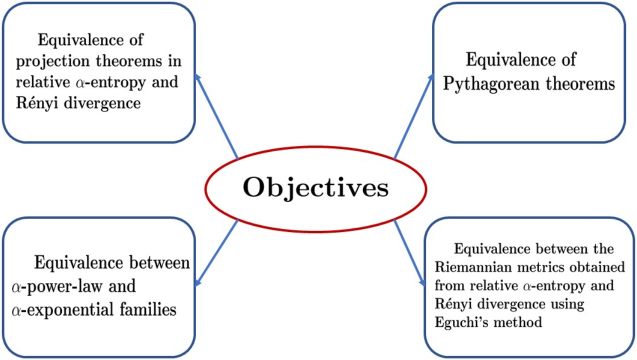
On the Equivalence of Projections in Relative α-Entropy and Rényi Divergence

P. N. Karthik
Department of Electrical Communication Engineering
Indian Institute of Science

Email: pnkarthik1992 @ gmail.com

December 15, 2017

Contents at a Glance



Notations

- $(\mathbb{X}, \mathcal{X})$ is an abstract measurable space, and μ is an arbitrary σ-finite measure on $(\mathbb{X}, \mathcal{X})$
- Consider $\alpha > 0, \alpha \neq 1$
- We work with the space of all probability measures on $(\mathbb{X}, \mathcal{X})$ absolutely continuous with respect to μ; the corresponding Radon-Nikodym derivatives (or μ-densities) are assumed to belong to $L^\alpha(\mu)$
- For a subset E of probability measures, we use $\mathcal{E}$ to denote the corresponding set of μ-densities, i.e.,

$$\mathcal{E} := \left\{ p = \frac{dP}{d\mu} : P \in E \right\}$$

Definitions-1

- Definition (α-scaled measure)

  Given a probability measure $P \ll \mu$ with μ-density p, its α-scaled measure $P^{(\alpha)}$ is the probability measure whose μ-density $p^{(\alpha)}$ is [1]

$$p^{(\alpha)} := \frac{p^\alpha}{\int p^\alpha d\mu} = \left(\frac{p}{||p||}\right)^\alpha \tag{1}$$

- Definition ($p \longleftrightarrow p^{(\alpha)}$ correspondence)

  Given a probability measure $P \ll \mu$ with μ-density p, a function $p^{(\alpha)}$ is said to be in correspondence with p, denoted as $p \longleftrightarrow p^{(\alpha)}$, if (1) holds

[1] We shall use the notation $||h|| = (\int h^\alpha d\mu)^{1/\alpha}$, even though $||\cdot||$, as defined, is not a norm when $\alpha < 1$. The dependence of $||\cdot||$ on α is suppressed for convenience.

Definitions-2

- **Definition (Relative α-entropy)**
  The relative α-entropy of P with respect to Q is defined as

$$\mathcal{I}_\alpha(P,Q) := \frac{\alpha}{1-\alpha}\log\left(\int \frac{p}{||p||}\left(\frac{q}{||q||}\right)^{\alpha-1} d\mu\right)$$

- **Definition (Rényi divergence)**
  The Rényi divergence of order α between P and Q is defined as

$$D_\alpha(P||Q) := \frac{1}{\alpha-1}\log\left(\int p^\alpha q^{1-\alpha} d\mu\right)$$

  - The key relation between the above quantities is

$$\mathcal{I}_\alpha(P,Q) = D_{1/\alpha}(P^{(\alpha)}||Q^{(\alpha)})$$

Definitions-3

- Definition ((α, λ)-mixture)

  Given two probability measures $P_0, P_1 \ll \mu$ and $\lambda \in [0,1]$, the (α, λ)-mixture of P_0 and P_1 is the probability measure $P_{\alpha,\lambda}$ whose μ-density $p_{\alpha,\lambda}$ is

$$p_{\alpha,\lambda} := \frac{\left(\lambda(p_1)^\alpha + (1-\lambda)(p_0)^\alpha\right)^{1/\alpha}}{\int \left(\lambda(p_1)^\alpha + (1-\lambda)(p_0)^\alpha\right)^{1/\alpha} d\mu}$$

- Definition (α-convex set)

  A set E of probability measures is said to be α-convex if for any $P_0, P_1 \in E$ and $\lambda \in [0,1]$, the (α, λ)-mixture of P_0 and P_1 belongs to E

Definitions-4

Let Q be a probability measure. Given $k \in \{1, 2, \dots\}$,
$\Theta = \{\theta = (\theta_1, \dots, \theta_k) : \theta_i \in \mathbb{R}\} \subset \mathbb{R}^k$, and functions $f_i : \mathbb{X} \to \mathbb{R}$, $1 \leq i \leq k$,

- Definition (α-power-law family)

  The α-power-law family generated by Q and $f_1, \dots, f_k$ is defined as the set of probability measures $\{P_\theta : \theta \in \Theta\}$, where

$$P_\theta(x)^{-1} = M(\theta) \left((Q(x))^{\alpha - 1} + (1 - \alpha) \sum_{i=1}^{k} \theta_i f_i(x) \right)^{\frac{1}{1-\alpha}}$$

- Definition (α-exponential family)

  The α-exponential family generated by Q and $f_1, \dots, f_k$ is defined as the set of probability measures $\{P_\theta : \theta \in \Theta\}$, where

$$P_\theta(x) = (N(\theta))^{-1} \left((Q(x))^{1 - \alpha} + (1 - \alpha) \sum_{i=1}^{k} \theta_i f_i(x) \right)^{\frac{1}{1-\alpha}}$$

for all $x \in \mathbb{X}$

Projection Theorems

- Consider a space $\mathbb{H}$ with a notion of a divergence $\mathcal{I}(P, Q)$ between any two points $P, Q \in \mathbb{H}$ that satisfies

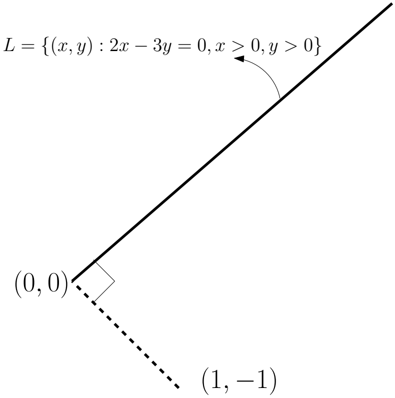
$$\mathcal{I}(P, Q) \geq 0, \text{ with equality if and only if } P = Q$$

- The (forward) projection of a point Q onto a set $E \subset \mathbb{H}$ is a member $P_* \in \mathbb{H}$ that satisfies

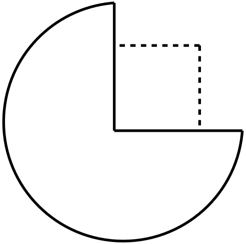
$$\mathcal{I}(P_*, Q) = \inf_{P \in E} \mathcal{I}(P, Q),$$

  and may be viewed as the best approximant of Q from the set E
- Projection theorems provide sufficient conditions on E for the existence and uniqueness of projections

Examples Where Existence or Uniqueness is Not Guaranteed



Convex but not closed



Closed but not convex

What's Known

- If $\mathbb{H}$ is a Hilbert space, $\mathcal{I}$ is the usual notion of distance $\langle P - Q, P - Q \rangle^{\frac{1}{2}}$ where $\langle \cdot, \cdot \rangle$ denotes the inner product, and if E is closed and convex, then it is known that a projection exists and is unique
- If $\mathbb{H}$ is the space of probability measures on an abstract measurable space, $\mathcal{I}$ is the relative entropy, and E is convex and closed with respect to the total variation metric, then a projection exists and is unique (result due to Csiszár [1])
- There are extensions in the latter context

Two Projection Problems

Problem considered in Kumar and Sundaresan [2]

- Fix $\alpha > 0$, $\alpha \neq 1$. Let Q be any probability measure, $Q \ll \mu$, and E be a set of probability measures whose set of μ-densities is $\mathcal{E}$. Solve

$$\inf_{P \in E} \mathcal{I}_\alpha(P, Q) \qquad (2)$$

- A sufficient condition proposed for the existence and uniqueness of solution is that E is convex and $\mathcal{E}$ is closed in $L^\alpha(\mu)$

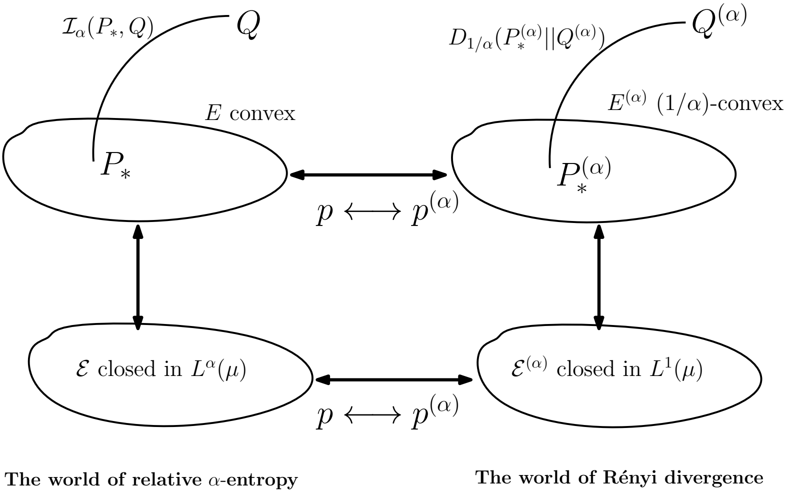
Problem considered in Kumar and Sason [3]

- Fix $\alpha > 0$, $\alpha \neq 1$. Let Q be any probability measure, $Q \ll \mu$, and E_1 be a set of probability measures whose set of μ-densities is $\mathcal{E}_1$. Solve

$$\inf_{P \in E_1} D_\alpha(P||Q) \qquad (3)$$

- A sufficient condition proposed for the existence and uniqueness of solution is that E_1 is α-convex and $\mathcal{E}_1$ is closed in $L^1(\mu)$

Equivalence of the Projection Problems



Equivalence of the Pythagorean Theorems

- Kumar and Sundaresan establish that if E is convex, then the projection P_* of Q onto E, if it exists, satisfies

$$\mathcal{I}_\alpha(P, Q) \geq \mathcal{I}_\alpha(P, P_*) + \mathcal{I}_\alpha(P_*, Q) \text{ for all } P \in E$$

- $E^{(\alpha)}$ is $(1/\alpha)$-convex
- $\mathcal{I}_\alpha(P, Q) = D_{1/\alpha}(P^{(\alpha)}||Q^{(\alpha)})$
- Thus, $P_*^{(\alpha)}$ is the $D_{1/\alpha}$-projection of $Q^{(\alpha)}$ onto $E^{(\alpha)}$ and this projection satisfies

$$D_{1/\alpha}(P^{(\alpha)}, Q^{(\alpha)}) \geq D_{1/\alpha}(P^{(\alpha)}, P_*^{(\alpha)}) + D_{1/\alpha}(P_*^{(\alpha)}, Q^{(\alpha)})$$

  for all $P^{(\alpha)} \in E^{(\alpha)}$
- This recovers the Pythagorean property for Rényi divergence, with $1/\alpha$ replacing α.

Equivalence of α-Power-Law and α-Exponential Families

- Suppose that P_θ. $\theta \in \Theta$, is a member of the α-power-law family generated by Q
- Then,

$$
\begin{aligned}
P_\theta^{(\alpha)}(x) &\propto (P_\theta(x))^\alpha \\
&\propto \left((Q(x))^{\alpha-1} + (1-\alpha) \sum_{i=1}^{k} \theta_i f_i(x) \right)^{-\frac{\alpha}{1-\alpha}} \\
&\propto \left(\left(\frac{Q(x)}{||Q||} \right)^{\alpha-1} + (1-\alpha) \sum_{i=1}^{k} \frac{\theta_i}{||Q||^{\alpha-1}} f_i(x) \right)^{-\frac{\alpha}{1-\alpha}} \\
&\propto \left((Q^{(\alpha)}(x))^{1-\frac{1}{\alpha}} + \left(1 - \frac{1}{\alpha}\right) \sum_{i=1}^{k} \theta_i' f_i(x) \right)^{\frac{1}{1-\frac{1}{\alpha}}},
\end{aligned}
$$

where $\theta_i' := \frac{(-\alpha)\theta_i}{||Q||^{\alpha-1}}$, $1 \leq i \leq k$

Statistical Manifolds and Riemannian Metrics

- A *(Riemannian) metric* at a point $p \in S$ is an inner product defined between any two tangent vectors in $T_p(S)$
- A metric is completely characterized by a matrix whose entries are the inner products between the basis tangent vectors, i.e., it is characterized by the matrix
$$G(\phi) = [g_{i,j}(\phi)]_{i,j=1,\ldots,n},$$
where $g_{i,j} = \langle \partial_i, \partial_j \rangle$

Eguchi's Method of Characterizing Metrics From Divergences

- Let S be a manifold with a coordinate system $\phi = (\phi_1, \ldots, \phi_n)$, and let $\mathcal{I}$[2] be a divergence function on $S \times S$

- Eguchi [4] showed that there is a metric

$$G^{(\mathcal{I})}(\phi) = [g_{i,j}^{(\mathcal{I})}(\phi)]_{i,j=1,\ldots,n}$$

  with

$$g_{i,j}^{(\mathcal{I})}(\phi) = -\frac{\partial}{\partial \phi_i} \frac{\partial}{\partial \phi_j^{'}} \mathcal{I}(p_\phi, p_{\phi'}) \bigg|_{\phi' = \phi},$$

  where $\phi = (\phi_1, \ldots, \phi_n)$ and $\phi^{'} = (\phi_1^{'}, \ldots, \phi_n^{'})$

[2] When $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$, we write $\mathcal{I}(P, Q) = \mathcal{I}(p, q)$

Riemannian Metrics for Relative Entropy and Rényi Divergence

- It is well known that when $\mathcal{I}(p,q) = I(p||q)$, the relative entropy between p and q, (16) can be written as

$$g_{i,j}^{(I)}(\phi) = E_{p_\phi}[\partial_i \log p_\phi \cdot \partial_j \log p_\phi], \tag{4}$$

  where E_{p_ϕ} denotes the expectation with respect to p_ϕ

- The quantity in (4) is the (i,j) entry of the Fisher information matrix

- On similar lines, when $\mathcal{I}(p,q) = D_\alpha(p||q)$, the Rényi divergence of order α between p and q, where $\alpha > 0$, $\alpha \neq 1$, it can be shown that

$$\begin{aligned} g_{i,j}^{(D_\alpha)}(\phi) &= \alpha \cdot E_{p_\phi}[\partial_i \log p_\phi \cdot \partial_j \log p_\phi] \\ &= \alpha \cdot g_{i,j}^{(I)}(\phi) \\ &= g_{i,j}^{(\alpha I)}(\phi) \end{aligned}$$

Main Results

Theorem (1)

The minimization problem (2) *for a given* $\alpha > 0$, $\alpha \neq 1$, *is equivalent to* (3) *with* α *replaced by* $1/\alpha$ *and* E_1 *replaced by* $E^{(\alpha)}$, *the set of* α*-scaled measures corresponding to* E, *and* Q *replaced by* $Q^{(\alpha)}$. *Moreover, the hypotheses are identical under the* $p \longleftrightarrow p^{(\alpha)}$ *correspondence.*

Theorem (2)

Let $\mathbb{X}$ *be a finite alphabet. Fix* $\alpha > 0$, $\alpha \neq 1$, $k \in \{1, 2, \ldots\}$, *and* $\Theta = \{\theta = (\theta_1, \ldots, \theta_k) : \theta_i \in \mathbb{R}\} \subset \mathbb{R}^k$. *Let* $f_i : \mathbb{X} \to \mathbb{R}$, $1 \leq i \leq k$, *be specified. Given a probability measure* Q, *for every member of the* α*-power-law family generated by* Q, $f_1, \ldots, f_k$ *and* Θ, *its* α*-scaled measure is a member of the* $(1/\alpha)$*-exponential family generated by* $Q^{(\alpha)}$, $f_1, \ldots, f_k$ *and* Θ', *where* Θ' *is a scalar modification of* Θ *that depends on* α *and* Q.

Theorem (3)

Consider a finite alphabet $\mathbb{X}$ *and fix* $\alpha > 0$, $\alpha \neq 1$. *Let* S *be a statistical manifold equipped with a coordinate system* $\phi = (\phi_1, \ldots, \phi_n)$, *and let* $S^{(\alpha)}$ *denote the statistical manifold of the corresponding* α*-scaled measures. Then, for every* $p \in S$, *the Riemannian metric specified by relative* α*-entropy on* $T_p(S)$ *is equivalent to that specified by Rényi divergence of order* $1/\alpha$ *on* $T_{p^{(\alpha)}}(S^{(\alpha)})$.

References

I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Annals of Probability*, vol. 3, pp. 146–158, 1975.

M. A. Kumar and R. Sundaresan, "Minimization problems based on relative α-entropy I: Forward projection," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 5063–5080, September 2015.

M. A. Kumar and I. Sason, "Projection theorems for the Rényi divergence on α-convex sets," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 4924–4935, September 2016.

S. Eguchi, "Geometry of minimum contrast," *Hiroshima Math. J*, vol. 22, no. 3, pp. 631–647, 1992.