46. *On the Theory of Continuous Information.*

By Hirosi ITÔ.
Department of Applied Physics, Faculty of Engineering, Osaka University.
(Comm. by M. MASIMA, M.J.A., April 12, 1952.)

1. Introduction. The information theory was first expounded by C.E. Shannon and is now attacked by some authors. I suppose, however, it is sufficiently completed in the case of the discrete system. But it seems to me somewhat vaguely in the continuous system. The most difficulty is that, continuous information lacks the unit of measure. Gabor[1] and some authors have noticed that there was such a relation between the bandwidth and the time duration as the uncertainty relation in quantum mechanics. I think, this relation plays an essential rôle in continuous information.

The definition of the entropy of the system will be most properly defined as the measure of the uncertainty which this system owns, or the power to transmit the information. In other words, information will be defined as the measure of the decrease of uncertainty. When we have some information about the system in question, the uncertainty of the system must be decreased. Assuming the certainty to correspond to the zero uncertainty, the measure of uncertainty must have difinite sign. In any case, we could not have more information than that which the first *a priori* uncertainty has. Therefore, I think, we can not consider the negative value of the entropy when defining the entropy to have the positive sign[2].

2. The entropy produced by the linear transformations. Now we consider the simplest case where the number of the random variables is only one, and the following transformation from x to a new variable y:

$$y=ax, \tag{2.1}$$

where a is a positive constant.

The entropy of the system is defined by Shannon as follows,

$$H(x)=-\int p(x)\log p(x)\;dx. \tag{2.2}$$

The new entropy of the system which is induced by (2.1) is given by

---

1) Phil. Mag. 41 (1949) p. 1161.

2) Rather we may consider the negative value of the information which increases the uncertainty of the system than the first expected one. For example, if we had the exact information which denied any law in physics, we should give up the every concepts that were introduced by this law and the uncertainty would increase.

$$H(y)=H(x)+\log a. \tag{2.3}$$

The relation (2.1) may be considered to have the following meaning that, if we determine the domain of the random variable beforehand, the new random variable will have the different domain from that of x. Corresponding to $a \gtrless 1$, y has the larger or smaller domain than that of x. The measurement to obtain some information which is whatever, restricts the domain in which this random variable varies. In other words, the information which will be expected to decrease the uncertainty must limit the domain of the random variable in question.

The value of a which is larger than **1** gives the larger domain for the new variable than that for the old one. This tells us that we must sacrify the certainty in order to obtain the larger value than the original one, as in the case of the amplifier. If we continue the process of the transformation (2.1), we have after the nth transformation

$$H(x_n)=H(x)-n\log a, \tag{2.4}$$

where $a>\mathbf{1}$. Note that in this case we must use the following transformation

$$y=\frac{1}{a}x. \tag{2.5}$$

The second term of (2.4) will be negative infinity as n grows infinitely large and the information I will be positive infinity by the relation

$$I=H(x)-H(x_n)=n\log a. \tag{2.6}$$

This seems queer because we could obtain infinite information by repeating the measurement. Whatever large *a pri-ori* value we may assume about $H(x)$, the value of $H(x_n)$ must take the negative value in this process. On the other hand, $H(x)$ will become infinity when we stick to the view-point where the entropy must be positive essentially. Both cases contradict to the preceding discussion. Any message signal has two random variables; its time duration and the bandwidth as in the case where the description of the motion of a particle needs its position vector and conjugate momentum.

When we consider the two independent variables, the entropy will be

$$H(x, y)=H(x)+H(y),$$

where x denotes the position vector of the particle or the frequency per second and y does the momentum or the time. Let us now consider the signal which has the bandwidth σ_1 and the time duration σ_2. If we assume the Gauss' distributions for both time and frequency, σ_1 and σ_2 are defined by

$$\sigma_1^2=\overline{(x-\bar{x})^2} \quad \sigma_2^2=\overline{(y-\bar{y})^2}, \tag{2.7}$$

and have the important relation from the Fourier transformation.

$$\sigma_1\sigma_2=1. \tag{2.8}$$

When we assume $\bar{x}$ and $\bar{y}$ to be zero for simplicity, the entropy of this signal is given by

$$\begin{aligned} H(x, y) &= 2\left(\log\sqrt{2\pi}+\frac{1}{2}\right)+\log\sigma_1\sigma_2 \\ &= 2\left(\log\sqrt{2\pi}+\frac{1}{2}\right). \end{aligned} \tag{2.9}$$

If we transform x and y by

$$\xi=ax \quad \text{and} \quad \eta=by, \tag{2.10}$$

the new entropy $H(\xi, \eta)$ is given by

$$\begin{aligned} H(\xi, \eta) &= H(\xi)+H(\eta) \\ &= H(x)+H(y)+\log a+\log b. \end{aligned}$$

But we have the following relation when the type of the distribution does not change,

$$\overline{x^2}=\frac{\overline{\xi^2}}{a^2}=\sigma_1^2, \quad \overline{y^2}=\frac{\overline{\eta^2}}{b^2}=\sigma_2^2. \tag{2.11}$$

As ξ and η have also Gauss' distributions, we have

$$\overline{\xi^2}\,\overline{\eta^2}=a^2b^2\sigma_1^2\sigma_2^2=a^2b^2=1.$$

Consequently, $b=\dfrac{1}{a}$.

By this relation, the new entropy $H(\xi, \eta)$ is

$$\begin{aligned} H(\xi, \eta) &= H(x)+H(y)+\log a-\log a \\ &= H(x)+H(y)=2\left(\log\sqrt{2\pi}+\frac{1}{2}\right). \end{aligned} \tag{2.12}$$

Therefore, we may conclude that while the partial entropy $H(x)$ or $H(y)$ will take negative value by some transformation, total entropy $H(x, y)$ must keep up the positive value (2.12), which may be considered as zero when assuming the uniform distribution.

Any entropy of the system can be defined by

$$H(x, y)=\log(\alpha\beta)+2\left(\log\sqrt{2\pi}+\frac{1}{2}\right) \tag{2.13}$$

where α is the permissible a pri-ori domain of x and β is that of y. The extension to n variables is easily obtained. Let us consider the $2n$ linear relations

$$\begin{aligned} \xi_i &= \sum_{k=1}^{n} a_{ik} x_k \\ \eta_i &= \sum_{k=1}^{n} b_{ik} y_k. \end{aligned} \qquad (i=1, 2, \dots n) \tag{2.14}$$

The new entropy $H(\xi_1\xi_2 \dots \xi_n;\ \eta_1\eta_2 \dots \eta_n)$ is

$$\begin{aligned} &H(\xi_1\xi_2 \dots \xi_n;\ \eta_1\eta_2 \dots \eta_n) \\ &\quad = H(x_1x_2 \dots x_n;\ y_1y_2 \dots y_n) \\ &\qquad -\log \frac{\partial(x_1x_2 \dots x_n)}{\partial(\xi_1, \xi_2 \dots \xi_n)} - \log \frac{\partial(y_1y_2 \dots y_n)}{\partial(\eta_1\eta_2 \dots \eta_n)} \\ &\quad = H(x_1x_2 \dots x_n;\ y_1y_2 \dots y_n) \\ &\qquad + \log |a_{ik}| + \log |b_{ik}| \ , \end{aligned} \tag{2.15}$$

where $|a_{ik}|$ and $|b_{ik}|$ are the determinants of this transformations. If the values of these determinants are equal to **1** respectively, entropy does not change its value by this transformation. By means of this fact, we can obtain the diagonal representation as follows. Let us consider the linear transformations

$$\xi = Ax, \quad \xi = S\xi' \text{ and } x = Sx',$$

where ξ, ξ', x and x' are considered as the vectors which have n components respectively and S is the orthogonal transformation which transforms the matrix (a_{ik}) into diagonal form, and its determinant has the value **1.**

$$\xi' = S^{-1}ASx' = \lambda x' \tag{2.16}$$

is the relation between new variables x' and ξ'. About the other variable η, we have also the similar relation

$$\eta' = T^{-1}BTy' = \mu y'. \tag{2.17}$$

We have

$$H(\xi_1\xi_2 \dots \xi_n;\ \eta_1\eta_2 \dots \eta_n) = H(\xi_1'\xi_2' \dots \xi_n';\ \eta_1'\eta_2' \dots \eta_n')$$

because of the values of $\det(S)$ and $\det(T)$ to be one. Now we obtain

$$\begin{aligned} &H(\xi_1\xi_2 \dots \xi_n;\ \eta_1\eta_2 \dots \eta_n) \\ &\quad = H(x_1x_2 \dots x_n;\ y_1y_2 \dots y_n) \\ &\qquad + \sum_{i=1}^{n} \log \lambda_i + \sum_{i=1}^{n} \log \mu_i. \end{aligned} \tag{2.18}$$

This result shows that the information is represented by the logarithm of the diagonal matrix.

3. **The entropy of ensembles.** Shannon has given the expres-

sion[3] of a function which is limitted to the band from 0 to W cycles per second and to a time duration T as follows.

$$f(t)=\sum_{-n}^{n} X_k \frac{\sin \pi (2Wt-k)}{\pi (2Wt-k)} \tag{3.1}$$

where

$$X_k=f\left\{\frac{k}{2W}\right\},$$

and n is defined by the time duration and the frequency band as $n=2WT$.

Let us consider the average energy of this function. If T is very large, this will be given approximately by

$$\bar{E}=\frac{1}{T}\int_0^T f(t)^2 dt=\frac{1}{4WT}\sum_{k=-n}^{n} X_k^2. \tag{3.3}$$

When we consider the average amplitude r^2; which is

$$2nr^2=\sum_{-n}^{n} X_k^2, \tag{3.4}$$

(3.4) becomes

$$\bar{E}\frac{n}{2WT}r^2=r^2$$

by the definition of n.

This result shows the ergodic property of this random series roughly.

r^2 may be considered as the average power of the elementary signal by which the total signal is constructed. Taking account of this fact, the entropy of the system introduced by (2.14) will be given by

$$H(x, y)=\log n+2\left(\log \sqrt{2\pi}+\frac{1}{2}\right). \tag{2.5}$$

This result will be considered as the quantization of the Shannon's result, which is

$$H=\frac{1}{2}\log N+\left(\log\sqrt{2\pi}+\frac{1}{2}\right),$$

where N is the average power of the white noise. The result which is given by (3.7) will be mainfest, if the average power of the elementary signal is counted as 1[4].

3) Mathematical Theory of Communication (1949) P. 53.

4) In order to obtain the same result as that by Shannon, it seems preferable to use the definition

$$H(x, y)=\frac{1}{2}(H(x)+H(y)).$$