Entropy **2005**, *7[1]*, 15-37

Entropy
ISSN 1099-4300
www.mdpi.org/entropy/

Full paper

The entropy of a mixture of probability distributions

Alexis De Vos
Imec v.z.w., Vakgroep voor elektronika en informatiesystemen, Universiteit Gent, Sint Pietersnieuwstraat 41, B - 9000 Gent, Belgium
E-mail: alex@elis.UGent.be

Received: 13 September 2004 / Accepted: 20 January 2005 / Published: 20 January 2005

Abstract: If a message can have n different values and all values are equally probable, then the entropy of the message is $\log(n)$. In the present paper, we investigate the expectation value of the entropy, for arbitrary probability distribution. For that purpose, we apply mixed probability distributions. The mixing distribution is represented by a point on an infinite dimensional hypersphere in Hilbert space. During an 'arbitrary' calculation, this mixing distribution has the tendency to become uniform over a flat probability space of ever decreasing dimensionality. Once such smeared-out mixing distribution is established, subsequent computing steps introduce an entropy loss expected to equal $\frac{1}{m+1} + \frac{1}{m+2} + ... + \frac{1}{n}$, where n is the number of possible inputs and m the number of possible outcomes of the computation.

Keywords: probability distribution; mixture distribution; Bhattacharyya space; Hilbert space.
PACS codes: 02.50

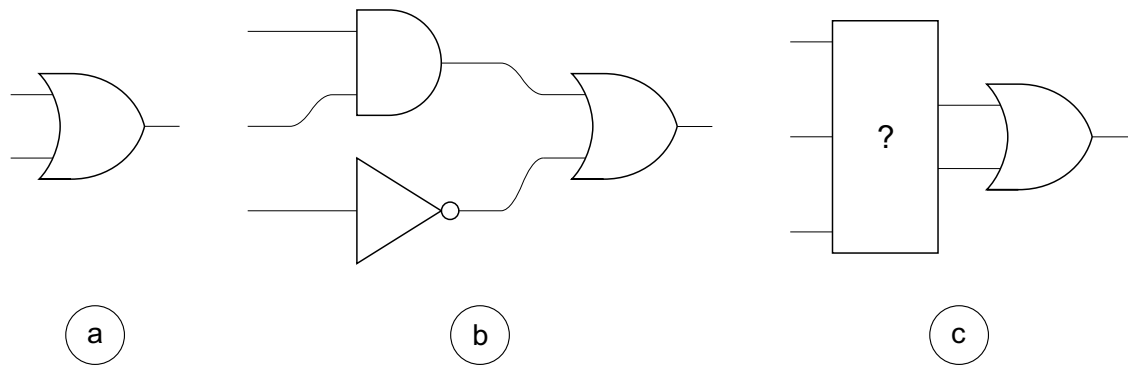

Figure 1: A logic `OR` gate: (a) standing alone, (b) after an `AND` and a `NOT` gate, (c) after an undetermined logic unit.

1 Introduction: distributions

Figure 1a shows an `OR` logic gate. Table 1a shows the truth table of the gate. The input can have four different values: either 00, or 01, or 10, or 11. We denote by p_i the probability that the input equals the i th of these four possibilities. If all four inputs are equally probable, i.e. if $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$, then the input will be called equiponderant and its entropy

$$S = -\sum_{i=1}^{4} p_i \log(p_i)$$

equals exactly $2\log(2)$. We say: the input entropy equals two bits. The output can have only two different values: either 0 or 1. We denote by p'_j the probability that the output equals the j th of these two alternatives. The output entropy

$$S' = -\sum_{j=1}^{2} p'_j \log(p'_j)$$

follows from the observation that $p'_1 = p_1$ and $p'_2 = p_2 + p_3 + p_4$ and thus $p'_1 = \frac{1}{4}$ and $p'_2 = \frac{3}{4}$. We find $S' = 2\log(2) - \frac{3}{4}\log(3)$ or 0.811 bit. We may conclude that an `OR` gate causes a loss of 1.189 bit. Because of the deterministic nature of a logic gate, the entropy loss $S - S'$ can also be interpreted as a conditional entropy. See Appendix A. The reader will easily verify that, whatever the logic gate, we always have $S' \leq S$. Indeed, it is a general phenomenon that entropy of a function of a random variable is less than or equal to the entropy of the random variable itself [1]. The conclusion that $S - S'$ equals 1.189 bits for an `OR` gate, is a direct consequence of our assumption that $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$, i.e. that the input message has been chosen equiponderantly. Our example shows that the output is not equiponderant, as both p'_1 and p'_2 are different from $\frac{1}{2}$. The assumption $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$ is not as selfevident as it looks like. Indeed, often the input of a logic gate is itself the outcome of a previous logic calculation. Figure 1b shows an example. If (and this again is a big 'if') the eight possible values (i.e. 000, 001, ..., 111) of the input of Figure 1b have equal probability (i.e. probability $\frac{1}{8}$), then the intermediate result (i.e. the input

Table 1: Two ways of writing down the truth table of the OR gate: (a) traditional way, (b) standard way.

in	out
0 0	0
0 1	1
1 0	1
1 1	1

(a)

in	out
1	1
2	2
3	2
4	2

(b)

of the OR gate) is not equiponderant: $p_1 = p_2 = \frac{3}{8}$ and $p_3 = p_4 = \frac{1}{8}$. As a result, the OR gate has $S = 3\log(2) - \frac{3}{4}\log(3)$ and $S' = 3\log(2) - \frac{3}{8}\log(3) - \frac{5}{8}\log(5)$, such that the entropy loss is not 1.189 bit, but rather 0.857 bit. We can thus conclude that the entropy loss introduced by a logic operation, depends on the 'history' of the incoming bits. According to the probability distribution of the incoming message, the entropy loss $S - S'$ can have different values. The reader can easily verify that, for the example of the OR operation, it can have any value between 0 and log(3), i.e. between 0 and 1.586 bit. In order to have an objective measure of the entropy loss caused by a logic gate, we should average over all possible input probabilities. We thus need a probability distribution of the probability distribution (p_1, p_2, p_3, p_4). Such 'distribution of distributions' is called a mixing distribution [2, 3].

In the general case, we assume a message that can only have n different forms. The number n is positive, integer and finite; therefore we talk of a 'digital' message. Sometimes the number n is small. E.g. if the message is just either 'yes' or 'no', then $n = 2$. In the above example of the incoming message of the OR gate, we have $n = 4$. Often the number n is very large. E.g. a 'telegram' of only 100 characters from an alphabet of only 27 characters (the 26 Roman upper-case letters plus the blank) can take $n = 27^{100} \approx 1.4 \times 10^{143}$ different 'values'. The probability of each possible message is denoted $p_1, p_2, ..., p_{n-1}$, and p_n. For an equiponderantly generated message, each message is equally likely and thus $p_1 = p_2 = ... = p_n = \frac{1}{n}$. The entropy

$$S = -\sum_{i=1}^{n} p_i \log(p_i)$$

of an equiponderant message is therefore $\log(n)$. However, most messages are not equiponderant: some possible messages are more likely than others. For arbitrary p_is (with, of course, the

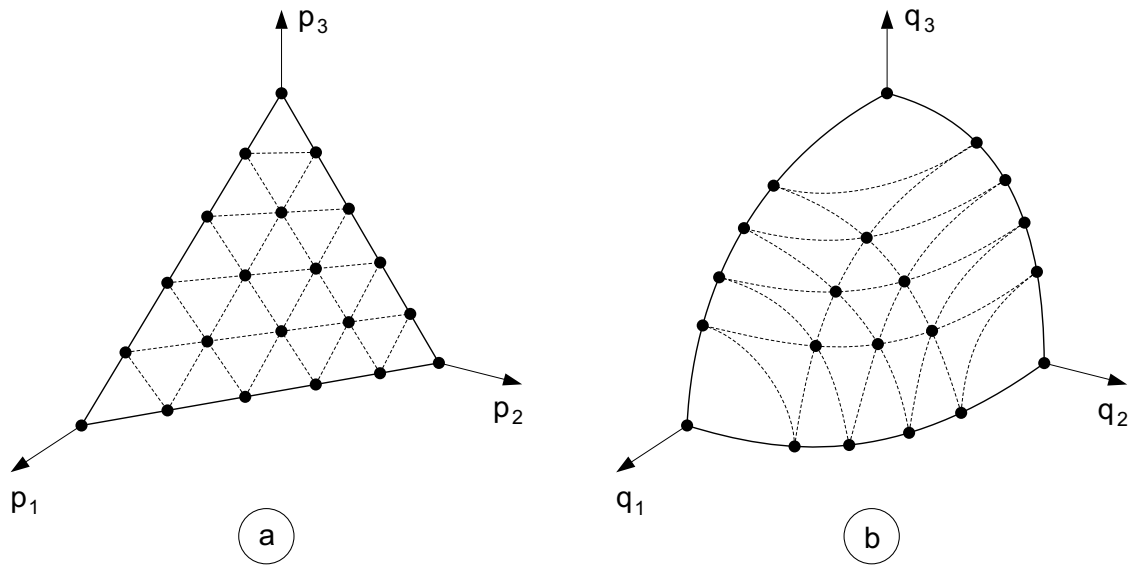

Figure 2: Probability space: (a) as a hypertriangle and (b) as a hyperoctant.

restriction $\sum_{i=1}^{n} p_i = 1$), one can easily demonstrate that

$$0 \leq S \leq \log(n) \ .$$

The lower bound $S = 0$ applies to each of the n deterministic messages, where $(p_1, p_2, ..., p_n)$ equals either (1, 0, ..., 0) or (0, 1, ..., 0) or ... or (0, 0, ..., 1).
Figure 2a shows the space subtended by the coordinates p_1, p_2, ..., and p_n. For sake of 'drawability', the figure represents the case $n = 3$. Our space is the n-dimensional triangle, defined by the n-dimensional plane $\sum_{i=1}^{n} p_i = 1$ together with the n conditions $0 \leq p_i \leq 1$. Such figure sometimes is referred to as the n-dimensional simplex. One point P within the hypertriangle represents one probability distribution. The corners of the hypertriangle represent the deterministic messages, whereas the centre point represents the equiponderant message.
If the input of Figure 1b has probabilities $(p_1, p_2, .., p_8) = (\frac{1}{8}, \frac{1}{8}, ..., \frac{1}{8})$, the intermediate result, i.e. the input of the `OR` gate is represented by the point $P = (\frac{3}{8}, \frac{3}{8}, \frac{1}{8}, \frac{1}{8})$. Let us however assume that we do not know the circuit preceding the `OR` gate. I.e. we only know that the logic unit preceding the `OR` gate in Figure 1c is a logic gate with three binary inputs and two binary outputs. Then the input can have $2^3 = 8$ possible values and the output can have $2^2 = 4$ possible values. Now, there exist 4^8 different truth tables with 8 possible inputs and 4 possible outputs. Thus the logic `?` gate is just one of the $4^8 = 65{,}536$ possible logic gates. If the input of Figure 1c has probability distribution $(p_1, p_2, ..., p_8) = (\frac{1}{8}, \frac{1}{8}, ..., \frac{1}{8})$, the input of the `OR` gate can be any of the different points $(\frac{a_1}{8}, \frac{a_2}{8}, \frac{a_3}{8}, \frac{a_4}{8})$, where (a_1, a_2, a_3, a_4) is one of the 165 ordered partitions of 8 into four integers: $8 = a_1 + a_2 + a_3 + a_4$. Note that these 165 points are points on the intersection of the hyperplane $p_1 + p_2 + p_3 + p_4 = 1$ with the hypersquare lattice with lattice constant $\frac{1}{8}$. For sake of 'drawability', Figure 2a shows another example: the points $(\frac{a_1}{5}, \frac{a_2}{5}, \frac{a_3}{5})$ in the plane $p_1 + p_2 + p_3 = 1$.

If each of the 65,536 logic gates can occur with equal probability 1/65,536, we can calculate the probability that P is in each of its 165 possible positions. These probabilities of a particular probability distribution, are called the mixing probabilities or mixing proportions. The mixing proportions vary from 1/65,536 in the corners of the hypertriangle (e.g. the point $(0,0,0,1)$) to 2,520/65,536 in the centre point $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$. They constitute the 165 weights of a finite mixture of 165 distributions. In the limit where the input of the logic ? gate of Figure 1c has not eight but l possible values and where $l \to \infty$ (i.e. in the limit of 'long history'), we obtain a smooth mixing distribution $\sigma(p_1, p_2, p_3, p_4)$ over the hypertriangle. This results in a so-called infinite mixture.
Figure 2b shows the space (p_1, p_2, p_3) after the non-linear coordinate transformation $q_i = \sqrt{p_i}$. Now the probability distributions are represented by points on the first hyperoctant of the unit hypersphere $\sum_{i=1}^{n} q_i^2 = 1$. Bhattacharyya's hyperoctant forms the statistical space. The use of the square root of the probabilities is justified, because it gives the 'natural' geometric structure for comparing probability distributions [4, 5, 6, 7, 8].
For an arbitrary probability distribution p with a set of k parameters, the statistical distance between two distributions, i.e. one with parameter set $(t_1, t_2, ..., t_k)$ and the other with set $(t_1^*, t_2^*, ..., t_k^*)$, is defined by the length $\int \sqrt{ds^2}$ of the shortest arc connecting the two points in the parameter space [5]. Here ds^2 is the expectation value of

$$\sum_i \sum_j \frac{1}{p} \frac{\partial p}{\partial t_i} \frac{1}{p} \frac{\partial p}{\partial t_j} \, dt_i dt_j$$

and therefore expresses how strongly p is affected by differences dt_i. In our case, the distribution has $n-1$ parameters, e.g. p_1, p_2, ..., p_{n-1}. After a few calculations, one finds [5, 8] that the element ds^2 is the expectation value of

$$\sum_i \frac{1}{p_i} \, (dp_i)^2 \ .$$

The Bhattacharyya choice $q_i = \sqrt{p_i}$ therefore guarantees that the geometric (Euclidean) distance between two points is equal to the statistical distance between the two statistics [8]. Indeed, $(dq_i)^2$ automatically equals $\frac{1}{p_i} \, (dp_i)^2$, with all due deference to a constant factor of $\frac{1}{4}$.
Two different statistics $(p_1, p_2, ..., p_n)$ and $(p_1^*, p_2^*, ..., p_n^*)$ are represented by two different points on the hypersphere: $Q = (q_1, q_2, ..., q_n)$ and $Q^* = (q_1^*, q_2^*, ..., q_n^*)$. The statistical distance between the two probability distributions equals the length of the geodesic line between the two points Q and Q^*. For this simple metric we have moreover that the geodesic is a large hypercircle and this distance equals the angle subtending the arc:

$$\arccos(q_1 q_1^* + q_2 q_2^* + ... + q_n q_n^*) \ .$$

Equally spaced pairs of points therefore are said to represent equally 'distinguishable' statistics [6] or equally 'dissimilar' statistics [7].
In **Section 2**, we will introduce data with arbitrary mixing distribution $\sigma(p_1, p_2, ..., p_n)$ and then will focuse on two special cases:

- the mixing distribution uniform in the $(p_1, p_2, ..., p_n)$ space and

- the mixing distribution uniform in Bhattacharyya's $(\sqrt{p_1}, \sqrt{p_2}, ..., \sqrt{p_n})$ space.

In both cases, the expectation value of the entropy is calculated. In **Section 3**, we will introduce data processing, i.e. computation. During such manipulation, an input mixing distribution $\sigma(p_1, p_2, ..., p_n)$ is transformed into an output mixing distribution $\sigma'(p'_1, p'_2, ..., p'_m)$, with $m \leq n$. Accordingly, the expectation value of the entropy changes from the value S to a new value S'. The entropy loss $S - S'$ is calculated. The computational process is regarded as a smooth transformation of the mixing distribution, from its initial shape σ to its final shape σ'. Therefore, the process is not a path followed by a single point in the finite-dimensional probability space, be it either the $(p_1, p_2, ..., p_n)$ or the $(q_1, q_2, ..., q_n)$ space, but a path followed by a single point in the mixing space, which is an infinite-dimensional Hilbert space. In **Section 4**, some mathematical properties of this space will be unfolded.

2 Mixture distributions

We now consider all possible statistics of order n. We assume they are either distributed over the hypertriangle of dimension n or distributed over the hyperoctant. Thus, in the former case we assume distributions $\sigma(p_1, p_2, ..., p_n)$; in the latter case we consider distributions $\mu(q_1, q_2, ..., q_n)$. Both quantities are, of course, correlated. Indeed, we have

$$\sigma \; d\omega_n = \mu \; d\Omega_n \; ,$$

where $d\omega_n$ and $d\Omega_n$ are the infinitesimal surface areas. From

$$d\omega_n = \sqrt{n} \; dp_1 dp_2 ... dp_{n-1}$$

and

$$d\Omega_n = \frac{1}{q_n} \; dq_1 dq_2 ... dq_{n-1} \; ,$$

one can easily deduce the following relationship between $d\omega_n$ and $d\Omega_n$:

$$\frac{d\omega_n}{d\Omega_n} = 2^{n-1}\sqrt{n} \; q_1 q_2 ... q_n \; .$$

Therefore, we get

$$\mu = 2^{n-1}\sqrt{n} \; q_1 q_2 ... q_n \; \sigma \; . \tag{1}$$

Such mixing distributions (either σ or μ) allow us to calculate the expectation value of any property of a probability distribution. We e.g. can calculate the expectation value of the entropy of a probability distribution.
The expectation value S_n of the entropy S is given by

$$S_n = \int_1 \int_2 ... \int_{n-1} \; \mu S d\Omega_n \; ,$$

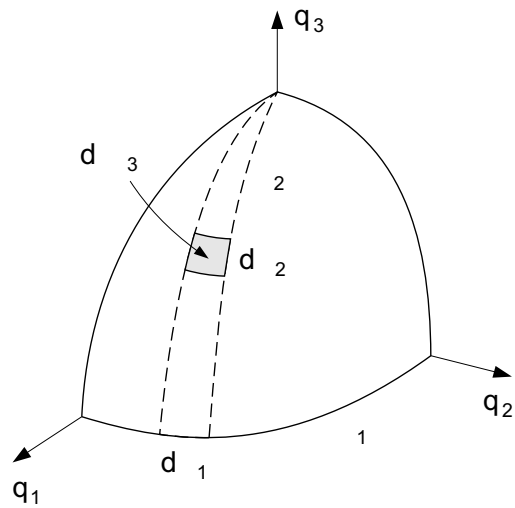

Figure 3: Probability space as a 3-dimensional octant.

where $\int_i$ is a short-hand notation for $\int_{\vartheta_i=0}^{\vartheta_i=\pi/2}$ and the function $\mu(\vartheta_1, \vartheta_2, ..., \vartheta_{n-1})$ is the assumed mixing function. The polar coordinates $(r, \vartheta_1, \vartheta_2, ..., \vartheta_{n-1})$ are given by $r = 1$ and

$$\begin{aligned}
q_n &= \cos\vartheta_{n-1} \\
q_{n-1} &= \sin\vartheta_{n-1}\cos\vartheta_{n-2} \\
q_{n-2} &= \sin\vartheta_{n-1}\sin\vartheta_{n-2}\cos\vartheta_{n-3} \\
... & \quad ... \\
q_3 &= \sin\vartheta_{n-1}\sin\vartheta_{n-2}\sin\vartheta_{n-3}...\cos\vartheta_2 \\
q_2 &= \sin\vartheta_{n-1}\sin\vartheta_{n-2}\sin\vartheta_{n-3}...\sin\vartheta_2\cos\vartheta_1 \\
q_1 &= \sin\vartheta_{n-1}\sin\vartheta_{n-2}\sin\vartheta_{n-3}...\sin\vartheta_2\sin\vartheta_1 \ .
\end{aligned}$$

The infinitesimal surface area $d\Omega_n$ of the n-dimensional unit sphere is given by

$$d\Omega_n = \sin\vartheta_2 \sin^2\vartheta_3 \sin^3\vartheta_4 ... \sin^{n-2}\vartheta_{n-1} d\vartheta_1 d\vartheta_2 d\vartheta_3 d\vartheta_4 ... d\vartheta_{n-1} \ .$$

See e.g. References [9] and [10]. Figure 3 shows the 3-dimensional case, where we recover the conventional spherical coordinates: $q_3 = \cos\vartheta_2$, $q_2 = \sin\vartheta_2\cos\vartheta_1$, $q_1 = \sin\vartheta_2\sin\vartheta_1$, and $d\Omega_3 = \sin\vartheta_2 d\vartheta_1 d\vartheta_2$. Note that for explicite calculations, we prefer the spherical probability space, as its $n-1$ integration intervals have simple (constant) values.
We now consider three special cases:

(a) the equiponderant message ($p_1 = p_2 = ... = p_n = \frac{1}{n}$) has a μ function that is a delta peak located in the 'middle' of the octant:

$$\mu = \delta(\, \vartheta_1 - \arccos\frac{1}{\sqrt{2}}\,,\ \vartheta_2 - \arccos\frac{1}{\sqrt{3}}\,,\ ...,\ \vartheta_{n-1} - \arccos\frac{1}{\sqrt{n}}\,)\ ;$$

(b) an arbitrary message which has a σ function that is uniform over the flat probability triangle:

$$\begin{aligned} \sigma &= \frac{1}{\omega_n} \\ &= \frac{(n-1)!}{\sqrt{n}} \ , \end{aligned}$$

where ω_n is the surface area of the hypertriangle in n-dimensional space (its value being derived in Appendix B), and thus (after (1)):

$$\mu = 2^{n-1}(n-1)! \ q_1 q_2 ... q_n \ ; \tag{2}$$

(c) an arbitrary message which has a μ function that is uniform over the spherical probability octant:

$$\mu = \frac{1}{\Omega_n} \ ,$$

where Ω_n is the surface area of the n-dimensional hyperoctant (its value being derived in Reference [11]).

We remark that equation (2) reveals the fact that a uniform σ results in a μ with a maximum in the centre of Bhattacharyya's hypersurface, in accordance with Figure 2, where the points, which are uniformly distributed on Figure 2a, are clustered around the centre of Figure 2b.
In the first case, we, of course, have $S_n = \log(n)$. See Table 2a. In the second case, we have

$$S_n = 2^{n-1}(n-1)! \int_1 \int_2 ... \int_{n-1} q_1 q_2 ... q_n \ S \ d\Omega_n \ . \tag{3}$$

Evaluating this quantity involves the calculation of integrals of the form

$$-\int_1 \int_2 ... \int_{n-1} q_1 q_2 ... q_n \ p_i \log(p_i) \ d\Omega_n \ .$$

After some calculations (Appendix C), one finds:

$$S_n = \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n} \ .$$

Table 2b gives the values for n up to 8. Figure 4b displays the results for n up to 32. For the important case of large n, one finds easily that

$$\begin{aligned} \lim_{n \to \infty} S_n &= \log(n) - (1 - \gamma) \\ &= \log(n) - 0.4228... \ , \end{aligned}$$

where γ is Euler's constant (also known as the Euler–Mascheroni constant).
In the third case, we have

$$S_n = \frac{1}{\Omega_n} \int_1 \int_2 ... \int_{n-1} S \ d\Omega_n \ .$$

Evaluating this quantity involves the calculation of integrals of the form

$$-\int_1 \int_2 ... \int_{n-1} p_i \log(p_i) \ d\Omega_n \ .$$

After laborious calculations [11], one finds:

Table 2: The expectation value S_n of the entropy of an n-valued digital message, for three different mixing distributions: (a) centered, (b) uniformly distributed over flat space, (c) uniformly distributed over spherical space.

n	(a)	(b)	(c)
1	$\log(1) = 0.000$	$0 = 0.000$	$0 = 0.000$
2	$\log(2) = 0.693$	$\frac{1}{2} = 0.500$	$2\log(2) - 1 = 0.386$
3	$\log(3) = 1.099$	$\frac{5}{6} = 0.833$	$\frac{2}{3} = 0.667$
4	$\log(4) = 1.386$	$\frac{13}{12} = 1.083$	$2\log(2) - \frac{1}{2} = 0.886$
5	$\log(5) = 1.609$	$\frac{77}{60} = 1.283$	$\frac{16}{15} = 1.067$
6	$\log(6) = 1.792$	$\frac{29}{20} = 1.450$	$2\log(2) - \frac{1}{6} = 1.220$
7	$\log(7) = 1.946$	$\frac{223}{140} = 1.593$	$\frac{142}{105} = 1.352$
8	$\log(8) = 2.079$	$\frac{481}{280} = 1.718$	$2\log(2) + \frac{1}{12} = 1.470$
...	...	...	...
$n \gg 1$	$\log(n)$	$\log(n) - 0.423$	$\log(n) - 0.730$

- if n is even:

$$S_n = 2\log(2) - 2 + 2\,(\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + ... + \frac{1}{n})\ ;$$

- if n is odd:

$$S_n = -2 + 2\,(1 + \frac{1}{3} + \frac{1}{5} + ... + \frac{1}{n})\ .$$

Table 2c gives the values for n up to 8. Figure 4c displays the results for n up to 32. For the important case of large n, one finds (both for n even and for n odd) that

$$\begin{aligned} \lim_{n\to\infty} S_n &= \log(n) - [2 - \log(2) - \gamma] \\ &= \log(n) - 0.7296...\ . \end{aligned}$$

3 An application: computations

Deterministic data processing can be regarded as a truth table. All possible 'messages in' are listed and the table gives the corresponding 'messages out'. In the truth table there are n rows with the n different input messages. As not necessarily all output messages are different, we have only m possible outgoing messages: $1 \leq m \leq n$. Table 1 gives the simple example of the `OR` gate. There are $n = 4$ possible input messages and only $m = 2$ possible output messages. Whereas

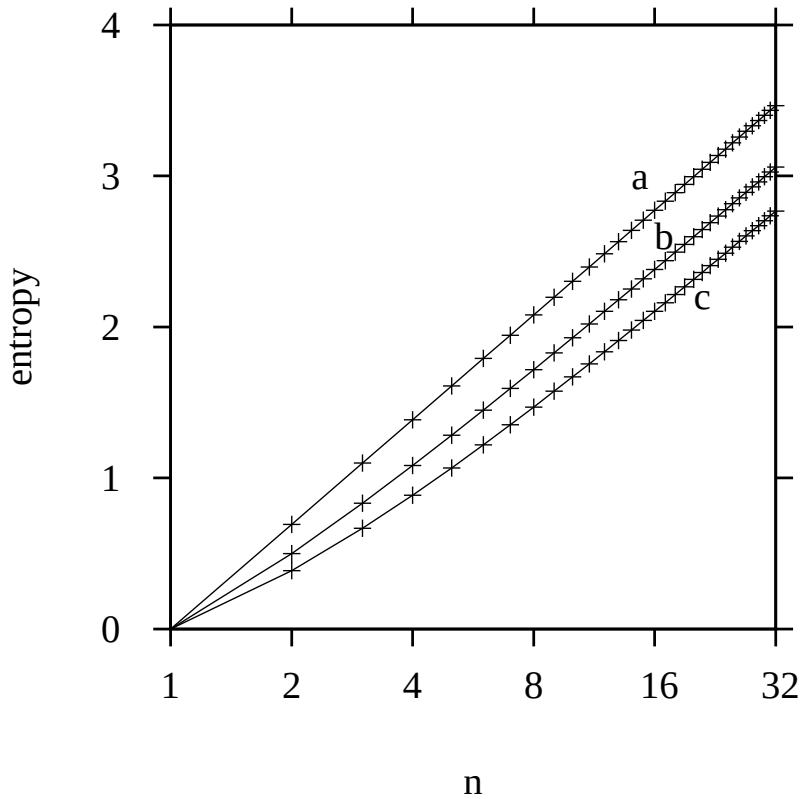

Figure 4: The expectation values S_n of the entropy of an n-valued digital message, for three different mixing distributions: (a) centered, (b) uniformly distributed over flat space, (c) uniformly distributed over spherical space.

Table 1a shows the traditional display of the truth table, Table 1b shows a standard form. Here, the left column consists of the ordered input message numbers $\{1, 2, 3, ..., n\}$, whereas the output column shows the corresponding output message numbers. Note that the right-hand side column can have equal numbers and that the highest number is m. Note that the standard form of thruth table is also applicable in case n or m or both are no integer power of 2.
We can generate arbitrary mixings $\mu(p_1, p_2, ..., p_n)$ and calculate the expectation value of both the input entropy $S = -\sum_{i=1}^{n} p_i \log(p_i)$ and the output entropy $S' = -\sum_{j=1}^{m} p'_j \log(p'_j)$.

3.1 Reversible computing

If $m = n$, then there is a one-to-one mapping from the input point $P = (p_1, p_2, ..., p_n)$ to the output point $P' = (p'_1, p'_2, ..., p'_n)$, or equivalently from the input point $Q = (q_1, q_2, ..., q_n)$ to the output point $Q' = (q'_1, q'_2, ..., q'_n)$. This case is nothing else but reversible computing [12]. Each of the p'_i equals some p_j (with j either equal to i or not). This has two consequences:

- the entropy S' of P' equals the entropy S of P;

- the position of P' is obtained from the position of P by a simple symmetry operation (either a rotation or a mirroring).

However, it would be wrong to suggest that the computation is a walk from point P to point P', because along such path the entropy is not constant (in spite of equal values at the two ends of the trajectory). Instead, a computation is a smooth change from mixing function σ (or μ) to mixing function σ' (or μ'). After completion of the shape change, it looks as if the mixing distribution density is merely rotated and/or mirrored. Therefore uniformity of the mixing is conserved.

3.2 Irreversible computing

If $m < n$, we have a computation which is logically irreversible. Then, more than one point of the n-dimensional hyperoctant is mapped to a single point of the m-dimensional hyperoctant. In fact, a whole subspace of the n-dimensional sphere is mapped to a single point on the m-dimensional sphere. Such subspaces obey equations such as $p_1 + p_2 + ... + p_a =$ constant, and therefore are hyperoctants of dimension a, however of non-unitary radius r. These 'small hyperspheres' (with surface area $\Omega_a r^{a-1}$) are multidimensional generalizations of the 'circles of latitude' on the 3-dimensional sphere. Thus complete 'small hyperoctants' of the n-dimensional sphere are mapped to a single point on the m-dimensional sphere.
If we assume that a_1 different input values lead to output value 1, that a_2 different input values lead to output value 2, ..., and that a_m different input values lead to output value m, then the logic truth table realizes a partition of the number n :

$$n = a_1 + a_2 + ... + a_m \ .$$

For the uniform input mixing measure

$$\mu(q_1, q_2, ..., q_n) = \frac{1}{\Omega_n} \ ,$$

the resulting output measure on the m-dimensional hyperoctant is non-uniform, but equal to

$$\mu'(q'_1, q'_2, ..., q'_m) = \frac{1}{\Omega_n} \ \Omega_{a_1}\Omega_{a_2}...\Omega_{a_m} \ (q'_1)^{a_1-1}(q'_2)^{a_2-1}...(q'_m)^{a_m-1} \tag{4}$$

and therefore a product of powers of sines and cosines of the polar coordinates ϑ'_1, ϑ'_2, ..., and ϑ'_{m-1}.
E.g. for the `OR` gate (where the partition $n = a_1 + a_2 + ... + a_m$ is $4 = 1 + 3$), the uniform mixing distribution

$$\mu(\vartheta_1, \vartheta_2, \vartheta_3) = \frac{1}{\Omega_4} = \frac{\pi^2}{8}$$

gives rise to the non-uniform mixing

$$\mu'(\vartheta'_1) = \frac{\Omega_1\Omega_3}{\Omega_4} \ (q'_2)^2 = \frac{4}{\pi} \ \cos^2\vartheta'_1 \ .$$

The expectation value $S' = \int_1 \left[-p'_1 \log(p'_1) - p'_2 \log(p'_2) \right] \mu' d\vartheta'_1$ equals 0.557 bit (See Table 3c).
For another gate with $n = 4$ and $m = 2$, the result can be different. Indeed, there exist 14 different logic gates with four input values and two output values:

Table 3: Mixing distribution and expected entropy, for three different input mixings: (a) centered, (b) uniformly distributed over flat space, (c) uniformly distributed over spherical space.

	(a)	(b)	(c)
before gate	$\mu = \delta(\vartheta_1 - \arccos\frac{1}{\sqrt{2}}, \vartheta_2 - \arccos\frac{1}{\sqrt{3}}, \vartheta_3 - \arccos\frac{1}{\sqrt{4}})$ $S = 2\log(2) = 1.386 = 2.000$ bit	$\mu = 48\, q_1 q_2 q_3 q_4$ $S = \frac{13}{12} = 1.083 = 1.562$ bit	$\mu = \frac{1}{\Omega_4}$ $S = 2\log(2) - \frac{1}{2} = 0.886 = 1.278$ bit
after `OR` gate	$\mu' = \delta(\vartheta'_1 - \frac{\pi}{6})$ $S' = 2\log(2) - \frac{3}{4}\log(3) = 0.562 = 0.811$ bit	$\mu' = 6\sin\vartheta'_1\cos^5\vartheta'_1$ $S' = \frac{11}{24} = 0.458 = 0.661$ bit	$\mu' = \frac{4}{\pi}\cos^2\vartheta'_1$ $S' = 2\log(2) - 1 = 0.386 = 0.557$ bit

- 4 of these (e.g. the `OR` gate) correspond with the partition $4 = 1 + 3$, whereas
- 6 of them (e.g. the `XOR` gate) realize the partition $4 = 2 + 2$, and finally
- 4 of them (e.g. the `AND` gate) realize the partition $4 = 3 + 1$.

In contrast to the mixing uniform over the hyperoctant, the mixing distribution, which is uniform over the 'flat probability space', i.e. over the hypertriangle ω_n, leads to conservation of uniformity. As mentioned above, the mixing distribution density of this uniform distribution is:

$$\mu(q_1, q_2, ..., q_n) = 2^{n-1}(n-1)!\; q_1 q_2 ... q_n \;.$$

Now, one can prove (See Appendix D), that, when we average over all logic paths with n different input messages and m different output messages, this distribution automatically leads to the new distribution

$$\mu'(q'_1, q'_2, ..., q'_m) = 2^{m-1}(m-1)!\; q'_1 q'_2 ... q'_m \;,$$

i.e. the mixing uniform over the hypertriangle in m-dimensional space. The reason why uniformity over the hypertriangle is conserved, whereas uniformity over the hyperoctant is not, is visualized geometrically in Figure 5. For sake of drawability, we have here $n = 3$ and $m = 2$. By the logic gates,

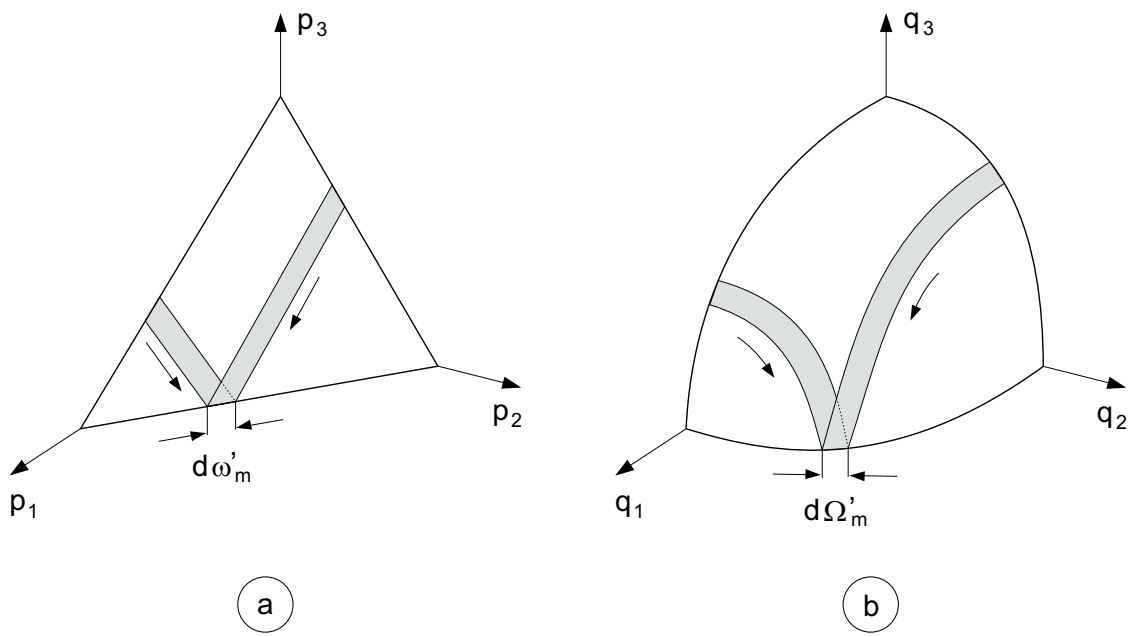

Figure 5: Projection from 3-dimensional probability space to 2-dimensional probability space: (a) in flat space and (b) in spherical space.

- the shaded areas on the triangle are projected towards the same infinitesimal space on the line segment, and
- the shaded areas on the sphere are projected towards the same infinitesimal space on the circle quadrant.

Now, the shaded surface area on the sphere is not constant, when we scan along the unit circle in the (q_1, q_2)-plane, whereas the shaded surface area on the triangle is constant, when we scan along the line segment in the (p_1, p_2)-plane.
In our example $n = 4$ and $m = 2$, the distribution

$$\mu(q_1, q_2, q_3, q_4) = 48\ q_1 q_2 q_3 q_4$$

is transformed into the distribution

$$\mu'(q'_1, q'_2) = 2\ q'_1 q'_2\ ,$$

provided we average over all possible logics to come from a 4-valued input to a 2-valued output. The expected entropy loss $S - S'$ then is $S_4 - S_2 = \frac{13}{12} - \frac{1}{2} = \frac{7}{12}$ or 0.842 bit. For arbitrary n and m, the expected loss is $S_n - S_m = \frac{1}{m+1} + \frac{1}{m+2} + ... + \frac{1}{n}$.
The entropy loss $S_n - S_m$ in a logically irreversible gate is fundamental in the framework of thermodynamics of computing. Indeed, it lays at the origin of unavoidable heat dissipation in irreversible computing. According to the Landauer principle [13, 14], an amount $(S_n - S_m)kT$ of heat will be generated during the computational step.

4 Polynomials

From the result (4), we can conclude that the monomials

$$\begin{aligned} & Q_{e_1,e_2,...,e_n}(\vartheta_1, \vartheta_2, ..., \vartheta_{n-1}) \\ = \; & q_1^{e_1} q_2^{e_2} ... q_n^{e_n} \\ = \; & \sin^{e_1} \vartheta_1 \cos^{e_2} \vartheta_1 \sin^{e_1+e_2} \vartheta_2 \cos^{e_3} \vartheta_2 ... \sin^{e_1+e_2+...+e_{n-1}} \vartheta_{n-1} \cos^{e_n} \vartheta_{n-1} \ , \end{aligned}$$

with e_i integer (zero or positive) exponents, naturally form a basis for the functions $\mu(\vartheta_1, \vartheta_2, ..., \vartheta_{n-1})$. However, these vectors are neither normal nor orthogonal. Indeed,

$$\int_1 \int_2 ... \int_{n-1} Q_{e_1,e_2,...,e_n} Q_{f_1,f_2,...,f_n} \ d\Omega_n = \delta_{e_1,f_1} \delta_{e_2,f_2} ... \delta_{e_n,f_n} \tag{5}$$

is not satisfied. In order to construct an orthonormal basis, we replace the functions $\sin^{k_1} \vartheta \cos^{k_2} \vartheta$ by the functions $\sin(k\vartheta)$ and $\cos(k\vartheta)$. The basis vectors are defined as the following products :

$$Q_{e_1,e_2,...,e_{n-1}}(\vartheta_1, \vartheta_2, ..., \vartheta_{n-1}) = M_{1,e_1}(\vartheta_1) M_{2,e_2}(\vartheta_2) ... M_{n-1,e_{n-1}}(\vartheta_{n-1}) \ ,$$

where the functions $M(\vartheta)$ are sums of $\sin(k\vartheta)$ and $\cos(k\vartheta)$ and satisfy the orthonormality conditions

$$\begin{aligned} \int_{\vartheta=0}^{\vartheta=\pi/2} M_{1,i}(\vartheta) \, M_{1,j}(\vartheta) \, d\vartheta &= \delta_{i,j} \\ \int_{\vartheta=0}^{\vartheta=\pi/2} M_{2,i}(\vartheta) \, M_{2,j}(\vartheta) \sin \vartheta \, d\vartheta &= \delta_{i,j} \\ ... \qquad & ... \\ \int_{\vartheta=0}^{\vartheta=\pi/2} M_{n-1,i}(\vartheta) \, M_{n-1,j}(\vartheta) \sin^{n-2} \vartheta \, d\vartheta &= \delta_{i,j} \ , \end{aligned}$$

which together guarantee that (5) is fulfilled. Appendix E gives some details.
We now have the following vector decomposition for the mixing distribution μ:

$$\mu(\vartheta_1, \vartheta_2, ..., \vartheta_{n-1}) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} ... \sum_{i_{n-1}=0}^{\infty} \mu_{i_1,i_2,...,i_{n-1}} \ Q_{i_1,i_2,...,i_{n-1}}(\vartheta_1, \vartheta_2, ..., \vartheta_{n-1}) \ .$$

The coefficients $\mu_{i_1,i_2,...,i_{n-1}}$ form coordinates in a Hilbert space subtended by the orthonormal basis mixing distributions $Q_{i_1,i_2,...,i_{n-1}}(\vartheta_1, \vartheta_2, ..., \vartheta_{n-1})$. Thus, whereas a particular distribution $(p_1, p_2, ..., p_n)$ is represented by a point in an n-dimensional space, a particular mixing distribution $\mu(p_1, p_2, ..., p_n)$ is represented by a point in a space with a (denumerable) infinitum of dimensions. However, because

$$\int_1 \int_2 ... \int_{n-1} \mu \ d\Omega_n = 1 \ ,$$

it may be even better to consider the function $\nu = \sqrt{\mu}$ instead of the function μ itself, as this new quantity is then automatically normalized. We can thus represent mixing distributions by points on the unit sphere of the Hilbert space subtended by the basis vectors. The quantities ν relate to the quantities μ like the numbers q relate to the numbers p.

Any computation, be it reversible or irreversible, can be considered as a walk from a point $(\mu_{0,0,...,0,0}, \mu_{0,0,...,0,1}, ...)$ to another point $(\mu'_{0,0,...,0,0}, \mu'_{0,0,...,0,1}, ...)$ on the Hilbert sphere. In the special case of reversible computing, the entropy remains constant all along the path. In the case of irreversible computing, the entropy monotonically decreases along the path.
The uniform distribution $\mu = 1/\Omega_n$ (and thus $\nu = 1/\sqrt{\Omega_n}$) is then represented by the first basis vector of the Hilbert space. Its coordinates are thus

$$(1, 0, 0, ...) \ ,$$

i.e. a single 1 followed by an infinite number of 0s. The representation of the uniform distribution $\sigma = 1/\omega_n$ or $\mu = 2^{n-1}(n-1)!\, q_1 q_2 ... q_n$ is dependent upon the parameter n. Suffice it here to give the coordinates in the simple case $n = 2$, thus $\mu = 2q_1q_2$ or $\nu = \sqrt{2}\, \sin^{\frac{1}{2}} \vartheta_1 \cos^{\frac{1}{2}} \vartheta_1$:

$$a \ (1, 0, -\frac{\sqrt{2}}{5}, 0, -\frac{\sqrt{2}}{15}, 0, -\frac{7\sqrt{2}}{195}, 0, ...) \ ,$$

where a is a constant:

$$a = \frac{2}{\pi} \ \left[\ \Gamma \left(\frac{3}{4} \right) \ \right]^2 = 0.9553... \ .$$

The statistical distance between two arbitrary mixing distributions ν and ν^* is the arc

$$\begin{aligned} & \arccos(\sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} ... \sum_{i_{n-1}=0}^{\infty} \nu_{i_1,i_2,...,i_{n-1}} \nu^*_{i_1,i_2,...,i_{n-1}}) \\ = \ & \arccos(\int_1 \int_2 ... \int_{n-1} \nu\nu^* d\Omega_n) \ . \end{aligned}$$

It should be stressed that the above Hilbert spaces are introduced for an arbitrary (finite) value of the integer n, and thus have nothing to do whatsoever with a limit case $n \to \infty$.

5 Conclusion

Bhattacharyya's statistical space and metric can be used to calculate properties of messages with an 'arbitrary information'. It consists of the first hyperoctant of the unit-radius hypersphere. A single point on this hyperoctant represents a probability distribution. We assume that an arbitrary message is represented by a distribution over this probability space. This 'distribution of probability distributions' allows us to calculate the expectation value of the information contained in an 'arbitrary message'. This content is smaller than for an equiponderant message of the same length. A distribution of distributions or mixing distribution is represented by a point on a unitary hypersphere in Hilbert space. Whereas uniformity of the mixing distribution in spherical probability space is not conserved, uniformity of mixing distribution over the corresponding flat probability space is conserved during arbitrary digital computations. Introduction of the Hilbert space has the following motivation: it allows a visualization of data processing (i.e. computing) as merely the movement of a point on its unit hypersphere. During such displacement, the mixing function changes shape accordingly.

Table 4: Joint and marginal probabilities of the OR gate.

i	j	$p_{i,j}$	p_i	p'_j
1	1	p_1	p_1	p_1
1	2	0	p_1	$p_2+p_3+p_4$
2	1	0	p_2	p_1
2	2	p_2	p_2	$p_2+p_3+p_4$
3	1	0	p_3	p_1
3	2	p_3	p_3	$p_2+p_3+p_4$
4	1	0	p_4	p_1
4	2	p_4	p_4	$p_2+p_3+p_4$

We make a final remark: in the above paper no complex numbers are introduced nor is introduced (in contrast to Wootters [6]) any physics. Hilbert spaces are deduced naturally from the mathematics. Nevertheless, many concepts and results strangely remind us of quantum mechanics. E.g. a point in the n-dimensional probability space resembles a state in classical physics, whereas a point in the Hilbert mixing space resembles a quantummechanical state. Further, averageing over all possible logic histories reminds us of a Feynman path integral. Such remarkable/puzzling resemblance between purely classical considerations and quantum mechanics are also observed by Diósi and Salamon [8].

Appendix A: Conditional entropy

Let $p_{i,j}$ be the probability that the input has the value i and the output has value j. The two marginal probabilities then satisfy $p_i = \sum_{j=1}^{m} p_{i,j}$ and $p'_j = \sum_{i=1}^{n} p_{i,j}$. The conditional entropies [1, 15] are defined as

$$\begin{aligned} S(\text{out}|\text{in}) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,j}\,\log(\frac{p_{i,j}}{p_i}) \\ S(\text{in}|\text{out}) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,j}\,\log(\frac{p_{i,j}}{p'_j})\ . \end{aligned}$$

Because of its deterministic character, a logic gate has probabilities $p_{i,j}$ which either equal p_i or equal zero. E.g. for the OR gate (Table 1b), we have the $p_{i,j}$ values as shown in Table 4.

As a result, we obtain automatically

$$\begin{aligned} S(\text{out}|\text{in}) &= -\sum_{i=1}^{n} p_i \log(1) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} S(\text{in}|\text{out}) &= -\sum_{i=1}^{n} p_i \log(p_i) + \sum_{j=1}^{m} p'_j \log(p'_j) \\ &= S - S' \ . \end{aligned}$$

Thus the entropy loss $S - S'$ is none else but the input entropy once the output is known. The entropy S' is both a measure for the information at the output and for the mutual information of the output with the input.
In case of a sequence of logic gates, we have $S(\text{out}|\text{in})$ again equal zero and $S(\text{in}|\text{out})$ equal to the sum of the subsequent entropy losses along the logic cascade. We basically have a sequence of Markov processes, however with special transition matrices T: each of the n rows consists of $m-1$ elements equal 0 and one element equal 1. In the example of the `OR` gate, we have

$$(T_{i,j}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \ .$$

Appendix B: Surface area of hypertriangle

The 'surface area' ω_n is constructed by stacking parallel slabs of ever decreasing size (Figure 6):

$$\omega_n = \int_0^1 \ \omega_{n-1} \, (1-p_n)^{n-2} \sqrt{\frac{n}{n-1}} \ dp_n \ ,$$

leading to the recursion formula

$$\omega_n = \frac{n^{1/2}}{(n-1)^{3/2}} \ \omega_{n-1} \ .$$

Together with $\omega_2 = \sqrt{2}$, this yields

$$\omega_n = \frac{\sqrt{n}}{(n-1)!} \ .$$

One can find the same result as follows:

$$\begin{aligned} \omega_n &= \int_1 \int_2 \ldots \int_{n-1} \ d\omega_n \\ &= \int_1 \int_2 \ldots \int_{n-1} \ 2^{n-1} \sqrt{n} \ q_1 q_2 \ldots q_n \ d\Omega_n \\ &= 2^{n-1} \sqrt{n} \int_1 \ \sin\vartheta_1 \cos\vartheta_1 \ d\vartheta_1 \int_2 \ \sin^3\vartheta_2 \cos\vartheta_2 \ d\vartheta_2 \ \ldots \int_{n-1} \ \sin^{2n-3}\vartheta_{n-1} \cos\vartheta_{n-1} \ d\vartheta_{n-1} \end{aligned}$$

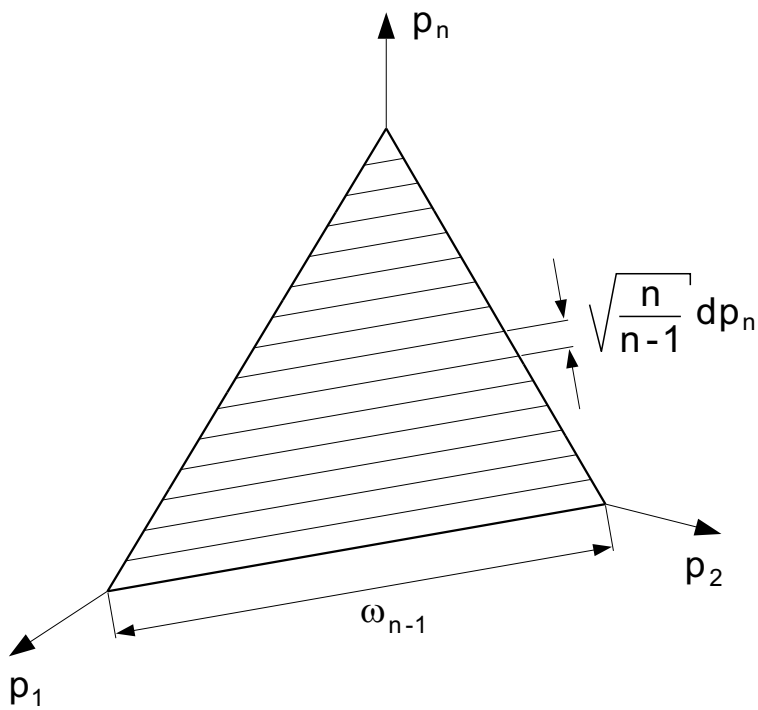

Figure 6: A hypertriangle as a stack of lower-dimensional hypertriangles.

$$
\begin{aligned}
&= 2^{n-1}\sqrt{n}\,\frac{1}{2}\,\frac{1}{4}\,\ldots\,\frac{1}{2n-2} \\
&= \frac{\sqrt{n}}{(n-1)!}\,.
\end{aligned}
$$

Appendix C: Calculation of an integral

Because of n-fold symmetry, all n integrals

$$\int_1\int_2\ldots\int_{n-1} q_1 q_2 \ldots q_n\; p_i \log(p_i) d\Omega_n$$

are equal, such that expression (3) leads to

$$
\begin{aligned}
S_n &= -2^{n-1}(n-1)!\;\; n\int_1\int_2\ldots\int_{n-1} q_1 q_2 \ldots q_n\; p_n \log(p_n) d\Omega_n \\
&= -2^{n-1} n! \int_1 \sin\vartheta_1 \cos\vartheta_1\; d\vartheta_1 \int_2 \sin^3\vartheta_2 \cos\vartheta_2\; d\vartheta_2\; \ldots \\
&\quad \int_{n-2} \sin^{2n-5}\vartheta_{n-2} \cos\vartheta_{n-2}\; d\vartheta_{n-2} \int_{n-1} \sin^{2n-3}\vartheta_{n-1} \cos\vartheta_{n-1} \cos^2\vartheta_{n-1} \log(\cos^2\vartheta_{n-1})\; d\vartheta_{n-1} \\
&= -2^{n-1} n!\;\frac{1}{2}\,\frac{1}{4}\ldots\frac{1}{2n-4} \int_{n-1} \sin^{2n-3}\vartheta_{n-1} \cos^3\vartheta_{n-1} \log(\cos^2\vartheta_{n-1})\; d\vartheta_{n-1} \\
&= -2(n-1)n\;\left(D_{2n-3} - D_{2n-1}\right)\,,
\end{aligned}
$$

where D_m stands for the integral

$$\begin{aligned} D_m &= \int_0^{\pi/2} \sin^m \vartheta \cos\vartheta \log(\cos^2\vartheta)\ d\vartheta \\ &= \int_0^1 x^m \log(1-x)\ dx + \int_0^1 x^m \log(1+x)\ dx\ . \end{aligned}$$

By recursive procedure, one can find that the former integral

$$\int_0^1 x^m \log(1-x)\ dx = -(1+\frac{1}{2}+\frac{1}{3}+...+\frac{1}{m}+\frac{1}{1+m})\ \frac{1}{m+1}\ .$$

Applying similar procedure to the latter integral, leads to distinct results according to the parity of m. For odd m, we find:

$$\int_0^1 x^m \log(1+x)\ dx = (1-\frac{1}{2}+\frac{1}{3}-...+\frac{1}{m}-\frac{1}{1+m})\ \frac{1}{m+1}$$

and thus

$$D_m = -(1+\frac{1}{2}+\frac{1}{3}+...+\frac{1}{(m+1)/2})\ \frac{1}{m+1}\ .$$

As a result, we get

$$S_n = \frac{1}{2}+\frac{1}{3}+...+\frac{1}{n}\ .$$

Appendix D: Conservation of uniformity

We first assume a particular truth table with $m = n-1$: see Table 5. The relation between output mixing distribution μ' and input mixing distribution μ is:

$$\mu'(q'_1, q'_2, ..., q'_{n-1})\ d\Omega'_{n-1} = \int_1 \mu(q_1, q_2, ..., q_n)\ d\Omega_n\ ,$$

with

$$\begin{aligned} q'_1 &= \sqrt{q_1^2+q_2^2} \\ q'_2 &= q_3 \\ q'_3 &= q_4 \\ ... &\quad ... \\ q'_{n-2} &= q_{n-1} \\ q'_{n-1} &= q_n \end{aligned}$$

and thus $\vartheta'_1 = \vartheta_2$, $\vartheta'_2 = \vartheta_3, ..., \vartheta'_{n-2} = \vartheta_{n-1}$. This yields:

$$\begin{aligned} \mu'\ d\Omega'_{n-1} &= \sin\vartheta_2 \sin^2\vartheta_3 ... \sin^{n-2}\vartheta_{n-1}\ d\vartheta_2 d\vartheta_3 ... d\vartheta_{n-1} \int_1 \mu\ d\vartheta_1 \\ &= \sin\vartheta'_1 \sin^2\vartheta'_2 ... \sin^{n-2}\vartheta'_{n-2}\ d\vartheta'_1 d\vartheta'_2 ... d\vartheta'_{n-2} \int_1 \mu\ d\vartheta_1 \\ &= q'_1 d\Omega'_{n-1} \int_1 \mu\ d\vartheta_1 \end{aligned}$$

Table 5: The truth table of the gate.

in	out
1	1
2	1
3	2
4	3
...	...
$n-1$	$n-2$
n	$n-1$

or

$$\mu' = q_1' \int_1 \mu \; d\vartheta_1 \; .$$

We now consider an n-dimensional mixing function of the particular form

$$\mu(q_1, q_2, ..., q_n) = c_{n,k} \, (q_1 q_2 ... q_n)^k \; ,$$

where the exponent k is not necessarily integer and where $c_{n,k}$ is the appropriate normalization constant. This yields

$$\begin{aligned}
\mu' &= q_1' \; c_{n,k} (q_2' ... q_{n-1}')^k \quad \int_1 (q_1 q_2)^k \; d\vartheta_1 \\
&= c_{n,k} \, q_1' (q_2' ... q_{n-1}')^k \quad (\sin\vartheta_1' \sin\vartheta_2' ... \sin\vartheta_{n-2}')^{2k} \int_1 \sin^k\vartheta_1 \cos^k\vartheta_1 \; d\vartheta_1 \\
&= c_{n,k} \frac{\sqrt{\pi}}{2^{k+1}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k+2}{2})} (\sin\vartheta_1' \sin\vartheta_2' ... \sin\vartheta_{n-2}')^{2k} \; q_1' (q_2' ... q_{n-1}')^k \\
&= c_{n,k} \frac{\sqrt{\pi}}{2^{k+1}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k+2}{2})} (q_1')^{2k} \; q_1' (q_2' ... q_{n-1}')^k \\
&= c_{n,k} \frac{\sqrt{\pi}}{2^{k+1}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k+2}{2})} (q_1')^{k+1} \; (q_1' q_2' ... q_{n-1}')^k \; .
\end{aligned}$$

We now consider all possible truth tables with $m = n-1$. If we average over all these, the factor $(q_1')^{k+1}$ is replaced by

$$\frac{1}{n-1} \; [\; (q_1')^{k+1} + (q_2')^{k+1} + ... + (q_{n-1}')^{k+1} \;] \; .$$

Only for $k = 1$, this expression is a constant. If $k = 1$ and only if $k = 1$, we get μ' proportional to $(q'_1 q'_2 ... q'_{n-1})^k$:

$$\mu' = c_{n,1} \frac{1}{2} \frac{1}{n-1} q'_1 q'_2 ... q'_{n-1} \ .$$

With $c_{n,1} = 2^{n-1}(n-1)!$ (from $\int_1 \int_2 ... \int_{n-1} \ \mu \ d\Omega_n = 1$ and in accordance with (2)), we finally conclude that

$$\mu = c_{n,1} \ q_1 q_2 ... q_n$$

leads to

$$\mu' = c_{n-1,1} \ q'_1 q'_2 ... q'_{n-1} \ .$$

If $m < n - 1$, we can decompose the logic gate into a cascade of subsequent logic gates, each satisfying the condition that the number of possible output messages is equal to or one less than the number of possible input messages. In each step the property of conservation of uniformity (over flat probability space) is fulfilled. Thus it is also fulfilled for the whole chain. This completes the proof of the theorem, for arbitrary n and arbitrary m, provided we average over all possible cascades leading from an n-valued input to an m-valued output.

Appendix E: Vector basis

The functions $M_{1,i}(\vartheta)$ have to satisfy the orthonormality conditions

$$\int_0^{\pi/2} M_{1,i}(\vartheta) M_{1,j}(\vartheta) \ d\vartheta = \delta_{i,j} \ .$$

We choose that all $M_{1,i}$ are extrapolated beyond the interval $0 \leq \vartheta_1 \leq \frac{\pi}{2}$ such that the two (hypermeridian) boundaries $\vartheta_1 = 0$ and $\vartheta_1 = \frac{\pi}{2}$ are hyperplanes of mirror symmetry. Then, the functions are simple Fourier functions:

$$\begin{aligned} M_{1,0} &= \frac{\sqrt{2}}{\sqrt{\pi}} \\ M_{1,i} &= \frac{2}{\sqrt{\pi}} \cos 2i\vartheta \qquad\qquad \text{for } i \geq 1 \ . \end{aligned}$$

The functions $M_{2,i}(\vartheta)$ have to satisfy the conditions

$$\int_0^{\pi/2} M_{2,i}(\vartheta) M_{2,j}(\vartheta) \sin\vartheta \ d\vartheta = \delta_{i,j} \ .$$

We choose that all $M_{2,i}$ are extrapolated beyond the interval $0 \leq \vartheta_2 \leq \frac{\pi}{2}$ such that the (hyperequator) boundary $\vartheta_2 = \frac{\pi}{2}$ is a hyperplane of mirror symmetry and the (hyperaxis) boundary $\vartheta_2 = 0$ is a hyperline of symmetry. We have:

$$\begin{aligned} M_{2,0} &= 1 \\ M_{2,1} &= \frac{\sqrt{5}}{4} \ (1 + 3\cos 2\vartheta) \\ M_{2,2} &= \frac{3}{64} \ (9 + 20\cos 2\vartheta + 35\cos 4\vartheta) \\ ... \quad & \quad ... \ . \end{aligned}$$

Apart from the classical normalization factor and an additional factor $\sqrt{2}$ (caused by the fact we have here a normalization interval $(0, \frac{\pi}{2})$ instead of $(0, \pi)$), these are the even Legendre polynomials $P_{2i}(\cos\vartheta)$.
We can proceed further to calculate any $M_{k,i}(\vartheta)$. Suffice it to mention the first two functions:

$$\begin{aligned} M_{k,0} &= \frac{1}{\sqrt{A_{k-1}}} \\ M_{k,1} &= \frac{1}{2}\sqrt{\frac{k+3}{2kA_{k-1}}}\ [\ (k-1)+(k+1)\cos 2\vartheta\ ]\ , \end{aligned}$$

where A_k is defined as

$$A_k = \int_0^{\pi/2} \sin^k\vartheta\ d\vartheta\ ,$$

its value being given e.g. in Reference [11] and being equal to the ratio $\Omega_{k+2}/\Omega_{k+1}$. Systematic application of the Gram–Schmidt orthonormalization procedure automatically results in the even-numbered Gegenbauer polynomials C_{2i}^{α} (i.e. the multidimensional generalizations [10] of the Legendre polynomials):

$$M_{k,i}(\vartheta) = \sqrt{\frac{4i+k-1}{k-1}\ \frac{(2i)!(k-2)!}{(2i+k-2)!}\ \frac{1}{A_{k-1}}}\ \ C_{2i}^{\frac{k-1}{2}}(\cos\vartheta)\ .$$

Note that the 'first' basis function is

$$\prod_{k=1}^{n-1} M_{k,0}(\vartheta_k) = \prod_{k=1}^{n-1} \frac{1}{\sqrt{A_{k-1}}} = \frac{1}{\sqrt{\prod_{k=0}^{n-2}\ A_k}} = \frac{1}{\sqrt{\Omega_n}}\ .$$

Acknowledgement

The author thanks the reviewers and editors of the journal 'Entropy' for their valuable suggestions, in particular for pointing his attention towards the literature on mixture distributions.

References

[1] Cover, T.; Thomas, J. *Elements of information theory*; Wiley, New York, 1991.

[2] Teicher, H. On the mixture of distributions. *The Annals of Mathematical Statistics* **1960**, *31*, 55-73.

[3] Everitt, B.; Hand, D. *Finite mixture distributions*; Chapman and Hall, London, 1981.

[4] Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society* **1943**, *35*, 99-109.

[5] Atkinson, C.; Mitchell, A. Rao's distance. *Sankhyā : The Indian Journal of Statistics* **1981**, *43*, 345-365.

[6] Wootters, W. Statistical distance and Hilbert space. *Physical Review D* **1981**, *23*, 357-362.

[7] Aherne, F.; Thacker, N.; Rockett, P. The Bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika* **1998**, *32*, 363-368.

[8] Diósi, L.; Salamon, P. From statistical distances to minimally dissipative processes. In *Thermodynamics of energy conversion and transport*; Sieniutycz, S.; De Vos, A., Eds.; Springer Verlag, New York, 2000; pp. 286-318.

[9] Kendall, M. *A course in the geometry of n dimensions*; Charles Griffin & co, London, 1961.

[10] Avery, J. *Hyperspherical harmonics – applications in quantum theory*; Kluwer Academic Publishers, Dordrecht, 1989.

[11] De Vos, A. The expectation value of the entropy of a digital message. *Open Systems & Information Dynamics* **2002**, *9*, 97-113.

[12] De Vos, A. Reversible computing. *Progress in Quantum Electronics* **1999**, *23*, 1-49.

[13] Landauer, R. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development* **1961**, *5*, 183-191.

[14] Bennett, C. The thermodynamics of computation – a review. *International Journal of Theoretical Physics* **1982**, *21*, 905-940.

[15] Shannon, C. A mathematical theory of communication. *Bell Systems Technical Journal* **1948**, *27*, 379-423.