BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **12**, No 2

Sofia • 2012

Some New Information Inequalities Involving *f*-Divergences

Amit Srivastava

Department of Mathematics, Jaypee Institute of Information Technology, NOIDA (Uttar Pradesh), INDIA
Email: raj_amit377@yahoo.co.in

Abstract: *New information inequalities involving f-divergences have been established using the convexity arguments and some well known inequalities such as the Jensen inequality and the Arithmetic-Geometric Mean (AGM) inequality. Some particular cases have also been discussed.*

Keywords: *f-divergence, convexity, parameterization, arithmetic mean, geometric mean.*

2000 Mathematics Subject Classification: 94A17, 26D15.

1. The concept of *f*-divergences

Let $\mathbb{F}$ be the set of convex functions $f:[0,\infty)\to(-\infty,\infty)$ which are finite on $(0,\infty)$ and continuous at point $0 (f(0)=\lim_{u\downarrow 0} f(u), \mathbb{F}_0=\{f\in\mathbb{F}: f(1)=0\}$. Further if $f\in\mathbb{F}$, then f^* is defined by

$$f^*(u)=\begin{cases} u\,f\left(\dfrac{1}{u}\right) & \text{for} \quad u\in(0,\infty), \\ \lim\limits_{v\to\infty}\dfrac{f(v)}{v} & \text{for} \quad u=0, \end{cases}$$

is also in $\mathbb{F}$ and is called the **-conjugate* (convex) function of *f*.

Definition 1.1. Let

(1.1) $\Delta_n=\{(p_1,p_2,\dots,p_n): p_i\ge 0, i=1,2,\dots,n, \sum_{i=1}^n p_i=1\}, n=2,3,\dots$

denote the set of all finite discrete (*n*-ray) complete probability distributions. For a convex function $f\in\mathbb{F}$, the *f-divergence* of the probability distributions P and Q is given by

(1.2) $$I_f(P,Q) = \sum_{i=1}^{n} q_i\, f\left(\frac{p_i}{q_i}\right)$$
where $P = (p_1, p_2, \dots, p_n) \in \Delta_n$ and $Q = (q_1, q_2, \dots, q_n) \in \Delta_n$. In Δ_n, we have taken all $p_i > 0$. If we take all $p_i \geq 0$ for $i = 1, 2, \dots, n$ then we have to suppose that $0 \ln 0 = 0 \ln\left(\frac{0}{0}\right) = 0$. It is generally common to take logarithms with base of 2, but here we have taken only natural logarithms.

These divergences were introduced and studied independently by Csiszár [5, 6] and Ali and Silvey [1] and are sometimes known as Csiszár *f-divergences* or Ali-Silvey distances. The *f-divergence* given by (1.1) is a versatile functional form, which with a suitable choice of the function involved, leads to some well known divergence measures. Some examples as $f(u) = u \ln u (f^*(u) = -\ln u)$ provide the Kullback-Leibler's measure [13], $f(u) = |u-1| = f^*(u)$ results in the variational distance [12], $f(u) = (u-1)^2$ $\left(f^*(u) = \frac{(u-1)^2}{u}\right)$ yields the χ^2 divergence [15] and many more. These measures have been applied in a variety of fields, such as economics and political science [18, 19], biology [16], the analysis of contingency tables [7] , approximation of probability distributions [4, 11] , signal processing [9, 10] and pattern recognition [2, 3, 8]. The *f-divergence* satisfies a large number of properties which are important from an information theoretic point of view. Österreicher [14] has discussed the basic general properties of *f*-divergences including their axiomatic properties and some important classes.

The *f-divergence* defined by (1.2) is generally asymmetric in P and Q. Nevertheless, the convexity of $f(u)$ implies that of
$$f^*(u) = u\, f\left(\frac{1}{u}\right)$$
and with this function we have
$$I_f(P,Q) = I_{f^*}(P,Q).$$
Hence, it follows, in particular, that the symmetrised *f-divergence*
$$I_f(P,Q) + I_f(Q,P)$$
is again an *f-divergence*, with respect to the convex function $f(u) + f^*(u)$.

In the present work, we have established new information inequalities involving *f-divergences* using the convexity arguments and some well known inequalities, such as the jensen inequality and the Arithmetic-Geometric Mean (AGM) inequality. Further we have used these inequalities in establishing relationships among some well-known divergence measures. Without essential loss of insight, we restrict ourselves to discrete probability distributions and note that the extension to the general case relies strongly on the Lebesgue–Radon–Nikodym Theorem.

2. Information inequalities

Result 2.1. If $\varphi: (0,\infty) \to \mathbb{R}$ is convex, then the function
$$\psi_1(u,v) = v\varphi\left(\frac{u+v}{2v}\right)$$
of two variables is convex on the domain $(u,v) \in (0,\infty)^2$.

P r o o f. Consider $\lambda \in (0,1)$ and two points $x_i = (u_i, v_i)$ from the domain of the function φ. For

$$w = \frac{\lambda v_1}{\lambda v_1 + (1-\lambda)v_2} \quad \text{and} \quad t_i = \frac{u_i + v_i}{2v_i}$$

we get $\varphi(wt_1 + (1-w)t_2) \le w\,\varphi(t_1) + (1-w)\varphi(t_2)$,
so that

$$(\lambda v_1 + (1-\lambda)v_2)\varphi\left(\frac{\lambda u_1 + (1-\lambda)u_2 + \lambda v_1 + (1-\lambda)v_2}{2(\lambda v_1 + (1-\lambda)v_2)}\right) \le$$
$$\le \lambda v_1 \varphi\left(\frac{u_1+v_1}{2v_1}\right) + (1-\lambda)v_2\varphi\left(\frac{u_2+v_2}{2v_2}\right)$$

or, equivalently $\psi(\lambda x_1 + (1-\lambda)x_2) \le \lambda\,\psi(x_1) + (1-\lambda)\varphi(x_2)$ which completes the proof.

Result 2.2. If $\varphi\colon (0,\infty) \to \mathbb{R}$ is convex then the function

$$\psi_n(u,v) = v\varphi\left(\frac{u+nv}{(n+1)v}\right), n > 0,$$

of two variables is convex on the domain $(u,v) \in (0,\infty)^2$.

P r o o f. The proof follows on similar lines as in the previous result except the choice of t_i which can be taken as

$$t_i = \frac{u_i + nv_i}{(n+1)v_i}, \ \ n > 0.$$

We, therefore have the following divergence functionals of *f-divergence* type:

(2.1) $$I_{f(n)}(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i + nq_i}{(1+n)q_i}\right)$$

where $P = (p_1, p_2, \dots, p_n) \in \Delta_n$ and $Q = (q_1, q_2, \dots, q_n) \in \Delta_n$. For $n = 0$, the function (2.1) is reduced to the Csiszár *f-divergence* given by (1.2). Replacing *n* by 1/*n* in (2.1), we obtain

(2.2) $$I_{f\left(\frac{1}{n}\right)}(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{np_i + q_i}{(1+n)q_i}\right).$$

Relationship with Csiszár *f-divergence* follow.

Result 2.3. Let $f\colon I \subseteq [0,\infty) \to \mathbb{R}$ be a differentiable convex function on the interval I, $x_i \in \tilde{I}$ ($\tilde{I}$ is interior of *I*). Further we assume that $f(1) = 0$. Then for all $P, Q \in \Delta_n$ we have

(2.3) $$I_{f(2n+1)}(P,Q) \le \frac{1}{2}\, I_{f(n)}(P,Q),$$

(2.4) $$I_{f\left(\frac{1}{2n+1}\right)}(P,Q) \le \frac{1}{2}\left(I_{f\left(\frac{1}{n}\right)}(P,Q) + I_f(P,Q)\right)$$

where $I_f(P,Q)$ and $I_{f(n)}(P,Q)$ are measures given by (1.2) and (2.1) respectively. The equality holds in the above inequalities if $p_i = q_i$ for each *i*.

P r o o f. Let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \Delta_n$. Then it is well known that

(2.5) $$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \le \sum_{i=1}^{n} \lambda_i f(x_i).$$

If *f* is strictly convex, then the equality holds if and only if all $x_1 = x_2 = \dots = x_n$.

The above inequality is famous as *Jensen inequality*. If f is a concave function, then the inequality sign will change. If we assume $\lambda_1 = \lambda_2 = \frac{1}{2}$ with all other $\lambda_i^{,}s$ zero, then we obtain

(2.6) $$f\left(\frac{x_1+x_2}{2}\right) \leq \frac{1}{2}[f(x_1) + f(x_2)].$$

Choosing $x_1 = x$ and $x_2 = 1$, we obtain

(2.7) $$f\left(\frac{x+1}{2}\right) \leq \frac{1}{2}[f(x)] \text{ since } f(1) = 0.$$

Substituting $x = \frac{p_i}{q_i}$ in the above inequality, multiplying by q_i and then summing over all *i*, we obtain

(2.8) $$2\, I_{f(1)}(P,Q) \leq I_f(P,Q).$$

A choice of $x_1 = \frac{x+1}{2}$ and $x_2 = 1$will give $2I_{f(3)}(P,Q) \leq I_{f(1)}(P,Q)$.

A choice of $x_1 = \frac{x+3}{4}$ and $x_2 = 1$ will give $2I_{f(7)}(P,Q) \leq I_{f(3)}(P,Q)$.

Finally a choice of $x_1 = \frac{x+n}{n+1}$ and $x_2 = 1$ will yield (2.3).

Combining the above choices of x_1 and x_2, we obtain

$$2^{n+1}\, I_{f(2^{n+1}-1)}(P,Q) \leq 2^n\, I_{f(2^n-1)}(P,Q) \leq \ldots \leq 16\, I_{f(15)}(P,Q) \leq$$
$$\leq 8\, I_{f(7)}(P,Q) \leq 4\, I_{f(3)}(P,Q) \leq 2\, I_{f(1)}(P,Q) \leq I_f(P,Q).$$

Also a choice of $x_1 = \frac{x+\frac{1}{n}}{\frac{1}{n}+1}$ and $x_2 = x$ will yield (2.4). The inequalities given by (2.3) and (2.4) can be used in establishing relationship among some well known divergence measures. For example, if $f(u) = -\ln u$ and $n = 0$ in (2.3), we obtain

$$\sum_{i=1}^{n} q_i \ln\left(\frac{2q_i}{p_i+q_i}\right) \leq \frac{1}{2}\sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right), \text{which gives } F(Q,P) \leq \frac{1}{2}K(Q,P).$$

Here $F(P,Q)$ and $K(P,Q)$ denote the Relative Jensen-Shannon divergence measure [17] and the Kullback–Leibler divergence measure [13] respectively.

3. Parameterization of *f*-divergences

Let us consider the set of all those divergence measures for which the associated convex functions f satisfy the functional equation

(3.1) $$f(u) = u\, f\left(\frac{1}{u}\right)$$

and for which $f(1) = 0$ (i.e., $f \in \mathbb{F}_0$). Now for any such solution *f*, set

(3.2) $$g(u) = \frac{u^{1/2}}{(u-1)^2}\, f(u) \text{ for } u > 1 \text{ and}$$

$$\varphi(t) = g\left(\left(t + \sqrt{t^2+1}\right)^2\right) \text{for } t > 1$$

(and define $\varphi(1)$ arbitrarily). One can easily check that

$$\varphi\left(\frac{u+1}{2u^{1/2}}\right) = g\left(\max\left\{u, \frac{1}{u}\right\}\right)$$

therefore, if $u > 1$,

$$f(u) = \frac{(u-1)^2}{u^{1/2}}\, g(u) = \frac{(u-1)^2}{u^{1/2}}\, g\left(\max\left\{u, \frac{1}{u}\right\}\right) = \frac{(u-1)^2}{u^{1/2}}\, \varphi\left(\frac{u+1}{2u^{1/2}}\right)$$

and, if $u < 1$,

$$f(u) = u\, f\left(\frac{1}{u}\right) = \frac{(u-1)^2}{u^{1/2}}\, g(u) = \frac{(u-1)^2}{u^{1/2}}\, g\left(\max\left\{u, \frac{1}{u}\right\}\right)$$
$$= \frac{(u-1)^2}{u^{1/2}}\, \varphi\left(\frac{u+1}{2u^{1/2}}\right).$$

Thus (3.1) holds (obviously, also for $u = 0$) for the function $\varphi(.)$ defined by (3.2). Therefore, it is very much clear that every solution of (3.1) satisfying $f(1) = 0$ can be written in the form

$$f(u) = \frac{(u-1)^2}{u^{1/2}}\, \varphi\left(\frac{u+1}{2u^{1/2}}\right)$$

for a suitable $\varphi(.)$.

We, therefore consider the following symmetric divergence functional

(3.3) $$I_{f(\varphi)}(P,Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{(p_i q_i)^{1/2}}\, f\left(\frac{p_i + q_i}{2(p_i q_i)^{1/2}}\right).$$

It should be noted that (3.3) represents a parameterization of the set of all such divergence measures which satisfy (3.1). But here the function $\varphi(.)$can be both convex and concave. Table 1 shows various choices of $\varphi(.)$ and the corresponding divergence functionals.

Table 1. New symmetric divergence measures

S. No	$\varphi(u)$	$f(u)$	$I_{f(\varphi)}(P,Q)$
1.	$\varphi(u) = k$ (a positive constant)	$k\,\frac{(u-1)^2}{u^{1/2}}$	$k\sum_{i=1}^{n} \frac{(p_i-q_i)^2}{(p_i q_i)^{1/2}} = E_k^*(P,Q)$
2.	$\varphi(u) = u^k$, $k = 1,2,3,\ldots$	$\frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^k$, $k = 1,2,3,\ldots$	$\sum_{i=1}^{n} \frac{(p_i-q_i)^2}{(p_i q_i)^{1/2}}\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}}\right)^k = L_k^*(P,Q)$, $k = 1,2,3,\ldots$
3.	$\varphi(u) = u^k \ln u$, $k = 1,2,3,\ldots$	$\frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^k \ln\left(\frac{u+1}{2u^{1/2}}\right)$, $k = 1,2,3,\ldots$	$\sum_{i=1}^{n} \frac{(p_i-q_i)^2}{(p_i q_i)^{1/2}}\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}}\right)^k \ln\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}}\right) = M_k^*(P,Q)$, $k = 1,2,3,\ldots$
4.	$\varphi(u) = \ln u$	$\frac{(u-1)^2}{u^{1/2}}\ln\left(\frac{u+1}{2u^{1/2}}\right)$	$\sum_{i=1}^{n} \frac{(p_i-q_i)^2}{(p_i q_i)^{1/2}}\ln\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}}\right) = G(P,Q)$
5.	$\varphi(u) = u^k - 1$, $k = 1,2,3,\ldots$	$\frac{(u-1)^2}{u^{1/2}}\left(\left(\frac{u+1}{2u^{1/2}}\right)^k - 1\right)$, $k = 1,2,3,\ldots$	$\sum_{i=1}^{n} \frac{(p_i-q_i)^2}{(p_i q_i)^{1/2}}\left(\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}}\right)^k - 1\right) = N_k^*(P,Q)$, $k = 1,2,3,\ldots$

Result 3.1. Consider the measures $E_1^*(P,Q), L_k^*(P,Q), G(P,Q)$ and $M_k^*(P,Q)$ as defined in Table 1. Then the following inequalities measures hold

(3.4) $$E_1^*(P,Q) \le L_1^*(P,Q) \le L_2^*(P,Q) \le L_3^*(P,Q) \le L_4^*(P,Q) \le \ldots$$

and

(3.5) $$G(P,Q) \le M_1^*(P,Q) \le M_2^*(P,Q) \le M_3^*(P,Q) \le M_4^*(P,Q) \le \ldots$$

The equality holds in the above inequalities if $p_{\mathrm{i}} = q_i$ for each i.

P r o o f. To start with, let us consider the arithmetic-geometric mean inequality given by

$$\frac{u+1}{2u^{1/2}} \geq 1 \text{ for all } u > 0.$$

The equality holds for $u = 1$. In general we have

$$1 \leq \frac{u+1}{2u^{1/2}} \leq \left(\frac{u+1}{2u^{1/2}}\right)^2 \leq \left(\frac{u+1}{2u^{1/2}}\right)^3 \leq \left(\frac{u+1}{2u^{1/2}}\right)^4 \leq \ \ldots$$

which is equivalent to

(3.6) $$\frac{(u-1)^2}{u^{1/2}} \leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right) \leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^2 \leq$$

$$\leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^3 \leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^4 \leq \ \ldots$$

Substituting $u = \frac{p_i}{q_i}$ in the above inequality, multiplying by q_i and then summing over all *i*, we obtain (3.4).

Again from (3.6), we have

$$\frac{(u-1)^2}{u^{1/2}}\ln\left(\frac{u+1}{2u^{1/2}}\right) \leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)\ln\left(\frac{u+1}{2u^{1/2}}\right) \leq$$

$$\leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^2 \ln\left(\frac{u+1}{2u^{1/2}}\right) \leq$$

$$\leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^3 \ln\left(\frac{u+1}{2u^{1/2}}\right) \leq$$

$$\leq \frac{(u-1)^2}{u^{1/2}}\left(\frac{u+1}{2u^{1/2}}\right)^4 \ln\left(\frac{u+1}{2u^{1/2}}\right) \leq \ \ldots$$

Substituting $u = \frac{p_i}{q_i}$ in the above inequality, multiplying by q_i and then summing over all *i*, we obtain (3.5). It is interesting to note that

$$L_1^*(P,Q) = \frac{1}{2}\psi(P,Q) = \frac{1}{2}\{\chi^2(P,Q) + \chi^2(Q,P)\}$$

where $\chi^2(P,Q)$ is the well known χ^2 divergence [15]

Result 3.2. Consider a differentiable function $\varphi(.) : (0,\infty) \to \mathbb{R}$ as defined in (3.3). Then the following inequalities hold

(3.7) $$\varphi'(1)N_1^*(P,Q) \leq I_{f(\varphi)}(P,Q) - \varphi(1)E_1^*(P,Q) \leq$$

$$\leq \sum_{i=1}^{n} \frac{\left(p_i^{1/2} - q_i^{1/2}\right)^2 (p_i - q_i)^2}{2\,(p_i q_i)} \varphi'\left(\frac{p_i + q_i}{2(p_i q_i)^{1/2}}\right)$$

if $\varphi(.)$ is convex, and

(3.8) $$\sum_{i=1}^{n} \frac{\left(p_i^{1/2} - q_i^{1/2}\right)^2 (p_i - q_i)^2}{2\,(p_i q_i)} \varphi'\left(\frac{p_i + q_i}{2(p_i q_i)^{1/2}}\right) \leq$$

$$\leq I_{f(\varphi)}(P,Q) - \varphi(1)E_1^*(P,Q) \leq \varphi'(1)N_1^*(P,Q)$$

if $\varphi(.)$ is concave.

P r o o f. First we assume that the function $\varphi(.)\colon (0,\infty) \to \mathbb{R}$ is differentiable and convex, then we have the following inequality

(3.9) $\varphi'(x)(y-x) \le \varphi(y) - \varphi(x) \le \varphi'(y)(y-x)$ for $x, y \in \mathbb{R}$.

Replacing y by $\frac{p_i + q_i}{2(p_i q_i)^{1/2}}$ and x by 1 in the above inequality, we obtain

$$\varphi'(1)\left(\frac{p_i + q_i}{2(p_i q_i)^{1/2}} - 1\right) \le \varphi\left(\frac{p_i + q_i}{2(p_i q_i)^{1/2}}\right) - \varphi(1) \le$$
$$\le \varphi'\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}}\right)\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}} - 1\right).$$

Multiplying both sides by and summing over all I in the above inequality, we obtain (3.7).

In addition, if we have $\varphi(1) = \varphi'(1) = 0$, then from (3.7), we have

(3.10) $0 \le I_{f(\varphi)}(P,Q) \le \sum_{i=1}^{n} \frac{\left(p_i^{1/2} - q_i^{1/2}\right)^2 (p_i - q_i)^2}{2\,(p_i q_i)} \varphi'\left(\frac{p_i+q_i}{2(p_i q_i)^{1/2}}\right).$

Again if we assume that the function $\varphi(.)\colon (0,\infty) \to \mathbb{R}$ to be differentiable and concave, then the inequality given by (3.9) gets reversed and as such the proof of (3.8) follows on similar lines as above.

Remark. The measure $E_1^*(P,Q)$ offers the following extension:

$$I_{f(\alpha)}(P,Q) = \sum_{i=1}^{n} \frac{|p_i - q_i|^{\alpha+1}}{(p_i q_i)^{\alpha/2}}, \alpha \in (0,\infty),$$

which includes the variation norm for $\alpha = 1$. Note that

$$f_\alpha(0) = \begin{Bmatrix} 1 & \alpha = 1 \\ \infty & \alpha \in (0,\infty) \end{Bmatrix}.$$

By virtue of arithmetic-geometric mean inequality, we have

$$\frac{2}{(u+1)^\alpha} \le \frac{1}{u^{\alpha/2}} \text{for all } u > 0, \quad \alpha \in (0,\infty)$$

which is equivalent to

$$\frac{|u-1|^{\alpha+1}}{(u+1)^\alpha} \le \frac{|u-1|^{\alpha+1}}{u^{\alpha/2}} \text{for all } u > 0, \alpha \in (0,\infty).$$

Substituting $u = \frac{p_i}{q_i}$ in the above inequality, multiplying by q_i and then summing over all i, we obtain

$$\vartheta_{f(\alpha)}(P,Q) \le I_{f(\alpha)}(P,Q), \quad \alpha \in (0,\infty).$$

Here $\vartheta_{f(\alpha)}(P,Q)$ are a class of symmetric divergences studied by Puri and Vincze which includes the triangular discrimination for $\alpha = 2$.

R e f e r e n c e s

1. A l i, S. M., S. D. S i l v e y. A General Class of Coefficients of Divergence of One Distribution from Another. – Jour. Roy. Statist. Soc. Series B, Vol. **28**, 1966, 131-142.
2. B a s s a t, M. B. *f*-Entropies, Probability of Error and Feature Selection. – Inform. Control, Vol. **39**, 1978, 227-242.
3. C h e n, H. C. Statistical Pattern Recognition. Hoyderc Book Co., Rocelle Park, New York, 1973.
4. C h o w, C. K., C. N. L i n. Approximating Discrete Probability Distributions with Dependence Trees. – IEEE Trans. Inform. Theory, Vol. **14**, 1968, No 3, 462-467.

5. C s i s z á r, I. Information – Type Measures of Difference of Probability Distributions and Indirect Observations. – Studia Sci. Math. Hungar., Vol. **2**, 1967, 299-318.
6. C s i s z á r, I. Information Measures: A Critical Survey. in Trans. – In: Seventh Prague Conf. on Information Theory, Academia, Prague, 1974, 73-86.
7. G o k h a l e, D. V., S. K u l l b a c k. Information in Contingency Tables. New York, Marcel Dekker, 1978.
8. J o n e s, L., C. B y r n e. General Entropy Criteria for Inverse Problems, with Applications to Data Compression, Pattern Classification, and Cluster Analysis. – IEEE Trans. Inform. Theory, Vol. **36**, 1990, 23-30.
9. K a d o t a, T. T., L. A. S h e p p. On the Best Finite Set of Linear Observables for Discriminating Two Gaussian Signals. – IEEE Trans. Inform. Theory, Vol. **13**, 1967, 288-294.
10. K a i l a t h, T. The Divergence and Bhattacharyya Distance Measures in Signal Selection. – IEEE Trans. Comm. Technology, Vol. **COM-15**, 1967, 52-60.
11. K a z a k o s, D., T. C o t s i d a s. A Decision Theory Approach to the Approximation of Discrete Probability Densities. – IEEE Trans. Perform. Anal. Machine Intell, Vol. **1**, 1980, 61-67.
12. K o l m o g o r o v, A. N. A New Invariant for Transitive Dynamical System. Dokl. Akad. Nauk. USSR, Vol. **119**, 1957, 861-869.
13. K u l l b a c k, S., A. S. L e i b l e r. On Information and Sufficiency. – Ann. Math. Stat., Vol. **22**, 1951, 79-86.
14. Ö s t e r r e i c h e r, F. Csiszar's *f* - Divergences – Basic Properties. Res. Report Collection, 2002. **http://rgmia.vu.edu.au/monographs/csiszar.htm**
15. P e a r s o n, K. On the Criterion That a Given System of Deviations from the Probable in the Case of Correlated System of Variables in Such That It Can Be Reasonable Supposed to Have Arisen from Random Sampling. – Phil. Mag., 1900, 157-172.
16. P i e l o u, E. C. Ecological Diversity. New York, Wiley, 1975.
17. S i b s o n, R. Information Radius. – Z. Wahrs. und Verw Geb., Vol. **14**, 1969, 149-160.
18. T h e i l, H. Economics and Information Theory. Amsterdam, North-Holland, 1967.
19. T h e i l, H. Statistical Decomposition Analysis. Amsterdam, North-Holland, 1972.