Upper Bounds on the Relative Entropy and Rényi Divergence as a Function of Total Variation Distance for Finite Alphabets

Igal Sason (Technion) Sergio Verdú (Princeton)

IEEE 2015 Information Theory Workshop
Jeju Island, Korea
October 12–15, 2015

Cast of Characters

- A finite set $\mathcal{A}$ with σ-algebra $\mathscr{F}$
- Probability measures P, Q defined on the measurable space $(\mathcal{A}, \mathscr{F})$.
- $X \sim P$.
- $Y \sim Q$.

Cast of Characters

- A finite set $\mathcal{A}$ with σ-algebra $\mathscr{F}$
- Probability measures P, Q defined on the measurable space $(\mathcal{A}, \mathscr{F})$.
- $X \sim P$.
- $Y \sim Q$.

Total Variation (TV) Distance

$$|P - Q| = 2 \sup_{\mathcal{F} \in \mathscr{F}} \big(P(\mathcal{F}) - Q(\mathcal{F})\big)$$
$$= \sum_{x \in \mathcal{A}} \big|P(x) - Q(x)\big|.$$

Csiszár-Kemperman-Kullback-Pinsker Inequality

$$D(P\|Q) \geq \tfrac{1}{2}\,|P-Q|^2\,\log e$$

and the constant is tight in the sense that

$$\inf_{P\neq Q} \frac{D(P\|Q)}{|P-Q|^2} = \tfrac{1}{2}\log e.$$

Csiszár-Kemperman-Kullback-Pinsker Inequality

$$D(P\|Q) \geq \tfrac{1}{2}\,|P - Q|^2 \,\log e$$

and the constant is tight in the sense that

$$\inf_{P \neq Q} \frac{D(P\|Q)}{|P - Q|^2} = \tfrac{1}{2} \log e.$$

An Implication of Pinsker's Inequality

Convergence in relative entropy $\Longrightarrow$ convergence in TV distance.

Question

Is there a reverse Pinsker inequality that provides an upper bound on the relative entropy as a function of the TV distance ?

Question

Is there a reverse Pinsker inequality that provides an upper bound on the relative entropy as a function of the TV distance ?

No, for every $\varepsilon > 0$ there exist P, Q s.t. $|P - Q| \le \varepsilon, \; D(P\|Q) = \infty$.

A Reverse Pinsker Inequality [Csiszár and Talata, 2006]

If $Q_{\min} \triangleq \min_{x \in \mathcal{A}} Q(x) > 0$, then

$$D(P\|Q) \leq \frac{\log e}{Q_{\min}} \cdot |P - Q|^2.$$

A Reverse Pinsker Inequality [Csiszár and Talata, 2006]

If $Q_{\min} \triangleq \min_{x \in \mathcal{A}} Q(x) > 0$, then

$$D(P\|Q) \leq \frac{\log e}{Q_{\min}} \cdot |P - Q|^2.$$

A Reverse Pinsker Inequality [Verdú, ITA 2014]

If $P \ll Q$, let $\beta_1 \in [0, 1]$ be given by $\beta_1^{-1} \triangleq \sup_{x \in \mathcal{A}} \frac{\mathrm{d}P}{\mathrm{d}Q}(x)$.
Then, if $\beta_1 < 1$ as above,

$$D(P\|Q) \leq \frac{\log \frac{1}{\beta_1}}{2(1 - \beta_1)} \cdot |P - Q|.$$

Hence, if the relative information $i_{P\|Q}(x) \triangleq \log \frac{\mathrm{d}P}{\mathrm{d}Q}(x)$ is bounded from above, a reverse Pinsker inequality exists.

New Reverse Pinsker Inequality

Theorem

a) *If* $P \ll Q$

$$D(P\|Q) \le \log\left(1 + \frac{|P-Q|^2}{2Q_{\min}}\right). \tag{1}$$

b) *Furthermore, if* $Q \ll P$ *and* $\beta_2 \in [0,1]$ *is given by*

$$\beta_2 = \min_{x \in \mathcal{A}} \frac{P(x)}{Q(x)}$$

then the following tightened bound holds:

$$D(P\|Q) \le \log\left(1 + \frac{|P-Q|^2}{2Q_{\min}}\right) - \frac{\beta_2 \log e}{2} \cdot |P-Q|^2. \tag{2}$$

Note on the Weaker Version of This Bound

This already improves the Csiszár-Talata inequality since

$$\log\left(1+\frac{|P-Q|^2}{2Q_{\min}}\right) \leq \frac{\log e}{\mathbf{2}Q_{\min}} \cdot |P-Q|^2.$$

Proof

The idea is to obtain upper and lower bounds on the χ^2-divergence

$$\chi^2(P,Q) \triangleq \sum_{x \in \mathcal{A}} \frac{(P(x) - Q(x))^2}{Q(x)}.$$

Proof

The idea is to obtain upper and lower bounds on the χ^2-divergence

$$\chi^2(P, Q) \triangleq \sum_{x \in \mathcal{A}} \frac{(P(x) - Q(x))^2}{Q(x)}.$$

Bounds on the χ^2-divergence

$$\begin{aligned} \chi^2(P\|Q) &= e^{D_2(P\|Q)} - 1 \\ &\geq e^{D(P\|Q)} - 1. \end{aligned}$$

$$\begin{aligned} \chi^2(P\|Q) &\leq \frac{\sum_{x \in \mathcal{A}} \big(P(x) - Q(x)\big)^2}{Q_{\min}} \\ &\leq \frac{|P - Q|^2}{2Q_{\min}}. \end{aligned}$$

Combining the bounds yields the weaker version of this inequality.

Distance From the Equiprobable Distribution

If P is a distribution on a finite set $\mathcal{A}$, $H(P)$ gauges the "distance" from U, the equiprobable distribution defined on $\mathcal{A}$, since

$$H(P) = \log |\mathcal{A}| - D(P \| \mathsf{U}).$$

It is of interest to explore the relationship between $H(P)$ and $|P - \mathsf{U}|$.

Distance From the Equiprobable Distribution

If P is a distribution on a finite set $\mathcal{A}$, $H(P)$ gauges the "distance" from U, the equiprobable distribution defined on $\mathcal{A}$, since

$$H(P) = \log|\mathcal{A}| - D(P\|\mathsf{U}).$$

It is of interest to explore the relationship between $H(P)$ and $|P - \mathsf{U}|$.

Next, we determine the exact locus of the points $(H(P), |P - \mathsf{U}|)$ among all probability measures P defined on $\mathcal{A}$, and this region is compared to upper and lower bounds on $|P - \mathsf{U}|$ as a function of $H(P)$.

Distance From the Equiprobable Distribution

If P is a distribution on a finite set $\mathcal{A}$, $H(P)$ gauges the "distance" from U, the equiprobable distribution defined on $\mathcal{A}$, since

$$H(P) = \log |\mathcal{A}| - D(P \| \mathsf{U}).$$

It is of interest to explore the relationship between $H(P)$ and $|P - \mathsf{U}|$.

Next, we determine the exact locus of the points $(H(P), |P - \mathsf{U}|)$ among all probability measures P defined on $\mathcal{A}$, and this region is compared to upper and lower bounds on $|P - \mathsf{U}|$ as a function of $H(P)$.

Notation

- $h(x) = -x \log x - (1-x) \log(1-x)$ for $x \in [0,1]$ denotes the binary entropy function;
- $d(x \| y) = x \log \frac{x}{y} + (1-x) \log \frac{1-x}{1-y}$ for $x, y \in [0,1]$ denotes the binary relative entropy.

Theorem

Let U *be the equiprobable distribution on a* $\{1, \ldots, |\mathcal{A}|\}$, *with* $1 < |\mathcal{A}| < \infty$.

a) *For* $\Delta \in (0, 2(1 - |\mathcal{A}|^{-1})]$,

$$\max_{P\colon |P - \mathsf{U}| = \Delta} H(P) = \log |\mathcal{A}| - \min_{m} d\left(|\mathcal{A}|^{-1} m + \tfrac{1}{2}\Delta \,\|\, |\mathcal{A}|^{-1} m\right) \quad (3)$$

where the minimum in the right side of (3) *is over all positive integers* m *not exceeding* $\min\left\{|\mathcal{A}| - 1, |\mathcal{A}|\,(1 - \tfrac{1}{2}\Delta)\right\}$. *Denoting such an integer by* m_Δ, *the maximum in the left side of* (3) *is attained by*

$$P_\Delta(\ell) = \begin{cases} |\mathcal{A}|^{-1} + \frac{\Delta}{2 m_\Delta} & \ell \in \{1, \ldots, m_\Delta\}, \\ |\mathcal{A}|^{-1} - \frac{\Delta}{2(|\mathcal{A}| - m_\Delta)}, & \ell \in \{m_\Delta + 1, \ldots, |\mathcal{A}|\}. \end{cases} \quad (4)$$

Theorem (cont.)

b) Let

$$h_k = \begin{cases} 0, & k = 0 \\ h\big(|\mathcal{A}|^{-1}k\big) + |\mathcal{A}|^{-1}\, k \log k, & k \in \{1, \ldots, |\mathcal{A}| - 2\} \\ \log |\mathcal{A}|, & k = |\mathcal{A}| - 1. \end{cases} \tag{5}$$

If $H \in [h_{k-1}, h_k)$ for $k \in \{1, \ldots, |\mathcal{A}| - 1\}$, then

$$\min_{P\colon H(P)=H} |P - \mathsf{U}| = 2\big(1 - (k+\theta)\, |\mathcal{A}|^{-1}\big) \tag{6}$$

which is achieved by

$$P_\theta^{(k)}(\ell) = \begin{cases} 1 - (k - 1 + \theta)\, |\mathcal{A}|^{-1}, & \ell = 1 \\ |\mathcal{A}|^{-1}, & \ell \in \{2, \ldots, k\}, \\ \theta\, |\mathcal{A}|^{-1}, & \ell = k + 1 \\ 0, & \ell \in \{k+2, \ldots, |\mathcal{A}|\} \end{cases} \tag{7}$$

where $\theta \in [0, 1)$ is chosen so that $H(P_\theta^{(k)}) = H$.

Bounds

Most well-known bound on the entropy difference in terms of the total variation distance is

$$|H(P) - H(Q)| \leq |P - Q| \log \left(\frac{|\mathcal{A}|}{|P - Q|} \right) \tag{8}$$

which holds if P, Q are probability measures defined on a finite set $\mathcal{A}$ with $\big|P(a) - Q(a)\big| \leq \frac{1}{2}$ for all $a \in \mathcal{A}$.

Bounds

Most well-known bound on the entropy difference in terms of the total variation distance is

$$|H(P) - H(Q)| \leq |P - Q| \log \left(\frac{|\mathcal{A}|}{|P - Q|} \right) \tag{8}$$

which holds if P, Q are probability measures defined on a finite set $\mathcal{A}$ with $\left|P(a) - Q(a)\right| \leq \frac{1}{2}$ for all $a \in \mathcal{A}$.

Particularizing (8) to the case where $Q = \mathsf{U}$, and $\left|P(a) - \frac{1}{|\mathcal{A}|}\right| \leq \frac{1}{2}$ for all $a \in \mathcal{A}$ yields

$$H(P) \geq \log |\mathcal{A}| - |P - \mathsf{U}| \, \log \left(\frac{|\mathcal{A}|}{|P - \mathsf{U}|} \right). \tag{9}$$

Bounds (cont.)

- The Bretagnolle-Huber inequality (1979)

$$\tfrac{1}{4}|P-Q|^2 \leq 1 - \exp\bigl(-D(P\|Q)\bigr). \quad (10)$$

- Particularizing Pinsker's inequality, (10) and the reverse Pinsker inequality yield

$$H(P) \leq \log|\mathcal{A}| - \tfrac{1}{2}\,|P-\mathsf{U}|^2\,\log e, \quad (11)$$

$$H(P) \leq \log|\mathcal{A}| + \log\left(1 - \tfrac{1}{4}\,|P-\mathsf{U}|^2\right), \quad (12)$$

$$H(P) \geq \log|\mathcal{A}| - \log\left(1 + \frac{|\mathcal{A}|}{2}\,|P-\mathsf{U}|^2\right). \quad (13)$$

- It can be checked that the lower bound on $H(P)$ in the previous slide is worse than (13), irrespectively of $|P-\mathsf{U}|$, if either $|\mathcal{A}|=2$ or $8 \leq |\mathcal{A}| \leq 102$ (note that $0 \leq |P-\mathsf{U}| \leq 2(1-|\mathcal{A}|^{-1})$).

Exact Locus versus Bounds

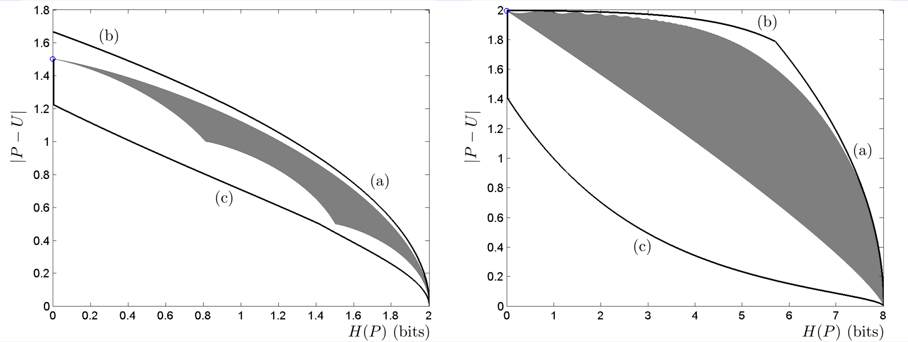

Figure: The exact locus of $(H(P), |P - \mathsf{U}|)$ among all the probability measures P defined on a finite set $\mathcal{A}$, and bounds on $|P - \mathsf{U}|$ as a function of $H(P)$ for $|\mathcal{A}| = 4$ (left plot), and $|\mathcal{A}| = 256$ (right plot). The point $(H(P), |P - \mathsf{U}|) = (0,\, 2(1 - |\mathcal{A}|^{-1}))$ is depicted on the y-axis. In the two plots, Curves (a), (b) and (c) refer, respectively, to (11), (12) and (13); the exact locus (shaded region) refers to Theorem 2.

The Rényi Divergence

In the discrete case, we have for $\alpha \in (0,1) \cup (1,\infty)$

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \left(\sum_{x \in \mathcal{A}} P^\alpha(x)\, Q^{1-\alpha}(x) \right).$$

The Rényi Divergence

In the discrete case, we have for $\alpha \in (0,1) \cup (1,\infty)$

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \left(\sum_{x \in \mathcal{A}} P^\alpha(x)\, Q^{1-\alpha}(x) \right).$$

Extreme cases:

- If $\alpha = 0$ then $D_0(P||Q) = -\log Q(\text{Support}(P))$,
- If $\alpha = +\infty$ then $D_\infty(P||Q) = \log \left(\text{ess sup } \frac{P}{Q} \right)$.
- If $D(P||Q) < \infty$, then $D(P||Q) = \lim_{\alpha \to 1^-} D_\alpha(P||Q)$.

New Bounds on the Rényi Divergence

Theorem

Let $P \ll\gg Q$ *be defined a common* finite set $\mathcal{A}$, *and assume that* P, Q *are strictly positive with minimum masses denoted by* $P_{\min}$ *and* $Q_{\min}$, *respectively. Let* $\beta_1 = \min_{a \in \mathcal{A}} Q(a)/P(a)$ *and* $\beta_2 = \min_{a \in \mathcal{A}} P(a)/Q(a)$. *Abbreviate* $\delta \triangleq \frac{1}{2}|P - Q| \in [0, 1]$. *Then, the Rényi divergence of order* $\alpha \in [0, \infty]$ *satisfies*

$$D_\alpha(P\|Q) \leq \begin{cases} f_1, & \alpha \in (2, \infty] \\ f_2, & \alpha \in [1, 2] \\ \min\left\{f_2, f_3, f_4\right\}, & \alpha \in \left(\frac{1}{2}, 1\right) \\ \min\left\{2 \log\left(\frac{1}{1-\delta}\right), f_2, f_3, f_4\right\}, & \alpha \in \left[0, \frac{1}{2}\right] \end{cases} \tag{14}$$

Theorem (cont.)

For $\alpha \in [0, \infty]$,

$$f_1(\alpha, \beta_1, \delta) \triangleq \begin{cases} \frac{1}{\alpha-1} \, \log\left(1 + \frac{\delta(\beta_1^{1-\alpha}-1)}{1-\beta_1}\right), & \alpha \in [0,1) \cup (1,\infty) \\ \frac{\delta}{1-\beta_1} \, \log\frac{1}{\beta_1}, & \alpha = 1, \\ \log\frac{1}{\beta_1}, & \alpha = \infty \end{cases} \tag{15}$$

for $\alpha \in [0, 2]$

$$f_2(\alpha, \beta_1, Q_{\min}, \delta) \triangleq \min\left\{ f_1(\alpha, \beta_1, \delta), \; \log\left(1 + \frac{2\delta^2}{Q_{\min}}\right)\right\} \tag{16}$$

and, for $\alpha \in [0, 1)$, f_3 and f_4 are given by

$$f_3(\alpha, P_{\min}, \beta_1, \delta) \triangleq \left(\frac{\alpha}{1-\alpha}\right)\left[\log\left(1 + \frac{2\delta^2}{P_{\min}}\right) - 2\beta_1\delta^2 \, \log e\right], \tag{17}$$

$$f_4(\beta_2, Q_{\min}, \delta) \triangleq \log\left(1 + \frac{2\delta^2}{Q_{\min}}\right) - 2\beta_2\delta^2 \, \log e. \tag{18}$$

Relation with a Reverse Pinsker Inequality

By letting $\alpha \to 1$ in Theorem 3, we get

$$D(P\|Q) \le \frac{\log \frac{1}{\beta_1}}{2(1-\beta_1)} \cdot |P-Q|$$

which coincides with the bound by Verdú (ITA '14). The bound on the Rényi divergence in Theorem 3 therefore generalizes the above bound on the relative entropy.

Plot: Bounds on the Rényi Divergence

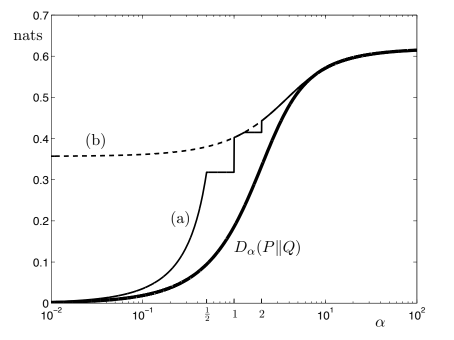

Figure: The Rényi divergence $D_\alpha(P\|Q)$ for P and Q which are defined on a binary alphabet with $P(0) = Q(1) = 0.65$, compared to (a) its upper bound in Theorem 3, and (b) its upper bound: $D_\alpha(P\|Q) \leq \frac{1}{\alpha-1} \log\left(1 + \frac{|P-Q|}{2} \frac{\beta_1^{1-\alpha}-1}{1-\beta_1}\right)$. These bounds coincide here when $\alpha \in (1, 1.291) \cup (2, \infty)$.

Summary

- This talk presents in part bounds among f-divergences which are derived for finite alphabets.
- The full paper version, which is currently under review and is available at `http://arxiv.org/abs/1508.00335`, derives bounds among f-divergences (and Rényi divergences) for general alphabets.