that, we study the state transition of a neural network composed of Hebbian assemblies within the framework of the conditional mixture model. We first show how covariances change during the state transition of the mixture model. When the state changes from one dominating cell assembly to another, the pairwise covariances increase midway in the firing sequence for two neurons in one and the same assembly, but they decrease for two neurons belonging to different assemblies. We further study the dynamics of Hebbian cell assemblies, in which the mixing parameters among cell assemblies are controlled by competitive dynamics.

This letter is organized as follows. Section 2 describes the joint probability distributions of correlated spikes as preliminaries. Section 3 explains the conventional models of generating and analyzing correlated spikes and shows that they are versions of our conditional mixture model. Section 4 describes properties of the mixture model and discusses the capabilities and limitations of our model. Section 5 describes algorithmic aspects of designing the mixture model for generation of prescribed correlational spikes

and of statistical estimation of a mixture model from observed data. Section 6 discusses the dynamics of state transitions using the mixture model. This describes competition among Hebbian cell assemblies to control the dynamics of the model's mixing parameters. Section 7 contains conclusions.

2 Joint Firing Probabilities

We consider a pool of n neurons and denote its state by a vector $x = (x_1, \ldots, x_n)$. Here, x_i is a binary random variable taking a value of 1 or 0: $x_i = 1$ when the ith neuron fires, and 0 otherwise. We assume that the state x is determined stochastically.

Let $p(x)$ be the joint probability of state x in a time bin. There are 2^n states, and

$$\sum_x p(x) = 1, \tag{2.1}$$

so that the distribution $p(x)$ has $2^n - 1$ degrees of freedom. We consider the set $S_n = \{p(x)\}$ of all probability distributions. Then it forms a $(2^n - 1)$-dimensional manifold. We introduce a coordinate system in it.

The firing probability (or shortly the firing rate) of neuron i is given by

$$r_i = \mathrm{E}[x_i] = \mathrm{Prob}\ \{x_i = 1\}, \tag{2.2}$$

where E denotes the expectation with respect to $p(x)$. The joint firing rate of neurons i and j is

$$r_{ij} = \mathrm{E}[x_i x_j] = \mathrm{Prob}\ \{x_i = x_j = 1\}, \tag{2.3}$$

and the joint firing rate of neurons i, j, and k is

$$r_{ijk} = \mathrm{E}[x_i x_j x_k] = \mathrm{Prob}\ \{x_i = x_j = x_k = 1\}. \tag{2.4}$$

Similarly, the firing rate of all neurons firing at the same time is

$$r_{12,\ldots,n} = \mathrm{E}[x_1 \ldots x_n]. \tag{2.5}$$

All of the above quantities can be summarized in a vector $r = \{r_i, r_{ij}, \ldots, r_{12,\ldots,n}\}$ having $2^n - 1$ components, and $p(x)$ is uniquely determined by it. Hence, r plays the role of a coordinate system in the space $S_n = \{p(x)\}$ consisting of all probability distributions. We call it the joint firing coordinate system (which is called the mixture or m-coordinate system in information geometry; Amari & Nagaoka, 2000).
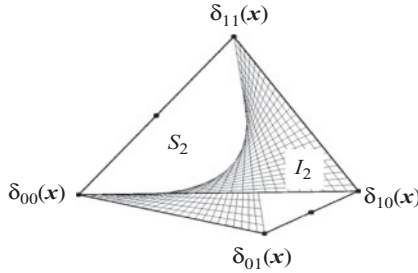
Figure 1: Set of probability distributions $S_2 = \{p(x_1, x_2)\}$ and the set I_2 of independent probability distributions.

Consider the case of $n = 2$, where x takes on only four states: $(1, 1), (1, 0), (0, 1), (0, 0)$. S_2 is represented by a three-dimensional simplex (see Figure 1), where the four corners (extreme points) represent special probability distributions denoted by $\delta_{11}(x), \delta_{10}(x), \delta_{01}(x), \delta_{00}(x)$. Here, $\delta_{11}(x) = 1$ if and only if $x_1 = x_2 = 1$; $\delta_{10}(x) = 1$ if and only if $x_1 = 1, x_2 = 0$; and so on. Hence, these distributions are deterministic. Any $p(x)$ can be written as a mixture of these extreme points, $p(x) = \sum p_{i_1,\dots,i_n} \delta_{i_1,\dots,i_n}(x)$. In the case of $n = 2$, $r = (r_1, r_2, r_{12})$ is a coordinate system of the simplex S_2.

When $p(x)$ is an independent distribution, it is a function of only the firing rates of neurons, $r_1, \dots, r_n$, and the other coordinates are functions of r_i, that is,

$$r_{ij} = r_i r_j, \tag{2.6}$$

$$r_{ijk} = r_i r_j r_k, \tag{2.7}$$

and so on. Hence, the set I_n of all independent distributions has only n degrees of freedom and forms an n-dimensional subspace I_n in S_n. When $n = 2$, the set I_2 of independent distributions is two-dimensional, and it can be represented by a surface in S_2, as shown in Figure 1.

The covariance of x_i and x_j is given by

$$c_{ij} = E[x_i x_j] - E[x_i]E[x_j] = r_{ij} - r_i r_j. \tag{2.8}$$

The covariance is 0 on I_n. However, even when all the pairwise covariances are 0, $p(x)$ might not be an independent distribution. There are higher-order correlations among more than two neurons (triple-wise correlations and higher), which are not reduced to pairwise correlations (Martignon et al., 1995; Amari et al., 2003).

3 Conditional Mixture Model

3.1 Models for Generating Correlated Spikes. The theoretical approach of studying S_n is not useful in practice because it includes too many parameters, $2^n - 1$. S_n is too complicated to be applied to experimental data. We need much simpler and more tractable models (Nakahara, Amari, & Richmond, 2006).

A number of models have been proposed for the purpose of generating correlated spike trains. They realize probability distributions $p(x)$ of correlated spikes that give correlated Poisson trains when spikes are temporally independent (Cox & Isham, 1980). As mentioned in section 1, they have been discussed in many papers (Kuhn et al., 2002, 2003; Feng & Brown, 2000; Bahjat, 2003; Stroeve & Gielen, 2001; Macke, Berens, Ecker, Tolias, & Bethge, 2008). In particular, Niebur (2007) proposed a new model that generates correlated spikes for stimulating a pool of neurons. Recently Brette (2009) proposed a more general model.

We prepare n independent Poisson processes $\tilde{x}_i(t)$, $i = 1, \dots, n; t = 1, \dots, N$, of n neurons and use a vector $\boldsymbol{p} = (p_1, \dots, p_n)$ for representing the firing rate p_i of neuron i. The spikes are spatially and temporally independent. To introduce correlations, we use another independent Poisson process $y(t)$, called a reference process, with firing probability w, and modify the independent $\tilde{x}_i(t)$ to give correlated $x_i(t)$. New processes $x_i(t)$ are generated from $\tilde{x}_i(t)$ and $y(t)$ by

$$x_i(t) = f\{\tilde{x}_i(t), y(t)\}, \tag{3.1}$$

where f is a deterministic function or a stochastic correspondence that will be explained below. The model Brette (2009) proposed uses a number of independent reference spike trains $y_i(t), i = 1, \dots, m$. We denote it in vector form $\boldsymbol{y}(t)$.

3.1.1 Additive Interaction Model. The additive interaction model generates $x_i(t)$ in such a way that, when $y(t) = 1$, an additional spike is randomly inserted in $\tilde{x}_i(t)$ with probability q_i independently for each neuron and each time. In other words, when $y(t) = 1$ and $\tilde{x}_i(t) = 0$, $\tilde{x}_i(t)$ changes into $x_i(t) = 1$ with probability q_i and $x_i(t) = \tilde{x}_i(t)$ otherwise. The firing rates for the new sequences are $r_i = (1 - wq_i)p_i + wq_i$, and each $x_i(t)$ is a Poisson process. They generate a correlated Poisson process, because the x_i's are correlated since $y(t)$ is common.

When $q_i = 1$, this model is deterministic, and has been used by Kuhn et al. (2003). This is called the single interaction process (SIP), where

$$x_i(t) = \tilde{x}_i(t) + y(t), \tag{3.2}$$

with the convention $1 + 1 = 1$. Therefore, when $y(t) = 1$, all $x_i(t)$ are 1, and synchronized spikes appear.

3.1.2 Eliminating Interaction Model. The eliminating interaction takes place in this model. When $y(t) = 1$, if $\tilde{x}_i(t)$ is 1, this is changed to $x_i(t) = 0$ with probability q_i independently. This elimination was called the multiple interaction by Kuhn et al. (2003). (See also Kuhn et al., 2002.)

3.1.3 Niebur Replacement Model. Niebur (2007) proposed a more powerful model. Here, $\tilde{x}_i(t)$ is replaced by $y(t)$ with probability q_i independently, irrespective of $y(t) = 0$ or $y(t) = 1$. Hence,

$$x_i(t) = \begin{cases} y(t), & \text{with probability } \sqrt{q_i} \\ \tilde{x}_i(t), & \text{with probability } 1 - \sqrt{q_i}, \end{cases} \tag{3.3}$$

where the replacements are independent for all i and t.

3.1.4 Brette Mixture Method. Brette (2009) proposed a new method, that uses m independent reference spikes $y_i, i = 1, \ldots, m$ and generates correlated spikes by

$$x_i = \sum c_{ik} y_k, \tag{3.4}$$

where $c_{ik} = 1$ with probability $p_{i,k}$ and 0 otherwise.

3.2 Conditional Mixture Model. We propose a new mixture model, a conditional mixture model, that includes all of the above models. Let y be a random variable, taking m values, $1, \ldots, m$, which specifies a state of a pool of neurons. The probability distribution of x is assumed to be $p(x \mid s)$, depending on the state $y = s, s = 1, \ldots, m$. Let w_s be the probability that $y = s$. Then the overall probability distribution is given by

$$p(x; w) = \sum w_s \, p(x \mid s), \tag{3.5}$$

where $w = (w_1, \ldots, w_m), \sum w_s = 1$.
    The model

$$M = \{p(x); w\} \tag{3.6}$$

is a mixture model, having m component distributions, $p(x \mid s), s = 1, \ldots, m$. Mixture models have been used in statistics (see, e.g., Titterington et al., 1985; McLachlan & Peel, 2000). The multilayer perceptron is also

regarded as a mixture model. However, one should be careful, because mixture models are singular where the Fisher information degenerates at some regions (see Amari, Park, & Ozeki, 2006).

When $p(x \mid s)$ is an independent distribution with a firing rate vector u^s,

$$p(x \mid s) = \prod_{i=1}^{n} p(x_i; u_i^s) = p(x; u^s),\qquad(3.7)$$

where $p(x, u)$ denotes the probability distribution such that $x = 1$ with probability u and $x = 0$ with probability $\bar{u} = 1 - u$, M consists of mixtures of independent distributions (see equation 3.7).

Here we show possible ideas as to how the condition is controlled. Some cortical neurons are known to be of two states; the up and down states. The firing rates differ depending on the states, and this difference may generate a mixture of probability distributions.

A mixture may also emerge from Hebbian cell assemblies. When Hebbian assemblies $H_1, \ldots, H_m$ coexist in a pool of neurons, we can represent the activity of a Hebbian assembly H_s by $p(x \mid s) = p(x, u^s)$, where u_i^s is large when x_i belongs to H_s. These assemblies coexist with weights (probabilities) w, and the weights w develop over time through dynamic interactions among the assemblies. We discuss this dynamics in a later section.

3.3 Examples of Mixture Models. We shall show that all models described in section 3.1 are versions of the conditional mixture model. Let us consider the case of $m = 2$, y taking on values of 0 and 1. We put

$$u_i^1 = \mathrm{Prob}\{x_i = 1 \mid y = 1\},\qquad(3.8)$$
$$u_i^2 = \mathrm{Prob}\{x_i = 1 \mid y = 0\}.\qquad(3.9)$$

Since the probability of $y = 1$ is $w_1 = w$ and that of $y = 0$ is $w_2 = \bar{w} = 1 - w$, the unconditional probability of $x_i = 1$ is given by

$$\mathrm{Prob}\{x_i = 1\} = w u_i^1 + \bar{w} u_i^2.\qquad(3.10)$$

This is a mixture of two independent distributions $p(x, u^1)$ and $p(x, u^2)$ specified by the firing rate vectors $u^1 = (u_1^1, \ldots, u_n^1)$ and $u^2 = (u_1^2, \ldots, u_n^2)$, respectively, and the new firing rate vector is given by

$$r = w u^1 + \bar{w} u^2.\qquad(3.11)$$

The three models **a, b, c** (Kuhn et al., 2002; Kuhn et al., 2003; Niebur, 2007) are mixture models with the following u^1 and u^2 and $w = \mathrm{Prob}\{y = 1\}$.

1. **Additive interaction model:**

$$u_i^1 = q_i + (1 - q_i)p_i \qquad (3.12)$$
$$u_i^2 = p_i \qquad (3.13)$$

2. **Eliminating interaction model:**

$$u_i^1 = p_i(1 - q_i) \qquad (3.14)$$
$$u_i^2 = p_i \qquad (3.15)$$

3. **Replacement model:**

$$u_i^1 = p_i + \sqrt{q_i}(1 - p_i) \qquad (3.16)$$
$$u_i^2 = p_i(1 - \sqrt{q_i}) \qquad (3.17)$$

4. **Brette model.** In this case, the state is specified by a vector y, which has 2^m states (Brette, 2009). When the state is y, the firing rates $r = E[x \mid y]$ are given by

$$u_i(y) = \sum_{k=1}^{m} p_{i,k} y_k. \qquad (3.18)$$

Hence,

$$p(x \mid y) = p(x, u(y)), \qquad (3.19)$$

and their mixtures are written as

$$p(x; w) = \sum w(y) p(x \mid y). \qquad (3.20)$$

4 Properties of Mixture Models

We need a simple but reasonably powerful model of distributions, with which we can easily generate correlated spike sequences for stimulating a pool of neurons and also for analyzing correlational structures from the observed data. The conditional mixture model is useful for this purpose.

4.1 Mixture of Independent Distributions. Let $p(x, u)$ be an independent distribution of n neurons with a firing rate vector $u = (u_1, \ldots, u_n)$. Our model $p(x; w, u^1, \ldots, u^s)$ given by equation 3.5 is a mixture of m independent distributions $p(x, u^m)$ with mixing rate w

$$p(x; w) = \sum w_s p(x, u^s). \qquad (4.1)$$

In order to analyze the properties of the mixture model, we need the following lemma. Any function $f(x)$ of a binary vector x can be written in a reduced polynomial as

$$f(x) = \sum_k \sum_{i_1,\ldots,i_k} c_{i_1,\ldots,i_k} x_{i_1}, \ldots, x_{i_k}, \tag{4.2}$$

where the indices $i_1, \ldots, i_k$ are all different, because of $x_i^2 = x_i$.

Lemma 1. *When $f(x)$ is a reduced form,*

$$E\left[f(x)\right] = \sum w_s f\left(u^s\right). \tag{4.3}$$

Proof. Since $p(x \mid s)$ are independent distributions, from

$$E[f(x)] = E_y E_{x|y}[f(x) \mid y] = \sum w_s E[f(x) \mid s], \tag{4.4}$$

where E_y is expectation with respect to y and $E_{x|y}$ is conditional expectation over x conditioned on y, we have the theorem.

From this we get the firing rate vector:

$$r = E[x] = \sum w_s u^s. \tag{4.5}$$

The joint firing rates are also given similarly, for example,

$$r_{i_1 \cdots i_k} = \sum w_s u_{i_1}^s, \ldots, u_{i_k}^s. \tag{4.6}$$

To study the higher-order moments, let us decompose u^s as a sum of the mean firing rate r and deviation δ^s from the mean,

$$u^s = r + \delta^s, \tag{4.7}$$

where

$$\sum w_s \delta^s = 0. \tag{4.8}$$

The kth order central moments of x are

$$c_{i_1,\ldots,i_k} = E[(x_{i_1} - r_{i_1}) \cdots (x_{i_k} - r_{i_k})]. \tag{4.9}$$

Theorem 1. *The kth order central moments of x are given by*

$$c_{i_1,\ldots,i_k} = \sum w_s \delta_{i_1}^s, \ldots, \delta_{i_k}^s. \tag{4.10}$$

The theorem describes a means of calculating moments in an independent mixture model. Obviously when $p(x)$ is independent, all central moments vanish.

When $m = 2$, we may put

$$\delta^1 = \frac{1}{w_1}\delta, \quad \delta^2 = -\frac{\delta}{w_2}. \tag{4.11}$$

Hence, the moments are, for $w = w_1$ and $\bar{w} = w_2$

$$c_{i_1 \cdots i_k} = \left(\frac{1}{w^{k-1}} + (-1)^k \frac{1}{\bar{w}^{k-1}} \right) \delta_{i_1} \delta_{i_2}, \ldots, \delta_{i_k}. \tag{4.12}$$

In particular, the covariances become

$$c_{ij} = \frac{1}{w\bar{w}} \delta_i \delta_j. \tag{4.13}$$

Hence, the covariance of x_i and x_j is positive when δ_i and δ_j have the same sign, that is,

$$\left(u_i^1 - u_i^2 \right) \left(u_j^1 - u_j^2 \right) > 0, \tag{4.14}$$

and negative, otherwise.

4.2 Mixture with Two Neurons. We will discuss the simple case of two neurons. In the case of $n = 2$, $m = 2$, the mixture model M_2 includes $2n + 1 = 5$ parameters, while the manifold of all distributions S_2 is three-dimensional. We first show that any distribution of S_2 can be realized in our mixture model M_2. It should be noted that the additive, elimination, Niebur replacement, and Brette mixture models cannot generate some of distributions in S_2.

Theorem 2. *M_2 includes all the distributions in S_2.*

Proof.   Let us consider the mapping of π from M_2 to S_2 given by equation 4.1. The image of $\pi(M_2)$ consists of mixtures of two independent distributions. S_2 is a simplex having four faces and six edges (see Figure 1). Of the six edges, four are composed of independent distributions. For example,

any distribution $p(x)$ connecting the vertices $\delta_{11}(x)$ and $\delta_{10}(x)$ can be written in product form as

$$w\delta_{11}(x) + \bar{w}\delta_{10(x)} = p(x_1, 1)\, p(x_2, w). \tag{4.15}$$

However, the distributions connecting the vertices of $\delta_{11}(x)$ and $\delta_{00}(x)$ are not independent. All faces of S_2 are surrounded by three edges, where two are composed of independent distributions. Therefore, any distribution in the faces of S_2 can be expressed as a linear mixture of two independent distributions. This means that the four faces are in the image of π. Since M_2 has no holes, its image given by a continuous mapping, equation 4.1, cannot have holes because of homotopy. Hence, the whole of S_2 is covered by the image of π, proving the theorem.

Remark. One can easily see that δ_i are always positive or always negative in the additive, elimination, and replacement models. Hence, $\delta_i\delta_j > 0$, so that negative covariances $c_{ij} < 0$ cannot be realized in these models. Even the more general Brette model cannot realize negative correlations.

4.3 Limitations of the Mixture Model. Since M_n includes only $mn + m - 1$ parameters, M_n includes only a very limited range of probability distributions. The limitation is given by theorem 1 that all the moments $c_{i_1,\dots,i_k}$ are determined by the deviation vectors δ^s. When $m = 2$, they are determined by a vector δ.

In the case with three neurons, $n = 3$ and $m = 2$, the number of dimensions is $2n + 1 = 7$, the same number as S_3. However, not all distributions in S_3 can be represented in M_3 with $m = 2$. The covariances c_{ij} that can be represented by using the mixture model satisfy $c_{ij} = w\bar{w}\delta_i\delta_j$, as shown in equation 4.13. Hence, when $c_{12} < 0, c_{23} < 0, c_{31} < 0$, we cannot find a set of $\delta_1, \delta_2, \delta_3$ satisfying these inequalities. This shows that a distribution for which all three pairs are negatively correlated cannot be represented in our model. Even when all the covariances are positive, there are still distributions that cannot be represented because equation 4.13 would be violated.

4.4 Higher-Order Restrictions. A mixture connects a number of distributions in a mixture-flat submanifold (Amari & Nagaoka, 2000) M_n. Since M_n is not exponentially flat, higher-order correlations exist. For example, for a distribution $p(x_1, x_2, x_3)$, the third-order correlation orthogonal to the pairwise firing rates is given by

$$\theta_{123} = \log\frac{p(1, 1, 1)p(1, 0, 0)p(0, 1, 0)p(0, 0, 1)}{p(1, 1, 0)p(0, 1, 1)p(1, 0, 1)p(0, 0, 0)}. \tag{4.16}$$

Although we can calculate θ_{123} in terms of w, r, δ^s, it is a very complicated function of the parameters. It would be interesting to see what type of higher-order correlations emerges from a mixture model.

Here, we consider the higher-order moments realized by a mixture model. Since the moments $c_{i_1,\ldots,i_k}$ are determined by w and deviations $\delta^1, \ldots, \delta^m$, they are within the restricted form of equation 4.10 in M_n.

When $m = 2$, we have

$$c_{i_1,\ldots,i_k} = a_k(w)\delta_{i_1}\delta_{i_2}, \ldots, \delta_{i_k}, \tag{4.17}$$

where

$$a_k(w) = \frac{1}{w^{k-1}} + (-1)^k \frac{1}{\bar{w}^{k-1}}. \tag{4.18}$$

For example,

$$c_{ij} = \left(\frac{1}{w} + \frac{1}{\bar{w}}\right)\delta_i\delta_j, \tag{4.19}$$

$$c_{ijk} = \left(\frac{1}{w^2} - \frac{1}{\bar{w}^2}\right)\delta_i\delta_j\delta_k, \tag{4.20}$$

and so on. They show that higher-order moments are related to lower-order moments through common δ.

4.5 State Transitions Through a Mixture. When an ensemble of neurons processes information, its state changes dynamically. Let us assume that the initial state is represented by a probability distribution $p_1(x)$ and the final state by $p_2(x)$. Here, $p_1(x)$ and $p_2(x)$ do not need to be independent distributions. One may assume that they correspond to two different Hebbian cell assemblies coexisting in a network. At $t = 0$, the state is $p_1(x)$, and it changes into the final state $p_2(x)$ at $t = 1$. Information is processed in the network while the state transition takes place.

We assume that the state transition is represented by the mixture model,

$$p(x, t) = (t)p_2(x) + (1 - t)p_1(x), \tag{4.21}$$

and check how the covariances among neurons change during the state transition. This is a mixture model in which the mixing rate $w = t$ changes from 0 to 1 over time.

Sakamoto et al. (2008) observed increases in pairwise covariances among some neurons during state transitions. They conjectured that this fact indicates an increase in synchronous firing in information processing. The mixture model provides a good tool for analyzing such phenomena. Let

c_{1ij} and c_{2ij} be the covariances of neurons i and j of $p_1(x)$ and $p_2(x)$, and let u^1 and u^2 be their firing rate vectors. Then we have

$$c_{ij}(t) = (1 - t)c_{1ij}(u) + tc_{2ij}(v) + t(1 - t)\left(u_i^1 - u_i^2\right)\left(u_j^1 - u_j^2\right). \tag{4.22}$$

When $p_1(x)$ and $p_2(x)$ are independent distributions, we have

$$c_{ij}(t) = t(1 - t)\left(u_i^1 - u_i^2\right)\left(u_j^2 - u_j^2\right). \tag{4.23}$$

The covariances $c_{ij}(t)$ between two neurons i and j are not only a mixture of those of the initial and final states, but they have an additional term due to the change in firing rates. The additional effect is maximized at $t = 1/2$. It is interesting to see that the covariance increases when the firing rates of neurons i and j increase together or decrease together through the state transition. On the other hand, the covariance decreases when the firing rate of one neuron increases while the other decreases.

5 Applications of Conditional Mixture Model

The mixture model will be used for generating correlated spikes and analyzing observed data. This section discusses these problems.

5.1 Generating Correlated Spikes. A mixture model is specified by the mixing vector w, the firing rate vector r, and deviation vectors δ^s, $s = 1, \ldots, m$, satisfying equation 4.8. We may use them as a new coordinate system to specify the probability

$$p(x, w, r, \delta^s) = \sum w_s p(x; r + \delta^s). \tag{5.1}$$

When the firing rates $\tilde{r}$ and covariances $\tilde{c}_{ij}$, $i \neq j$, are specified, we need to search for w and δ^s that satisfy the requirements approximately. The firing rate r is given by $r = \tilde{r}$. Since a mixture model is limited, we can only approximate given $\tilde{c}_{ij}$. Since the covariances of a mixture model are given by

$$c_{ij} = \sum w_s \delta_i^s \delta_j^s, \tag{5.2}$$

we may use the least squares method to design w and δ^s. The cost function is

$$L(w, \delta^1, \ldots, \delta^m) = \frac{1}{2} \sum_{ij} (c_{ij} - \tilde{c}_{ij})^2, \tag{5.3}$$

under constraint 4.8 and

$$0 < r_i + \delta_i^s < 1. \tag{5.4}$$

This is a nonlinear programming problem. We may use the gradient descent method or others to solve it.

Let us consider a special case in which neurons are divided into two groups. Neurons in each group are homogeneous, having the same firing rate and the same covariance. The first group of neurons is assumed to be independent. Let c be the covariances between two neurons in the second group. We realize such a distribution by a mixture model with $m = 2$. Since the first group consists of independent neurons, the deviation vector $\boldsymbol{\delta}^{(1)} = 0$ in this group. The deviation vector of the second group has the same component, $\boldsymbol{\delta}^2 = (\delta, \dots, \delta)$. It is determined from

$$c = \frac{1}{w\bar{w}}\delta^2. \tag{5.5}$$

When we have three homogeneous neuron groups, we can realize the probability distribution with $m = 3$.

5.2 Analyzing Observed Data by Statistical Estimation. We can use a mixture model to analyze experimental data for the purpose of estimating the underlying stochastic structure. Given experimental observations $x(1), \dots, x(t)$, we can estimate w, r, $\boldsymbol{\delta}^s$ of the parameters of the mixture model, equation 5.1. This is an ordinary estimation problem. However, a mixture model is singular, and there are computational difficulties in estimation. (See Amari et al., 2006; Wei & Amari, 2008; and Cousseau, Ozeki, & Amari, 2008, for such singular models.)

When data are summarized in the observed firing rates $\tilde{r} = (\tilde{r}_i)$ and covariances $\tilde{c}_{ij}$, we can solve the estimation problem by the least-square method of minimizing equation 5.3. This is known as the moment method of estimation. The estimator is usually not Fisher efficient but consistent (Wu, Amari, & Nakahara, 2002).

We may use the EM algorithm to solve the estimation problem (Dempster, Laird, & Rubin, 1977; Csiszár & Tusnády, 1984; Amari, 1995). Let us consider random variables (x, z), where $z = 1, \dots, m$, shows that x is generated from independent distribution $p(x, u^z)$, $u^z = r + \boldsymbol{\delta}^z$. If we know $z(t)$ together with $x(t)$, $t = 1, \dots, T$, the estimation is simple and easy. However, $z(t)$ are hidden. The EM algorithm is useful for estimation when some random variables are hidden. It is an alternative minimization method (Csiszár & Tusnády, 1984; Amari, 1995) consisting of the E-step and M-step.

When a candidate distribution $p(x, z; \hat{w}, \hat{r}, \hat{\delta}^s)$ is given, we calculate the conditional estimate of

$$L(w, r, \delta^s) = E\left[\sum_t \log p(x(t), z(t); w, r, \delta^s)\right], \tag{5.6}$$

where E is the conditional expectation with respect to $z(t)$ by using

$$p(z \mid x; \hat{w}, \hat{r}, \hat{\delta}^s). \tag{5.7}$$

The M-step searches for the next estimate of w, r, δ^s by maximizing $L(w, r, \delta^s)$.

The EM algorithm does not always guarantee the convergence to the true minimizer, but it is computationally easy to obtain the maximum likelihood estimator.

6 Discussions on Dynamics of Hebbian Assemblies

We discuss here a possible application of the conditional mixture model to the dynamics of competition among Hebbian assemblies. Let us assume that a pool of n neurons is composed of m Hebbian assemblies $H_1, \ldots, H_m$. When assembly H_s is active, the probability distribution is $p(x \mid s)$, in which the neurons belonging to H_s have high firing rates. Let u_i^s be the firing rate of neuron i when H_s is active:

$$p(x \mid s) = p(x; u^s). \tag{6.1}$$

We assume that $p(x \mid s)$ is independent, but we may treat the case with a correlated distribution for each Hebbian assembly.

A neuron may belong to a number of Hebbian assemblies at the same time. Which neuron belongs to which Hebbian assembly is fuzzy, and we may consider a_{is} the membership value of neuron i belonging to assembly H_s, $0 \leq a_{is} \leq 1$. When a_{is} is large, the degree of neuron i belonging to H_s is large. We may use u_i^s as the membership value.

The conditional mixture model is given by

$$p(x, w) = \sum w_s p(x \mid s), \tag{6.2}$$

where w_s is the probability or the degree that H_s is activated. The weight vector $w = (w_1, \ldots, w_m)$ changes depending on the current firing pattern x of neurons. We assume that there is a neural system that controls the level

w_s of activation of H_s. The total amount of stimuli that the system receives for H_s is

$$X_s = \sum a_{is} x_i, \tag{6.3}$$

which is a weighted sum of x_i with weight a_{is} of x_i belonging to H_s. Since the average firing rate of x_i is $\sum w_s u_i^s$, the average of X_s is

$$E[X_s] = \sum_{r,i} w_r u_i^r a_{is}. \tag{6.4}$$

It is assumed that the activity w_s of H_s is controlled by a continuous-time dynamics of the type

$$\tau \frac{dw_s}{dt} = -w_s + f[\bar{X}_s] - I_0 + I_s. \tag{6.5}$$

Here, $\bar{X}_s = E[X_s]$ is the temporal average of X_s, f is a sigmoidal function, and τ is a time constant. It increases in proportion to a nonlinear function of the short-time temporal average of the activity X_s of the Hebbian assembly. I_s is an input to H_s from the outside, and

$$I_0 = \frac{1}{m} \sum (f(\bar{X}_s) + I_s - 1) \tag{6.6}$$

is a common inhibitory term, which makes it possible to preserve the total probabilities,

$$\sum \frac{dw_s}{dt} = 0. \tag{6.7}$$

Therefore equation 6.6 preserves the constraint

$$\sum w_s = 1. \tag{6.8}$$

Since the activation function $f(u)$ is a sigmoid function, equation 6.8 is a typical competition model. We illustrate this in a simple example in which the membership function a_{is} is hard,

$$a_{is} = \begin{cases} 1, & \text{when } x_i \in H_s, \\ 0, & \text{otherwise.} \end{cases} \tag{6.9}$$

and $u_i^s = 0$ for $x_i \notin H_s$. Then,

$$\bar{X}_s = w_s \sum_i u_i^s = w_s u^s, \tag{6.10}$$

where $u^s = \sum u_i^s$ is a constant. Then the dynamics reduces to

$$\tau \frac{d w_s}{d t} = -w_s + f\left(w_s u^s\right) - \frac{1}{m} \sum \left\{ f(w_r u^r) - 1 + I_r \right\}. \tag{6.11}$$

This is a typical winner-takes-all or winners-share-all model (Amari & Arbib, 1977; Fukai & Tanaka, 1997), and eventually one or some of Hebbian assemblies win, depending the inputs I_r.

There are many other possibilities of dynamics in the conditional mixture model and it is an interesting problem for future research.

7 Conclusions

We showed that many existing models for generating correlated spikes are versions of our conditional mixture model. Although our model might not explain the mechanism responsible for dynamical interactions in a real neuron pool, it is a convenient statistical model for generating correlated spike sequences for stimulation (Niebur, 2007; Brette, 2009). Moreover, we may use it as a simple working model to analyze experimental data and estimate the intrinsic correlational structure.

We also showed a model in which each Hebbian assembly generates a component probability distribution and proposed a dynamical model of competition among Hebbian assemblies. This is a dynamical model in the space of a mixture family.

Acknowledgments

References

Abeles, M. (1991). *Corticonics: Neural circuit of cerebral cortex.* Cambridge: Cambridge University Press.

Amari, S. (1995). Information geometry of the EM and EM algorithms for neural networks. *Neural Networks, 8,* 1379–1408.

Amari, S. (2001). Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory, 47*(5), 1701–1711.

Amari, S. (2009). Measure of correlation orthogonal to change in firing rate. *Neural Computation, 21*, 960–972.

Amari, S., & Arbib, M. A. (1977). Competition and cooperation in neural nets. In J. Metzler (Ed.), *Systems neuroscience* (pp. 119–165). Orlando, FL: Academic Press.

Amari, S., & Nagaoka, H. (2000). *Methods of information geometry*. New York: AMS & Oxford University Press.

Amari, S., Nakahara, H., Wu, S., & Sakai, Y. (2003). Synchronous firing and higher-order interactions in neuron pool. *Neural Computation, 15*, 127–142.

Amari, S., Park, H., & Ozeki, T. (2006). Singularities affect dynamics of learning in neuromanifolds. *Neural Computation, 18*, 1007–1065.

Bahjat, F. Q. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika, 90*(2), 455–463.

Berger, D., Warren, D., Normann, R., Arieli, A., & Grün, S. (2007). Spatially organized spike correlation in cat visual cortex. *Neurocomputing, 70*, 2112–2116.

Brette, R. (2009). Generation of correlated spike trains. *Neural Computation, 21*, 188–215.

Cousseau, R., Ozeki, T., & Amari, S. (2008). Dynamics of learning in multilayer perceptrons near singularities. *IEEE Transactions on Neural Networks, 19*, 1313–1328.

Cox, D. R., & Isham, V. (1980). *Point processes*. London: Chapman & Hall.

Csiszár, I., & Tusnády, G. (1984). Information geometry and alternating minimization procedures. *Statistics and Decisions,* Suppl. Issue 1, 205–237.

Dayan, P., & Abbott, L. F. (2005). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

de la Rocha, J., Doiron, B., Shea-Brown, E., Josic, K., & Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature, 448*, 802–806.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, B-39*, 1–38.

Destexhe, A., & Pare, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J. Neurophysiol., 81*(4), 1531–1547.

Diesmann, M., Gewaltig, M. O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature, 402*, 529–533.

Feng, J., & Brown, D. (2000). Impact of correlated inputs on the output of the integrate-and-fire model. *Neural Computation, 12*, 671–692.

Fukai, T., & Tanaka, S. (1997). A simple neural network exhibiting selective activation of neuronal ensembles: From winner-take-all to winners-share-all. *Neural Computation, 9*, 77–97.

Gerstein, G. L., & Perkel, D. H. (1969). Simultaneously recorded trains of action potentials: Analysis and functional interpretation. *Science, 164*, 828–830.

Gutig, R., Aharonov, R., Rotter, S., & Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *J. Neurosci., 23*(9), 3697–3714.

Kass, R. E., Ventura, V., & Brown, E. N. (2005). Statistical issue in the analysis of neuronal data. *Journal of Neurophysiology, 94*, 8–25.

Kuhn, A., Aertsen, A., & Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Computation, 15*, 67–101.

Kuhn, A., Rotter, S., & Aertsen, A. (2002). Correlated input spike trains and their effects on the leaky integrate-and-fire neuron. *Neurocomputing, 44–46*, 121–126.

Macke, J. H., Berens, Ph., Ecker, A. S., Tolias, A. S., & Bethge, M. (2008). Generating spike trains with specified correlation coefficients. *Neural Computation, 21*, 397–423.

Martignon, L., von Hasseln, H., Grün, S., Aertsen, A., & Palm, G. (1995). Detecting higher-order interactions among the spiking events in a group of neurons. *Biological Cybernetics, 73*(1), 69–81.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. Hoboken, NJ: Wiley.

Nakahara, H., & Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Computation, 14*, 2269–2316.

Nakahara, H., Amari, S., & Richmond, B. J. (2006). A comparison of descriptive models of a single spike train by information-geometric measure. *Neural Computation, 18*, 545–568.

Niebur, E. (2007). Generation of synthetic spike trains with defined pairwise correlations. *Neural Computation, 19*, 1720–1738.

Sakamoto, K., Mushiake, M., Saito, N., Aihara, K., Yano, M., & Tanji, J. (2008). Discharge synchrony during the transition of behavioral encoded by discharge rates of prefrontal neurons. *Cerebral Cortex, 18*, 2036–2045.

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience, 2*(8), 539–550.

Stroeve, S., & Gielen, S. (2001). Correlation between uncoupled conductance-based integrate-and-fire neurons due to common and synchronous presynaptic firing. *Neural Computation, 13*, 2005–2029.

Tatsuno, M., Fellous, J.-M., & Amari, S. (2009). Information geometric measures as robust estimators of connection strength and external inputs. *Neural Computation, 21*, 2309–2335.

Titterington, D., Smith, A., & Makov, U. (1985). *Statistical analysis of finite mixture distributions*. Hoboken, NJ: Wiley.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., et al. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioral events. *Nature, 373*, 515–518.

Wei, H., & Amari, S. (2008). Dynamics of learning near singularities in radial basis function networks. *Neural Networks, 21*, 989–1005.

Wu, S., Amari, S., & Nakahara, H. (2002). Population coding and decoding in a neural field: A computational study. *Neural Computation, 14*, 999–1026.