Supplementary Material: Optimal Cooperative Inference

This supplementary material presents the additional details and proofs associated with the main paper.

1 Details of Remark 2.6

Suppose that $|\mathcal{H}|$ is countably infinite. Let $\mathbf{A} = (\mathbf{L}_{i,j}\mathbf{T}_{i,j})_{|\mathcal{D}|\times|\mathcal{H}|}$ be the matrix obtained from $\mathbf{L}$ and $\mathbf{T}$ by element-wise multiplication. Denote the sum of elements in the j-th column of $\mathbf{A}$ by C_j. Then $S_n = \sum_{j=1}^{n} C_j$ is the sum of elements in the first n columns of $\mathbf{A}$. Note that $0 \leq C_j = \sum_{i=1}^{|\mathcal{D}|} \mathbf{L}_{i,j}\mathbf{T}_{i,j} \leq \sum_{i=1}^{|\mathcal{D}|} \mathbf{T}_{i,j} = 1$ and so $0 \leq S_n \leq n$. Therefore, for any j, n, both C_j and S_n exist, and $\{\frac{S_n}{n}\}_{n=1}^{\infty}$ is a well-defined sequence whose limit is then called TI.

Regrading the existence of TI, there are two cases.

Case 1: The growth rate of S_n is strictly slower than any linear function. Thus, for any $k > 0$, there exists an integer $N(k) > 0$ (depends on k) such that $S_n < k \cdot n$ for any $n > N(k)$. Then for any $k > 0$, the following holds:

$$0 \leq \text{TI} = \lim_{n\to\infty} \frac{S_n}{n} \leq \lim_{n\to\infty} \frac{k \cdot n}{n} = k.$$

Thus, $\text{TI} = 0$.

Case 2: If the growth rate of S_n is not strictly slower than linear functions, then TI exists if and only if the sequence $\{C_j\}$ converges as $j \to \infty$. Suppose that $\{C_j\}$ converges to k. Then for any $\epsilon > 0$, there exists an integer $N(\epsilon)$ such that $|C_m - k| < \epsilon$ for any $m > N(\epsilon)$. Therefore, for n sufficiently large,

$$|\frac{S_n}{n} - k| = |\frac{S_n - n \cdot k}{n}| = |\frac{S_N - N \cdot k}{n} + \frac{\sum_{j=N}^{n} C_j - k}{n - N}| \leq |\frac{S_N - N \cdot k}{n}| + \epsilon \leq \epsilon'.$$

Thus, TI exists. Similarly the other direction also holds.

Moreover, when TI exists, Proposition 2.4 can also be generalized. $0 \leq S_n \leq n$ implies that the range of TI is $[0, 1]$, and $\text{TI} = 1$ if and only if C_j converges to 1.

2 Proof of Theorem 4.6

For convenience, we first write the fixed-point iteration of (2) explicitly in vector form. We denote the matrix with elements $P_L(h|D)$ by $\mathbf{L} \in [0,1]^{|\mathcal{D}|\times|\mathcal{H}|}$, the matrix with elements $P_T(D|h)$ by $\mathbf{T} \in [0,1]^{|\mathcal{D}|\times|\mathcal{H}|}$, and the matrix with elements $P_0(D|h)$ by $\mathbf{M} \in [0,1]^{|\mathcal{D}|\times|\mathcal{H}|}$. Further, denote the vectors consisting of $P_{L_0}(h)$ and $P_T(h)$ by $\mathbf{a}, \mathbf{d} \in [0,1]^{|\mathcal{H}|\times 1}$, vectors consisting of $P_{T_0}(D)$ and $P_L(D)$ by $\mathbf{b}, \mathbf{c} \in [0,1]^{|\mathcal{D}|\times 1}$, respectively. Given

$\mathbf{a}$, $\mathbf{b}$, and $\mathbf{M}$, the fixed-point iteration of the cooperative inference equations can be expressed as:

$$P_{L_1}(h|D) = \frac{P_0(D|h)\,P_{L_0}(h)}{P_{L_1}(D)} \iff \mathbf{L}^{(1)} = \mathrm{Diag}\left(\frac{1}{\mathbf{M}\,\mathbf{a}}\right)\mathbf{M}\,\mathrm{Diag}(\mathbf{a}) \tag{1a}$$

$$P_{T_{k+1}}(D|h) = \frac{P_{L_{k+1}}(h|D)\,P_{T_0}(D)}{P_{T_{k+1}}(h)} \iff \mathbf{T}^{(k+1)} = \mathrm{Diag}(\mathbf{b})\,\mathbf{L}^{(k+1)}\mathrm{Diag}\left(\frac{1}{\mathbf{d}^{(k+1)}}\right) \tag{1b}$$

$$P_{T_{k+1}}(h) = \sum_{D\in\mathcal{D}} P_{L_k}(h|D)\,P_{T_0}(D) \iff \mathbf{d}^{(k+1)} = (\mathbf{L}^{(k+1)})^\mathsf{T}\,\mathbf{b} \tag{1c}$$

$$P_{L_{k+1}}(h|D) = \frac{P_{T_k}(D|h)\,P_{L_0}(h)}{P_{L_{k+1}}(D)} \iff \mathbf{L}^{(k+1)} = \mathrm{Diag}\left(\frac{1}{\mathbf{c}^{(k+1)}}\right)\mathbf{T}^{(k)}\mathrm{Diag}(\mathbf{a}) \tag{1d}$$

$$P_{L_{k+1}}(D) = \sum_{h\in\mathcal{H}} P_{T_k}(D|h)\,P_{L_0}(h) \iff \mathbf{c}^{(k+1)} = \mathbf{T}^{(k)}\mathbf{a}, \tag{1e}$$

where k denotes the iteration step; $\mathrm{Diag}(\mathbf{z})$ denotes the diagonal matrix with elements of the vector $\mathbf{z}$ on its diagonal; and $\frac{1}{\mathbf{z}}$ denotes element-wise inverse of vector $\mathbf{z}$.

Note that (1b) and (1c) are the operations to column normalize $\mathrm{Diag}(\mathbf{b})\,\mathbf{L}^{(k)}$, and (1d) and (1e) are the operations to row normalize $\mathbf{T}^{(k)}\mathrm{Diag}(\mathbf{a})$. Zero rows in $\mathbf{L}^{(k)}$ and zero columns in $\mathbf{T}^{(k)}$ are fixed throughout the iteration of (1) if they exist. This is equivalent to removing the zero rows and zero columns of $\mathbf{M}$ for (1) and inserting them back at convergence or when the iteration is stopped.

Now we provide a version of the proof using the notations introduced in the paper. The original proof can be found in [2]. Remember that $\mathbf{a}$ and $\mathbf{b}$ are assumed to be uniform.

Proof. Let σ be a permutation of $\{1,\cdots,n\}$ that makes $\{\mathbf{M}_{i,\sigma(i)}\}_{i=1}^n$ a positive diagonal. Define

$$e^{(k)} := \prod_{i=1}^n \mathbf{L}^{(k)}_{i,\sigma(i)}; \quad f^{(k)} := \prod_{i=1}^n \mathbf{T}^{(k)}_{i,\sigma(i)}.$$

Applying (1a), $\mathbf{L}^{(1)}$ is a row-stochastic matrix, and $\{\mathbf{L}^{(1)}_{i,\sigma(i)}\}_{i=1}^n$ is a positive diagonal, hence $e^{(1)}$ is positive. Also, by applying (1b),

$$f^{(1)} = \prod_{i=1}^n \mathbf{T}^{(1)}_{i,\sigma(i)} = \prod_{i=1}^n \left(\mathbf{b}_i \frac{\mathbf{L}^{(1)}_{i,\sigma(i)}}{\mathbf{d}^{(1)}_{\sigma(i)}}\right) = \frac{e^{(1)}}{n^n \prod_{i=1}^n \mathbf{d}^{(1)}_{\sigma(i)}} = \frac{e^{(1)}}{n^n \prod_{i=1}^n \mathbf{d}^{(1)}_i}. \tag{2}$$

By the inequality of arithmetic and geometric means, $\left(\prod_{i=1}^n \mathbf{d}^{(1)}_i\right)^{\frac{1}{n}} \le \frac{1}{n}\sum_{i=1}^n \mathbf{d}^{(1)}_i$. Also, $\mathbf{L}^{(1)}$ is a row-stochastic matrix and we assumed uniform prior on data set space, and hence, by (1c)

$$n^n \prod_{j=1}^n \mathbf{d}^{(1)}_j \le \left(\sum_{j=1}^n \mathbf{d}^{(1)}_j\right)^n = \left(\sum_{i=1}^n\sum_{j=1}^n \mathbf{b}_j \mathbf{L}^{(1)}_{i,j}\right)^n = \left(\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \mathbf{L}^{(1)}_{i,j}\right)^n = 1. \tag{3}$$

The equality in (3) is achieved if and only if $\mathbf{d} = \left(\frac{1}{n},\ldots,\frac{1}{n}\right)$, or equivalently, $\mathbf{L}^{(1)}$ being a doubly stochastic matrix. Because $f^{(1)}$ is the product of n values between 0 and 1,

$$0 < e^{(1)} \underset{(a)}{\le} f^{(1)} \underset{(b)}{\le} 1, \tag{4}$$

with equality in (a) if and only if $\mathbf{L}^{(1)}$ is a doubly stochastic matrix, and equality in (b) if and only if $\mathbf{L}^{(1)}$ is a permutation matrix. Applying the same logic to equations (1d) and (1e), we have

$$0 < f^{(1)} \underset{(c)}{\le} e^{(2)} \underset{(d)}{\le} 1,$$

with equality in (c) if and only if $\mathbf{T}^{(1)}$ is a doubly stochastic matrix, and equality in (d) if and only if $\mathbf{T}^{(1)}$ is a permutation matrix. Repeating this argument, we get the increasing sequence

$$0 < e^{(1)} \le f^{(1)} \le e^{(2)} \le f^{(2)} \le \cdots \le 1.$$

Monotone convergence theorem of real numbers guarantees that this sequence converges to its supremum

$$\lim_{k\to\infty} e^{(k)} = \lim_{k\to\infty} f^{(k)} = \sup\{\mathbf{e}, \mathbf{f}\}.$$

Asymptotically, $e^{(k)} = f^{(k)} = e^{(k+1)}$; therefore, $\mathbf{L}^{(k)}$ and $\mathbf{T}^{(k)}$ are both doubly stochastic matrices. Because doubly stochastic matrices are stable under row and column normalization, $\mathbf{L}$ and $\mathbf{T}$ converge to the same doubly stochastic matrix,

$$\mathbf{M}^{(\infty)} := \lim_{k\to\infty} \mathbf{L}^{(k)} = \lim_{k\to\infty} \mathbf{T}^{(k)}.$$

□

3 Proof of Theorem 4.10

Proof. (1) $(a) \iff (b)$: We first prove that (a) $\text{CI}(\mathbf{M}) = 1$, and (b) $\mathbf{M}$ *has exactly one positive diagonal,* are equivalent. Since $\mathbf{M}$ is an $n \times n$ nonnegative matrix with at least one positive diagonal, Theorem 4.6 guarantees that the iteration of equation set (1) converges to a doubly stochastic matrix, $\mathbf{M}^{(\infty)}$. According to Birkhoff–von Neumann theorem [1, 3], there exist $\theta_1, \dots, \theta_k \in (0, 1]$ with $\sum_i \theta_i = 1$ and distinct permutation matrices $P_1, \dots, P_k$ such that $\mathbf{M}^{(\infty)} = \theta_1 P_1 + \cdots + \theta_k P_k$. To simplify, we adopt the *inner product* notation between matrices: $A \cdot B = \sum_{i,j} A_{i,j} B_{i,j}$, for any two $n \times n$ square matrices A and B. Then the following holds:

$$\text{CI} = \text{TI}(\mathbf{M}^{(\infty)}, \mathbf{M}^{(\infty)}) \underset{(I)}{=} \frac{1}{n}\mathbf{M}^{(\infty)} \cdot \mathbf{M}^{(\infty)} \underset{(II)}{=} \frac{1}{n}(\sum_i \theta_i P_i) \cdot (\sum_j \theta_j P_j) \underset{(III)}{=} \frac{1}{n}\sum_{i,j} \theta_i \theta_j P_i \cdot P_j.$$

Equality (I) comes from rewriting TI in the inner product notation. Equality (II) comes from substituting $\mathbf{M}^{(\infty)}$ by its Birkhoff–von Neumann decomposition. Equality (III) comes from distribution.

Further, as permutation matrices, $P_i \cdot P_j \leq n$, and the equality holds if and only if $P_i = P_j$. So we have

$$\text{CI}(\mathbf{M}) = \frac{1}{n}\sum_{i,j} \theta_i \theta_j P_i \cdot P_j \underset{(IV)}{\leq} \frac{1}{n}\sum_{i,j} \theta_i \theta_j n = \sum_{i,j} \theta_i \theta_j = (\sum_i \theta_i) \times (\sum_j \theta_j) = 1.$$

The equality in (IV) holds if and only if $P_i = P_j$ for any i, j. Note that $P_1, \dots, P_k$ are distinct, i.e., $P_i \neq P_j$ when $i \neq j$. So the equality in (IV) is achieved precisely when $k = 1$ and $\mathbf{M}^{(\infty)} = P_1$. Hence, $\text{CI}(\mathbf{M})$ is maximized if and only if $\mathbf{M}^{(\infty)}$ is a permutation matrix.

We then prove that $\mathbf{M}^{(\infty)}$ is a permutation matrix if and only if $\mathbf{M}$ has exactly one positive diagonal. This follows from this claim, **Claim** (1): elements of $\mathbf{M}$ that lie in a positive diagonal do not tend to zero during the cooperative inference iteration [2] (i.e., if $\mathbf{M}_{i,j} \neq 0$ lies in a positive diagonal, then $\mathbf{M}^{(\infty)}_{i,j} \neq 0$). Claim (1) implies that $\mathbf{M}^{(\infty)}$ and $\mathbf{M}$ have the same number of positive diagonals. Further, note that a doubly stochastic matrix has exactly one diagonal if and only it is a permutation matrix. So as a doubly stochastic matrix, $\mathbf{M}^{(\infty)}$ is a permutation matrix if and only if $\mathbf{M}$ has exactly one positive diagonal. Thus, CI is maximized if and only if $\mathbf{M}$ has exactly one positive diagonal.

To complete the proof for $(a) \iff (b)$, we only need to justify Claim (1). Note that the product of any positive diagonal converges to a positive number $\sup\{\mathbf{e}, \mathbf{f}\}$ (shown in the proof for Theorem 4.6) and all elements on the positive diagonal is upper-bounded by 1 and lower-bounded by $\sup\{\mathbf{e}, \mathbf{f}\}$. , elements on a diagonal of $\mathbf{M}$ cannot converge to 0.

(2) $(b) \iff (c)$: This follows immediately from a slightly more general claim below, where positive diagonals are generalized to non-zero diagonals (can have negative values).

Claim (2): Let A be an $n \times n$-square matrix (elements can be any real number). Then A has exactly one non-zero diagonal (i.e., a diagonal with no zero element) if and only if A is a permutation of an upper-triangular matrix.

We now prove Claim (2). The if direction is clear since an upper-triangular matrix always has exactly one non-zero diagonal, which is its main diagonal. The only if direction is proved by induction on the dimension n of A.

Step 1—Induction basis: When $n = 2$, it is easy to check that any 2×2 matrix with exactly one diagonal is either of the form $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ or $\begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$, where $a, c \neq 0$. So it is a permutation of an upper-triangular matrix.

Step 2—Inductive step: Suppose that the claim—an $n \times n$-square matrix A has exactly one non-zero diagonal if and only if it is a permutation of an upper-triangular matrix—holds for any $n < N$. We need to show that the claim also holds when $n = N$.

The following notation will be used. Let A be an $n \times n$-square matrix. $A_{i,j}$ denotes the element of A at row i and column j. $\widetilde{A}_{i,j}$ denotes the $(n-1) \times (n-1)$ sub-matrix obtained from A by crossing out row i and column j.

First, we will prove three handy observations.

Observation 1: If A has exactly one non-zero diagonal and $A_{i,j} \neq 0$, then $\widetilde{A}_{i,j}$ has at most one non-zero diagonal. In particular, if $A_{i,j}$ is on that non-zero diagonal, then $\widetilde{A}_{i,j}$ has exactly one non-zero diagonal.

Proof of Observation 1: Suppose that $\widetilde{A}_{i,j}$ has more than one diagonal. Then these diagonals for $\widetilde{A}_{i,j}$ along with $A_{i,j}$ form different diagonals for A, which is a contradiction.

Observation 2: If A has exactly one non-zero diagonal and A has a row or a column with exactly one non-zero element, then A is a permutation of an upper-triangular matrix.

Proof of Observation 2: Suppose that A has a column with exactly one non-zero element. Then by permutation, we may assume that it is the first column of A and the only non-zero element in column 1 is $A_{1,1}$. $A_{1,1}$ must be on the non-zero diagonal of A. Hence, according to observation 1, $\widetilde{A}_{1,1}$ is a $(N-1 \times N-1)$-square matrix with exactly one non-zero diagonal. Then by the inductive assumption, we may permute $\widetilde{A}_{1,1}$ into an upper-triangular matrix. Note that each permutation of $\widetilde{A}_{1,1}$ induces a permutation of A. So there exist permutations that convert A into A' such that $A'_{i,j} = 0$ when $j > 1$ and $i > j$. Moreover, permutations that convert A to A' never switch column 1 (row 1) of A with any other columns (rows). So $A'_{i,1} = 0$ for $i \neq 1$, as $A_{1,1}$ is the only non-zero element in the first column of A. Thus, we have $A'_{i,j} = 0$ when $i > j$, which implies that A' is an upper-triangular matrix.

If A has a row with exactly one non-zero element, then up to permutation, we may assume it is the last row of A and the only non-zero element is $A_{N,N}$. Following similar argument as above, we may show that $\widetilde{A}_{N,N}$ can be arranged into an upper-triangular matrix by permutations. The corresponding permutations of A will also convert A into an upper triangular matrix. So observation 2 holds.

Observation 3: If the main diagonal of A is the only non-zero diagonal of A, then $A_{t_1,t_2} A_{t_2,t_3} \cdots A_{t_{k-1},t_k} A_{t_k,t_1} = 0$ for any distinct $t_1, t_2, \ldots, t_k$.

Proof of Observation 3: Suppose that $A_{t_1,t_2} A_{t_2,t_3} \cdots A_{t_{k-1},t_k} A_{t_k,t_1} \neq 0$. Then a different non-zero diagonal for A other than the main diagonal is form by $\{A_{i,i} | i \neq t_1, \ldots, t_k\}$ and $A_{t_1,t_2}, A_{t_2,t_3}, \cdots, A_{t_{k-1},t_k}, A_{t_k,t_1}$.

Now back to the inductive step. Suppose that A is an $N \times N$-square matrix with exactly one non-zero diagonal. By permutation, we may assume that the main diagonal of A is the only non-zero diagonal. In particular, $A_{1,1} \neq 0$. According to Observation 1, $\widetilde{A}_{1,1}$ has exactly one non-zero diagonal and so can be arranged into an upper-triangular matrix by permutations. The corresponding permutations convert A into a new form, denoted by A^1, with the property that $A^1_{i,j} = 0$ when $j > 1$ and $i > j$. In particular, $A^1_{Nj} = 0$ when $j \neq 1$ and $j \neq N$. $\widetilde{A}^1_{1,1}$ is an upper-triangular matrix implies that $A^1_{N,N} \neq 0$. If $A^1_{N,1} = 0$, then the last row of A^1 contains only one non-zero element $A^1_{N,N}$. So by Observation 2, we are done.

Otherwise, according to Observation 1, $\widetilde{A}^1_{N,N}$ can be arranged into an upper-triangular matrix by permutation. Hence, after the corresponding permutations, we may convert A^1 into a new form, denoted by A^2 with the property that $A^2_{i,j} = 0$ when $i > j$ and $i \neq N$. Moreover, permutations that convert A^1 to A^2 never switch row N (column N) of A^1 with any other rows (columns). So only one of $\{A^2_{N,j} | j \neq N\}$ is not zero. If $A^2_{N,1} = 0$, along with $A^2_{i,1} = 0$ for $N > i > 1$, we have that the first column of A^2 contains exactly one non-zero element, $A^2_{1,1}$. So by Observation 2, we are done.

Otherwise, $A^2_{N,1} \neq 0$. According to Observation 3, $A^2_{N,1} A^2_{1,k} A^2_{k,N} = 0$, for $k = 2, \ldots, N-1$. So we have that $A^2_{1,k} A^2_{k,N} = 0$, for $k = 2, \ldots, N-1$. We will proceed by analyzing cases from $k = 2$ to $k = N-1$.

When $k = 2$, if $A^2_{1,2} = 0$, then column 2 of A^2 contains only one non-zero element $A^2_{2,2}$, and we are done by Observation 2. Otherwise, we may assume that $A^2_{1,2} \neq 0$ and $A^2_{2,N} = 0$.

When $k = 3$, if $A^2_{3,N} \neq 0$, then $A^2_{1,3} = 0$. According to Observation 3, $A^2_{N,1}A^2_{1,2}A^2_{2,3}A^2_{3,N} = 0$, and this implies that $A^2_{2,3} = 0$. Hence, column 3 of A^2 contains only one non-zero element, $A^2_{3,3}$, and again we are done by Observation 2. Otherwise, we may assume that $A^2_{3,N} = 0$, and one of $\{A^2_{1,3}, A^2_{2,3}\}$ is not zero.

When $k = k$, if $A^2_{4,N} \neq 0$, then $A^2_{1,4} = 0$. Similarly, as in the case where $k = 3$ (by Observation 3), $A^2_{N,1}A^2_{1,2}A^2_{2,4}A^2_{3,N} = 0$, and this implies that $A^2_{2,4} = 0$. One of $\{A^2_{1,3}, A^2_{2,3}\}$ is not zero $\Longrightarrow$ either $A^2_{N,1}A^2_{1,3}A^2_{3,4}A^2_{3,N} = 0$ or $A^2_{N,1}A^2_{1,2}A^2_{2,3}A^2_{3,4}A^2_{3,N} = 0 \Longrightarrow A^2_{3,4} = 0$. Hence, column 4 of A^2 contains only one non-zero element, $A^2_{4,4}$, and again we are done by Observation 2. Otherwise, we may assume that $A^2_{4,N} = 0$, and at least one of $\{A^2_{1,4}, A^2_{2,4}, A^2_{3,4}\}$ is not zero.

Inductively, either one of column k's of A^2 contains only one non-zero element, or $A^2_{k,N} = 0$ for all $k = 2, \ldots, N-1$. Note that the latter case implies that column N of A^2 contains only one non-zero element, $A^2_{N,N}$, as $A^2_{N,1} \neq 0 \Longrightarrow A^2_{1,N} = 0$. Either way, the proof is then completed by Observation 2.

□

4 Details to Example 4.11

To construct $\mathbf{M}$, first notice that if maximum likelihood is achieved, $\mathbf{M}_{1,1} = \mathbf{M}_{1,2}$ under all settings of Δ, a, and q. This is because a first- and second-order polynomial give the same fit to D_1.

For $\mathbf{M}_{2,1}$, by symmetry arguments we know that the maximum-likelihood fit of a first-order polynomial to D_2 is a horizontal line ($f(x) = b$). We can find this value of b through a grid search. Given this b,

$$\mathbf{M}_{2,1} = N_q(a;b)^2 N_q(-a;b)^2 N_q(\Delta + a;b) N_q(\Delta - a;b),$$

where

$$N_q(z;b) = \frac{\sqrt{\beta}}{C_q} e_q(-\beta(x_i - \mu)^2).$$

Here, $\beta = \frac{1}{5-3q}$ so that the variance is 1; $e_q(x)$ is the q-exponential function defined by $[1 + (1-q)x]^{\frac{1}{1-q}}$ when $q \neq 1$, and $\exp(x)$ when $q = 1$. The normalizing constant C_q is given by:

$$C_q = \begin{cases} \frac{2\sqrt{\pi}\Gamma(\frac{1}{1-q})}{(3-q)\sqrt{1-q}\Gamma(\frac{3-q}{2(1-q)})} & \text{for } -\infty < q < 1 \\ \sqrt{\pi} & \text{for } q = 1 \\ \frac{\sqrt{\pi}\Gamma(\frac{3-q}{2(q-1)}}{\sqrt{q-1}\Gamma\frac{1}{q-1}} & \text{for } 1 < q < 3. \end{cases}$$

For $\mathbf{M}_{2,2}$, again by symmetry arguments we know that the maximum-likelihood fit of a second order polynomial to D_2 is a parabola that passes through the middle of each of the three pairs of data points. Thus, $\mathbf{M}_{2,2} = N_q(a;0)^6$.

References

[1] Garrett Birkhoff. Three observations on linear algebra. *Univ. Nac. Tucumán. Revista A*, 5:147–151, 1946.

[2] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[3] John Von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games*, 2:5–12, 1953.