JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 97, 355–373 (1983)

Information of Relative Pairwise Comparisons

TETURO KAMAE

Department of Mathematics, Osaka City University, Sugimoto-cho, Osaka, 558 Japan

Submitted by G.-C. Rota

We study the amount of information obtained by a set of relative pairwise comparisons. From this point of view, two kinds of entropy are introduced for graphs.

1. INTRODUCTION

It is an interesting problem to ask the amount of information obtained by relative pairwise comparisons. For example, consider 3 balls σ, τ and ξ with weights x_σ, x_τ and x_ξ. Compare the weights of σ and τ to know which is heavier. Since we have no prior information on the weights, we expect 2 kinds of results $x_\sigma < x_\tau$ and $x_\sigma > x_\tau$ with same probability 1/2. Here, we assume that the probability of $x_\sigma = x_\tau$ is 0. Thus, the comparison has entropy 1 in the binary base. Now, consider the set of 2 comparisons σ to τ and τ to ξ. Since we have no information on the weights, we can assign same probability 1/6 to any of the 6 cases $x_\sigma < x_\tau < x_\xi$, $x_\sigma < x_\xi < x_\tau$, $x_\tau < x_\sigma < x_\xi$, $x_\tau < x_\xi < x_\sigma$, $x_\xi < x_\sigma < x_\tau$ and $x_\xi < x_\tau < x_\sigma$. Therefore, we have 4 kinds of results $x_\sigma < x_\tau < x_\xi$, $x_\sigma < x_\tau > x_\xi$, $x_\sigma > x_\tau < x_\xi$ and $x_\sigma > x_\tau > x_\xi$ with probability 1/6, 2/6, 2/6 and 1/6, respectively. Thus, the entropy of the comparisons is

$$
\begin{aligned}
&-\tfrac{1}{6}\log_2\tfrac{1}{6} - \tfrac{2}{6}\log_2\tfrac{2}{6} - \tfrac{2}{6}\log_2\tfrac{2}{6} - \tfrac{1}{6}\log_2\tfrac{1}{6} \\
&\doteqdot 1.918
\end{aligned}
$$

in the binary base. This set of comparisons can be represented as the graph with 3 vertices σ, τ, ξ and 2 edges $\{\sigma, \tau\}$ and $\{\tau, \xi\}$. The entropy of the graph is defined as the entropy of the corresponding comparisons as above. This entropy will be called the combinatorial entropy of the graph.

In other words, let G be a *graph* on a finite set Σ. That is to say that G is a family of two points subsets of Σ. Denote $\Sigma(G) = \bigcup_{\alpha \in G} \alpha$. An *orientation*

0022-247X/83 $3.00

on G is a pair $\theta=(\theta_1,\theta_2)$ of mappings $\theta_i: G\to\Sigma$ $(i=1,2)$ such that $\alpha=\{\theta_1(\alpha),\ \theta_2(\alpha)\}$ for any $\alpha\in G$. Throughout this paper, we fixed $\Sigma\neq\phi$, $G\neq\phi$ and an orientation θ on G as above. Let $X_\Sigma=\{X_\sigma;\sigma\in\Sigma\}$ be a family of independent and identically distributed random variables with the standard normal distribution $N(0,1)$. Define

$$Y_\alpha=\varphi(X_{\theta_2(\alpha)}-X_{\theta_1(\alpha)})$$

for any $\alpha\in G$, where

$$\varphi(x)=\begin{cases}1 & x>0\\ 0 & x=0\\ -1 & x<0.\end{cases}$$

Then, the *combinatorial entropy* $H(G)$ of G is defined as the Shanon's entropy (in the natural base) of the family of random variables $Y_G=\{Y_\alpha;\alpha\in G\}$. It is clear that the definition is independent of the orientation θ on G. In general, the relative information $I(X,Y)$ between random variables X and Y is defined by

$$I(X,Y)=\int\log\frac{dP_{X,Y}}{d(P_X\times P_Y)}\,dP_{X,Y},$$

where $P_{X,Y}$, P_X, P_Y are distributions of random variables (X,Y), X and Y, respectively. Then it holds that

$$H(G)=I(X_\Sigma,Y_G).$$

Let $\Xi\subset\Sigma$. We define the *combinatorial entropy* $H^\Xi(G)$ of G *in* Ξ by

$$H^\Xi(G)=I(X_\Xi,Y_G).$$

This definition is independent of the orientation θ on G. It should be remarked that the assumption that each X_σ is normally distributed is not essential at all. It may be replaced by any nonatomic distribution without changing the notion of the combinatorial entropy.

Another approach to the problem is that we assume that our observation in each comparison is the difference between the pair with a normally distributed error instead of the order. Define a family of random variables $Z_G=\{Z_\alpha;\alpha\in G\}$ by

$$Z_\alpha=X_{\theta_2(\alpha)}-X_{\theta_1(\alpha)}+\varepsilon_\alpha,$$

where $\{\varepsilon_\alpha;\alpha\in G\}$ is a family of independent and identically distributed random variables which is independent of X_Σ with distribution $N(0,1/\lambda)$ for some $\lambda > 0$. We define the *λ-Gaussian entropy* $H_\lambda(G)$ of G and the *λ-Gaussian entropy* $H_\lambda^{\Xi}(G)$ *of* G *in* $\Xi \subset \Sigma$ by

$$H_\lambda(G) = I(X_\Sigma, Z_G)$$

and

$$H_\lambda^{\Xi}(G) = I(X_\Xi, Z_G),$$

respectively. These definitions are independent of the orientation θ on G. A subgraph K of G (i.e., $K \subset G$) is called *sufficient at* σ in the combinatorial or Gaussian sense if $H^\sigma(K)$ $(=H^{\{\sigma\}}(K)) = H^\sigma(G)$ or $H_\lambda^\sigma(K) = H_\lambda^\sigma(G)$ for any $\lambda > 0$, respectively, where $\sigma \in \Sigma$.

Our aim is to solve the following problems:

Problem I. Given a class of graphs. Find a graph which maximizes the combinatorial or λ-Gaussian entropy for any $\lambda > 0$ in the class.

Problem II. Given a graph G and $\sigma \in \Sigma(G)$. Find the minimum sufficient subgraph of G at σ in the combinatorial or Gaussian sense.

Problem III. Find relations between the combinatorial entropy and the Gaussian entropy.

In this paper, we give partial solutions to the problems.

2. Problem I

Let $\Xi \subset \Sigma$. We define matrices $A_G(\lambda) = (a_{\alpha\beta})_{\alpha,\beta \in G}$ and $A_G^{\Xi}(\lambda) = (d_{\alpha\beta})_{\alpha,\beta \in G \cup \Xi}$ by

$$a_{\alpha\beta} = \begin{cases} 1 + 2\lambda & \text{if } \alpha = \beta \\ 0 & \text{if } \alpha \cap \beta = \phi \\ (-1)^{i+j}\lambda & \text{if } \#(\alpha \cap \beta) = 1 \text{ and } \theta_i(\alpha) = \theta_j(\beta) \text{ for some} \\ & i, j = 1, 2, \end{cases}$$

and

$$d_{\alpha\beta} = \begin{cases} a_{\alpha\beta} & \text{if } \alpha \in G \text{ and } \beta \in G \\ \lambda\delta_{\alpha\beta} & \text{if } \alpha \in \Xi \text{ and } \beta \in \Xi \\ (-1)^i\lambda & \text{if } \alpha = \theta_i(\beta) \text{ or } \beta = \theta_i(\alpha) \text{ for some } i = 1, 2. \end{cases}$$

THEOREM 1.

$$H_\lambda(G) = \tfrac{1}{2} \log \det A_G(\lambda)$$

$$H_\lambda^{\Xi}(G) = H_\lambda(G) - \tfrac{1}{2} \log \det A_G^{\Xi}(\lambda) + (\#\Xi/2) \log \lambda.$$

From this theorem, it follows that $\det A_G(\lambda)$ and $\det A_G^{\equiv}(\lambda)$ are independent of the orientation θ on G and $\det A_G(\lambda) \geqslant 1$ because of the nonegativity of the entropy. Denote

$$P_G(\lambda) = \det A_G(\lambda)$$

and call it the *information polynomial* of G.

Proof. Let $\# G = n$ and $\# \Sigma = m$. Let $x_\Sigma = \{x_\sigma ; \sigma \in \Sigma\}$ and $z_G = \{z_\alpha ; \alpha \in G\}$ be observed values of random variables X_Σ and Z_G, respectively. The density functions of X_Σ, Z_G and (X_Σ, Z_G) are denoted by $p(x_\Sigma)$, $p(z_G)$ and $p(x_\Sigma, z_G)$, respectively. The conditional density of Z_G given X_Σ is denoted by $p(z_G | x_\Sigma)$. Then we have

$$p(x_\Sigma) = \left(\frac{1}{2\pi}\right)^{m/2} e^{-(1/2)\sum_{\sigma \in \Sigma} x_\sigma^2}$$

and

$$p(z_G | x_\Sigma) = \left(\frac{\lambda}{2\pi}\right)^{n/2} e^{-(\lambda/2)\sum_{\alpha \in G}(z_\alpha + x_{\theta_1(\alpha)} - x_{\theta_2(\alpha)})^2}.$$

It holds that

$$\begin{aligned}
H_\lambda(G) &= \iint \log \frac{p(x_\Sigma, z_G)}{p(x_\Sigma)\, p(z_G)} \cdot p(x_\Sigma, z_G)\, dx_\Sigma\, dz_G \\
&= \iint \log \frac{p(z_G | x_\Sigma)}{p(z_G)} \cdot p(x_\Sigma, z_G)\, dx_\Sigma\, dz_G \\
&= \int \left(\int \log p(z_G | x_\Sigma) \cdot p(z_G | x_\Sigma)\, dz_G \right) p(x_\Sigma)\, dx \\
&\qquad - \int \log p(z_G) \cdot p(z_G)\, dz_G \\
&= I_1 - I_2.
\end{aligned}$$

Since

$$\begin{aligned}
&\int \log p(z_G | x_\Sigma) \cdot p(z_G | x_\Sigma)\, dz_G \\
&\quad = \frac{n}{2} \log \frac{\lambda}{2\pi} - \frac{\lambda}{2} \sum_{\alpha \in G} \int (z_\alpha + x_{\theta_1(\alpha)} - x_{\theta_2(\alpha)})^2 \left(\frac{\lambda}{2\pi}\right)^{1/2} \\
&\quad \times e^{-(\lambda/2)(z_\alpha + x_{\theta_1(\alpha)} - x_{\theta_2(\alpha)})^2}\, dz_\alpha \\
&\quad = \frac{n}{2} \log \frac{\lambda}{2\pi} - \frac{\lambda}{2} \sum_{\alpha \in G} \frac{1}{\lambda} = \frac{n}{2} \log \frac{\lambda}{2\pi} - \frac{n}{2},
\end{aligned}$$

we have

$$I_1 = \frac{n}{2} \log \frac{\lambda}{2\pi} - \frac{n}{2}.$$

On the other hand, since z_G is normally distributed with mean $\mathbb{O}$ and covariance matrix $(1/\lambda) A_G$, we have

$$p(z_G) = \left(\frac{\lambda}{2\pi}\right)^{n/2} (\det A_G)^{-1/2} e^{-(\lambda/2) A_G^{-1}[z_G]},$$

where

$$A_G^{-1}[z_G] = \sum_{\substack{\alpha \in G \\ \beta \in G}} a^{\alpha\beta} z_\alpha z_\beta$$

and

$$A_G^{-1} = (a^{\alpha\beta}).$$

Therefore,

$$\begin{aligned} I_2 &= \frac{n}{2} \log \frac{\lambda}{2\pi} - \frac{1}{2} \log \det A_G - \frac{\lambda}{2} \sum_{\substack{\alpha \in G \\ \beta \in G}} a^{\alpha\beta} \int z_\alpha z_\beta p(z_G)\, dz_G \\ &= \frac{n}{2} \log \frac{\lambda}{2\pi} - \frac{1}{2} \log \det A_G - \frac{\lambda}{2} \sum_{\substack{\alpha \in G \\ \beta \in G}} a^{\alpha\beta} \frac{a_{\alpha\beta}}{\lambda} \\ &= \frac{n}{2} \log \frac{\lambda}{2\pi} - \frac{1}{2} \log \det A_G - \frac{n}{2}. \end{aligned}$$

Thus,

$$H_\lambda(G) = \tfrac{1}{2} \log \det A_G.$$

The other equality can be proved similarly.

For $n \geqslant 1$, a graph which is isomorphic to

$$B_n = \{\{1, 2\}, \{2, 3\}, \ldots, \{n, n+1\}\}$$

is called an *n-line*. For $n \geqslant 3$, a graph which is isomorphic to

$$C_n = \{\{1, 2\}, \{2, 3\}, \ldots, \{n-1, n\}, \{n, 1\}\}$$

is called an *n-cycle*. For graphs G and K, the *direct sum* $G \oplus K$ is defined as a graph $H = H_1 \cup H_2$ such that $\Sigma(H_1) \cap \Sigma(H_2) = \varnothing$, H_1 is isomorphic to G

and H_2 is isomorphic to H. The above K is determined uniquely up to isomorphism. Let $B(n, k)$ be the class of graphs which are isomorphic to $B_{n_1} \oplus B_{n_2} \oplus \cdots \oplus B_{n_k}$ for some $n_1, n_2, \ldots, n_k$ with $n_1 + n_2 + \cdots + n_k = n$. Let $C(n)$ be the class of graphs which are isomorphic to $C_{n_1} \oplus C_{n_2} \oplus \cdots \oplus C_{n_k}$ for some k and $n_1, n_2, \ldots, n_k$ with $n_1 + n_2 + \cdots + n_k = n$.

THEOREM 2. *For any positive integers n, m, k and h with $n + m = k + h$ and $|n - m| < |k - h|$, it holds that*

$$H(B_n) + H(B_m) > H(B_k) + H(B_h),$$

and

$$H_\lambda(B_n) + H_\lambda(B_m) > H_\lambda(B_k) + H_\lambda(B_h)$$

for any $\lambda > 0$.

COROLLARY. *The following three conditions on a graph $G \in B(n, k)$ are equivalent*:

(1) $H(G) = \max\{H(K); K \in B(n, k)\}$,

(2) $H_\lambda(G) = \max\{H_\lambda(K); K \in B(n, k)\}$ for any $\lambda > 0$, and

(3) *G is isomorphic to*

$$\underbrace{B_d \oplus \cdots \oplus B_d}_{r} \oplus \underbrace{B_{d+1} \oplus \cdots \oplus B_{d+1}}_{k-r},$$

where $d = [n/k]$ and $r = k(d + 1) - n$.

Proof. For any positive integers n and m with $n < m$, we have

$$H(B_{n+1}) - H(B_n) = H(Y_{\{n+1, n+2\}} | Y_{B_n})$$

$$> H(Y_{\{m+1, m+2\}} | Y_{B_m}) = H(B_{m+1}) - H(B_m).$$

One-half of the theorem follows from this fact. The other half follows from the following Lemma 4.

We denote $b_n = P_{B_n}(\lambda)$ $(n = 1, 2, \ldots)$ and $c_n = P_{C_n}(\lambda)$ $(n = 1, 2, \ldots)$. Then $b_n \geqslant 1$ $(n = 1, 2, \ldots)$ and $c_n \geqslant 1$ $(n = 3, 4, \ldots)$ for any $\lambda > 0$. We also define $b_0 = 1$, $b_{-1} = 0$, $b_{-2} = -\lambda^{-2}$, $c_2 = 1 + 4\lambda$, $c_1 = 1$ and $c_0 = 0$.

LEMMA 1. *It holds that*

$$b_n = \begin{vmatrix} 1+2\lambda & -\lambda & & & 0 \\ -\lambda & 1+2\lambda & -\lambda & & \\ & -\lambda & & & \\ & 0 & & & -\lambda \\ & & & -\lambda & 1+2\lambda \end{vmatrix} \quad (n \geqslant 1)$$

(size n)

and

$$c_n = \begin{vmatrix} 1+2\lambda & -\lambda & & & -\lambda \\ -\lambda & 1+2\lambda & -\lambda & 0 & \\ & -\lambda & & & \\ & 0 & & & -\lambda \\ -\lambda & & & -\lambda & 1+2\lambda \end{vmatrix} \quad (n \geqslant 3).$$

(size n)

Proof. Clear from the definitions.

LEMMA 2. (1) *For any $n \geqslant 0$, we have*

$$c_n = b_n - \lambda^2 b_{n-2} - 2\lambda^n.$$

(2) *For any $n, m \geqslant -1$, we have*

$$b_{n+m} = b_n b_m - \lambda^2 b_{n-1} b_{m-1}.$$

Proof. Clear from Lemma 1 and the definitions of b_n ($n = 0, -1, -2$) and c_n ($n = 2, 1, 0$).

LEMMA 3. *For any $n, m, k, h \geqslant -1$ such that $n - k = h - m = p \geqslant -1$, we have*

$$b_n b_m - b_k b_h = \lambda^2 b_{p-1}(b_k b_{m-1} - b_{k-1} b_m).$$

Proof. By Lemma 2, we have

$$b_n = b_k b_p - \lambda^2 b_{k-1} b_{p-1},$$

and

$$b_h = b_m b_p - \lambda^2 b_{m-1} b_{p-1}.$$

Putting them into the left-hand side of our equation, we have the right-hand side.

LEMMA 4. *For any n, m, k, $h \geqslant -1$ such that $n+m=k+h$ and $k \leqslant \min\{n, m\}+1$, we have*

$$b_n b_m - b_k b_h = \lambda^{2k+2} b_{n-k-1} b_{m-k-1}.$$

Thus, if $k < \min\{n, m\}$, then $b_n b_m > b_k b_h$ for any $\lambda > 0$.

Proof. Applying Lemmas 2 and 3, we have

$$\begin{aligned} b_n b_m - b_k b_h &= \lambda^2 b_{n-k-1}(b_k b_{m-1} - b_{k-1} b_m) \\ &= \lambda^4 b_{n-k-1}(b_{k-1} b_{m-2} - b_{k-2} b_{m-1}) \\ &= \cdots \\ &= \lambda^{2k} b_{n-k-1}(b_1 b_{m-k} - b_0 b_{m-k+1}) \\ &= \lambda^{2k} b_{n-k-1}(b_1 b_{m-k} - b_{m-k+1}) \\ &= \lambda^{2k+2} b_{n-k-1} b_{m-k-1}. \end{aligned}$$

THEOREM 3. *For any $n, m \geqslant 3$ and $\lambda > 0$,*

$$H_\lambda(C_{n+m}) > H_\lambda(C_n) + H_\lambda(C_m).$$

Proof. Our theorem follows from the following Lemma 10.

COROLLARY. *For $G \in C(n)$,*

$$H_\lambda(G) = \max\{H_\lambda(K); K \in C(n)\}$$

for any $\lambda > 0$ if and only if G is isomorphic to C_n.

LEMMA 5. *For $1 \leqslant n \leqslant m$, we have*

$$c_{n+m} - c_n c_m = 2\lambda^n c_m + 2\lambda^m c_n - \lambda^{2n} c_{m-n} - 4\lambda^{n+m}.$$

Proof. By Lemma 2, we have

$$\begin{aligned} &c_{n+m} - c_n c_m \\ &\quad = b_{n+m} - \lambda^2 b_{n+m-2} - 2\lambda^{n+m} - (b_n - \lambda^2 b_{n-2})(b_m - \lambda^2 b_{m-2}) \\ &\qquad + 2\lambda^n c_m + 2\lambda^m c_n \\ &\quad = -2\lambda^2 b_{n-1} b_{m-1} + \lambda^2 b_n b_{m-2} + \lambda^2 b_{n-2} b_m - 2\lambda^{n+m} + 2\lambda^n c_m + 2\lambda^m c_n. \end{aligned}$$

Since by Lemma 4,

$$-b_{n-1}b_{m-1}+b_n b_{m-2}=\lambda^{2n}b_{m-n-2}$$

and

$$-b_{n-1}b_{m-1}+b_{n-2}b_m=-\lambda^{2n-2}b_{m-n},$$

we have

$$\begin{aligned} c_{n+m}&-c_n c_m\\ &=2\lambda^n c_m+2\lambda^m c_n-\lambda^{2n}(b_{m-n}-\lambda^2 b_{m-n-2})-2\lambda^{n+m}\\ &=2\lambda^n c_m+2\lambda^m c_n-\lambda^{2n}c_{m-n}-4\lambda^{n+m}. \end{aligned}$$

By $b_{n,k}$ and $c_{n,k}$, we denote the coefficients of λ^k in b_n and c_n, respectively. We denote $b_n \ll b_m$ if $b_{n,k} \leqslant b_{m,k}$ for any integer k.

LEMMA 6. *For any $n \geqslant -2$, we have $b_{n,n}=n+1$, and for any $n \geqslant 0$, we have $c_{n,n}=0$.*

Proof. By definition, $b_{-2,-2}=-1$ and $b_{-1,-1}=0$. Let $n \geqslant 0$. Assume that $b_{k,k}=k+1$ for any $-2 \leqslant k \leqslant n-1$. Since by Lemma 2,

$$\begin{aligned} b_n&=b_1 b_{n-1}-\lambda^2 b_0 b_{n-2}\\ &=(1+2\lambda)\,b_{n-1}-\lambda^2 b_{n-2}, \end{aligned}$$

we have

$$\begin{aligned} b_{n,n}&=2b_{n-1,n-1}-b_{n-2,n-2}\\ &=2n-(n-1)=n+1. \end{aligned}$$

Hence by Lemma 2, we have

$$\begin{aligned} c_{n,n}&=b_{n,n}-b_{n-2,n-2}-2\\ &=(n+1)-(n-1)-2=0 \end{aligned}$$

for any $n \geqslant 0$.

LEMMA 7. *For $-1 \leqslant n < m$ and $p \geqslant 1$, we have*

$$0 \ll \lambda^p(b_m-\lambda^{m-n}b_n) \ll b_{m+p}-\lambda^{m-n}b_{n+p}.$$

Proof. It is sufficient to prove that

$$0 \ll \lambda(b_{n+1}-\lambda b_n) \ll b_{n+2}-\lambda b_{n+1}$$

for any $n \geqslant -1$. We prove this by the induction on n. For $n=-1$, this is clear. Assume that $n \geqslant 0$ and this is true up to $n-1$. Then, we have $b_{n+1} \gg 0$ and $b_{n+1} - \lambda b_n \gg 0$. Since

$$b_{n+2} = (1+2\lambda)\, b_{n+1} - \lambda^2 b_n,$$

we have

$$b_{n+2} - \lambda b_{n+1} = \lambda(b_{n+1} - \lambda b_n) + b_{n+1} \gg \lambda(b_{n+1} - \lambda b_n) \gg 0,$$

which completes the proof.

LEMMA 8. *For any $n \geqslant 1$, we have $c_n \gg 0$.*

Proof. Since

$$c_n = b_n - \lambda^2 b_{n-2} - 2\lambda^n,$$
$$b_n - \lambda^2 b_{n-2} \gg 0,$$

and

$$c_{n,n} = 0,$$

we have $c_n \gg 0$.

LEMMA 9. *For any $n \geqslant -1$, we have*

$$b_{n,n-1} = \tfrac{1}{6} n^3 + \tfrac{1}{2} n^2 + \tfrac{1}{3} n,$$

and for any $n \geqslant 1$, we have $c_{n,n-1} = n^2$.

Proof. Since $b_{n,n} = n+1$ and

$$b_n = (1+2\lambda)\, b_{n-1} - \lambda^2 b_{n-2}$$

for any $n \geqslant -1$, we have

$$b_{n,n-1} = n + 2b_{n-1,n-2} - b_{n-2,n-3}$$

for any $n \geqslant 0$. Therefore,

$$b_{n,n-1} = \tfrac{1}{6} n^3 + \tfrac{1}{2} n^2 + \tfrac{1}{3} n$$

holds by the induction on n. Since $c_n = b_n - \lambda^2 b_{n-2}$, we have

$$\begin{aligned} c_{n,n-1} &= b_{n,n-1} - b_{n-2,n-3} \\ &= \tfrac{1}{6} n^3 + \tfrac{1}{2} n^2 + \tfrac{1}{3} n - \tfrac{1}{6}(n-1)^3 - \tfrac{1}{2}(n-1)^2 - \tfrac{1}{3}(n-1) \\ &= n^2. \end{aligned}$$

LEMMA 10. *For any $n, m \geqslant 1$, we have $c_{n+m} \gg c_n c_m$ and $c_{n+m} \neq c_n c_m$.*

Proof. We may assume $n \leqslant m$. By Lemmas 5 and 8, we have

$$c_{n+m} - c_n c_m \gg \lambda^n (c_m - \lambda^n c_{m-n}) - 4\lambda^{n+m}.$$

Since the $(n+m)$th term in $c_{n+m} - c_n c_m$ is 0 by Lemma 6, $c_{n+m} - c_n c_m \gg 0$ follows from $c_m - \lambda^n c_{m-n} \gg 0$. On the other hand, by Lemmas 2 and 7,

$$c_m - \lambda^n c_{m-n} = (b_m - \lambda^n b_{m-n}) - \lambda^2 (b_{m-2} - \lambda^n b_{m-n-2}) \gg 0.$$

The last statement follows from the fact that by Lemma 9, $c_{n+m, n+m-1} = (n+m)^2 > 0$, while the $(n+m-1)$th coefficient in $c_n c_m$ is 0 by Lemma 6.

3. PROBLEM II

For random variables X, Y and Z, we denote by $P_{X|Y}$ and $P_{X|Z}$ the conditional distributions of X given Y and Z, respectively. Suppose that Z is a function of Y. Then it holds that

$$\begin{aligned}
I(X, Y) &= \int \log \frac{dP_{X,Y}}{d(P_X \times P_Y)} \, dP_{X,Y} \\
&= \int \log \frac{dP_{X|Y}}{dP_X} \, dP_{X,Y} \\
&= \int \left(- \int \log \frac{dP_X}{dP_{X|Y}} \, dP_{X|X,Z} \right) dP_{X,Z} \\
&\geqslant \int \left(- \log \int \frac{dP_X}{dP_{X|Y}} \, dP_{Y|X,Z} \right) dP_{X,Z} \\
&= \int \log \frac{dP_{X|Z}}{dP_X} \, dP_{X,Z} \\
&= I(X, Z),
\end{aligned}$$

where the equality holds if and only if

$$\frac{dP_{X|Y}}{dP_X} = \frac{dP_{X|Z}}{dP_X}$$

holds almost surely with respect to $P_{X,Y}$. That is to say that $I(X, Y) = I(X, Z)$ if and only if $P_{X|Y}$ is a function of Z, or equivalently, X and Y are conditionally independent of each other given Z.

For a graph G, $\sigma \in \Sigma(G)$ and $\alpha \in G$, we call that α is *neutral* in the combinatorial or Gaussian sense for σ in G if X_σ and Y_α or X_σ and Z_α are conditionally independent of each other given $Y_{G\setminus\{\alpha\}}$ or $Z_{G\setminus\{\alpha\}}$ for any $\lambda > 0$, respectively.

LEMMA 11. *Let*

$G_1 = \{\alpha \in G$; *α is not neutral in the combinatorial sense for σ in G*$\}$ *and*

$G_2 = \{\alpha \in G$; *α is not neutral in the Gaussian sense for σ in G*$\}$.

Then, G_1 and G_2 are the minimum sufficient subgraphs of G at σ in the combinatorial and Gaussian sense, respectively.

Proof. For any $\alpha \notin G_1$, since

$$P_{X_\sigma | Y_G} = P_{X_\sigma | Y_{G\setminus\{\alpha\}}},$$

$P_{X_\sigma | Y_G}$ does not depend on Y_α. Hence, $P_{X_\sigma | Y_G}$ is a function of Y_{G_1}. Thus, $H^\sigma(G_1) = H^\sigma(G)$ by the above argument. To prove that G_1 is minimum, let K be a subgraph of G such that $H^\sigma(K) = H^\sigma(G)$. Take any $\alpha \notin K$. Then since

$$H^\sigma(K) \leqslant H^\sigma(G\setminus\{\alpha\}) \leqslant H^\sigma(G),$$

we have $H^\sigma(G\setminus\{\alpha\}) = H^\sigma(G)$. Hence, by the above argument, X_σ and Y_α are conditionally independent of each other given $Y_{G\setminus\{\alpha\}}$. Thus, $G_1 \subset K$. The other half of the lemma can be proved similarly.

For $\Sigma_1, \Sigma_2 \subset \Sigma$, define a graph $Q(\Sigma_1, \Sigma_2)$ by

$$Q(\Sigma_1, \Sigma_2) = \{\{\tau, \xi\}; \tau \in \Sigma_1, \xi \in \Sigma_2, \tau \neq \xi\}.$$

THEOREM 4. *Let G be a graph, $\sigma \in \Sigma(G)$ and $\alpha \in G$. Then, α is neutral for σ in G in the combinatorial sense if there exist Σ_1, $\Sigma_2, \Sigma_3 \subset \Sigma(G)$ such that*

(1) $\Sigma_1 \cup \Sigma_2 \cup \Sigma_3 = \Sigma(G)$ and $\Sigma_i \cap \Sigma_j = \phi$ $(i \neq j)$,

(2) $\sigma \in \Sigma_1$,

(3) $\alpha \subset \Sigma_3$,

(4) $G \cap Q(\Sigma_1, \Sigma_3) = \phi$,

(5) $Q(\Sigma_2, \Sigma_3) \subset G$.

Proof. Denote $H = Q(\Sigma_2, \Sigma_3)$, $K = Q(\Sigma_3, \Sigma_3)$ and $L = G\setminus(H \cup K)$. At first, we show that our theorem follows from the fact that $X_{\Sigma_1 \cup \Sigma_2}$ and Y_K are conditionally independent of each other given Y_H. Assume this fact. Since Y_L and (X_σ, Y_L) are functions of $X_{\Sigma_1 \cup \Sigma_2}$, we have

$$P_{Y_L | Y_H} = P_{Y_L | Y_{H \cup K}}$$

and

$$P_{X_\sigma, Y_L | Y_H} = P_{X_\sigma, Y_L | Y_{H \cup K}}.$$

Therefore,

$$P_{X_\sigma | Y_{H \cup L}} = \frac{P_{X_\sigma, Y_L | Y_H}}{P_{Y_L | Y_H}} = \frac{P_{X_\sigma, Y_L | Y_{H \cup K}}}{P_{Y_L | Y_{H \cup K}}} = P_{X_\sigma | Y_{H \cup K \cup L}}.$$

Since

$$H \cup L \subset G \backslash \{\alpha\} \subset G \subset H \cup K \cup L,$$

this implies that

$$P_{X_\sigma | G \backslash \{\alpha\}} = P_{X_\sigma | G},$$

and hence, α is neutral for σ in G. Since

$$X_{\Sigma_1} \amalg (X_{\Sigma_2}, Y_H, Y_K),$$

to prove that $X_{\Sigma_1 \cup \Sigma_2}$ and Y_K are conditionally independent of each other given Y_H, it is sufficient to prove that X_{Σ_2} and Y_K are conditionally independent of each other given Y_H. We call $x_{\Sigma_2} = \{x_\tau; \tau \in \Sigma_2\} \in \mathbb{R}^{\Sigma_2}$ *consistent* if $x_\tau \neq x_\xi$ for any $\tau, \xi \in \Sigma_2$ with $\tau \neq \xi$. A relation Λ on a set Σ (i.e., $\Lambda \subset \Sigma \times \Sigma$) is called an *order* on Λ if

(1) $(\tau, \tau) \notin \Lambda$ for any $\tau \in \Sigma$, and

(2) $(\tau, \xi) \in \Lambda$ and $(\xi, \eta) \in \Lambda$ imply $(\tau, \eta) \in \Lambda$ for any $\tau, \xi, \eta \in \Sigma$.

In this case, $(\tau, \xi) \in \Lambda$ is also denoted as $\tau \prec \xi[\Lambda]$. Let x_{Σ_2} be consistent. Then, it defines an order on Σ_2 denoted by $[x_{\Sigma_2}]$ by $\tau \prec \xi[x_{\Sigma_2}]$ if $x_\tau < x_\xi$. In general, for a graph E on Σ and an orientation ψ on E, we call $y_E = \{y_\zeta; \zeta \in E\} \in \{-1, 1\}^E$ *consistent* if there exists an order on Σ which contains the relation

$$\left\{ \begin{array}{c} (\tau, \xi) \in \Sigma \times \Sigma;\ \{\tau, \xi\} \in E, \tau = \psi_i(\{\tau, \xi\}) \text{ and } y_{\{\tau, \xi\}} = (-1)^{i+1} \\ \text{for some } i = 1, 2 \end{array} \right\}.$$

If y_E is consistent, then the minimum order as above is denoted by $[y_E]$. Let orientations on H and K be given. Let y_H be consistent. We call y_K *consistent with respect to* y_H if $y_{H \cup K} = (y_H, y_K)$ is consistent. We call x_{Σ_2} *consistent with respect to* y_H if x_{Σ_2} is consistent and if $\xi \prec \tau[y_H]$ implies $\xi \prec \tau[x_{\Sigma_2}]$ for any $\xi, \tau \in \Sigma_2$. It is clear that

$$P(Y_H \text{ is consistent, } X_{\Sigma_2} \text{ and } Y_K \text{ are consistent with respect to } Y_H) = 1.$$

Therefore, to prove that X_{Σ_2} and Y_K are conditionally independent of each other given Y_H, it is sufficient to prove that

$$P(Y_K = y_K | X_{\Sigma_2} = x_{\Sigma_2}, Y_H = y_H) = P(Y_K = y_K | Y_H = y_H)$$

for any consistent y_H, consistent x_{Σ_2} and y_K with respect to y_H. Let such y_H, x_{Σ_2} and y_K be given. For $\tau \in \Sigma_3$, define

$$m(\tau) = \max\{x_\xi ; \xi \in \Sigma_2, \xi \prec \tau [y_H]\}$$

and

$$M(\tau) = \min\{x_\xi ; \xi \in \Sigma_2, \tau \prec \xi [y_H]\}.$$

For $\tau, \xi \in \Sigma_3$, we denote

$$\tau \sim \xi \, [y_H]$$

if neither $\tau \prec \xi \, [y_H]$ nor $\xi \prec \tau \, [y_H]$. Then, it defines an equivalence relation on Σ_3. Let $\Xi_1, \Xi_2, ..., \Xi_n$ be the equivalence classes of the relation arranged so that $\tau \prec \xi \, [y_H]$ for any $\tau \in \Xi_i$ and $\xi \in \Xi_j$ with $i < j$. Then, it is easy to see that $m(\tau) = m(\xi)$ and $M(\tau) = M(\xi)$ hold if $\tau \sim \xi \, [y_H]$. We denote $m_i = m(\tau)$ and $M_i = M(\tau)$ for any $\tau \in \Xi_i$ and $i = 1, 2, ..., n$. It is also easy to see that $m_i \leqslant M_i \leqslant m_{i+1} \leqslant M_{i+1}$ holds for any $i = 1, 2, ..., n-1$. Then, we have

$$\begin{aligned} P(Y_K = y_K | x_{\Sigma_2}, y_H) &= \prod_{i=1}^{n} P(Y_{K_i} = y_{K_i} | x_{\Sigma_2}, y_H) \\ &= \prod_{i=1}^{n} P(Y_{K_i} = y_{K_i} | m_i \leqslant x_\tau \leqslant M_i \text{ for any } \tau \in \Xi_i) \\ &= \prod_{i=1}^{n} \frac{1}{(\#\Xi_i)!}, \end{aligned}$$

where $K_i = Q(\Xi_i, \Xi_i)$. Since the last term depends only on y_H, we complete the proof.

THEOREM 5. *For a graph G, $\sigma \in \Sigma(G)$ and $\alpha \in G$, α is neutral for σ in G in the Gaussian sense if and only if*

$$\det(d_{\beta\gamma})_{\beta \in G, \gamma \in (G \setminus \{\alpha\}) \cup \{\sigma\}} = 0,$$

where $d_{\beta\gamma}$ are as in the paragraph preceding Theorem 1 *with $\Xi = \{\sigma\}$.*

Proof. Since (X_σ, Z_G) is normally distributed with

$$\text{Var}(X_\sigma | Z_G) = \frac{\det A_G^\sigma(\lambda)}{\lambda \det A_G(\lambda)} \equiv v$$

there exists a linear function T of z_G such that

$$p(x_\sigma | z_G) = (2\pi v)^{-1/2} e^{-(1/2v)(x_\sigma - T(z_G))^2}.$$

Therefore, to prove our theorem, it is sufficient to prove that $T(z_G)$ does not depend on z_α. On the other hand, since

$$\begin{aligned} p(x_\sigma | z_G) &= \frac{p(x_\sigma, z_G)}{p(z_G)} \\ &= \left(\frac{\lambda}{2\pi}\right)^{1/2} \left(\frac{\det A_G(\lambda)}{\det A_G^\sigma(\lambda)}\right)^{1/2} e^{-(\lambda/2)(A_G^\sigma(\lambda)^{-1}[x_\sigma, z_G] - A_G(\lambda)^{-1}[z_G])}, \end{aligned}$$

we have

$x_\sigma^2 - 2x_\sigma T(z_G) = -\lambda v A_G^\sigma(\lambda)^{-1}[x_\sigma, z_G] +$ (a term depending only on z_G). Therefore, the term $-2x_\sigma T(z_G)$ comes from the first term in the right-hand side. Hence, to prove that $T(z_G)$ is independent of z_α, it is sufficient to prove that (α, σ) factor of $A_G^\sigma(\lambda)^{-1}$ is 0. This is equivalent to say that (α, σ) cofactor of $A_G^\sigma(\lambda)$ is 0, which completes the proof.

COROLLARY. *A sufficient condition for $\alpha \in G$ to be neutral for $\sigma \in \Sigma(G)$ in G is that there exists $\Xi \subset \Sigma(G)$ such that*

(1) *$\alpha \subset \Xi$, and*

(2) *for any τ, $\xi \in \Xi$, there exists an automorphism of $G \cup Q(\Xi, \Xi)$ which transforms σ to itself, τ to ξ and ξ to τ.*

Proof. Let $G' = G \cup Q(\Xi, \Xi)$. Take an arbitrary $\zeta = \{\tau, \xi\} \in Q(\Xi, \Xi)$ and an automorphism f of G' such that $f(\sigma) = \sigma, f(\tau) = \xi$ and $f(\xi) = \tau$. For $\beta = \{\rho, \nu\} \in G'$, we denote $f(\beta) = \{f(\rho), f(\nu)\} \in G'$. Take an orientation ψ on G'. For $\beta \in G' \cup \{\sigma\}$, define

$$g(\beta) = \begin{cases} 1 & \text{if } \beta = \sigma \text{ or if } \beta \in G' \text{ and } \psi_1(f(\beta)) = f(\psi_1(\beta)) \\ -1 & \text{else.} \end{cases}$$

Then, f is a permutation on $G' \cup \{\sigma\}$. Let $(d_{\beta\gamma})$ be as in the paragraph preceding Theorem 1 with G' and $\{\sigma\}$ for G and Ξ. Then, it holds that

$$d_{f(\beta)f(\gamma)} = g(\beta)g(\gamma)\, d_{\beta\gamma}$$

for any β, $\gamma \in G' \cup \{\sigma\}$. Let $k = \#\{\beta \in G'; g(\beta) = -1\}$. Note that $g(\zeta) = -1$. Therefore, if we permutate the rows and the columns of the matrix

$$(d_{\beta\gamma})_{\beta \in G', \gamma \in (G' \backslash \{\zeta\}) \cup \{\sigma\}}$$

according to f at the same time, then it changes the sign at k rows and at $k-1$ columns, succeedingly. This implies that the determinant of the matrix changes the sign, while it remains unchanged since the number of the transpositions of rows together with columns is even since $f(\sigma)=\sigma$ and $f(\zeta)=\zeta$. Thus, the determinant is 0, which implies that ζ is neutral for σ in G' by Theorem 5. Since $\zeta \in Q(\Xi, \Xi)$ was arbitrary, we have for any $\lambda > 0$,

$$H_\lambda^\sigma(G \backslash Q(\Xi, \Xi)) = H_\lambda^\sigma(G \cup Q(\Xi, \Xi))$$

by Lemma 11. Since

$$G \backslash Q(\Xi, \Xi) \subset G \backslash \{\alpha\} \subset G \subset G \cup Q(\Xi, \Xi),$$

this implies that for any $\lambda > 0$,

$$H_\lambda^\sigma(G) = H_\lambda^\sigma(G \backslash \{\alpha\}),$$

and hence α is neutral for σ in G by Lemma 11.

4. Examples

Take the orientations θ on B_{n-1} $(n \geqslant 2)$ or C_n $(n \geqslant 3)$ such that $\theta_1(\{i, i+1\}) = i$ for $i = 1, 2, \ldots, n-1$ and $\theta_1(\{n, 1\}) = n$. Here is an algorithm to calculate $H(B_n)$ due to Harriet Fell. For $n \geqslant 1$, $y_{B_n} \in \{-1, 1\}^{B_n}$ and $0 \leqslant k \leqslant n$, define $J_n(y_{B_n}, k)$ by the following equation:

$$J_1(1, 0) = J_1(-1, 1) = 0$$

$$J_1(1, 1) = J_1(-1, 0) = 1$$

$$J_n(y_{B_n}, k) = \begin{cases} \sum_{i<k} J_{n-1}(y_{B_{n-1}}, i) & (y_{\{n,n+1\}} = 1) \\ \sum_{i \geqslant k} J_{n-1}(y_{B_{n-1}}, i) & (y_{\{n,n+1\}} = -1). \end{cases}$$

Theorem 6 (Harriet Fell).

$$H(B_n) = \log(n+1)! - \frac{1}{(n+1)!} \sum_{y_{B_n}} \sum_{k=0}^{n} J_n(y_{B_n}, k) \times \log \sum_{k=0}^{n} J_n(y_{B_n}, k).$$

Proof. For $y_{B_n} \in \{-1, 1\}^{B_n}$, let $[y_{B_n}]$ be the order on $\{1, 2, \ldots, n+1\}$ defined in the proof of Theorem 4. Then, it can be proved by the induction on n that $J_n(y_{B_n}, k)$ is the number of total orders which are extensions of

$[y_{B_n}]$ such that the number of elements in $\{1, 2,..., n\}$ which is smaller than $n+1$ in the order is k. Hence,

$$P(Y_{B_n}=y_{B_n})=\frac{1}{(n+1)!}\sum_{k=0}^{n} J_n(y_{B_n}, k).$$

Our theorem follows from this fact together with the fact that

$$H(B_n)=-\sum_{y_{B_n}} P(Y_{B_n}=y_{B_n})\log(Y_{B_n}=y_{B_n}).$$

There is a similar algorithm to calculate $H(C_n)$ due to Mitsuru Fukui. For any $y_{C_n}\in\{-1, 1\}^{C_n}$ $(n\geqslant 3)$ and $j=0, 1,..., n-1$, define $\tau_j y_{C_n}\in\{-1, 1\}^{C_n}$ by

$$(\tau_j y_{C_n})_{\{i,i+1\}}=y_{\{i+j,i+j+1\}}$$

for any $i=1, 2,..., n$, where the additions involving i or j are considered in modulo n. Define $\eta y_{C_n}\in\{-1, 1\}^{C_{n-1}}$ $(n\geqslant 4)$ by

$$(\eta y_{C_n})_{\{i,i+1\}}=y_{\{i,i+1\}} \qquad (i=1, 2,..., n-2)$$

and

$$(\eta y_{C_n})_{\{n-1,1\}}=-1.$$

For $n\geqslant 3$, $y_{C_n}\in\{-1, 1\}^{C_n}$ and $1\leqslant k\leqslant n-1$, define $J_n(y_{C_n}, k)$ by the following equation:

$$J_3(y_{C_3}, k)=\begin{cases}1 & \text{if } y_{\{1,2\}}=1, y_{\{3,1\}}=-1 \text{ and } (y_{\{2,3\}}+3)/2=k\\ 0 & \text{else}\end{cases}$$

$$J_n(y_{C_n}, k)=\begin{cases}\sum_{i<k} J_{n-1}(\eta y_{C_n}, i) & \text{if } y_{\{n,1\}}=-1 \text{ and } y_{\{n-1,n\}}=1\\ \sum_{i\geqslant k} J_{n-1}(\eta y_{C_n}, i) & \text{if } y_{\{n,1\}}=y_{\{n-1,n\}}=-1\\ 0 & \text{else.}\end{cases}$$

THEOREM 7 (MITSURU FUKUI).

$$H(C_n)=\log n!-\frac{1}{n!}\sum_{y_{C_n}}\sum_{j=0}^{n-1}\sum_{k=1}^{n-1} J_n(\tau_j y_{C_n}, k)$$
$$\times\log\sum_{j=0}^{n-1}\sum_{k=1}^{n-1} J_n(\tau_j y_{C_n}, k).$$

EXAMPLE 1 (MITSURU FUKUI).

$$
\begin{array}{ll}
H(B_1) \doteqdot 1.0000 & \\
H(B_2) \doteqdot 1.9183 & \\
H(B_3) \doteqdot 2.8250 & H(C_3) \doteqdot 2.5850 \\
H(B_4) \doteqdot 3.7298 & H(C_4) \doteqdot 3.5849 \\
H(B_5) \doteqdot 4.6344 & H(C_5) \doteqdot 4.5106 \\
H(B_6) \doteqdot 5.5388 & H(C_6) \doteqdot 5.4222 \\
 & H(C_7) \doteqdot 6.3295 \\
 & H(C_8) \doteqdot 7.2351 \\
 & H(C_9) \doteqdot 8.1400
\end{array}
$$

(in the binary base)

EXAMPLE 2.

$$
\begin{array}{ll}
b_1 = 1 + 2\lambda & \\
b_2 = 1 + 4\lambda + 3\lambda^2 & c_3 = 1 + 6\lambda + 9\lambda^2 \\
b_3 = 1 + 6\lambda + 10\lambda^2 + 4\lambda^3 & c_4 = 1 + 8\lambda + 20\lambda^2 + 16\lambda^3 \\
b_4 = 1 + 8\lambda + 21\lambda^2 + 20\lambda^3 + 5\lambda^4 & c_5 = 1 + 10\lambda + 35\lambda^2 + 50\lambda^3 + 25\lambda^4 \\
b_5 = 1 + 10\lambda + 36\lambda^2 + 56\lambda^3 + 35\lambda^4 + 6\lambda^5 & \\
b_6 = 1 + 12\lambda + 55\lambda^2 + 120\lambda^3 + 126\lambda^4 + 56\lambda^5 + 7\lambda^6 &
\end{array}
$$

EXAMPLE 3. It holds that

$$H(B_3 \oplus C_3) = H(B_3) + H(C_3) \doteqdot 5.4100 < 5.4222 \doteqdot H(C_6),$$

while

$$
\begin{aligned}
H_\lambda(B_3 \oplus C_3) &= \log b_3 c_3 = \log(1 + 12\lambda + 55\lambda^2 + 118\lambda^3 + 114\lambda^4 + 36\lambda^5) \\
&> \log(1 + 12\lambda + 54\lambda^2 + 112\lambda^3 + 105\lambda^4 + 36\lambda^5) = \log c_6 = H_\lambda(C_6)
\end{aligned}
$$

for any $\lambda > 0$.

EXAMPLE 4. By the corollary to Theorem 5, $\{1, 5\}$ is neutral for 3 in C_5, while it is not in the combinatorial sense, since

$$P(Y_{\{5,1\}} = 1 \mid X_3 = x,\ Y_{\{1,2\}} = Y_{\{3,4\}} = Y_{\{4,5\}} = 1,\ Y_{\{2,3\}} = -1)$$

tends to 0 as $x \to \infty$ and tends to $1/6$ as $x \to -\infty$.

5. Unsolved Problems

Problem 1. Does

$$H(C_{n+m}) > H(C_n) + H(C_m)$$

hold for any $n, m \geqslant 3$?

Problem 2. Are the converses of Theorem 4 and the corollary to Theorem 5 true?

Problem 3. Does the combinatorial entropy determine the isomorphic class of graphs?

Acknowledgments

The author would like to thank Professor Gian-Carlo Rota (M.I.T.) for his encouragement, and Professor Harriet Fell (Northeastern Univ.) and Mr. Mitsuru Fukui (Osaka City Univ.) for their useful discussions on the subject.