Annales Univ. Sci. Budapest., Sect. Comp. 35 (2011) 239–253

SYMMETRIC DEVIATIONS AND DISTANCE MEASURES

Wolfgang Sander (Braunschweig, Germany)

Dedicated to Professor Antal Járai on his 60th birthday

Abstract. In this paper we characterize measurable information measures depending upon two probability distributions in a unified manner in order to get most of the existing information measures. Moreover it turns out that our characterization contains new, unexpected information measures.

1. Introduction

In this paper we investigate information measures on the open domain depending upon two probability distributions which are also called deviations (or similarity, affinity or divergence measures). Thus a deviation is a sequence (M_n) of functions, where

$$M_n : \Gamma_n^2 \to \mathbb{R},\ n \in \mathbb{N},\ n \geq 2.$$

Here

$$\Gamma_n = \left\{ P = (p_1, \ldots, p_n) \middle| \quad p_i \in I, \sum_{i=1}^{n} p_i = 1 \right\} \tag{1.1}$$

2000 Mathematics Subject Classification: Primary 94A17, Secondary 39B52.

Key words and phrases: Information measures, open domain, sum form, weighted additivity, polynomially additivity.

denotes the set of all discrete n-ary complete positive probability distributions and I denotes the open interval (0,1).

In Shore and Johnson [11] it is shown that each deviation (M_n) which satisfies the four desirable conditions of uniqueness, invariance, system independency and subset independency has a sum form representation

$$M_n(P,Q) = \sum_{i=1}^{n} f(p_i, q_i) \tag{1.2}$$

for some generating function $f : I^2 \to \mathbb{R}$. This result underlines the fact that each known deviation has a sum form, and it is thus natural to assume that a deviation has the sum form property (1.2) for some generating function f.

Many known deviations have a symmetric generating function f that is, $f(p,q) = f(q,p)$ for all $p, q \in I$. If a deviation (M_n) is not symmetric then going over to $M'_n(P,Q) = M_n(P,Q) + M_n(Q,P)$ means that M'_n has a symmetric generating function $f'(p,q) = f(p,q) + f(q,p)$.

The problem of how to characterize all sum form deviations, that is to find some natural conditions which imply the explicite form of the generating function, arises.

In Ebanks et al [3] (see chapter 5) two results were proven for information measures (M_n) depending upon two probability distributions $P, Q \in \Gamma_n$ satisfying a sufficient "fullness" of the range of (M_n) (the range $\{M_n(\Gamma_n^2) | n = 2, 3, ...\}$ has infinite cardinality):

1. For $P, Q \in \Gamma_n, U, V \in \Gamma_m$ we introduce $P * U, Q * U, P * V, Q * V \in \Gamma_{nm}$, where

$$(P * U, Q * V) =$$
$$= ((p_1u_1, ..., p_1u_m, ..., p_nu_1, ..., p_nu_m), (q_1v_1, ..., q_1v_m, ..., q_nv_1, ..., q_nv_m))).$$

Now, if (M_n) has the sum form property with some generating function f and if $M_{nm}(P*U, Q*V) = h(M_n(P,Q), M_m(U,V))$ for some polynomial $h : \mathbb{R}^2 \to \mathbb{R}$ and for all $m, n \geq 2$, then it is shown that h is a symmetric polynomial of degree at most one so that

$$\begin{aligned} &M_{nm}(P * U, Q * V) = \\ &= M_n(P,Q) + M_m(U,V) + \lambda M_n(P,Q) M_m(U,V) \end{aligned} \tag{1.3}$$

for some $\lambda \in \mathbb{R}$.

2. If (M_n) has the sum form property with some generating function f, and there are distributions $P', Q' \in \Gamma_n, U', V' \in \Gamma_m$ such that $I_n(P', Q') \neq 0$, respectively $I_m(U',V') \neq 0$ and

$$M_{nm}(P*U,Q*V) = \\ = A(U,V)M_n(P,Q) + B(P,Q)M_m(U,V) \tag{1.4}$$

for some "weights" A and B, then A and B have the sum form

$$A(U,V) = \sum_{j=1}^{m} M(u_j,v_j) \quad , \quad B(P,Q) = \sum_{i=1}^{n} M'(p_i,q_i) \tag{1.5}$$

for some generating multiplicative functions $M, M' : \mathbb{R}_+^2 \to \mathbb{R}$.

We remark that in Ebanks et al [3] the results in **1.** and **2.** were proven for information measures depending upon k probability distributions, but the special case k = 2 with the notation $(P,Q)*(U,V) = (P*U,Q*V)$ leads exactly to the above (nontrivial) results given in (1.3)–(1.5).

We now assume that the generating function f is symmetric in (1.3) and (1.4) and that $M = M'$ is symmetric so that $M(p,q) = M'(p,q) = M_1(p)M_1(q)$ for some multiplicative function $M_1 : I \to \mathbb{R}$ (since a multiplicative function of two variables is the product of two multiplicative functions in one variable).

Then we form the expression $M_{nm}(P*U,Q*V) + M_{nm}(P*V,Q*U)$ to get

$$M_{nm}(P*U,Q*V) + M_{nm}(P*V,Q*U) = \\ = 2M_n(P,Q) + 2M_m(U,V) + \lambda' M_n(P,Q)M_m(U,V) \tag{1.6}$$

and

$$M_{nm}(P*U,Q*V) + M_{nm}(P*V,Q*U) = \\ = 2A(U,V)\cdot M_n(P,Q) + 2A(P,Q)\cdot M_m(U,V), \tag{1.7}$$

from (1.3) and (1.4) respectively, where $\lambda' = 2\lambda$ and where

$$2A(P,Q) = \sum_{i=1}^{n} 2M_1(p_i)M_1(q_i) \ , \ 2A(U,V) = \sum_{j=1}^{m} 2M_1(u_j)M_1(v_j). \tag{1.8}$$

Thus a common generalization of the deviations given in (1.6) and (1.7) leads to the following class of deviations:

Definition 1.1. A deviation (M_n) is a symmetrically weighted compositive sum form deviation of additive-multiplicative type if (M_n) satisfies

$$M_{nm}(P*U,Q*V) + M_{nm}(P*V,Q*U) = \\ = G_m(U,V)M_n(P,Q) + G_n(P,Q)M_m(U,V) + \lambda M_n(P,Q)M_m(U,V), \tag{1.9}$$

for some $\lambda \in \mathbb{R}$, for all $m, n \geq 2$ and for all $P, Q \in \Gamma_n$, $U, V \in \Gamma_m$ with $P * U, Q * U, P * V, Q * V \in \Gamma_{nm}$, where M_n and G_n have the sum form

$$(1.10) \quad M_n(P,Q) = \sum_{i=1}^{n} f(p_i, q_i), \qquad G_n(P,Q) = \sum_{i=1}^{n} g(p_i, q_i), \quad P, Q \in \Gamma_n$$

for some symmetric functions $f, g : I^2 \to \mathbb{R}$, and where g satisfies

$$(1.11) \quad g(pu, qv) + g(pv, qu) = g(p,q)g(u,v) \quad , \quad p, q, u, v \in I.$$

We say that (M_n) is measurable if f and g are measurable in each variable. Moreover, every symmetric deviation (M_n) satisfying $M_n(P,P) = 0$ is called a distance measure.

Note that (1.9) and (1.10) with $g(p,q) = p + q$ and $g(p,q) = 2M_1(p)M_1(q)$ lead to (1.6) and (1.7), respectively, and that both functions g satisfy (1.11).

Thus the deviations (M_n) given by (1.9) and (1.10) satisfy the following fundamental functional equation

$$(1.12) \quad \sum_{i=1}^{n}\sum_{j=1}^{m}[\, f(p_i u_j, q_i v_j) + f(p_i v_j, q_i u_j) - \\ -g(u_j, v_j)f(p_i, q_i) - g(p_i, q_i)f(u_j, v_j) - \lambda f(p_i, q_i)f(u_j, v_j)\,] = 0,$$

where g satisfies (1.11).

In this paper we will present the measurable solutions of (1.11) and (1.12), generalizing the result in Chung et al [2] where the measurable solutions of functional equation (1.6) were given.

Let us finally consider some examples in this introduction.

Kerridge's inaccuracy K_n or the directed divergence F_n is given by

$$(1.13) \quad K_n(P,Q) = -\sum_{i=1}^{n} p_i \log q_i, \qquad F_n(P,Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}\,.$$

Note that $K_n(P,P) = H_n(P)$ and $F_n(P,Q) = K_n(P,Q) - K_n(P,P)$, where H_n is the well-known Shannon-entropy. K_n and F_n are indeed errors or deviations due to using $Q = (q_1, \ldots, q_n)$ as an estimation of the true probability distribution $P = (p_1, \ldots, p_n)$.

A 1-parametric generalization of (F_n) is given by (F_n^α), the directed divergence of degree α,

$$(1.14) \quad F_n^\alpha(P,Q) = \begin{cases} F_n(P,Q) & \alpha = 1 \\ \dfrac{1}{2^{\alpha-1}-1}\left(\displaystyle\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} - 1\right) & \alpha \in \mathbb{R}\setminus\{1\}. \end{cases}$$

We see immediately that $\lim_{\alpha\to 1} F_n^\alpha = F_n^1 = F_n$. F_n^α is not symmetric in P and Q, but F_n^α can be symmetrized by going over to

$$J_n^\alpha(P,Q) = F_n^\alpha(P,Q) + F_n^\alpha(Q,P) \qquad P,Q \in \Gamma_n \tag{1.15}$$

so that we arrive at the J-divergence (J_n^α) of degree α, $\alpha \in \mathbb{R}$, which satisfies $J_n^\alpha(P,Q) = J_n^\alpha(Q,P)$. Again we have $\lim_{\alpha\to 1} J_n^\alpha = J_n^1$ (because of $\lim_{\alpha\to 1} F_n^\alpha = = F_n^1$).

A further generalization of J_n^α is given by

$$L_n^{\alpha,\gamma}(P,Q) = \begin{cases} 2^{1-\alpha} \sum_{i=1}^{n} (p_i^\alpha - q_i^\alpha) \log \frac{p_i}{q_i} & \alpha = \gamma \\ \frac{1}{2^{\alpha-1} - 2^{\gamma-1}} \sum_{i=1}^{n} (p_i^\alpha - q_i^\alpha)\left(q_i^{\gamma-\alpha} - p_i^{\gamma-\alpha}\right) & \alpha \neq \gamma, \end{cases} \tag{1.16}$$

the J-divergence of degree (α,γ). We get $L_n^{\alpha,1} = J_n^\alpha$ and $\lim_{\gamma\to\alpha} L_n^{\alpha,\gamma} = L_n^{\alpha,\alpha}$, therefore $L_n^{\alpha,\gamma}$ can be considered as a 2-parametric generalization of J_n^1.

The sequences (J_n^α) and $(L_n^{\alpha,\gamma})$ satisfy (1.9) and (1.10) indeed: In the first case we choose $\lambda = 2^{\alpha-1} - 1$ and $g(p,q) = p + q$ and in the second case $\lambda = 2^{\alpha-1} - 2^{\gamma-1}$ and $g(p,q) = p^\gamma + q^\gamma$, respectively (and the obvious choices for f (see (1.13) and (1.14)). Moreover, $L_n^{\alpha,\gamma}$ is a distance measure since $L_n^{\alpha,\gamma}(P,P) = = 0$.

Note that for example (for $\lambda \neq 0$ and $\gamma = 2\alpha$)

$$\frac{2^{2\alpha-1} - 2^{\alpha-1}}{\lambda} L_n^{\alpha,2\alpha}(P,Q) = \frac{1}{\lambda} \sum_{i=1}^{n} (p_i^\alpha - q_i^\alpha)^2 =: \frac{1}{\lambda} D_n^\alpha(P,Q), \tag{1.17}$$

i.e. for $\alpha = \frac{1}{2}$ we arrive at Jeffreys distance in Jeffreys [5].

In the following Lemma we finally cite for the convenience of the reader Lemma 2 and Lemma 4 of Riedel and Sahoo [10] which are needed in the proof of Lemma 2.1.

Lemma 1.2. (1) *Let $M : I^2 \to \mathbb{C}$ be a given multiplicative function. The function $f : I^2 \to \mathbb{C}$ satisfies the functional equation*

$$f(pu,qv) + f(pv,qu) = 2M(uv)f(p,q) + 2M(pq)f(u,v) \tag{1.18}$$

if and only if

$$f(,p,q) = M(p)M(q)\Big[L(p) + L(q) + l(\frac{p}{q}.\frac{p}{q})\Big], \tag{1.19}$$

where $L : I \to \mathbb{C}$ is an arbitrary logarithmic map and $l : I^2 \to \mathbb{C}$ is a bilogarithmic function.

(2) *Let $M_1, M_2 : I \to \mathbb{C}$ be any two nonzero multiplicative maps with $M_1 \neq M_2$. Then the function $f : I^2 \to \mathbb{C}$ satisfies the functional equation*

$$\begin{aligned} f(pu, qv) + f(pv, qu) = [M_1(u)M_2(v) + M_1(v)M_2(u)]f(p, q)+ \\ + [M_1(p)M_2(q) + M_1(q)M_2(p)]f(u, v) \end{aligned} \tag{1.20}$$

if and only if

$$\begin{aligned} f(p, q) = \\ = M_1(p)M_2(q)[L_1(p) + L_2(q)] + M_1(q)M_2(p)[L_1(q) + L_2(p)], \end{aligned} \tag{1.21}$$

where $L_1, L_2 : I \to \mathbb{C}$ are logarithmic functions.

2. Symmetrically weighted compositive sum form deviations

In order to solve the functional equation (1.11) and (1.12) we first determine the general solution of (1.11) and the corresponding "functional equation without the sums"

$$f(pu, qv) + f(pv, qu) = g(u, v)f(p, q) + g(p, q)f(u, v) + \lambda f(p, q)f(u, v) \tag{2.1}$$

for all $p, q, u, v \in I$.

Lemma 2.1. *The functions $f, g : I^2 \to \mathbb{R}, f \neq 0$ satisfy (1.11) and (2.1) for all $p, q \in I$ if and only if for all $p, q \in I$:*
in the case $\lambda = 0$

$$\begin{aligned} f(p, q) &= M_1(p)M_2(q)[L_1(p) + L_2(q)] + M_1(q)M_2(p)[L_1(q) + L_2(p)], \\ g(p, q) &= M_1(p)M_2(q) + M_1(q)M_2(p), \qquad M_1 \neq M_2 \end{aligned} \tag{2.2}$$

or

$$\begin{aligned} f(p, q) &= M(p)M(q)[L_3(p) + L_3(q) + l(p, p) + l(q, q) - 2l(p, q)], \\ g(p, q) &= 2M(p)M(q); \end{aligned} \tag{2.3}$$

and in the case $\lambda \neq 0$

$$\begin{aligned} f(p, q) &= \frac{1}{\lambda}([M_3(p)M_4(q) + M_3(q)M_4(p)]- \\ &\qquad - [M_5(p)M_6(q) + M_5(q)M_6(p)]), \\ g(p, q) &= M_5(p)M_6(q) + M_5(q)M_6(p), \end{aligned} \tag{2.4}$$

where $c \neq 0$, $M : \mathbb{R}_+ \to \mathbb{R}$ *and* $M_i : \mathbb{R}_+ \to \mathbb{C}, 1 \leq i \leq 6$ *are multiplicative functions,* $L_1, L_2, L_3 : \mathbb{R}_+ \to \mathbb{R}$ *are logarithmic functions and* $l : \mathbb{R}_+ \to \mathbb{R}$ *is a bilogarithmic function, i.e.* l *is logarithmic in both variables. Moreover,* M_{2i-1} *and* M_{2i} *are both real-valued or* M_{2i} *is the complex conjugate of* M_{2i-1}, $i = 1, 2, 3$.

Finally, if f *and* g *are measurable then* M, M_i, L *and* L_i *are measurable, too.*

Proof. We start with the case $\lambda \neq 0$ in (2.1). By substituting

$$h(p,q) = g(p,q) + \lambda f(p,q)$$

we obtain from (2.1) that

$$h(pu, qv) + h(pv, qu) = h(p,q)h(u,v), \tag{2.5}$$

that is, g and h both satisfy (1.11).

Thus we get from the general solution of (1.11) (see Chung et al [2]) that

$$\begin{aligned} g(p,q) &= M_5(p)M_6(q) + M_5(q)M_6(p) \quad p, q \in I, \\ h(p,q) &= M_3(p)M_4(q) + M_3(q)M_4(p) \quad p, q \in I, \end{aligned} \tag{2.6}$$

where $M_i : \mathbb{R}_+ \to \mathbb{C}, 3 \leq i \leq 6$, M_{2i-1} and M_{2i} are both real-valued or M_{2i} is the complex conjugate of $M_{2i-1}, i = 2, 3$. Using now the substitution for h we arrive at (2.4).

Now we treat the case $\lambda = 0$. Then we have to solve (1.11) and

$$f(pu, qv) + f(pv, qu) = g(u,v)f(p,q) + g(p,q)f(u,v). \tag{2.7}$$

The idea is to extend f and g simultaneously to functions $\bar{f}$, $\bar{g} : \mathbb{R}_+ \to \mathbb{R}$, where $\bar{f}$, $\bar{g}$ satisfy (1.11) and (1.2), too. Then it is possible to solve (1.11) and (2.7). It turns out that indeed it is only important to have the point (1,1) in the domain of f and g : putting $q = v = 1$ in (1.11) and (2.7) we get

$$\begin{aligned} g(p,u) &= g(p,1)g(u,1) - g(pu,1), \\ f(p,u) &= g(u,1)f(p,1) + g(p,1)f(u,1) - f(pu,1), \end{aligned}$$

respectively (so that it is sufficient to determine the functions $p \to g(p,1)$ and $p \to f(p,1)$).

Let us define

$$M : I \to \mathbb{R} \quad \text{by} \quad M(t) := \frac{1}{2}\, g(t,t) \quad , \quad t \in I \quad \text{and} \tag{2.8}$$

$$\bar{g} : \mathbb{R}_+ \to \mathbb{R} \quad , \quad \bar{g}(p,q) = \frac{g(tp,tq)}{M(t)}, \quad p, q \in \mathbb{R}_+ \tag{2.9}$$

(here (2.9) means that for given $p, q \in \mathbb{R}_+$ there is $t \in I$ such that $(tp, tq) \in I^2$). Then M is a multiplicative function which is different from zero everywhere. Moreover $\bar{g}$ is well-defined, is uniquely determined, is a continuation of g and satisfies (1.11) on $\mathbb{R}_+^2$ (see Chung et al [2]).

Before we define $\bar{f}$ we need to do some calculations first. Putting $u = v = t$ into (2.7) we obtain (with $G(t) := g(t,t) = 2M(t)$ and $F(t) := \frac{1}{2}f(t,t)$)

$$2f(tp, tq) = g(t,t)f(p,q) + g(p,q)f(t,t) = G(t)f(p,q) + 2F(t)g(p,q)$$

or

$$f(tp, tq) = M(t)f(p,q) + F(t)g(p,q), \qquad p, q \in I. \tag{2.10}$$

Substituting $p = q = t$ and $u = v = w$ into (2.7) we arrive at

$$F(tw) = F(t)M(w) + M(t)F(w), \qquad t, w \in I.$$

Then we get, defining $L(t) := \frac{F(t)}{M(t)}$ and dividing the last equation by $M(tw)$,

$$L(tw) = L(t) + L(w), \qquad t, w \in I. \tag{2.11}$$

Thus L is logarithmic. We now define the continuation $\bar{f} : \mathbb{R}_+ \to \mathbb{R}$ by

$$\bar{f}(p,q) = \frac{f(tp, tq)}{M(t)} - L(t)\bar{g}(p,q), \qquad p, q \in \mathbb{R}_+\,, \tag{2.12}$$

where for each $p, q \in \mathbb{R}_+$ we choose $t \in I$ such that $tp, tq \in I$.

In order to show that $\bar{f}$ is well-defined, we choose (for given $p, q \in \mathbb{R}_+$) $t, w \in$ $\in I$, $t \neq w$ such that $tp, tq, wp, wq \in I$. We have to prove that

$$\frac{f(tp, tq)}{M(t)} - L(t)\bar{g}(p,q) = \frac{f(wp, wq)}{M(w)} - L(w)\bar{g}(p,q)$$

or, equivalently

$$M(w)f(tp, tq) - F(t)M(w)\bar{g}(p,q) = M(t)f(wp, wq) - F(w)M(t)\bar{g}(p,q),$$
$$M(w)f(tp, tq) + F(w)g(tp, tq) = M(t)f(wp, wq) + F(t)g(wp, wq).$$

But the last equation is equivalent with the obvious identity (see (2.10))

$$f(w(tp), w(tq)) = f(t(wp), t(wq)).$$

The function $\bar{f}$ is indeed a continuation of f: Choose $t = p \in I$ to get

$$\bar{f}(p,q) = \frac{f(p^2, pq)}{M(p)} - L(p)\bar{g}(p,q) =$$
$$= \frac{1}{M(p)}(M(p)f(p,q) + F(p)g(p,q)) - \frac{F(p)}{M(p)}\bar{g}(p,q) = f(p,q)$$

from (2.12) and (2.10) for $q \in I$

We show that $\bar{f}$ and $\bar{g}$ satisfy (2.7) for all $p, q \in \mathbb{R}_+$. For $p, q, u, v \in \mathbb{R}_+$ choose $t \in I$ such that $tp, tq, tu, tv \in I$. Using (2.10) and (2.7) we get (using $M(t^2) = M(t)^2$ and $L(t^2) = 2L(t)$)

$$\begin{aligned} &\bar{f}(pu, qv) + \bar{f}(pv, qu) = \\ &= \frac{f(tptu, tqtv)}{M(t^2)} - L(t^2)\bar{g}(pu, qv) + \frac{f(tptv, tqtu)}{M(t^2)} - L(t^2)\bar{g}(pv, qu) = \\ &= \bar{g}(u, v)(\frac{f(tp, tq)}{M(t)} - L(t)\bar{g}(p, q)) + \bar{g}(p, q)(\frac{f(tu, tv)}{M(t)} - L(t)\bar{g}(u, v)) = \\ &= \bar{g}(u, v)\bar{f}(p, q) + \bar{g}(p, q)\bar{f}(u, v). \end{aligned}$$

In order to prove, that f is uniquely determined, let us assume that $\tilde{f} : \mathbb{R}_+^2 \to \mathbb{R}$ is an extension of f satisfying also (2.7) for all $p, q, u, v \in \mathbb{R}_+$. Now choose for $p, q \in \mathbb{R}_+$ an element $t \in I$ such that $tp, tq \in I$ and put $u = v = t$ in (2.7). We get (since $\tilde{f} = f$ on I)

$$2\tilde{f}(tp, tq) = 2M(t)\tilde{f}(p, q) + 2\tilde{f}(t, t)g(p, q)$$

or, solving the last equation for $\tilde{f}(p, q)$ we see that

$$\tilde{f}(p, q) = \frac{\tilde{f}(tp, tq)}{M(t)} - g(p, q)\frac{F(t)}{M(t)} = \frac{f(tp, tq)}{M(t)} - L(t)g(p, q) = \bar{f}(p, q).$$

Simplifying the notation we don't distinguish f and $\bar{f}$, and g and $\bar{g}$ and suppose that f satisfies (2.7) for all $p, q, u, v \in \mathbb{R}_+$ and assume that g has the form

$$g(p, q) = M_1(p)M_2(q) + M_1(q)M_2(p), \qquad p, q \in \mathbb{R}_+\,, \tag{2.13}$$

for some multiplicative functions $M_1, M_2 : \mathbb{R}_+ \to \mathbb{C}_+$, where M_1 and M_2 are both real-valued or M_2 is the complex conjugate of M_1.

Now we consider two cases: $M_1 \neq M_2$ and $M_1 = M_2 = M'$ in (2.13), respectively.

In the first case we get the solution (2.2) from Lemma 4 in Riedel and Sahoo [10] and in the second case we get the solution (2.3) from Lemma 2 in Riedel and Sahoo [10] (in these Lemmas the domain of the functions f, M, M_1, M_2 is $(0, 1]$ or $(0, 1]^2$ and the range is $\mathbb{C}$, but the proofs can be taken over directly for our domains and ranges).

Moreover the proofs of the two Lemmas show that the measurability of f and g imply the measurability of the functions M, L, L_i and M_i. ■

Note that f and g are both symmetric although it was not supposed.

Theorem 2.2. *All measurable, symmetrically weighted compositive sum form deviations* (M_n) *of additive-multiplicative type are given as follows: in the case* $\lambda = 0$ *by*

$$M_n(P,Q) = \sum_{i=1}^{n} [p_i^\gamma q_i^\delta (a \log p_i + b \log q_i) + p_i^\delta q_i^\gamma (a \log q_i + b \log p_i)] \tag{2.14}$$

or

$$M_n(P,Q) = \sum_{i=1}^{n} p_i^\rho q_i^\rho \left[c \log(p_i q_i) + d \left(\log \frac{p_i}{q_i} \right)^2 \right], \tag{2.15}$$

and in the case $\lambda \neq 0$ *by*

$$M_n(P,Q) = -\frac{1}{\lambda} \sum_{i=1}^{n} \left(p_i^\gamma q_i^\delta + p_i^\delta q_i^\gamma \right) \tag{2.16}$$

or

$$M_n(P,Q) = -\frac{1}{\lambda} \sum_{i=1}^{n} 2 p_i^\rho q_i^\rho \cos \left(\sigma \log \frac{p_i}{q_i} \right) \tag{2.17}$$

or

$$M_n(P,Q) = \frac{1}{\lambda} \sum_{i=1}^{n} \left[\left(p_i^\alpha q_i^\beta + p_i^\beta q_i^\alpha \right) - \left(p_i^\gamma q_i^\delta + p_i^\delta q_i^\gamma \right) \right] \tag{2.18}$$

or

$$M_n(P,Q) = \frac{1}{\lambda} \sum_{i=1}^{n} \left[2 p_i^\rho q_i^\rho \cos \left(\sigma \log \frac{p_i}{q_i} \right) - \left(p_i^\gamma q_i^\delta + p_i^\delta q_i^\gamma \right) \right], \tag{2.19}$$

where $a, b, c, d, \alpha, \beta, \gamma, \delta, \rho, \sigma$ *are arbitrary real constants with* $\alpha \neq \beta$ *and* $\gamma \neq \delta$.

Proof. We start from the fundamental equation (1.12) and substitute

$$h(p,q) = g(p,q) + \lambda f(p,q) \tag{2.20}$$

in the case $\lambda \neq 0$ into (1.12). Then (using (2.5)) equation (1.12) turns into $\sum_{j=1}^{n} F(u_j, v_j) = 0$ for all $U, V \in \Gamma_m$ and for all $m, n \geq 2$ where for fixed $P, Q \in \Gamma_n$

$$F(u,v) = \begin{cases} \sum\limits_{i=1}^{n} (f(p_i u, q_i v) + f(p_i v, q_i u) - g(u,v) f(p_i, q_i) - g(p_i, q_i) f(u,v)) \\ \text{if } \lambda = 0, \\ \sum\limits_{i=1}^{n} \Big[h(p_i u, q_i v) + h(p_i v, q_i u) - h(u,v) h(p_i, q_i) \Big] \text{ if } \lambda \neq 0. \end{cases}$$

The fact that $F : I^2 \to \mathbb{R}$ is measurable and satisfies $\sum_{j=1}^{n} F(u_j, v_j) = 0$ for all $U, V \in \Gamma_n$ and for all $n \geq 2$ implies

$$F(u,v) = a(u-v), \quad u,v \in I^2 \text{for some real constant } a.$$

Indeed, for $n = 2$ we get with $U = (u, 1-u),\ V = (v, 1-v) \in \Gamma_2$

$$F(u,v) + F(1-u, 1-v) = 0 \ \text{ for all } \ u, v \in I.$$

For $n = 3$ we get with $U = (u_1, u_2, 1-(u_1+u_2)),\ V = (v_1, v_2, 1-(v_1+v_2)) \in \Gamma_3$ that

$$F(u_1, v_1) + F(u_2, v_2) + F(1-(u_1+u_2), 1-(v_1, v_2)) = 0.$$

But from last two equations result we obtain the 2-dimensional Cauchy-functional equation

$$F(u_1, v_1) + F(u_2, v_2) + F(u_1+u_2, v_1+v_2),$$

$$u_1, u_2, u_1+u_2, v_1, v_2, v_1+v_2 \in I.$$

Thus $F(u,v) = au + bv$ for some constants $a, b \in \mathbb{R}$. But then we obtain

$$\sum_{j=1}^{n} F(u_j, v_j) = \sum_{j=1}^{n} (au_j + bv_j) = a + b = 0.$$

Thus $a = -b$ and F has the form $F(u,v) = a(u-v)$. Since F is measurable and symmetric (since f and g are symmetric) we get $F(u,v) = a(u-v) = = F(v,u) = -a(u-v)$ for some constant a. Letting P, Q vary again we see that $a(P,Q) = -a(P,Q) = 0$ and so $F = 0$, too.

Now for fixed $u, v \in \Gamma_n$ we define

$$(2.21) \quad G(p,q) = \begin{cases} f(pu,qv) + f(pv,qu) - g(u,v)f(p,q) - g(p,q)f(u,v) \\ \text{if} \quad \lambda = 0, \\ \\ h(pu,qv) + h(pv,qu) - h(u,v)h(p,q) \quad \text{if} \quad \lambda \neq 0. \end{cases}$$

Again, G is measurable, symmetric and satisfies

$$(2.22) \qquad \sum_{i=1}^{n} G(p_i, q_i) = F(u,v) = 0,$$

and so that like above $G = 0$. This means that f satisfies

1. (2.7) (that is, g is given by (2.13)) and (1.11), or
2. $G(p,q) = 0$, where h satisfies (1.11) and g is given by (2.13) (see (2.20)).

CASE 1. From (2.2) in Lemma 2.1 we obtain (using that L_1 and L_2 are measurable)

$$f(p,q) = p^\gamma q^\delta (a\log p + b\log q) + p^\delta q^\gamma (a\log q + b\log p)\ p,q \in I \tag{2.23}$$

for some constants $a, b, \gamma, \delta, \gamma \neq \delta$.

From (2.2) in Lemma 2.1 we get for arbitrary, but fixed p, q that

$$L_3(p) = c\log p,\ c \in \mathbb{R}, \qquad l(p,q) = d(q)\log p = l(q,p) = d(p)\log q \tag{2.24}$$

which implies $d(p) = d\log p$ for some $d \in \mathbb{R}$. Using this we arrive at

$$f(p,q) = p^\rho q^\rho \Big(c\log(p\cdot q) + d\Big(\log^2 p + \log^2 q - 2\log p\log q\Big)\Big), \quad \rho \in \mathbb{R}. \tag{2.25}$$

Thus we get (2.14) and (2.15) by using the sum form of (M_n).

CASE 2. From (2.20) we get $f(p,q) = \frac{1}{\lambda}(h(p,q) - g(p,q))$, so Lemma 2.1 implies the representation (2.4) for f. Like in Chung et al [2] we get

$$g(p,q) = p^\alpha q^\beta + q^\alpha p^\beta \quad \text{or} \quad g(p,q) = 2p^\rho q^\rho \cos(\sigma \log \frac{p_i}{q_i}), \tag{2.26}$$

$$h(p,q) = p^\gamma q^\delta + q^\gamma p^\delta \quad \text{or} \quad h(p,q) = 2p^\mu q^\mu \cos(\nu \log \frac{p_i}{q_i}) \tag{2.27}$$

for some constants $\alpha, \beta, (\alpha \neq \beta), \gamma, \delta, (\gamma \neq \delta), \rho, \sigma, \mu, \nu$. Then the cases $h = 0$ and $h \neq 0$ lead to the solutions in (2.16) - (2.19).

Reversely, all solutions, given by (2.14)–(2.19) satisfy (1.9). ■

Theorem 2.3. *A deviation (M_n) fulfills the conditions of Theorem 2.2 and satisfies $M_n(P,P) = 0$ iff*

$$M_n(P,Q) = a\sum_{i=1}^n \left(p_i^\gamma q_i^\delta - p_i^\delta q_i^\gamma\right)\log\frac{p_i}{q_i} \quad ,\gamma \neq \delta, \quad \lambda = 0 \tag{2.28}$$

or

$$M_n(P,Q) = b\sum_{i=1}^n \left(\log\frac{p_i}{q_i}\right)^2, \qquad \lambda = 0 \tag{2.29}$$

or

$$M_n(P,Q) = \frac{1}{\lambda}\sum_{i=1}^n \left(p_i^\alpha q_i^\delta - q_i^\alpha p_i^\delta\right)\left(q_i^{\gamma-\alpha} - p_i^{\gamma-\alpha}\right),\ \lambda \neq 0 \tag{2.30}$$

or

(2.31)

$$M_n(P,Q) = \frac{1}{\lambda}\sum_{i=1}^{n}\left(2p_i^{\frac{\gamma+\delta}{2}}q_i^{\frac{\gamma+\delta}{2}}\cos\left(\sigma\log\frac{p_i}{q_i}\right) - \left(p_i^{\gamma}q_i^{\delta} + p_i^{\delta}q_i^{\gamma}\right)\right),\ \lambda \neq 0,$$

where $a, b, \alpha, \gamma, \delta, \sigma$ are arbitrary constants.

Proof. We put $P = Q$ into (2.14)–(2.19) to obtain

$$M_n(P,P) = \sum_{i=1}^{n} 2(a+b)p_i^{\gamma+\delta}\log p_i, \tag{2.32}$$

$$M_n(P,P) = \sum_{i=1}^{n} 2c\cdot p_i^{2\rho}\log p_i, \tag{2.33}$$

$$M_n(P,P) = -\frac{2}{\lambda}\sum_{i=1}^{n} p_i^{\gamma+\delta} \neq 0, \tag{2.34}$$

$$M_n(P,P) = -\frac{2}{\lambda}\sum_{i=1}^{n} p_i^{2\rho} \neq 0, \tag{2.35}$$

$$M_n(P,P) = \frac{2}{\lambda}\sum_{i=1}^{n}\left(p_i^{\alpha+\beta} - p_i^{\gamma+\delta}\right), \tag{2.36}$$

and

$$M_n(P,P) = \frac{2}{\lambda}\sum_{i=1}^{n}\left(p_i^{2\rho} - p_i^{\gamma+\delta}\right), \tag{2.37}$$

respectively. Now we consider $M_n(P,P) = 0$ in all cases. We get $b = -a$ in (2.32) and $c = 0$ in (2.33), which imply (2.28) and (2.29), respectively. Moreover, (2.34) and (2.35) lead to no solution, whereas (2.36) leads to $\alpha+\beta = \gamma+\delta$. Putting $\beta = \gamma+\delta-\alpha$ into (2.18) we have (2.30). Finally, $M_n(P,P) = 0$ in (2.37) implies $2\rho = \gamma+\delta$ which gives (2.31). ■

The above distance measures contain many known measures as special case. Let us mention the following examples:

(a) $\delta = 0$ in (2.28) gives

$$M_n(P,Q) = a2^{\gamma-1}L_n^{\gamma,\gamma}(P,Q).$$

(b) $\delta = 0$ in (2.29) results in

$$M_n(P,Q) = \frac{2^{\alpha-1} - 2^{\gamma-1}}{\lambda}L_n^{\alpha,\gamma}(P,Q).$$

(c) $\alpha = 0$ in (2.30) leads to

$$M_n(P,Q) = -\frac{1}{\lambda}\sum_{i=1}^{n}\left(\sqrt{p_i^{\gamma} q_i^{\delta}} - \sqrt{p_i^{\delta} q_i^{\gamma}}\right)^2.$$

(d) $(\gamma, \delta) \in (1,0), (0,1)$ in (c) yields

$$M_n(P,Q) = \frac{1}{\lambda}\sum_{i=1}^{n}\left(\sqrt{p_i} - \sqrt{q_i}\right)^2 = \frac{1}{\lambda} D_n^{\frac{1}{2}}(P,Q) \quad \text{(see (1.17))}.$$

(e) Note that

$$D_n^{\frac{1}{2}}(P,Q) = \frac{2}{\lambda}\Big[1 - B_n(P,Q)\Big],$$

where $B_n(P,Q) = \sum_{i=1}^{n}\sqrt{p_i q_i}$ is the Hellinger coefficient (see Hellinger [4]).

(f) If $\gamma = 2\alpha$ and $\delta = 1$ in (c) then we get

$$M_n(P,Q) = \frac{1}{\lambda}\sum_{i=1}^{n}\left(p_i^{\alpha} - q_i^{\alpha}\right)^2 = \frac{2^{2\alpha-1} - 2^{\alpha-1}}{\lambda} L_n^{\alpha,2\alpha}(P,Q) =$$
$$= \frac{1}{\lambda} D_n^{\frac{1}{2}}(P,Q).$$

References

[1] **Aczél, J.,** On different characterizations of entropies, In: *Probability and Information Theory, Proc. Internat. Sympos., McMaster Univ., Hamilton, Ontario,* 1968, Lecture Notes in Math., vol. **89**, Springer, New York, 1969, 1–11.

[2] **Chung, J.K., PL. Kannappan, C.T. Ng and P.K. Sahoo,** Measures of distance between probability distributions. *J. Math. Anal. Appl.,* **139** (1989), 280–292.

[3] **Ebanks, B., P.K. Sahoo and W. Sander,** *Characterizations of Information Measures,* World Scientific, Singapore, New Jersey, London, Hongkong (1998).

[4] **Hellinger, E.,** *Die Orthogonalvarianten quadratischer Formen von unendlich vielen Variablen,* Dissertation, Göttingen, 1907.

[5] **Jeffreys, H.,** An invariant form for the prior probability in estimation problems, *Proc. Roy. Soc. London, Ser. A,* **186** (1946), 453–461.

[6] **Kannappan, PL. and P.K. Sahoo,** Sum form distance measures between probability distributions and functional equations, *Intern. J. Math. Stat. Sci.,* **6** (1997), 91–105.

[7] **Kerridge, D.F.,** Inaccuracy and inference, *J. Roy. Statist. Soc. Ser. B,* **23** (1961), 184–194.

[8] **Kullback, S.,** *Information Theory and Statistics,* John Wiley and Sons, Inc., New York, 1959.

[9] **Rényi, A.,** On measures of entropy and information, In: *Proc. 4th Berkeley Sympos. Math.Statist. and Prob., Vol. I,* 547–561.

[10] **Riedel, T. and P.K. Sahoo,** On a generalization of a functional equation associated with the distance between the probability distributions, *Publ. Math. Debrecen,* **46** (1995), 125–135.

[11] **Shore, J.E. and R.W. Johnson,** Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inform. Theory, IT,* **26** (1980), 26–37.

Wolfgang Sander
Computational Mathematics
TU Braunschweig
38106 Braunschweig Pockelsstr. 14
Germany
w.sander@tu-bs.de