O divergenci a fluktuaci proměnných veličin a pravděpodobnostních distribucí

Igor Vajda

Výzkumná zpráva ÚTIA AV ČR Praha 2007/10

1 Proč f-divergence a co to je

Předpokládejme, že se měří určitá fyzikální veličina $p = p(t)$ proměnná v čase (nebo na určité trase) $t \in T$. Budeme předpokládat $p(t) \geq 0$, t. j. že měřené hodnoty jsou nezáporné a jejich proměny spojité v závislosti na t. Bez újmy na obecnosti můžeme položit $t = [0, 1]$ a předpokládat, že veličina není triviální nula a měřítko je nastavené tak, aby

$$\int_0^1 p(t)\,\mathrm{d}t = 1. \tag{1.1}$$

Chování veličiny tudíž zadává spojitou pravděpodobnostní distribuci na intervalu $[0, 1]$.

Spojitá funkce $p(t) : [0, 1] \to [0, \infty)$ je však jen ideální matematická představa o měření. Ve skutečnosti se zaznamenávají průměrné (předpokládejme, že nenulové) hodnoty

$$p_i = \int_{t_{i-1}}^{t_i} p(t)\,\mathrm{d}t, \quad 1 \leq i \leq n \tag{1.2}$$

na intervalech (t_{i-1}, t_i) odpovídajících určitému dělení

$$0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1 \tag{1.3}$$

celého pozorovacího intervalu $[0, 1]$. Skutečným záznamem měření je tudíž vektor

$$P = (p_1, \ldots, p_n), \quad p_i > 0, \quad \sum_{i=1}^{n} p_i = 1 \tag{1.4}$$

který představuje diskrétní pravděpodobnostní distribuci. Z tohoto důvodu existuje úzký vztah mezi divergencí (mírou neshody) $D(p, q)$ dvou proměnných veličin

$$p = p(t), \quad q = q(t), \quad 0 \leq t \leq 1 \tag{1.5}$$

a divergencí $D(P, Q)$ jim odpovídajících diskrétních distribucí

$$P = (p_1, \ldots, p_n), \quad Q = (q_1, \ldots, q_n) \tag{1.6}$$

získaných kvantováním (1.2). V dalším se napřed zaměříme na vlastnosti, které musí mít divergence diskrétních pravděpodobnostních distribucí $D(P,Q)$ a z nich pak odvodíme odpovídající vlastnosti divergence spojitých pravděpodobnostních distribucí $D(p,q)$.

Rozumné je požadovat divergenci v jednoduchém součtovém tvaru

$$D(P,Q) = \sum_{i=1}^{n} \delta(p_i, q_i), \tag{1.7}$$

kde δ je spojitá funkce na čtverci $(0,1) \otimes (0,1)$. Vlastnosti této funkce poněkud osvětlíme tím, že budeme uvažovat hrubší kvantování průběhu měřených veličin, kdy místo oddělených intervalů (t_{k-1}, t_k), (t_k, t_{k+1}) použijeme jen jeden interval $(t_{k-1},\, t_{k+1})$. To znamená, že místo dvou záznamů p_k, p_{k+1} a q_k, q_{k+1} o veličinách p, q budeme mít k dispozici jen jejich spojení

$$\tilde{p}_k = \int_{t_{k-1}}^{t_{k+1}} p(t)\,\mathrm{d}t = p_k + p_{k+1} \quad \text{a} \quad \tilde{q} = q_k + a_{k+1}.$$

V nových distribucích

$$\begin{aligned} \tilde{P} &= (p_1, \ldots, p_{k-1}, \tilde{p}_k, p_{k+2}, \ldots, p_n)\,, \\ \tilde{Q} &= (q_1, \ldots, q_{k-1}, \tilde{q}_k, q_{k+2}, \ldots, q_n) \end{aligned}$$

jsou původní rozdíly mezi záznamy p_k, q_k a p_{k+1}, q_{k+1} o průběhu veličin p, q na intervalech (t_{k-1}, t_k) a t_k, t_{k+1} rozpuštěny v souhrnech $p_k + p_{k+1}$, $q_k + q_{k+1}$ záznamů na sjednoceném intervalu (t_{k-1}, t_{k+1}). Z toho důvodu musí divergence

$$D(\tilde{P}, \tilde{Q}) = \sum_{i \neq k,\, k+1} \delta(p_i, q_i) + \delta(\tilde{p}_k, \tilde{q}_k)$$

být menší nebo nejvýše rovna původní divergenci $D(P,Q)$.

Jinými slovy, funkce δ musí splňovat nerovnost

$$\delta(\tilde{p}_k, \tilde{q}_k) \leq \delta(p_k, q_k) + \delta(p_{k+1}, q_{k+1}) \tag{1.8}$$

představující základní teoretický princip v oblasti zpracování dat (viz *data processing condition* na str. 1289 v práci Pardo a Vajdy [6] a odkazy tam na něj uvedené). Tento princip vyvozuje logický důsledek – zhoršení rozlišitelnosti – ze ztráty informace zapříčiněné dodatečnou redukcí dat (hrubším kvantováním veličin). Rovnost se v neostré nerovnosti (1.8) připustí tehdy, když oba věrohodnostní poměry

$$\frac{p_k}{q_k} \quad \text{a} \quad \frac{p_{k+1}}{q_{k+1}}$$

před redukcí se shodují navzájem a tudíž i s věrohodnostním poměrem

$$\frac{\tilde{p}_k}{\tilde{q}_k} = \frac{p_k + p_{k+1}}{q_k + q_{k+1}}$$

po redukci, t. j. když platí

$$\frac{p_k + p_{k+1}}{q_k + q_{k+1}} = \frac{p_k}{q_k} = \frac{p_{k+1}}{q_{k+1}}. \tag{1.9}$$

V tomto případě se v matematické statistice příslušná redukce nazývá postačující a dokazuje se o ní, že nesnižuje kvalitu rozhodování při přechodu od souborů dat P, Q k souborům $\tilde{P}$, $\tilde{Q}$.

V Teorému 1 zmíněné práce [6] bylo dokázáno, že splňuje-li míra divergence (1.7) základní princip zpracování dat (1.8), potom tato míra má tvar

$$D(P,Q) = D_f(P,Q) = \sum_{i=1}^{n} q_i \, f\left(\frac{p_i}{q_i}\right) \tag{1.10}$$

pro nezápornou spojitou funkci f proměnné $u > 0$ danou vztahem

$$f(u) = \delta(u,1)\, \boldsymbol{I}(0 < u \leq 1) + u\delta(1,\, 1/u)\, \boldsymbol{I}(u > 1), \tag{1.11}$$

kde $\boldsymbol{I}(\cdot)$ je indikátorová funkce. Po dosazení

$$\delta(p,q) = qf(p/q)$$

do nerovnosti (1.8) zjistíme, že má platit

$$(q_k + q_{k+1})\, f\left(\frac{p_k + p_{k+1}}{q_k + q_{k+1}}\right) \leq q_k f\left(\frac{p_k}{q_k}\right) + q_{k+1} f\left(\frac{p_{k+1}}{q_{k+1}}\right)$$

t. j.

$$f\left(\alpha u_k + (1-\alpha)\, u_{k+1}\right) \leq \alpha f(u_k) + (1-\alpha)\, f(u_{k+1}) \tag{1.12}$$

pro parametr $\alpha = q_k/(q_k + q_{k+1})$ z intervalu $(0,1)$ a pro argumenty

$$u_i = \frac{p_i}{q_i}, \quad i \in \{k,\, k+1\}$$

z intervalu $(0,\infty)$. Protože (1.12) se požaduje pro všechna $\alpha \in (0,1)$ a u_k, $u_{k+1} > 0$, funkce f musí být konvexní na intervalu $(0,\infty)$. Dále vidíme, že pokud základní princip zpracování dat zesílíme v tom smyslu, že rovnost (1.8) má platit jen když je redukce dat statisticky postačující ve smyslu (1.9), pak funkce f ze vztahu (1.10) musí být ryze konvexní v celém oboru $u > 0$.

Definice 1.1. Nechť $f : (0, \infty) \mapsto \mathbb{R}$ je libovolná konvexní funkce s vlastností $f(1) = 0$. Potom veličinu $D_f(P, Q)$ definovanou výrazem (1.10) nazveme f-divergencí diskrétních distribucí P, Q ze vztahu (1.6).

Poznámka 1.1. Vlastnost $f(1) = 0$ znamená, že shoda $P = Q$ má za následek nulovou f-divergenci. Dále, je-li $f : (0, \infty) \mapsto \mathbb{R}$ konvexní funkce a

$$f'_+ = \lim_{h \downarrow 0} \frac{f(u+h) - f(u)}{h} \tag{1.13}$$

je její pravá derivace v bodě $u > 0$, pak

$$\tilde{f}(u) = f(u) - f(1) - f'_+(1)\,(u-1)$$

je nezáporná konvexní funkce proměnné $u > 0$ (ryze konvexní v nějakém bodě právě když $f(u)$ je ryze konvexní v tomto bodě). Je-li $f(1)$ podle předpokladu Definice 1.1 nulové, pak z této definice pro naši funkci $\tilde{f}$ plyne

$$D_{\tilde{f}}(P, Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right),$$

t. j. shoda $\tilde{f}$-divergence s f-divergencí pro všechna P, Q. Tato shoda umožňuje rozšíření definice (1.10) na všechny konvexní funkce s vlastností $f(1) = 0$.

Poznámka 1.2. O distribucích P, Q se ve vztahu předpokládalo, že jejich komponenty p_i, q_i jsou kladné. Někdy je výhodné rozšířit Definici 1.1 také na distribuce s libovolnými komponentami p_i, $q_i \geq 0$. Za tím účelem uvažujeme funkce f^* adjungované k funkcím f z Definice 1.1 ve smyslu

$$f^*(u) \equiv u f(1/u), \quad u > 0, \tag{1.14}$$

které jsou také konvexní na intervalu $(0, 1)$ s vlastností $f^*(1) = 0$. Definici 1.1 rozšiřujeme na f-divergence libovolných distribucí pravděpodobnosti $P = (p_1, \ldots, p_n)$, $Q = (q_1, \ldots, q_n)$ vztahem

$$D_f(P, Q) = \sum_{i: p_i \leq q_i} q_i f\left(\frac{p_i}{q_i}\right) + \sum_{i: p_i > q_i} p_i f^*\left(\frac{q_i}{p_i}\right), \tag{1.15}$$

kde

$$f(0) = \lim_{u \downarrow 0} f(u), \quad f^*(0) = \lim_{u \downarrow 0} f^*(u), \quad f(0) + f^*(0) \geq 0 \tag{1.16}$$

a kde klademe $0f(0/0) = 0$. Snadno se přesvědčíme, že pro distribuce P, Q s kladnými komponentami p_i, q_i se vztah (1.15) zredukuje na (1.10).

Podle Teorému 6 v práci [8] platí pro všechny konvexní funkce z Definice 1.1 resp. Poznámky 1.2

$$\sup D_f(P, Q) = D_f(p, q), \tag{1.17}$$

kde supremum se bere přes všechny konečné rozklady (1.3) intervalu $(0, 1)$ a kde pravá strana je definovaná vztahem

$$D_f(p, q) = \int_0^1 q(t)\, f\left(\frac{p(t)}{q(t)}\right) \mathrm{d}t \tag{1.18}$$

resp.

$$D_f(p, q) = \int_{p \le q} q(t)\, f\left(\frac{p(t)}{q(t)}\right) \mathrm{d}t + \int_{p > q} p(t)\, f^*\left(\frac{q(t)}{p(t)}\right) \mathrm{d}t \tag{1.19}$$

podle toho, zda $p(t) = 0$ má za následek $q(t) = 0$ či nikoliv, přičemž $f(0)$, $f^*(u)$ pro $u \ge 0$ jsou definovány stejně jako v (1.14) – (1.16).

Definice 1.2. Každá konvexní funkce $f : (0, \infty) \mapsto \mathbb{R}$ s vlastností $f(1) = 0$ definuje f-divergenci $D_f(p, q)$ spojitých distribucí p, q vztahy (1.18), (1.19) a přidruženými konvencemi.

Poznámka 1.3. Z Definice 1.2 a vztahu (1.17) vyplývá pro libovolné spojité distribuce p, q uvažované v (1.5) a jim příslušné diskrétní distribuce P, Q odpovídající rozkladům (1.3) následující nerovnost:

$$D_f(P, Q) \le D_f(p, q). \tag{1.20}$$

Podle Teorému 1 v [10], rovnost zde platí, když věrohodností poměry $p(t)/q(t)$ splňují pro všechna $1 \le k \le n$ podmínku

$$\frac{p(t)}{q(t)} = \frac{p_k}{q_k}, \quad t \in (t_{k-1}, t_k). \tag{1.21}$$

Je-li navíc f ryze konvexní na intervalu $(0, \infty)$ a $D_f(p, q) < \infty$, pak podmínka (1.21) je nejen postačující, ale i nutná pro dosažení rovnosti (1.20). Tato podmínka představuje statistickou postačitelnost rozkladu (1.3) pro spojité distribuce p, q, t. j. statistickou postačitelnost příslušných kvantovaných distribucí (veličin) P, Q.

2 Základní vlastnosti f-divergencí

Nechť $f : (0, \infty) \mapsto \mathbb{R}$ je konvexní funkce s $f(1) = 0$. Dá se ukázat (viz [9]), že je-li $f(u)$ ryze konvexní v některém bodě $u > 0$, pak platí $f(0) + f^*(0) > 0$. Ve zbývající části práce budeme předpokládat, že f je ryze konvexní v bodě 1. Tato lokalizace ryzí konvexnosti umožňuje z nulovosti f-divergence vyvodit shodu příslušných distribucí. V dalším se omezujeme na vlastnosti f-divergencí $D_f(p, q)$ z Definice 1.2, které přísluší spojitým distribucím (1.5). Divergence $D_f(P, Q)$ diskrétních distribucí příslušných rozkladům (1.3) chápeme jako numerické aproximace teoretických hodnot $D_f(p, q)$, pro které si čtenář snadno může modifikovat všechny níže uvedené vlastnosti. Pokud jde o důkazy těchto vlastností, můžeme se odkázat na publikace [2,3] nebo [9,10], resp. v diskrétním případě na Reada a Cressieho [7].

Vlastnost 2.1 (Obor hodnot). Pro konstanty $f(0)$, $f^*(0)$ ze vztahu (1.16) platí

$$0 \leq D_f(p, q) \leq f(0) + f^*(0), \tag{2.1}$$

kde levá rovnost nastane právě když $p = q$ a pravá nastane když $p(t) = 0 \iff q(t) \neq 0$ (ortogonalita, symbolicky $p \perp q$). Je-li navíc $f(0) + f^*(0) < \infty$, pak pravá rovnost platí právě když $p \perp q$.

Z hlediska f-divergence je tedy nejmenší pravděpodobností distribucí jejich shoda a nejvyšší jejich ortogonalita.

Poznámka 2.1. Při shodných distribucích (shodných veličinách p, q) neodlišíme hypotézu $\mathcal{H} : p$ od alternativy $\mathcal{A} : q$ na základě žádných pozorování. V případě ortogonality $p \perp q$ toto rozlišení nastane s pravděpodobností 1 na základě pozorování veličiny v jediném bodě $0 < t < 1$ (když $p(t) > 0$, pak s pravděpodobností 1 platí $\mathcal{H}$, zatímco v případě $q(t) > 0$ s touto pravděpodobností platí $\mathcal{A}$). Je-li $f(0) = \infty$ (resp. $f^*(0) = \infty$), pak podle (1.19) máme maximální f-divergenci $D_f(p, q) = \infty$ i bez ortogonality, totiž v případě $p(t) = 0$, $q(t) > 0$ (resp. $p(t) > 0$, $q(t) = 0$) pro všechna t z některého neprázdného intervalu $(a, b) \subset (0, 1)$. V takovém případě na základě jediného pozorování veličiny odlišíme $\mathcal{H}$ od $\mathcal{A}$ s kladnou pravděpodobností

$$\int_a^b q(t)\,\mathrm{d}t \quad \left(\text{resp. } \int_a^b p(t)\,\mathrm{d}t\right),$$

která ovšem nemusí být 1.

Vlastnost 2.2 (Symetrie). Pro funkci f^* adjungovanou k f ve smyslu (1.14) platí

$$D_f(p,q) = D_{f^*}(q,p). \tag{2.2}$$

Je zřejmé, že rovnost (2.2) ve Vlatnosti 2.2 nastane i tehdy, když $f^*(u)$ nahradíme funkcí

$$\tilde{f}(u) = f^*(u) + \text{const.}(u-1).$$

Poznámka 2.2. Nezápornost $D_f(p,q) \geq 0$ z Vlastnosti 2.1 s rovností $D_f(p,q) = 0$ právě když $p = q$ znamená reflexivnost f-divergence na prostoru $\mathcal{P}$ uvažovaných spojitých veličin $p = p(t)$ splňujících (1.1). Z Vlastnosti 2.2 vyplývá symetrie f-divergencí, jejichž f jsou samoadjungovaná v rozšířeném smyslu

$$f(u) - f^*(u) = \text{const.}(u-1), \quad u > 0. \tag{2.3}$$

Mocniny symetrických f-divergencí $D_f(p,q)$ kladného řádu $\tau > 0$ jsou tedy metrikami na prostoru $\mathcal{P}$, pokud pro všechna p, $\tilde{p} \in \mathcal{P}$ platí trojuhelníková nerovnost

$$D_f(p,q)^\tau \leq D_f(p,\tilde{p})^\tau + D_f(\tilde{p},q)^\tau. \tag{2.4}$$

Nutnou podmínkou pro tuto nerovnost je omezenost f-divergencí: $f(0) + f^*(0) < \infty$. V opačném případě totiž lze najít p, $\tilde{p}$, $q \in \mathcal{P}$ pro která

$$D_f(p,q) = \infty, \quad D_f(p,\tilde{p}) < \infty, \quad D_f(\tilde{p},q) < \infty,$$

což protiřečí nerovnosti(2.4).

Poznámka 2.3. Podmínka omezenosti f-divergencí $f(0)+f^*(0) < \infty$ je splněna právě tehdy, když je funkce $f(u)/(1+u)$ omezená na $(0,\infty)$.

3 Proč robustní f-divergence a které to jsou

Definice 3.1. Řekneme, že f-divergence $D_f(p,q)$ je robustní vzhledem k nepřesnému určení distribucí (veličin) p, q, když existuje konstanta $C > 0$ pro kterou všechna p, q, $\tilde{p}$, $\tilde{q}$ z prostoru $\mathcal{P}$ zavedeného v Poznámce 2.2 poskytují Fréchetovy derivace

$$\lim_{\varepsilon \downarrow 0} \frac{D_f\left((1-\varepsilon)\,p + \varepsilon\tilde{p},\, (1-\varepsilon)\,q + \varepsilon\tilde{q}\right) - D_f(p,q)}{\varepsilon} \tag{3.1}$$

v absolutní hodnotě menší než C.

Vlastnost 3.1 (Robustnost). Je-li $f(0) + f^*(0) < \infty$, pak je příslušná f-divergence $D_f(p, q)$ robustní vzhledem k nepřesnému určení p, q ve smyslu Definice 3.1.

Důkaz. Podle Teorému 9.27 v práci [9] je výraz $D_f(p, q)$ konvexní v proměnných $(p, q) \in \mathcal{P} \otimes \mathcal{P}$. Tudíž

$$D_f\left(91 - \varepsilon\right) p + \varepsilon\tilde{p},\, \left(1 - \varepsilon\right) q + \varepsilon\tilde{q}\right) \leq \left(1 - \varepsilon\right) D_f(p, q) + \varepsilon D_f(\tilde{p}, \tilde{q}).$$

Odsud plyne, že podmínka Definice 3.1 platí pro

$$C = 2\left(f(0) + f^*(0)\right).$$

V následující Tabulce 3.1 uvádíme některé příklady f-divergencí. Kvůli úspoře místa za integrály chybí symbol dt.

Některé metrické a robustní divergence z této tabulky patří do třídy funkcí

$$f_\alpha(u) = \frac{2}{\alpha - 1}\left[\left(\frac{u^\alpha + 1}{2}\right)^{1/\alpha} - \frac{u + 1}{2}\right] \tag{3.2}$$

s parametrem $\alpha \in \mathbb{R}$, kde při $\alpha = 1$, $\alpha = 0$ se uvažují příslušné limity

$$f_1(u) = u \log \frac{2u}{u + 1} + \log \frac{2}{u + 1}, \quad \log = \log_e \tag{3.3}$$

$$f_0(u) = \left(\sqrt{u} - 1\right)^2 \tag{3.4}$$

a kde platí

$$f_{-1}(u) = \frac{(u - 1)^2}{u + 1}. \tag{3.5}$$

Zřejmě

$$D_{f_0}(p, q) = H^2(p, q) \quad \text{a} \quad D_{f_{-1}}(p, q) = L^2(p, q) \tag{3.6}$$

je Hellingerova a Le Camova divergence z Tabulky 3.1 a

$$D_{f_1}(p, q) = I(p,\, (p + q)/2) + I(q,\, (p + q)/2) \tag{3.7}$$

je metrická (s parametrem $\tau = 1/2$) a robustní divergence, jejíž jednotlivé komponenty $I(p,\, q)$ a $I(q,\, p)$ byly také uvedeny v Tabulce 3.1.

$f(u)$	$f^*(u)$	Formule Jméno	$f(0)+f^*(0)$	Metrika, τ	Robustnost
$\lvert u-1\rvert$	$\lvert u-1$	$V(p,q)=\int \lvert p-q\rvert$ Totální variace	2	ano, 1	ano
$u\log u$	$-\log u$	$I(p,q)=\int p\log\frac{p}{q}$ Kullback	∞	ne	ne
$-\log u$	$u\log u$	$I(q,p)$ - viz výše Obrácený Kullback	∞	ne	ne
$u\log\frac{2u}{u+1}$	$\log\frac{2}{u+1}$	$I(p,(p+q)/2)$	$\log 2$	ne	ano
$\log\frac{2}{u+1}$	$u\log\frac{2z}{u+1}$	$I(q,(p+q)/2)$	$\log 2$	ne	ano
$(u-1)^2$	$\frac{(u-1)^2}{u}$	$\chi^2(p,q)=\int\frac{(p-q)^2}{q}$ Pearson	∞	ne	ne
$\frac{(u-1)^2}{u}$	$(u-1)^2$	$\chi^2(q,p)$ - viz výše Neyman	∞	ne	ne
$\frac{(u-1)^2}{u+1}$	$\frac{(u-1)^2}{u+1}$	$L^2(p,q)=\int\frac{(p-q)^2}{p+q}$ Le Cam	2	ano, 1/2	ano
$\left(\sqrt{u}-1\right)^2$	$\left(\sqrt{u}-1\right)^2$	$H^2(p,q)=\int\left(\sqrt{p}-\sqrt{q}\right)^2$ Hellinger	2	ano, 1/2	ano

Tabulka 3.1.

Třída konvexních funkcí $2f_\alpha(u)$ ze vztahů (3.5) – (3.7) je z práce Kůs a spol. [4] a podtřídu

$$\frac{\alpha f_\alpha(u)}{2^{(\alpha-1)/2}}$$

s parametry $\alpha>0$ zavedli již předtím Österreicher a Vajda [5].

V následujících příkladech ilustrujeme rozdíl mezi nerobustními a robustními f-divergencemi.

Příklad 3.1. Máme-li dvě shodné pravděpodobnostní hustoty nebo veličiny

$$p(t)=q(t)=2\boldsymbol{I}(1/4<t<3/4) \tag{3.8}$$

potom podle definice Kullbackovy divergence s $f(t) = t \log t$

$$I(p,q) = \int_{1/2}^{3/4} 2 \log \frac{2}{2} \, \mathrm{d}t = 0 \quad \text{(viz (1.19))}. \tag{3.9}$$

Tento výsledek souhlasí s Vlastností 2.1, podle které všechny f-divergence detekují úplnou shodu p, q tím, že dosahují minimální možnou hodnotu

$$D_f(p,q) = 0. \tag{3.10}$$

Nyní uvažujme pro malé hodnoty $0 < \delta < 1/4$ hustoty (veličiny)

$$\tilde{p}(t) = p(t-\delta), \quad \tilde{q}(t) = p(t+\delta). \tag{3.11}$$

(i) Jsou-li $p = q$ pravděpodobností hustoty náhodné veličiny X, pak $\tilde{p}$, $\tilde{q}$ jsou hustoty náhodných veličin $X \pm \delta$ odpovídajících nepřímému pozorování veličiny X, které je zatížené systematickou aditivní chybou $\pm\delta$. Pravděpodobnosti přímého a nepřímého pozorování $1-\varepsilon$ a ε pak povedou ke dvěma různým náhodným veličinám X_ε a Y_ε s hodnotami pravděpodobnosti

$$\begin{aligned} p_\varepsilon(t) &= (1-\varepsilon)\, p(t) + \varepsilon \tilde{p}(t) \\ &= 2\left[(1-\varepsilon)\, \boldsymbol{I}(1/4 < t \le 1/4+\delta) + \boldsymbol{I}(1/4+\delta < t \le 3/4) + \varepsilon \boldsymbol{I}(3/4 < t < 3/4+\delta)\right] \end{aligned}$$

resp.

$$\begin{aligned} q_\varepsilon(t) &= (1-\varepsilon)\, q(t) + \varepsilon \tilde{q}(t) \\ &= 2\left[\varepsilon\, \boldsymbol{I}(1/4-\delta < t \le 1/4) + \boldsymbol{I}(1/4 < t \le 3/4-\delta) + (1-\varepsilon)\, \boldsymbol{I}(3/4-\delta, 3/4)\right]. \end{aligned}$$

(ii) Jsou-li naproti tom $p = q$ dvě shodné veličiny pozorované v čase (na trase) $0 < t < 1$, pak vzájemně různé veličiny $\tilde{p}, \tilde{q}$ ze vztahu (3.11) odpovídají nepřesné synchronizaci měření posunutého v čase (na trase) o konstantu $\pm\delta$. Veličiny $p_\varepsilon(t)$, $q_\varepsilon(t)$ z předchozích dvou formulí pak budou průměry z měření u kterých tato porucha synchronizace nastala s četností $0 < \varepsilon < 1$.

Nyní se budeme zajímat o Kullbackovu divergenci $I(p_\varepsilon, q_\varepsilon)$ hustot resp. veličin, ke kterým jsme zde dospěli. Podle (1.19) platí

$$I(p_\varepsilon, q_\varepsilon) = \int_{(0,3/4-\delta)\cup(3/4+\delta,1)} q_\varepsilon f\left(\frac{p_\varepsilon}{q_\varepsilon}\right) \mathrm{d}t + \int_{(3/4-\delta,3/4+\delta)} p_\varepsilon f^*\left(\frac{q_\varepsilon}{p_\varepsilon}\right) \mathrm{d}t \tag{3.12}$$

kde $f^*(t) = -\log t$ podle Tabulky 3.1. Protože však na sjednocení intervalů $(0, 1/4-\delta] \cup (3/4+\delta, 1)$ platí $p_\varepsilon = q_\varepsilon = 0$ a na intervalu $(1/4+\delta,\, 3/4-\delta)$ je $p_\varepsilon + q_\varepsilon = 1$, stačí v

(3.12) integrovat na intervalech $(1/4-\delta,\, 1/4+\delta)$ a $(3/4-\delta,\, 3/4+\delta)$. Proto po dosazení z definice p_ε a q_ε dostaneme

$$\begin{aligned} I(p_\varepsilon, q_\varepsilon) &= \int_{1_4-\delta}^{1/4} 2\varepsilon\, f(0)\,\mathrm{d}t + \int_{1/4}^{1/4+\delta} 2f\left(\frac{1-\varepsilon}{1}\right)\mathrm{d}t \\ &\qquad + \int_{3/4-\delta}^{3/4} 2f^*\left(\frac{1-\varepsilon}{1}\right)\mathrm{d}t + \int_{3/4}^{3/4+\delta} 2\varepsilon\, f^*(0)\,\mathrm{d}t \\ &= 2\delta\left[\varepsilon\cdot 0 + f(1-\varepsilon) + f^*(1-\varepsilon) + \varepsilon\cdot\infty\right] \\ &= 2\delta\varepsilon\left[\log\frac{1}{1-\varepsilon} + \infty\right] = \infty \end{aligned}$$

nezávisle na hodnotě chyby $0 < \delta < 1/4$ a pravděpodobnosti $0 < \varepsilon < 1$. Proto také Fréchetova derivace (3.1) je pro Kullbackovu divergenci nekonečná, t. j.

$$\lim_{\varepsilon\downarrow 0} \frac{I(p_\varepsilon, q_\varepsilon) - I(p,q)}{\varepsilon} = \infty. \tag{3.13}$$

Tento výsledek znamená, že tato divergence je hodně nerobustní vzhledem k nepřesnému určení distribucí (veličin) p, q. Jinými slovy, při libovolně malé chybě určení $\delta > 0$ dostaneme místo nulové správné hodnoty (3.9) nekonečně velkou nesprávnou hodnotu (3.13) a to i tehdy, když k tomuto nepatrně nepřesnému určení dojde s nepatrně malou pravděpodobností nebo četností $\varepsilon > 0$. Tímto se osvětluje, co v Tabulce 2.1 znamená *ne* ve sloupci "Robustnost" u Kullbackovy divergence $f(u) = u\log u$ a také dalších tam uvedených divergencí.

Příklad 3.2. Nyní v Příkladě 3.1 místo nerobustní Kullbackovy divergence použijeme Hellingerovu divergenci odpovídající funkci $f(u) = (\sqrt{u}-1)^2$ resp. $f(u) = 2(1-\sqrt{u})$, která je podle Tabulky 3.1 robustní. Podíváme se, jak se Hellingerova divergence $H^2(p_\varepsilon, q_\varepsilon)$ odchýlí od správné hodnoty

$$H^2(p,q) = 0 \qquad\qquad (\text{viz } (3.10)) \tag{3.14}$$

když jak nepřesnost $\delta > 0$ tak i pravděpodobnost $\varepsilon > 0$ se kterou k nepřesnosti dojde budou malé. Podobně jako v (3.12), ze vztahu (1.19) a formulí pro p_ε, q_ε, a $f = f^*$

obdržíme

$$\begin{aligned} H^2(p_\varepsilon, q_\varepsilon) &= \int_{1/4-\delta}^{1/4+\delta} q_\varepsilon \, f\left(\frac{p_\varepsilon}{q_\varepsilon}\right) \mathrm{d}t + \int_{3/4-\delta}^{3/4+\delta} p_\varepsilon \, f^*\left(\frac{q_\varepsilon}{p_\varepsilon}\right) \mathrm{d}t \\ &= 2\left[\int_{1/4-\delta}^{1/4+\delta} (q_\varepsilon - \sqrt{p_\varepsilon q_\varepsilon}) \, \mathrm{d}t + \int_{3/4-\delta}^{3/4+\delta} (p_\varepsilon - \sqrt{p_\varepsilon q_\varepsilon}) \, \mathrm{d}t\right] \\ &= 8\delta\left[\varepsilon + 1 - \sqrt{1-\varepsilon}\right] \\ &= 8\delta\varepsilon \, \frac{2+\sqrt{1-\varepsilon}}{1+\sqrt{1-\varepsilon}} \leq 16\delta\varepsilon. \end{aligned}$$

Proto je Fréchetova derivace (3.2) v tomto případě konečná,

$$\lim_{\varepsilon \downarrow 0} \frac{H^2(p_\varepsilon, q_\varepsilon) - H^2(p, q)}{\varepsilon} = 16\delta,$$

a klesající k nule při $\delta \downarrow 0$. Dále, odchylka Hellingerovy divergence $H^2(p_\varepsilon, q_\varepsilon)$ od správné hodnoty $H^2(p, q) = 0$ je menší než $16\delta\varepsilon$, t. j. klesá k nule jak v případě, že nepřesnost δ klesá k nule, tak i v případě, že pravděpodobnost ε výskytu takové nepřesnosti klesá k nule.

4 Proč f-fluktuace a co to je

Přes veškerou svoji výjimečnost a dobré vlastnosti, f-divergence někdy nepostihuje žádoucí rozdíly mezi omezenými veličinami nebo distribucemi. Například signály $p(t)$, $q(t)$ mohou být po celý čas $0 < t < 1$ nenulové a vzájemně velmi blízké, ale jeden u nich může přitom s malými amplitudami rychle kmitat kolem pomalu se měnící střední hodnoty dané druhým z nich. Divergence bude reagovat na amplitudu tohoto kmitání, ale už nemusí být schopna zachytit jeho frekvenci, na které nám může v některých situacích především záležet.

Může nám například velmi záležet na zrnitosti plochy, která je při zběžném pohledu stejně šedá, jako vedlejší zcela hladká plocha (viz Filip a Havran [1]). Nebo může být velmi důležité, zda je ve spojitém spektru převážně bílého šumu zastoupena slabá diskrétní komponenta hřebenového typu, která se projeví slabým zvlněním převážně konstantní spektrální hustoty. Pro fotografa může být zajímavé rozlišit situace, kdy je hladina vody jen v pomalém pohybu v důsledku ustáleného stojatého vlnění, a kdy je navíc zčeřena slabým vánkem. Tyto a další podobné situace budeme ilustrovat jedním konkrétním příkladem.

Příklad 4.1. Nechť distribuce $q(t) = 1$ je ustálená na celé jednotce prostoru nebo času $0 < t < 1$ a uvažujme dále ne této jednotce soustavu distribucí

$$p_k(t) = 1 + a \sin 2\pi k t$$

které kolem $q(t)$ kmitají s amplitudou $0 < a < 1$ a frekvencemi $k = 0, 1, 2, \ldots$. Použijeme-li k rozlišení těchto distribucí klasickou Pearsonovu divergenci $\chi^2(p, q)$ z Tabulky 3.1, pak pro $k = 0$ obdržíme $\chi^2(p_0, q) = 0$ a pro ostatní k

$$\begin{aligned}
\chi^2(p_k, q) &= \int_0^1 \frac{(p_k(t) - q(t))^2}{q(t)} \,\mathrm{d}t \\
&= \int_0^1 (p_k(1) - 1)^2 \,\mathrm{d}t \\
&= a^2 \int_0^1 \sin^2 2\pi k t \,\mathrm{d}t \\
&= \frac{a^2}{2\pi k} \int_0^{2\pi k} \sin^2 t \,\mathrm{d}t \\
&= \frac{a^2}{2\pi} \int_0^{2\pi} \sin^2 t \,\mathrm{d}t.
\end{aligned}$$

Tudíž

$$\chi^2(p_k, q) = \frac{a^2}{2} \quad \text{pro všechna } k = 1, 2, \ldots \tag{4.1}$$

t. j. Pearsonova divergence není schopna rozlišovat mezi nenulovými frekvencemi.

Protože pro všechna $k > 0$ v celém oboru $0 < t < 1$ platí

$$1 - a \le \frac{p_k(t)}{q(t)} \le 1 + a, \tag{4.2}$$

výsledek Příkladu 4.1 můžeme ještě zesílit. Totiž, pro všechny f-divergence dostaneme ze (4.2) meze

$$0 \le D_f(p_k, q) \le \max\left\{ f(1 + a) + f'_+(1)\, a,\ (1 - a) - f'_+(1)\, a \right\} \tag{4.3}$$

platné pro všechna $k = 1, 2, \ldots$ Zde $f'_+(1)$ je pravá derivace funkce $f(u)$ v bodě $u = 1$ (viz (1.13)), přičemž celý výsledek (4.3) vyplývá z toho, že $f(u) + f'_+(1)\,(u - 1)$ je nerostoucí v oblasti $u \le 1$ a neklesající v oblasti $u > 1$. Žádná f-divergence tudíž nemůže rozpoznat v modelu z Příkladu 4.1 například to, že frekvence k roste nade všechny meze.

Přirozenou mírou fluktuace distribuce (veličiny) p kolem q je norma derivace věrohodnostního poměru

$$\rho(t) = \frac{p(t)}{q(t)} \quad \left(\text{kde } \frac{0}{0} = 0 \right), \tag{4.4}$$

například L_1-norma

$$\|\rho'\| = \int_0^1 |\rho'(t)|\,\mathrm{d}t. \tag{4.5}$$

Abychom předešli matematickým nejasnostem, zde i v dalším se omezíme na situace, kdy věrohodnostní poměr $\rho(t)$ je vztahem (4.4) dobře definovaný v celém oboru $(0,1)$ s výjimkou konečně mnoha hodnot $t \in (0,1)$ a v tomto oboru je taktéž po částech spojitě diferencovatelný. Protože připouštíme, že derivace

$$\rho'(t) = \frac{\mathrm{d}\rho(t)}{\mathrm{d}t} \tag{4.6}$$

nebude pro některá $t \in (0,1)$ definována, budeme integrál (4.5) chápat ve smyslu Lebesgueově, kde stačí, aby tato derivace byla definována skoro všude na $(0,1)$.

Výhodou normy $\|\rho'\|$ jakožto míry fluktuace funkce $p(t)$ kolem $q(t)$ je její nezápornost s tím, že

$$\|\rho'\| = 0 \quad \Leftrightarrow \quad \rho' \equiv 0, \tag{4.7}$$

t. j. míra fluktuace je nulová tehdy a jen tehdy, když věrohodnostní poměr $p(t)/q(t)$ je na intervalu $(0,1)$ po částech konstantní. Konstantní poměr $p(t)/q(t)$ vystihuje totiž tu situaci, kdy $p(t)$ nevykazuje žádné kmitání kolem $q(t)$ a ke změně poměru mezi $p(t)$ a $q(t)$ dojde při nejvýše konečně mnoha přechodech z jednoho intervalu do druhého. V extrémním případě, kdy je počet těchto přechodů nulový platí pro všechna $t \in (0,1)$ pravá úměra $p(t) = \mathrm{const.}\, q(t)$, t. j.

$$p = q \quad \text{na } (0,1). \tag{4.8}$$

Nevýhodou normy $\|\rho'\|$ jakožto míry fluktuace funkce $p(t)$ kolem $q(t)$ je fakt, že všechny hodnoty $|\rho'(t)| \in [0,\infty)$ se do hodnoty integrálu (4.5) promítají se stejnou váhou. Intuice napovídá, že velké odchylky derivace $|\rho'(t)|$ od jejího průměru $\|\rho'\|$ by se měly do celkové míry fluktuace promítat s větší vahou než odchylky malé. Toto uspořádání dosáhneme, když za míru fluktuace vezmeme f-divergenci pravděpodobnostních hustot

$$\tilde{p}(t) = \frac{|\rho'(t)|}{\|\rho'\|} \quad \text{a} \quad \tilde{q}(t) = 1 \tag{4.9}$$

na intervalu $(0,1)$, t. j. veličinu

$$D_f(\tilde{p},\tilde{q}) = \int_0^1 \|\rho'\|\, f\left(\frac{|\rho'(t)|}{\|\rho'\|}\right) \mathrm{d}t \tag{4.10}$$

(viz definici f-divergence v (1.2)). V literatuře se totiž dokazuje (viz např. Appendix v práci [2]), že pro funkce f z Definice 1.2 poměr $f(u)/u$ v oblasti $u \geq 1$ s rostoucím u

roste a v oblasti $0 < u \leq 1$ klesá , t. j. čím dále je $|\rho'(t)|$ od průměru $\|\rho'\|$, tím větší je vzájemný poměr

$$f\left(\frac{|\rho'(t)|}{\|\rho'\|}\right) \Big/ \frac{|\rho'(t)|}{\|\rho'\|}.$$

Matematická nekorektnost spočívající v tom, že v definici $\tilde{p}$ v (4.9) připouštíme i nulový jmenovatel $\|\rho'\| = 0$ se řeší konvencí $0f(0/0) = 0$, která byla součástí zmíněné Definice 1.2. Je-li totiž $\|\rho'\| = 0$, pak podle předpokladů platí $\rho'(t) = 0$ skoro jistě na celém intervalu $(0, 1)$ a tudíž zcela korektním způsobem dostaneme implikaci

$$\|\rho'\| = 0 \quad \Rightarrow \quad D_f(\tilde{p}, \tilde{q}) = 0. \tag{4.11}$$

Na druhé straně, podle Vlastnosti 2.1 platí pro f-divergenci hustot (4.9) ekvivalence

$$D_f(\tilde{p}, \tilde{q}) = 0 \quad \Leftrightarrow \quad |\rho'| = \text{const.}, \tag{4.12}$$

t. j. f-divergence je nulová tehdy a jen tehdy, když absolutní derivace $|\rho'(t)|$ je na intervalu $(0, 1)$ konstantní s výjimkou konečně mnoha $t \in (0, 1)$ (tam podle předpokladů derivace nemusí existovat). To znamená, že věrohodnostní poměr $p(t)/\rho(t)$ je na $(0, 1)$ pila s konstantním sklonem všech (možná nestejně vysokých) zubů. Tento sklon nulový v případě (4.7) a jenom v tomto případě, kdy se věrohodnostní poměr $p(t)/q(t)$ stane na intervalu $(0, 1)$ po částech konstantní.

Příklad 4.2. Je-li

$$p(t) = 1 + a\left[(ut - 1)\,\boldsymbol{I}(0 < t \leq 1/2) - (4t - 3)\,\boldsymbol{I}(1/2 < t < 1)\right]$$

a $q(t) = \boldsymbol{I}(0 < t < 1)$, pak věrohodnostní poměr (4.4) je pila $\rho|t| = p(t)$ s jedním zubem, který má na obě strany konstantní sklon $|\rho'| = |p'(t)| = 4a$. Pila jako taková (t.j. její zub) vymizí právě když $a = 0$, t. j. když p splyne s q.

Podle výše řečeného je míra fluktuace nulová tehdy a jen tehdy, když const. v (4.8) je nula. To ovšem poukazuje na skutečnost, že $D_f(\tilde{p}, \tilde{q})$ sama o sobě nestačí k žádoucí charakteristice fluktuace $p(t)$ kolem $q(t)$. Současně z výše řečeného vyplývá, že takovou žádoucí charakterizaci nabízí míra uvedená v následující definici, t. j. že pro tuto míru platí věta, kterou uvádíme hned za touto definicí.

Definice 4.1. Platí-li pro $p(t)$, $q(t)$ a jejich věrohodnostní poměr $\rho(t)$ předpoklady uvedené na začátku této sekce, pak f-fluktuace $p(t)$ kolem $q(t)$ je definovaná vztahem

$$\Delta_f(p,q) = \|\rho'\| \left(\int_0^1 f\left(\frac{|\rho'(t)|}{\|\rho'\|} \right) \mathrm{d}t + 1 \right) \tag{4.13}$$

kde za integrálem klademe $f(0/0) = 0$.

Věta 4.1. Pro všechny funkce p, q z Definice 4.1 je jejich f-fluktuace nezáporná,

$$\Delta_f(p,q) \geq 0, \tag{4.14}$$

přičemž nulovost nastane právě když jr věrohodnostní poměr $p(t)/q(t)$ po částech konstantní.

Důkaz. Protože $\Delta_f(p,q)$ je součet L_1-normy (4.5) a f-divergence (4.8), její nezápornost plyne z nezápornosti jak normy tak i f-divergence. Rovnost $\Delta_f(p,q) = 0$ nastane právě když

$$D_f(\tilde{p},\tilde{q}) = 0 \quad \text{a} \quad \|\rho'\| = 0.$$

První rovnost je podle (4.8) ekvivalentní po částech konstantnosti funkce $|\rho'(t)|$ a druhá je podle (4.7) ekvivalentní po částech nulovosti funkce $|\rho'(t)|$, t. j. po částech konstantnosti věrohodnostního poměru $p(t)/q(t)$. Protože z druhé podmínky plyne podmínka první, je druhá podmínka ekvivalentní rovnosti $\Delta_f(p,q) = 0$, což bylo dokázat.

Nyní ukážeme, že f-fluktuace (4.11) postihuje jemné ale přitom významné rozdíly mezi p a q z Příkladu 4.1, které nastanou, když amplituda a je malá a frekvence k je velká. Viděli jsme, že Pearsonova divergence není schopna tyto rozdíly zaznamenat žádoucím způsobem.

Příklad 4.3. V modelu z Příkladu 4.1 máme pro všechna $k = 0, 1, 2, \ldots$ věrohodnostní poměry

$$\rho_k(t) = p_k(t), \quad \rho_k'(t) = p_k'(t) = 2\pi k a \cos 2\pi k t$$

a tedy L_1-normy

$$\begin{aligned} \|\rho_k'\| &= 2\pi a k \int_0^1 |\cos 2\pi k t| \, \mathrm{d}t \\ &= a \int_0^{2\pi k} |\cos t| \, \mathrm{d}t \\ &= 4ak. \end{aligned}$$

Dále

$$\frac{|\rho_k'(t)|}{\|\rho_k'\|} = \frac{\pi}{2}\,|\cos 2\pi k t|,$$

takže pro Pearsonovu funkci $f(u) = u^2 - 1$, která je podle Poznámky 1.1 ekvivalentní funkci $(u-1)^2$, dostaneme

$$\begin{aligned}\int_0^1 f\left(\frac{|\rho_k'(t)|}{\|\rho_k'\|}\right)\mathrm{d}t + 1 &= \frac{\pi^2}{4}\int_0^1 \cos^2 2\pi k t\,\mathrm{d}t \\ &= \frac{\pi^2}{4}\,\frac{1}{2\pi k}\int_0^{2\pi k}\cos^2 t\,\mathrm{d}t \\ &= \frac{\pi^2}{4}\,\frac{k\pi}{2\pi k} = \frac{\pi^2}{8}.\end{aligned}$$

Lze tedy výpočet uzavřít formulí

$$\Delta_f(p_k, q) = 4ak\,\frac{\pi^2}{8} = \frac{\pi^2 ak}{2} \tag{4.15}$$

pro příslušnou Pearsonovskou f-fluktuaci. Pearsonovská míra fluktuace funkcí p_k a q tudíž roste lineárně s frekvencí $k = 0, 1, 2, \ldots$.

Poznámka 4.1. Z Věty 4.1 je vidět, že míra fluktuace $\Delta_f(p, q)$ je reflexivní jen v případě, kdy věrohodnostní poměr $\rho(t)$ je diferencovatelný ve všech bodech $t \in (0, 1)$. Jen v tomto případě totiž $|\rho'(t)| = \text{const.}$ znamená, že $|\rho'(t)| = 0$ všude na $(0, 1)$, což je za uvedeného předpokladu možné jen když $\rho(t) = 1$ (t. j. $p = q$) všude na $(0, 1)$. O tom, zda $\Delta_f(p, q) = 0$ znamená totožnost $p = q$ anebo jen nulovou fluktuaci, se musíme přesvědčit tak, že vezmeme v potaz rovněž hodnotu f-divergence $D_f(p, q)$: totožnost $p = q$ nastane tehdy a jen tehdy, když $D_f(p, q) = 0$.

Poznámka 4.2. Omezme se na p, q se všude diferencovatelným a nenulovým věrohodnostním poměrem ρ. Zajímat nás budou míry fluktuace $\tilde{\Delta}(p, q)$, které vedle reflexivnosti z předchozí poznámky budou též symetrické ve smyslu

$$\tilde{\Delta}(p, q) = \tilde{\Delta}(q, p) \tag{4.16}$$

a budou splňovat i trojúhelníkovou nerovnost

$$\tilde{\Delta}(p, q) \le \Delta(p, \tilde{q}) + \Delta(\tilde{q}, q), \tag{4.17}$$

t. j. budou metriky na prostoru příslušných p, q. Takovou mírou zřejmě bude

$$\tilde{\Delta}(p, q) = \int_0^1 \left|\frac{\rho'(t)}{\rho(t)}\right|\mathrm{d}t = \left\|\frac{p'}{p} - \frac{q'}{q}\right\|, \tag{4.18}$$

kde

$$\left\| \frac{p'}{p} - \frac{q'}{q} \right\| = \int_0^1 \left| \frac{p'}{p} - \frac{q'}{q} \right| \mathrm{d}t$$

je příslušná L_1-norma rozdílu derivací logaritmických funkcí

$$\log p(t) - \log q(t) = \log \rho(t).$$

Toto nás přivádí k celé třídě metrických ϕ-fluktuací

$$\tilde{\Delta}_\phi(p,q) = \int_0^1 |(\phi(p))' - (\phi(q))'| \, \mathrm{d}t = \|(\phi(p))' - (\phi(q))'\|$$

příslušných monotónním diferencovatelným funkcím $\phi : (0, \infty] \to \mathbb{R}$. Výzkum takovýchto metrických měr fluktuace si zaslouží systematickou pozornost.

Poděkování Tato práce je součástí výzkumu podporovaného granty MŠMT 1M0572 a GAČR 102/07/1311.

Reference

[1] J. Filip and V. Havran, *Bidirectional texture function compression based on multi-level quantization*, Rukopis, 2007.

[2] F. Liese and I. Vajda, *Convex Statistical Distances*, Leipzig: Teubner, 1987.

[3] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, *IEEE Trans. on Inform. Theory*, vol. 52, No. 10, pp. 4394–4412, 2006.

[4] V. Kůs, D. Morales a I. Vajda, Extensions of the parametric families of divergences used in statistical inference, *Kybernetika*, vol. 43, in print, 2007.

[5] F. Österreicher and I. Vajda, A new class of metric divergences on probability spaces and its applicability in statistics, *Ann. Inst. Statist. Math.*, vol. 55, No. 3, pp. 639–653, 2003.

[6] M. C. Pardo and I. Vajda, About distances of discrete distributions satisfying the data processing theorem of information theory, *IEEE Trans. on Inform. Theory*, vol. 43, pp. 1288–1293, 1997.

[7] M. R. C. Read and N. A. C. Cressie, *Goodness-of-Fit Statistics for Discrete Multivariate Data*,

[8] I. Vajda, On the f-divergence and singularity of probability measures, *Periodica Math. Hungar.*, vol. 2, pp. 223-234, 1972.

[9] I. Vajda, *Theory of Statistical Inference and Information*, Boston: Kluwer, 1989.

[10] I. Vajda, *Divergence pravděpodobnostních distribucí a statistická informace*, Res. Report 2168, Inst. of Inform. Theory, Prague 2006.