FORMALIZED MATHEMATICS
2007, Vol. **15**, No. **3**, Pages 111–119
DOI: 10.2478/v10037-007-0012-9

Definition and some Properties of Information Entropy

Bo Zhang
Shinshu University
Nagano, Japan

Yatsuka Nakamura
Shinshu University
Nagano, Japan

Summary. In this article we mainly define the information entropy [3], [11] and prove some its basic properties. First, we discuss some properties on four kinds of transformation functions between vector and matrix. The transformation functions are LineVec2Mx, ColVec2Mx, Vec2DiagMx and Mx2FinS. Mx2FinS is a horizontal concatenation operator for a given matrix, treating rows of the given matrix as finite sequences, yielding a new finite sequence by horizontally joining each row of the given matrix in order to index. Then we define each concept of information entropy for a probability sequence and two kinds of probability matrices, joint and conditional, that are defined in article [25]. Further, we discuss some properties of information entropy including Shannon's lemma, maximum property, additivity and super-additivity properties.

MML identifier: `ENTROPY1`, version: `7.8.05 4.84.971`

The papers [21], [23], [1], [20], [24], [6], [14], [8], [4], [22], [17], [7], [9], [2], [5], [15], [16], [12], [10], [13], [18], [25], and [19] provide the terminology and notation for this paper.

1. Preliminaries

For simplicity, we use the following convention: D denotes a non empty set, i, j, k, l denote elements of $\mathbb{N}$, n denotes a natural number, a, b, c, r, r_1, r_2 denote real numbers, p, q denote finite sequences of elements of $\mathbb{R}$, and M_1, M_2 denote matrices over $\mathbb{R}$.

Next we state several propositions:

(1) If $k \neq 0$ and $i < l$ and $l \leq j$ and $k \mid l$, then $i \div k < j \div k$.

ISSN 1426–2630

(2) If $r > 0$, then $(\log\_(e))(r) \leq r - 1$ and $r = 1$ iff $(\log\_(e))(r) = r - 1$ and $r \neq 1$ iff $(\log\_(e))(r) < r - 1$.

(3) If $r > 0$, then $\log_e r \leq r - 1$ and $r = 1$ iff $\log_e r = r - 1$ and $r \neq 1$ iff $\log_e r < r - 1$.

(4) If $a > 1$ and $b > 1$, then $\log_a b > 0$.

(5) If $a > 0$ and $a \neq 1$ and $b > 0$, then $-\log_a b = \log_a(\frac{1}{b})$.

(6) If $a > 0$ and $a \neq 1$ and $b \geq 0$ and $c \geq 0$, then $b \cdot c \cdot \log_a(b \cdot c) = b \cdot c \cdot \log_a b + b \cdot c \cdot \log_a c$.

(7) Let q, q_1, q_2 be finite sequences of elements of $\mathbb{R}$. Suppose $\operatorname{len} q_1 = \operatorname{len} q$ and $\operatorname{len} q_1 = \operatorname{len} q_2$ and for every k such that $k \in \operatorname{dom} q_1$ holds $q(k) = q_1(k) + q_2(k)$. Then $\sum q = \sum q_1 + \sum q_2$.

(8) Let q, q_1, q_2 be finite sequences of elements of $\mathbb{R}$. Suppose $\operatorname{len} q_1 = \operatorname{len} q$ and $\operatorname{len} q_1 = \operatorname{len} q_2$ and for every k such that $k \in \operatorname{dom} q_1$ holds $q(k) = q_1(k) - q_2(k)$. Then $\sum q = \sum q_1 - \sum q_2$.

(9) Suppose $\operatorname{len} p \geq 1$. Then there exists q such that $\operatorname{len} q = \operatorname{len} p$ and $q(1) = p(1)$ and for every k such that $0 \neq k$ and $k < \operatorname{len} p$ holds $q(k+1) = q(k) + p(k+1)$ and $\sum p = q(\operatorname{len} p)$.

Let us consider p. Let us observe that p is non-negative if and only if:

(Def. 1) For every i such that $i \in \operatorname{dom} p$ holds $p(i) \geq 0$.

Let us note that there exists a finite sequence of elements of $\mathbb{R}$ which is non-negative.

The following proposition is true

(10) If p is non-negative and $r \geq 0$, then $r \cdot p$ is non-negative.

Let us consider p, k. We say that p has only one value in k if and only if:

(Def. 2) $k \in \operatorname{dom} p$ and for every i such that $i \in \operatorname{dom} p$ and $i \neq k$ holds $p(i) = 0$.

Next we state four propositions:

(11) If p has only one value in k and $i \neq k$, then $p(i) = 0$.

(12) If $\operatorname{len} p = \operatorname{len} q$ and p has only one value in k, then $p \bullet q$ has only one value in k and $(p \bullet q)(k) = p(k) \cdot q(k)$.

(13) If p has only one value in k, then $\sum p = p(k)$.

(14) If p is non-negative, then for every k such that $k \in \operatorname{dom} p$ and $p(k) = \sum p$ holds p has only one value in k.

Let us observe that every finite sequence of elements of $\mathbb{R}$ which is finite probability distribution is also non empty and non-negative.

One can prove the following propositions:

(15) Let p be finite probability distribution finite sequence of elements of $\mathbb{R}$ and given k such that $k \in \operatorname{dom} p$ and $p(k) = 1$. Then p has only one value in k.

(16) Let i be a non empty natural number. Then $i \mapsto \frac{1}{i}$ is finite probability distribution finite sequence of elements of $\mathbb{R}$.

One can check that every matrix over $\mathbb{R}$ which is summable-to-1 is also non empty yielding and every matrix over $\mathbb{R}$ which is joint probability is also non empty yielding.

The following propositions are true:

(17) For every matrix M over $\mathbb{R}$ such that $M = \emptyset$ holds $\mathrm{SumAll}\, M = 0$.

(18) For every matrix M over D and for every i such that $i \in \mathrm{dom}\, M$ holds $\mathrm{dom}\, M(i) = \mathrm{Seg}\, \mathrm{width}\, M$.

(19) M_1 is nonnegative iff for every i such that $i \in \mathrm{dom}\, M_1$ holds $\mathrm{Line}(M_1, i)$ is non-negative.

2. Properties of Transformations between Vector and Matrix

Next we state four propositions:

(20) For every j such that $j \in \mathrm{dom}\, p$ holds $(\mathrm{LineVec2Mx}\, p)_{\square,j} = \langle p(j) \rangle$.

(21) Let p be a non empty finite sequence of elements of $\mathbb{R}$, q be a finite sequence of elements of $\mathbb{R}$, and M be a matrix over $\mathbb{R}$. Then $M = \mathrm{ColVec2Mx}\, p \cdot \mathrm{LineVec2Mx}\, q$ if and only if the following conditions are satisfied:

  (i) $\mathrm{len}\, M = \mathrm{len}\, p$,

  (ii) $\mathrm{width}\, M = \mathrm{len}\, q$, and

  (iii) for all i, j such that $\langle i, j \rangle \in$ the indices of M holds $M_{i,j} = p(i) \cdot q(j)$.

(22) Let p be a non empty finite sequence of elements of $\mathbb{R}$, q be a finite sequence of elements of $\mathbb{R}$, and M be a matrix over $\mathbb{R}$. Then $M = \mathrm{ColVec2Mx}\, p \cdot \mathrm{LineVec2Mx}\, q$ if and only if the following conditions are satisfied:

  (i) $\mathrm{len}\, M = \mathrm{len}\, p$,

  (ii) $\mathrm{width}\, M = \mathrm{len}\, q$, and

  (iii) for every i such that $i \in \mathrm{dom}\, M$ holds $\mathrm{Line}(M, i) = p(i) \cdot q$.

(23) Let p, q be finite probability distribution finite sequences of elements of $\mathbb{R}$. Then $\mathrm{ColVec2Mx}\, p \cdot \mathrm{LineVec2Mx}\, q$ is joint probability.

Let us consider n and let M_1 be a matrix over $\mathbb{R}$ of dimension n. We say that M_1 is diagonal if and only if:

(Def. 3) For all i, j such that $\langle i, j \rangle \in$ the indices of M_1 and $(M_1)_{i,j} \neq 0$ holds $i = j$.

Let us consider n. Observe that there exists a matrix over $\mathbb{R}$ of dimension n which is diagonal.

The following proposition is true

(24) Let M_1 be a matrix over $\mathbb{R}$ of dimension n. Then M_1 is diagonal if and only if for every i such that $i \in \operatorname{dom} M_1$ holds $\operatorname{Line}(M_1, i)$ has only one value in i.

Let us consider p. The functor $\operatorname{Vec2DiagMx} p$ yielding a diagonal matrix over $\mathbb{R}$ of dimension $\operatorname{len} p$ is defined as follows:

(Def. 4) For every j such that $j \in \operatorname{dom} p$ holds $(\operatorname{Vec2DiagMx} p)_{j,j} = p(j)$.

One can prove the following propositions:

(25) $M_1 = \operatorname{Vec2DiagMx} p$ iff $\operatorname{len} M_1 = \operatorname{len} p$ and $\operatorname{width} M_1 = \operatorname{len} p$ and for every i such that $i \in \operatorname{dom} M_1$ holds $\operatorname{Line}(M_1, i)$ has only one value in i and $\operatorname{Line}(M_1, i)(i) = p(i)$.

(26) Suppose $\operatorname{len} p = \operatorname{len} M_1$. Then $M_2 = \operatorname{Vec2DiagMx} p \cdot M_1$ if and only if the following conditions are satisfied:
  (i) $\operatorname{len} M_2 = \operatorname{len} p$,
  (ii) $\operatorname{width} M_2 = \operatorname{width} M_1$, and
  (iii) for all i, j such that $\langle i, j \rangle \in$ the indices of M_2 holds $(M_2)_{i,j} = p(i) \cdot (M_1)_{i,j}$.

(27) If $\operatorname{len} p = \operatorname{len} M_1$, then $M_2 = \operatorname{Vec2DiagMx} p \cdot M_1$ iff $\operatorname{len} M_2 = \operatorname{len} p$ and $\operatorname{width} M_2 = \operatorname{width} M_1$ and for every i such that $i \in \operatorname{dom} M_2$ holds $\operatorname{Line}(M_2, i) = p(i) \cdot \operatorname{Line}(M_1, i)$.

(28) Let p be finite probability distribution finite sequence of elements of $\mathbb{R}$ and M be a non empty yielding conditional probability matrix over $\mathbb{R}$. If $\operatorname{len} p = \operatorname{len} M$, then $\operatorname{Vec2DiagMx} p \cdot M$ is joint probability.

(29) Let M be a matrix over D and p be a finite sequence of elements of D^*. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) ^\frown M(k+1)$. Let given k. If $k \in \operatorname{dom} p$, then $\operatorname{len} p(k) = k \cdot \operatorname{width} M$.

(30) Let M be a matrix over D and p be a finite sequence of elements of D^*. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) ^\frown M(k+1)$. Let given i, j. If $i \in \operatorname{dom} p$ and $j \in \operatorname{dom} p$ and $i \leq j$, then $\operatorname{dom} p(i) \subseteq \operatorname{dom} p(j)$.

(31) Let M be a matrix over D and p be a finite sequence of elements of D^*. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) ^\frown M(k+1)$. Then $\operatorname{len} p(1) = \operatorname{width} M$ and for every j such that $\langle 1, j \rangle \in$ the indices of M holds $j \in \operatorname{dom} p(1)$ and $p(1)(j) = M_{1,j}$.

(32) Let M be a matrix over D and p be a finite sequence of elements of D^*. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) ^\frown M(k+1)$. Let given j. If $j \geq 1$ and $j < \operatorname{len} p$, then for every l such that $l \in \operatorname{dom} p(j)$ holds $p(j)(l) = p(j+1)(l)$.

(33) Let M be a matrix over D and p be a finite sequence of elements of D^*.

Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) \frown M(k+1)$. Let given i, j. Suppose $i \in \operatorname{dom} p$ and $j \in \operatorname{dom} p$ and $i \leq j$. Let given l. If $l \in \operatorname{dom} p(i)$, then $p(i)(l) = p(j)(l)$.

(34) Let M be a matrix over D and p be a finite sequence of elements of D^*. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) \frown M(k+1)$. Let given j. Suppose $j \geq 1$ and $j < \operatorname{len} p$. Let given l. If $l \in \operatorname{Seg} \operatorname{width} M$, then $j \cdot \operatorname{width} M + l \in \operatorname{dom} p(j+1)$ and $p(j+1)(j \cdot \operatorname{width} M + l) = M(j+1)(l)$.

(35) Let M be a matrix over D and p be a finite sequence of elements of D^*. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) \frown M(k+1)$. Let given i, j. Suppose $\langle i, j \rangle \in$ the indices of M. Then $(i-1) \cdot \operatorname{width} M + j \in \operatorname{dom} p(i)$ and $M_{i,j} = p(i)((i-1) \cdot \operatorname{width} M + j)$.

(36) Let M be a matrix over D and p be a finite sequence of elements of D^*. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) \frown M(k+1)$. Let given i, j. Suppose $\langle i, j \rangle \in$ the indices of M. Then $(i-1) \cdot \operatorname{width} M + j \in \operatorname{dom} p(\operatorname{len} M)$ and $M_{i,j} = p(\operatorname{len} M)((i-1) \cdot \operatorname{width} M + j)$.

(37) Let M be a matrix over $\mathbb{R}$ and p be a finite sequence of elements of $\mathbb{R}^*$. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) \frown M(k+1)$. Let given k. If $k \geq 1$ and $k < \operatorname{len} M$, then $\sum p(k+1) = \sum p(k) + \sum M(k+1)$.

(38) Let M be a matrix over $\mathbb{R}$ and p be a finite sequence of elements of $\mathbb{R}^*$. Suppose $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) \frown M(k+1)$. Then $\operatorname{SumAll} M = \sum p(\operatorname{len} M)$.

Let D be a non empty set and let M be a matrix over D. The functor $\operatorname{Mx2FinS} M$ yields a finite sequence of elements of D and is defined by:

(Def. 5)(i) $\operatorname{Mx2FinS} M = \emptyset$ if $\operatorname{len} M = 0$,

(ii) there exists a finite sequence p of elements of D^* such that $\operatorname{Mx2FinS} M = p(\operatorname{len} M)$ and $\operatorname{len} p = \operatorname{len} M$ and $p(1) = M(1)$ and for every k such that $k \geq 1$ and $k < \operatorname{len} M$ holds $p(k+1) = p(k) \frown M(k+1)$, otherwise.

We now state several propositions:

(39) For every matrix M over D holds $\operatorname{len} \operatorname{Mx2FinS} M = \operatorname{len} M \cdot \operatorname{width} M$.

(40) Let M be a matrix over D and given i, j. If $\langle i, j \rangle \in$ the indices of M, then $(i-1) \cdot \operatorname{width} M + j \in \operatorname{dom} \operatorname{Mx2FinS} M$ and $M_{i,j} = (\operatorname{Mx2FinS} M)((i-1) \cdot \operatorname{width} M + j)$.

(41) Let M be a matrix over D and given k, l. Suppose $k \in \operatorname{dom} \operatorname{Mx2FinS} M$

and $l = k - 1$. Then $\langle (l \div \mathrm{width}\, M) + 1, (l \bmod \mathrm{width}\, M) + 1 \rangle \in$ the indices of M and $(\mathrm{Mx2FinS}\, M)(k) = M_{(l \div \mathrm{width}\, M)+1,(l \bmod \mathrm{width}\, M)+1}$.

(42) $\mathrm{SumAll}\, M_1 = \sum \mathrm{Mx2FinS}\, M_1$.

(43) M_1 is nonnegative iff $\mathrm{Mx2FinS}\, M_1$ is non-negative.

(44) M_1 is joint probability iff $\mathrm{Mx2FinS}\, M_1$ is finite probability distribution.

(45) Let p, q be finite probability distribution finite sequences of elements of $\mathbb{R}$. Then $\mathrm{Mx2FinS}(\mathrm{ColVec2Mx}\, p \cdot \mathrm{LineVec2Mx}\, q)$ is finite probability distribution.

(46) Let p be finite probability distribution finite sequence of elements of $\mathbb{R}$ and M be a non empty yielding conditional probability matrix over $\mathbb{R}$. If $\mathrm{len}\, p = \mathrm{len}\, M$, then $\mathrm{Mx2FinS}(\mathrm{Vec2DiagMx}\, p \cdot M)$ is finite probability distribution.

3. Information Entropy

Let us consider a, p. Let us assume that $a > 0$ and $a \neq 1$ and p is non-negative. The functor $\overrightarrow{\log_a} p$ yields a finite sequence of elements of $\mathbb{R}$ and is defined by:

(Def. 6) $\mathrm{len}\, \overrightarrow{\log_a} p = \mathrm{len}\, p$ and for every k such that $k \in \mathrm{dom}\, \overrightarrow{\log_a} p$ holds if $p(k) > 0$, then $(\overrightarrow{\log_a} p)(k) = \log_a p(k)$ and if $p(k) = 0$, then $(\overrightarrow{\log_a} p)(k) = 0$.

Let us consider p. The functor $\overrightarrow{\mathrm{id}\ \log}\, p$ yields a finite sequence of elements of $\mathbb{R}$ and is defined by:

(Def. 7) $\overrightarrow{\mathrm{id}\ \log}\, p = p \bullet \overrightarrow{\log_2} p$.

The following propositions are true:

(47) Let p be a non-negative finite sequence of elements of $\mathbb{R}$ and given q. Then $q = \overrightarrow{\mathrm{id}\ \log}\, p$ if and only if the following conditions are satisfied:
- (i) $\mathrm{len}\, q = \mathrm{len}\, p$, and
- (ii) for every k such that $k \in \mathrm{dom}\, q$ holds $q(k) = p(k) \cdot \log_2 p(k)$.

(48) Let p be a non-negative finite sequence of elements of $\mathbb{R}$ and given k such that $k \in \mathrm{dom}\, p$. Then
- (i) if $p(k) = 0$, then $(\overrightarrow{\mathrm{id}\ \log}\, p)(k) = 0$, and
- (ii) if $p(k) > 0$, then $(\overrightarrow{\mathrm{id}\ \log}\, p)(k) = p(k) \cdot \log_2 p(k)$.

(49) Let p be a non-negative finite sequence of elements of $\mathbb{R}$ and given q. Then $q = -\overrightarrow{\mathrm{id}\ \log}\, p$ if and only if the following conditions are satisfied:
- (i) $\mathrm{len}\, q = \mathrm{len}\, p$, and
- (ii) for every k such that $k \in \mathrm{dom}\, q$ holds $q(k) = p(k) \cdot \log_2(\frac{1}{p(k)})$.

(50) Let p be a non-negative finite sequence of elements of $\mathbb{R}$. Suppose $r_1 \geq 0$ and $r_2 \geq 0$. Let given i. If $i \in \mathrm{dom}\, p$ and $p(i) = r_1 \cdot r_2$, then $(\overrightarrow{\mathrm{id}\ \log}\, p)(i) = r_1 \cdot r_2 \cdot \log_2 r_1 + r_1 \cdot r_2 \cdot \log_2 r_2$.

(51) For every non-negative finite sequence p of elements of $\mathbb{R}$ such that $r \geq 0$ holds $\overrightarrow{\mathrm{id}\,\log}\, r \cdot p = r \cdot \log_2 r \cdot p + r \cdot (p \bullet \overrightarrow{\log_2} p)$.

(52) Let p be a non empty finite probability distribution finite sequence of elements of $\mathbb{R}$ and given k. If $k \in \operatorname{dom} p$, then $(\overrightarrow{\mathrm{id}\,\log}\, p)(k) \leq 0$.

Let us consider M_1. Let us assume that M_1 is nonnegative. The functor $\overrightarrow{\mathrm{id}\,\log}\, M_1$ yields a matrix over $\mathbb{R}$ and is defined as follows:

(Def. 8) $\operatorname{len} \overrightarrow{\mathrm{id}\,\log}\, M_1 = \operatorname{len} M_1$ and $\operatorname{width} \overrightarrow{\mathrm{id}\,\log}\, M_1 = \operatorname{width} M_1$ and for every k such that $k \in \operatorname{dom} \overrightarrow{\mathrm{id}\,\log}\, M_1$ holds $(\overrightarrow{\mathrm{id}\,\log}\, M_1)(k) = \mathrm{Line}(M_1, k) \bullet \overrightarrow{\log_2} \mathrm{Line}(M_1, k)$.

The following two propositions are true:

(53) For every nonnegative matrix M over $\mathbb{R}$ and for every k such that $k \in \operatorname{dom} M$ holds $\mathrm{Line}(\overrightarrow{\mathrm{id}\,\log}\, M, k) = \overrightarrow{\mathrm{id}\,\log}\, \mathrm{Line}(M, k)$.

(54) Let M be a nonnegative matrix over $\mathbb{R}$ and M_3 be a matrix over $\mathbb{R}$. Then $M_3 = \overrightarrow{\mathrm{id}\,\log}\, M$ if and only if the following conditions are satisfied:
  (i) $\operatorname{len} M_3 = \operatorname{len} M$,
  (ii) $\operatorname{width} M_3 = \operatorname{width} M$, and
  (iii) for all i, j such that $\langle i, j \rangle \in$ the indices of M_3 holds $(M_3)_{i,j} = M_{i,j} \cdot \log_2(M_{i,j})$.

Let p be a finite sequence of elements of $\mathbb{R}$. The functor $\mathrm{Entropy}\, p$ yields a real number and is defined by:

(Def. 9) $\mathrm{Entropy}\, p = -\sum \overrightarrow{\mathrm{id}\,\log}\, p$.

We now state several propositions:

(55) For every non empty finite probability distribution finite sequence p of elements of $\mathbb{R}$ holds $\mathrm{Entropy}\, p \geq 0$.

(56) Let p be a non empty finite probability distribution finite sequence of elements of $\mathbb{R}$. If there exists k such that $k \in \operatorname{dom} p$ and $p(k) = 1$, then $\mathrm{Entropy}\, p = 0$.

(57) Let p, q be non empty finite probability distribution finite sequences of elements of $\mathbb{R}$ and p_1, q_3 be finite sequences of elements of $\mathbb{R}$. Suppose that
  (i) $\operatorname{len} p = \operatorname{len} q$,
  (ii) $\operatorname{len} p_1 = \operatorname{len} p$,
  (iii) $\operatorname{len} q_3 = \operatorname{len} q$, and
  (iv) for every k such that $k \in \operatorname{dom} p$ holds $p(k) > 0$ and $q(k) > 0$ and $p_1(k) = -p(k) \cdot \log_2 p(k)$ and $q_3(k) = -p(k) \cdot \log_2 q(k)$.

  Then
  (v) $\sum p_1 \leq \sum q_3$,
  (vi) for every k such that $k \in \operatorname{dom} p$ holds $p(k) = q(k)$ iff $\sum p_1 = \sum q_3$, and
  (vii) there exists k such that $k \in \operatorname{dom} p$ and $p(k) \neq q(k)$ iff $\sum p_1 < \sum q_3$.

(58) Let p be a non empty finite probability distribution finite sequence of elements of $\mathbb{R}$. Suppose that for every k such that $k \in \operatorname{dom} p$ holds $p(k) > 0$. Then

(i) $\operatorname{Entropy} p \leq \log_2 \operatorname{len} p$,

(ii) for every k such that $k \in \operatorname{dom} p$ holds $p(k) = \frac{1}{\operatorname{len} p}$ iff $\operatorname{Entropy} p = \log_2 \operatorname{len} p$, and

(iii) there exists k such that $k \in \operatorname{dom} p$ and $p(k) \neq \frac{1}{\operatorname{len} p}$ iff $\operatorname{Entropy} p < \log_2 \operatorname{len} p$.

(59) For every nonnegative matrix M over $\mathbb{R}$ holds $\operatorname{Mx2FinS} \overrightarrow{\mathrm{id}\ \log} M = \overrightarrow{\mathrm{id}\ \log} \operatorname{Mx2FinS} M$.

(60) Let p, q be finite probability distribution finite sequences of elements of $\mathbb{R}$ and M be a matrix over $\mathbb{R}$. If $M = \operatorname{ColVec2Mx} p \cdot \operatorname{LineVec2Mx} q$, then $\operatorname{SumAll} \overrightarrow{\mathrm{id}\ \log} M = \sum \overrightarrow{\mathrm{id}\ \log} p + \sum \overrightarrow{\mathrm{id}\ \log} q$.

Let us consider M_1. The entropy of joint probability of M_1 yields a real number and is defined as follows:

(Def. 10) The entropy of joint probability of $M_1 = \operatorname{Entropy} \operatorname{Mx2FinS} M_1$.

Next we state the proposition

(61) Let p, q be finite probability distribution finite sequences of elements of $\mathbb{R}$. Then the entropy of joint probability of $\operatorname{ColVec2Mx} p \cdot \operatorname{LineVec2Mx} q = \operatorname{Entropy} p + \operatorname{Entropy} q$.

Let us consider M_1. The entropy of conditional probability of M_1 yields a finite sequence of elements of $\mathbb{R}$ and is defined by the conditions (Def. 11).

(Def. 11)(i) len (the entropy of conditional probability of M_1) $= \operatorname{len} M_1$, and

(ii) for every k such that $k \in$ dom (the entropy of conditional probability of M_1) holds (the entropy of conditional probability of M_1)$(k) = \operatorname{Entropy} \operatorname{Line}(M_1, k)$.

One can prove the following propositions:

(62) Let M be a non empty yielding conditional probability matrix over $\mathbb{R}$ and p be a finite sequence of elements of $\mathbb{R}$. Then $p =$ the entropy of conditional probability of M if and only if $\operatorname{len} p = \operatorname{len} M$ and for every k such that $k \in \operatorname{dom} p$ holds $p(k) = -\sum(\overrightarrow{\mathrm{id}\ \log} M)(k)$.

(63) Let M be a non empty yielding conditional probability matrix over $\mathbb{R}$. Then the entropy of conditional probability of $M = -\operatorname{LineSum} \overrightarrow{\mathrm{id}\ \log} M$.

(64) Let p be finite probability distribution finite sequence of elements of $\mathbb{R}$ and M be a non empty yielding conditional probability matrix over $\mathbb{R}$. Suppose $\operatorname{len} p = \operatorname{len} M$. Let M_3 be a matrix over $\mathbb{R}$. If $M_3 = \operatorname{Vec2DiagMx} p \cdot M$, then $\operatorname{SumAll} \overrightarrow{\mathrm{id}\ \log} M_3 = \sum \overrightarrow{\mathrm{id}\ \log} p + \sum(p \bullet \operatorname{LineSum} \overrightarrow{\mathrm{id}\ \log} M)$.

(65) Let p be finite probability distribution finite sequence of elements of $\mathbb{R}$ and M be a non empty yielding conditional probability matrix over

$\mathbb{R}$. Suppose $\operatorname{len} p = \operatorname{len} M$. Then the entropy of joint probability of $\mathrm{Vec2DiagMx}\, p \cdot M = \mathrm{Entropy}\, p + \sum(p \bullet$ the entropy of conditional probability of $M)$.

References

[1] Grzegorz Bancerek. The ordinal numbers. *Formalized Mathematics*, 1(**1**):91–96, 1990.
[2] Grzegorz Bancerek and Krzysztof Hryniewiecki. Segments of natural numbers and finite sequences. *Formalized Mathematics*, 1(**1**):107–114, 1990.
[3] P. Billingsley. *Ergodic Theory and Information.* John Wiley & Sons, 1964.
[4] Czesław Byliński. Binary operations. *Formalized Mathematics*, 1(**1**):175–180, 1990.
[5] Czesław Byliński. Finite sequences and tuples of elements of a non-empty sets. *Formalized Mathematics*, 1(**3**):529–536, 1990.
[6] Czesław Byliński. Functions and their basic properties. *Formalized Mathematics*, 1(**1**):55–65, 1990.
[7] Czesław Byliński. Functions from a set to a set. *Formalized Mathematics*, 1(**1**):153–164, 1990.
[8] Czesław Byliński. Some basic properties of sets. *Formalized Mathematics*, 1(**1**):47–53, 1990.
[9] Czesław Byliński. The sum and product of finite sequences of real numbers. *Formalized Mathematics*, 1(**4**):661–668, 1990.
[10] Agata Darmochwał. The Euclidean space. *Formalized Mathematics*, 2(**4**):599–603, 1991.
[11] Shigeichi Hirasawa. *Information Theory.* Baifukan CO., 1996.
[12] Katarzyna Jankowska. Matrices. Abelian group of matrices. *Formalized Mathematics*, 2(**4**):475–480, 1991.
[13] Artur Korniłowicz. On the real valued functions. *Formalized Mathematics*, 13(**1**):181–187, 2005.
[14] Jarosław Kotowicz. Real sequences and basic operations on them. *Formalized Mathematics*, 1(**2**):269–272, 1990.
[15] Rafał Kwiatek. Factorial and Newton coefficients. *Formalized Mathematics*, 1(**5**):887–890, 1990.
[16] Yatsuka Nakamura, Nobuyuki Tamaura, and Wenpai Chang. A theory of matrices of real elements. *Formalized Mathematics*, 14(**1**):21–28, 2006.
[17] Library Committee of the Association of Mizar Users. Binary operations on numbers. *To appear in Formalized Mathematics.*
[18] Konrad Raczkowski and Andrzej Nędzusiak. Real exponents and logarithms. *Formalized Mathematics*, 2(**2**):213–216, 1991.
[19] Yasunari Shidama. The Taylor expansions. *Formalized Mathematics*, 12(**2**):195–200, 2004.
[20] Andrzej Trybulec. Subsets of complex numbers. *To appear in Formalized Mathematics.*
[21] Andrzej Trybulec. Tarski Grothendieck set theory. *Formalized Mathematics*, 1(**1**):9–11, 1990.
[22] Wojciech A. Trybulec. Vectors in real linear space. *Formalized Mathematics*, 1(**2**):291–296, 1990.
[23] Zinaida Trybulec. Properties of subsets. *Formalized Mathematics*, 1(**1**):67–71, 1990.
[24] Edmund Woronowicz. Relations and their basic properties. *Formalized Mathematics*, 1(**1**):73–83, 1990.
[25] Bo Zhang and Yatsuka Nakamura. The definition of finite sequences and matrices of probability, and addition of matrices of real elements. *Formalized Mathematics*, 14(**3**):101–108, 2006.

Received July 9, 2007