SOME GENERAL DIVERGENCE MEASURES FOR PROBABILITY DISTRIBUTIONS

S. S. DRAGOMIR (Victoria)

Abstract. General divergence measures for probability distributions are introduced and their main properties established. Connections with \( f \)-divergence corresponding to a convex function \( f \) are explored.

1. Introduction

Let \((X, \mathcal{A})\) be a measurable space satisfying \(|\mathcal{A}| > 2\) and \(\mu\) be a \(\sigma\)-finite measure on \((X, \mathcal{A})\). Let \(\mathcal{P}\) be the set of all probability measures on \((X, \mathcal{A})\) which are absolutely continuous with respect to \(\mu\). For \(P, Q \in \mathcal{P}\), let \(p = \frac{dP}{d\mu}\) and \(q = \frac{dQ}{d\mu}\) denote the Radon–Nikodym derivatives of \(P\) and \(Q\) with respect to \(\mu\). Two probability measures \(P, Q \in \mathcal{P}\) are said to be orthogonal and we denote this by \(Q \perp P\) if

\[ P\{q = 0\} = Q\{p = 0\} = 1. \]

Let \(f : [0, \infty) \to (-\infty, \infty]\) be a convex function that is continuous at 0, i.e., \(f(0) = \lim_{u \downarrow 0} f(u)\).

In 1963, I. Csiszár [2] introduced the concept of \(f\)-divergence as follows.

**Definition 1.** Let \(P, Q \in \mathcal{P}\). Then

\[ I_f(Q, P) = \int_X p(x) f\left(\frac{q(x)}{p(x)}\right) d\mu(x), \tag{1.1} \]

is called the \(f\)-divergence of the probability distributions \(Q\) and \(P\).
Remark 1. Observe that, the integrand in the formula (1.1) is undefined when \( p(x) = 0 \). The way to overcome this problem is to postulate for \( f \) as above that

\[
0 f \left[ \frac{q(x)}{0} \right] = q(x) \lim_{u \to 0} \left[ u f \left( \frac{1}{u} \right) \right], \quad x \in X.
\]

We now give some examples of \( f \)-divergences that are well-known and often used in the literature (see also [3]).

1.1. The class of \( \chi^\alpha \)-divergences. The \( f \)-divergences of this class, which is generated by the function \( \chi^\alpha = |u - 1|^{\alpha}, u \in [0, \infty), \alpha \in [1, \infty) \), have the form

\[
I_f(Q, P) = \int_X \left| \frac{q}{p} - 1 \right|^{\alpha} d\mu = \int_X p^{1-\alpha} |q - p|^{\alpha} d\mu.
\]

From this class only the parameter \( \alpha = 1 \) provides a distance in the topological sense, namely the total variation distance \( V(Q, P) = \int_X |q - p| d\mu \). The most prominent special case of this class is, however, the Karl Pearson’s \( \chi^2 \)-divergence.

1.2. Dichotomy class. From this class, generated by the function

\[
f_\alpha(u) = \begin{cases} 
    u - 1 - \ln u & \text{for } \alpha = 0; \\
    \frac{1}{\alpha(1-\alpha)}[\alpha u + 1 - \alpha - u^{\alpha}] & \text{for } \alpha \in \mathbb{R}\backslash\{0,1\}; \\
    1 - u + u \ln u & \text{for } \alpha = 1;
\end{cases}
\]

only the parameter \( \alpha = \frac{1}{2} \left( f_{\frac{1}{2}}(u) = 2(\sqrt{u} - 1)^2 \right) \) provides a distance, namely, the Hellinger distance

\[
H(Q, P) = \left[ \int_X (\sqrt{q} - \sqrt{p})^2 d\mu \right]^{\frac{1}{2}}.
\]

Another important divergence is the Kullback-Leibler divergence obtained for \( \alpha = 1 \):

\[
KL(Q, P) = \int_X q \ln \left( \frac{q}{p} \right) d\mu.
\]

1.3. Matsushita’s divergences. The elements of this class, which is generated by the function

\[
\varphi_\alpha(u) := |1 - u^{\alpha}|^{\frac{1}{\alpha}}, \quad u \in [0, \infty),
\]

Acta Mathematica Hungarica 109, 2005