166. On "Amount of Information"

By Kōmei SUZUKI

(Comm. by K. KUNUGI, M.J.A., Dec. 13, 1956)

We have pursued the mathematical characters of the "amount of information" propounded by Norbert Wiener[1] and Claude E. Shannon.[3] And considering it, we were always to specify at least one partition of the probability space.[2]

Thus we have designed to discuss it without discriminating the types of probability distributions.

§ 1. $\log_2\frac{1}{P}$ and $P\log_2\frac{1}{P}$

As the "state" (A) considered about some object is defined by the possible K cases, by the K attributes or more generally by the number K, the capability of the "source" for causing the state (A) to occur is measured by $\log_2 K$.

Mathematically this corresponds with the fact that the subsets of the finite set consisting of M elements are 2^M in all, while K and $\log_2 K$ correspond to 2^M and M respectively.

Further, if the capability is expressed statistically, the probability P in which the state (A) occurs is used for K or 2^M and it will be measured by $\log_2\frac{1}{P}$ avoiding negative.

This quantity $\log_2\frac{1}{P}$ will be called the amount of information for the source due to the state (A) that happens.

Hence putting conveniently $0\log\frac{1}{0}=0$, we have easily the following proposition.

(1.1) $P\log\frac{1}{P}$, $P\geq 0$, is concave, and attains its maximum at $P=\frac{1}{e}$, thus it is an increasing function in $0\leq P\leq\frac{1}{e}$, (e is the base of the natural logarithm); and for $\varDelta P\geq 0$ and $1\geq P_k$, $P_k+(-1)^{k+1}\varDelta P\geq 0$, $k=1,2$,

$$(P_1+\varDelta P)\log\frac{1}{P_1+\varDelta P}-(P_2-\varDelta P)\log\frac{1}{P_2-\varDelta P}\gtreqqless P_1\log\frac{1}{P_1}+P_2\log\frac{1}{P_2}$$

accordingly as $|(P_1+\varDelta P)-(P_2-\varDelta P)|\gtreqqless|P_1-P_2|$.

Further, as a great number of states are sometimes supposed to exist in observing some object, the average amount of information for the capability of the source due to the sequence of states such as $\{A_i\,|\,i=0,1,2,\cdots\}$, where (A_i) is defined by the probability P_i,

$P_i \geq 0$, $\sum_{i=0}^{\infty} P_i = 1$, may also be taken into consideration and represented as, $H = \sum_{i=0}^{\infty} P_i \log \frac{1}{P_i}$. And this quantity H is usually called the amount of information for the source due to the sequence of states $\{A_i \mid i=0, 1, 2, \cdots\}$.

§ 2. The partition of the probability space

By $(R, \mathfrak{X}, \lambda)$ we denote, as usual, the probability space (or the probability distribution), i.e. R is a non-empty set, $\mathfrak{X}$ an additive class of subsets in R and λ is non-negative measure defined for the set of $\mathfrak{X}$, and $\lambda(R)=1$.

A partition Λ of $(R, \mathfrak{X}, \lambda)$ shall be defined as follows:

$$\Lambda:\ R = \bigcup_{i=0}^{\infty} A_i;\ A_i \in \mathfrak{X},\ A_i \cap A_j = 0,\ i \neq j.$$

For a given partition Λ, we put $\lambda(A_i) = P_i$ and $H = H(\lambda; \Lambda) = \sum_{i=0}^{\infty} P_i \log \frac{1}{P_i}$.

Thus we may consider the above as an amount of information for a partition Λ.

When a partition Λ_1 is given as follows:

$$\Lambda_1:\ R = \bigcup_{j=0}^{\infty} B_j;\ B_j \in \mathfrak{X},\ B_j \cap B_{j'} = 0,\ j \neq j',$$

$$B_j = \bigcup_{\nu} A_{i_\nu},\ \nu = 1, 2, \cdots, k_j,\ A_{i_\nu} \in \{A_i \mid i = 0, 1, \cdots\},$$

we have

$$P_j = \lambda(B_j) = \lambda(\bigcup_{\nu} A_{i_\nu}) = \sum_{\nu} P_{i_\nu}$$

and

$$P_j \log \frac{1}{P_j} \leq \sum_{\nu=1}^{k_j} P_{i_\nu} \log 1/P_{i_\nu}$$

thus

$$H(\lambda; \Lambda_1) = \sum_{j=0}^{\infty} P_j \log \frac{1}{P_j} \leq \sum_{j=0}^{\infty} \sum_{\nu=1}^{k_j} P_{i_\nu} \log \frac{1}{P_{i_\nu}} = H(\lambda; \Lambda).$$

This gives the following proposition.

(2.1) The unification of some components of a partition causes the decrease of the amount of information.

§ 3. Convergence of $H = \sum_{i=0}^{\infty} P_i \log \frac{1}{P_i}$

(3.1) If $\lim_{i \to \infty} \frac{P_{i+1}}{P_i} < 1$, then $H = \sum_{i=0}^{\infty} P_i \log \frac{1}{P_i}$ is convergent.

From $\lim_{i \to \infty} \frac{P_{i+1}}{P_i} < 1$, we could select a positive number ρ_0 and a sufficiently large integer n_0 such that $0 < \frac{P_{i+1}}{P_i} \leq \rho_0 < 1$ for $i \geq n_0$, and then, referring to (1.1), we have

$$0<\sum_{i=m}^{n} P_i \log \frac{1}{P_i} \leq \sum_{\nu=m-n_0}^{n-m_0} \rho_0^\nu P_{n_0} \log (1/P_{n_0}\rho_0^\nu)$$

$$\leq P_{n_0} \sum_{\nu=m-n_0}^{n-n_0} \rho_0^\nu \Big(\log \frac{1}{P_{n_0}} + \nu \log \frac{1}{\rho_0} \Big)$$

for $0 \leq P_i \leq \frac{1}{e}$ and for any n, m such that $n > m \geq n_0$.

(3.2) If $\lim_{i\to\infty} \frac{P_{i+1}}{P_i} = 1$, then $H = \sum_{i=0}^{\infty} P_i \log \frac{1}{P_i}$ converges or diverges concurrently with the series $\sum_{\nu=0}^{\infty} \nu \eta_\nu$,

where $\eta_\nu = \sum_{e^{-1}(1-e^{-1})^{i_\nu} > p_i \geq e^{-1}(1-e^{-1})^{i_\nu+1}} P_i$.

Let m_ν be the number of terms of η_ν, from the definition of η_ν,

$$e^{-1} \sum_{\nu=0}^{\infty} m_\nu r^{i_\nu} \geq \sum_{\nu=0}^{\infty} \eta_\nu \geq e^{-1} \sum_{\nu=0}^{\infty} m_\nu r^{i_\nu+1}, \quad r = 1 - e^{-1},$$

then the series

$$e^{-1} \sum_\nu m_\nu r^{i_\nu}, \quad e^{-1} \sum_\nu m_\nu r^{i_\nu+1}$$

are convergent since $\sum_\nu \eta_\nu$ is convergent. And for a sufficiently large number ν_0, $i_{\nu+1} = i_\nu + 1$, $\nu \geq \nu_0$.

Thus we shall obtain two integers ν_1, ν_2 $(\geq \nu_0)$ such that

$$(1-e^{-1}) \sum_{\mu=1}^{\infty} \eta_{\nu_1+\mu} \Big(1 + (i_{\nu_1} + \mu + 1) \log \frac{e}{e-1} \Big)$$

$$\leq \sum_{i=N}^{\infty} P_i \log \frac{1}{P_i} \leq \frac{e}{e-1} \sum_{\mu=1}^{\infty} \eta_{\nu_2+\mu} \Big(1 + (i_{\nu_2} + \mu) \log \frac{e}{e-1} \Big)$$

for a sufficiently large integer N. Hence the proposition (3.2) may be proved.

§ 4. Comparison between two amounts of information

Henceforth we assume that the series

$$H(\lambda; A) = \sum_{i=0}^{\infty} P_i \log \frac{1}{P_i}$$

and

$$H(\lambda_1; A) = \sum_{i=0}^{\infty} (P_i + \varDelta P_i) \log 1/(P_i + \varDelta P_i)$$

are convergent; and that there exist some positive numbers α, k such that

$$(4.1) \qquad -1+\alpha \leq \frac{\varDelta P_i}{P_i} \leq k, \quad i = 0, 1, 2, \cdots, \ 0 < \alpha < 1.$$

Thus we see that the series $\sum_{i=0}^{\infty} \varDelta P_i \log \frac{1}{P_i}$ and $\sum_{i=0}^{\infty} \varDelta P_i \log \frac{1}{P_i + \varDelta P_i}$ are also absolute convergent. Therefore we have as follows:

$$\varDelta H = H(\lambda_1; A) - H(\lambda; A)$$

$$= \sum_{i=0}^{\infty} \Big((P_i + \varDelta P_i) \log \frac{1}{P_i + \varDelta P_i} - P_i \log \frac{1}{P_i} \Big)$$

$$(4.1.1)\qquad =\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i}-\sum_{i=0}^{\infty}(P_i+\Delta P_i)\log\frac{P_i+\Delta P_i}{P_i}$$

$$=\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i+\Delta P_i}+\sum_{i=0}^{\infty}P_i\log\frac{P_i}{P_i+\Delta P_i}.$$

And on the other hand it is always true that

$$(4.1.2)\qquad \sum_{i=0}^{\infty}\Delta P_i=0,\quad \text{(absolute convergent).}$$

Thus we get easily

$$\sum_{i=0}^{\infty}(P_i+\Delta P_i)\log\frac{P_i+\Delta P_i}{P_i}=\sum_{i=0}^{\infty}\left(\frac{(\Delta P_i)^2}{P_i}-(P_i+\Delta P_i)\int_0^{\Delta P_i/P_i}(t/(1+t))dt\right)\geq 0$$

(4.1.4) and

$$\sum_{i=0}^{\infty}P_i\log\frac{P_i}{P_i+\Delta P_i}=\sum_{i=0}^{\infty}P_i\int_0^{\Delta P_i/P_i}(t/(1+t))dt\geq 0.$$

And finally we get

$$(4.2)\qquad \sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i+\Delta P_i}\leq \Delta H\leq\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i}.$$

From this result the following propositions are clear.

$$\Delta H\leq 0\rightarrow\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i+\Delta P_i}\leq 0$$

(4.2.1) or

$$\Delta H\geq 0\rightarrow\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i}\geq 0$$

and

$$\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i+\Delta P_i}\geq 0\rightarrow\Delta H\geq 0$$

or

$$\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i}\leq 0\rightarrow\Delta H\leq 0.$$

§ 5. $\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i}$ and $\sum_{i=0}^{\infty}\Delta P_i\log\frac{1}{P_i+\Delta P_i}$

According to $\Delta P_i<0$ or $\Delta P_i\geq 0$, we classify the sequence $\{(P_i, \Delta P_i)\mid i=0,1,2,\cdots\}$ into two groups;

1) group α: $\{(P_{\alpha_\mu}, -\Delta P_{\alpha_\mu})\mid \Delta P_{\alpha_\mu}>0,\ \mu=1,2,\cdots\}$

2) group β: $\{(P_{\beta_\nu}, +\Delta P_{\beta_\nu})\mid \Delta P_{\beta_\nu}\geq 0,\ \nu=1,2,\cdots\}$.

Hence we may consider the mass $\Delta P_{\mu\nu}\geq 0$ for a pair (μ,ν), $\mu,\nu=1,2,\cdots$. The masses $\Delta P_{\mu\nu}$, $\mu,\nu=1,2,\cdots$, are interpreted, for example, as the ones remove from A_{α_μ} to A_{β_ν} when by the probability distribution, which is shifting in the passage of time, the types represented as $(R,\mathfrak{X},\lambda)$ and $(R,\mathfrak{X},\lambda_1)$ are taken at the time t and t_1 respectively.

Then it follows that if we put

$$\Delta P_{\alpha_\mu}=\sum_{\nu=1}^{\infty}\Delta P_{\mu\nu}\quad\text{(absolute convergent),}$$

$\varDelta P_{\beta_\nu}$ are to be defined as $\varDelta P_{\beta_\nu}=\sum_{\mu=1}^{\infty}\varDelta P_{\mu\nu}$
since $\varDelta P_{\mu\nu}\leq \varDelta P_{\alpha_\mu}$ and $\sum_{\mu=1}^{\infty}\varDelta P_{\alpha_\mu}$ is absolute convergent.

Thus

$$(5.1)\qquad \begin{aligned}\sum_{i=0}^{\infty}\varDelta P_i\log\frac{1}{P_i}&=\sum_{\nu=1}^{\infty}\Big(\varDelta P_{\beta_\nu}\log\frac{1}{P_{\beta_\nu}}\Big)-\sum_{\mu=1}^{\infty}\Big(\varDelta P_{\alpha_\mu}\log\frac{1}{P_{\alpha_\mu}}\Big)\\&=\sum_{\mu,\nu}\varDelta P_{\mu\nu}\Big(\log\frac{1}{P_{\beta_\nu}}-\log\frac{1}{P_{\alpha_\mu}}\Big).\end{aligned}$$

If $\varDelta P_{\mu\nu}>0$, then

$$(5.1.1)\qquad \varDelta P_{\mu\nu}\Big(\log\frac{1}{P_{\beta_\nu}}-\log\frac{1}{P_{\alpha_\mu}}\Big)\begin{cases}>0\leftrightarrow P_{\beta_\nu}<P_{\alpha_\mu}\\=0\leftrightarrow P_{\beta_\nu}=P_{\alpha_\mu}\\<0\leftrightarrow P_{\beta_\nu}>P_{\alpha_\mu}.\end{cases}$$

Similarly

$$(5.2)\qquad \sum_{i=0}^{\infty}\varDelta P_i\log\frac{1}{P_i+\varDelta P_i}=\sum_{\mu\ \nu}\varDelta P_{\mu\nu}\Big(\log\frac{1}{P_{\beta_\nu}+\varDelta P_\beta}-\log\frac{1}{P_{\alpha_\mu}-\varDelta P_{\alpha_\mu}}\Big)$$

and if $\varDelta P_{\mu\nu}>0$, then

$$(5.2.1)\qquad \varDelta P_{\mu\nu}\Big(\log\frac{1}{P_{\beta_\nu}+\varDelta P_{\beta_\nu}}-\log\frac{1}{P_{\alpha_\mu}-\varDelta P_{\alpha_\mu}}\Big)\begin{cases}>0\leftrightarrow P_{\beta_\nu}+\varDelta P_{\beta_\nu}<P_{\alpha_\mu}-\varDelta P_{\alpha_\mu}\\=0\leftrightarrow P_{\beta_\nu}+\varDelta P_{\beta_\nu}=P_{\alpha_\mu}-\varDelta P_{\alpha_\mu}\\<0\leftrightarrow P_{\beta_\nu}+\varDelta P_{\beta_\nu}>P_{\alpha_\mu}-\varDelta P_{\alpha_\mu}.\end{cases}$$

Therefore if $P_{\beta_\nu}>P_{\alpha_\mu}$ for all μ,ν, referring to (4.2), we have $\varDelta H<0$; and also if $P_{\beta_\nu}+\varDelta P_{\beta_\nu}<P_{\alpha_\mu}-\varDelta P_{\alpha_\mu}$ for all μ,ν, $\varDelta H>0$.
Thus we may infer that
(5.3) the concentration or the divergence of masses causes to decrease or increase the amount of information respectively.

References

1) Wiener, N.,: Cybernetics (1948).
2) Darrow, C. K.,: Statistical theories of matter radiation and electricity, Phys. Rev. (1929).
3) Shannon, E. C., and Weaver, W.,: The Mathematical Theory of Communication, University of Illinois Press (1949).