Received 24/02/11

λ-Renyi Relative Entropy Maximization

B.L.S. PRAKASA RAO [1]
University of Hyderabad, Hyderabad, India

Abstract: It is shown that every probability density fnction with respect to a σ-finite measure μ on a measure space $(\Omega, \mathcal{F}, \mu)$ is the unique maximizer of λ-Renyi relative entropy in an appropriate class.

Key words: λ-Renyi entropy power, λ-Renyi relative entropy; λ-Renyi relative entropy maximization.

AMS 2010 Subject Classification: Primary 62E10.

1 Introduction

In a recent paper, Athreya (2009) has shown that every probability density function is the unique maximizer of relative entropy in an appropriate class of probability density functions. We extend this result to λ-Renyi entropy studied by Lutwak et al.(2005). Lutwik et al. (2004) have considered the problem of maximizing the λ-Renyi entropy for probability density functions on R^n. We prove the result for general measure spaces.

2 Preliminaries

Entropy :

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space where μ is σ-fintie. A nonnegative $\mathcal{F}$-measurable function such that

$$\int_\Omega f \, d\mu = 1$$

is called a probability density function. For such an f, let

$$P_f(A) = \int_A f \, d\mu, A \in \mathcal{F}.$$

[1] blsprao@gmail.com

Then P_f is a probaility measure on the measurable space $(\Omega, \mathcal{F})$. The *Shannon entropy* of the probability measure P_f with respect to the measure μ is defined by

$$h(f) \equiv - \int_\Omega f \log f \, d\mu \tag{2. 1}$$

provided the integral exists. We define $0 \log 0 = 0$.

For $\lambda > 0$, the λ-*Renyi entropy power* of the probability measure P_f with respect to the measure μ is defined to be

$$\begin{aligned} N_\lambda(f) &= [\int_\Omega f^\lambda \;\; d\mu]^{1/(1-\lambda)} \text{ if } \lambda \neq 1, \\ &= e^{h(f)} \text{ if } \lambda = 1 \end{aligned} \tag{2. 2}$$

provided that the integral defined above exists. The λ-*Renyi entropy* of the probability measure P_f with respect to the measure μ is defined to be

$$h_\lambda(f) = \log N_\lambda(f). \tag{2. 3}$$

Relative Entropy :

Given two probability densities f and g with respect to the measure μ, their *relative Shannon entropy* or *Kullback-Leibler information* is defined by

$$h_1(f, g) = \int_\Omega f \log(\frac{f}{g}) \;\; d\mu \tag{2. 4}$$

provided the integral exists.

λ-Renyi Relative Entropy :

For $\lambda > 0$, the *relative* λ-*Renyi entropy power* of the probability measures P_f and P_g with respect to the measure μ is defined to be

$$\begin{aligned} N_\lambda(f, g) &= \frac{[\int_\Omega g^{\lambda-1} f \;\; d\mu]^{1/(1-\lambda)} [\int_\Omega g^\lambda \;\; d\mu]^{1/\lambda}}{[\int_\Omega f^\lambda \;\; d\mu]^{1/\lambda(1-\lambda)}} \text{ if } \lambda \neq 1, \\ &= e^{h_1(f,g)} \text{ if } \lambda = 1 \end{aligned} \tag{2. 5}$$

provided the integrals on the right side exist. The λ-*Renyi relative entropy* of the probability measures P_f and P_g with respect to the measure μ is defined to be

$$h_\lambda(f,g) = \log N_\lambda(f,g). \tag{2. 6}$$

The following Theorem is due to Lutwak et al. (2005)

Theorem 2.1: If f and g are probabiliy densities with respect to a measure μ such that $h_\lambda(f), h_\lambda(g)$ and $h_\lambda(f,g)$ are finite, then

$$h_\lambda(f,g) \geq 0 \tag{2. 7}$$

equality holds if and only if $f = g$ a.e $[\mu]$ for any $\lambda > 0$.

The theorem stated above is well known if $\lambda = 1$ and Ω is the real line and μ is the Lebesgue measure (cf. Cover and Thomas (1999), p.234). For $\lambda > 0, \lambda \neq 1$, the theorem was proved by Lutwak et al. (2005) applying the Holder's inequality when Ω is the real line and μ is the Lebesgue measure. The same arguments will prove Theorem 2.1 when f and g are probability densities with respect to a measure μ on a measurable space $(\Omega, \mathcal{F})$ by applying the Holder's inequality for a measure space $(\Omega, \mathcal{F}, \mu)$. We give the proof here for completeness.

Proof : Case (i) Suppose $\lambda = 1$. Following Athreya (2009) and observing that the function $h(x) = x - 1 - \log x$ is nonnegative and has a unique minimum at $x = 1$, it follows that

$$f(\omega) \log g(\omega) - f(\omega) \log f(\omega) \leq g(\omega) - f(\omega), \omega \in \Omega.$$

Integrating with repect to the measure μ, we get that

$$\int_\Omega f(\omega) \log g(\omega) \mu(d\omega) - \int_\Omega f(\omega) \log f(\omega) \mu(d\omega) \leq \int_\Omega [g(\omega) - f(\omega)] \mu(d\omega) = 0$$

since f and g are probability density functions with respect to the measure μ. This proves that

$$h_1(f,g) \geq 0$$

equality occuring if and only if $f = g$ a.e. $[\mu]$.

Case (ii) Suppose $\lambda > 1$. Applying the Holder's inequality, we get that

$$\int_\Omega g^{\lambda-1} f d\mu \leq [\int_\Omega g^\lambda d\mu]^{\frac{\lambda-1}{\lambda}} [\int_\Omega f^\lambda d\mu]^{\frac{1}{\lambda}}$$

equality occuring if and only if $f = g$ a.e. $[\mu]$ since f and g are probability density functions with respect to the measure μ. Hence $N_\lambda(f, g) \geq 1$ or equivalently $h_\lambda(f, g) \geq 0$ for any two probability density functions with repect to the measure μ equality occuring if and $f = g$ a.e. $[\mu]$.

Case (iii) Suppose $\lambda < 1$. Then

$$\begin{aligned} \int_\Omega f^\lambda d\mu &= \int_\Omega (g^{\lambda-1} f)^\lambda g^{\lambda(1-\lambda)} d\mu \\ &\leq [\int_\Omega (g^{\lambda-1} f) d\mu]^\lambda [\int_\Omega g^\lambda d\mu]^{(1-\lambda)} \end{aligned} \tag{2. 8}$$

equality occuring if and only if $f = g$ a.e. $[\mu]$ since f and g are probability density functions with respect to the measure μ. Hence $N_\lambda(f, g) \geq 1$ or equivalently $h_\lambda(f, g) \geq 0$ for any two probability density functions with repect to the measure μ equality occuring if and $f = g$ a.e. $[\mu]$.

As a consequence of Theorem 2.1, we get that

$$[\int_\Omega g^{\lambda-1} f \;\; d\mu]^{1/(1-\lambda)} [\int_\Omega g^\lambda \;\; d\mu]^{1/\lambda} \geq [\int_\Omega f^\lambda \;\; d\mu]^{1/[\lambda(1-\lambda)]}$$

whenever $\lambda > 0, \lambda \neq 1$, and equality occurs if and only if $f = g$ a.e. $[\mu]$.

Let f_0 be a probability density function with repect to the measure μ such that $\gamma = [\int_\Omega f_0^\lambda d\mu]^{\frac{1}{\lambda(1-\lambda)}} < \infty$. Let

$$\zeta_\gamma \equiv \{g : \int_\Omega g d\mu = 1 \text{ and } \int_\Omega g f_0^{\lambda-1} \;\; d\mu = \gamma^{\lambda(1-\lambda)}\}.$$

For any $g \in \zeta_\gamma$,

$$H_\lambda(g) \equiv [\int_\Omega g^\lambda \;\; d\mu]^{1/\lambda(1-\lambda)} \leq [\int_\Omega g f_0^{\lambda-1} \;\; d\mu]^{1/(1-\lambda)} [\int_\Omega f_0^\lambda \;\; d\mu]^{1/\lambda} = \gamma = [\int_\Omega f_0^\lambda d\mu]^{\frac{1}{\lambda(1-\lambda)}} < \infty.$$

We have the following theorem generalizing Corollary 1 in Athryea (2009).

Theorem 2.2:

$$\sup\{H_\lambda(g); g \in \zeta_\lambda\} = H_\lambda(f_0)$$

and f_0 is the unique maxiimizer.

3 Examples

For the case $\lambda = 1$, Athreya (2009) has discussed several examples as applications of his result. He showed that, in the class of all probability density functions f that satisfy the conditions

$$\int_\Omega f h_i d\mu = \gamma_i, i = 1, 2, \ldots, k,$$

the maximizer of entropy is a probability density f_0 that is proportional to the function

$$\exp\{\sum_{i=1}^{k} c_i h_i\}$$

for some choice of $c_i, i = 1, 2, \ldots, k$. Suppose $0 < \lambda \neq 1$. Let h be a real valued measurable function defined on the measurable space $(\Omega, \mathcal{F})$, such that

$$\psi(c) = \int_\Omega e^{ch} d\mu < \infty; \ \int_\Omega e^{ch\lambda} d\mu < \infty$$

for some real γ and c such that

$$f_0 = \frac{e^{ch}}{\psi(c)}$$

is a probability density fnction and

$$\int_\Omega f_0^\lambda d\mu = \gamma^{\lambda(1-\lambda)}.$$

We ilustrate our results on λ-Renyi entropy by two examples.

Example 1: Suppose $\Omega = \{1, 2, \ldots, N\}$ where μ is the counting measure. Let $h \equiv 1$. Suppose $\gamma > 0$ such that γ^λ is a positive integer. Let f_0 be the uniform distribution on the integers $\{1, 2, \ldots, N\}$ where $N = \gamma^\lambda$. It is easy to check that, f_0 is the unique maximizer among all discrete distributions with support contained in Ω with $N = \gamma^\lambda$.

Example 2: Suppose $\Omega = R$, $\mathcal{F}$ is the Borel σ-algebra and μ is the Lebesgue measure. Suppose $\lambda > 1$. Let $h(x) = x^2$. It is easy to see that $\int_R e^{cx^2} dx < \infty$ for any $c < 0$. Let

$$f_0(x) = \frac{e^{cx^2}}{\psi(c)}$$

where $\psi(c) = \int_R e^{cx^2} dx$. Observe that f_0 is the Gaussian probability density function with mean zero and variance (-c/2). For any $\gamma > 0$, the density f_0 is the unique maximizer among all probability density functions g satisfying the confition

$$\int_R g(x) e^{(\lambda-1)cx^2} \ dx = [\psi(c)]^{\lambda-1} \gamma^{\lambda(1-\lambda)}.$$

Suppose $\lambda < 1$. Let $h(x) = -x^2$. It is easy to see that $\int_R e^{-cx^2} dx < \infty$ for any $c > 0$. Let

$$f_0(x) = \frac{e^{-cx^2}}{\psi(c)}$$

where $\psi(c) = \int_R e^{-cx^2} dx$. Observe that f_0 is the Gaussian probability density function with mean zero and variance (c/2). For any $\gamma > 0$, the density f_0 is the unique maximizer among all probability density functions g satisfying the confition

$$\int_R g(x) e^{-(\lambda-1)cx^2} \; dx = [\psi(c)]^{\lambda-1} \gamma^{\lambda(1-\lambda)}.$$

References :

Athreya, K.B. (2009) Entropy maximization, *Proc. Indian Acad. Sci. (Math. Sci.)*, **119**, 531-539.

Cover, T.M and Thomas, J.A. (1991) *Elements of Information Theory*, Wiley, New York.

Lutwak, E., Yang, D. and Zhang, G. (2004) Moment-entropy inequalities, *Ann. Probab.*, **32**, 757-774.

Lutwak, E., Yang, D. and Zhang, G. (2005) Cramer-Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information, *IEEE Transactions on Information Theory*, **51**, 473-478.