SIAM J. APPL. MATH.
Vol. 33, No. 1, July 1977

A MATHEMATICAL FOUNDATION FOR STATISTICAL NEURODYNAMICS*

SHUN-ICHI AMARI†, KIYONORI YOSHIDA‡ AND KEN-ICHI KANATANI†

Abstract. The brain is a large-scale system composed of an enormous number of neurons. In order to understand its functioning, we need to know the macroscopic behavior of a nerve net as a whole. Statistical neurodynamics treats an ensemble of nets of randomly connected neurons and derives macroscopic equations from the microscopic state transition laws of the nets. There arises, however, a theoretical difficulty in deriving the macroscopic state equations, because of possible correlations among the microscopic states. The situation is similar to that encountered in deriving the Boltzmann equation in statistical mechanics of gases.

We first elucidate the stochastic structures of random nerve nets. We then derive macroscopic state equations which apply to a wide range of ensembles of random nets. These equations are shown to hold in a weak sense: we prove that the probability that these equations are valid within an arbitrarily small error and for an arbitrarily long time converges to 1 as the number n of the component elements in a random net tends to infinity. We also derive the macroscopic equations describing the dynamic behavior of the state correlations, and prove that these equations hold in the weak sense. The strong assertion which states the uniform convergence in time, is proved for a special class of random nets.

1. Introduction.

The study of the human brain is probably one of the most active areas of science today. The challenge is to understand the fundamental functioning of the highly intellectual activities characteristic of human beings. Difficulties arise, however, due to the great complexity of the brain and its enormous number of elements. Our main concern in this paper is the study of such large scale systems. When one is dealing with a system consisting of a large number of elements, one is, in many cases, not interested in the detailed behavior of the individual elements. Instead, one is usually interested in macroscopic behavior of the system. The problem is to deduce the equations which describe the macroscopic behavior of the system from a knowledge of the individual elements. Thus our task is to find macroscopic state equations on the basis of the microscopic state equations. This problem has much in common with those encountered in the study of statistical mechanics [14], [24].

We study nets composed of idealized neurons interconnected with one another. Following the method of statistical mechanics, we consider an ensemble of nerve nets instead of treating any specific net. We then introduce a probability measure on the ensemble in such a manner that the specific net can be regarded as a typical sample from the ensemble. The probability measure is naturally introduced when we treat an ensemble of nets in which neuron elements are connected in a random manner subject to some prescribed probability distribution. In this case, a sample net is called a random nerve net. We study the dynamic behavior of nerve nets which is valid for almost all typical sample nets of the ensemble. That is, in terms of probability theory, we study those properties whose probability converges to 1 as the number n of the components elements tends to infinity.

* Received by the editors July 22, 1975, and in revised form May 15, 1976.

† Department of Mathematical Engineering and Instrumentation Physics, University of Tokyo, Tokyo 113, Japan.

‡ Department of Electrical Engineering, Yamanashi University, Kofu, Yamanashi 400, Japan.

Random nets consisting of formal neurons have been investigated by many authors, and many computer-simulated experiments have been done (e.g., [1], [7], [8], [11], [13], [16], [17], [22]). The macroscopic state equations are derived by Rozonoer [20], Amari [2] and others for various types of nets. Amari [2], [3], [6] has studied the macroscopic behavior of random nets in detail. He derived the catastrophe curve for random nerve nets in the sense of Thom [23]. When the structural parameters of a net change across the catastrophe curve, a sudden qualitative change appears in its behavior.

There is, however, a certain difficulty in deriving the macroscopic state equation, as Rozonoer [19] has pointed out (see also Amari [5]). In order to obtain the macroscopic equations, one must assume the statistical postulate that possible correlations between the states of the neuron elements can be neglected, provided the number n of the elements is sufficiently large. The situation is again the same as in statistical mechanics; the so-called assumption of molecular chaos or "Stosszahlansatz" is employed in the derivation of the famous Boltzmann equation in the kinetic theory of gases. This postulate states that the correlation between particles after collisions can be neglected. The postulate gave rise to a major controversy in the history of statistical mechanics (see Kac [14]). The analogy between statistical mechanics and random nerve nets was discussed by Rozonoer with the help of Kac's work. He presented many important concepts and hypotheses in his discussion [19]. This was the first important step in establishing the theoretical foundations of statistical neurodynamics.

The statistical postulate is said to hold in the strong sense, if the correlations vanish uniformly in time t as the number n tends to infinity. It is said to hold in the weak sense, if the convergence is not necessarily uniform. Rotenberg [18] first proved the weak postulate for a very special ensemble of nerve nets.

In the present paper, we will give, along the lines proposed in [5], [18], [19], a rigorous mathematical foundation of statistical neurodynamics. In the beginning we will treat ensembles consisting of stochastically homogeneous random nerve nets of formal neurons of the McCulloch–Pitts type. We will prove the statistical postulate in the weak sense for these ensembles. It will then be shown that our method is applicable to more general ensembles of nerve nets. The macroscopic state equations will be derived for these general random nets. Moreover, generalizing Amari's method of approach [5], we will show that the correlations of a number of different states themselves can be treated as macroscopic quantities characterizing the net behavior as a whole. The dynamical equations will be derived for these generalized states of correlations and the statistical postulate in the weak sense will also be proved. The postulate in the strong sense will be proved only for a special class of ensembles.

2. Macroscopic description of a large system. In this section we will discuss the condition under which the macroscopic description is possible in a system composed of a large number, say n, of elements. Let us denote the state of the ith element at time t by $x_i(t)$ (which may be a vector quantity). Then, the state of the system at time t is described by a vector of n components

$$\mathbf{x}(t) = (x_1(t), x_2(t), \cdots, x_n(t)).$$

We assume that the system operates in discrete time, $t=0,1,2,\cdots$, in accordance with the state transition equation

$$x_i(t+1)=T_i(x_1(t),\cdots,x_n(t);\omega),\qquad i=1,2,\cdots,n,$$

or in vector notation

$$\mathbf{x}(t+1)=T_\omega\mathbf{x}(t), \tag{2.1}$$

where ω denotes a set of parameters specifying the structure of the system and T_ω is the state transition operator. Then we can write

$$\mathbf{x}(t)=T_\omega^t\mathbf{x}(0).$$

It should be noted that $\mathbf{x}(t)$ depends on ω, i.e., on the structure of the net in which the state transition takes place successively. Therefore, $\mathbf{x}(t)$ should be written as $\mathbf{x}(t;\omega)$, when we need to show the explicit dependence on the net parameters.

If the explicit form of the transition equation (2.1) is known, the behavior of the whole system is understood in principle. However, it is almost impossible to obtain analytic solutions of (2.1) if the number n is very large. Even if numerical solutions are obtainable by the use of big computers, our interest is in many cases not to know the detailed behaviors of the individual elements, but to know the macroscopic properties of the system as a whole. In such cases, it is most appropriate to introduce the statistical method. To this end, we treat an ensemble $\mathcal{N}_n$ consisting of similar systems instead of considering a specific system, where n is the number of elements of every system. Then the probability of selecting a particular system from the ensemble is defined in such a manner that the system of our interest is regarded as a typical sample from the ensemble.

Now we introduce a probability measure on the ensemble $\mathcal{N}_n$ of systems described by the transition equation (2.1), by considering the set of the system parameters ω as random variables subject to some probability distribution. What we are most concerned with is the dynamical behavior of the system which is common to almost all systems of the ensemble, when n is very large. In mathematical terminology, we are interested in the asymptotic behavior, whose probability converges to 1 in the limit $n\to\infty$. Therefore, we consider an infinite sequence $\mathcal{N}_n$ of the ensemble, $n=1,2,3,\cdots$. The distribution functions F_n of the system parameters ω of $\mathcal{N}_n$ need to be determined in such a way that the asymptotic structure of the system has a definite meaning.

In order to know the macroscopic properties of a system in $\mathcal{N}_n$, we investigate the dynamical behavior of certain macroscopic quantities which characterize the state of the system. Let such quantities be given by a set of functions of the state variables $\mathbf{x}=(x_1,\cdots,x_n)$

$$X_1(\mathbf{x}),\cdots,X_r(\mathbf{x})$$

or in vector notation

$$\mathbf{X}=\mathbf{X}(\mathbf{x}).$$

We call $\mathbf{X}(\mathbf{x})$ a macroscopic state corresponding to the microscopic state $\mathbf{x}$. If the state $\mathbf{x}$ is transformed by the state transition operator T_ω to a new state $T_\omega\mathbf{x}$, the macroscopic state becomes $\mathbf{X}(T_\omega\mathbf{x})$. Obviously, the quantities $\mathbf{X}$, $\mathbf{x}$, ω and operator T_ω are defined in relation to the ensemble $\mathcal{N}_n$, so that they should be written as $\overset{n}{\mathbf{X}}$, $\overset{n}{\mathbf{x}}$, $\overset{n}{\omega}$, $\overset{n}{T}_\omega$, etc., and we study the corresponding limits of these quantities as $n \to \infty$. However, the superscript n is often neglected solely for simplicity's sake. Hence, $\mathbf{X}$, for example, implies the sequence $\overset{n}{\mathbf{X}}$.

Now, in order that the macroscopic state deserves its name, it is desired that the value $\mathbf{X}(T_\omega \mathbf{x})$ be determined only by the value $\mathbf{X}(\mathbf{x})$ for almost all systems of the ensemble, if the number n is very large. Hence the following condition is required for a macroscopic quantity $\mathbf{X}(\mathbf{x})$ to be a macroscopic state [5]. The condition is related to the representative hypothesis of Rozonoer [9].

Macroscopic state condition.. A quantity $\mathbf{X}(\mathbf{x})$ (or more precisely a sequence of quantities $\overset{n}{\mathbf{X}}(\overset{n}{\mathbf{x}})$) is said to satisfy the macroscopic state condition, if there exist a set of functions $\Phi_1(\mathbf{X}), \cdots, \Phi_r(\mathbf{X})$ such that, for any state $\mathbf{x}$ satisfying $\mathbf{X}(\mathbf{x}) = \mathbf{X}_0$ (or for any sequence $\overset{n}{\mathbf{x}}$ of state such that $\lim_{n\to\infty} \overset{n}{\mathbf{X}}(\overset{n}{\mathbf{x}}) = \mathbf{X}_0$)

$$\lim_{n\to\infty} E[X_\alpha(T_\omega \mathbf{x})] = \Phi_\alpha(\mathbf{X}_0) \tag{2.2}$$

$$\lim_{n\to\infty} V[X_\alpha(T_\omega \mathbf{x})] = 0, \tag{2.3}$$

or in short form

$$\lim_{n\to\infty} E[|X_\alpha(T_\omega \mathbf{x}) - \Phi_\alpha(\mathbf{X})|^2] = 0, \qquad \alpha = 1, 2, \cdots, r, \tag{2.4}$$

where E and V denote, respectively, the operations of taking the expectation and variance with respect to the random variables ω.

When $\mathbf{X}(\mathbf{x})$ satisfies the macroscopic state condition, it seems to be expected that the macroscopic state $\mathbf{X}^t = \mathbf{X}^t(\omega)$ at time t,

$$\mathbf{X}^t = \mathbf{X}(x_1(t), \cdots, x_n(t)) \tag{2.5}$$

satisfies the equation

$$\mathbf{X}^{t+1} = \mathbf{\Phi}(\mathbf{X}^t), \tag{2.6}$$

where $\mathbf{\Phi} = (\Phi_1, \cdots, \Phi_r)$, within an arbitrary small error, if n is sufficiently large. But it is not necessarily true [5], [19]. If we substitute $\mathbf{x}(t) = \mathbf{x}(t; \omega)$, the microscopic state of a net at time t, for $\mathbf{x}$ in (2.4), equality (2.4) does not always hold, because $\mathbf{x}(t) = T_\omega^t \mathbf{x}$ is also a random variable depending on the parameters ω. In other words, $\mathbf{x}(t)$ depends on which net in the ensemble the initial value is given to. This is a very crucial point of the theory. Therefore, we need to prove the following proposition in order that equation (2.6) be regarded as the macroscopic state equation.

Strong proposition. There exists a set of functions $\Phi_1(\mathbf{X}), \cdots, \Phi_r(\mathbf{X})$ such that, if $\tilde{\mathbf{X}}^t$ is the solution of the equation

$$\mathbf{X}^{t+1} = \mathbf{\Phi}(\mathbf{X}^t)$$

with the initial condition $\tilde{\mathbf{X}}^0 = \mathbf{X}(\mathbf{x}(0))$, then the true macroscopic state $\mathbf{X}^t = \mathbf{X}^t(\omega)$ given by (2.5) satisfies

$$\lim_{n\to\infty} \sup_t E[|X_\alpha^t - \tilde{X}_\alpha^t|^2] = 0, \qquad \alpha = 1, 2, \cdots, r. \tag{2.7}$$

The functions Φ_α in the proposition, if they exist, are the same as those in (2.4). As is well known, the mean-square convergence (2.7) implies the convergence in probability

$$\lim_{n\to\infty} \text{Prob}\,\{\sup_t |X_\alpha^t - \tilde{X}_\alpha^t| > \varepsilon\} = 0 \tag{2.7$'$}$$

for an arbitrarily small positive ε. It means that, when n is sufficiently large, the probability measure of those nets for which the macroscopic equation (2.6) holds within an arbitrary small error ε converges to 1. However, as will be seen later, it is very difficult to prove this proposition in most cases of interest. Therefore, we consider the following weak proposition.

Weak proposition. For an arbitrary time t, the macroscopic state satisfies

$$\lim_{n\to\infty} E\{|X_\alpha^t - \tilde{X}_\alpha^t|^2\} = 0. \tag{2.8}$$

If this proposition is satisfied, the macroscopic state equation (2.6) holds up to an arbitrary fixed t within an arbitrarily small error bound, if n is sufficiently large. However, for a large but a fixed number n, there may exist T such that (2.6) does not necessarily hold for $t > T$. The time T depends on n and $T \to \infty$ as n tends to infinity. The strong proposition requires the uniform convergence in t of (2.8) as $n \to \infty$, whereas the weak proposition does not.

If n random variables $x_1(t; \omega), x_2(t; \omega), \cdots, x_n(t; \omega)$ are stochastically independent, we can prove the weak proposition from the macroscopic state condition. They are, however, not independent. Instead of the rigorous independence of these variables, it is useful to consider the asymptotic independence among a finite subset of these variables. This requires that, for arbitrary number k and arbitrary k indices $i_1, i_2, \cdots, i_k$, the k variables $x_{i_1}(t), x_{i_2}(t), \cdots, x_{i_k}(t)$ are independently distributed as n tends to infinity. Kac [14] called this the Boltzmann property, and Rozonoer [19] treated this proposition as the independence hypothesis.

We prove the weak proposition by showing the following proposition of similar character.

Asymptotic independence proposition. We say that the asymptotic independence proposition is satisfied, when, for each n, there exist independently distributed random variables $\tilde{x}_1(t), \cdots, \tilde{x}_n(t)$ which satisfy:

(i) the $\tilde{x}_i(t)$'s are independent of ω and their distribution satisfies

$$\lim_{n\to\infty} E[|X_\alpha(\tilde{\mathbf{x}}(t)) - X_\alpha^t|^2] = 0, \qquad \alpha = 1, 2, \cdots, r.$$

(ii) For an arbitrary finite set of indices $i_1, i_2, \cdots, i_k$, the joint probability distribution of $x_{i_1}(t), \cdots, x_{i_k}(t)$ converges to that of $\tilde{x}_{i_1}(t), \cdots, \tilde{x}_{i_k}(t)$ as n tends to infinity.

Statistical description of a dynamical system has been studied by many people in the field of statistical mechanics (Kac[14]). In the following, we will prove the weak proposition for macroscopic description of various kinds of random nerve nets by proving the asymptotic independence proposition. The strong proposition is proved only for a special class of ensembles.

3. Probability measure on an ensemble of homogeneous random nerve nets. We treat in this section a random net of stochastically homogeneous structure, in which every neuron has the same stochastic property. However, our method can be applied to more general nonhomogeneous nets or complex systems obtained by connecting these homogeneous nets, as will be shown in later sections.

Let us consider a net consisting of mutually connected McCulloch–Pitts formal neurons. The behavior of a McCulloch–Pitts neuron is described as follows: A neuron has at most n input terminals. A quantity called the synaptic weight is associated with each input terminal. Let $x_1, x_2, \cdots, x_n$ be input signals arriving at the input terminals, and let $w_1, w_2, \cdots, w_n$ be the corresponding synaptic weights. These inputs cause a change in the membrane potential of the neuron. The change is given by the weighted sum $\sum w_i x_i$ of the inputs, which we call the postsynaptic potential (PSP). When the PSP exceeds a threshold value h of the neuron, it is excited and emits an output pulse, so that the output x is equal to 1. When the PSP does not exceed h, it is not excited and the output x is equal to 0. Therefore, the input-output relation of a McCulloch–Pitts neuron is described by

$$x = 1\left(\sum_{i=1}^{n} w_i x_i - h\right),$$

where $1(u)$ is the unit step function defined by

$$1(u) = \begin{cases} 1, & u > 0, \\ 0, & u \leqq 0, \end{cases}$$

and the $n+1$ quantities w_i and h characterize the structure of the neuron.

Let us consider a nerve net consisting of n mutually connected McCulloch–Pitts neurons. Let $x_i(t)$ $(i = 1, 2, \cdots, n)$ be the output of the ith neuron at time t. All the neurons are interconnected in such a manner that the output signal of every neuron is fed back to inputs of all the neurons. We assume that the net works synchronously at discrete times, so that the output $x_j(t)$ enters into the jth terminals of all the neurons, becoming their jth input signal at time $t+1$. Let w_{ij} be the synaptic weight of the jth input terminal of the ith neuron. Then, the PSP of the ith neuron becomes at time $t+1$

$$\sum_{j=1}^{n} w_{ij} x_j(t).$$

When the jth output is not connected to the ith neuron, we may simply put $w_{ij} = 0$.

Let h_i be the threshold value of the ith neuron. Then the output $x_i(t+1)$ of the ith neuron is given by the state transition equation

$$x_i(t+1) = 1\left(\sum_{j=1}^{n} w_{ij}x_j(t) - h_i\right), \qquad i = 1, \cdots, n. \tag{3.1}$$

In this case, a net has n^2+n system parameters composed of n^2 synaptic weights w_{ij} $(i, j = 1, \cdots, n)$ and n thresholds h_i $(i = 1, \cdots, n)$, i.e. $\omega = \{w_{ij}, h_i\}$. Thus the structure of the net is completely specified by ω.

Now we introduce a probability measure on an ensemble of nerve nets by considering ω, i.e. the n^2+n parameters w_{ij} and h_i, as random variables. Since we consider the limiting ideal case $n \to \infty$, we need to determine the distribution of the w_{ij}'s in such a manner that the PSP $\sum_{j=1}^{n} w_{ij}x_j$ of a neuron converges to a certain random variable in distribution as $n \to \infty$, with the fraction of the number of excited inputs (i.e., those for which $x_j = 1$) to n kept constant. In this case, the contribution of the weight w_{ij} of a single input to the whole PSP decreases as n tends to infinity, since the number of the excited inputs increases in proportion to n. Thus, the distribution of w_{ij}'s in $\mathcal{N}_n$ depends on the number n of the elements. We denote by $W(\theta)$ the characteristic function of a random variable $\lim_{n\to\infty} \sum_{j=1}^{n} w_{ij}$, which represents the PSP when all the inputs are excited. Now we put the following assumptions on the probability measure of the w_{ij}'s and h_i's in the $\mathcal{N}_n$'s.

Assumption 1. All the weights w_{ij} and thresholds h_i are independently distributed.

Assumption 2. All the w_{ij} in an $\mathcal{N}_n$ are identically distributed. The characteristic function $W_n(\theta)$ of w_{ij} in $\mathcal{N}_n$ is analytic at the origin.

Assumption 3. All the threshold h_i's are identically distributed irrespective of n and have a continuous density function. The characteristic function $H(\theta)$ of h_i is analytic at the origin.

Assumption 4. The characteristic function $W(\theta)$ of the limit distribution of $\sum_{j=1}^{n} w_{ij}$ is analytic at the origin.

Assumption 5.

$$\lim_{n\to\infty} E\left[\sum_{j=1}^{n} |\overset{n}{w}_{ij}|\right] < \infty$$

or equivalently

$$E[|\overset{n}{w}_{ij}|] = O\left(\frac{1}{n}\right) \tag{3.2}$$

where $O(1/n)$ denotes a term of order $1/n$, the superscript n indicating that $\overset{n}{w}_{ij}$ is defined on $\mathcal{N}_n$.

The Assumptions 1, 2 and 3 imply that all the component neurons have the same stochastic properties. Hence, a random net of this ensemble may be said to be stochastically homogeneous under these assumptions. The independence assumption plays a very important role in the following. Since the w_{ij}'s in $\mathcal{N}_n$ are identically and independently distributed, the characteristic function of $\sum_{j=1}^{n} w_{ij}$ is given by $(W_n(\theta))^n$. Therefore, we have

$$\lim_{n\to\infty} (W_n(\theta))^n = W(\theta). \tag{3.3}$$

This shows that $W(\theta)$ must be a characteristic function of an infinitely divisible distribution.

It is easily seen that if the characteristic function of a random variable is analytic at the origin, the Taylor expansion converges uniformly in an interval including the origin, and hence the moment of any order exists. Moreover, the distribution of the random variable is uniquely determined by the series of these moments.

From (3.3), or equivalently from

$$\lim_{n\to\infty} n \log W_n(\theta) = \log W(\theta), \tag{3.4}$$

we see that $W_n(\theta)$ converges to 1 as $n \to \infty$. If we put

$$W_n(\theta) = 1 + \frac{1}{n} U_n(\theta), \qquad U_n(0) = 0, \tag{3.5}$$

we have

$$\lim_{n\to\infty} U_n(\theta) = \log W(\theta).$$

From (3.5), we can easily prove that

$$E[(w_{ij})^k] = O\left(\frac{1}{n}\right), \qquad k = 1, 2, 3, \cdots. \tag{3.6}$$

Moreover, we obtain

$$E[|w_{ij}|^k] = O\left(\frac{1}{n}\right), \qquad k = 1, 2, 3, \cdots \tag{3.7}$$

by the use of Schwarz' inequality.

It is known (Doob [10]) that the characteristic function $W(\theta)$ of an infinitely divisible distribution has the following Levy–Khinchin representation,

$$\log W(\theta) = i\mu\theta + \int_{-\infty}^{\infty} \left(e^{iz\theta} - 1 - \frac{iz\theta}{1+z^2}\right) \frac{z^2+1}{z^2} dG(z),$$

where $G(z)$ is a bounded monotonically nondecreasing function. We show some typical examples of infinitely divisible distributions. When $G(z) = 0$, we have

$$\log W(\theta) = i\mu\theta,$$

which represents a constant distribution. Put

$$G(z) = \begin{cases} 0, & z < c, \\ \sigma^2 & z \geqq c. \end{cases}$$

When $c \neq 0$, putting $\mu = 1/c$, we obtain a Poisson distribution

$$\log W(\theta) = \sigma^2 \frac{1+c^2}{c^2}(e^{ic\theta} - 1).$$

In the case of $c = 0$, we obtain a normal distribution

$$\log W(\theta) = i\mu\theta - \frac{\sigma^2}{2}\theta.$$

Thus the Levy–Khinchin representation shows that an infinitely divisible distribution can be decomposed into the convolution of a constant distribution, a normal distribution and an infinitely large aggregate of Poisson distributions.

Among the above three typical distributions, however, the normal distribution $N(\mu, \sigma^2)$ does not satisfy Assumption 5, because, if we put $W_n(\theta) = (W(\theta))^{1/n}$, w_{ij} is also normally distributed, $N(\mu/n, \sigma^2/n)$, and we have

$$E[|w_{ij}|] = O\left(\frac{1}{\sqrt{n}}\right),$$

which does not satisfy (3.2).

We now show some examples of distribution functions satisfying the assumptions.

Example 1. Poisson distribution. Consider a case in which every w_{ij} takes only on two values 1 and 0 such that

$$w_{ij} = \begin{cases} 1 & \text{with probability } \dfrac{\mu}{n}, \\ 0 & \text{with probability } 1 - \dfrac{\mu}{n} \end{cases}$$

where μ is a fixed constant.

Example 2. Γ-distribution. Consider a case in which w_{ij} is subject to a distribution whose density function has the form

$$p_n(w) = \frac{w^{p/n-1} e^{-w/\sigma}}{\Gamma(p/n)\sigma^{p/n}}, \qquad \sigma > 0, \quad p > 0,$$

where $w \geqq 0$.

In the case of Example 1, all but a finite number of the w_{ij}'s are exactly zero with probability 1, even when n tends to infinity. Each element is connected with only μ other elements on the average. The values of nonzero w_{ij}'s are always 1 and do not decrease as $n \to \infty$. In the case of Example 2, on the other hand, almost all the w_{ij}'s are nonzero and hence each element is connected with infinitely many elements as n goes to infinity, whereas w_{ij} becomes infinitesimally small as $n \to \infty$. We can say that a random nerve net with a Poisson distribution is sparsely connected, and that a net with Γ-distribution is densely connected.

It should be noted that when w_{ij} has a distribution satisfying the assumptions, so does

$$w'_{ij} = \varepsilon_{ij} w_{ij},$$

where the ε_{ij}'s are independently and identically distributed random variables, independent of n and having a characteristic function analytic at the origin. Therefore, various distributions are derived from those in the above examples. For example, if we replace w_{ij} in Example 1 by $\varepsilon_{ij}w_{ij}$, $W(\theta)$ becomes a compound Poisson distribution. In this case, each element is also connected with μ other elements on the average, but their weights are not unity but randomly distributed.

Lastly, we show a family of rather trivial distributions.

Example 3.

$$w_{ij} = \begin{cases} \dfrac{\varepsilon_{ij}}{n^{\alpha}} & \text{with probability } \dfrac{1}{n^{1-\alpha}}, \\ 0 & \text{with probability } 1 - \dfrac{1}{n^{1-\alpha}} \end{cases}$$

where α $(0 < \alpha \leqq 1)$ is a constant and the ε_{ij}'s are independently and identically distributed random variables as before. The characteristic function of $\sum_{j=1}^{n} w_{ij}$ converges to

$$W(\theta) = e^{iw\theta},$$

where w is the expectation of ε_{ij}. This case is rather trivial in the sense that $\sum_{j=1}^{n} w_{ij}$ converges to a constant w in distribution, i.e.,

$$\lim_{n\to\infty} V\left[\sum_{j=1}^{n} w_{ij}\right] = 0.$$

Rotenberg [18] proved the weak proposition in the case with $\alpha = 1$. We prove the strong proposition for any $0 < \alpha \leqq 1$.

Since the normal distribution does not satisfy the assumptions, even the weak proposition cannot be proved in this case by our method. However, when all of the w_{ij}'s are nonnegative or nonpositive, the assumption (3.2) is automatically satisfied, so that our method is applicable. It is known as Dale's law that the sign of a synaptic weight is determined uniquely by the type of the presynaptic neurons. This implies that the synaptic weights stemming from one and the same type of neurons have the same sign. Our method is applicable not only to a homogeneous net but also to a complex net composed of various types of excitatory and inhibitory neurons. Hence, Assumption 5, though it excludes the normal distribution, does not impose any serious restriction. It is only from the theoretical point of view that we have an interest in proving whether the weak proposition holds or not in the case of the normal distribution.

4. Macroscopic equation of a homogeneous random nerve net. The activity level of a nerve net is defined by

$$X(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{4.1}$$

It is a quantity representing the fraction of excited neurons in state $\mathbf{x} = (x_1, \cdots, x_n)$. This activity level satisfies the conditions (2.2) and (2.3). In fact, we have

THEOREM 1. *The activity level $X(\mathbf{x})$ satisfies the macroscopic state condition: For any state* $\mathbf{x}$ *such that* $X(\mathbf{x}) = X^0$

$$\lim_{n\to\infty} E[X(T_\omega \mathbf{x})] = \Phi(X^0), \tag{4.2}$$

$$\lim_{n\to\infty} V[X(T_\omega \mathbf{x})] = 0 \tag{4.3}$$

where

$$\Phi(X) = \frac{1}{2\pi}\int_0^\infty \int_{-\infty}^\infty e^{-iu\theta}(W(\theta))^X H(-\theta)\, d\theta\, du. \tag{4.4}$$

Proof. We have

$$E[X(T_\omega \mathbf{x})] = \frac{1}{n}\sum_{i=1}^n E\Big[1\Big(\sum_{j=1}^n w_{ij}x_j - h_i\Big)\Big]$$

$$= \frac{1}{n}\sum_{i=1}^n E\Big[1\Big(\sum_{j=1}^{nX^0} w_{ij} - h_i\Big)\Big]$$

$$= \text{Prob}\Big\{\sum_{j=1}^{nX^0} w_{ij} - h_i > 0\Big\}.$$

The characteristic function of $\sum_{j=1}^{nX^0} w_{ij} - h_i$ is $(W_n(\theta))^{nX^0} H(-\theta)$ which converges to $(W(\theta))^{X^0} H(-\theta)$ as $n \to \infty$. Thus, using the inversion formula, we obtain (4.2). On the other hand, each $1[\sum_{j=1}^n w_{ij}x_j - h_i]$ is independently and identically distributed with a finite variance, and hence

$$V[X(T\omega \mathbf{x})] = \frac{1}{n^2}\sum_{i=1}^n V\Big[1\Big(\sum_{j=1}^n w_{ij}x_j - h_i\Big)\Big] \to 0.$$

It is expected that the equation

$$X^{t+1} = \Phi(X^t) \tag{4.5}$$

well describes the behavior of the activity level in a random net. The characteristics of the above equation, especially its dependence on the macroscopic parameters specifying the stochastic structure of the net, is investigated in detail [2] under the normal approximation. The weak proposition is proved in the next section.

Assume that $\tilde{x}_i$'s $(i = 1, \cdots, n)$ are independently and identically distributed random variables taking on two values, 1 and 0. Moreover, we assume that the $\tilde{x}_i$'s are independent of ω, and that their probabilities are specified by

$$\tilde{x}_i = \begin{cases} 1 & \text{with probability } X, \\ 0 & \text{with probability } 1 - X. \end{cases}$$

Then we have

$$E[\tilde{x}_i] = X.$$

Furthermore, we can easily prove that $X(\tilde{\mathbf{x}})$ converges to X in the mean-square sense,

$$\lim_{n\to\infty} E[|X(\tilde{\mathbf{x}})-X|^2]=0.$$

Therefore, we may regard $\tilde{\mathbf{x}}$ as a typical microscopic state whose macroscopic state is X.

We have in this connection the following lemma.

LEMMA 1. *The random variable*

$$\tilde{u}_i=\sum_{j=1}^{n} w_{ij}\tilde{x}_j-h_i$$

converges in distribution to a random variable v_i whose characteristic function is $(W(\theta))^X H(-\theta)$ and hence

$$\text{Prob}\,\{v_i>0\}=\Phi(X),$$

where $X=E[\tilde{x}_i]$.

Proof. The characteristic function of $w_{ij}\tilde{x}_j$ is given by

$$E[e_{ij}^{iw\tilde{x}_j}]=(1-X)+XW_n(\theta)=1+\frac{X}{n}U_n(\theta),$$

and then that of $\sum_{j=1}^{n} w_{ij}\tilde{x}_j-h_i$ is given by

$$\left(1+\frac{X}{n}U_n(\theta)\right)^n H(-\theta),$$

which converges to $(W(\theta))^X H(-\theta)$.

5. Proof of the weak proposition. In the present section, the weak proposition is proved for the activity level of the homogeneous random nerve net. The method used in the proof is also applicable to more general nerve nets as will be shown later. The proof owes to the moment theorem due to Kendall and Rao [15], which states that the convergence of a sequence of distribution functions is assured by the convergence of the sequences of the corresponding moments of all orders. We state the theorem in the form convenient to us.

MOMENT THEOREM. *Let $\mathbf{Y}_1, \mathbf{Y}_2, \cdots$ be a sequence of random vector variables having s components*

$$\mathbf{Y}_n=(Y_{n,1}, Y_{n,2}, \cdots, Y_{n,s}), \qquad n=1,2,\cdots.$$

Let

$$A=(a_1,a_2,\cdots,a_s)$$

be a set of s nonnegative integers, and let $\mu_{n,A}$ be the moment of $\mathbf{Y}_n$ defined by

$$\mu_{n,A}=E[(Y_{n,1})^{a_1}(Y_{n,2})^{a_2}\cdots(Y_{n,s})^{a_s}].$$

If (i) *$\mu_{n,A}$ exists for all n and A,* (ii) *all the $\mu_{n,A}$'s converge to μ_A*

$$\lim_{n\to\infty}\mu_{n,A}=\mu_A$$

and moreover (iii) *all these μ_A's uniquely determine a distribution function $F(\mathbf{Y})$ of random variable $\mathbf{Y}$, then $\mathbf{Y}_n$ converges to $\mathbf{Y}$ in distribution, i.e. the distribution function $F_n(\mathbf{Y})$ of $\mathbf{Y}_n$ converges to $F(\mathbf{Y})$ at every point of continuity.*

If the distribution function $F_n(\mathbf{Y}; \mathbf{z})$ of $\mathbf{Y}_n$ contains a set of parameters $\mathbf{z}$, and if the corresponding moment $\mu_{n,A}(\mathbf{z})$ converges to μ_A for all A uniformly in $\mathbf{z}$, then $F_n(\mathbf{Y}; \mathbf{z})$ converges to $F(\mathbf{Y})$ at every point of continuity uniformly in $\mathbf{z}$.

The former part of the theorem is well known in probability theory ([25]). The latter part is a direct consequence of the former (Appendix A).

Let $\tilde{X}^t$ be the solution of the equation

$$\tilde{X}^{t+1} = \Phi(\tilde{X}^t)$$

with the initial condition $\tilde{X}^0 = X^0$. Note that $\tilde{X}^t$ is not a random variable, while

$$X^t = X(\mathbf{x}(t))$$

is a random variable depending on all the w_{ij}'s and h_i's.

Now we introduce random variables $\tilde{x}_i(t)$, $i = 1, 2, \cdots, n$, which are independently and identically distributed, independent of the w_{ij}'s and h_i's and have the following distribution.

$$\tilde{x}_i(t) = \begin{cases} 1 & \text{with probability } \tilde{X}^t, \\ 0 & \text{with probability } 1 - \tilde{X}^t. \end{cases}$$

Let us put

$$\tilde{u}_i(t) = \sum_{j=1}^{n} w_{ij}\tilde{x}_j(t-1) - h_i,$$

$$u_i(t) = \sum_{j=1}^{n} w_{ij}x_j(t-1) - h_i,$$

the latter of which is the true PSP minus the threshold of the ith neuron at time t. We have already seen from Lemma 1 that $\tilde{u}_i(t)$ converges to a random variable $v_i(t)$ whose characteristic function is given by $(W(\theta))^{X^{t-1}}H(-\theta)$.

The following notations will simplify the description:

$$x_I = \{x_i \mid i \in I\},$$

$$h_I = \{h_i \mid i \in I\}, \text{ etc.,}$$

where I is a finite set of integers. By $|I|$, we denote the number of elements in I, i.e. if $I = \{i_1, \cdots, i_p\}$, then $|I| = p$. Let $A = (a_1, \cdots, a_p)$ be a p-tuple of nonnegative integers. We introduce the following abbreviation:

$$x_I^A = (x_{i_1})^{a_1}(x_{i_2})^{a_2} \cdots (x_{i_p})^{a_p}.$$

The set A is called a power set corresponding to the index set I. We denote by $|A|$ the number of nonzero elements in A.

Similar abbreviation is also used for w_{ij} as follows. Let K be a finite integer set. We associate a finite integer set L_i with every integer $i \in K$,

$$L_i = \{l_{i,1}, l_{i,2}, \cdots, l_{i,p_i}\},$$

where

$$p_i = |L_i|.$$

Then, we write as

$$W_{KL} = \{w_{ij} | i \in K, j \in L_i\},$$

where L is the family of integer sets L_i $(i \in K)$. Let

$$C_i = (c_{i1}, c_{i2}, \cdots, c_{ip_i})$$

be a power set corresponding to L_i, and let C be the family of those power sets C_i $(i \in K)$. Then, we use the following abbreviation:

$$W_{KL}^C = \prod_{i \in K} (w_{iL_i})^{C_i} = \prod_{i \in k} \prod_{j \in L_i} (w_{ij})^{c_{ij}},$$

$$|C| = \sum_{i \in K} |C_i|$$

denotes the number of nonzero c_{ij}'s.

Now the weak proposition is proved by verifying the following fundamental lemmas (Lemmas 2, 3, and 4).

LEMMA 2. *The asymptotic independence proposition is satisfied, and any finite set of random variables $x_I(t)$ converges to $\tilde{x}_I(t)$ in distribution as $n \to \infty$.*

LEMMA 3. *For any finite index sets I, J, K and a family of finite integer sets L such that*

$$I \cap J = \varnothing, \qquad I \cap K = \varnothing,$$

where $\varnothing$ is the empty set, the conditional probability distribution of $(u_I(t), h_J)$ under the condition $w_{KL} = z_{KL}$ converges to that of $(v_I(t), h_J)$ uniformly in z_{KL} as $n \to \infty$, where v_i's are defined in Lemma 1, i.e.,

$$\lim_{n \to \infty} \text{Prob}\,\{u_I(t), h_J | w_{KL} = z_{KL}\} = \text{Prob}\,\{v_I(t)\} \cdot \text{Prob}\,\{h_J\}.$$

LEMMA 4. *For any index set I, J, K, M, N and a family of finite integer sets L such that*

$$I \cap (J \cup K) = \varnothing, \qquad M \cap (I \cup J \cup K) = \varnothing,$$

and for arbitrary corresponding power sets A, B and a family of power sets C, we have

$$\lim_{n \to \infty} n^{|C|}(E[x_I^A(t) h_J^B w_{KL}^C | w_{MN} = z_{MN}] - E[\tilde{x}_I^A(t) h_J^B w_{KL}^C]) = 0$$

uniformly in z_{MN}, where $E[\,\cdot\,|\,\cdot\,]$ denotes the conditional expectation.

Lemma 2 is a direct consequence of Lemma 3, since $u_i(t)$ has a continuous distribution. The proof of Lemmas 3 and 4 are given in Appendix B. Lemma 2 shows that the independence hypothesis of Rozonoer or the Boltzmann property of Kac holds for this net.

We now prove the weak proposition.

THEOREM 2.

$$\lim_{n\to\infty} E[|X^t - \tilde{X}^t|^2] = 0.$$

Proof. From Lemma 2,

$$\lim_{n\to\infty} E[X^t] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E[x_i(t)]$$
$$= E[\tilde{x}_i(t)] = \tilde{X}^t.$$

Moreover, from

$$\lim_{n\to\infty} \mathrm{Cov}\,[x_i(t), x_j(t)] = \mathrm{Cov}\,[\tilde{x}_i(t), \tilde{x}_j(t)] = 0,$$

we have

$$\lim_{n\to\infty} V[X^t] = 0,$$

where $\mathrm{Cov}\,[x, y]$ is the covariance between x and y.

6. Macroscopic state equations of various types of random nerve nets.

6.1. Nonhomogeneous random nets. Many types of neurons are mutually interconnected in the brain. Therefore, we need to treat nonhomogeneous random nets or systems composed of a number of different homogeneous subnets. We can derive the macroscopic state equations in this case, too.

Let us consider a random net consisting of m types of McCulloch–Pitts neurons with different stochastic structures. This net can be regarded as a system composed of m subnets, each subnet being a homogeneous random net. Let the number of neurons of type α, $\alpha = 1, \cdots, m$, be n_α. We consider the case in which the number $n = \sum_\alpha n_\alpha$ of all the neurons increases in such a manner that the ratios n_α/n, $\alpha = 1, \cdots, m$, are always fixed. Let the output and the threshold of the ith neuron of type α be denoted by x_i^α and h_i^α, $i = 1, \cdots, n_\alpha$, respectively. Let $w_{ij}^{\alpha\beta}$ be the synaptic weight of the ith neuron of type α, with which weight the output of the jth neuron of type β is connected to the former element. Then, we obtain the following state transition equation:

$$x_i^\alpha(t+1) = 1\left(\sum_{\beta=1}^{m} \sum_{j=1}^{n_\beta} w_{ij}^{\alpha\beta} x_j^\beta(t) - h_i^\alpha\right). \tag{6.1}$$

The synaptic weight $w_{ij}^{\alpha\beta}$ is a random variable whose distribution depends on the types β and α of the presynaptic and postsynaptic neurons. Therefore, we assume that for each pair (α, β), the $w_{ij}^{\alpha\beta}$'s are independently and identically distributed. We denote their characteristic function by $W_n^{\alpha\beta}(\theta)$. The distribution is assumed to satisfy the assumptions of § 3. We denote the limit characteristic function of a random variable $\sum_{j=1}^{n_\beta} w_{ij}^{\alpha\beta}$ by $W^{\alpha\beta}(\theta)$.

We can choose the m activity levels X_α corresponding to each of the m types of neurons,

$$X_\alpha(\mathbf{x}) = \frac{1}{n_\alpha} \sum_{i=1}^{n_\alpha} x_i^\alpha, \qquad \alpha = 1, \cdots, m, \tag{6.2}$$

as macroscopic quantities characterizing the state of a net. Let $\mathbf{X}^t = (X_1^t, X_2^t, \cdots, X_m^t)$ be the macroscopic state at time t defined by

$$X_\alpha^t = X_\alpha(\mathbf{x}(t)).$$

Then, we can prove the following theorem.

THEOREM 3. *The activity levels* $\mathbf{X}(\mathbf{x})$ *satisfy the macroscopic state condition, with the macroscopic state transition function* $\mathbf{\Phi} = (\Phi_1, \cdots, \Phi_m)$,

$$\Phi_\alpha(\mathbf{X}) = \int_0^\infty \int_{-\infty}^\infty \prod_{\beta=1}^m (W^{\alpha\beta}(\theta))^{X\beta} H(-\theta)\, e^{-iu\theta}\, d\theta\, du. \tag{6.3}$$

The macroscopic state equation

$$\mathbf{X}^{t+1} = \mathbf{\Phi}(\mathbf{X}^t) \tag{6.4}$$

satisfies the weak proposition.

The proofs are obtained by the arguments similar to those in previous sections. Some characteristics of the above macroscopic state equations are analyzed by Amari [2]. Models of association and concept formations are proposed in [2], [4], [6] by the use of nonhomogeneous random nets.

6.2. Random nets consisting of neurons with refractory. A McCulloch–Pitts formal neuron is said to have absolute refractory of period r, if it cannot be excited in the successive r times after its excitation. We treat a homogeneous random net consisting of formal neurons with absolute refractory, where random variables w_{ij} and h_i are considered to satisfy the assumptions of § 3.

In this net, the state $\mathbf{x}(t+1)$ at time $t+1$ cannot be determined by $\mathbf{x}(t)$ alone. It depends on the previous r successive states $\mathbf{x}(t), \mathbf{x}(t-1), \cdots, \mathbf{x}(t-r+1)$. The state transition equation is written as

$$x_i(t+1) = \begin{cases} 0 & \text{if } \sum_{k=0}^{r-1} x_i(t-k) \neq 0, \\ 1\left(\sum_{j=1}^n w_{ij} x_j(t) - h_i\right) & \text{otherwise.} \end{cases}$$

We treat the activity level

$$X^t = \frac{1}{n} \sum_{i=1}^n x_i(t)$$

as a macroscopic quantity describing the state of the net at time t. Let $\mathbf{x}(0), \mathbf{x}(1), \cdots, \mathbf{x}(r-1)$ be the initial r successive states satisfying

$$X(\mathbf{x}(k)) = X^k, \qquad k = 0, 1, \cdots, r-1.$$

We assume, for the sake of consistency, that

$$\sum_{k=0}^{r-1} x_i(k)=0 \quad \text{or} \quad 1.$$

Then we can prove the following theorem.

THEOREM 4. *Let $X^r = X(\mathbf{x}(r))$. Then*

$$\lim_{n\to\infty} E[X^r]=\Big(1-\sum_{k=0}^{r-1} X^k\Big)\Phi(X^{r-1}),$$

$$\lim_{n\to\infty} V[X^r]=0$$

where $\Phi(X)$ is the function given by (4.4).

Let $\mathbf{X}^t$ denote the following r dimensional vector composed of r successive activities:

$$\mathbf{X}^t=(X^{t-r+1}, X^{t-r+2}, \cdots, X^t).$$

Then, Theorem 4 implies that the macroscopic state condition is satisfied for this vector quantity. Therefore, it is expected that $X^t = X(\mathbf{x}(t))$ satisfies the following difference equation of order r,

$$X^{t+1}=\Big(1-\sum_{k=0}^{r-1} X^{t-k}\Big)\Phi(X^t). \tag{6.5}$$

The behavior of this equation is studied by Yoshizawa [26].

Let the solution of the above equation be $\tilde{X}^t$, with the initial conditions $\tilde{X}^k = X^k$, $k=0, 1, \cdots, r-1$. Then we can prove the asymptotic independence proposition, which leads to the weak proposition.

THEOREM 5. *The weak proposition holds for the macroscopic state equation of the net with absolute refractory,*

$$\lim_{n\to\infty} E[|X^t-\tilde{X}^t|^2]=0. \tag{6.6}$$

6.3. Random nets of multi-threshold threshold elements and analogue neurons. The McCulloch–Pitts formal neuron can easily be generalized to a multi-threshold threshold element whose output takes $s+1$ different values $d_0, d_1, \cdots, d_s$. The real axis R^1 is divided into $s+1$ mutually disjoint intervals $D_0, D_1, \cdots, D_s$. The output of each element is d_α, if the PSP minus its threshold is included in the interval D_α. Thus, we obtain the following microscopic state transition function:

$$x_i(t+1)=\chi\Big(\sum_{j=1}^{n} w_{ij}x_j(t)-h_i\Big), \tag{6.7}$$

where $\chi(u)$ is a multi-threshold function of the form

$$\chi(u)=d_\alpha, \quad \text{when } u \in D_\alpha \quad (\alpha=0, 1, \cdots, s).$$

If the function χ is replaced by a continuous function in (6.7), the element is called an analog neuron element, which is considered to be the limiting case of the multi-threshold threshold element as $s\to\infty$.

Now let us consider the following macroscopic quantities [18]

$$X_\alpha(\mathbf{x})=\frac{1}{n}\sum_{i=1}^{n}\delta_\alpha(x_i),\qquad \alpha=0,1,\cdots,s, \tag{6.8}$$

where δ_α is defined by

$$\delta_\alpha(u)=\begin{cases}1 & u=d_\alpha,\\ 0 & \text{otherwise.}\end{cases}$$

Then $X_\alpha(\mathbf{x})$ expresses the fraction of the neuron elements whose output is d_α. We write $\mathbf{X}=(X_0,\cdots,X_s)$. The quantities $X_0,\cdots,X_s$, are called the occupation rates. Note that only s of them are algebraically independent, since we have

$$\sum_{\alpha=0}^{s}X_\alpha(\mathbf{x})=1.$$

We can derive the macroscopic state equation of this net and prove the weak proposition. The proof is omitted.

THEOREM 6. *The macroscopic quantity* $\mathbf{X}(\mathbf{x})$ *satisfies the macroscopic state condition with the macroscopic state transition function* $\mathbf{\Psi}=(\Psi_0,\cdots,\Psi_s)$.

$$\Psi_\alpha(\mathbf{X})=\int_{D_\alpha}\int_{-\infty}^{\infty}(W(d_0\theta))^{X_0}\cdots(W(d_s\theta))^{X_s}H(-\theta)\,e^{-iu\theta}\,d\theta\,du, \tag{6.9}$$

$$\alpha=0,\cdots,s.$$

The macroscopic state $\mathbf{X}^t$ *satisfies the weak proposition.*

The occupation rates are not the only macroscopic quantities characterizing the net composed of multi-threshold threshold elements. An alternative choice is the following quantities:

$$Q_\alpha(\mathbf{x})=\frac{1}{n}\sum_{i=1}^{n}(x_i)^\alpha,\qquad \alpha=1,2,\cdots,s.$$

When the distribution of the sum $\sum_{j=1}^n w_{ij}$ can be approximated by a normal distribution, the distribution of the sum $\sum_{j=1}^n w_{ij}x_j(t)$ is approximated by a normal distribution. In this case, only two macroscopic quantities Q_1 and Q_2 are sufficient to constitute the macroscopic state. The approximated macroscopic equations are

$$Q_\alpha^{t+1}=\eta_\alpha(Q_1^t,Q_2^t),\qquad \alpha=1,2,$$

where $\bar{w}=\lim_{n\to\infty}E[nw_{ij}]$,

$$\eta_\alpha(Q_1,Q_2)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\chi(u))^\alpha e^{-i\theta u}$$

$$\cdot\exp\{i\bar{w}(Q_1-\sqrt{Q_2})\theta\}W(\sqrt{Q_2}\theta)H(-\theta)\,d\theta\,du,\qquad \alpha=1,2.$$

The same argument also holds for a net consisting of analogue neuron elements, where the multi-threshold function $\chi(u)$ is replaced by a continuous function. Behaviors of random nets of analog neurons have been investigated in [3], [9].

7. Macroscopic description of state correlations. A state of a McCulloch–Pitts neuron net is represented by a vector whose n components consist of 0 and 1. There are 2^n states all together. A sequence of states $\mathbf{x}, T_\omega\mathbf{x}, T_\omega^2\mathbf{x}, \cdots, T_\omega^k\mathbf{x}$ is called a cycle of period k, when k is the least positive integer satisfying $T_\omega^k\mathbf{x}=\mathbf{x}$. Every infinite sequence of state transition

$$\mathbf{x}, T_\omega\mathbf{x}, T_\omega^2\mathbf{x}, \cdots$$

falls into a cycle after a finite number of transient periods. It is both interesting and meaningful to study the characteristic features of the state transition diagram, e.g., the number of cycles, the periods of cycles, the average length of transient periods, etc. Amari [5] tried to attack this problem, investigating the macroscopic law which gives the distance (in the sense of Hamming) between two states $T_\omega\mathbf{x}$ and $T_\omega\mathbf{y}$ in terms of the distance between $\mathbf{x}$ and $\mathbf{y}$. Rozonoer [21] also treated the problem in the special case of Example 3 in § 3. The distance of two states is considered as one of the quantities that describe the correlation of states. Here we will generalize the method so that we can treat the dynamics of correlation of many states.

Let us consider a set of r states $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_r$ and let $R=\{1, 2, \cdots, r\}$ be the set of indices of the states. Let S be a subset of R. There are 2^r subsets of R. The set of all the subsets of R is denoted by 2^R. We define the following quantity related to a subset $S=\{\alpha_1, \alpha_2, \cdots, \alpha_p\}$ of R:

$$Y_S(\mathbf{x}_1, \cdots, \mathbf{x}_r)=\frac{1}{n}\sum_{i=1}^{n} x_{\alpha_1,i}\cdots x_{\alpha_p,i}(1-x_{\beta_1,i})\cdots(1-x_{\beta_q,i}), \tag{7.1}$$

where $x_{\alpha,i}$ is the ith component of $\mathbf{x}_\alpha$, and $\bar{S}=\{\beta_1, \beta_2, \cdots, \beta_q\}$ is the complement of S, $\bar{S}=R-S$.

The meaning of the quantity Y_S becomes clear, by considering r random $\mathbf{x}_\alpha$'s. We assume that the ith components $x_{1,i}, x_{2,i}, \cdots, x_{r,i}$ of $\mathbf{x}_1, \cdots, \mathbf{x}_r$ are mutually correlated. Moreover, we assume that their joint distribution is the same for all i's $(i=1, \cdots, n)$ and they are independently distributed for different i. We associate with the set S the probability

$$P(S)=\text{Prob}\,\{x_{\alpha_1,i}=1, \cdots, x_{\alpha_p,i}=1;\ x_{\beta_1,i}=0, \cdots, x_{\beta_q,i}=0\}. \tag{7.2}$$

By virtue of the law of large numbers, we have

$$\lim_{n\to\infty} E[|Y_S(\mathbf{x}_1, \cdots, \mathbf{x}_r)-P(S)|^2]=0.$$

This shows that, if the $\mathbf{x}_\alpha$'s are random, Y_S gives a good approximation to the probability $P(S)$. The set of all the 2^r quantities $P(S)$ $(S\in 2^R)$ completely determines the correlational structure of $\mathbf{x}_1, \cdots, \mathbf{x}_r$. There are many kinds of quantities representing various types of correlations (e.g., pairwise correlation, triple correlation, and so on), and their structures are investigated by Han [12] from the information-theoretical point of view. Since all of the correlational quantities can be calculated from the $P(S)$'s or Y_S's, we treat, in the present paper, the Y_S's as the quantities representing the various kinds of correlations among the r states. We call the 2^r-dimensional vector

(7.3) $$\mathbf{Y}(\mathbf{x}_1, \cdots, \mathbf{x}_r) = (Y_S(\mathbf{x}_1, \cdots, \mathbf{x}_r), S \in 2^R)$$

the correlation vector of $\mathbf{x}_1, \cdots, \mathbf{x}_r$.

Let $\mathbf{Y}$ be the correlation vector of r states $\mathbf{x}_1, \cdots, \mathbf{x}_r$. Then how are their next states $T_\omega \mathbf{x}_1, T_\omega \mathbf{x}_2, \cdots, T_\omega \mathbf{x}_r$ correlated? Is there any law by which $\mathbf{Y}(\mathbf{x}_1, \cdots, \mathbf{x}_r)$ is related to $\mathbf{Y}(T_\omega \mathbf{x}_1, \cdots, T_\omega \mathbf{x}_r)$? More generally, put

$$\mathbf{x}_\alpha(t) = T_\omega \mathbf{x}_\alpha(t-1) = T_\omega^t \mathbf{x}_\alpha(0),$$

and

$$\mathbf{Y}^t = \mathbf{Y}(\mathbf{x}_1(t), \cdots, \mathbf{x}_r(t)).$$

By treating the quantity $\mathbf{Y}$ as a macroscopic state in the generalized sense, we want to obtain, if any, the common dynamical law by which $\mathbf{Y}^{t+1}$ is determined from $\mathbf{Y}^t$ in almost all nets in the ensemble.

We can prove that $\mathbf{Y}(\mathbf{x}_1, \cdots, \mathbf{x}_r)$ satisfies the macroscopic state condition and the dynamics satisfies the weak proposition.

THEOREM 7.

$$\lim_{n\to\infty} E[|Y_S(T_\omega \mathbf{x}_1, \cdots, T_\omega \mathbf{x}_r) - \Phi_S(\mathbf{Y})|^2] = 0,$$

where $\Phi_S(\mathbf{Y})$ *is given by*

(7.4) $$\Phi_S(\mathbf{Y}) = \frac{1}{(2\pi)^r} \int_{\mathscr{D}_S} du_1 \cdots du_r \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} d\theta_1 \cdots d\theta_r \cdot \exp\left\{-i \sum_{\alpha=1}^{r} u_\alpha \theta_\alpha\right\} \cdot \prod_{Q \in 2^R} \left(W\left(\sum_{\alpha \in Q} \theta_\alpha\right)\right)^{Y_Q} H\left(-\sum_{\alpha=1}^{r} \theta_\alpha\right),$$

(7.5) $$\mathscr{D}_S = \{(u_1, \cdots, u_r) | u_\alpha > 0, \alpha \in S; u_\beta \leqq 0, \beta \in \bar{S}\}.$$

The correlation $\mathbf{Y}^t$ *satisfies the weak proposition.*

We have so far considered the dynamics with r initial states $\mathbf{x}_1(0), \cdots, \mathbf{x}_r(0)$. Now we consider a state transition sequence

$$\mathbf{x}(0), \quad \mathbf{x}(1) = T_\omega \mathbf{x}(0), \quad \mathbf{x}(2) = T_\omega^2 \mathbf{x}(0), \cdots$$

beginning with a single initial state $\mathbf{x}(0)$. We want to know the correlation

(7.6) $$\mathbf{Y}(t) = \mathbf{Y}(\mathbf{x}(t), \mathbf{x}(t+1), \cdots, \mathbf{x}(t+r-1))$$

of the r successive states $\mathbf{x}(t), \cdots, \mathbf{x}(t+r-1)$ in this sequence. If we define r states $\mathbf{x}_1(0), \cdots, \mathbf{x}_r(0)$ by

(7.7) $$\mathbf{x}_\alpha(0) = T_\omega^{\alpha-1} \mathbf{x}(0) = \mathbf{x}(\alpha - 1), \qquad \alpha = 1, \cdots, r,$$

then the correlation $\mathbf{Y}(t)$ can be written as

$$\mathbf{Y}(t) = \mathbf{Y}(T_\omega^t \mathbf{x}_1(0), \cdots, T_\omega^t \mathbf{x}_r(0)).$$

Therefore, it is expected that $\mathbf{Y}(t)$ also satisfies the macroscopic state equation

$$\mathbf{Y}(t+1)=\mathbf{\Phi}(\mathbf{Y}(t)). \tag{7.8}$$

However, in this case, the r initial states $\mathbf{x}_1(0), \cdots, \mathbf{x}_r(0)$ defined by (7.7) are not independent of the w_{ij}'s and h_i's. Hence, Theorem 7 cannot be applied to this case. However, we can also prove the weak proposition via the asymptotic independence proposition.

THEOREM 8.

$$\lim_{n\to\infty} E[|Y_S(t)-\tilde{Y}_S(t)|^2]=0.$$

We note lastly that, if the strong proposition holds for (7.8), then $\mathbf{Y}(t)$ represents the correlation among cyclic states in the state transition diagram, for sufficiently large t, because all the states $\mathbf{x}(t), \mathbf{x}(t+1), \cdots$ fall into cyclic states for large t. This suggests that the structure of the state transition diagram of a nerve net can be studied by the present approach (see Amari [5]).

8. Proof of the strong proposition. We prove the strong proposition for the dynamics of the activity level of homogeneous random nets of McCulloch–Pitts neurons, in the case of the distributions shown in Example 3 of § 3. In this special case, the macroscopic state transition function (4.4) reduces to

$$\Phi(X)=F(wX), \tag{8.1}$$

where F is the distribution function of the threshold h_i (i.e., the inverse Fourier transform of H).

We first study the property of the macroscopic state equation

$$X^{t+1}=\Phi(X^t). \tag{8.2}$$

When $w \geqq 0$, the function $\Phi(X)$ is continuous and monotonically nondecreasing. Since $\Phi(X)$ is bounded, every solution $\tilde{X}^0, \tilde{X}^1, \tilde{X}^2, \cdots$ is convergent. It converges to an equilibrium X^* satisfying

$$X^*=\Phi(X^*).$$

An equilibrium X^* is said to be stable when $\Phi(X)$ is a contraction in a neighborhood of X^*, i.e., when there exists an $\varepsilon>0$ and a $c<1$ such that

$$|\Phi(X)-\Phi(Y)|\leqq c|X-Y| \tag{8.3}$$

holds for arbitrary X and Y satisfying $|X-X^*|<\varepsilon$, $|Y-X^*|<\varepsilon$. A solution $\tilde{X}^t$ of (8.2) is said to be stable, when it converges to a stable X^*. Unless initial X^0 is an unstable equilibrium, every solution is stable, except for such a pathological case that the set of the equilibrium states has an accumulating point.

In the case of $w<0$, similar arguments hold, if we consider the following subsequence

$$\tilde{X}^0, \tilde{X}^2, \tilde{X}^4, \cdots.$$

If we put

$$\Psi(X)=\Phi(\Phi(X)), \tag{8.4}$$

every subsequence $\tilde{X}^{2t}$ converges to an equilibrium X^* of $\Psi(X)$ satisfying

$$X^* = \Psi(X^*).$$

When the equilibrium X^* of $\Psi(X)$ also satisfies

$$X^* = \Phi(X^*),$$

the original solution $\tilde{X}^t$ surely converges to X^*. However, when

$$X^{*\prime} = \Phi(X^*)$$

does not coincide with X^*, the original solution $\tilde{X}^t$ converges to the oscillatory solution alternating X^* and $X^{*\prime}$ with period 2. We also say that solution $\tilde{X}^t$ is stable, if the subsequence $\tilde{X}^{2t}$ converges to a stable equilibrium of $\Psi(X)$.

We now state the main theorem which guarantees that (8.2) is the macroscopic equation in the strong sense.

THEOREM 9. *Let $\tilde{X}^t$ be the stable solution of* (8.2) *with initial condition $\tilde{X}^0 = X^0$. Then, for an arbitrary $\varepsilon > 0$,*

$$\lim_{n\to\infty} \operatorname{Prob}\{\sup_t |X^t - \tilde{X}^t| < \varepsilon\} = 1. \tag{8.5}$$

To prove Theorem 9, we first prove the following lemma.

LEMMA 5. *For an arbitrary $\varepsilon > 0$,*

$$\lim_{n\in\infty} \operatorname{Prob}\left\{\max_{\mathbf{x}} |X(T_\omega \mathbf{x}) - \Phi(X(\mathbf{x}))| > \varepsilon\right\} = 0. \tag{8.6}$$

Proof. Put

$$u_i = \sum_{j=1}^n w_{ij}x_j - h_i.$$

Then, the next macroscopic state $X' = X(T\omega\mathbf{x})$ is written as

$$X' = \frac{1}{n}\sum_{i=1}^n 1(u_i).$$

By putting

$$a_i = n\bar{w}_n X - h_i,$$

$$b_i = \sum_{j=1}^n (w_{ij} - \bar{w}_n)x_j,$$

where

$$E[w_{ij}] = \bar{w}_n,$$

$$X = X(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^n x_j,$$

u_i is decomposed into the sum

$$u_i = a_i + b_i.$$

Putting

$$D=\max_{x}\left|X'-\frac{1}{n}\sum_{j=1}^{n} 1(a_i)\right|,$$

$$C=\max_{x}\left|\frac{1}{n}\sum_{j=1}^{n} 1(a_i)-\Phi(X)\right|,$$

we obtain

$$\max_{x}|X'-\Phi(X)|\leqq D+C.$$

We see that nD is equal to or less than the number of those components for which a_i and $u_i=a_i+b_i$ have different signs. When there exists a δ such that

$$|a_i|>\delta>|b_i|,$$

a_i and u_i surely have the same sign. Therefore, putting

$$nA=\max_{x}|\{i|\,|a_i|<\delta\}|,$$

$$nB=\max_{x}|\{i|\,|b_i|>\delta\}|$$

where $|\{i|Z(i)\}|$ means the number of those i's for which $Z(i)$ holds, we have

$$D\leqq A+B.$$

Combining the above inequalities, we have

$$\max_{x}|X'-\Phi(X)|\leqq A+B+C.$$

When

$$\max_{x}|X'-\Phi(X)|>\varepsilon=3\varepsilon'$$

holds, it is impossible that all of A, B and C are smaller than or equal to ε' at the same time, and therefore one of them is greater than ε'. Hence, we have

$$\text{Prob}\left\{\max_{x}|X'-\Phi(X)|>3\varepsilon'\right\}\leqq\text{Prob}\{A>\varepsilon'\}+\text{Prob}\{B>\varepsilon'\}+\text{Prob}\{C>\varepsilon'\}.$$

Therefore, the lemma is proved by showing that, for an arbitrary ε',

(A) $$\lim_{n\to\infty}\text{Prob}\{A>\varepsilon'\}=0;$$

(B) $$\lim_{n\to\infty}\text{Prob}\{B>\varepsilon'\}=0;$$

(C)
$$\lim_{n\to\infty} \text{Prob}\{C > \varepsilon'\} = 0.$$

The proofs are given in Appendix C.

LEMMA 6. *For an arbitrary* $\varepsilon > 0$,

$$\lim_{n\to\infty} \text{Prob}\left\{\sup_t |X^{t+1} - \Phi(X^t)| > \varepsilon\right\} = 0. \tag{8.7}$$

Proof. Taking the limit $n \to \infty$ of

$$\text{Prob}\left\{\sup_t |X^{t+1} - \Phi(X^t)| > \varepsilon\right\} \leqq \text{Prob}\left\{\max_{\mathbf{x}} |X(T_\omega \mathbf{x}) - \Phi(X(\mathbf{x}))| > \varepsilon\right\},$$

we have (8.7) from Lemma 5.

Proof of the theorem. We prove the theorem in the case of $w \geqq 0$. (The case with $w < 0$ can be proved in a similar way by using $\Psi(X)$ instead of $\Phi(X)$.) Since $\Phi(X)$ is continuous and bounded, we can easily prove from Lemma 7 that the weak proposition

$$\lim_{n\to\infty} \text{Prob}\{|X^t - \tilde{X}^t| > \varepsilon\} = 0, \qquad t = 0, 1, \cdots, T \tag{8.8}$$

holds, where T is an arbitrary finite number and ε is an arbitrary positive number.

Let $\tilde{X}^t$ be the stable solution of (8.2) with the initial condition $\tilde{X}^0 = X^0$, converging to a stable equilibrium X^*. Let ε' be such a number that $\Phi(X)$ is a contraction in the ε'-neighborhood $U_{\varepsilon'}$ of X^*. Let ε be an arbitrary positive number satisfying

$$2\varepsilon < \varepsilon'.$$

Since $\tilde{X}^t$ converges to X^*, there exists a T such that

$$|X^t - X^*| < \varepsilon, \qquad t \geqq T,$$

holds.

When the two conditions

$$\text{(i)} \quad \sup_t |X^{t+1} - \Phi(X^t)| < (1-c)\varepsilon,$$

$$\text{(ii)} \quad |X^t - \tilde{X}^t| < \varepsilon, \qquad t = 1, 2, \cdots, T,$$

where $c < 1$ is a positive constant defined in (8.3), hold, we can prove

$$\sup_t |X^t - \tilde{X}^t| < \varepsilon \tag{8.9}$$

by induction. We show that the condition $|X^t - \tilde{X}^t| < \varepsilon$ implies $|X^{t+1} - \tilde{X}^{t+1}| < \varepsilon$ for any $t \geqq T$. When $|X^t - \tilde{X}^t| < \varepsilon$ holds, X^t belongs to the ε'-neighborhood of X^*, because

$$|X^t - X^*| \leqq |X^t - \tilde{X}^t| + |\tilde{X}^t - X^*| < 2\varepsilon < \varepsilon'.$$

Hence, applying (8.3), we have

$$|\Phi(X^t)-\Phi(\tilde{X}^t)|\leqq c|X^t-\tilde{X}^t|<c\varepsilon.$$

Then

$$\begin{aligned}|X^{t+1}-\tilde{X}^{t+1}|&\leqq|X^{t+1}-\Phi(X^t)|+|\Phi(X^t)-\Phi(\tilde{X}^t)|\\&<(1-c)\varepsilon+c|X^t-\tilde{X}^t|<(1-c)\varepsilon+c\varepsilon\\&=\varepsilon.\end{aligned}$$

Thus, (i) and (ii) imply (8.9). In other words, if

$$\sup_t |X^t-\tilde{X}^t|>\varepsilon$$

holds, then at least one of (i) and (ii) must be violated. Hence,

$$\begin{aligned}&\text{Prob}\left\{\sup_t |X^t-\tilde{X}^t|>\varepsilon\right\}\\&\quad\leqq \text{Prob}\left\{\sup_t |X^{t+1}-\Phi(X^t)|>(1-c)\varepsilon\right\}+\text{Prob}\left\{\max_{0\leqq t\leqq T}|X^t-\tilde{X}^t|>\varepsilon\right\}.\end{aligned}$$

According to (8.7) and (8.8), the right-hand side converges to 0 as n tends to infinity, which completes the proof of the theorem.

Conclusion. We have derived the macroscopic state equations for various kinds of random nerve nets and proved that the equations hold in the weak sense under some stochastic assumptions on the distribution of synaptic weights. However, the assumptions are not satisfied in the case where synaptic weights are normally distributed. Therefore, we cannot yet prove the weak proposition in this case. We have proved the strong proposition only in a special case.

It should be noted that our method will be applicable not only to random nerve nets but also to various systems of random structure composed of a large number of elements.

Appendix A. Proof of the moment theorem. The latter part of the theorem is proved by the use of the former part. If the convergence $F_n(\mathbf{Y};\mathbf{z})\to F(\mathbf{Y})$ is not uniform in $\mathbf{z}$, there must exist a sequence $\mathbf{z}_n(\mathbf{Y}_0)$ such that $F_n(\mathbf{Y};\mathbf{z}_n(\mathbf{Y}_0))$ does not converge to $F(\mathbf{Y})$ at $\mathbf{Y}=\mathbf{Y}_0$. However, the convergence $\mu_{n,A}(\mathbf{z})\to\mu_A$ is uniform in $\mathbf{z}$, and hence $\mu_{n,A}(\mathbf{z}_n(\mathbf{Y}_0))$ is convergent to μ_A. This fact contradicts the result of the former part.

Appendix B. Proofs of Lemma 3 and Lemma 4. We prove Lemmas 3 and 4 by the method of mathematical induction. Let $[L3]_t$ and $[L4]_t$ be the propositions that Lemma 3 and Lemma 4 holds at time t, respectively. The proof is completed by showing the following three propositions.

1. $[L4]_{t=0}$ is true.
2. $[L3]_t$ implies $[L4]_t$.
3. $[L4]_{t-1}$ implies $[L3]_t$.

Since the initial quantities $x_i(0)$ are independent of the w_{ij}'s and h_i's, [L4] trivially holds at time 0. We prove 2 and 3 in the following.

I. *Proof of* $[L3]_t \Rightarrow [L4]_t$. We first note the following three preliminary lemmas.

LEMMA B.1. *When a sequence of random variables* $u_1, u_2, \cdots$ *converges to* $\tilde{u}$ *in distribution, and when* $\tilde{u}=0$ *is not a discontinuous point of the distribution of* $\tilde{u}$, *then*

$$\lim_{n\to\infty} E[x_n]=E[\tilde{x}],$$

where

$$x_n = 1(u_n), \qquad \tilde{x}=1(\tilde{u}).$$

The proof is obvious, because we have $E[x_n]=\text{Prob}\{u_n>0\}$ and $E[\tilde{x}]=\text{Prob}\{u>0\}$, and $\tilde{u}=0$ is not a discontinuous point.

LEMMA B.2. *The distribution of* $v_i(t)$ *is continuous at* $v_i(t)=0$.

Proof. Since h_i has a continuous distribution, and is independent of $\sum_{j=1}^n w_{ij}\tilde{x}_j(t)$ and its limit, the event $v_i(t)=0$ occurs with probability 0.

LEMMA B.3. *For a finite index set K and a family of finite index sets L,*

$$E[|w_{KL}^C|]=O\left(\frac{1}{n^{|C|}}\right).$$

Proof. The lemma follows from the fact that $E[|w_{ij}|^c]=O(1/n)$ for $c>0$, all the w_{ij}'s are independent and from the definition that $|C|$ denotes the number of nonzero c_{ij}'s.

Proof of $[L3]_t \Rightarrow [L4]_t$. Assume that Lemma 3 holds at time t. Then, we obtain, by virtue of the uniform convergence,

$$\lim_{n\to\infty} E[x_I^A(t)h_J^B|w_{MN}=z_{MN}]=E[\tilde{x}_I^A(t)h_J^B].$$

In other words, by putting

$$E[x_I^A(t)h_J^B|w_{MN}=z_{MN}]=E[\tilde{x}_I^A(t)h_J^B]+\varepsilon_n(z_{MN}), \tag{B.1}$$

we can choose δ_n such that

$$|\varepsilon_n(z_{MN})|<\delta_n$$

irrespective of z_{MN} and

$$\lim_{n\to\infty}\delta_n=0.$$

Let $F_n(y_{KL})$ be the distribution function of a set of random variables w_{KL}, where n denotes the number of elements. Then, the conditional expectation can be written as

$$\begin{aligned}&E[x_I^A(t)h_J^B w_{KL}^C|w_{MN}=z_{MN}]\\&\qquad=\int y_{KL}^C E[x_I^A(t)h_J^B|w_{MN}=z_{MN}, w_{KL}=y_{KL}]dF_n(y_{KL}).\end{aligned} \tag{B.2}$$

Since $(K\cup M)\cap I=0$, taking account of (B.1), we can write

$$E[x_I^A(t)h_J^B|w_{KL}=y_{KL}, w_{MN}=z_{MN}]=E[\tilde{x}_I^A(t)h_J^B]+\varepsilon_n(y_{KL}, z_{MN}),$$

where we can find a sequence δ_n such that

$$|\varepsilon_n(y_{KL}, z_{MN})| < \delta_n,$$

$$\lim_{n\to\infty} \delta_n = 0.$$

Therefore, we have

$$E[x_I^A(t)h_J^B w_{KL}^C | w_{MN} = z_{MN}]$$
$$= \int y_{KL}^C \{E[x_I^A(t)h_J^B] + \varepsilon_n(y_{KL}, z_{MN})\} dF_n(y_{KL})$$
$$= E[\tilde{x}_I^A(t)h_J^B]E[w_{KL}^C] + \int y_{KL}^C \varepsilon_n(y_{KL}, z_{MN}) dF_n(y_{KL}).$$

However, we have

$$\left| n^{|C|} \int y_{KL}^C \varepsilon_n(y_{KL}, z_{MN}) dF_n(y_{KL}) \right|$$
$$\leqq \delta_n n^{|C|} \int |y_{KL}^C| dF_n(y_{KL}) = \delta_n n^{|C|} E[|w_{KL}^C|] = \delta_n n^{|C|} O\left(\frac{1}{n^C}\right) = \delta_n O(1),$$

where Lemma B.3 is taken into account. This converges to 0 uniformly in z_{MN} as n tends to infinity. This completes the proof that Lemma 4 holds at time t, or $[L4]_t$.

II. *Proof of* $[L4]_{t-1} \Rightarrow [L3]_t$. We have the following two preliminary lemmas.

LEMMA B.4. *For any integers t and m, the moment $E[(u_i(t))^m]$ and its limit $n \to \infty$ exist.*

Proof. We have

$$E[(u_i(t))^m] \leqq E\left[\left(\sum_{j=1}^n |w_{ij}x_j(t-1)| + |h_i|\right)^m\right]$$
$$\leqq E\left[\left(\sum_{j=1}^n |w_{ij}| + |h_i|\right)^m\right]$$

Since $\sum_{j=1}^n w_{ij}$ and h_i are mutually independent, and since $E[(\sum_{j=1}^n |w_{ij}|)^m]$, its limit, and $E[|h_i|^m]$ always exist, both $E[(u_i(t))^m]$ and its limit exist.

LEMMA B.5. *Let J be a finite set of integers and let $u_i'(t)$ be*

$$u_i'(t) = \sum_{j\in J} w_{ij}x_j(t-1) - h_i.$$

Then, the moment of $u_i'(t)$ of any order converges to the corresponding moment of $u_i(t)$, in the limit $n \to \infty$.

Proof. We have

$$E[(u_i(t))^m - (u_i'(t))^m] \tag{B.3}$$
$$\leqq \sqrt{E[|u_i(t) - u_i'(t)|^2]} \sqrt{E[(\textstyle\sum_{k=0}^m (u_i(t))^{m-k}(u_i'(t))^k)^2]},$$

where Schwarz' inequality is taken into account. We have, moreover,

$$E[(u_i(t))^a(u_i'(t))^b] \leqq \sqrt{E[(u_i(t))^{2a}]E[(u_i'(t))^{2b}]}.$$

Since $u_i(t)$ and $u_i'(t)$ have the moments of any order in the limit $n \to \infty$,

$$\lim_{n\to\infty} E[((u_i(t))^{m-k}(u_i'(t))^k)^2]$$

surely exists. On the other hand, since $|J|$ is finite,

$$E[|u_i(t)-u_i'(t)|^2]=E[|\sum_{j\in J} w_{ij}x_j(t-1)|^2]\leqq E[|\sum_{j\in J} w_{ij}|^2]=O\left(\frac{1}{n}\right).$$

Therefore, the right-hand side of (B.3) vanishes in the limit, and we have

$$\lim_{n\to\infty} E[(u_i(t))^m]=\lim_{n\to\infty} E[(u_i'(t))^m].$$

Before proving the proposition, we need some preliminaries on abbreviated notations. Let us consider the following expansion:

$$\left(\sum_{i=1}^n x_i\right)^a=\sum_{i_1}(x_{i_1})^a+\sum_{a_1+a_2=a}\ \sum_{i_1<i_2}(x_{i_1})^{a_1}(x_{i_2})^{a_2}$$
$$+\sum_{a_1+a_2+a_3=a}\ \sum_{i_1<i_2<i_3}(x_{i_1})^{a_1}(x_{i_2})^{a_2}(x_{i_3})^{a_3}+\cdots.$$

We adopt the following abbreviation for the above expression:

$$\left(\sum_{i=1}^n x_i\right)^a=\sum_{\kappa=1}^n \sum_{A_\kappa}\sum_{I_\kappa} x_{I_\kappa}^{A_\kappa}$$

where I_κ stands for a set of κ integers $(i_1, i_2, \cdots, i_\kappa)$ satisfying $1\leqq i_1<i_2<\cdots< i_\kappa\leqq n$, A_κ stands for a power set $(a_1, a_2, \cdots, a_\kappa)$ corresponding to I_κ, satisfying

$$\sum_{i=1}^{\kappa} a_i=a, \qquad a_i>0,$$
$$x_I^{A_\kappa}=x_{i_1}^{a_1}\cdots x_{i_\kappa}^{a_\kappa}$$

and the summation is taken over all these I_κ and A_κ. It should be noted that the number of the sets of integers represented by the I_κ's is ${}_nC_\kappa$, the binomial coefficient, and when n is large,

$${}_nC_\kappa=O(n^\kappa).$$

The ath power of $u_i(t)$ is written, in this notation, as follows

$$(u_i(t))^a=\left(\sum_{j=1}^n w_{ij}x_j(t-1)-h_i\right)^a$$
(B.4)
$$=\sum_{d=0}^a {}_aC_d(-h_i)^{a-d}\sum_{\kappa=1}^d \sum_{A_\kappa}\sum_{I_\kappa} w_{iI_\kappa}^{A_\kappa}x_{I_\kappa}^{A_\kappa}(t-1),$$

where

$$w_{iI_\kappa}^{A_\kappa}=(w_{ii_1})^{a_1}(w_{ii_2})^{a_2}\cdots(w_{ii_\kappa})^{a_\kappa}.$$

So far as the limit $n\to\infty$ of the nth moment $E[(u_i(t))^a]$ is concerned, we may replace I_κ in (B.4) by I'_κ,

$$(u_i(t))^a = \sum_{d=0}^{a} {}_aC_d(-h_i)^{a-d} \sum_\kappa \sum_{A_\kappa} \sum_{I'_\kappa} w_{iI'_\kappa}^{A_\kappa} x_{I'_\kappa}^{A_\kappa}(t-1),$$

where I'_κ represents a set of κ integers which does not include any integers belonging to K, where $|K|$ is finite. The number of the sets of these I'_κ's is also of the order $O(n^\kappa)$.

Similarly, the product of the $(u_i(t))^{a_i}$'s over $i \in I$ is symbolically written as

$$u_I^A(t) = \prod_{i\in I} (u_i(t))^{a_i}$$

$$= \sum_D {}_AC_D(-h_I)^{A-D} \sum_\kappa \sum_{A_\kappa} \sum_{J_\kappa} w_{IJ_\kappa}^{A_\kappa} x_{J_\kappa}^{A_\kappa}(t-1)$$

where $A = (a_i; i \in I)$, $D = (d_i; i \in I, 0 \leqq d_i \leqq a_i)$,

$${}_AC_D = \prod_{i\in I} {}_{a_i}C_{d_i},$$

$$(-h_I)^{A-D} = \prod_{i\in I} (-h_i)^{a_i-d_i},$$

$\kappa = (\kappa_i, i \in I)$, $\sum_\kappa$ stands for $\sum_{i\in I}\sum_{\kappa_i=1}^{d_i}$ and J_κ denotes the family of sets J_{κ_i} $(i \in I)$ of integers and A_κ denotes the corresponding family of power sets.

Proof of $[\mathrm{L4}]_{t-1} \Rightarrow [\mathrm{L3}]_t$. Assume that Lemma 4 holds at time $t-1$. The conditional expectation of the moments of $u_I(t)$ and h_J is calculated as

$$E[u_I^A(t)h_J^B | w_{KL} = z_{KL}]$$

(B.5)
$$= \sum_D {}_AC_D \sum_\kappa \sum_{A_\kappa} \sum_{J_\kappa} E[h_J^B(-h_I)^{A-D} w_{IJ_\kappa}^{A_\kappa} x_{J_\kappa}^{A_\kappa}(t-1) | w_{KL} = z_{KL}]$$

for $I \cap J = \varnothing$ and $I \cap K = \varnothing$. So far as the limit $n \to \infty$ is concerned, the above J_κ in (B.5) may be replaced by the J'_κ's, where J'_κ consists of those integer sets J'_{κ_i} which do not include integers in I, J and K, i.e., $J'_{\kappa_i} \cap (I \cup J \cup K) = 0$. Since Lemma 4 holds at time $t-1$, we obtain

$$E[h_J^B(-h_I)^{A-D} w_{IJ'_\kappa}^{A_\kappa} x_{J'_\kappa}^{A_\kappa}(t-1) | w_{KL} = z_{KL}]$$

$$= E[h_J^B(-h_I)^{A-D} w_{IJ'_\kappa}^{A_\kappa} \tilde{x}_{J'_\kappa}^{A_\kappa}(t-1)] + O\left(\frac{1}{n^{|A_\kappa|}}\right)\varepsilon_n,$$

where

$$\lim_{n\to\infty} \varepsilon_n = 0$$

uniformly in z_{KL}. Obviously, $|A_\kappa| = |J'_\kappa| = \sum_i |J'_{\kappa_i}|$ holds. Hence, the number of terms summed over under the symbol $\sum_{J'_\kappa}$ is of order $O(n^{|A_\kappa|})$. Therefore, we have

$$\Big|\sum_{J'_\kappa} E[h_J^B(-h_I)^{A-D} w_{IJ'_\kappa}^{A_\kappa} x_{IJ'_\kappa}^{A_\kappa}(t-1) | w_{KL} = z_{KL}]$$

$$- E[h_J^B(-h_I)^{A-D} w_{IJ'_\kappa}^{A_\kappa} \tilde{x}_{J'_\kappa}^{A_\kappa}(t-1)]\Big| = \varepsilon_n$$

with

$$\lim_{n\to\infty} \varepsilon_n = 0.$$

Since the summation with respect to D, κ and A_κ includes only a finite number of terms, the conditional expectation converges to

$$E[v_I^A(t)h_J^B]$$

in the limit. We have thus proved that all the moments of $u_I(t)$ and h_J under the condition $w_{KL} = z_{KL}$

$$E[u_I^A(t)h_J^B | w_{KL} = z_{KL}]$$

converge to those of $v_I(t)$ and h_J

$$E[v_I^A(t)h_J^B]$$

uniformly in z_{KL}. By virtue of the moment theorem, this implies that Lemma 3 holds at time t or $[L3]_t$.

Appendix C. Proofs of (A), (B) and (C).

Proof of (A). Let

$$p_\delta(X) = \text{Prob}\{|a_i| < \delta\}.$$

Since h_i has a continuous distribution, for any X, $p_\delta(X)$ converges to 0 as δ tends to 0. In other words

$$p_\delta = \max_{0 \leqq X \leqq 1} p_\delta(X)$$

can be made arbitrarily small by taking a sufficiently small δ. Since a_i depends on x only through X, we have

$$\text{Prob}\{A > \varepsilon'\} = \text{Prob}\left\{\max_{0 \leqq X \leqq 1} |\{i|\,|a_i| < \delta\}| > n\varepsilon'\right\}.$$

When $\max_X |\{i|\,|a_i| < \delta\}| > n\varepsilon'$ occurs, $|\{i|\,|a_i| < \delta\}| > n\varepsilon'$ occurs for at least one X. Hence,

$$\text{Prob}\{A > \varepsilon'\} \leqq \sum_X \text{Prob}\{|\{i|\,|a_i| < \delta\}| > n\varepsilon'\}.$$

Since the events $|a_i| < \delta$ are independent for all different i's, we have

$$\text{Prob}\{|\{i|\,|a_i| < \delta\}| > n\varepsilon'\} = \sum_{j > n\varepsilon'} {}_nC_j (p_\delta(X))^j (1 - p_\delta(X))^{n-}$$

$$\leqq \sum_{j > n\varepsilon'} {}_nC_j (p_\delta)^j (1 - p_\delta)^{n-j},$$

where ${}_nC_j$ are the binomial coefficients. Since X takes only $n+1$ values $0, 1/n, 2/n, \cdots, 1$, we finally have

$$\text{Prob}\{A > \varepsilon'\} \leqq (n+1) \sum_{j > n\varepsilon'} {}_nC_j (p_\delta)^j (1 - p_\delta)^{n-j},$$

which converges to 0 as n tends to infinity.

Proof of (B). Put

$$q_n(x) = \text{Prob}\,\{|b_i| > \delta\}.$$

It is easy to show that b_i converges to 0 in distribution as n tends to infinity. Therefore, for an arbitrary δ, $q_n(\mathbf{x})$ converges to 0 as n tends to infinity. Moreover,

$$q_n = \max_{\mathbf{x}} q_n(\mathbf{x})$$

becomes arbitrarily small as n tends to infinity.

$$\begin{aligned} \text{Prob}\,\{B > \varepsilon'\} &= \text{Prob}\,\{\max_{\mathbf{x}} |\{i|\,|b_i| > \delta\}| > n\varepsilon'\} \\ &\leqq \sum_{\mathbf{x}} \text{Prob}\,\{|\{i|\,|b_i| > \delta\}| > n\varepsilon'\} \\ &\leqq 2^n \sum_{j > n\varepsilon'} {}_nC_j (q_n)^j (1-q_n)^{n-j}, \end{aligned}$$

for there are 2^n x's and the events $|b_i| > \delta$ are independent. When n is large, q_n becomes smaller than ε'. In this case, we have

$$2^n \sum_{j > n\varepsilon'} {}_nC_j (q_n)^j (1-q_n)^{n-j} \leqq n 2^n {}_nC_{[n\varepsilon']} (q_n)^{[n\varepsilon']},$$

where $[n\varepsilon']$ is the least integer greater than $n\varepsilon'$. Hence,

$$\lim_{n \to \infty} \text{Prob}\,\{B > \varepsilon'\} = 0.$$

Proof of (C). Since a_i depends on $\mathbf{x}$ through $X = X(\mathbf{x})$, we put

$$r_n(X) = \text{Prob}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} 1(a_i) - \Phi(X)\right| > \varepsilon'\right\}.$$

From

$$\text{Prob}\,\{1(a_i) = 1\} = \text{Prob}\,\{h_i \leqq n\bar{w}_n X\},$$

which converges to $F(wX) = \Phi(X)$ as $n \to \infty$, we see that $r_n(X)$ converges to 0 as n tends to infinity. Part (C) is proved from

$$\begin{aligned} \text{Prob}\,\{C > \varepsilon'\} &= \text{Prob}\left\{\max_X \left|\frac{1}{n}\sum_{i=1}^{n} 1(a_i) - \Phi(X)\right| > \varepsilon'\right\} \\ &\leqq \sum_X \text{Prob}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} 1(a_i) - \Phi(X)\right| > \varepsilon'\right\} \\ &= (n+1) \max_X r_n(X). \end{aligned}$$

REFERENCES

[1] J. T. ALLANSON, *Some properties of randomly connected neural networks*, Proc. 3rd London Symp. on Information Theory, C. Cherry, ed., Butterworths, London, 1956.

[2] S. AMARI, *Characteristics of randomly connected threshold element networks and network systems*, Proc. IEEE, 59 (1971), pp. 35–47.

[3] ———, *Characteristics of random nets of analog neuron-like elements*, IEEE Trans. Systems, Man Cybernet., SMC-2 (1972), pp. 643–657.

[4] ———, *Learning patterns and pattern sequences by self-organizing nets of threshold elements*, IEEE Trans. Computers, C-21 (1972), pp. 1197–1206.

[5] ———, *A method of statistical neurodynamics*, Kybernetik, 14 (1974), pp. 201–215.

[6] ———, *A mathematical theory of nerve nets*, Advances in Biophysics, 6, M. Kotani, ed., University of Tokyo Press, Tokyo, 1974, pp. 75–120.

[7] P. A. ANNIOS, *Evoked potential in artificial neural nets*, Kybernetik, 13 (1973), pp. 24–29.

[8] W. R. ASHBY, F. v. FOERSTER AND C. C. WALKER, *Instability of pulse activity in a net with threshold*, Nature, 196 (1962), pp. 561–562.

[9] D. YA. AVERBUKH, *Random nets of analog neurons*, Automat. Remote Control, (1969), pp. 116–123.

[10] J. L. DOOB, *Stochastic Processes*, John Wiley, New York, 1953.

[11] J. L. FELDMAN AND J. D. COWAN, *Large-scale activity in neural nets II*, Biol. Cybernet., 17 (1975), pp. 39–51.

[12] T. S. HAN, *Linear dependence structure of the entropy space*, Information and Control, 29 (1975), pp. 337–368.

[13] E. M. HARTH, T. J. CSERMELY, B. BEEK AND R. D. LINDSAY, *Brain functions and neural dynamics*, J. Theoret. Biol., 26 (1970), pp. 93–120.

[14] M. KAC, *Probability and Related Topics in Physical Sciences*, Interscience, London, 1959.

[15] M. G. KENDALL AND K. S. RAO, *On the generalized second limit-theorem in the calculus of probabilities*, Biometrika, 37 (1950), pp. 224–230.

[16] R. J. MACGREGOR AND R. L. PALASEK, *Computer simulation of rhythmic oscillation in neuron pools*, Kybernetik, 16 (1974), pp. 79–86.

[17] A. RAPOPORT, *Ignition phenomena in random nets*, Bull. Math. Biophys., 14 (1952), pp. 35–44.

[18] A. R. ROTENBERG, *Behavior of Markovian statistical ensembles of finite automata*, Automat. Remote Control, (1971), pp. 84–92.

[19] L. I. ROZONOER, *Random logical nets I*, Ibid., (1969), pp. 137–147.

[20] ———, *Random logical nets II*, Ibid., (1969), pp. 99–109.

[21] ———, *Random logical nets III*, Ibid., (1969), pp. 129–136.

[22] D. R. SMITH AND C. H. DAVIDSON, *Maintained activity in neural nets*, J. Assoc. Comput. Mach., 9 (1962), pp. 268–278.

[23] R. THOM, *Stabilities structurelles et morphogenese*, W. A. Benjamin, Reading, MA, 1972.

[24] R. C. TOLMAN, *The Principles of Statistical Mechanics*, Oxford University Press, Oxford, 1938.

[25] S. S. WILKS, *Mathematical Statistics*, John Wiley, New York and London, 1962.

[26] S. YOSHIZAWA, *Some properties of randomly connected networks of neuron-like elements with refractory*, Kybernetik, 16 (1974), pp. 173–182.