Distributional solutions in information theory, I

by A. Kamiński (Katowice), P.L. Kannappan (Waterloo, Canada) and J. Mikusiński (Katowice)

Abstract. The measures directed divergence and inaccuracy are characterized with the help of a functional equation, which is solved by means of distributions.

1. Let $P = (p_1, \ldots, p_n)$ with $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$ be a finite discrete probability distribution. Let us consider another finite discrete probability distribution $Q = (q_1, \ldots, q_n)$ with $q_i \geq 0$, $\sum_{i=1}^{n} q_i = 1$ such that there is a 1-1 correspondence between the elements of $P$ and $Q$ given by their suffices. Then the measure of information directed divergence [10] or information gain [13] is given by

$$D_n (p_1, \ldots, p_n) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

and inaccuracy [9] is given by

$$I_n (p_1, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log q_i.$$

Remark 1. In (1) and (2) it is used that $q_i = 0$ (or 1) implies $p_i = 0$ (or 1) and $0 \cdot \log 0 / 0 = 0$ and the base of the logarithm is 2.

The quantities (1) and (2) have been characterized by using various systems of postulates ([3], [4], [6]–[8], [12]). In this paper we are characterizing (1) and (2) through a single functional equation, which is solved by the method of differentiation, consequently by the theory of distributions, along similar methods used in [2] to characterize Shannon's entropy.

Let $K^n (P || Q) (n \geq 2)$ be a system of functions defined on the sets $D_n$, where

$$D_n = \{(P, Q): p_i, q_i \geq 0, \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1\}$$
and satisfying the following axioms:

(a) $K^n(P|Q)$ is symmetric in $\left(\frac{p_i}{q_i}\right) (i = 1, 2, \ldots, n),$

(b) $K^n$ is continuous,

(c) $K^n\left(\frac{p_1, \ldots, p_n}{q_1, \ldots, q_n}\right) = K^{n-1}\left(\frac{p_1 + p_2, p_3, \ldots, p_n}{q_1 + q_2, q_3, \ldots, q_n}\right) +$

\[+ (p_1 + p_2) K^2\left(\frac{p_1}{q_1}, \frac{p_2}{q_2}\right),\]

whenever $p_1 + p_2, q_1 + q_2 > 0.$

**Theorem 1.** If the functions $K^n (n \geq 2)$ satisfy (a), (b), and (c), then

\[
K^n(P|Q) = A \sum_{i=1}^{n} p_i \log p_i + B \sum_{i=1}^{n} p_i \log q_i,
\]

where $A$ and $B$ are arbitrary constants.

**Remark 2.** It is easy to see (from the references cited) that (c) for $n = 3$ and $K^3$ symmetric lead to the functional equation

\[
F(x, y) + (1-x) F\left(\frac{u}{1-x}, \frac{v}{1-y}\right)
= F(u, v) + (1-u) F\left(\frac{x}{1-u}, \frac{y}{1-v}\right),
\]

where

\[
F(x, y) = K^2\left(\frac{x}{y}, \frac{1-x}{1-y}\right),
\]

and the solution of (4) by [5] with the use of (a) results in (3). Here we are using theory of distributions to achieve this result (1).

Analogously as in [2], Theorem 2 follows from a theorem on a single function $K$, which we shall formulate below.

By using (a) for $n = 3$, we get

\[
K^3\left(\frac{p, 1-p, 0}{q, 1-q, 0}\right) = K^2\left(\frac{1, 0}{1, 0}\right) + K^2\left(\frac{p, 1-p}{q, 1-q}\right),
\]

\[
K^3\left(\frac{p, 0, 1-p}{q, 0, 1-q}\right) = K^2\left(\frac{p, 1-p}{q, 1-q}\right) + pK^2\left(\frac{1, 0}{1, 0}\right) \quad \text{for } 0 < p, q < 1,
\]

which by the use of \((a_1)\) for \(n = 3\) results in \(K^2\begin{pmatrix} 1, 0 \\ 1, 0 \end{pmatrix} = 0\). Thus

\[
K^3\begin{pmatrix} p_1, p_2, 0 \\ q_1, q_2, 0 \end{pmatrix} = K^2\begin{pmatrix} p_1, p_2 \\ q_1, q_2 \end{pmatrix}.
\]

Again using \((a_3)\) for \(n = 3\) and \((5)\), we obtain

\[
K^3\begin{pmatrix} p_1, p_2, p_3 \\ q_1, q_2, q_3 \end{pmatrix} = K^3\begin{pmatrix} p_1 + p_2, 0, p_2 \\ q_1 + q_2, 0, q_3 \end{pmatrix} + (p_1 + p_2)K^3\begin{pmatrix} p_1, p_2, 0 \\ p_1 + p_2, q_1, q_2 \end{pmatrix},
\]

which by defining

\[
K\begin{pmatrix} x, y, z \\ u, v, w \end{pmatrix} = (x + y + z)K^3\begin{pmatrix} x \\ y \\ z \\ x + y + z \\ u \\ v \\ w \\ u + v + w \end{pmatrix}
\]

gives

\[
K\begin{pmatrix} p_1, p_2, p_3 \\ q_1, q_2, q_3 \end{pmatrix} = K\begin{pmatrix} p_1 + p_2, 0, p_3 \\ q_1 + q_2, 0, q_3 \end{pmatrix} + K\begin{pmatrix} p_1, p_2, 0 \\ q_1, q_2, 0 \end{pmatrix}.
\]

Further, for arbitrary \(\lambda > 0\), we have from \((6)\),

\[
K\begin{pmatrix} \lambda x, \lambda y, \lambda z \\ u, v, w \end{pmatrix} = \lambda K\begin{pmatrix} x, y, z \\ u, v, w \end{pmatrix},
\]

and

\[
K\begin{pmatrix} x, y, z \\ \lambda u, \lambda v, \lambda w \end{pmatrix} = K\begin{pmatrix} x, y, z \\ u, v, w \end{pmatrix},
\]

that is, \(K\) is positively homogeneous of order 1 with respect to the variables \(x, y, z\) and of order 0 with respect to the variables \(u, v, w\).

Thus, in order to prove Theorem 1, it is enough to solve the functional equation \((7)\) under the additional conditions \((8)\) and \((9)\).

2. Let \(D\) be the following domain in \(D^3\): \(x_1, x_2, x_3, y_1, y_2, y_3 \geq 0, x_1y_1 \cdot x_2y_2 \cdot x_3y_3 + x_1y_2 + x_2y_3 + x_3y_1 > 0\) (that is, at least two of the pair of variables have positive elements).

It is clear that the interior \(D^3\) of \(D\) coincides with the interior of \(D_3\).

Let \(K\) be a function defined on \(D\) and be positively homogeneous in the sense that \((8)\) and \((9)\) hold. Then we prove the following theorem.
THEOREM 2. If a function $K\left(\frac{x}{u}, \frac{y}{v}, \frac{z}{w}\right)$ is continuous, pairwise symmetric and positively homogeneous in $D$ and satisfies, in $D^0$, the functional equation (7), then

\[
K\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{x_3}{y_3}\right) = a \left[ \left( \sum_{i=1}^{3} \frac{x_i}{y_i} \right) \log \left( \sum_{i=1}^{3} \frac{x_i}{y_i} \right) - \sum_{i=1}^{3} \frac{x_i \log x_i}{y_i} \right] + \\
+ \beta \left[ \left( \sum_{i=1}^{3} \frac{x_i}{y_i} \right) \log \left( \sum_{i=1}^{3} \frac{y_i}{x_i} \right) - \sum_{i=1}^{3} \frac{x_i \log y_i}{x_i} \right],
\]

holds in $D$, where $a$ and $\beta$ are arbitrary constants.

In order to use the method of distributional differentiation we prove the more general theorem.

THEOREM 3. If a distribution $K\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{x_3}{y_3}\right)$ defined in an open set $S$ containing $D$, is pairwise symmetric and positively homogeneous in $D^0$ and satisfies equation (7) in $D^0$, then it is of the form (10) in $D^0$.

It is now evident how Theorem 2 follows from Theorem 3, namely, each function which is continuous in $D$ can be continuously extended to the domain $S$ and then it can be considered as a distribution. On the other hand each function which is continuous in $D$ and has the form (10) in $K^0$ has this form (10) also in $D$.

The proof of Theorem 3 is given in section 5.

3. The symbols

\[
K\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{0}{0}\right) \quad \text{and} \quad K\left(\frac{x_1 + x_2}{y_1 + y_2}, \frac{0}{0}, \frac{x_3}{y_3}\right)
\]

in (7) are meant as in [2], i.e., in the sense of generalized operations on distributions (see also [1] and [11]). In solving (7), we shall use differentiation. We shall denote the partial derivatives of $K$ with respect to $x_i$ and $y_i$ by $K_i^{(i)}$ and $K_i^{(j)}$, respectively.

The symbols

\[
K\left(\frac{x_1 + x_2}{y_1 + y_2}, \frac{0}{0}, \frac{x_3}{y_3}\right), \quad K^{(i)}\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{0}{0}\right), \quad K^{(j)}\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{0}{0}\right)
\]

are defined uniquely, because they do not depend on a succession of performing operations (see [5]).

In [2], the following lemma is proved.

LEMMA 1. If $f(x, y, z) = g(x + y, z) = h(x, y + z)$, where $f$ is a distribution of three variables and $g$ and $h$ are distributions of two variables, then there is a distribution $I$ of one variable such that $f(x, y, z) = I(x + y + z)$.

The proof of Lemma 1 is based on a simple substitution. In [2], Lemma 1 is formulated for $x, y, z > 0$, but it is true also in the case where $x, y \in R^2, x, y > 0$. If in particular $q = 3$, we obtain the following lemma.
Lemma 2. If \( f \begin{pmatrix} x_1, x_2, x_3 \\ y_1, y_2, y_3 \end{pmatrix} = g \begin{pmatrix} x_1, x_2 + x_3 \\ y_1, y_2 + y_3 \end{pmatrix} = h \begin{pmatrix} x_1 + x_2, x_3 \\ y_1 + y_2, y_3 \end{pmatrix} \) \((x_1, x_2, x_3, y_1, y_2, y_3 > 0)\), where \( f \) is a distribution of six variables and \( g, h \) are distributions of four variables, then there is a distribution \( I \) of two variables such that \( f \begin{pmatrix} x_1, x_2, x_3 \\ y_1, y_2, y_3 \end{pmatrix} = I \begin{pmatrix} x_1 + x_2 + x_3 \\ y_1 + y_2 + y_3 \end{pmatrix} \).

4. Before proving Theorem 3, we briefly sketch and consequently correct a slip made in [2] (in section 4, proof of Theorem 3).

If \( H \) is a distribution satisfying

\[
(11) \quad H(x, y, z) = H(x + y, 0, z) + H(x, y, 0)
\]

and \( H(\lambda x, \lambda y, \lambda z) = \lambda H(x, y, z) \)
on appropriate domain, as in [2], we obtain

\[
H_1(x, y, z) = G(x + y + z) - a(x), \quad H_2(x, y, z) = G(x + y + z) - \beta(y),
\]

\[
H_3(x, y, z) = G(x + y + z) - \gamma(z),
\]

where \( a, \beta, \gamma \) are different distributions \( (H_1(x, y, z) \) is the partial derivative of \( H \) with respect to \( x \) etc.) and

\[
(12) \quad H(x, y, z) = (x + y + z)G(x + y + z) - x\alpha(x) - y\beta(y) - z\gamma(z),
\]

\[
(13) \quad f_1(x) + f_2(y) + f_3(z) = g(x + y + z),
\]

where \( f_1(x) = x^2\alpha'(x), f_2(x) = x^2\beta'(x), f_3(x) = x^2\gamma'(x) \) and

\[
(14) \quad f_i(x) = ax + b_i \quad (i = 1, 2, 3)
\]

(cf. (12)–(14) above with (13)–(15) in [2] for corrections). Consequently,

\[
a(x) = a \log x - \frac{b_1}{x} + c_1, \quad \beta(x) = a \log x - \frac{b_2}{x} + c_2,
\]

\[
\gamma(x) = a \log x - \frac{b_3}{x} + c_3 \quad \text{and} \quad G(x) = a \log x - \frac{b_1 + b_2 + b_3}{x} + d,
\]

where \( c_i \) \((i = 1, 2, 3)\) and \( d \) are constants, so that (11) becomes

\[
H(x, y, z) = a[(x + y + z)\log(x + y + z) - x\log x - y\log y - z\log z] +
\]

\[
+ d(x + y + z) - c_1 x - c_2 y - c_3 z.
\]

Because of the symmetry of \( H \), we get \( c_1 = c_2 = c_3 = 0 \). Thus,

\[
H(x, y, z) = a[(x + y + z)\log(x + y + z) - x\log x - y\log y - z\log z] +
\]

\[
+ (d - c)(x + y + z),
\]

same as in [2]. This \( H \) satisfies (11) if \( d - c = 0 \), that is,

\[
(15) \quad H(x, y, z) = a[(x + y + z)\log(x + y + z) - x\log x - y\log y - z\log z].
\]
5. Proof of Theorem 3. Let \((x_1, x_2, x_3) \in D^6\). First of all we note that (9) leads to the Euler formula

\begin{equation}
0 = y_1 K_1(x_1, x_2, x_3) + y_2 K_2(x_1, x_2, x_3) + y_3 K_3(x_1, x_2, x_3).
\end{equation}

Differentiating (7) with respect to \(y_3\), we get

\begin{equation}
K_3(x_1, x_2, x_3) + K_3(x_1 + x_2, 0, x_3) y + K_3(x_1, y_2, x_3) + K_3(y_1 + y_2, 0, y_3).
\end{equation}

Next on differentiating (17) with regard to \(x_1\) and \(x_2\), we find

\begin{equation}
K_{31}(x_1, x_2, x_3) = K_{31}(x_1 + x_2, 0, y_3),
\end{equation}

so that \(K_{31} = K_{32}\) and because of the symmetry of \(K^3\) and so of \(K\), we obtain

\begin{equation}
K_{31} = K_{32} = K_{23} = K_{21} = K_{12} = K_{31} = K_{12} = K_{11} = K''
\end{equation}

which implies that \(K''\) is also symmetric.

Similarly we can prove that

\begin{equation}
K_{31}^{(1)} = K_{32}^{(2)} = K_{23}^{(3)} = K_{21}^{(1)} = K_{12}^{(2)} = K_{31}^{(3)}
\end{equation}

From (18), it follows that

\begin{equation}
K''(x_1, x_2, x_3) + K''(x_1 + x_2, 0, x_3) y + K''(y_1 + y_2, 0, x_3) y = K''(y_2 + y_3, 0, y_3).
\end{equation}

In view of Lemma 2, there is a distribution \(I\) of two variables such that

\begin{equation}
K''(x_1, x_2, x_3) = I(x_1 + x_2 + x_3).
\end{equation}

Using (19) and (21), since \(K'' = K_{31} = K_{32}\), we have

\begin{equation}
K_{31}(x_1, x_2, x_3) = G(x_1 + x_2 + x_3) + P(x_1 + x_2 + x_3),
\end{equation}

where \(G(x)\) is a primitive distribution of \(K(x)\) with respect to \(y\) and \(P, Q\) are distributions of five variables. Hence, there is a distribution of four
variables $F_3$ such that

\begin{equation}
K_3(x_1, x_2, x_3) = G(x_1 + x_2 + x_3, y_1 + y_2 + y_3) - F_3(x_1, x_2, x_3, y_3).
\end{equation}

Similarly, there exist distributions $F_1$ and $F_2$ of four variables such that

\begin{align}
K_1(x_1, x_2, x_3) &= G(x_1 + x_2 + x_3, y_1 + y_2 + y_3) - F_1(x_1, x_2, x_3, y_1), \\
K_2(x_1, x_2, x_3) &= G(x_1 + x_2 + x_3, y_1 + y_2 + y_3) - F_2(x_1, x_2, x_3, y_2).
\end{align}

Now (16), (22) and (23) yield

\begin{equation}
f_1(x_1, x_2, x_3, y_1) + f_2(x_1, x_2, x_3, y_2) + f_3(x_1, x_2, x_3, y_3) = \frac{g(x_1 + x_2 + x_3, y_1 + y_2 + y_3)}{y_1 + y_2 + y_3},
\end{equation}

where

\begin{equation}
f_i(x_1, x_2, x_3) = y_i F_i(x_1, x_2, x_3, y_i), \quad g(x) = yG(x).
\end{equation}

Differentiating (24) successively with respect to $y_1, y_2, y_3$ and denoting the derivatives by $f'_i, g'$, we have

\begin{equation}
f_1(x_1, x_2, x_3) = f'_1(x_1, x_2, x_3) = f'_2(x_1, x_2, x_3) = f'_3(x_1, x_2, x_3) = g'(x_1 + x_2 + x_3, y_1 + y_2 + y_3)
\end{equation}

which gives rise to a distribution $A$ of one variable such that

\begin{equation}
f_i(x_1, x_2, x_3) = y_i A(x_1 + x_2 + x_3) + B_i(x_1, x_2, x_3) \quad (i = 1, 2, 3),
\end{equation}

\begin{equation}
g(x_1 + x_2 + x_3, y_1 + y_2 + y_3) = (y_1 + y_2 + y_3) A(x_1 + x_2 + x_3) + B(x_1, x_2, x_3),
\end{equation}

where $B_1, B_2, B_3, B$ are distributions of three variables and $B = B_1 + B_2 + + B_3$. Consequently (22), (23), (25) and (26) lead to

\begin{equation}
K_{(i)}(x_1, x_2, x_3) = \frac{B(x_1, x_2, x_3)}{y_1 + y_2 + y_3} - \frac{B_j(x_1, x_2, x_3)}{y_j},
\end{equation}

which on integration with respect to $y_j$, results to

\begin{equation}
K(x_1, x_2, x_3) = B(x_1 + x_2 + x_3) \log(x_1 + x_2 + x_3) - B_j(x_1, x_2, x_3) y_j + C_j,
\end{equation}

where $C_j$ is a distribution of the variables $x_1, x_2, x_3$ and $y_k$ for $k \neq j$. 

\textit{Information theory, I}
Now (27) can be rewritten as
\[
K \left( x_1, x_2, x_3 \middle| y_1, y_2, y_3 \right) = B \log(y_1 + y_2 + y_3) - B_1 \log y_1 - B_2 \log y_2 - B_3 \log y_3 + D_j,
\]
where \( D_j = c_j + D \sum_{k \neq j} B_k \log y_k \), i.e., \( D_j \) does not depend on \( y_j \), so that it is easy to see that \( D_1 = D_2 = D_3 \), all depending upon \( x_1, x_2, x_3 \) only, say equal to \( B_3 \). Thus
\[
(28) \quad K \left( x_1, x_2, x_3 \middle| y_1, y_2, y_3 \right) = B(x_1, x_2, x_3) \log(y_1 + y_2 + y_3) - \sum_{i=1}^{3} B_i(x_1, x_2, x_3) \log y_i + B_0(x_1, x_2, x_3).
\]

From (28) results
\[
(29) \quad K^{(i)}_{(j)} \left( x_1, x_2, x_3 \middle| y_1, y_2, y_3 \right) = \frac{B^{(i)}(x_1, x_2, x_3)}{y_1 + y_2 + y_3} - \frac{B^{(i)}_j(x_1, x_2, x_3)}{y_j}
\]
for \( i, j = 1, 2, 3 \) which with (20) give
\[
\frac{B^{(i)}_j(x_1, x_2, x_3)}{y_j} = \frac{B^{(i)}_k(x_1, x_2, x_3)}{y_k} \quad (k, j \neq i)
\]
for arbitrary indices \( i, j, k = 1, 2, 3 \). Since \( y_j, y_k \) are linearly independent, we have \( B^{(i)}_j = 0 \) for \( i \neq j \). Thus
\[
(30) \quad B_j(x_1, x_2, x_3) = b_j(x_j) \quad \text{for} \quad j = 1, 2, 3,
\]
where \( b_j \) is a distribution of one variable.

Since \( B = B_1 + B_2 + B_3 \), (29) and (30) give
\[
(31) \quad K^{(i)}_{(j)} \left( x_1, x_2, x_3 \middle| y_1, y_2, y_3 \right) = \frac{b_i(x_i)}{y_1 + y_2 + y_3} \quad \text{for} \quad j \neq i.
\]

In the same way, from (20) and (31), we obtain
\[
b_i'(x_i) = b_2'(x_2) = b_3'(x_3) = \beta,
\]
where \( \beta \) is a constant.

Hence, \( B_i(x_1, x_2, x_3) = b_i(x_i) = \beta x_i + d_i \), where \( d_i \) \((i = 1, 2, 3)\) are constants, and \( B(x_1, x_2, x_3) = \beta (x_1 + x_2 + x_3) + d_1 + d_2 + d_3 \) so that (28) becomes
\[
(32) \quad K \left( x_1, x_2, x_3 \middle| y_1, y_2, y_3 \right) = B_0(x_1, x_2, x_3) + \beta (x_1 + x_2 + x_3) \log(y_1 + y_2 + y_3) +
+ (d_1 + d_2 + d_3) \log(y_1 + y_2 + y_3) - \beta \sum_{i=1}^{3} x_i \log y_i - \sum_{i=1}^{3} d_i \log y_i.
\]
Symmetry of \( K \) yields \( d_1 = d_2 = d_3 = d \) (say). Then the substitution
of (32) in (7) gives \( d = 0 \) and

\[
B_0(x_1, x_2, x_3) = B_0(x_1 + x_2, 0, x_3) + B_0(x_1, x_2, 0)
\]

with \( B_0 \) symmetric.

Now, it is easy to see from (32), since \( d = 0 \), that \( B_0 \) is positively homogeneous. Thus, by [2] or by using section 4 and (15) we have

\[
B_0(x_1, x_2, x_3) = a \left[ \left( \sum_{i=1}^{3} x_i \right) \log \left( \sum_{i=1}^{3} x_i \right) - \sum_{i=1}^{3} x_i \log x_i \right].
\]

Hence \( K \) has the required form (10) and this completes the proof of Theorem 3.

6. Proof of Theorem 1. By Theorem 3, \( K \) satisfying (7), (8) and (9) has the form (10). Hence from (6) and (10), we get

\[
K^2 \left( p_1, p_2, p_3 \right) = (p_1 + p_2 + p_3)^2 \left( \frac{p_1}{p_1 + p_2 + p_3}, \frac{p_2}{p_1 + p_2 + p_3}, \frac{p_3}{p_1 + p_2 + p_3} \right) \left( \frac{p_1}{q_1 + q_2 + q_3}, \frac{p_2}{q_1 + q_2 + q_3}, \frac{p_3}{q_1 + q_2 + q_3} \right)
\]

\[
= K \left( p_1, p_2, p_3 \right) = A \left( -\sum_{i=1}^{3} p_i \log p_i \right) + B \left( -\sum_{i=1}^{3} p_i \log q_i \right),
\]

which with (5) gives

\[
K^2 \left( p_1, p_2 \right) = A \left( -p_1 \log p_1 - p_2 \log p_2 \right) + B \left( -p_1 \log q_1 - p_2 \log q_2 \right).
\]

Now \( (a_3) \) and (34) yield the sought for result (3). This proves Theorem 1.

7. Directed divergence and inaccuracy. In order to obtain directed divergence, use the initial conditions

\[
K^2 \left( \frac{1}{2}, 0 \right) = 1 \quad \text{and} \quad K^2 \left( \frac{1}{2}, \frac{1}{2} \right) = 0
\]

(refer [6], [4], [7]).

In order to obtain inaccuracy, use the initial conditions,

\[
K^2 \left( \frac{1}{2}, 0 \right) = 1 \quad \text{and} \quad K^2 \left( \frac{1}{2}, \frac{1}{2} \right) = 1
\]

(refer [4]).

References


INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES, KATOWICE
UNIVERSITY OF WATERLOO, WATERLOO (CANADA)
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES, KATOWICE

*Reçu par la Rédaction le 3. 5. 1976*