On Projections of the Rényi Divergence on Generalized Convex Sets

M. Ashok Kumar*

Joint work with Igal Sason**

*Department of Mathematics, Indian Institute of Technology Indore.

**Andrew and Erna Viterbi Faculty of Electrical Engineering,
Technion-Israel Institute of Technology.

July 15, 2016

Outline

- Forward & reverse projections of Rényi divergence on generalized convex sets.

Outline

- Forward & reverse projections of Rényi divergence on generalized convex sets.
- Forward projections - large deviations theory, maximum entropy principle.

Outline

- Forward & reverse projections of Rényi divergence on generalized convex sets.
- Forward projections - large deviations theory, maximum entropy principle.
- Reverse projections - maximum likelihood estimation, robust statistics.

Outline

- Forward & reverse projections of Rényi divergence on generalized convex sets.
- Forward projections - large deviations theory, maximum entropy principle.
- Reverse projections - maximum likelihood estimation, robust statistics.
- Orthogonality of α-linear and α-exponential families for the Rényi divergence.

Rényi Divergence

- $\mathcal{M} := \mathcal{M}(\mathcal{A})$ - the set of all probability measures on $\mathcal{A}$.

Rényi Divergence

- $\mathcal{M} := \mathcal{M}(\mathcal{A})$ - the set of all probability measures on $\mathcal{A}$.
- For $P, Q \in \mathcal{M}$ and $\alpha \in (0,1) \cup (1,\infty)$, the **Rényi divergence** from P to Q is

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \left(\sum_a P(a)^\alpha Q(a)^{1-\alpha} \right).$$

Rényi Divergence

- $\mathcal{M} := \mathcal{M}(\mathcal{A})$ - the set of all probability measures on $\mathcal{A}$.
- For $P, Q \in \mathcal{M}$ and $\alpha \in (0,1) \cup (1,\infty)$, the **Rényi divergence** from P to Q is

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \left(\sum_a P(a)^\alpha Q(a)^{1-\alpha} \right).$$

For $\alpha = 1$, by continuous extension

$$D_1(P\|Q) = D(P\|Q).$$

Rényi Divergence

- $\mathcal{M} := \mathcal{M}(\mathcal{A})$ - the set of all probability measures on $\mathcal{A}$.
- For $P, Q \in \mathcal{M}$ and $\alpha \in (0,1) \cup (1,\infty)$, the **Rényi divergence** from P to Q is

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \left(\sum_a P(a)^\alpha Q(a)^{1-\alpha} \right).$$

For $\alpha = 1$, by continuous extension

$$D_1(P\|Q) = D(P\|Q).$$

  - $D_\alpha(P\|Q) \geq 0$ and $D_\alpha(P\|Q) = 0$ iff $P = Q$.

Information Projections for the Rényi Divergence

- Let $\mathcal{P} \subset \mathcal{M}$ and $Q \in \mathcal{M}$. Any $P^* \in \mathcal{P}$ satisfying

$$\min_{P \in \mathcal{P}} D_\alpha(P \| Q) = D_\alpha(P^* \| Q)$$

is called **forward** D_α**-projection** of Q on $\mathcal{P}$.

Information Projections for the Rényi Divergence

- Let $\mathcal{P} \subset \mathcal{M}$ and $Q \in \mathcal{M}$. Any $P^* \in \mathcal{P}$ satisfying

$$\min_{P \in \mathcal{P}} D_\alpha(P \| Q) = D_\alpha(P^* \| Q)$$

  is called **forward** D_α**-projection** of Q on $\mathcal{P}$.

- Let $\mathcal{Q} \subset \mathcal{M}$ and $P \in \mathcal{M}$. Any $Q^* \in \mathcal{Q}$ satisfying

$$\min_{Q \in \mathcal{Q}} D_\alpha(P \| Q) = D_\alpha(P \| Q^*)$$

  is called **reverse** D_α**-projection** of P on $\mathcal{Q}$.

Literature on Information Projections

- I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Annals of Probability*, 1975.
- I. Csiszár, "Sanov property, Generalized I-projection and a conditional limit theorem," *Annals of Probability*, 1984.
- I. Csiszár and F. Matúš, "Information projections revisited," *IEEE Trans. on IT*, 2003.
- I. Csiszár and P. C. Shields, "Information Theory and Statistics: A Tutorial," *FnT in COM & IT*, 2004.
- T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. on IT*, 2014.
- M. A. Kumar and R. Sundaresan, "Minimization problems based on α-relative entropy I & II: Forward & Reverse Projections," *IEEE Trans. on IT*, 2015.

FORWARD D_α-PROJECTION

Forward Projection - Motivation

- **Sanov's theorem**:
  - Suppose that $X_1, X_2, \ldots$ are i.i.d. and $X_1 \sim Q$. Then, if $m > \mathbb{E}\big[g(X_1)\big]$, for large n

    $$\mathsf{Pr}\Big(\frac{1}{n}\sum_{i=1}^{n} g(X_i) \geq m\Big) \approx \exp\{-nD(P^* \| Q)\},$$

    where

    $$P^* = \arg\min_{P \in \mathscr{L}} D(P\|Q)$$

    $$\mathscr{L} = \{P \in \mathcal{M} : \sum_a P(a)g(a) \geq m\}.$$

Forward Projection - Motivation

- **Sanov's theorem**:
  - Suppose that $X_1, X_2, \ldots$ are i.i.d. and $X_1 \sim Q$. Then, if $m > \mathbb{E}\big[g(X_1)\big]$, for large n

$$\Pr\Big(\frac{1}{n}\sum_{i=1}^{n} g(X_i) \geq m\Big) \approx \exp\{-nD(P^*\|Q)\},$$

    where

$$P^* = \arg\min_{P \in \mathscr{L}} D(P\|Q)$$

$$\mathscr{L} = \{P \in \mathcal{M} : \sum_a P(a)g(a) \geq m\}.$$

- **Conditional limit theorem**:
  - Suppose that $X_1, X_2, \ldots$ are i.i.d. and $X_1 \sim Q$. Then

$$\lim_{n\to\infty} P\Big\{X_1 = a \,\Big|\, \frac{1}{n}\sum_{i=1}^{n} g(X_i) \geq m\Big\} = P^*(a).$$

Tsallis' Maximum Entropy Problem

$$\arg\max_{(p_i)} \; S_\alpha(P) := \frac{1}{\alpha - 1}\Big(1 - \sum_a P(a)^\alpha\Big) \tag{1}$$

$$\text{subject to} \quad \frac{\sum_a P(a)^\alpha \epsilon(a)}{\sum_a P(a)^\alpha} = U^{(\alpha)}, \tag{2}$$

Tsallis' Maximum Entropy Problem

$$\arg\max_{(p_i)} \; S_\alpha(P) := \frac{1}{\alpha - 1}\Big(1 - \sum_a P(a)^\alpha\Big) \tag{1}$$

$$\text{subject to} \quad \frac{\sum_a P(a)^\alpha \epsilon(a)}{\sum_a P(a)^\alpha} = U^{(\alpha)}, \tag{2}$$

- The functional $S_\alpha(P)$ in (1) is called **Tsallis entropy**

Tsallis' Maximum Entropy Problem

$$\arg\max_{(p_i)} \; S_\alpha(P) := \frac{1}{\alpha - 1}\Big(1 - \sum_a P(a)^\alpha\Big) \tag{1}$$

$$\text{subject to} \quad \frac{\sum_a P(a)^\alpha \epsilon(a)}{\sum_a P(a)^\alpha} = U^{(\alpha)}, \tag{2}$$

- The functional $S_\alpha(P)$ in (1) is called **Tsallis entropy**
- The constraint in (2) corresponds to an α-**convex set**

Tsallis' Maximum Entropy Problem

$$\arg\max_{(p_i)} \; S_\alpha(P) := \frac{1}{\alpha-1}\Big(1-\sum_a P(a)^\alpha\Big) \tag{1}$$

$$\text{subject to} \quad \frac{\sum_a P(a)^\alpha \epsilon(a)}{\sum_a P(a)^\alpha} = U^{(\alpha)}, \tag{2}$$

- The functional $S_\alpha(P)$ in (1) is called **Tsallis entropy**
- The constraint in (2) corresponds to an α-**convex set**
- If $Q = U$ is uniform,

$$D_\alpha(P\|U) = \log|\mathcal{A}| + \frac{1}{\alpha-1}\,\log\big(1-(\alpha-1)S_\alpha(P)\big).$$

Thus maximization of $S_\alpha(P)$ is equivalent to minimization of $D_\alpha(P\|U)$.

α-Convex Sets

Definition ((α, λ)-**mixture**)

Let $P_0, P_1 \in \mathcal{M}$. The (α, λ)-**mixture** of (P_0, P_1) is the probability measure $S_{0,1}$ defined by

$$S_{0,1}(a) := \frac{1}{Z}\Big[(1-\lambda)P_0(a)^\alpha + \lambda P_1(a)^\alpha\Big]^{\frac{1}{\alpha}} \quad \forall a \in \mathcal{A},$$

where Z is a normalizing constant.

α-Convex Sets

Definition ((α, λ)-**mixture**)

Let $P_0, P_1 \in \mathcal{M}$. The (α, λ)-**mixture** of (P_0, P_1) is the probability measure $S_{0,1}$ defined by

$$S_{0,1}(a) := \frac{1}{Z}\Big[(1-\lambda)P_0(a)^\alpha + \lambda P_1(a)^\alpha\Big]^{\frac{1}{\alpha}} \quad \forall a \in \mathcal{A},$$

where Z is a normalizing constant.

Definition (α-**convex set**)

$\mathcal{P} \subset \mathcal{M}$ is said to be an α-**convex** set if, for every $P_0, P_1 \in \mathcal{P}$ and $\lambda \in (0,1)$, the (α, λ)-mixture $S_{0,1} \in \mathcal{P}$.

Result of van Erven and Harremoës (2014)

Theorem

If P^* is the forward D_α-projection of Q on an α-convex set $\mathcal{P}$, then the following **Pythagorean inequality** *holds:*

$$D_\alpha(P\|Q) \geq D_\alpha(P\|P^*) + D_\alpha(P^*\|Q) \quad \forall P \in \mathcal{P}.$$

Result of van Erven and Harremoës (2014)

Theorem

If P^* is the forward D_α-projection of Q on an α-convex set $\mathcal{P}$, then the following **Pythagorean inequality** *holds:*

$$D_\alpha(P\|Q) \geq D_\alpha(P\|P^*) + D_\alpha(P^*\|Q) \quad \forall P \in \mathcal{P}.$$

Note that the existence of the forward projection P^* is **not assured** in this theorem.

A Sufficient Condition for Existence

Theorem

Let $\alpha \in (0, \infty)$, $Q \in \mathcal{M}$, $\mathcal{P} \subseteq \mathcal{M}$ be α-convex & closed under total variation distance. If there exists $P \in \mathcal{P}$ such that $D_\alpha(P \| Q) < \infty$, then there exists a forward D_α-projection of Q on $\mathcal{P}$.

A Sufficient Condition for Existence

Theorem

Let $\alpha \in (0, \infty)$, $Q \in \mathcal{M}$, $\mathcal{P} \subseteq \mathcal{M}$ be α-convex & closed under total variation distance. If there exists $P \in \mathcal{P}$ such that $D_\alpha(P\|Q) < \infty$, then there exists a forward D_α-projection of Q on $\mathcal{P}$.

Proof outline:

- $\alpha > 1$: Similar to Csiszár 1975, but relies on a new Apollonius theorem for the Hellinger divergences:

$$\begin{aligned}(1-\lambda)\big(\mathscr{H}_\alpha(P_0\|Q) - \mathscr{H}_\alpha(P_0\|S_{0,1})\big) \\ + \lambda\big(\mathscr{H}_\alpha(P_1\|Q) - \mathscr{H}_\alpha(P_1\|S_{0,1})\big) \geq \mathscr{H}_\alpha(S_{0,1}\|Q),\end{aligned}$$

  where

$$\mathscr{H}_\alpha(P\|Q) := \frac{1}{\alpha - 1}\left(\sum_a P(a)^\alpha Q(a)^{1-\alpha} - 1\right).$$

A Sufficient Condition for Existence

Theorem

Let $\alpha \in (0, \infty)$, $Q \in \mathcal{M}$, $\mathcal{P} \subseteq \mathcal{M}$ be α-convex & closed under total variation distance. If there exists $P \in \mathcal{P}$ such that $D_\alpha(P\|Q) < \infty$, then there exists a forward D_α-projection of Q on $\mathcal{P}$.

Proof outline:

- $\alpha > 1$: Similar to Csiszár 1975, but relies on a new Apollonius theorem for the Hellinger divergences:

$$\begin{aligned}&(1-\lambda)\big(\mathscr{H}_\alpha(P_0\|Q) - \mathscr{H}_\alpha(P_0\|S_{0,1})\big)\\&+\lambda\big(\mathscr{H}_\alpha(P_1\|Q) - \mathscr{H}_\alpha(P_1\|S_{0,1})\big) \geq \mathscr{H}_\alpha(S_{0,1}\|Q),\end{aligned}$$

  where

$$\mathscr{H}_\alpha(P\|Q) := \frac{1}{\alpha-1}\Big(\sum_a P(a)^\alpha Q(a)^{1-\alpha} - 1\Big).$$

- $\alpha < 1$: Exploits Banach-Aloaglu theorem from functional analysis (for asserting compactness of a set).

Forward projection on α-linear Family

We focus on α-**linear family**:

$$\mathscr{L}_\alpha = \Big\{ P \in \mathcal{M} \colon \sum_a P(a)^\alpha f_i(a) = 0, \quad i = 1, \ldots, k \Big\}.$$

Forward projection on α-linear Family

We focus on α-**linear family**:

$$\mathscr{L}_\alpha = \Big\{ P \in \mathcal{M} \colon \sum_a P(a)^\alpha f_i(a) = 0, \quad i = 1, \ldots, k \Big\}.$$

Theorem

If $\mathcal{A}$ is finite and $\alpha > 1$, then Supp$(P^*) =$ Supp$(\mathscr{L}_\alpha)$ and the Pythagorean equality holds.

$$D_\alpha(P \| Q) = D_\alpha(P \| P^*) + D_\alpha(P^* \| Q) \quad \forall P \in \mathscr{L}_\alpha.$$

Forward projection on α-linear Family

We focus on α-**linear family**:

$$\mathscr{L}_\alpha = \Big\{ P \in \mathcal{M} \colon \sum_a P(a)^\alpha f_i(a) = 0, \quad i = 1, \ldots, k \Big\}.$$

Theorem

If $\mathcal{A}$ is finite and $\alpha > 1$, then Supp$(P^*) =$ Supp$(\mathscr{L}_\alpha)$ and the Pythagorean equality holds.

$$D_\alpha(P\|Q) = D_\alpha(P\|P^*) + D_\alpha(P^*\|Q) \quad \forall P \in \mathscr{L}_\alpha.$$

Theorem

If P^* is the forward D_α-projection on $\mathscr{L}_\alpha$, and if Supp$(P^*) =$ Supp$(\mathscr{L}_\alpha)$, then

$$P^*(a) = Z^{-1} \left[Q(a)^{1-\alpha} + (1-\alpha) \sum_{i=1}^{k} \theta_i^* f_i(a) \right]^{\frac{1}{1-\alpha}},$$

for some $\theta^* = (\theta_1^*, \ldots, \theta_k^*) \in \mathbb{R}^k$, and a normalizing constant Z.

α-Exponential Family

- Can write

$$P^*(a) = Z^{-1}\, e_\alpha\Big(\ln_\alpha(Q(a)) + \sum_{i=1}^{k} \theta_i f_i(a)\Big),$$

where e_α and $\ln_\alpha$ are, respectively, the α-**exponential** and α-**logarithmic** functions:

$$e_\alpha(x) := \begin{cases} \exp(x) & \text{if } \alpha = 1, \\ \Big(\max\big\{1 + (1-\alpha)x,\ 0\big\}\Big)^{\frac{1}{1-\alpha}} & \text{if } \alpha \in (0,1) \cup (1,\infty), \end{cases}$$

$$\ln_\alpha(x) := \begin{cases} \ln(x) & \text{if } \alpha = 1, \\ \frac{x^{1-\alpha}-1}{1-\alpha} & \text{if } \alpha \in (0,1) \cup (1,\infty). \end{cases}$$

α-Exponential Family

- Can write

$$P^*(a) = Z^{-1}\, e_\alpha\Big(\ln_\alpha(Q(a)) + \sum_{i=1}^{k} \theta_i f_i(a)\Big),$$

where e_α and $\ln_\alpha$ are, respectively, the α-**exponential** and α-**logarithmic** functions:

$$e_\alpha(x) := \begin{cases} \exp(x) & \text{if } \alpha = 1, \\ \Big(\max\big\{1 + (1-\alpha)x,\, 0\big\}\Big)^{\frac{1}{1-\alpha}} & \text{if } \alpha \in (0,1) \cup (1,\infty), \end{cases}$$

$$\ln_\alpha(x) := \begin{cases} \ln(x) & \text{if } \alpha = 1, \\ \frac{x^{1-\alpha}-1}{1-\alpha} & \text{if } \alpha \in (0,1) \cup (1,\infty). \end{cases}$$

- α-**exponential family** extends the usual exponential family:

$$\mathscr{E}_\alpha := \Big\{P \in \mathcal{M} \colon P(a) = Z(\theta)^{-1}\, e_\alpha\Big(\ln_\alpha(Q(a)) + \sum_{i=1}^{k} \theta_i f_i(a)\Big)\Big\}.$$

Convergence of an Iterative Process

Theorem

Let $\alpha \in (1,\infty)$. Suppose that $\mathscr{L}_\alpha^{(1)}, \ldots, \mathscr{L}_\alpha^{(m)}$ are α-linear families, and let

$$\mathscr{L}_\alpha := \bigcap_{n=1}^{m} \mathscr{L}_\alpha^{(n)}.$$

Let $P_0 = Q$, and let P_n be the forward D_α-projection of P_{n-1} on $\mathscr{L}_\alpha^{(i_n)}$ with $i_n = n \bmod (m)$ for $n = 1, 2, \ldots$. Then, $P_n \to P^*$.

- Similar to Csiszár 1975.

REVERSE D_α-PROJECTION

Maximum Likelihood Estimation

- Assume $X_1, \ldots, X_n$ i.i.d. samples drawn according to some member of $\mathbb{E} = \{P_\theta : \theta \in \Theta\}$.

Maximum Likelihood Estimation

- Assume $X_1, \ldots, X_n$ i.i.d. samples drawn according to some member of $\mathbb{E} = \{P_\theta : \theta \in \Theta\}$.
- Let $\hat{P}$ be the empirical measure of $X_1, \ldots, X_n$.

Maximum Likelihood Estimation

- Assume $X_1, \ldots, X_n$ i.i.d. samples drawn according to some member of $\mathbb{E} = \{P_\theta : \theta \in \Theta\}$.
- Let $\hat{P}$ be the empirical measure of $X_1, \ldots, X_n$.

$$
\begin{aligned}
\frac{\prod_{i=1}^n P_\theta(X_i)}{\prod_{i=1}^n \hat{P}(X_i)} &= \prod_{a \in \mathcal{A}} \left(\frac{P_\theta(a)}{\hat{P}(a)} \right)^{n\hat{P}(a)} \\
&= \exp\left\{ n \sum_{a \in \mathcal{A}} \hat{P}(a) \log \left(\frac{P_\theta(a)}{\hat{P}(a)} \right) \right\} \\
&= \exp\{-nD(\hat{P} \| P_\theta)\}.
\end{aligned}
$$

Maximum Likelihood Estimation

- Assume $X_1, \ldots, X_n$ i.i.d. samples drawn according to some member of $\mathbb{E} = \{P_\theta : \theta \in \Theta\}$.
- Let $\hat{P}$ be the empirical measure of $X_1, \ldots, X_n$.

$$\frac{\prod_{i=1}^n P_\theta(X_i)}{\prod_{i=1}^n \hat{P}(X_i)} = \prod_{a \in \mathcal{A}} \left(\frac{P_\theta(a)}{\hat{P}(a)} \right)^{n\hat{P}(a)}$$

$$= \exp\left\{ n \sum_{a \in \mathcal{A}} \hat{P}(a) \log \left(\frac{P_\theta(a)}{\hat{P}(a)} \right) \right\}$$

$$= \exp\{-nD(\hat{P} \| P_\theta)\}.$$

- Thus MLE is a **reverse projection**

Maximum Likelihood Estimation

- Assume $X_1, \ldots, X_n$ i.i.d. samples drawn according to some member of $\mathbb{E} = \{P_\theta : \theta \in \Theta\}$.
- Let $\hat{P}$ be the empirical measure of $X_1, \ldots, X_n$.

$$\frac{\prod_{i=1}^n P_\theta(X_i)}{\prod_{i=1}^n \hat{P}(X_i)} = \prod_{a \in \mathcal{A}} \left(\frac{P_\theta(a)}{\hat{P}(a)} \right)^{n\hat{P}(a)}$$
$$= \exp\left\{ n \sum_{a \in \mathcal{A}} \hat{P}(a) \log \left(\frac{P_\theta(a)}{\hat{P}(a)} \right) \right\}$$
$$= \exp\{-nD(\hat{P} \| P_\theta)\}.$$

- Thus MLE is a **reverse projection**
- Reverse projection of Rényi divergence on α-convex sets corresponds to a **robust** version of MLE when some fraction of samples are outliers (Pardo 2006, Basu et al. 2011).

Robust estimation on α-exponential family and the duality

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. samples drawn according to a distribution from $\mathscr{E}_\alpha$, an α-exponential family, and let $\hat{P}$ be its empirical probability measure.

Robust estimation on α-exponential family and the duality

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. samples drawn according to a distribution from $\mathscr{E}_\alpha$, an α-exponential family, and let $\hat{P}$ be its empirical probability measure. Let P^* be the forward D_α-projection of Q on

$$\mathscr{L}_\alpha = \{P \in \mathcal{M} : \sum_a P(a)^\alpha [f_i(a) - \hat{\eta}_i{}^{(n)} Q(a)^{1-\alpha}] = 0, i = 1, \ldots, k\},$$

$$\hat{\eta}_i{}^{(n)} = \frac{\sum_a \hat{P}(a)^\alpha f_i(a)}{\sum_a \hat{P}(a)^\alpha Q(a)^{1-\alpha}}.$$

Robust estimation on α-exponential family and the duality

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. samples drawn according to a distribution from $\mathscr{E}_\alpha$, an α-exponential family, and let $\hat{P}$ be its empirical probability measure. Let P^* be the forward D_α-projection of Q on

$$\mathscr{L}_\alpha = \{P \in \mathcal{M} : \sum_a P(a)^\alpha [f_i(a) - \hat{\eta}_i{}^{(n)} Q(a)^{1-\alpha}] = 0, i = 1, \ldots, k\},$$

$$\hat{\eta}_i{}^{(n)} = \frac{\sum_a \hat{P}(a)^\alpha f_i(a)}{\sum_a \hat{P}(a)^\alpha Q(a)^{1-\alpha}}.$$

- **α-linear and α-exponential families are orthogonal**: *If Supp(P^*) = Supp$(\mathscr{L}_\alpha)$, then $\mathscr{L}_\alpha \cap \mathscr{E}_\alpha = \{P^*\}$, and*

  $$D_\alpha(P\|P_\theta) = D_\alpha(P\|P^*) + D_\alpha(P^*\|P_\theta) \quad \forall P \in \mathscr{L}_\alpha \quad \forall P_\theta \in \mathscr{E}_\alpha.$$

Robust estimation on α-exponential family and the duality

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. samples drawn according to a distribution from $\mathscr{E}_\alpha$, an α-exponential family, and let $\hat{P}$ be its empirical probability measure. Let P^* be the forward D_α-projection of Q on

$$\mathscr{L}_\alpha = \{P \in \mathcal{M} : \sum_a P(a)^\alpha [f_i(a) - \hat{\eta_i}^{(n)}\, Q(a)^{1-\alpha}] = 0, i = 1, \ldots, k\},$$

$$\hat{\eta_i}^{(n)} = \frac{\sum_a \hat{P}(a)^\alpha f_i(a)}{\sum_a \hat{P}(a)^\alpha Q(a)^{1-\alpha}}.$$

- **α-linear and α-exponential families are orthogonal**: *If Supp(P^*) = Supp$(\mathscr{L}_\alpha)$, then $\mathscr{L}_\alpha \cap \mathscr{E}_\alpha = \{P^*\}$, and*

  $$D_\alpha(P \| P_\theta) = D_\alpha(P \| P^*) + D_\alpha(P^* \| P_\theta) \quad \forall P \in \mathscr{L}_\alpha \quad \forall P_\theta \in \mathscr{E}_\alpha.$$

- Thus, $\arg\min_{P_\theta \in \mathscr{E}_\alpha} D_\alpha(\hat{P} \| P_\theta)$

Robust estimation on α-exponential family and the duality

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. samples drawn according to a distribution from $\mathscr{E}_\alpha$, an α-exponential family, and let $\hat{P}$ be its empirical probability measure. Let P^* be the forward D_α-projection of Q on

$$\mathscr{L}_\alpha = \{P \in \mathcal{M} : \sum_a P(a)^\alpha [f_i(a) - \hat{\eta}_i^{(n)} Q(a)^{1-\alpha}] = 0, i = 1, \ldots, k\},$$

$$\hat{\eta}_i^{(n)} = \frac{\sum_a \hat{P}(a)^\alpha f_i(a)}{\sum_a \hat{P}(a)^\alpha Q(a)^{1-\alpha}}.$$

- **α-linear and α-exponential families are orthogonal**: *If Supp*(P^*) = *Supp*$(\mathscr{L}_\alpha)$, *then* $\mathscr{L}_\alpha \cap \mathscr{E}_\alpha = \{P^*\}$, *and*

  $$D_\alpha(P \| P_\theta) = D_\alpha(P \| P^*) + D_\alpha(P^* \| P_\theta) \quad \forall P \in \mathscr{L}_\alpha \quad \forall P_\theta \in \mathscr{E}_\alpha.$$

- Thus, $\arg \min_{P_\theta \in \mathscr{E}_\alpha} D_\alpha(\hat{P} \| P_\theta) = P^*$

Robust estimation on α-exponential family and the duality

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. samples drawn according to a distribution from $\mathscr{E}_\alpha$, an α-exponential family, and let $\hat{P}$ be its empirical probability measure. Let P^* be the forward D_α-projection of Q on

$$\mathscr{L}_\alpha = \{P \in \mathcal{M} : \sum_a P(a)^\alpha [f_i(a) - \hat{\eta}_i^{(n)} Q(a)^{1-\alpha}] = 0, i = 1, \ldots, k\},$$

$$\hat{\eta}_i^{(n)} = \frac{\sum_a \hat{P}(a)^\alpha f_i(a)}{\sum_a \hat{P}(a)^\alpha Q(a)^{1-\alpha}}.$$

- **α-linear and α-exponential families are orthogonal**: *If Supp(P^*) = Supp$(\mathscr{L}_\alpha)$, then $\mathscr{L}_\alpha \cap \mathscr{E}_\alpha = \{P^*\}$, and*

  $$D_\alpha(P \| P_\theta) = D_\alpha(P \| P^*) + D_\alpha(P^* \| P_\theta) \quad \forall P \in \mathscr{L}_\alpha \quad \forall P_\theta \in \mathscr{E}_\alpha.$$

- Thus, $\arg\min_{P_\theta \in \mathscr{E}_\alpha} D_\alpha(\hat{P} \| P_\theta) = P^* = \arg\min_{P \in \mathscr{L}_\alpha} D_\alpha(P \| Q)$.

Summary

- Sufficient condition for the existence of forward projection on α-convex sets
- Pythagorean equality on α-linear family for $\alpha > 1$
- Convergence of iterated projection on an intersection of α-linear families
- Form of forward projection on α-linear family
- Orthogonality of α-linear and α-exponential families
- Full version: `http://arxiv.org/abs/1512.02515`.