

# Kybernetika

VOLUME 39 (2003), NUMBER 1

The Journal of the Czech Society for  
Cybernetics and Information Sciences

Published by:

Institute of Information Theory  
and Automation of the Academy  
of Sciences of the Czech Republic

Editor-in-Chief:

Milan Mareš

Managing Editors:

Karel Sladký

Editorial Board:

Jiří Anděl, Marie Demlová, Petr Hájek,  
Martin Janžura, Jan Ježek, Radim Jiroušek,  
George Klir, Ivan Kramosil, Friedrich Liese,  
Jean-Jacques Loiseau, František Matúš,  
Radko Mesiar, Jiří Outrata, Jan Štecha,  
Olga Štěpánková, Igor Vajda, Pavel Zítek,  
Pavel Žampa

Editorial Office:

Pod Vodárenskou věží 4, 182 08 Praha 8

*Kybernetika* is a bi-monthly international journal dedicated for rapid publication of high-quality, peer-reviewed research articles in fields covered by its title.

*Kybernetika* traditionally publishes research results in the fields of Control Sciences, Information Sciences, System Sciences, Statistical Decision Making, Applied Probability Theory, Random Processes, Fuzziness and Uncertainty Theories, Operations Research and Theoretical Computer Science, as well as in the topics closely related to the above fields.

The Journal has been monitored in the Science Citation Index since 1977 and it is abstracted/indexed in databases of Mathematical Reviews, Current Mathematical Publications, Current Contents ISI Engineering and Computing Technology.

---

*Kybernetika*. Volume 39 (2003)

ISSN 0023-5954, MK ČR E 4902.

Published bi-monthly by the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. — Address of the Editor: P. O. Box 18, 182 08 Prague 8, e-mail: kybernetika@utia.cas.cz. — Printed by PV Press, Pod vrstevnicí 5, 140 00 Prague 4. — Orders and subscriptions should be placed with: MYRIS TRADE Ltd., P. O. Box 2, V Štíhlách 1311, 142 01 Prague 4, Czech Republic, e-mail: myris@myris.cz. — Sole agent for all “western” countries: Kubon & Sagner, P. O. Box 34 01 08, D-8 000 München 34, F.R.G.

Published in February 2003.

© Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, Prague 2003.

**BLENDED  $\phi$ -DIVERGENCES WITH EXAMPLES**

VÁCLAV KŮS

Several new examples of divergences emerged in the recent literature called blended divergences. Mostly these examples are constructed by the modification or parametrization of the old well-known  $\phi$ -divergences. Newly introduced parameter is often called *blending parameter*. In this paper we present compact theory of blended divergences which provides us with a generally applicable method for finding new classes of divergences containing any two divergences  $D_0$  and  $D_1$  given in advance. Several examples of blends of well-known divergences are given.

*Keywords:* divergences of probability distributions, blended divergences, statistical applications

*AMS Subject Classification:* 62B10, 62F35, 62G35

**1. INTRODUCTION AND BASIC CONCEPTS**

Lindsay [5] introduced a new class of divergences by the modification of weights inside the integral expression of Pearson's  $\chi^2$ -divergence. He called this divergence "blended weight chi-squared disparity",  $BWCS(\beta)$ , and the weight parameter  $\beta \in [0, 1]$  called *blending parameter*. Similarly, he obtained "blended weight Hellinger disparity"  $BWHD(\beta)$ . Lindsay used these blended classes of disparities to achieve better efficiency and robustness of estimators based on  $BWCS$  or  $BWHD$ .

Park and Basu [8] deal with two further modifications of blended Hellinger disparity and they called it "combined" and "penalized" variant of the Hellinger distance. They presented computer simulation study for the corresponding estimators and tests in case of some discrete models, in particular for the Poisson and geometric distributions and their mixtures.

Kůs [3] introduced several examples of new classes of divergences based on a method of normalization of a convex or concave functions. Some of these divergences were shown to have a blend interpretation if we use a reparametrization by means of blending parameter  $\beta$ .

In general, all these new classes of disparities have the following common property. If the blending parameter is equal to the limiting values  $\beta = 0$  or  $\beta = 1$ , then the two original divergences, on which the blend was based, are achieved in this class of blended divergences.

Menéndez et al [6] introduced a general method for obtaining such blended divergences and they stated some theoretical results concerning these blends without proofs. They also proposed to use blends as new disparity statistics for grouped data on which the goodness of fit testing procedures are based. Moreover, asymptotic distribution for appropriately scaled  $\phi$ -disparity statistic was proved to be the  $\chi_{m+1}^2$  distribution, where  $m+1$  denotes the number of intervals for a given partition of  $\mathbb{R}$ .

In this paper, which is based on Menéndez et al [6], we present a compact theory of blended divergences which provides us with a generally applicable method for finding new classes of divergences connecting any two divergences  $D_0$  and  $D_1$  given in advance. We use this method to obtain blended divergences originated from the family of  $I_a$ -divergences.

First let us define  $\phi$ -divergences with its basic properties. For a systematic theory of  $\phi$ -divergences we refer to Vajda [10] and, for some additional recent results on  $\phi$ -divergences, also to KÛs [3].

**Definition 1.** Let  $\mathcal{P}$  be the set of all probability measures on a measurable space  $(\mathcal{X}, \mathcal{A})$ . We define  $\phi$ -divergence of two measures  $P$  and  $Q$  from  $\mathcal{P}$  by

$$D_\phi(P, Q) = \int_{\mathcal{X}} q \phi\left(\frac{p}{q}\right) d\mu, \quad (1)$$

where  $\mu$  is a  $\sigma$ -finite measure on  $(\mathcal{X}, \mathcal{A})$  such that  $\{P, Q\} \ll \mu$ , and  $p = dP/d\mu$ ,  $q = dQ/d\mu$  denote the Radon–Nikodym derivative of  $P$ ,  $Q$  with respect to  $\mu$ . We assume that divergence function  $\phi : (0, \infty) \rightarrow \mathbb{R}$  is convex on  $(0, \infty)$  and strictly convex at  $t = 1$ , with  $\phi(1) = 0$ . On the boundary of the open domain  $p, q > 0$  we extend the definition by  $q \phi(p/q) = q \phi(0)$  if  $p = 0$  and  $q \phi(p/q) = p \phi(\infty)/\infty$  if  $q = 0$ , where  $\phi(0) = \lim_{t \rightarrow 0^+} \phi(t)$  and  $\phi(\infty)/\infty = \lim_{t \rightarrow \infty} \phi(t)/t$  with the convention “ $0 \cdot \infty = 0$ ”.

Divergences (1) are all reflexive and the range of  $D_\phi(P, Q)$  is

$$0 \leq D_\phi(P, Q) \leq \phi(0) + \phi(\infty)/\infty, \quad P, Q \in \mathcal{P},$$

where the upper bound is achieved if  $P, Q$  are two singular measures. The values of  $D_\phi(P, Q)$  do not depend on a linear term of the form  $c(t-1)$  added to, or extracted from, divergence function  $\phi$ . It means that every  $\phi$  has its nonnegative version

$$\tilde{\phi}(t) = \phi(t) - \phi'_+(1)(t-1), \quad t \in (0, \infty), \quad (2)$$

equivalent to  $\phi$  with respect to the same values of  $D_\phi$  divergence. ( $\phi'_+(1)$  denotes the derivatives of  $\phi$  at  $t = 1$  from the right.)

In this paper we make use of a *conjugated* divergence functions  $\phi^*$  which usually results in *reversed*  $\phi$ -divergences.

**Proposition 1.** Let  $\phi$  be a divergence function. Then the *conjugated function*  $\phi^*(t) = t\phi(1/t)$ ,  $t \in (0, \infty)$ , is also a divergence function,  $\phi^*(0) = \phi(\infty)/\infty$ ,  $\phi^*(\infty)/\infty = \phi(0)$ , and  $D_{\phi^*}(P, Q) = D_{\phi}(Q, P)$  for all  $P, Q \in \mathcal{P}$ . Moreover, the divergence  $D_{\phi}$  is symmetric if and only if there exists a real constant  $c$  such that  $\phi(t) = \phi^*(t) + c(t-1)$ ,  $t \in (0, \infty)$ .

The examples of blended divergences, presented in Section 3, are taken mostly from the class of power  $I_a$ -divergences

$$I_a(P, Q) = \frac{1}{a(a-1)} \left( \int p^a q^{1-a} d\mu - 1 \right), \quad P, Q \in \mathcal{P},$$

defined for  $a \neq 0, 1$  by means of the divergence function

$$\phi_a(t) = \frac{t^a - a(t-1) - 1}{a(a-1)}, \quad t \in (0, \infty),$$

with

$$\phi_a(0) = \begin{cases} 1/a & \text{if } a > 0, a \neq 1, \\ \infty & \text{if } a \leq 0, \end{cases} \quad \phi_a(\infty)/\infty = \begin{cases} 1/(1-a) & \text{if } a < 1, \\ \infty & \text{if } a \geq 1. \end{cases}$$

Note that the class of  $I_a$ -divergences contains twice Hellinger divergence  $H^2(P, Q)$  for  $a = 1/2$ , half of the Pearson's  $\chi^2(P, Q)$  divergence for  $a = 2$ , and half of the Neyman's  $\chi^2(Q, P)$  divergence for  $a = -1$ . The limits of  $I_a$  at  $a = 1$  and  $a = 0$  provide us with the Kullback–Leibler divergence  $I_0(P, Q) = I(P, Q)$  and the reversed Kullback–Leibler divergence  $I_1(P, Q) = I(Q, P)$ . The conjugated function of Proposition 1 is  $\phi_a^*(t) = \phi_{1-a}(t)$ ,  $a \neq 0, 1$ . Further properties and applications of  $I_a$ -divergences can be found, for example, in Vajda [10], Lindsay [5], Cressie and Read [1], and Read and Cressie [9].

## 2. BLENDS OF DIVERGENCES

**Theorem 1.** Let  $L : [0, \infty) \rightarrow [0, \infty)$  be a linear function ( $L \not\equiv 0$ ) with  $L_{\infty} := \lim_{t \rightarrow \infty} L(t)/t \geq 0$  and  $\phi(y)$  be a divergence function strictly convex at  $y = L(1)$ . The function  $\phi_L(t) : (0, \infty) \rightarrow \mathbb{R}$ , defined by

$$\phi_L(t) = \phi(L(t)) - \phi(L(1)), \quad t \in (0, \infty),$$

is a divergence function,  $\phi_L(0) = \phi(L(0)) - \phi(L(1))$ ,  $\phi_L(\infty)/\infty = L_{\infty} \cdot \phi(\infty)/\infty$ , and

$$D_{\phi_L}(P, Q) = D_{\phi}(L_{\infty} P + L(0) Q, Q) - \phi(L(1)), \quad P, Q \in \mathcal{P}.$$

**Proof.** It is clear that  $\phi_L : \mathbb{R}^+ \rightarrow \mathbb{R}$  and  $\phi_L(1) = 0$ . Since for all  $\alpha \in (0, 1)$  and  $t_1, t_2 \in \mathbb{R}^+$

$$\begin{aligned} \phi_L(\alpha t_1 + (1-\alpha)t_2) &= \phi(L(\alpha t_1 + (1-\alpha)t_2)) - \phi(L(1)) \\ &\leq \alpha\phi(L(t_1)) + (1-\alpha)\phi(L(t_2)) - \phi(L(1)) = \alpha\phi_L(t_1) + (1-\alpha)\phi_L(t_2) \end{aligned}$$

then  $\phi_L$  is a convex function on  $(0, \infty)$ . Further, strict convexity of  $\phi(y)$  at  $y = L(1)$  implies strict convexity of  $\phi_L(t)$  at  $t = 1$ . The two assertions concerning limits  $\phi_L(0)$  and  $\phi_L(\infty)/\infty$  are trivial consequences of the definition of  $\phi_L$ . To prove the last assertion, assume first that  $L(0) \neq 0$  and  $L_\infty \neq 0$ . Then

$$\begin{aligned} D_\phi(L_\infty P + L(0) Q, Q) &= \int_{A_+ \cap \{q>0\}} q \phi\left(\frac{L_\infty p + L(0) q}{q}\right) d\mu \\ &+ \phi(0) \int_{A_0} q d\mu + \phi(\infty)/\infty \int_{\{q=0\}} (L_\infty p + L(0) q) d\mu \end{aligned} \quad (3)$$

where

$$A_+ = \{L_\infty p + L(0) q > 0\} \quad \text{and} \quad A_0 = \{L_\infty p + L(0) q = 0\}.$$

Let us denote the first integral in (3) by  $I_1$ , the second integral by  $I_2$ , and the third one by  $I_3$ . Then

$$\begin{aligned} I_2 &= 0, \\ I_3 &= L_\infty \phi(\infty)/\infty \int_{\{q=0\}} p d\mu = \phi_L(\infty)/\infty \int_{\{q=0\}} p d\mu, \\ I_1 &= \int_{\{pq>0\}} q \phi\left(\frac{L_\infty p + L(0) q}{q}\right) d\mu + \int_{\{p=0, q>0\}} q \phi(L(0)) d\mu \\ &= \int_{\{pq>0\}} q \phi\left(L\left(\frac{p}{q}\right)\right) d\mu + \phi(L(0)) \int_{\{p=0\}} q d\mu \\ &= \int_{\{pq>0\}} q \left[ \phi\left(L\left(\frac{p}{q}\right)\right) - \phi(L(1)) \right] d\mu \\ &\quad + \left[ \phi(L(0)) - \phi(L(1)) \right] \int_{\{p=0\}} q d\mu + \phi(L(1)) \int_{\{p \geq 0\}} q d\mu \\ &= \int_{\{pq>0\}} q \phi_L\left(\frac{p}{q}\right) d\mu + \phi_L(0) \int_{\{p=0\}} q d\mu + \phi(L(1)). \end{aligned}$$

Thus we obtain from (3) that

$$D_\phi(L_\infty P + L(0) Q, Q) = D_{\phi_L}(P, Q) + \phi(L(1)).$$

If  $L_\infty = 0$  then  $\phi_L(t) = \phi(L(0)) - \phi(L(1))$  and  $D_{\phi_L}(P, Q) = \phi(L(0)) - \phi(L(1))$ . Further,

$$D_\phi(L_\infty P + L(0) Q, Q) - \phi(L(1)) = D_\phi(L(0) Q, Q) - \phi(L(1)) = \phi(L(0)) - \phi(L(1)).$$

If  $L(0) = 0$  then

$$D_{\phi_L}(P, Q) = \int q \left[ \phi \left( L_\infty \frac{p}{q} \right) - \phi(L(1)) \right] d\mu = D_\phi(L_\infty P, Q) - \phi(L(1)).$$

Thus all assertions of the theorem are proved.  $\square$

Note that if  $L(0) = 0$  then  $D_{\phi_L}$  is bounded iff  $D_\phi$  is bounded. If  $L(0) \neq 0$  then  $D_{\phi_L}$  is bounded iff  $\phi(\infty)/\infty < \infty$ . Thus  $D_{\phi_L}$  can be bounded while  $D_\phi$  is not bounded. It means that if we apply suitable linear transformation  $y = L(t)$  on  $\phi(y)$ , we can obtain bounded divergence in spite of the fact that the original  $D_\phi$  was unbounded in the sense that  $\phi(0) = \infty$  and  $\phi(\infty)/\infty < \infty$ .

**Corollary 1.** Let  $\phi$  be a divergence function and  $\phi^*(t)$  is conjugated to  $\phi$ . If we define for all  $\beta \in [0, 1]$  the functions  $\phi_{S,\beta}(t)$  by

$$\phi_{S,\beta}(t) = \phi(1 - \beta + \beta t) = (1 - \beta + \beta t) \phi^* \left( \frac{1}{1 - \beta + \beta t} \right), \quad t \in (0, \infty),$$

then all  $\phi_{S,\beta}$  are divergence functions,  $\phi_{S,\beta}(0) = \phi(1 - \beta)$ ,  $\phi_{S,\beta}(\infty)/\infty = \beta \phi(\infty)/\infty$ , and

$$D_{\phi_{S,\beta}}(P, Q) = D_\phi(\beta P + (1 - \beta)Q, Q), \quad P, Q \in \mathcal{P}.$$

If  $\phi(\infty)/\infty < \infty$  and  $\beta \neq 0$  then  $D_{\phi_{S,\beta}}$  are bounded.

**Corollary 2.** Let  $\phi$  be a divergence function. If we define for all  $\beta \in [0, 1]$  the functions  $\phi_{R,\beta}(t)$  by

$$\phi_{R,\beta}(t) = (1 - \beta + \beta t) \phi \left( \frac{t}{1 - \beta + \beta t} \right), \quad t \in (0, \infty),$$

then all  $\phi_{R,\beta}$  are divergence functions,  $\phi_{R,\beta}(0) = (1 - \beta) \phi(0)$ ,  $\phi_{R,\beta}(\infty)/\infty = \beta \phi(1/\beta)$  (where we take  $\beta \phi(1/\beta) = \phi(\infty)/\infty$  if  $\beta = 0$ ), and

$$D_{\phi_{R,\beta}}(P, Q) = D_\phi(P, \beta P + (1 - \beta)Q), \quad P, Q \in \mathcal{P}.$$

If  $\phi(0) < \infty$  and  $\beta \neq 0$  then  $D_{\phi_{R,\beta}}(P, Q)$  are bounded.

*Proof.* Observe that the following relation exists between  $\phi_R$  and  $\phi_S$  of Corollaries 1 and 2,

$$\begin{aligned} \phi_{R,\beta}(t) &= t \phi^* \left( \frac{1 - \beta + \beta t}{t} \right) = t \phi^* \left( \beta + (1 - \beta) \frac{1}{t} \right) \\ &= \left( \phi^*(\beta + (1 - \beta) t) \right)^*(t) = \left( (\phi^*)_{S,1-\beta} \right)^*(t). \end{aligned}$$

Thus, Proposition 1 and Corollary 1 with  $L(t) = \beta + (1 - \beta)t$  imply that the function  $\phi_{R,\beta}(t)$  is a divergence function and further

$$\begin{aligned} D_{\phi_{R,\beta}}(P, Q) &= D_{(\phi^*)_{S,1-\beta}}(Q, P) && \text{(Proposition 1)} \\ &= D_{\phi}^*(\beta P + (1 - \beta)Q, P) && \text{(Corollary 1)} \\ &= D_{\phi}(P, \beta P + (1 - \beta)Q) && \text{(Proposition 1)} \end{aligned}$$

The same reasoning can be applied to prove the remaining assertions of Corollary 2.  $\square$

Corollaries 1 and 2 can serve to construct a new, possibly bounded,  $\phi_{S,\beta}$  or  $\phi_{R,\beta}$ -divergences. Both the corollaries also justify the correctness of the following definition of blended divergences.

**Definition 2.** Let  $\phi_0$  and  $\phi_1$  be divergence functions and  $\beta \in [0, 1]$ . The function

$$\phi_{\beta}(t) = (1 - \beta + \beta t) \phi_0 \left( \frac{t}{1 - \beta + \beta t} \right) + \phi_1(1 - \beta + \beta t), \quad t \in (0, \infty),$$

is said to be *blended divergence function* and the corresponding  $\phi_{\beta}$ -divergence

$$D_{\beta}(P, Q) := D_{\phi_{\beta}}(P, Q) = D_{\phi_0}(P, \beta P + (1 - \beta)Q) + D_{\phi_1}(\beta P + (1 - \beta)Q, Q)$$

is called *blended divergence*, more precisely  $\beta$ -blend of  $D_{\phi_0} = D_0$  and  $D_{\phi_1} = D_1$ .

Corollary 1 and 2 imply that

$$\begin{aligned} \phi_{\beta}(0) &= (1 - \beta) \phi_0(0) + \phi_1(1 - \beta), \\ \phi_{\beta}(\infty)/\infty &= \beta \phi_0(1/\beta) + \beta \phi_1(\infty)/\infty, \end{aligned}$$

where we take  $\beta \phi_0(1/\beta) = \phi_0(\infty)/\infty$  if  $\beta = 0$ . The order of blended divergences  $D_{\phi_0}$  and  $D_{\phi_1}$  is substantial, i. e. the blend of  $D_{\phi_1}$  and  $D_{\phi_0}$  can differ from the blend of Definition 2 (see Examples 1 and 2 in the next section).

**Specification 1.** For a given  $\phi$ , if we take into account the special case  $\phi_0(t) = \phi(t)$  and  $\phi_1(t) = \phi^*(t)$ , then we get from Definition 2

$$\begin{aligned} \phi_{\beta}(t) &= (1 - \beta + \beta t) \phi \left( \frac{t}{1 - \beta + \beta t} \right) + (1 - \beta + \beta t) \phi \left( \frac{1}{1 - \beta + \beta t} \right) \\ &= t \phi^* \left( \beta + (1 - \beta) \frac{1}{t} \right) + \phi^*(1 - \beta + \beta t) \end{aligned}$$

with the corresponding blended divergence

$$D_{\beta}(P, Q) = D_{\phi}(P, \beta P + (1 - \beta)Q) + D_{\phi}(Q, \beta P + (1 - \beta)Q).$$

Thus, we obtain the *blend* of  $D_\phi(P, Q)$  and reversed  $D_\phi^*(P, Q) = D_\phi(Q, P)$ . The order of blending is again substantial. Further,

$$\begin{aligned}\phi_\beta(0) &= (1 - \beta)\phi(0) + (1 - \beta)\phi\left(\frac{1}{1 - \beta}\right), \\ \phi_\beta(\infty)/\infty &= \beta\phi(1/\beta) + \beta\phi(0),\end{aligned}$$

where we take  $\beta\phi(1/\beta) = \phi(\infty)/\infty$  if  $\beta = 0$ . Therefore, if  $\phi(0) = \infty$  then for all  $\beta \in [0, 1]$  the blended divergences  $D_\beta(P, Q)$  are unbounded. However, provided that  $\phi(0) < \infty$  and  $\beta \in (0, 1)$ , we get bounded blended divergences  $D_\beta$  irrespectively of the value of  $\phi(\infty)/\infty$ .

**Theorem 2.** Let  $\phi$  be a divergence function,  $\phi_0(t) = \phi(t)$  and  $\phi_1(t) = \phi^*(t)$ . Then for all  $\beta \in [0, 1]$  the blends  $D_\beta$  of  $D_\phi(P, Q)$  and reversed  $D_\phi(Q, P)$  are skew symmetric about  $\beta = 1/2$

$$D_\beta(P, Q) = D_{1-\beta}(Q, P), \quad \beta \in [0, 1], P, Q \in \mathcal{P},$$

with the symmetric blend

$$D_{\frac{1}{2}}(P, Q) = D_\phi\left(P, \frac{P+Q}{2}\right) + D_\phi\left(Q, \frac{P+Q}{2}\right).$$

*Proof.* By direct computation it can be verified that  $\phi_\beta^*(t) = \phi_{1-\beta}(t)$ ,  $t \in (0, \infty)$  which proves the skew symmetry. The symmetry of  $D_{1/2}(P, Q)$  follows immediately from Proposition 1, since  $\phi_{1/2}^*(t) = \phi_{1/2}(t)$ .  $\square$

If we intend to construct blended divergence, we can use either the given  $\phi(t)$  or  $\tilde{\phi}(t)$  or another similar variant of  $\phi(t)$  with the same divergence. However, the  $D_\beta(P, Q)$  blend of reversed divergences does not depend on the variant used since

$$\left(\tilde{\phi}\right)_\beta(t) = \phi_\beta(t) - (1 - 2\beta)\phi'_+(1)(t - 1)$$

and thus  $D_{(\tilde{\phi})_\beta}(P, Q) = D_{\phi_\beta}(P, Q) = D_\beta(P, Q)$ . Furthermore, provided  $\phi$  is twice differentiable at  $t = 1$ , the symmetric blend of Theorem 2 for  $\beta = 1/2$  is always based on nonnegative divergence function  $\phi_\beta$ , since  $\phi'_\beta(1) = (1 - 2\beta)\phi'(1)$ .

The natural question arises whether a certain power  $D_{1/2}^\alpha(P, Q)$ ,  $\alpha > 0$ , of the reflexive and symmetric blend  $D_{1/2}$  of Theorem 2 can represent a metric distance on  $\mathcal{P}$ . Kafka, Österreicher and Vincze [2] proved that a certain power  $D_\phi^\alpha(P, Q)$  of the symmetric  $\phi$ -divergence satisfies the triangle inequality if the function  $(1 - t^\alpha)^{1/\alpha}/\phi(t)$  is nonincreasing in the domain  $0 < t < 1$ . Unfortunately, to verify this sufficient condition for a given blended metric divergence  $D_{1/2}$  of Theorem 2 can be quite difficult (see Österreicher [7]).



### 3. EXAMPLES OF BLENDED DIVERGENCES

To illustrate the presented theory of blended divergences we give several examples taken from the family of  $I_a$ -divergences. However, two arbitrary  $\phi$ -divergences or even blends, given in advance, can be blended to achieve required statistical properties of estimators or tests based on the corresponding blend. The best way, how to apply the theory of blends, would be to design a blend fitted to the real statistical application.

**Example 1.** (Pearson–Neyman blend) If we choose

$$\phi_0(t) = \phi(t) = (t-1)^2 \quad \text{and} \quad \phi_1(t) = \phi^*(t) = \frac{(t-1)^2}{t}$$

then, applying Specification 1, we come to the blend of Pearson's  $D_0(P, Q) = \chi^2(P, Q)$  and Neyman's  $D_1(P, Q) = \chi^2(Q, P)$  divergences defined by means of blended divergence function

$$\begin{aligned} \phi_\beta(t) &= (1-\beta+\beta t) \left[ \left( \frac{t}{1-\beta+\beta t} - 1 \right)^2 + \left( \frac{1}{1-\beta+\beta t} - 1 \right)^2 \right] \\ &= \left[ (1-\beta)^2 + \beta^2 \right] \frac{(t-1)^2}{1-\beta+\beta t}, \end{aligned}$$

for all  $\beta \in [0, 1]$ . We restrict ourselves to a normalized blended divergence function  $\phi_\beta$  with  $\phi_\beta''(a) = 1$ , i. e.

$$\phi_\beta(t) = \frac{1}{2} \frac{(t-1)^2}{1-\beta+\beta t}, \quad t \in (0, \infty), \beta \in [0, 1],$$

with the corresponding blended divergence

$$D_\beta(P, Q) = \frac{1}{2} \int \frac{(p-q)^2}{\beta p + (1-\beta)q} d\mu, \quad P, Q \in \mathcal{P}.$$

This blend  $D_\beta$  coincides with the *generalized Le Cam divergence*  $LC_\beta$  investigated in Kùs [3]. This blend was found also by Lindsay [5] by the modification of weights inside the integral expression for Pearson's  $\chi^2$ -divergence,  $BWCS(\beta)$ , mentioned already in Section 1. Thus, the present example shows that blended divergences developed earlier in the literature remain to be  $\phi$ -divergences. Symmetric blend of Theorem 2,  $D_{\frac{1}{2}}(P, Q)$ , corresponds to the squared Le Cam distance  $LC^2(P, Q)$ .

**Example 2.** (Neyman–Pearson blend) On the other side, if we exchange  $\phi_0(t)$  with  $\phi_1(t)$  in Example 1 we get a blend of Neyman's  $D_0(P, Q) = \chi^2(Q, P)$  and Pearson's  $D_1(P, Q) = \chi^2(P, Q)$ , with

$$\begin{aligned} \widetilde{\phi}_\beta(t) &= t \left( \beta + (1-\beta) \frac{1}{t} - 1 \right)^2 + (1-\beta+\beta t - 1)^2 \\ &= \left[ (1-\beta)^2 + \beta^2 t \right] \frac{(t-1)^2}{t} = (1-\beta)^2 \phi^*(t) + \beta^2 \phi(t). \end{aligned}$$

Consequently,

$$\widetilde{D}_\beta(P, Q) = (1 - \beta)^2 \chi^2(Q, P) + \beta^2 \chi^2(P, Q).$$

The symmetric divergence of Theorem 2

$$\widetilde{D}_{\frac{1}{2}}(P, Q) = \frac{1}{4} \left( \chi^2(P, Q) + \chi^2(Q, P) \right) = \frac{1}{4} \int \frac{(p+q)(p-q)^2}{pq} d\mu$$

coincides with the symmetrized divergence  $J_2(P, Q)/2$  defined in Vajda [11]. Note that  $\phi_\beta$  of Example 1 differs from  $\widetilde{\phi}_\beta$  of Example 2 since for all  $\beta \in (0, 1)$  there is  $\phi_\beta(0) < \infty$  but  $\widetilde{\phi}_\beta(0) = +\infty$ . It means that the Neyman–Pearson blend produces unbounded divergences for all  $\beta \in [0, 1]$ , while the Pearson–Neyman blends are bounded for all  $\beta \in (0, 1)$ .

**Example 3.** (Blended power divergences – variant A) Both Examples 1 and 2 are the special cases (for  $a = 2$  and  $a = -1$ ) of blended  $I_a(P, Q)$  and  $I_{1-a}(P, Q)$  power divergences. For  $a \in \mathbb{R} - \{0, 1\}$  and

$$\phi_a(t) = \frac{t^a - 1}{a(a-1)}, \quad \phi_a^*(t) = \phi_{1-a}(t) - \frac{t-1}{a(a-1)}$$

we obtain by Specification 1

$$\phi_{a,\beta}(t) = \frac{1}{a(a-1)} \left[ \frac{t^a + 1}{(1 - \beta + \beta t)^{a-1}} - 2\beta(t-1) - 2 \right], \quad a \neq 0, 1,$$

where the linear term  $2\beta(t-1)$  can be omitted as it does not alter the value of the  $\phi_{a,\beta}$ -divergence

$$I_{a,\beta}(P, Q) = \frac{1}{a(a-1)} \left( \int \frac{p^a + q^a}{(\beta p + (1-\beta)q)^{a-1}} d\mu - 2 \right), \quad a \neq 0, 1,$$

as the blend of  $I_{a,0}(P, Q) = I_a(P, Q)$  and  $I_{a,1}(P, Q) = D_{\phi_a^*}(P, Q) = I_{1-a}(P, Q)$ . Note that the reversed order of blending  $I_{1-a}$  and  $I_a$  divergences is also included in the expressions for all  $a \neq 0, 1$ . If we use the following limits of  $\phi_a(t)$  as  $a \rightarrow 0$  and  $a \rightarrow 1$ ,

$$\phi_0(t) = -\ln t, \quad \phi_1(t) = \phi_0^*(t) = t \ln t,$$

then we get

$$\phi_{0,\beta}(t) = 2(1 - \beta + \beta t) \ln(1 - \beta + \beta t) - (1 - \beta) \ln t - \beta t \ln t$$

with the corresponding reversed *Kullback–Kullback blend*

$$\begin{aligned} I_{0,\beta}(P, Q) &= - \int (\beta p + (1 - \beta)q) \left[ \ln \frac{\beta p + (1 - \beta)q}{p} + \ln \frac{\beta p + (1 - \beta)q}{q} \right] d\mu \\ &= 2 \int (\beta p + (1 - \beta)q) \ln \frac{\beta p + (1 - \beta)q}{q} d\mu + (1 - \beta) I(Q, P) - \beta I(P, Q) \end{aligned}$$

where the last equality holds if the expression on the right hand side is meaningful. However, all the blended divergences  $I_{0,\beta}(P, Q)$  are unbounded since for all  $\beta \in [0, 1)$  there is  $\phi_{0,\beta}(0) = \infty$  and for  $\beta = 1$  it is  $\phi_{0,1}(\infty)/\infty = \infty$ .

On the other side, the interchanging of the order of the functions  $\phi_0$  and  $\phi_1$  results in the *Kullback–reversed Kullback blend*

$$\begin{aligned}\phi_{1,\beta}(t) &= t \ln t - (t+1) \ln(1-\beta+\beta t), \\ I_{1,\beta}(P, Q) &= \int \left( p \ln \frac{p}{\beta p + (1-\beta)q} + q \ln \frac{q}{\beta p + (1-\beta)q} \right) d\mu \\ &= I(P, Q) - \int (p+q) \ln \frac{\beta p + (1-\beta)q}{q} d\mu\end{aligned}$$

provided the last expression is meaningful. For all  $\beta \in (0, 1)$  the Kullback–reversed Kullback blend  $I_{1,\beta}$  is bounded since  $\phi_{1,\beta}(0) = -\ln(1-\beta)$  and  $\phi_{1,\beta}(\infty)/\infty = -\ln \beta$ . The symmetric blend of Theorem 2,

$$I_{1,\frac{1}{2}}(P, Q) = I\left(P, \frac{P+Q}{2}\right) + I\left(Q, \frac{P+Q}{2}\right)$$

coincides with the  $f_\alpha$ -divergence of Österreicher [7] for  $\alpha = 1$  which has been proved to be a squared metric distance.

**Example 4.** (Blended power divergences – variant B) To blend  $I_a$  and  $I_{-a}$  divergences we set, for  $0 < |a| < 1$ ,

$$\phi_{a,0}(t) = \frac{t^a - 1}{\text{sign}[a(a-1)]} \quad \text{and} \quad \phi_{a,1}(t) = \frac{t^{-a} - 1}{\text{sign}[a(a+1)]}$$

in place of  $\phi_0$  and  $\phi_1$  in Definition 2, respectively. Then for all  $\beta \in [0, 1]$

$$\phi_{a,\beta}(t) = -\text{sign}(a) \frac{(1-\beta+\beta t)t^a - 1}{(1-\beta+\beta t)^a}, \quad t \in (0, \infty), \quad 0 < |a| < 1,$$

where we have already omitted the linear term  $\beta(t-1)$  which has no influence on the corresponding blended divergence

$$D_{a,\beta}(P, Q) = -\text{sign}(a) \int \frac{(\beta p + (1-\beta)q) p^a - q^{a+1}}{(\beta p + (1-\beta)q)^a} d\mu, \quad 0 < |a| < 1.$$

If  $a \in (0, 1)$  then these blended divergences are bounded for all  $\beta \in (0, 1)$ . Note that the blend

$$D_{a,\frac{1}{2}}(P, Q) = -\text{sign}(a) 2^{a-1} \int \frac{(p+q)p^a - 2q^{a+1}}{(p+q)^a} d\mu$$

given by

$$\phi_{a,\frac{1}{2}}(t) = -\text{sign}(a) 2^{a-1} \frac{t^a(1+t) - 2}{(1+t)^a}, \quad t \in (0, \infty), \quad 0 < |a| < 1$$

is bounded for all  $a \in (0, 1)$ , but it is not symmetric (Theorem 2 is applicable only to the blends of mutually reversed divergences). Another variant of “blended”  $I_a$  and  $I_{-a}$  divergences was presented in KÛs [3].

**Example 5.** (Self-blended power divergences) If we put

$$\phi_{a,0}(t) = \phi_{a,1}(t) = \frac{t^a - 1}{a(a-1)}, \quad a \neq 0, 1,$$

then we obtain from Definition 2 *self-blended  $I_a$ -divergence* defined by the blended divergence function

$$\phi_{a,\beta}(t) = \frac{1}{a(a-1)} \left[ \frac{t^a + (1-\beta+\beta t)^{2a-1}}{(1-\beta+\beta t)^{a-1}} - 2 \right], \quad a \neq 0, 1, \beta \in [0, 1],$$

where we have omitted the linear term  $\beta(t-1)$  again. For example, the parameter  $a = 1/2$  defines a *self-blended Hellinger divergence*

$$H_\beta^S(P, Q) := D_{\frac{1}{2},\beta}(P, Q) = 4 \left( 2 - \int (\sqrt{p} + \sqrt{q}) \sqrt{\beta p + (1-\beta)q} d\mu \right)$$

given by means of divergence function  $\phi_{\frac{1}{2},\beta}(t) = 4(2 - (\sqrt{t} + 1)\sqrt{1-\beta+\beta t})$ ,  $t \in (0, \infty)$ ,  $\beta \in [0, 1]$ . The blends  $H_\beta^S$  are bounded divergences for all  $\beta \in [0, 1]$ . The symmetric divergence in this class is

$$H_{\frac{1}{2}}^S(P, Q) = 4 \left( 2 - \frac{\sqrt{2}}{2} \int (\sqrt{p} + \sqrt{q}) \sqrt{p+q} d\mu \right).$$

Similarly, for  $a = 2$  we obtain a *self-blended Pearson divergence* and  $a = -1$  leads to a *self-blended Neyman divergence*. The limiting case  $\phi_{0,0}(t) = -\ln t$  defines a *self-blended reversed Kullback divergence*  $KL_\beta(P, Q)$  using the blended divergence function

$$\phi_{0,\beta}(t) = \beta(t-1) \ln(1-\beta+\beta t) - (1-\beta) \ln t - \beta t \ln t$$

with corresponding divergences

$$\begin{aligned} KL_\beta(P, Q) &= KL(P, \beta P + (1-\beta)Q) + KL(\beta P + (1-\beta)Q, Q) \\ &= \beta \int (p-q) \ln \frac{\beta p + (1-\beta)q}{q} d\mu + (1-\beta)I(Q, P) - \beta I(P, Q), \end{aligned}$$

provided the last expression is meaningful. If we consider the second limit  $\phi_{1,0}(t) = t \ln t$ , we get a *self-blended Kullback divergence*  $I_\beta(P, Q)$  defined by means of

$$\phi_{1,\beta}(t) = t \ln t - (1-\beta)(t-1) \ln(1-\beta+\beta t), \quad \beta \in [0, 1],$$

as follows

$$\begin{aligned} I_\beta(P, Q) &= I(P, \beta P + (1-\beta)Q) + I(\beta P + (1-\beta)Q, Q) \\ &= I(P, Q) - (1-\beta) \int (p-q) \ln \frac{\beta p + (1-\beta)q}{q} d\mu. \end{aligned}$$

## ACKNOWLEDGMENT

We thank for the hospitality of the Operational Research Center at the Miguel Hernández University in Elche (Alicante) where the paper was finalized for publication. This research was partially supported through the project FRVS 596/2000 and through the grant No. 201/02/1391 of the Grant Agency of the Czech Republic.

(Received January 11, 2002.)

## REFERENCES

- 
- [1] N. Cressie and T. R. C. Read: Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. A* *46* (1984), 440–464.
  - [2] P. Kafka, F. Österreicher and I. Vincze: On powers of  $f$ -divergences defining a distance. *Studia Sci. Math. Hungar.* *26* (1991), 415–422.
  - [3] V. Kús: Divergences and Generalized Score Functions in Statistical Inference. Ph.D. Thesis, Czech Technical University, Prague 1998.
  - [4] L. Le Cam: *Asymptotic Methods in Statistical Decision Theory*. Springer, New York 1986.
  - [5] B. G. Lindsay: Efficiency versus robustness. The case for minimum Hellinger distance and other methods. *Ann. Statist.* *22* (1994), 1081–1114.
  - [6] M. Menéndez, D. Morales, L. Pardo, and I. Vajda: Two approaches to grouping of data and related disparity statistics. *Comm. Statist. A – Theory Methods* *27* (1998), 609–633.
  - [7] F. Österreicher.: On a class of perimeter-type distances of probability distributions. *Kybernetika* *32* (1996), 389–393.
  - [8] Ch. Park, A. Basu and S. Basu: Robust minimum distance inference based on combined distances. *Comm. Statist. B – Simulation Comput.* *24* (1995), 653–673.
  - [9] R. C. Read and N. A. C. Cressie: *Goodness-of-fit Statistics for Discrete Multivariate Data*. Springer, Berlin 1988.
  - [10] I. Vajda: *Theory of Statistical Inference and Information*. Kluwer, Boston 1989.
  - [11] I. Vajda: *Information–Theoretic Methods in Statistics*. Research Report No. 1834, ÚTIA, Academy of Sciences of the Czech Republic, Prague 1995.

*Ing. Václav Kús, Ph.D., Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering – Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2. Czech Republic.*  
*e-mail: Kus@km1.fjfi.cvut.cz*