KŌDAI MATH. SEM. REP
20 (1968), 355-373

PATTERN RECOGNITION BY RANDOM NET

BY YASUICHI HORIBE

1. Introduction.

It might be said that the significance of the role played by random transmission of information or signal transmission through a random net in the large scale information processing systems such as pattern recognizing learning systems has become to be recognized through a kind of simulation of a certain part of the information processing mechanism in some living organisms. Especially in the perceptron theory [2], which, getting out of the experimental stages, has attracted various theoretical attentions, e.g. [1], [5], [6], [7], [8], [12], one considers a random connection structure by such a simulation. The theory, however, does not use any essential property of "randomness", i.e. mathematically this random connection only transforms input stimuli (or patterns) into another forms of patterns, and therefore one may consider only on these transformed patterns as if they were original input patterns.

The work of the random net in such systems might be said not to be clarified particularly from the theoretical view point.

Generally a random net consists of, like neural nets, a large number of elements which might be called "basic organs" corresponding to neurons in neural nets and emit signals after transforming their input signals by certain operations, and a large number of random "connections" which transmit signals among the basic organs. A subset I of the set of all basic organs contained in the random net receives stimulus signals from the outer world of the net. The signals, then, propagate along connections through the whole net. They are transformed many times by basic organs, and eventually reach another specified subset O of basic organs which output final signals. The net in the whole, therefore, has a general character of transforming input signals into output ones except for the random signal transformations.

Now suppose that a signal has a multi-dimensional vector form and each component signal in the vector signal is received by each basic organ in the subset I of our random net. Correlations or dependences among component signals may easily be seen to be generally strong if we only consider our visual patterns as vector signals. One of the important roles of the random net is "to tear these dependences into independent pieces" to gain statistical independences among component signals. In other words, as a vector signal passes through the net, it

Received December 7, 1967.

gradually loses its initial strong component correlations because of the random connections in the net.

It must however be remarked that the acquisition of independences should not lose the original "information" which the input signals have, i.e. the equivocation should be maintained near zero or below a certain desirable level. It seems to be for this requirement that the size or the dimension of a random net must be large, i.e. it must have a large number of basic organs and complicated connections among them. This seems to be similar to the general notion that to gain the reliability of information through an unreliable medium one must increase the dimension of the system involved, for example, the coding problem in information theory [4], or von Neumann's reliable construction of automata from less reliable basic organs [13].

The object of the present paper is to consider a reason of why "from dependence to independence" in the light of the pattern recognition problem, and to evaluate a certain kind of random net, which will be shown to be capable of recognizing patterns.

The pattern recognition methods by perceptron-type iterative error-correction learning which have been proposed in various ways, e.g. [1], [5], [6], [7], [8], [12], never uses random net structures and presupposes that the input pattern space is decomposed (disjointly) into a finite number of category classes. In the present paper, however, we treat patterns with a more general character as reflected in setting a general statistical structure on the input pattern space, and show that in the random net which has as its all basic organs (except for I) AND-basic organs and OR-basic organs, if the number of input points (i.e. the number of elements in I) and the number of layers of basic organs are taken sufficiently large, then the random net becomes pattern recognizable, where the pattern recognizability is the direct consequence that the random net which satisfies the above mentioned requirements (i.e. statistical independence and information losslessness) can recognize patterns by *a posteriori* probability method, assuming a certain learning mechanism that stores a set of frequencies or probabilities gained from the sample training patterns taken from the input pattern space.

2. General concept and motivation.

The random net we shall consider here has n *input points* each of which receives signal 0 or 1 (non-stimulus or stimulus, or, white or black) from an outer world of the net, and output basic organs each of which emits signal also 0 or 1 (as well as other basic organs in the net) to the outer world or a certain observer. Let the set of input points be $I=\{1, 2, \cdots, n\}$ and the set of output basic organs be $O=\{1, 2, \cdots, N\}$. A random net with I and O has of course a *randomly fixed structure* between I and O (Fig. 1).

Suppose that an input signal is an n-dimensional vector whose n components are signals 0 and/or 1 which are received simultaneously by I, each component corresponding respectively to each input point in I. Suppose also that possible

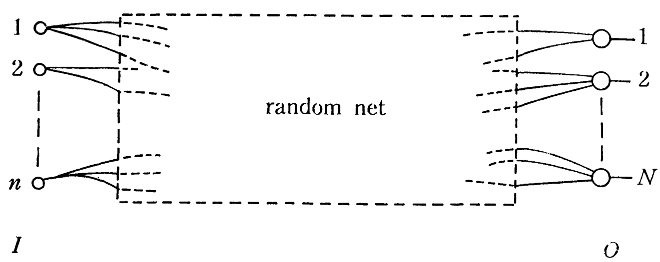

Fig. 1

time delays produced in the course of signal transmission through the net are properly synchronized so that we can observe an output signal as an N-dimensional vector with components 0 and/or 1. We call an n-dimensional input vector *input pattern* or simply *pattern*. The set F of all 2^n possible input patterns will be called (*input*) *pattern space*.

A pattern $f=(\sigma_1, \cdots, \sigma_n)$ or simply written as $f=\sigma_1, \cdots, \sigma_n$, with $\sigma_i=0$ or 1, may be a binary expression of a continuous signal like a speech sound. Or a pattern may be a direct white-black visual image (on a "retina") of an object like a character, for instance the image of the number 2 shown in the Fig. 2 has its pattern 000011000011011001100110000110000011000001111110 with $n=6\times8=48$ (this 48 is too small) if the image is scanned horizontally from left to right.

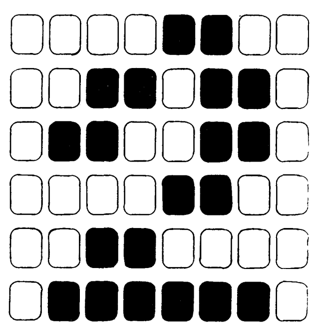

Fig. 2

Now to say that there are given K *categories on the pattern space* F to which each pattern belongs is to say that there are given K probability distributions $P^{(1)}, P^{(2)}, \cdots, P^{(K)}$ on F. Denote by $C=\{1, 2, \cdots, K\}$ the set of categories and put $P^{(k)}\{f_i\}=p_i^{(k)}$. Then the $K\times2^n$-matrix $(p_i^{(k)})$, $k\in C$, $i=0, 1, \cdots, 2^n-1$, may be regarded as a channel matrix of "pattern generating channel" with the sending signal space C and the receiving signal space F. The value $p_i^{(k)}$ is the conditional probability that f_i is presented to I, given k.

The pattern classification (recognition) means the decision that to which category the current input pattern belongs.

A special case of the pattern generating channel is the lossless channel [4], i.e. it is possible to decompose the pattern space F into K disjoint subsets, $F=\cup_{k=1}^{K}F^{(k)}$ such that $P^{(k)}\{F^{(k)}\}=1$ for every $k\in C$. This case has been treated as a problem of pattern recognition by means of a suitable learning algorithm or an iterative accompolishment of a correct classification of input patterns, but never essentially used any property of random net [1], [2], [5], [6], [7], [8], [12].

In order to make clear the possible differences between this iterative procedure (or usually called perceptron-type learning) and the random net approach which concerns us in the present work, we briefly sketch the essential mathematical features of the former for the case, $K=2$ and the binary patterns, which never loses generality. The input pattern space F is transformed by an N-dimensional vector valued function φ into a (transformed) pattern space G,

$$\varphi=(\varphi_1,\cdots,\varphi_N):\quad F\rightarrow G,\quad \varphi(F)=G.$$

φ might be a product $\varphi^L\cdots\varphi^2\varphi^1$ of L mappings for a suitably chosen integer L. Suppose that we could have taken φ such that the two sets $\{G^{t(1)},G^{t(2)}\}$, $G^{t(k)}=\varphi(F^{t(k)})\subset G$ for "training input pattern sets" $F^{t(k)}\subset F^{(k)}$, $k\in\{1,2\}$, is (strictly) linearly separable, i.e. there exists $(N+1)$-dimensional vector s^* such that $(s^*,g^*)>0$ for any $g^*\in G^{t(1)*}\cup(-G^{t(2)*})$, where $G^{t(1)*}=\{g^*=(g,1);\ g\in G^{t(1)}\}$ and $-G^{t(2)*}=\{-g^*=(-g,-1);\ g\in G^{t(2)}\}$. Then the learning algorithm (e.g. [2]) is applicable, which iteratively "improves" the $(N+1)$-dimensional "state" vector s_n as follows: set s_0 arbitrarily, and put $s_{n+1}=s_n+g_n^*$ if $(s_n,g_n^*)\leqq 0$, $s_{n+1}=s_n$ if $(s_n,g_n^*)>0$, where $\{g_n^*,\ n=0,1,\cdots\}$, $g_n^*\in G^{t(1)*}\cup(-G^{t(2)*})$ is generated according to a training pattern sequence $\{f_n,\ n=0,1,\cdots\}$, each training pattern in $F^t=F^{t(1)}\cup F^{t(2)}$ being presented to I infinitely often in this sequence. This completes the recognition ability for any pattern in the training pattern set $F^t\subset F$, with "finite" learning, i.e. there exists a finite integer n^* such that $\{s_n\}$ converges to s_{n^*} so that $s_{n^*}=s_{n^*+1}=s_{n^*+2}=\cdots$ and $(s_{n^*},g^*)>0$ for every $g^*\in G^{t(1)*}\cup(-G^{t(2)*})$. (An identification problem for the "solution" state s_{n^*} is discussed in [9]). If $\{F^{t(1)},F^{t(2)}\}$ itself is linearly separable, we may take φ as the identity mapping. If not, we must generally take N larger than n. Specifically, for example, we may take $N=\sum_{j=1}^{s}\binom{n}{j}\leqq 2^n-1$, and take the component functions of φ (taken $L=1$) in such forms as $\varphi_i(f)=\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_j}$ $(1\leqq j\leqq s)$, $f=\sigma_1\cdots\sigma_n$, for non-negative integers $i_1,i_2,\cdots,i_j$ satisfying $1\leqq i_1<i_2<\cdots<i_j\leqq n$, choosing a suitable positive integer s such that $\{G^{t(1)},G^{t(2)}\}$ is linearly separable. Since N becomes very large when $\{F^{(1)},F^{(2)}\}$ has a complicated separating structure, the speed of learning convergence, which strongly depends on the unknown pattern structure, seems to be redundantly very slow. And furthermore the extrapolating or generalizing ability (i.e. the ability to recognize patterns in $F-F^t$) seems to be weak except for cases of very simple pattern structure. These defects might be more undesirable if either the number of categories becomes large or the condition that the pattern generating channel is lossless is modified.

In the present work we treat patterns with more general character as reflected in setting a general statistical structure on the input pattern space, fully applying a randomly interconnected system, and indicate a potential direction toward systems capable of recognizing a wider class of patterns.

Now let us not ask for a while what kind of structure the random net between I and O in the Fig. 1 has, but only regard it as a "black box" or an unknown mapping φ: $F \rightarrow G$, where G is the output pattern space comprising of all possible 2^N output patterns or transformed patterns observed at O. (Although φ might be a "noisy" or probabilistic transformation of input patterns due to, for example, malfunctioning of basic organs in the net, we shall later assume φ to be deterministic, hence we can only observe at most 2^n distinct output patterns each with positive probability.) The random net may be considered as a kind of information channel Γ_1 characterized by φ with sending signal space F and receiving signal space G. Denoting by Γ_0 the pattern generating channel, then the two channels are cascaded: $\Gamma_1\Gamma_0$: $C \rightarrow F \rightarrow G$. The channel matrix $\mathcal{M}_{01}=(q_j^{(k)})$, $k \in C$, $j=0, 1, \cdots, 2^N-1$, for this cascaded channel is, therefore, determined by the matrix $\mathcal{M}_0=(p_i^{(k)})$ and φ. We must take N as $N \geqq n$, for otherwise $\Gamma_1\Gamma_0$ may become more noisy that the original Γ_0 due to the transformation φ, hence more erroneous in pattern classification decision (such as a decision method described soon). And even if we planned to perform the pattern classification decision based on $\mathcal{M}_0$ without using random net, it would be absolutely impractical, when n is large, to learn by a proper mechanism the gigantic matrix $\mathcal{M}_0$ and to use it for real decisions. More impractical for the decision based on $\mathcal{M}_{01}$, since $N \geqq n$.

Now for each category $k \in C$, put the output of the l-th basic organ on the output level O ($l=1, 2, \cdots, N$) as a random variable $x_l^{(k)}$, and put $Q_{l\sigma}^{(k)}=P^{(k)}\{x_l^{(k)}=\sigma\}$ ($\sigma=0, 1$), which is a conditional probability that the output symbol of l-th basic organ is σ, given category k. Hence

$$Q_{l0}^{(k)}+Q_{l1}^{(k)}=1, \quad k \in C, \quad l=1, 2, \cdots, N.$$

Suppose now that the $K \times N$-matrix $\mathcal{M}=(Q_{l0}^{(k)})$ is calculated (or estimated) *by some learning mechanism* which for example might be a method of mere sampling of patterns from F followed by indicating or *teaching* to which category each sample pattern (or *training pattern*) belongs. Generally however $\mathcal{M}$ can not completely specify $\mathcal{M}_{01}$, since the k-th row of $\mathcal{M}_{01}$, $(q_0^{(k)}, q_1^{(k)}, \cdots, q_{2^N-1}^{(k)})$, is a joint probability distribution of N random variables $x_1^{(k)}, x_2^{(k)}, \cdots, x_N^{(k)}$, and $(Q_{10}^{(k)}, Q_{20}^{(k)}, \cdots, Q_{N0}^{(k)})$ is a marginal probability distribution of $(q_0^{(k)}, q_1^{(k)}, \cdots, q_{2^N-1}^{(k)})$. But if, for each category k, random variables $x_1^{(k)}, x_2^{(k)}, \cdots, x_N^{(k)}$ are mutually independent, then $\mathcal{M}_{01}$ is completely determined by $\mathcal{M}$, since any element $q_j^{(k)}$ in $\mathcal{M}_{01}$ is represented by all elements in the k-th row of $\mathcal{M}$, i.e.

$$(1) \qquad q_j^{(k)}=\prod_{l=1}^{N}[Q_{l0}^{(k)}]^{1-\sigma_l}\cdot[1-Q_{l0}^{(k)}]^{\sigma_l},$$

if the observed pattern $g_j \in G$ is $\sigma_1\sigma_2 \cdots \sigma_N$ ($\sigma_l=0, 1$).

This is why we use random net to gain independent output component signals

$x_1^{(k)}, x_2^{(k)}, \cdots, x_N^{(k)}$ for each $k \in C$. The *complete* statistical independence, however, should be considered to be unable to be obtained by any *finite* scale random net, because, generally speaking, "distance" between any two out of K distributions $(p_0^{(k)}, p_1^{(k)}, \cdots, p_{2^n-1}^{(k)})$, $k \in C$, on F is not small and each distribution itself is a fairly sufficiently informative reflection of each category, and thus, for each category, strong correlations or dependences among n components constituting input patterns exist. Therefore we have to show in a certain type of random net that we can attain arbitrarily near independence to the complete one by suitably enlarging the scale or dimension of the net. This is our first requirement for our random net.

Next consider the decision problem how the pattern classification can be performed. Our pattern classification method is simply the usual *a posteriori* probability method [3]. We know the matrix $\mathscr{M}$ and furthermore the probability distribution $(p^{(1)}, p^{(2)}, \cdots, p^{(K)})$ on the category space C, both of which are supposed to be estimated by learning. We can then procede the decision as follows:

(A) When an unknown input pattern is presented to I of the random net, the corresponding output pattern $g_j \in G$ is observed.

(B) Compute $p^{(1)} \cdot q_j^{(1)}, p^{(2)} \cdot q_j^{(2)}, \cdots, p^{(K)} \cdot q_j^{(K)}$ by (1) using the matrix $\mathscr{M}$.

(C) Find the maximal subset $C(g_j) \subset C$ such that every category k in $C(g_j)$ gives the maximum value among $p^{(1)} q_j^{(1)}, p^{(2)} q_j^{(2)}, \cdots, p^{(K)} q_j^{(K)}$.

(D) Decide that the unknown input pattern belongs to one of categories in $C(g_j)$

Let us denote this decision on g_j by $\mathscr{D}[g_j] = C(g_j)$. Note that since

$$\mathrm{Prob}(k|g_j) \cdot \mathrm{Prob}(g_j) = p^{(k)} q_j^{(k)},$$

where $\mathrm{Prob}(g_j)$, the absolute probability of the pattern g_j, is independent of k, $C(g_j)$ is the set of all categories which give the same maximum value among *a posteriori* probabilities, $\mathrm{Prob}(k|g_j)$, $k \in C$.

Although the direct decision $\mathscr{D}[f_i]$ on input pattern f_i without using our random net can not be performed for the reason as noted above when n, the number of input points, becomes large (as usual case), if we could take $\varphi \colon F \to G$ (or equivalently the net structure) to be one-to-one function, we should have $\mathscr{D}[f_i] = \mathscr{D}[\varphi(f_i)]$ as a trivial corollary of the following general (i.e. permitting φ to be probabilistic) proposition.

PROPOSITION. *If the pattern transforming channel Γ_1 is lossless, then $\mathscr{D}[f] = \mathscr{D}[g]$ for $f \in F$ and $g \in \varphi(f) = \{g\colon \mathrm{Prob}\,(g|f) \neq 0\}$.*

Proof. Since Γ_1 is lossless, we have $G = \cup_{f \in F} \varphi(f)$, $\varphi(f) \cap \varphi(f') = 0$ $(f \neq f')$, and $\mathrm{Prob}(\varphi(f)|f) = 1$. Suppose $\mathscr{D}[f^*] = C(f^*)$ for $f^* \in F$, then $p^{(k)} \cdot p_{f^*}^{(k)} > p^{(k')} \cdot p_{f^*}^{(k')}$ for $k \in C(f^*)$, $k' \in C - C(f^*)$, where $p_{f^*}^{(k)}$ is the conditional probability of f^*, given k. For arbitrary pattern $g^* \in \varphi(f^*)$ we have

$$\begin{aligned} p^{(k)} \cdot q_{g^*}^{(k)} &= p^{(k)} \cdot \sum_{f \in F} \mathrm{Prob}(g^*|f) \cdot p_f^{(k)} \\ &= p^{(k)} \cdot \mathrm{Prob}(g^*|f^*) \cdot p_{f^*}^{(k)} > p^{(k')} \cdot \mathrm{Prob}(g^*|f^*) p_{f^*}^{(k')} \\ &= p^{(k')} \sum_{f \in F} \mathrm{Prob}(g^*|f) p_f^{(k')} = p^{(k')} \cdot q_{g^*}^{(k')}. \end{aligned}$$

Similarly we have $p^{(k)}q_{g^*}^{(k)}=p^{(k')}q_{g^*}^{(k')}$ if $k, k'\in C(f^*)$. This means that $\mathcal{D}[g^*]=C(f^*)$ for every $g^*\in\varphi(f^*)$.

But we should consider that because of the characteristics which the random net has the *complete* one-to-one φ could not always be obtained. Hence it is desirable that the probability for φ to be one-to-one can be arbitrarily near one. This is the second requirement for our random net.

In summary we want to have a random net which has the following two properties, assuming that the net has a certain learning mechanism which can estimate matrix $\mathcal{M}$ and $K-1$ probabilities $p^{(1)}, p^{(2)}, \cdots, p^{(K-1)}$, hence $KN+(K-1)$ probabilities in all.

(a) statistically almost independence among output component signals.
(b) statistically almost one-to-one correspondence between input pattern space and output pattern space.

3. The random net with AND-neurons and OR-neurons.

We shall call basic organs contained in a random net simply *neurons*. The random net considered in the present work has as its neurons only AND-neurons and OR-neurons. Each of these two kinds of neuron has two "input lines" which receive stimuli (signals) 0 and/or 1, and one "output line" which emits signal 0 or 1 by performing the following Boolean operation on the input signals. AND-neuron receives signals σ_1, σ_2 ($=0$ or 1) and emits 1 if and only if $\sigma_1=\sigma_2=1$ (hence threshold value 2) and emits 0 otherwise, i.e. emits $\sigma_1\wedge\sigma_2$, while OR-neuron emits 0 if and only if $\sigma_1=\sigma_2=0$ and emits 1 otherwise (hence threshold value 1), i.e. emits $\sigma_1\vee\sigma_2$.

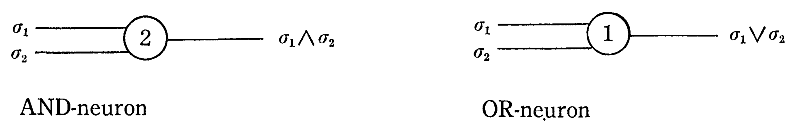

Fig. 3

But the random net does not have these elements in order to do some logical calculations, but use them only for transforming signals (compare with [13] or with the role of these basic organs in the general automata or logical net theory in this respect.).

Now our random net has the following structure having the equal number of AND-neurons and OR-neurons (Fig. 4). The layer $\mathcal{L}_1$ comprises of $N_1/2$ AND-neurons and $N_1/2$ OR-neurons. Each neuron in this layer has its two input lines (uniformly) randomly connected to two (with repetition) out of n input points of $I\equiv\mathcal{L}_0$. We have thus a "random subnet" between the layer $\mathcal{L}_0$ with n input points and the layer $\mathcal{L}_1$ with N_1 neurons in all. The third layer $\mathcal{L}_2$ has $N_2/2$

AND-neurons and $N_2/2$ OR-neurons, each of which has its two input lines (uniformly) randomly connected to two (with repetition) out of N_1 output lines from the layer $\mathcal{L}_1$ as if these output lines were "new" input points. We thus have a new random subnet between $\mathcal{L}_0$ and $\mathcal{L}_2$. Continue this construction, then we have our random net which has $L+1$ layers, $\mathcal{L}_0, \mathcal{L}_1, \cdots, \mathcal{L}_L$. This random net in the whole has n input points and N_L output points.

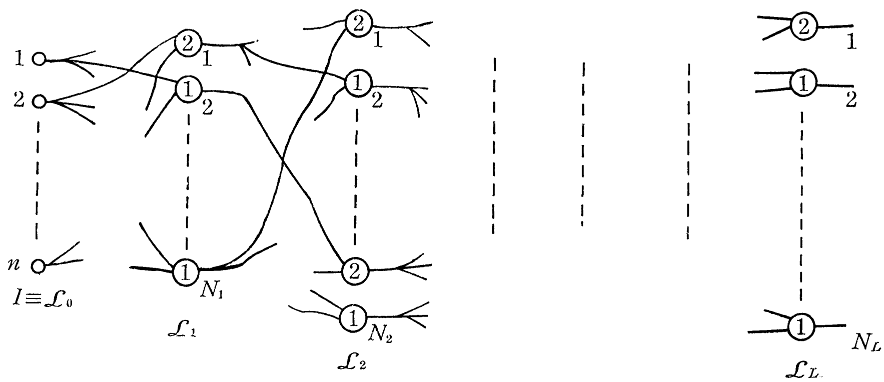

Fig. 4

Note that we must take $n \leqq N_1 \leqq N_2 \leqq \cdots \leqq N_L$ to gain (b) in the preceding section. Note also the input pattern space may be successively mapped into L pattern spaces as: $F \to G_1 \to G_2 \to \cdots \to G_L$ where G_r is the pattern space observed at the layer $\mathcal{L}_r$, $1 \leqq r \leqq L$. Since the net can be considered as the net which has "independent" L random subnets in series, by considering only one of them we might know inductively the whole net characteristics. Thus the problem is to investigate what properties about the net can be obtained if n and L become large.

The random subnet between the layers $\mathcal{L}_{r_1}$, $\mathcal{L}_{r_2}$ $(0 \leqq r_1 < r_2 \leqq L)$ will be denoted by $\mathfrak{N}(r_1, r_2)$ in the following sections.

4. An assumption on the pattern space structure.

Any input pattern $f \in F$ is completely defined by the set of all stimulated points in the set I of input points, hence a subset itself of I will sometimes be called a pattern.

We assume the following structure on the pattern space F, but this assumption never loses the generality of the pattern variety, if we take n a little larger.

Now decompose the set I into four disjoint non-empty subsets: $I = I_1 \cup I_2 \cup I_3 \cup I_4$, where

$$I_1=\{1, \cdots, n_1\},$$

$$I_2=\{n_1+1, \cdots, n_1+n_2\},$$

$$I_3=\{n_1+n_2+1, \cdots, n_1+n_2+n_3\},$$

$$I_4=\{n_1+n_2+n_3+1, \cdots, n_1+n_2+2n_3\},$$

$n_1+n_2+2n_3=n$, and n_2 will be specified by the following definitions.

If generally we denote by $\pi(A)$ the number of elements in a finite set A, then an *external* (*input*) *pattern* will be defined as a subset $J\subset I_1$ which satisfies $0\leqq n_1'\leqq\pi(J)\leqq n_1$ where n_1' is a pre-defined constant integer. In other words, any external pattern J with $0\leqq\pi(J)<n_1'$ is observed (or presented to I) only with probability 0 for each category, and any external pattern J with $n_1'\leqq\pi(J)\leqq n_1$ is observed with strictly positive probability for *some* category.

All of the n_3 input points in I_3 are always stimulated whichever external pattern is presented, and the ratio n_3/n will be called *constantly stimulating area* or simply *c.s-area*. While all of the n_3 input points in I_4 are always non-stimulated irrespective to which external pattern is presented to I, and the ratio n_3/n will be called *constantly non-stimulating area* or simply c.$\bar{\text{s}}$-area, hence c.s-area $=$c.$\bar{\text{s}}$-area$\equiv\tau$.

The subset I_2 might be called a regulating or normalizing part of I, in which every external pattern is "normalized" as follows. If an external pattern J with $\pi(J)=\nu$ is observed, input points $n_1+1, n_1+2, \cdots, n_1+(n_1-\nu)\in I_2$ are then stimulated (if $\nu=n_1$, any point in I_2 is not stimulated) and remaining points in I_2 are not stimulated, hence we shall take $n_2=n_1-n_1'$.

Therefore, as the whole, to any external pattern J there uniquely corresponds input pattern $I(J)\subset I$, such that $\pi(I(J))=\pi(J)+(n_1-\pi(J))+n_3=n_1+n_3$: constant. The ratio $\rho=(n_1+n_3)/n$ will be called *stimulating area* or simply *s-area*, and the value $1-\rho$ will be called *non-stimulating area* or simply *$\bar{s}$-area*. Note $\rho+\tau+\gamma=1$ where $\gamma=(n_1-n_1')/n$.

Now by above mentioned structure we newly define the input pattern space F as the set of all patterns each having s-area$=\rho>0$ and c.s-area$=$c.$\bar{\text{s}}$-area$=\tau>0$.

5. Changes of the stimulating area.

Before we consider how the s-area of an input pattern will change as it passes through the random net, we remark a simple (but essential in our theory) property of each AND-neuron and OR-neuron. If input signals σ_1, σ_2 received through the two input lines of a neuron are mutually independent and satisfy: $\text{Prob}(\sigma_1=1)=\text{Prob}(\sigma_2=1)=\rho$, then its output obeys $\text{Prob}(\sigma_1\wedge\sigma_2=1)=\rho^2$ for AND-neuron and $\text{Prob}(\sigma_1\vee\sigma_2=1)=2\rho-\rho^2$ for OR-neuron, hence $(1/2)\rho^2+(1/2)(2\rho-\rho^2)=\rho$ (Fig. 5).

Now consider the random subnet $\mathfrak{N}(0, 1)$. Divide the $\mathcal{L}_1$ layer into two sublayers, *AND-layer* consisting of $N_1/2$ AND-neurons and *OR-layer* consisting of $N_1/2$ OR-neurons. Suppose an arbitrary pattern $f\in F$ (which has s-area$=\rho$) is presented

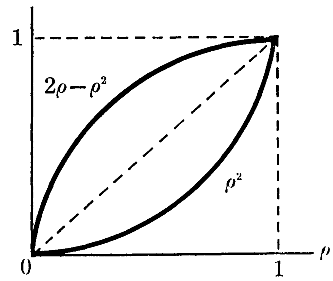

Fig. 5

to I of the net, then the corresponding pattern observed within the AND-layer can be considered as a realization of a Bernoulli trial of success probability ρ^2, because of the random connections between I and the AND-layer. In other words the number of 1 (stimulated) in the pattern within AND-layer is a realization of a random variable which obeys normal ditribution

$$\boldsymbol{N}\left(\frac{N_1}{2}\rho^2, \quad \frac{N_1}{2}\rho^2(1-\rho^2)\right),$$

when n, hence N_1, is taken sufficiently large. Similarly the number of 1 in the pattern observed within OR-layer is a realization of another random variable which obeys normal distribution

$$\boldsymbol{N}\left(\frac{N_1}{2}(2\rho-\rho^2), \quad \frac{N_1}{2}(2\rho-\rho^2)(1-2\rho+\rho^2)\right).$$

Since the above two random variables are clearly independent, the number of stimulated points in the pattern observed at the whole $\mathcal{L}_1$ layer, is only a realization of a certain random variable obeying the convolution of the above two normal distributions, hence obeying

$$\boldsymbol{N}\left(\frac{N_1}{2}\rho^2+\frac{N_1}{2}(2\rho-\rho^2), \quad \frac{N_1}{2}\rho^2(1-\rho^2)+\frac{N_1}{2}(2\rho-\rho^2)(1-2\rho+\rho^2)\right)$$
$$=\boldsymbol{N}(N_1\rho, \quad N_1\rho(1-\rho)(1-\rho+\rho^2)).$$

Therefore, when any pattern is presented to I, then the corresponding pattern observed at $\mathcal{L}_1$ has an s-area (i.e. the number of 1's in the pattern divided by N_1) which is a realization of a variable obeying $\boldsymbol{N}(\rho, (1/N_1)\rho(1-\rho)(1-\rho+\rho^2))$.

Let us denote by $\pi(f)$ the number of 1's (stimulated components) in a pattern $f\in F$, and by $\rho(f)$ the s-area of a pattern $f\in F$ (hence $\rho(f)=\rho$ for every $f\in F$). Then we may also write $\pi(g)$, $\rho(g)=\pi(g)/N_1$ for a pattern $g\in G_1$, where G_1 is the pattern space comprising of all patterns observable at $\mathcal{L}_1$. Denote, furthermore, symbolically by $h\mathcal{R}^*\boldsymbol{N}(m, \sigma^2)$ the proposition that the value h is a realization of a variable which obeys a certain distribution with the same mean m but with variance less than or equal to σ^2, when this distribution is compared with normal one $\boldsymbol{N}(m, \sigma^2)$.

Then the above argument gives that for every $g \in G_1$,

$$\pi(g)\mathcal{R}^*\boldsymbol{N}\left(N_1\rho, \frac{3}{16}N_1\right), \qquad \rho(g)\mathcal{R}^*\boldsymbol{N}\left(\rho, \frac{3}{16}\cdot\frac{1}{N_1}\right),$$

since $v(\rho)=\rho(1-\rho)(1-\rho+\rho^2)\leqq 3/16=v(1/2)$, $0\leqq\rho\leqq 1$.

Now proceed to the random net $\mathcal{N}(0, 2)$. Every input pattern $f\in F$ is first transformed by the $\mathcal{N}(0,1)$ into a pattern $g_1\in G_1$ with s-area $\rho(g_1)\mathcal{R}^*\boldsymbol{N}(\rho, (3/16)/(1/N_1))$, and then transformed by the net $\mathcal{N}(1, 2)$ into a pattern $g_2\in G_2$ which is observed at $\mathcal{L}_2$ with N_2 neurons. Since, by an analogy to the above argument, $\pi(g_2)$ is a realization of a variable obeying

$$\boldsymbol{N}(N_2\rho(g_1), \quad N_2\rho(g_1)(1-\rho(g_1))(1-\rho(g_1)+\rho^2(g_1))),$$

hence $\pi(g_2)\mathcal{R}^*\boldsymbol{N}(N_2\rho(g_1), (3/16)N_2)$, we need the following lemma.

LEMMA 1. *If a random variable x obeys $\boldsymbol{N}(m, \sigma^2)$, and if, for a fixed x, another random variable $y(x)$ obeys $\boldsymbol{N}(ax+b, \delta^2)$, a, b constants, then the mean of $y(x)$ with respect to x obeys $\boldsymbol{N}(am+b, (a\sigma)^2+\delta^2)$.*

Proof.

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}\cdot\frac{1}{\sqrt{2\pi}\,\delta}\exp\left\{-\frac{1}{2}\left(\frac{y-ax-b}{\delta}\right)^2\right\}dx$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{(a\sigma)^2+\delta^2}}\exp\left\{-\frac{1}{2}\left(\frac{y-(am+b)}{\sqrt{(a\sigma)^2+\delta^2}}\right)^2\right\}.$$

In this lemma, put $m=\rho N_1$, $\sigma^2=(3/16)N_1$, $a=N_2/N_1$, $b=0$, $\delta^2=(3/16)N_2$, then we easily have $\pi(g_2)\mathcal{R}^*\boldsymbol{N}(N_2\rho, (3/16)N_2^2(1/N_1+1/N_2))$, hence $\rho(g_2)\mathcal{R}^*\boldsymbol{N}(\rho, (3/16)(1/N_1+1/N_2))$.

The argument is now readily extended inductively to the whole net $\mathcal{N}(0, L)$. That is, the solution of the system of the following difference equations,

$$\begin{cases} m_{r+1}=\left(\dfrac{N_{r+1}}{N_r}\right)m_r, \\[2ex] \sigma_{r+1}^2=\left(\dfrac{N_{r+1}}{N_r}\right)^2\sigma_r^2+\dfrac{3}{16}N_{r+1} \end{cases}$$

with $m_1=\rho N_1$, $\rho_1^2=(3/16)N_1$, gives:

LEMMA 2. *For the input pattern space F each of whose patterns has s-area ρ, every output pattern $g \in G_L$ of the random net $\mathfrak{N}(0, L)$ has its s-area $\rho(g)$ such that*

$$\rho(g)\mathcal{R}^*\boldsymbol{N}\left(\rho, \frac{3}{16}\left(\frac{1}{N_1}+\frac{1}{N_2}+\cdots+\frac{1}{N_L}\right)\right).$$

We conclude this section with the following remark. We know that in our random net the s-areas of the patterns observed at any layer are distributed around ρ (and their deviations from ρ will be found in the next section to be arbitrarily small when n is taken large, by showing the convergence of the series $1/N_1+1/N_2+\cdots$). This is of course due to our choice of AND-neurons and OR-neurons with equal numbers among possible neurons. If instead we choose only majority neurons each of which has $2s+1$ ($s \geqq 1$) input lines and one output line, and is excited (output 1) if and only if at least $s+1$ input lines are stimulated. Therefore, when every input line is stimulated independently of others with probability ρ, the probability that the output becomes 1 is

$$u(\rho)=\sum_{\nu=0}^{s}\binom{2s+1}{s+1+\nu}\rho^{s+1+\nu}(1-\rho)^{s-\nu}$$

(see Fig. 6). When a pattern $f \in F$ is transformed into $g_1 \in G_1$ by the net $\mathfrak{N}(0, 1)$, its corresponding patterns $g_2 \in G_2, g_3 \in G_3, \cdots$ through layers $\mathcal{L}_2, \mathcal{L}_3, \cdots$ have respective s-areas $u(\rho(g_1))$, $u(u(\rho(g_1))), \cdots$ in the means. But this sequence will rapidly converge to 0 or 1 according as $\rho(g_1)<1/2$ or $\rho(g_1)>1/2$. This "degenerating" phenomenon seems to be quite inappropriate for our random net approach to pattern recognition problems. It is interesting, in this respect, to note that the Fig. 6 shows, conversely, the fundamental importance in controls of transient malfunctions of neurons in automata [13].

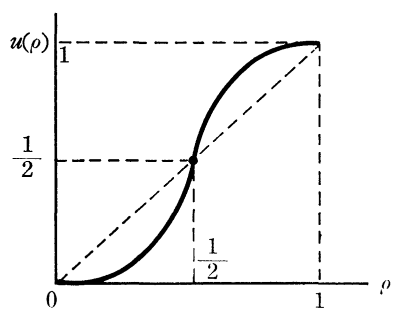

Fig. 6

6. One-to-oneness.

First consider the random subnet $\mathfrak{N}(0, 1)$, as usual, and remark that any input pattern $f \in F$ is completely specified by the subset $I_1 \cup I_2 \subset I$ of input points, since I_3 and I_4 are always stimulated and non-stimulated respectively irrespective to which pattern is presented to I.

A neuron in the layer $\mathcal{L}_1$ will be called *proper to the input point* $\nu \in I_1 \cup I_2$, if and only if one input line of the neuron is connected to ν, and another input line is connected to *some* input point in I_3 or I_4 according as this neuron is AND or OR. We have then,

LEMMA 3. *In the random subnet $\mathcal{L}_1(0,1)$ if each input point in $I_1 \cup I_2$ has at least one proper neuron in $\mathcal{L}_1$, then the pattern space F and the corresponding (transformed) pattern space G_1 can be made one-to-one correspondent.*

Proof. Suppose the neuron ν' in the layer $\mathcal{L}_1$ is a proper one to the input point ν. Then the set of such proper neurons $\{1', 2', \cdots, n_1', (n_1+1)', \cdots, (n_1+n_2)'\}$ gives the complete replica of every input pattern restricted to $I_1 \cup I_2$.

Denote by Prob(1−1) the probability of one-to-one correspondence between F and G_1, and by Prob(proper) the probability that the assumption of the lemma 3 is true, then Prob(1−1)$\geqq$Prob(proper). Now Prob(proper) can be calculated as follows. It is easily seen that the probability for an input point in $I_1 \cup I_2$ to have at least one proper neuron out of N_1 neurons in the layer $\mathcal{L}_1$ is approximately

$$1-\left(1-2\tau \cdot \frac{1}{n}\right)^{N_1},$$

where τ is c.s-area or c.s.-area defined in Section 4 according as the proper neuron is AND or OR. Since each neuron in layer $\mathcal{L}_1$ is connected to I independently of other neurons in $\mathcal{L}_1$, we have

$$\begin{aligned} \text{Prob(proper)} &\approx \left[1-\left(1-2\tau \cdot \frac{1}{n}\right)^{N_1}\right] \cdot \left[1-\left(1-2\tau \cdot \frac{1}{n}\right)^{N_1-1}\right] \cdots \\ &\quad \cdot \left[1-\left(1-2\tau \cdot \frac{1}{n}\right)^{N_1-(1-2\tau)n+1}\right] \\ &\geqq \left[1-\left(1-2\tau \cdot \frac{1}{n}\right)^{N_1-(1-2\tau)n}\right]^{(1-2\tau)n} \\ &\approx [1-e^{-(2\tau/n)N_1+2\tau(1-2\tau)}]^{(1-2\tau)n}, \end{aligned}$$

if n is taken large. Now take $N_1=(1/\tau)n \log n$ (estimated sufficiently large) then, neglecting $2\tau(1-2\tau) \leqq 1/4$,

$$\text{Prob(proper)} \geqq \left(1-\frac{1}{n^2}\right)^n \approx e^{-1/n}.$$

Take

$$N_{r+1}=\frac{1}{\tau}N_r \log N_r, \quad 0\leqq r\leqq L-1, \quad N_0=n, \tag{2}$$

and get back to lemma 2 in Section 5. Then it is easily seen that the series $1/N_1+1/N_2+\cdots$ converges to a limit which can be made arbitrarily small if n taken large. Therefore for n sufficiently large the s-area for any input pattern will be maintained to be ρ through the entire net $\mathfrak{N}(0, L)$, even if L becomes infinitely large.

By noting Fig. 5 and applying the argument in Section 5 to c.s-area and c.$\bar{\text{s}}$-area, we may also conclude that both of these areas can be maintained to be τ through the whole net. This means that by (2) the probability of one-to-one correspondence between F and the output pattern space G_L has a lower bound of order

$$e^{-1/n}\cdot e^{-1/N_1}\cdot e^{-1/N_2}\cdots e^{-1/N_{L-1}}=e^{-(1/n+1/N_1+1/N_2+\cdots+1/N_{L-1})},$$

which can be made arbitrary near 1 if n is taken large, and even if L becomes infinitely large. In summary we have

LEMMA 4. *If the number of neurons in the $\mathcal{L}_{r+1}$ is taken as $N_{r+1}=(1/\tau)N_r \log N_r$, $0\leqq r\leqq L-1$, $N_0=n$, and if n is sufficiently large, then the s-area for any input pattern is maintained to be ρ up to an arbitrary small variance through the whole net with arbitrary number of layers, and the probability of one-to-one correspondence between F and G_L can be made arbitrarily near* 1.

7. Changes of the intersecting stimulating area.

The lemma 4 in the preceding section says that, if n is large, the s-area of any input pattern is maintained always to be ρ up to an arbitrarily small variance irrespective to how many times the pattern might be transformed through the net. We therefore in the following sections may regard the s-area to be maintained precisely to be ρ in order not to make our following argument inessentially complicated.

First, again, consider the random subnet $\mathfrak{N}(0, 1)$. Denote by $\sigma(f\cap f')$ the *intersecting s-area* (simply denoted as i.s-area) of the two s-areas of arbitrary input patterns $f, f'\in F$. The object of the present section is to investigate how this i.s-area changes when the patterns pass through the whole net. The answer to this problem will be applied to our "statistical independence" problem discussed in Section 8.

We consider only two fixed input patterns f, f' whose i.s-area is $\sigma(f\cap f')=\sigma$. Now just as in Section 5 we divide the layer $\mathcal{L}_1$ into AND-layer and OR-layer. The output of any AND-neuron will be stimulated by both f and f' with probability σ^2 through the random connection operation. Hence the number of AND-neurons which are stimulated by both f and f' is a realization of a random variable which obeys $\boldsymbol{N}((N_1/2)\sigma^2, (N_1/2)\sigma^2(1-\sigma^2))$. While OR-neuron in the OR-layer may be seen, by a simple combinatorial argument, to be stimulated by both f and f' with probability $2\sigma-\sigma^2+2(\rho-\sigma)^2$ through the random connection operation. Hence the number of OR-neurons which are stimulated by both f and f' is a realization of another random variable which obeys

$$\boldsymbol{N}\left(\frac{N_1}{2}(2\sigma-\sigma^2+2(\rho-\sigma)^2), \quad \frac{N_1}{2}(2\sigma-\sigma^2+2(\rho-\sigma)^2)(1-2\sigma+\sigma^2-2(\rho-\sigma)^2)\right).$$

Therefore, if $g, g'\in G_1$ corresponds to f, f' respectively, the number $\pi(g\cap g')$ of neurons in the whole layer $\mathcal{L}_1$ which are stimulated by both f and f' is only a realization of the random variable, which is the sum of the above two independent random variables, obeying

$$\boldsymbol{N}\left(N_1(\sigma+(\rho-\sigma)^2), \quad N_1\left(\frac{1}{2}w_1(\sigma)(1-w_1(\sigma))+\frac{1}{2}w_2(\sigma)(1-w_2(\sigma))\right)\right),$$

where $w_1(\sigma)=\sigma^2$, $w_2(\sigma)=2\sigma-\sigma^2+2(\rho-\sigma)^2$. We therefore have $\pi(g\cap g')\mathcal{R}^*\boldsymbol{N}(N_1(\sigma+(\rho-\sigma)^2), (1/4)N_1)$, since $(1/2)w_1(1-w_1)+(1/2)w_2(1-w_2)\leqq 1/4$ in the square $0\leqq w_1\leqq 1$, $0\leqq w_2\leqq 1$, hence $\sigma(g\cap g')\mathcal{R}^*\boldsymbol{N}(\sigma+(\rho-\sigma)^2, (1/4)\cdot(1/N_1))$.

The mean value of i.s-area then changes as

$$\sigma_{r+1}=\sigma_r+(\rho-\sigma_r)^2, \quad 0\leqq r\leqq L-1, \quad \sigma_0=\sigma, \tag{3}$$

where σ_r is the mean i.s-area of two patterns in $\mathcal{L}_r$ corresponding to f, f', hence $\sigma_L\to\rho$ as $L\to\infty$. But the variance of the i.s-area is also less than $(1/4)(1/N_1+1/N_2+\cdots)$ with "good approximation", since the non-linear transformation (3) is well approximated by a linear transformation in the small neighbourhood of σ_r, hence lemma 1 may be applicable with an argument similar to that given in Section 5. We now have the following lemma.

LEMMA 5. *For n sufficiently large the i.s-area σ for input patterns f, f' is changed through the net by the following recursive formula with arbitrarily small deviations*:

$$\sigma_{r+1}=\sigma_r+(\rho-\sigma_r)^2, \quad 0\leqq r\leqq L-1, \quad \sigma_0=\sigma.$$

And the limit $\sigma_L\to\rho(L\to\infty)$ is independent of the choice of input patterns f, f'.

8. Independence.

Although there are defined K categories on the input pattern space F, it is clear from the fact that K is finite that the possible considerations for answering the question (a) given at the end of Section 2 should preferably be performed under the condition that one but arbitrarily fixed category is given, i.e. one probability distribution $P\in\{P^{(1)},\cdots,P^{(K)}\}$ is given. The net symbol $\mathcal{N}(0, L)$, in this section, will also be meant the set of all possible random nets (including extremely degenerated ones) constructed between layers $\mathcal{L}_0$ and $\mathcal{L}_L$. The set $\mathcal{N}(0, L)$ is called the *net space*. On the net space $\mathcal{N}(0, L)$ there is defined a natural probability measure which is introduced onto the space by the random connection operation. We thus see, for example, that the subset $\subset\mathcal{N}(0, L)$ of all possible nets each of which has the property that the first neuron in the layer $\mathcal{L}_3$ has its first input line connected to the 10th neuron and the second input line connected to the 23th neuron in the layer $\mathcal{L}_2$ has probability mass $(1/N_2)^2$, etc.. Thus on the product space $F\otimes\mathcal{N}(0, L)$ there is defined the product probability measure of two measures: P on F and the above measure on $\mathcal{N}(0, L)$. Therefore the output value $x(=0, 1)$ of any neuron contained in the layer $\mathcal{L}_L$ is a random variable defined on the space $F\otimes\mathcal{N}(0, L)$ with the corresponding probability measure.

Now take $s(2\leqq s\leqq N_L)$ neurons from $\mathcal{L}_L$ with s_1 AND-neurons and s_2 OR-neurons, $s=s_1+s_2$, and denote their output random variables by $x_1, x_2, \cdots, x_{s_1}$ for AND-neurons and by $x_{s_1+1},\cdots, x_{s_1+s_2}$ for OR-neurons. We denote, furthermore, by E the expectation operation over the pattern space F for a fixed net in $\mathcal{N}(0, L)$, and by $\mathcal{E}$ that over the net space $\mathcal{N}(0, L)$ for a fixed pattern in F. We are interested in evaluating the "average behavior" of the absolute value of

$$U^{(s)}=E[x_1\cdots x_s]-E[x_1]\cdots E[x_s]$$

over the net space, i.e. $\mathcal{E}[|U^{(s)}|]$. For this we first evaluate $\mathcal{E}[U^{(s)}]^2$, since $\mathcal{E}[|U^{(s)}|]\leqq\sqrt{\mathcal{E}[U^{(s)}]^2}$.

$$\mathcal{E}[U^{(s)}]^2=\mathcal{E}[E^2[x_1\cdots x_s]]-2\mathcal{E}[E[x_1\cdots x_s]E[x_1]\cdots E[x_s]]+\mathcal{E}[E^2[x_1]\cdots E^2[x_s]]. \tag{4}$$

We are going to evaluate three terms in (4) separately. Denoting by $x_i(f)$ the random variable x_i on the net space $\mathcal{N}(0, L)$ given pattern f, then we see

$$\begin{aligned}\mathcal{E}[E^2[x_1\cdots x_s]]&=\mathcal{E}[\sum_{f\in F}P(f)x_1(f)\cdots x_s(f)]^2\\&=\sum_{f,f'}P(f)P(f')\mathcal{E}[x_1(f)x_1(f')]\cdots\mathcal{E}[x_s(f)x_s(f')],\end{aligned}$$

since for *fixed* patterns f, f', $x_1, x_2, \cdots, x_s$ are mutually independent. By a similar fashion, we have

$$\begin{aligned}&\mathcal{E}[E[x_1\cdots x_s]E[x_1]\cdots E[x_s]]\\&=\sum_{f,f',\cdots,f^{(s)}}P(f)P(f')\cdots P(f^{(s)})\mathcal{E}[x_1(f)x_1(f')]\cdots\mathcal{E}[x_s(f)x_s(f^{(s)})]\end{aligned}$$

and

$$\mathcal{E}[E^2[x_1]\cdots E^2[x_s]]$$
$$=\sum_{f,f'} P(f)P(f')\mathcal{E}[x_1(f)x_1(f')]\cdots \sum_{f,f'} P(f)P(f')\mathcal{E}[x_s(f)x_s(f')].$$

Now the value $\mathcal{E}[x_i(f)x_i(f')]=\text{Prob}\{x_i(f)=1,\ x_i(f')=1\}$, $1\leqq i\leqq s$, is the probability that the output of the i-th neuron is stimulated by both f and f', in the case $f\neq f'$, and by only f, in the case $f=f'$. Therefore we have, from Section 7, if $f\neq f'$,

$$\mathcal{E}[x_i(f)x_i(f')]=\begin{cases}\sigma_{L-1}^2\equiv p_L(f,f') & \text{if } i\text{-th neuron is AND,}\\ 2\sigma_{L-1}-\sigma_{L-1}^2+(\rho-\sigma_{L-1})^2\equiv q_L(f,f') & \text{if } i\text{-th neuron is OR}\end{cases}$$

and, if $f=f'$,

$$\mathcal{E}[x_i(f)x_i(f')]=\begin{cases}\rho^2 & \text{if } i\text{-th neuron is AND,}\\ 2\rho-\rho^2 & \text{if } i\text{-th neuron is OR.}\end{cases}$$

But by lemma 5 it is readily seen that, if we take $p_L=\min_{f\neq f'}p_L(f,f')$ and $q_L=\min_{f\neq f'}q_L(f,f')$, then

$$p_L\uparrow \rho^2,\qquad q_L\uparrow 2\rho-\rho^2\qquad \text{as}\qquad L\to\infty.$$

Thus we have

$$\mathcal{E}[E^2[x_1\cdots x_s]]\leqq(\rho^2)^{s_1}(2\rho-\rho^2)^{s_2},$$
$$\mathcal{E}[E[x_1\cdots x_s]E[x_1]\cdots E[x_s]]\geqq p_L^{s_1}q_L^{s_2},$$
$$\mathcal{E}[E^2[x_1]\cdots E^2[x_s]]\leqq(\rho^2)^{s_1}(2\rho-\rho^2)^{s_2},$$

hence

$$\mathcal{E}[U^{(s)}]^2\leqq 2\{(\rho^2)^{s_1}(2\rho-\rho^2)^{s_2}-p_L^{s_1}q_L^{s_2}\}\to 0,\quad L\to\infty,$$

therefore

$$\mathcal{E}[|U^{(s)}|]\to 0,\quad L\to\infty.$$

Now this time conversely take s arbitrarily but fixed, and consider the net space $\mathfrak{N}(0, L)$ where $L\geqq s$. And fix a net in $\mathfrak{N}(0, L)$. For this fixed net the probability distribution P on the pattern space F defines a joint probability distribution $p(x_1=\varepsilon_1,\cdots,x_s=\varepsilon_s)$, $\varepsilon_i=0,1$, and a product probability distribution $p_1(x_1=\varepsilon_1)\cdots p_s(x_s=\varepsilon_s)$, $\varepsilon_i=0,1$, on the s output random variables $x_1, x_2,\cdots,x_s$. Denote by $\varphi(t_1,\cdots,t_s)$, $\psi(t_1,\cdots,t_s)$ the s-dimensional characteristic functions for $p(\cdots)$, $p_1(\cdot)\cdots p_s(\cdot)$ respectively, i.e.

$$\varphi(t_1,\cdots,t_s)=\sum_{\varepsilon_1,\cdots,\varepsilon_s} e^{it_1\varepsilon_1+\cdots+it_s\varepsilon_s}p(x_1=\varepsilon_1,\cdots,x_s=\varepsilon_s),$$
$$\psi(t_1,\cdots,t_s)=\sum_{\varepsilon_1,\cdots,\varepsilon_s} e^{it_1\varepsilon_1+\cdots+it_s\varepsilon_s}p_1(x_1=\varepsilon_1)\cdots p_s(x_s=\varepsilon_s).$$

Then we see that

$$\begin{aligned}&|\varphi(t_1, \cdots, t_s)-\psi(t_1, \cdots, t_s)|\\ &\leqq \sum_{\nu=2}^{\infty}\frac{1}{\nu!}\sum_{i_1, \cdots, i_\nu}|t_{i_1}\cdots t_{i_\nu}|\cdot|E[x_{i_1}\cdots x_{i_\nu}]-E[x_{i_1}]\cdots E[x_{i_\nu}]|.\end{aligned}$$

Note that both sides of this inequality are considered as random variables on the net space $\mathfrak{N}(0, L)$ and to depend on L. Then taking expectation $\mathcal{E}$ over the entire net space, we have

$$\mathcal{E}|\varphi(t_1, \cdots, t_s)-\psi(t_1, \cdots, t_s)|\leqq\alpha_L\cdot\sum_{\nu=2}^{\infty}\frac{1}{\nu!}(|t_1|+|t_2|+\cdots+|t_s|)^\nu$$

where

$$\alpha_L=\max_{1\leqq i_1, \cdots, i_\nu\leqq s}\mathcal{E}[|E[x_{i_1}\cdots x_{i_\nu}]-E[x_{i_1}]\cdots E[x_{i_\nu}]|].$$

But for arbitrary but fixed bounded domain $|t_1|+|t_2|+\cdots+|t_s|\leqq T$, we have

$$\mathcal{E}|\varphi(t_1, \cdots, t_s)-\psi(t_1, \cdots, t_s)|\leqq\alpha_L\cdot\sum_{\nu=2}^{\infty}\frac{T^\nu}{\nu!}=\alpha_L(e^T-1-T)\to 0 \quad (L\to\infty),$$

which is valid for arbitrary s.

Hence by Levy's continuity theorem [10] we have

LEMMA 6. *For a given category, the output component signals constituting output patterns in the layer $\mathcal{L}_L$ becomes mutually independent as $L\to\infty$, in the sense of the average over the net space $\mathfrak{N}(0, L)$.*

9. Conclusion.

As a summary of our study and several lemmas we have made so far, we might give the following result as a form of theorem. Since a random net which satisfies our two requirements (a), (b) stated at the end of Section 2 is *pattern recognizable*, we have

THEOREM. *In the random net which has as its all basic organs AND-neurons and OR-neurons, if n, the number of input points, and L, the number of layers, of the net, are taken sufficiently large, then the random net becomes pattern recognizable.*

We note lastly that in actual situations, however, it seems plausible that the value $\sum_{f\in F}P^2(f)$ becomes rapidly 0 as n taken large. Considering this in estimating (4), it is seen without difficulty that the "almost independence" may be obtained before L becomes large, since i.s-areas for pairs of patterns will soon be almost identical. (Refer to my forthcoming paper which is to appear in the Reports.)

ACKNOWLEDGMENT. The author wishes to thank Prof. K. Kunisawa deeply for his valuable suggestions and criticisms. His constant encouragement through several years made the present study possible.

REFERENCES

[1] ARKADEV, A. G., AND E. M. BRAVERMAN, Teaching computers to recognize patterns. Academic Press (1967).

[2] BLOCK, H. D., The perceptron, a model for brain functioning. Rev. Modern Phys. **34** (1962) 123–135.

[3] FANO, R. M., Transmission of information. 1961, M. I. T. Press.

[4] FEINSTEIN, A., Foundations of information theory. McGraw-Hill (1958).

[5] GLUSHKOV, V. M., Teaching theory of one class of perceptrons. Intern. Journ. Computer Math. **1** (1965) 199–220.

[6] GREENBERG, H. J., AND A. G. KONHEIM, Linear and nonlinear methods in pattern classification. IBM Journ. Research and Development **8** (1964) 299–307.

[7] HORIBE, Y., On an adaptive process for learning finite patterns. Kōdai Math. Sem. Rep. **19** (1967) 43–52.

[8] HORIBE, Y., On a learning network. Kōdai Math. Sem. Rep. **19** (1967) 95–102.

[9] HORIBE, Y., On a state identification by learning. Kōdai Math. Sem. Rep. **19** (1967) 470–473.

[10] KUNISAWA, K., Modern probability theory. 12th Edition, Iwanami Shoten (1963). (in Japanese).

[11] MINSKY, M., Steps toward artificial intelligence. Proc. IRE Jan. (1961) 8–30.

[12] NOVIKOFF, A. B. J., On convergence proofs for perceptrons. Proc. Symp. on Math. Theory of Automata (1962), 615–622, Polytechnic Press.

[13] VON NEUMANN, J., Probabilistic logics and the synthesis of reliable organisms from unreliable components. In *Automata Studies*, Ed. C. E. Shannon and J. McCarthy, Princeton University Press, (1956).

DEPARTMENT OF MATHEMATICS,
TOKYO INSTITUTE OF TECHNOLOGY.