Abstract—This paper considers derivation of f-divergence inequalities via the approach of functional domination. Bounds on an f-divergence based on one or several other f-divergences are introduced, dealing with pairs of probability measures defined on arbitrary alphabets. In addition, a variety of bounds are shown to hold under boundedness assumptions on the relative information.

Index Terms – f-divergence, relative entropy, relative information, reverse Pinsker inequalities, reverse Samson’s inequality, total variation distance, χ² divergence.

I. BASIC DEFINITIONS

We assume throughout that the probability measures P and Q are defined on a common measurable space (A, ℱ), and P ≪ Q denotes that P is absolutely continuous with respect to Q.

Definition 1: If P ≪ Q, the relative information provided by a ∈ A according to (P, Q) is given by

\[ \iota_{P\|Q}(a) = \log \frac{dP}{dQ}(a). \]  (1)

Introduced by Ali-Silvey [1] and Csiszár ([4]), a useful generalization of the relative entropy, which retains some of its major properties (and, in particular, the data processing inequality), is the class of f-divergences. A general definition of f-divergence is given in [14, p. 4398], specialized next to the case where P ≪ Q.

Definition 2: Let f : (0, ∞) → ℝ be a convex function, and suppose that P ≪ Q. The f-divergence from P to Q is given by

\[ D_f(P\|Q) = \int f\left(\frac{dP}{dQ}\right) dQ = \mathbb{E}[f(Z)] \]  (2)

with

\[ Z = \exp(\iota_{P\|Q}(Y)), \quad Y \sim Q. \]  (3)

In (2), we take the continuous extension\(^3\)

\[ f(0) = \lim_{t\downarrow 0} f(t) \in (-\infty, +\infty]. \]  (4)

If p and q denote, respectively, the densities of P and Q with respect to a σ-finite measure μ (i.e., \( p = \frac{dP}{d\mu} \), \( q = \frac{dQ}{d\mu} \)), then we can write (2) as

\[ D_f(P\|Q) = \int q f\left(\frac{p}{q}\right) d\mu. \]  (5)

Remark 1: Different functions may lead to the same f-divergence for all (P, Q): if for an arbitrary b ∈ ℜ, we have

\[ f_b(t) = f_0(t) + b(t - 1), \quad t \geq 0 \]  (6)

then

\[ D_{f_b}(P\|Q) = D_{f_0}(P\|Q). \]  (7)

Relative entropy is \( D_r(P\|Q) \) where r is given by

\[ r(t) = t \log t + (1 - t) \log e, \]  (8)

and the total variation distance \( |P - Q| \) and χ² divergence \( \chi^2(P\|Q) \) are f-divergences with \( f(t) = (t - 1)^2 \) and \( f(t) = |t - 1| \), respectively.

The following key property of f-divergences follows from Jensen’s inequality.

Proposition 1: If f : (0, ∞) → ℜ is convex and f(1) = 0, P ≪ Q, then

\[ D_f(P\|Q) \geq 0. \]  (9)

If, furthermore, f is strictly convex at t = 1, then equality in (9) holds if and only if P = Q.

The reader is referred to [19] for a survey on general properties of f-divergences, and also to the textbook by Liese and Vajda [13].

The numerical optimization of an f-divergence subject to simultaneous constraints on f_i-divergences (i = 1, ..., L) was recently studied in [12], which showed that for that purpose it is enough to restrict attention to alphabets of cardinality L + 2.

The full paper version of our work, which includes several approaches for the derivation of f-divergence inequalities, is available in [17].

\(^3\)The convexity of f : (0, ∞) → ℜ implies its continuity on (0, ∞).
II. FUNCTIONAL DOMINATION

Let \( f \) and \( g \) be convex functions on \((0, \infty)\) with \( f(1) = g(1) = 0 \), and let \( P \) and \( Q \) be probability measures defined on a measurable space \((\mathcal{A}, \mathcal{F})\). If, for \( \alpha > 0 \), \( f(t) \leq \alpha g(t) \) for all \( t \in (0, \infty) \) then, it follows from Definition 2 that

\[
D_f(P||Q) \leq \alpha D_g(P||Q). \tag{10}
\]

This simple observation leads to a proof of several inequalities with the aid of Remark 1.

A. Basic Tool

We start this section by proving a general result, which will be helpful in proving various tight bounds among \( f \)-divergences.

**Theorem 1**: Let \( P \ll Q \), and assume
- \( f \) is convex on \((0, \infty)\) with \( f(1) = 0 \);
- \( g \) is convex on \((0, \infty)\) with \( g(1) = 0 \);
- \( g(t) > 0 \) for all \( t \in (0, 1) \cup (1, \infty) \).

Denote the function \( \kappa : (0, 1) \cup (1, \infty) \rightarrow \mathbb{R} \)

\[
\kappa(t) = \frac{f(t)}{g(t)}, \quad t \in (0, 1) \cup (1, \infty) \tag{11}
\]

and

\[
\bar{\kappa} = \sup_{t \in (0, 1) \cup (1, \infty)} \kappa(t). \tag{12}
\]

Then,

a) \( D_f(P||Q) \leq \bar{\kappa} D_g(P||Q). \tag{13} \)

b) If, in addition, \( f'(1) = g'(1) = 0 \), then

\[
\sup_{P \neq Q} \frac{D_f(P||Q)}{D_g(P||Q)} = \bar{\kappa}. \tag{14}
\]

**Proof**: See [17, Theorem 1].

**Remark 2**: Beyond the restrictions in Theorem 1a), the only operative restriction imposed by Theorem 1b) is the differentiability of the functions \( f \) and \( g \) at \( t = 1 \). Indeed, we can invoke Remark 1 and add \( f'(1)(1-t) \) to \( f(t) \), without changing \( D_f \) and likewise with \( g \) and thereby satisfying the condition in Theorem 1b); the stationary point at \( 1 \) must be a minimum of both \( f \) and \( g \) because of the assumed convexity, which implies their non-negativity on \( (0, \infty) \).

**Remark 3**: It is useful to generalize Theorem 1b) by dropping the assumption on the existence of the derivatives at \( 1 \). As it is explained in [17], it is enough to require that the left derivatives of \( f \) and \( g \) at \( 1 \) be equal to 0. Analogously, if \( \bar{\kappa} = \sup_{0 < t < 1} \kappa(t) \), it is enough to require that the right derivatives of \( f \) and \( g \) at \( 1 \) be equal to 0.

B. Relationships Among \( D(P||Q) \), \( \chi^2(P||Q) \) and \( |P - Q| \)

**Theorem 2**: \( P \ll Q \)

a) If \( P \ll Q \) and \( c_1, c_2 \geq 0 \), then

\[
D(P||Q) \leq (c_1 |P - Q| + c_2 \chi^2(P||Q)) \log e \tag{15}
\]

holds if \( (c_1, c_2) = (0, 1) \) and \( (c_1, c_2) = (\frac{1}{2}, \frac{1}{2}) \). Furthermore, if \( c_1 = 0 \) then \( c_2 = 1 \) is optimal, and if \( c_2 = \frac{1}{2} \) then \( c_1 = \frac{1}{2} \) is optimal.

b) If \( P \ll Q \) and \( P \neq Q \), then

\[
\frac{D(P||Q) + D(Q||P)}{\chi^2(P||Q) + \chi^2(Q||P)} \leq \frac{1}{2} \log e \tag{16}
\]

and the constant in the right side of (16) is the best possible.

**Proof**: See [17, Theorem 2].

**Remark 4**: Inequality (15) strengthens the bound in [9, (2.8)].

\[
D(P||Q) \leq \frac{1}{2} (|P - Q| + \chi^2(P||Q)) \log e. \tag{17}
\]

Note that the short outline of the suggested proof in [9, p. 710] leads not (17) but to the weaker upper bound \(|P - Q| + \frac{1}{2} \chi^2(P||Q) \) nats.

C. An Alternative Proof of Samson’s Inequality

For the purpose of this sub-section, we introduce Marton’s divergence [15]:

\[
d^2_2(P, Q) = \min E [\mathbb{P}^2(X \neq Y | Y)] \tag{18}
\]

where the minimum is over all probability measures \( P_{XY} \) with respective marginals \( P_X = P \) and \( P_Y = Q \). From [15, pp. 558–559]

\[
d^2_2(P, Q) = D_s(P||Q) \tag{19}
\]

with

\[
s(t) = (t - 1)^2 1 \{ t < 1 \}. \tag{20}
\]

Note that Marton’s divergence satisfies the triangle inequality [15, Lemma 3.1], and \( d_2(P, Q) = 0 \) implies \( P = Q \); however, due to its asymmetry, it is not a distance measure.

An analog of Pinsker’s inequality, which comes in handy for the proof of Marton’s conditional transportation inequality [3, Lemma 8.4], is the following bound due to Samson [16, Lemma 2]:

**Theorem 3**: If \( P \ll Q \), then

\[
d^2_2(P, Q) + d^2_2(Q, P) \leq \frac{2}{\log e} D(P||Q). \tag{21}
\]

In [17, Section 3.D], we provide an alternative proof of Theorem 3, in view of Theorem 1b), with the following advantages:

a) This proof yields the optimality of the constant in (21), i.e., we prove that

\[
\sup_{P \neq Q} \frac{d^2_2(P, Q) + d^2_2(Q, P)}{D(P||Q)} = \frac{2}{\log e} \tag{22}
\]
where the supremum is over all probability measures \( P, Q \) such that \( P \neq Q \) and \( P \ll Q \).

b) A simple adaptation of this proof results in a reverse inequality to (21), which holds under the boundedness assumption of the relative information (see Section III-D).

D. Ratio of \( f \)-Divergence to Total Variation Distance

Let \( f: (0, \infty) \rightarrow \mathbb{R} \) be a convex function with \( f(1) = 0 \), and let \( f^*: (0, \infty) \rightarrow \mathbb{R} \) be given by

\[
f^*(t) = tf\left(\frac{1}{t}\right)
\]

(23) for all \( t > 0 \). Note that \( f^* \) is also convex, \( f^*(1) = 0 \), and \( D_f(P\|Q) = D_{f^*}(Q\|P) \) if \( P \ll Q \). By definition, we take

\[
f^*(0) = \lim_{t \downarrow 0} f^*(t) = \lim_{u \to \infty} \frac{f(u)}{u}.
\]

(24) Vajda [18, Theorem 2] showed that the range of an \( f \)-divergence is given by

\[
0 \leq D_f(P\|Q) \leq f(0) + f^*(0)
\]

(25) where every value in this range is attainable by a suitable pair of probability measures \( P \ll Q \). Recalling Remark 1, note that \( f_0(0) + f^*_0(0) = f(0) + f^*(0) \) with \( f_0(\cdot) \) defined in (6). Basu et al. [2, Lemma 11.1] strengthened (25), showing that

\[
D_f(P\|Q) \leq \frac{1}{2} (f(0) + f^*(0)) |P - Q|.
\]

(26) If \( f(0) \) and \( f^*(0) \) are finite, (26) yields a counterpart to a result by Csiszár (see [6, Theorem 3.1]) which implies that if \( f: (0, \infty) \rightarrow \mathbb{R} \) is a strictly convex function, then there exists a real-valued function \( \psi_f \) such that \( \lim_{x \to 0} \psi_f(x) = 0 \), and

\[
|P - Q| \leq \psi_f(D_f(P\|Q)).
\]

(27) Next, we demonstrate that the constant in (26) cannot be improved.

Theorem 4: If \( f: (0, \infty) \rightarrow \mathbb{R} \) is convex with \( f(1) = 0 \), then

\[
\sup_{P \neq Q} \frac{D_f(P\|Q)}{|P - Q|} = \frac{1}{2} (f(0) + f^*(0))
\]

(28) where the supremum is over all probability measures \( P, Q \) such that \( P \ll Q \) and \( P \neq Q \).

Proof: See [17, Theorem 5].

Remark 5: Csiszár [5, Theorem 2] showed that if \( f(0) \) and \( f^*(0) \) are finite and \( P \ll Q \), then there exists a constant \( C_f > 0 \) which depends only on \( f \) such that \( D_f(P\|Q) \leq C_f \sqrt{|P - Q|} \). Note that, if \( |P - Q| < 1 \), then this inequality is superseded by (26) where the constant is not only explicit but is the best possible according to Theorem 4.

A direct application of Theorem 4 yields

Corollary 1: \[
\sup_{P \neq Q} \frac{d_f^2(P, Q)}{|P - Q|} = \frac{1}{2}
\]

(29) where the supremum in (29) is over all \( P \ll Q \) with \( P \neq Q \), and the supremum in (30) is over all \( P \ll Q \) with \( P \neq Q \).

Proof: See [17, Corollary 1].

Remark 6: The results in (29) and (30) form counterparts of (22).

III. BOUNDED RELATIVE INFORMATION

In this section we show that it is possible to find bounds among \( f \)-divergences without requiring a strong condition of functional domination (see Section II) as long as the relative information is upper and/or lower bounded almost surely.

A. Definition of \( \beta_1 \) and \( \beta_2 \)

The following notation is used throughout the rest of the paper. Given a pair of probability measures \( (P, Q) \) on the same measurable space, denote \( \beta_1, \beta_2 \in [0, 1] \) by

\[
\beta_1 = \exp\left(-D_\infty(P\|Q)\right),
\]

(31) \[
\beta_2 = \exp\left(-D_\infty(Q\|P)\right)
\]

(32) with the convention that if \( D_\infty(P\|Q) = \infty \), then \( \beta_1 = 0 \), and if \( D_\infty(Q\|P) = \infty \), then \( \beta_2 = 0 \). Note that if \( \beta_1 > 0 \), then \( P \ll Q \), while \( \beta_2 > 0 \) implies \( Q \ll P \). Furthermore, if \( P \ll Q \), then with \( Y \sim Q \),

\[
\beta_1 = \text{ess inf} \frac{dQ}{dP}(Y) = \left(\text{ess sup} \frac{dP}{dQ}(Y)e^{-1}\right),
\]

(33) \[
\beta_2 = \text{ess inf} \frac{dP}{dQ}(Y) = \left(\text{ess sup} \frac{dQ}{dP}(Y)e^{-1}\right).
\]

(34) The following example illustrates an important case in which \( \beta_1 \) and \( \beta_2 \) are positive.

Example 1: (Shifted Laplace distributions.) Let \( P \) and \( Q \) be the probability measures whose probability density functions are, respectively, given by \( f_\lambda(\cdot - a_0) \) and \( f_\lambda(\cdot - a_1) \) with

\[
f_\lambda(x) = \frac{1}{2} \exp(-\lambda |x|), \quad x \in \mathbb{R}
\]

(35) where \( \lambda > 0 \). In this case, (35) yields

\[
\beta_1 = \beta_2 = \exp(-\lambda |a_1 - a_0|) \in (0, 1].
\]

(36)

B. Basic Tool

Since \( \beta_1 = 1 \Leftrightarrow \beta_2 = 1 \Leftrightarrow P = Q \), it is advisable to avoid trivialities by excluding that case.

Theorem 5: Let \( f \) and \( g \) satisfy the assumptions in Theorem 1, and assume that \((\beta_1, \beta_2) \in [0, 1]^2\). Then,

\[
D_f(P\|Q) \leq \kappa^* D_g(P\|Q)
\]

(37) where

\[
\kappa^* = \sup_{\beta \in (\beta_1, 1) \cup (1, \beta_1)^{-1}} \kappa(\beta)
\]

(38) and \( \kappa(\cdot) \) is defined in (11).

Proof: See [17, Theorem 5].

Note that if \( \beta_1 = \beta_2 = 0 \), then Theorem 5 does not improve upon Theorem 1a).
Remark 7: In the application of Theorem 5, it is often convenient to make use of the freedom afforded by Remark 1 and choose the corresponding offsets such that:

- the positivity property of $g$ required by Theorem 5 is satisfied;
- the lowest $\kappa^*$ is obtained.

Remark 8: Similarly to the proof of Theorem 1b), under the conditions therein, one can verify that the constants in Theorem 5 are the best possible among all probability measures $P, Q$ with given $(\beta_1, \beta_2) \in (0, 1)^2$.

Remark 9: Note that if we swap the assumptions on $f$ and $g$ in Theorem 5, the same result translates into

$$\inf_{\beta \in (\beta_1, \beta_2) \cup (1, \beta_2^{-1})} \kappa(\beta) \cdot D_g(P\|Q) \leq D_f(P\|Q).$$

Furthermore, provided both $f$ and $g$ are positive (except at $t = 1$) and $\kappa$ is monotonically increasing, Theorem 5 and (39) result in

$$\kappa(\beta_2) D_g(P\|Q) \leq D_f(P\|Q) \leq \inf \kappa(\beta_2^{-1}) D_g(P\|Q).$$

In this case, if $\beta_1 > 0$, sometimes it is convenient to replace $\kappa(\beta_2) > 0$ with $\beta_2$ at the expense of loosening the bound. A similar observation applies to $\beta_2$.

Example 2: If $f(t) = (t - 1)^2$ and $g(t) = |t - 1|$, we get

$$\chi^2(P\|Q) \leq \max \{\beta_1^{-1} - 1, 1 - \beta_2\} |P - Q|.$$ (42)

C. Bounds on $D_f(P\|Q)$ and $D_g(P\|Q)$

The remaining part of this section is devoted to various applications of Theorem 5. From this point, we make use of the definition of $\kappa: (0, \infty) \to [0, \infty]$ in (8).

An illustrative application of Theorem 5 gives upper and lower bounds on the ratio of relative entropies.

Theorem 6: Let $P \ll Q, P \neq Q$, and $(\beta_1, \beta_2) \in (0, 1)^2$. Let $\kappa: (0, 1) \cup (1, \infty) \to (0, \infty)$ be defined as

$$\kappa(t) = \frac{t \log t + (1 - t) \log e}{(t - 1) \log e - \log t}.$$ (43)

Then,

$$\kappa(\beta_2) \leq \frac{D_f(P\|Q)}{D_g(P\|Q)} \leq \kappa(\beta_1^{-1}).$$ (44)

Proof: See [17, Theorem 6].

D. Reverse Samson’s Inequality

The next result gives a counterpart to Samson’s inequality (21).

Theorem 7: Let $(\beta_1, \beta_2) \in (0, 1)^2$. Then,

$$\inf \frac{D_2^2(P, Q) + D_2^2(Q, P)}{D_f(P\|Q)} = \min \{\kappa(\beta_1^{-1}), \kappa(\beta_2)\}$$ (45)

where the infimum is over all $P \ll Q$ with given $(\beta_1, \beta_2)$, and where $\kappa: (0, 1) \cup (1, \infty) \to (0, \frac{2}{\log e})$ is given by

$$\kappa(t) = \frac{(t - 1)^2}{rt(t) \max \{1, t\}}, \quad t \in (0, 1) \cup (1, \infty).$$ (46)

Proof: See [17, Theorem 7].

E. Local Behavior of $f$-Divergences

Another application of Theorem 5 shows that the local behavior of $f$-divergences differs by only a constant, provided that the first distribution approaches the reference measure in a certain strong sense.

Theorem 8: Suppose that $\{P_n\}$, a sequence of probability measures defined on a measurable space $(A, \mathcal{F})$, converges to $Q$ (another probability measure on the same space) in the sense that, for $Y \sim Q$,

$$\lim_{n \to \infty} \text{ess sup} \frac{dP_n}{dQ}(Y) = 1$$

where it is assumed that $P_n \ll Q$ for all sufficiently large $n$. If $f$ and $g$ are convex on $(0, \infty)$ and they are positive except at $t = 1$ (where they are 0), then

$$\lim_{n \to \infty} D_f(P_n\|Q) = \lim_{n \to \infty} D_g(P_n\|Q) = 0,$$ (48)

and

$$\min \{\kappa(1^-), \kappa(1^+)\} \leq \lim_{n \to \infty} D_f(P_n\|Q) \leq \max \{\kappa(1^-), \kappa(1^+)\}$$ (49)

where we have indicated the left and right limits of the function $\kappa(\cdot)$, defined in (11), at 1 by $\kappa(1^-)$ and $\kappa(1^+)$, respectively.

Proof: See [17, Theorem 9].

Corollary 2: Let $\{P_n \ll Q\}$ converge to $Q$ in the sense of (47). Then, $D(P_n\|Q)$ and $D(Q\|P_n)$ vanish as $n \to \infty$ with

$$\lim_{n \to \infty} \frac{D(P_n\|Q)}{D(Q\|P_n)} = 1.$$ (50)

Corollary 3: Let $\{P_n \ll Q\}$ converge to $Q$ in the sense of (47). Then, $\chi^2(P_n\|Q)$ and $D(P_n\|Q)$ vanish as $n \to \infty$ with

$$\lim_{n \to \infty} \chi^2(P_n\|Q) = \frac{1}{2} \log e.$$ (51)

Note that (51) is known in the finite alphabet case [7, Theorem 4.1].

F. Strengthened Jensen’s inequality

Bounding away from zero a certain density between two probability measures enables the following strengthened version of Jensen’s inequality, which generalizes a result in [11, Theorem 1].

Lemma 1: Let $f: \mathbb{R} \to \mathbb{R}$ be a convex function, $P_1 \ll P_0$ be probability measures defined on a measurable space $(A, \mathcal{F})$, and fix an arbitrary random transformation $P_{Z|X}: A \to \mathbb{R}$. Denote $P_0 \to P_{Z|X} \to P_{Z_0}$, and $P_1 \to P_{Z|X} \to P_{Z_1}$.

Then,

$$\beta \left[ E[f(E[Z_0|X_0])] - f(E[Z_0]) \right] \leq E[f(E[Z_{1|X_1}]) - f(E[Z_1])].$$ (52)
where \( X_0 \sim P_0, X_1 \sim P_1, \) and \( \beta \triangleq \text{ess inf} \frac{dP_1}{dP_0}(X_0). \)

**Proof:** See [17, Lemma 1].

**Remark 10:** Letting \( Z = X, \) and choosing \( P_0 \) so that \( \beta = 0 \) (e.g., \( P_1 \) is a restriction of \( P_0 \) to an event of \( P_0 \)-probability less than 1), (52) becomes Jensen’s inequality \( f(\mathbb{E}[X_1]) \leq \mathbb{E}[f(X_1)]. \)

Lemma 1 finds the following application to the derivation of \( f \)-divergence inequalities.

**Theorem 9:** Let \( f: (0, \infty) \to \mathbb{R} \) be a convex function with \( f(1) = 0 \). Fix \( P \ll Q \) on the same space with \( (\beta_1, \beta_2) \in [0, 1)^2 \) and let \( X \sim P \). Then,

\[
\beta_2 D_f(P\|Q) = \mathbb{E}[f(\exp(tP\|Q(X))) - f(1 + \chi^2(P\|Q)) \leq \beta_2^{-1} D_f(P\|Q). \tag{54}
\]

Specializing Theorem 9 to the convex function on \( (0, \infty) \) where \( f(t) = -\log t \) sharpens the inequality

\[
D(P\|Q) \leq \log (1 + \chi^2(P\|Q)) \leq \chi^2(P\|Q) \log e. \tag{56}
\]

under the assumption of bounded relative information.

**Theorem 10:** Fix \( P \ll Q \) such that \( (\beta_1, \beta_2) \in (0, 1)^2 \). Then,

\[
\beta_2 D_f(P\|Q) = \mathbb{E}[f(\exp(tP\|Q(X))) - f(1 + \chi^2(P\|Q)) \leq \beta_2^{-1} D_f(Q\|P). \tag{58}
\]

**IV. REVERSE PINSKER INEQUALITIES**

It is not possible to lower bound \( |P - Q| \) solely in terms of \( D_f(P\|Q) \) since for an arbitrary small \( \epsilon > 0 \) and an arbitrary large \( \lambda > 0 \), we can construct examples with \( |P - Q| < \epsilon \) and \( \lambda < D_f(P\|Q) < \infty \). As in Section III, the following result involves the bounds on the relative information.

**Theorem 11:** If \( \beta_1 \in (0, 1) \) and \( \beta_2 \in [0, 1), \) then

\[
D_f(P\|Q) \leq \frac{1}{2} \left( \varphi(\beta_1^{-1}) - \varphi(\beta_2) \right) |P - Q| \tag{59}
\]

where \( \varphi : [0, \infty) \to [0, \infty) \) is given by

\[
\varphi(t) = \begin{cases} 0 & t = 0 \\ \frac{t \log t}{t - 1} & t \in (0, 1) \cup (1, \infty) \\ \log e & t = 1. \end{cases} \tag{60}
\]

**Proof:** See [17, Theorem 23].

**Remark 11:** Note that for Theorem 11 to give a nontrivial result, it is necessary that the relative information be upper bounded, namely \( \beta_1 > 0 \). However, we still get a nontrivial bound if \( \beta_2 = 0 \).

In the following, we assume that \( P \) and \( Q \) are probability measures defined on a common finite set \( \mathcal{A} \), and \( Q \) is strictly positive on \( \mathcal{A} \) with \( |\mathcal{A}| \geq 2 \).

**Theorem 12:** Let \( Q_{\min} = \min_{a \in \mathcal{A}} Q(a) \), then

\[
D_f(P\|Q) \leq \log \left( 1 + \frac{|P - Q|^2}{2Q_{\min}} \right). \tag{61}
\]

Furthermore, if \( Q \ll P \) and \( \beta_2 \) is defined as in (32), then the following tightened bound holds:

\[
D_f(P\|Q) \leq \log \left( 1 + \frac{|P - Q|^2}{2Q_{\min}} \right) - \frac{1}{2}\beta_2 |P - Q|^2 \log e. \tag{62}
\]

**Proof:** See [17, Theorem 25].

**Remark 12:** The result in Theorem 12 improves the inequality by Csiszar and Talata [8, p. 1012]:

\[
D_f(P\|Q) \leq \left( \frac{\log e}{Q_{\min}} \right) |P - Q|^2. \tag{62}
\]

For further reverse Pinsker Inequalities and some of their implications, see [17, Section 6].

**REFERENCES**

[1] S. M. Ali and S. D. Silvey, “A general class of coefficients of divergence of one distribution from another,” Journal of the Royal Statistical Society, series B, vol. 28, no. 1, pp. 131–142, 1966.

[2] A. Basu, H. Shioya and C. Park, “Statistical Inference: The Minimum Distance Approach,” Chapman & Hall/CRC Monographs on Statistics and Applied Probability, vol. 120, CRC Press, Boca Raton, Florida, USA, 2011.

[3] S. Boucheron, G. Lugosi, and P. Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence, Oxford University Press, 2013.

[4] I. Csiszar, “Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten,” Publ. Math. Inst. Hungar. Acad. Sci., vol. 8, pp. 85–108, January 1963.

[5] I. Csiszar, “On topological properties of \( f \)-divergences,” Studia Scientiarum Mathematicarum Hungarica, vol. 2, pp. 329–339, 1967.

[6] I. Csiszar, “A class of measures of informativity of observation channels,” Periodica Mathematica Hungarica, vol. 2, no. 1, pp. 191–23, March 1972.

[7] I. Csiszar and P. C. Shields, “Information Theory and Statistics: A Tutorial”, Foundations and Trends in Communications and Information Theory, vol. 1, no. 4, 2004.

[8] I. Csiszar and Z. Talata, “Context tree estimation for not necessarily finite memory processes, via BIC and MDL,” IEEE Trans. on Information Theory, vol. 52, no. 3, pp. 1007–1016, March 2006.

[9] P. Diaconis and L. Saloff-Coste, “Logarithmic Sobolev inequalities for finite Markov chains,” Annals of Applied Probability, vol. 6, pp. 695–750, 1996.

[10] S. S. Dragomir, “Upper and lower bounds for Csiszar \( f \)-divergence in terms of the Kullback-Leibler distance and applications,” Inequalities for Csiszar \( f \)-Divergence in Information Theory, RGMIA Monographs, 2000.

[11] S. S. Dragomir, “Bounds for the normalized Jensen functional,” Bulletin of the Australian Mathematical Society, vol. 74, no. 3, pp. 471–478, 2006.

[12] A. Guntuboyina, S. Saha and G. Schiebinger, “Sharp inequalities for \( f \)-divergences,” IEEE Trans. on Information Theory, vol. 60, no. 1, pp. 104–121, Jan. 2014.

[13] F. Liese and I. Vajda, Convex Statistical Distances, Teubner-Texte Zur Mathematik, vol. 95, Germany, 1987.

[14] F. Liese and I. Vajda, “On divergences and informations in statistics and information theory,” IEEE Trans. on Information Theory, vol. 52, no. 10, pp. 4394–4412, October 2006.

[15] S. Martin, “A measure concentration inequality for contracting Markov chains,” Geometric and Functional Analysis, vol. 6, pp. 556–571, 1996.

[16] P. M. Massart, “Concentration of measure inequalities for Csiszar f-Divergence in Information Theory, RGMIA Monographs, 2000.

[17] P. M. Samson, “Concentration of measure inequalities for Markov chains,” Probability Theory and Related Fields, vol. 110, no. 3, pp. 465–483, 1998.

[18] I. Vajda, “On f-divergence and singularity of probability measures,” Periodica Mathematica Hungarica, vol. 2, no. 1–4, pp. 223–234, 1972.

[19] I. Vajda, “On metric divergences of probability measures,” Kybernetika, vol. 45, no. 6, pp. 885–900, 2009.

[20] S. Verdú, Information Theory, in preparation.