Yokohama Mathematical
Journal Vol. 64, 2018

ON RELATIVE RÉNYI ENTROPY CONVERGENCE OF THE MAX DOMAIN OF ATTRACTION

By

Ali Saeb

(Received November 30, 2017; Revised April 16, 2018)

Abstract. The differential Rényi entropy and Rényi divergence are perhaps the two fundamental quantities in information theory and its applications. The differential Rényi entropy and Rényi divergence are generalized form of the differential Shannon entropy and Kullback Leibler divergence, respectively. In this article, we prove that, if differential Rényi entropy of max domain of attraction laws exists then, Rényi divergence between the density function of linearly normalized partial maxima of iid random variables and density function of max stable converges to zero. The order of differential Rényi entropy has important role to solve the problem.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (iid) random variables (rvs) with common distribution function (df) F and density function f. The differential Rényi entropy of order β is defined as

$$h_\beta(F) = \frac{\beta}{1-\beta} \log\left(\|f(x)\|_\beta\right), \tag{1.1}$$

where, $\|f(x)\|_\beta = \left(\int_A (f(x))^\beta\, dx\right)^{\frac{1}{\beta}}$ for $A = \{x \in \mathbb{R} : f(x) > 0\}$ and $0 < \beta < \infty$, $\beta \neq 1$, (Rényi, 1961). It is one of a family of functionals for quantifying the diversity, uncertainty or randomness of a system. Here, we assume that the goodness of the model F is assessed in terms of the closeness as a probability distribution to true distribution G. As a measure of this closeness, for any simple order β, Rényi divergence (relative Rényi entropy) of order β of F and G is defined as

$$D_\beta(F\|G) = \frac{1}{\beta-1} \log\left(\int_A (f(x))^\beta (g(x))^{1-\beta} dx\right). \tag{1.2}$$

2010 Mathematics Subject Classification: 60F10
Key words and phrases: Rényi entropy, Rényi divergence, Max stable laws, Max domain of attraction

It is straight forward to justify this as an extension by continuity; as β tends to 1, the relative Rényi entropy tends to Kullback Leibler divergence. Gilardoni (2010) shows that Rényi divergence is related to the total variation distance $V(F,G) = \int_A |f(x) - g(x)|dx$ by a generalization of Pinskers inequality:

$$\frac{\beta}{2}V^2(F,G) \leq D_\beta(F\|G), \text{ for } \beta \in (0,1].$$

Gibbs and Su (2002) study the differential Rényi entropy for $\beta = 1/2$ is function of squared Hillinger distance $Hel^2(F,G) = \int_A (f^{1/2}(x) - g^{1/2}(x))^2 dx$ which is $D_{1/2}(F\|G) = -2\log\left(1 - \frac{Hel^2(F,G)}{2}\right)$. Similarly, for $\beta = 2$ it is given by $D_2(F\|G) = \log(1 + \chi^2(F,G))$, where, $\chi^2(F,G) = \int_A \frac{(f(x)-g(x))^2}{g(x)}dx$. Therefore, by using these equations and $\log(t) \leq t - 1$, imply that

$$Hel^2(F,G) \leq D_{1/2}(F\|G) \leq D_1(F\|G) \leq D_2(F\|G) \leq \chi^2(F,G).$$

The properties of Rényi divergence is studied by van Erven and Harremoës (2014) and they show that Rényi divergence is nondecreasing in its order. The idea of tracking the central limit theorem using Shannon entropy goes back to Linnik (1959) and Shimizu (1975), who used it to give a particular proof of the central limit theorem. Brown (1982), Barron (1986) and Takano (1987) discuss the central limit theorem with convergence in the sense of Shannon entropy and relative entropy. Artstein et al.(2004) and Johnson and Barron (2004) obtained the rate of convergence under some conditions on the density. Johnson (2006) is a good reference to the application of information theory to limit theorems, especially the central limit theorem. Cui and Ding (2010) show that the convergence of Rényi entropy of the normalized sums of iid rvs and obtain the corresponding rate of convergence.

Extremes. The limit laws of linearly normalized partial maxima $M_n = \max\{X_1, \ldots, X_n\}$ of iid rvs $X_1, X_2, \ldots$, with common df F, namely,

$$\lim_{n\to\infty} \Pr(M_n \leq a_n x + b_n) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \quad x \in \mathcal{C}(G), \tag{1.3}$$

where, $a_n > 0$, $b_n \in \mathbb{R}$, are norming constants, G is a nondegenerate distribution function, $\mathcal{C}(G)$ is the set of all continuity points of G, are called max stable laws. If, for some nondegenerate distribution function G, a distribution function F satisfies (1.3) for some norming constants $a_n > 0$, $b_n \in \mathbb{R}$, then we say that F belongs to the max domain of attraction of G under linear normalization and denote it by $F \in \mathcal{D}(G)$. Limit distribution functions G satisfying (1.3) are well known extreme value types of distributions, or max stable laws. Fisher and

Tippett (1928) show that max stable laws can only be one of three types, namely,

$$\text{the Fréchet law:}\Phi_\alpha(x) = \begin{cases} 0, & x < 0, \\ \exp(-x^{-\alpha}), & x \geq 0; \end{cases}$$

$$\text{the Weibull law: } \Psi_\alpha(x) = \begin{cases} \exp(-|x|^\alpha), & x < 0, \\ 1, & x \geq 0; \end{cases}$$

$$\text{and the Gumbel law: } \Lambda(x) = \exp(-\exp(-x)); \qquad x \in \mathbb{R};$$

$\alpha > 0$ being a parameter. Criteria for $F \in \mathcal{D}(G)$ are well known (see, for example, Galambos, 1987; Resnick, 1987; Embrechts et al., 1997).

Under von Mises type conditions (see, theorems B.2 and B.3), de Haan and Resnick (1982) and Sweeting (1985) prove the density of the normalized maximum converges to density of appropriate extreme value distribution in space L_β. They show that, if the differential Rényi entropy of original distribution of max domain of attraction exist then Rényi entropy convergence of max domain attraction and max stable laws hold. In recent research works on information theory and extreme value theory, Ravi and Saeb (2013), study the convergence theory of Shannon entropy for max domain of attraction. Saeb (2014) investigate the rate of convergence of Rényi entropy in max domain of attraction.

In this article, our main interest is to investigate conditions under which the relative Rényi entropy distance between the density function of the normalized partial maxima of iid rvs and density function of max stable converges to zero. In the next section, we give our main results. Rényi entropy of the extreme value distributions are illustrated in the appendix A and appendix B containing results used in this article.

Throughout the manuscript, we shall denote the right extremity of F by $r(F) = \sup\{x : F(x) < 1\} \leq \infty$ and left extremity of F is $l(F) = \inf\{x : F(x) > 0\} \leq \infty$, we define the inverse of F as $F^{\leftarrow}(x) = \inf\{s : F(s) \geq x\}$, survival function is $\bar{F}(\cdot) = 1 - F(\cdot)$. Also, we employ the notation, $\phi_\alpha(x) = \alpha x^{-(\alpha+1)} e^{-x^{-\alpha}},\ x > 0$; is the density of Fréchet, $\psi_\alpha(x) = \alpha |x|^{\alpha-1} e^{-|x|^\alpha},\ x < 0$; is the density of Weibull, and $\lambda(x) = e^{-x} e^{-e^{-x}},\ x \in \mathbb{R}$ is the density of the Gumbel. The distribution and density function of the normalized maximum are $G_n(x) = F^n(a_n x + b_n)$ and $g_n(x) = n a_n f(a_n x + b_n) F^{n-1}(a_n x + b_n)$, respectively.

2. Main Results

2.1 Fréchet case.

Suppose $F \in \mathcal{D}(\Phi_\alpha)$, and $1 - F$ is regularly varying with index $-\alpha$ (written $\bar{F} \in RV_{-\alpha}$) so that $\lim_{n\to\infty} \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} = x^{-\alpha}$, and from (1.3), $\lim_{n\to\infty} F^n(a_n x) =$

$\Phi_\alpha(x),\ x>0$, with $a_n = F^{\leftarrow}(1-\frac{1}{n}),\ n \geq 1$ and $b_n = 0$. Now, we switch to function $a(n)$. Since $\bar{F} \in RV_{-\alpha}$, Resnick (1987) shows $a(n) \in RV_{\frac{1}{\alpha}}$. This means that for any fixed $x>0$, the function $a(n)$, consider as a function of n, is also in $RV_{\frac{1}{\alpha}}$.

We note that, the proof of the main theorems are different from the proof for the relative Rényi entropy of the normalized sums of iid rvs. In our proofs, the properties of normalized partial maxima such as, von Mises condition and density convergence, plays an important role. By considering this assumptions, we now justify the relative Renyi entropy convergence.

Theorem 2.1. *Let $X_1, X_2, \ldots$ be iid random variables with df $F \in \mathcal{D}(\Phi_\alpha)$ which is absolutely continuous with density function f. If f is nonincreasing function, $h_\beta(F) < \infty$ for $0<\beta<\frac{\alpha}{1+\alpha}$, and left extremity of F, $l(F) \neq 0$ then*

$$\lim_{n\to\infty} D_\beta(G_n \| \Phi_\alpha) = 0.$$

Proof. From definition (1.2)

$$D_\beta(G_n\|\Phi_\alpha) = \frac{1}{\beta-1}\log\left[I_1(n,v)+I_2(n,v)+I_3(n,v)\right]. \tag{2.1}$$

where, $I_1(n,v)=\int_v^\infty (\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta\,dx$, and $I_2(n,v)=\int_{-\infty}^{v^{-1}}(\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta\,dx$, and $I_3(n,v)=\int_{v^{-1}}^{v}(\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta\,dx$. It is enough to show that

$$\lim_{v\to\infty}\lim_{n\to\infty}\left(I_1(n,v)+I_2(n,v)\right)=0.$$

We set,

$$0<I_1(n,v)\leq L(n;\beta)\int_v^\infty\left(\frac{f(a_nx)}{f(a_n)}\right)^\beta(\phi_\alpha(x))^{1-\beta}dx,$$

where, $L(n;\beta)=\left(\frac{a_nf(a_n)n\bar{F}(a_n)}{\bar{F}(a_n)}\right)^\beta$. From (B. 2) and the fact that $n\bar{F}(a_n)=1$ we have $L(n;\beta)\to\alpha^\beta$, as $n\to\infty$. Under von Mises conditions (Theorem B.2-(b)), f is nonincreasing function and $\left(\frac{f(a_nx)}{f(a_n)}\right)^\beta\leq 1$ for $x\geq 1$. We have

$$I_1(n,v)\leq L(n;\beta)\left(\frac{f(a_nv)}{f(a_n)}\right)^\beta\int_v^\infty(\phi_\alpha(x))^{1-\beta}dx.$$

From Lemma A.1-(i), $\int_{\mathbb{R}}(\phi_\alpha(x))^{1-\beta}dx<\infty$, for $0<\beta<\frac{\alpha}{\alpha+1}$. Hence,

$$\lim_{v\to\infty}\lim_{n\to\infty}I_1(n,v)=0. \tag{2.2}$$

Now, we choose ξ_n by $-\log F(\xi_n) \simeq n^{-1/2}$, and $t_n = \frac{\xi_n}{a_n}$. If $\frac{\xi_n}{a_n} \to c > 0$ then $n^{1/2} \simeq -n(\log F(t_n\, a_n)) \to c^{-\alpha}$ and this contradicts the fact that $n^{1/2} \to \infty$. Therefore, $t_n \to 0$ as $\xi_n \to \infty$ for large n. Setting,

$$\begin{aligned} I_2(n,u) &= \int_{l(F)/a_n}^{t_n} (\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta dx + \int_{t_n}^{v^{-1}} (\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta dx, \\ &= I_{21}(n) + I_{22}(n,v), \end{aligned} \tag{2.3}$$

where, $l(F) \neq 0$. We write,

$$\begin{aligned} I_{21}(n) &= \int_{l(F)}^{\xi_n} (nF^{n-1}(s)f(s))^\beta (\phi_\alpha(sa_n^{-1}))^{1-\beta}\, ds, \quad \text{(where, } a_n x = s\text{)}, \\ &\leq (nF^{n-1}(\xi_n))^\beta \left(\left(\frac{l(F)}{a_n}\right)^{-\alpha-1} \alpha \Phi_\alpha(t_n)\right)^{1-\beta} \int_0^\infty (f(s))^\beta\, ds, \text{ if } 0<\beta<1, \\ &\simeq \left(\frac{n a_n^{(\alpha+1)(1-\beta)/\beta}}{\exp\left((n-1)n^{-1/2}\right)}\right)^\beta (c_\alpha(l(F))\Phi_\alpha(t_n))^{1-\beta} \int_0^\infty (f(s))^\beta\, ds, \end{aligned} \tag{2.4}$$

where, $c_\alpha(l(F)) = \alpha l(F)^{-\alpha-1}$. Since, $a(n) \in RV_{1/\alpha}$ and combining this with (B. 1) we have

$$a(n) = c(n) \exp\left(\int_N^n \rho(t)t^{-1}dt\right),$$

where, $\rho(n) \to \frac{1}{\alpha}$ and $c(n) \to c$ as $n \to \infty$. On the other hand, for $n > N$ given $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $\rho(n) < \frac{1+\epsilon_1}{\alpha}$ and $c(n) < (1+\epsilon_2)c = c'$ then

$$\begin{aligned} a(n) &= c(n) \exp\left(\int_N^n \rho(t)t^{-1}dt\right), \\ &< \left(\frac{n}{N}\right)^{\frac{1+\epsilon_1}{\alpha}} c', \end{aligned}$$

From (2.4) we write,

$$0 < I_{21}(n) \leq \left(\frac{n(c'\, n^{\frac{(1+\epsilon_1)}{\alpha}} N^{-\frac{(1+\epsilon_1)}{\alpha}})^{(\alpha+1)(1-\beta)/\beta}}{\exp(n^{1/2} - n^{-1/2})}\right)^\beta (c_\alpha(l(F))\Phi_\alpha(t_n))^{1-\beta} \int_0^\infty (f(s))^\beta\, ds.$$

If $\int_0^\infty (f(s))^\beta\, ds < \infty$ and $l(F) \neq 0$ then,

$$\lim_{n\to\infty} I_{21}(n) = 0. \tag{2.5}$$

Next, we write $I_{22}(n,v) = \int_{t_n}^{v^{-1}} (\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta\, dx$. Here, we need to find the upper bound for g_n ultimately. Now $U = 1/(1-F) \in RV_\alpha$ so that $U(a_n)/U(a_n x)$

$= U(a_n x x^{-1})/U(a_n x)$ we get from Theorem B.1 for given $0 < \epsilon_1 < \alpha$, and large n we have

$$(1-\epsilon_1)x^{-(\alpha-\epsilon_1)} \leq U(a_n)/U(a_n x) \leq (1+\epsilon_1)x^{-(\alpha+\epsilon_1)}.$$

Since $n\bar{F}(a_n x) \simeq \bar{F}(a_n x)/\bar{F}(a_n) = U(a_n)/U(a_n x)$ we have for large n

$$(1-\epsilon_1)x^{-(\alpha-\epsilon_1)} \leq n\bar{F}(a_n x) \leq (1+\epsilon_1)x^{-(\alpha+\epsilon_1)}.$$

From (B. 2), for given $\epsilon_2 > 0$ we write $f(a_n x) \leq \frac{\alpha+\epsilon_2}{a_n x}\bar{F}(a_n x)$ ultimately. Hence, for sufficiently large n such that $n > N$ we have

$$\begin{aligned} g_n(x) &< (\alpha+\epsilon_2)n\bar{F}(a_n x)x^{-1}\exp\left(-n\bar{F}(a_n x)\right), \\ &< (\alpha+\epsilon_2)(1+\epsilon_1)x^{-1-\alpha-\epsilon_1}\exp\left(-(1-\epsilon_1)x^{-(\alpha-\epsilon_1)}\right), \end{aligned}$$

by choosing suitable ϵ_1, ϵ_2 and α, we define

$$k(x) = c\alpha' x^{-1-\alpha'}\exp\left(-cx^{-\alpha'}\right). \tag{2.6}$$

where, $c, \alpha' > 0$. Hence, $g_n(x) < k(x)$, for large n, we obtain

$$0 < I_{22}(n,v) < \int_0^{v^{-1}} (\phi_\alpha(x))^{1-\beta}(k(x))^\beta dx.$$

Since $\int_0^\infty (\phi_\alpha(x))^{1-\beta}(k(x))^\beta dx < \infty$ so that,

$$\lim_{v\to\infty}\lim_{n\to\infty} I_{22}(n,v) = 0. \tag{2.7}$$

From, (2.3), (2.5) and (2.7) we have

$$\lim_{v\to\infty}\lim_{n\to\infty} I_2(n,v) = 0. \tag{2.8}$$

Finally, $I_3(n,v) = \int_{v^{-1}}^v (\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta\, dx$. Since, for large n, $g_n(x) < k(x)$, and it is well known that, $\int_{v^{-1}}^v (\phi_\alpha(x))^{1-\beta}(k(x))^\beta dx < \infty$, by using Theorem B.4-i, $\lim_{n\to\infty} g_n(x) = \phi_\alpha(x)$, locally uniformly convergence in $x \in [v^{-1}, v]$ by dominated convergence theorem,

$$\lim_{v\to\infty}\lim_{n\to\infty}\int_{v^{-1}}^v (\phi_\alpha(x))^{1-\beta}(g_n(x))^\beta\, dx = \int_0^\infty \phi_\alpha(x)\, dx = 1. \tag{2.9}$$

From (2.1), (2.2), (2.8) and (2.9) imply,

$$\lim_{n\to\infty} D_\beta(G_n \| \Phi_\alpha) = 0.$$

□

2.2 Weibull case.

Suppose $F \in \mathcal{D}(\Psi_\alpha)$, a necessary and sufficient condition for the existence of $a_n > 0$, and $b_n = r(F) < \infty$ such that (1.3) holds and $F^*(x) = F(r(F) - 1/x)$ is in the domain of attraction of Φ_α. Results for Rényi entropy convergence in Ψ_α may be derived directly from the corresponding results for $F^* \in \mathcal{D}(\Phi_\alpha)$. So, we concentrate on Ψ_α by using the relationship between Fréchet and Weibull.

Theorem 2.2. *Let $X_1, X_2, \ldots$ be iid random variables with df $F \in \mathcal{D}(\Psi_\alpha)$ which is absolutely continuous with density function f. If f is nonincreasing function, $h_\beta(F) < \infty$ for $0 < \beta < \frac{\alpha}{1+\alpha}$, and left extremity of F, $l(F) \neq 0$ then*

$$\lim_{n\to\infty} D_\beta(G_n \| \Psi_\alpha) = 0.$$

Proof. Suppose $F \in \mathcal{D}(\Psi_\alpha)$ iff $r(F) < \infty$ and $\bar{F}(r(F) - x^{-1}) \in RV_{-\alpha}$, is regularly varying. In this case, we may set $\tau_n = F^{\leftarrow}(1 - \frac{1}{n})$ and $a_n = r(F) - \tau_n$, and then

$$\lim_{n\to\infty} F^n(a_n x + r(F)) = \Psi_\alpha(x),$$

for $x \in \mathbb{R}$. Now, from definition (1.2) we write,

$$\begin{aligned} D_\beta(G_n \| \Psi_\alpha) &= \frac{1}{\beta - 1} \log \left(\int_A \left(g_n^*(y)\right)^\beta \left(\phi_\alpha\left(y\right)\right)^{1-\beta} dy \right), \quad \text{(setting, } y = -1/x) \\ &= D_\beta(G_n^* \| \Phi_\alpha). \end{aligned}$$

where, $G_n^*(y) = F^n(r(F) - a_n/y)$ and $g_n^*(y) = \frac{n a_n}{y^2} f(r(F) - a_n/y) F^{n-1}(r(F) - a_n/y)$. Under conditions Theorem 2.1, if $h_\beta(F) < \infty$, $l(F) \neq 0$ and f is nonincreasing function then

$$\lim_{n\to\infty} D_\beta(G_n \| \Psi_\alpha) = 0.$$

□

2.3 Gumbel case.

Suppose $F \in \mathcal{D}(\Lambda)$ and $\bar{F}$ is Γ varying so that $\lim_{n\to\infty} \frac{\overline{F}(b_n + x a_n)}{\overline{F}(b_n)} = e^{-x}$, and from (1.3) $\lim_{n\to\infty} F^n(a_n x + b_n) = \Lambda(x),\ x \in \mathbb{R}$, where the function $a_n = u(b_n) = \int_{b_n}^{r(F)} \overline{F}(s) ds / \overline{F}(b_n)$ is called an auxiliary function and $b_n = F^{\leftarrow}(1 - \frac{1}{n})$. With this case in mind we state and proof the following lemata which will be used subsequently.

Lemma 2.1. *Suppose $F \in \mathcal{D}(\Lambda)$. For $x \in \mathbb{R}$, there exists a large N such that for $n > N$ and $\epsilon, c > 0$,*

$$n(1 - F(a_n x + b_n)) \leq c(1 + \epsilon\,|x|)^{-\epsilon^{-1}};$$

Proof. Case $x \geq 0$: For sufficient large n, and $\epsilon > 0$ we have $(1-\epsilon) < n\bar{F}(b_n) < (1+\epsilon)$. From Theorem B.6, (recall $a_n = u(b_n)$), we write

$$\begin{aligned} n(1 - F(a_n x + b_n)) &\leq (1+\epsilon)\frac{1 - F(b_n + a_n x)}{1 - F(b_n)}, \\ &= (1+\epsilon)\frac{c(b_n + a_n x)}{c(b_n)} \exp\left(-\int_{b_n}^{b_n + a_n x} \frac{dy}{u(y)}\right), \\ &\leq (1+\epsilon)^2 \exp\left(-\int_0^x \frac{u(b_n)ds}{u(b_n + a_n s)}\right), \qquad (2.10) \end{aligned}$$

where, $y = b_n + a_n s$ and $\lim_{t\to\infty} \frac{c(t+xu(t))}{c(t)} = 1$. Since, $\lim_{t\to r(F)} u'(t) = 0$, for sufficient large n such that $|u'(a_n t + b_n)| \leq \epsilon$, and

$$\begin{aligned} \left|\frac{u(a_n x + b_n)}{u(b_n)} - 1\right| &= \left|\int_{b_n}^{a_n x + b_n} \frac{u'(s)}{u(b_n)} ds\right|, \\ &\leq \int_0^x |u'(a_n t + b_n)|\, dt, \quad \text{where } s = a_n t + b_n, \\ &\leq \epsilon\, x. \end{aligned}$$

Consequently,

$$\frac{u(b_n)}{u(a_n x + b_n)} > \frac{1}{1 + \epsilon\, x}. \qquad (2.11)$$

From (2.10) and (2.11) for large n and $\epsilon > 0$ we have

$$\begin{aligned} n(1 - F(a_n x + b_n)) &\leq (1+\epsilon)^2 \exp\left(-\int_0^x \frac{ds}{1 + \epsilon s}\right), \\ &< c(1 + \epsilon x)^{-\epsilon^{-1}}. \end{aligned}$$

where, $c > 0$. With similar argument, for the second statement $x < 0$ and sufficient large n

$$\frac{u(b_n)}{u(a_n x + b_n)} < \frac{1}{1 + \epsilon\, |x|}. \qquad (2.12)$$

where,

$$\begin{aligned} \left|1 - \frac{u(a_n x + b_n)}{u(b_n)}\right| &\leq \int_x^0 |u'(a_n t + b_n)|\, dt, \quad (s = a_n t + b_n), \\ &\leq \epsilon\, |x|. \end{aligned}$$

From (2.10) and (2.12) we get,

$$\begin{aligned} n(1 - F(a_n x + b_n)) &\leq (1-\epsilon)^2 \exp\left(-\int_x^0 \frac{u(t)dt}{u(b_n + a_n t)}\right), \\ &\leq (1-\epsilon)^2 \exp\left(-\int_x^0 \frac{dt}{1 + \epsilon\, |t|}\right), \\ &< c(1 + \epsilon\, |x|)^{-\epsilon^{-1}}, \end{aligned}$$

where, $c > 0$. □

LEMMA 2.2. *Suppose $F \in \mathcal{D}(\Lambda)$ with auxiliary function u and $\epsilon, c > 0$. There exists a large N such that for $x \in \mathbb{R}$, and $n > N$,*

$$g_n(x) < c(1 + \epsilon\,|x|)^{-\frac{1}{\epsilon}}.$$

Proof. From Theorem B.3, for $\epsilon_1 > 0$ we have $f(a_n x + b_n) \leq (1 + \epsilon_1)\frac{\bar{F}(a_n x + b_n)}{u(a_n x + b_n)}$ ultimately. We write,

$$\begin{aligned} g_n(x) &= n a_n f(a_n x + b_n) F^{n-1}(a_n x + b_n), \\ &< \frac{u(b_n)}{u(a_n x + b_n)} n(1 + \epsilon_1)\bar{F}(a_n x + b_n), \end{aligned} \tag{2.13}$$

Using the resukt of theorem B.5, for $\epsilon_2 > 0$ and $x > 0$ such that $a_n x + b_n \geq N$, $\frac{u(b_n)}{u(a_n x + b_n)} < \frac{1}{1-\epsilon_2}\left[\frac{-\log F(b_n)}{-\log F(a_n x + b_n)}\right]^{\epsilon_2}$. Now we apply

$$\begin{aligned} g_n(x) &< \frac{u(b_n)}{u(a_n x + b_n)}(1 + \epsilon_1) n\bar{F}(a_n x + b_n), \\ &< \frac{1 + \epsilon_1}{1 - \epsilon_2}(n\bar{F}(a_n x + b_n))^{1-\epsilon_2}. \end{aligned}$$

From Lemma 2.1 for $\epsilon_3 > 0$ and $x > 0$ we have $n\bar{F}(a_n x + b_n) < c(1 + \epsilon_3 x)^{-\epsilon_3^{-1}}$, we have

$$g_n(x) < c(1 + \epsilon_3 x)^{-\frac{1-\epsilon_2}{\epsilon_3}} \simeq c(1 + \epsilon_3 x)^{-\frac{1}{\epsilon_3}},$$

where, $1 - \epsilon_2 \simeq 1$.

Similarly, from (2.13), Theorem B.5 and Lemma 2.1 for $x < 0$, we justified that,

$$g_n(x) < c(1 + \epsilon_3\,|x|)^{-\frac{1}{\epsilon_3}}.$$

□

Here, we show that under some conditions, the relative Rényi entropy distance between density function of max domain of attraction of Gumbel under linear normalization with norming constants $a_n > 0$ and $b_n \in \mathbb{R}$ and density function of Gumbel converges to zero.

THEOREM 2.3. *Let $\{X_i, i \geq 1\}$ be iid random variables with df $F \in \mathcal{D}(\Lambda)$ which is absolutely continuous with density function f. If f is nonincreasing function, $h_\beta(F) < \infty$, for $0 < \beta < 1$ and $l(F)$ exists then*

$$\lim_{n\to\infty} D_\beta(G_n \| \Lambda) = 0.$$

Proof. From (1.2) we write,

$$D(G_n\|\Lambda) = \frac{1}{\beta-1}\log\left[I_1(n,v)+I_2(n,v)+I_3(n,v)\right]. \tag{2.14}$$

where, $I_1(n,v)=\int_v^\infty(\lambda(x))^{1-\beta}(g_n(x))^\beta\,dx$, and $I_2(n,v)=\int_{-\infty}^{-v}(\lambda(x))^{1-\beta}(g_n(x))^\beta\,dx$, and $I_3(n,v)=\int_{-v}^{v}(\lambda(x))^{1-\beta}(g_n(x))^\beta\,dx$. It is enough to show that,

$$\lim_{v\to\infty}\lim_{n\to\infty}\left[I_1(n,v)+I_2(n,v)\right]=0.$$

Set, $I_1(n,v)=\int_v^\infty(\lambda(x))^{1-\beta}(g_n(x))^\beta dx$, we have

$$0<I_1(n,v)\le J(n;\beta)\int_v^\infty\left(\frac{f(a_nx+b_n)}{f(b_n)}\right)^\beta(\lambda(x))^{1-\beta}dx.$$

where, $J(n;\beta)=\left(n\bar F(b_n)\frac{f(b_n)a_n}{\bar F(b_n)}\right)^\beta$. From (B. 4) and the facet that $n\bar F(b_n)\to 1$, as $n\to\infty$, we get $J(n;\beta)\to 1$ as $n\to\infty$. From Theorem B.3-(b), f is nonincreasing function and $\frac{f(a_nv+b_n)}{f(b_n)}\le 1$ for $v\ge 1$. We write,

$$0<I_1(n,v)\le J(n;\beta)\left(\frac{f(a_nv+b_n)}{f(b_n)}\right)^\beta\int_v^\infty(\lambda(x))^{1-\beta}dx.$$

Since, $\int_v^\infty(\lambda(x))^{1-\beta}dx<\infty$, for $0<\beta<1$, we get

$$\lim_{v\to\infty}\lim_{n\to\infty}I_1(n,v)=0. \tag{2.15}$$

Now, we choose ξ_n satisfying $-\log F(\xi_n)\simeq n^{-1/2}$. If $t_n=\frac{\xi_n-b_n}{a_n}\to c$, then $n^{1/2}\simeq -n\log F(a_nt_n+b_n)\to e^{-c}$, and this contradicts with $n^{1/2}\to\infty$. Therefore, $t_n=\frac{\xi_n-b_n}{a_n}\to-\infty$, as $\xi_n\to r(F)$ for large n.
We now decompose the integral,

$$I_2(n,v)=\int_{\frac{l(F)-b_n}{a_n}}^{t_n}(\lambda(x))^{1-\beta}(g_n(x))^\beta dx+\int_{t_n}^{-v}(\lambda(x))^{1-\beta}(g_n(x))^\beta dx=I_{21}(n)+I_{22}(n,v).$$

Set

$$\begin{aligned}0<I_{21}(n)&<(nF^{n-1}(\xi_n))^\beta(a_n\Lambda(\xi_n))^{1-\beta}e^{-(1-\beta)l(F)}\int_{\mathbb{R}}(f(s))^\beta dx,\ \text{where } a_nx+b_n=s,\\&\simeq\frac{n^\beta(a_n\Lambda(\xi_n))^{1-\beta}}{\exp(\beta(n-1)n^{-1/2})}e^{-(1-\beta)l(F)}\int_{\mathbb{R}}(f(s))^\beta dx,\end{aligned} \tag{2.16}$$

It is well known that, if $F\in\mathcal{D}(\Lambda)$ with $a_n=u(b_n)$ and denominator is exponential form of n. If $\int_{\mathbb{R}}(f(s))^\beta dx<\infty$ then,

$$\lim_{n\to\infty}I_{21}(n)=0. \tag{2.17}$$

Now, set $I_{22}(n,v) = \int_{t_n}^{-v} (\lambda(x))^{1-\beta}(g_n(x))^{\beta}\,dx$. By Lemma 2.2, we define

$$k(x) = (1+\epsilon\,|x|)^{-\frac{1}{\epsilon}},$$

and using Hölder inequality, for $0<\beta<1$, we get,

$$I_{22}(n,v) < \left(\int_{t_n}^{-v} (\lambda(x))^{p(1-\beta)}dx\right)^{1/p} \left(\int_{t_n}^{-v} (k(x))^{q\beta}\,dx\right)^{1/q}.$$

where, $\frac{1}{p}+\frac{1}{q}=1$. Since, $\left(\int_{\mathbb{R}}(\lambda(x))^{p(1-\beta)}dx\right)^{1/p}\left(\int_{\mathbb{R}}(k(x))^{q\beta}\,dx\right)^{1/q}<\infty$ we have

$$\lim_{v\to\infty}\lim_{n\to\infty} I_{22}(n,v) = 0. \tag{2.18}$$

From, (2.17) and (2.18)

$$\lim_{v\to\infty}\lim_{n\to\infty} I_{2}(n,v) = 0. \tag{2.19}$$

Finally, we consider $I_3(n,v) = \int_{-v}^{v} (\lambda(x))^{1-\beta}(g_n(x))^{\beta}\,dx$. From Lemma 2.2, for large n, we have $g_n(x) < k(x)$, and $\int_{-v}^{v} k(x)dx < \infty$ by using Theorem B.4-iii $\lim_{n\to\infty} g_n(x) = \lambda(x)$, for $x\in[-v,v]$, we get

$$\lim_{v\to\infty}\lim_{n\to\infty}\int_{-v}^{v} (\lambda(x))^{1-\beta}(g_n(x))^{\beta}\,dx = \int_{-\infty}^{\infty}\lambda(x)\,dx = 1. \tag{2.20}$$

From (2.14), (2.15), (2.19) and (2.20),

$$\lim_{n\to\infty} D_\beta(G_n\|\Lambda) = 0.$$

□

REMARK 2.1. Lemma A.2 shows, Rényi entropy does not depend on the location and scale parameters.

Appendix A.

LEMMA A.1. *The Rényi entropy of*

(i) *Fréchet law:*

$$h_\beta(\Phi_\alpha) = -\log\alpha + \frac{\alpha+1}{\alpha}\log\beta - \frac{1}{1-\beta}\left(\log\beta - \log\Gamma\left(\frac{\alpha+1}{\alpha}(\beta-1)+1\right)\right);$$

where, $\frac{1}{\alpha+1}<\beta$.

(ii) *Weibull law:*

$$h_\beta(\Psi_\alpha) = -\log\alpha + \frac{\alpha-1}{\alpha}\log\beta - \frac{1}{1-\beta}\left(\log\beta - \log\Gamma\left(\frac{\alpha-1}{\alpha}(\beta-1)+1\right)\right);$$

where, $\max\left(0, \frac{\beta-1}{\beta}\right) < \alpha$, *for* $\beta > 0$.

(iii) *Gumbel law:*

$$h_\beta(\Lambda) = \frac{1}{1-\beta}\log\frac{\Gamma(\beta)}{\beta^\beta};$$

where, $\beta > 0$.

Proof. (i) The Rényi entropy of Fréchet distribution is

$$h_\beta(\Phi_\alpha) = \frac{1}{1-\beta}\log\int_0^\infty \left(\alpha x^{-\alpha-1}e^{-x^{-\alpha}}\right)^\beta dx.$$

Putting, $\beta x^{-\alpha} = u$, $-\beta\alpha x^{-\alpha-1}dx = du$,

$$\begin{aligned} &h_\beta(\Phi_\alpha)\\ &= \frac{1}{1-\beta}\log\int_0^\infty \alpha^{\beta-1}u^{(\beta-1)(\frac{\alpha+1}{\alpha})}\beta^{-((\beta-1)(\frac{\alpha+1}{\alpha})+1)}e^{-u}du,\\ &= \frac{1}{1-\beta}\left((\beta-1)\log\alpha - \left((\beta-1)\frac{\alpha+1}{\alpha}+1\right)\log\beta + \log\Gamma\left(\frac{\alpha+1}{\alpha}(\beta-1)+1\right)\right), \end{aligned}$$

where, $\frac{\alpha+1}{\alpha}(\beta-1)+1 > 0$, so $\frac{1}{\beta} < \alpha+1$.
(ii) By a similar argument for Rényi entropy of Weibull distribution we have

$$\begin{aligned} &h_\beta(\Psi_\alpha)\\ &= \frac{1}{1-\beta}\log\int_0^\infty \left(\alpha x^{\alpha-1}e^{-x^\alpha}\right)^\beta dx,\\ &= \frac{1}{1-\beta}\left((\beta-1)\log\alpha - \left((\beta-1)\frac{\alpha-1}{\alpha}+1\right)\log\beta + \log\Gamma\left(\frac{\alpha-1}{\alpha}(\beta-1)+1\right)\right), \end{aligned}$$

where, $\frac{\alpha-1}{\alpha}(\beta-1)+1 > 0$. If $\beta > 1$ then $\alpha > \frac{\beta-1}{\beta}$, and for $\beta < 1$, we have $0 < \alpha$, therefore, $\max\left(0, \frac{\beta-1}{\beta}\right) < \alpha$, for all $\beta > 0$.
(iii) The Rényi entropy of Gumbel distribution

$$h_\beta(\Lambda) = \frac{1}{1-\beta}\log\int_{-\infty}^\infty \left(e^{-x}e^{-e^{-x}}\right)^\beta dx,$$

Taking $\frac{u}{\beta} = e^{-x}$ and $\frac{du}{\beta} = -e^{-x}dx$

$$\begin{aligned} h_\beta(\Lambda) &= \frac{1}{1-\beta}\log\int_0^\infty u^{\beta-1}e^{-u}\beta^{-\beta}du, \\ &= \frac{1}{1-\beta}\left(\log\Gamma(\beta) - \beta\log\beta\right). \end{aligned}$$

where, $\beta > 0$. □

Lemma A.2. *If $Y = \frac{X-b}{a}$, for $b \in \mathbb{R}$ and $a > 0$, then Rényi's entropy of Y is given by*

$$h_\beta(F_Y) = -\log a + h_\beta(F_X).$$

Proof. We have $F_Y(y) = \Pr\left(X \le ay+b\right) = F_X\left(ay+b\right)$, and $f_Y(y) = af_X(ay+b)$, so that from (1.1),

$$\begin{aligned} h_\beta(F_Y) &= \frac{1}{1-\beta}\log\int_{-\infty}^{\infty}\left(af_X(ay+b)\right)^\beta dy \;= \frac{1}{1-\beta}\log\int_{-\infty}^{\infty} f_X^\beta(z)a^{\beta-1}dz, \\ &= -\log a + h_\beta(F_X). \end{aligned}$$

□

Appendix B.

Theorem B.1. (Proposition 0.8, Resnick (1987)) *Suppose $U \in RV_\rho$, $\rho \in \mathbb{R}$. Take $\epsilon > 0$. Then there exists t_0 such that for $x \ge 1$ and $t \ge t_0$*

$$(1-\epsilon)x^{\rho-\epsilon} < \frac{U(tx)}{U(t)} < (1+\epsilon)x^{\rho+\epsilon}.$$

Remark B.1. (Remark, Page 19, Resnick (1987)) If $U \in RV_\rho$ then U has representation

$$U(x) = c(x)\exp\left(\int_1^x t^{-1}\rho(t)dt\right), \tag{B. 1}$$

where, $\lim_{x\to\infty} c(x) = c$ and $\lim_{t\to\infty}\rho(t) = \rho$.

Theorem B.2. (Proposition 1.15 and 1.16, Resnick (1987))

(i) *Suppose that distribution function F is absolutely continuous with positive density f in some neighborhood of ∞.*

(*a*) *If for some* $\alpha > 0$

$$\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} = \alpha, \tag{B.2}$$

then $F \in \mathcal{D}(\Phi_\alpha)$.

(*b*) *If* f *is nonincreasing and* $F \in \mathcal{D}(\Phi_\alpha)$ *then* $(B.2)$ *holds.*

(ii) *Suppose* F *has finite right endpoint* $r(F)$ *and is absolutely continuous in a left neighborhood of* $r(F)$ *with positive density* f.

(*a*) *If for some* $\alpha > 0$

$$\lim_{x\to r(F)} \frac{(r(F)-x)f(x)}{\overline{F}(x)} = \alpha \tag{B.3}$$

then $F \in \mathcal{D}(\Psi_\alpha)$.

(*b*) *If* f *is nonincreasing and* $F \in \mathcal{D}(\Psi_\alpha)$ *then* $(B.3)$ *holds.*

THEOREM B.3. (Proposition 1.17, Resnick (1987)) (*a*) *Let* F *be absolutely continuous in a left neighborhood of* $r(F)$ *with density* f. *If*

$$\lim_{x\uparrow r(F)} f(x) \int_x^{r(F)} \overline{F}(t)dt/\overline{F}(x)^2 = 1, \tag{B.4}$$

then $F \in \mathcal{D}(\Lambda)$. *In this case we may take,*

$$u(x) = \int_x^{r(F)} \overline{F}(t)dt/\overline{F}(x),\ b_n = F^{\leftarrow}(1-1/n),\quad a_n = u(b_n).$$

(*b*) *If* f *is nonincreasing and* $F \in \mathcal{D}(\Lambda)$ *then* $(B.4)$ *holds.*

THEOREM B.4. (Theorem 2.5, Resnick (1987)) *Suppose that* F *is absolutely continuous with pdf* f. *If* $F \in \mathcal{D}(G)$ *and*

(i) $G = \Phi_\alpha$, *then* $g_n(x) \to \phi_\alpha(x)$ *locally uniformly on* $(0,\infty)$ *iff* $(B.2)$ *holds;*

(ii) $G = \Psi_\alpha$, *then* $g_n(x) \to \psi_\alpha(x)$ *locally uniformly on* $(-\infty, 0)$ *iff* $(B.3)$ *holds;*

(iii) $G = \Lambda$, *then* $g_n(x) \to \lambda(x)$ *locally uniformly on* $\mathbb{R}$ *iff* $(B.4)$ *holds.*

THEOREM B.5. (Lemma 2, de Haan and Resnick (1982)) *Suppose* $F \in \mathcal{D}(\Lambda)$ *with auxiliary function* u *and* $\epsilon > 0$. *There exists a* t_0 *such that for* $x \geq 0$, $t \geq t_0$

$$(1-\epsilon)\left[\frac{-\log F(t)}{-\log F(t+xu(t))}\right]^{-\epsilon} \leq \frac{u(t+xu(t))}{u(t)} \leq (1+\epsilon)\left[\frac{-\log F(t)}{-\log F(t+xu(t))}\right]^{\epsilon},$$

and for $x < 0$, $t + xu(t) \geq t_0$

$$(1-\epsilon)\left[\frac{-\log F(t+xu(t))}{-\log F(t)}\right]^{-\epsilon} \leq \frac{u(t+xu(t))}{u(t)} \leq (1+\epsilon)\left[\frac{-\log F(t+xu(t))}{-\log F(t)}\right]^{\epsilon}.$$

Theorem B.6. (Corollary, Balkema and de Haan (1972)) *A distribution function* $F \in \mathcal{D}(\Lambda)$ *if and only if there exist a positive function* c *satisfying* $\lim_{x\to r(F)} c(x) = 1$ *and a positive differentiable function* $u(t)$ *satisfying* $\lim_{x\to r(F)} u'(x) = 0$ *such that*

$$\bar{F}(x) = c(x)\exp\left(-\int_{-\infty}^{x}\frac{dt}{u(t)}\right) \quad \textit{for } x < r(F).$$

Acknowledgement. I thank my professors who provided insight and expertise that greatly assisted the research.

References

[1] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, On the rate of convergence in the entropic central limit theorem. *Probability Theory Related Fields*, **129** (2004), 381–390.

[2] A. R. Barron, Entropy and the Central Limit Theorem. *The Annals of Probability*, **14**(1986), 336–342.

[3] A. A. Balkema, and de L. Haan, On R. von Mises condition for the domain of attraction of $\exp(-e^{-x})$. *The Annals of Mathematical Statistics*, **43**(1972), 1352–1354.

[4] L. D. Brown, A proof of the central limit theorem motivated by the Cramer-Rao inequality. *Statistics and probability: Essays in Honour of C. R. Rao.* (1982), 141–148.

[5] H. Cui, and Y. Ding, The convergence of the Rényi entropy of the normalized sums of IID random variables. *Statistics and Probability Letters*, **80** (2010), 1167–1173.

[6] de L. Haan, and S. I. Resnick, Local limit theorems for sample extremes. *The Annals of Probability*, **10**(1982), 396–413.

[7] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modeling Extremal Events for Insurance and Finance*, 1997, Springer Verlag.

[8] R. A. Fisher, and L. H. C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Procs. Cambridge Philos. SOC.*, **24**, 180–190.

[9] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, 1987, Krieger Pub. Co.

[10] A. L. Gibbs, and F. E. Su, On choosing and bounding probability metrics. *Int. Statist. Rev.*, **70**(2002), 419–435.

[11] G. L. Gilardoni, On Pinsker's and Vajda's type inequalities for Csiszár's f-divergence. *IEEE Trans. Inf. Theory*, **56**(2010), 5377–5386.

[12] O. Johnson, *Information Theory and the Central Limit Theorem*, 2006, Imperical College Press.

[13] O. Johnson, and A. Barron, Fisher information inequalities and the central limit theorem. *Probab. Theory Relat. Fields*, **129** (2004), 391–409.

[14] J. V. Linnik, An information theoretic proof of the central limit theorem with Lindeberg conditions, *Theory Probability Application*, **4** (1959), 288–299.

[15] S. Ravi, and A. Saeb, *On convergence of entropy of distribution functions in the max domain of attraction of max stable laws*, 2013, (arXiv).

[16] S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes*, 1987, Springer Verlag.

[17] A. Renyi, On Measures of Entropy and Information. *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability, University of California Press*, **1** (1961), 547–561.

[18] A. Saeb, *Rates of convergence for Rényi entropy in extreme value theory*, 2014, (arXiv).

[19] R. Shimizu, On Fisher's amount of information for location family. *Statistical distributions in scientific work*, **3** (1975), 305–312.

[20] T. J. Sweeting, On domains of uniform local attraction in extreme value theory. *The Annals of Probability*, **13**(1985), 196–205.

[21] S. Takano, Convergence entropy in the central limit theorem. *Yokohama Mathematical Journal*, **35** (1987).

[22] van T. Erven, and P. Harremoës, Rényi divergence and KullbackLeibler divergence. *IEEE Transactions on Information Theory*, **60**(2014).

Department of Economic Sciences,
Indian Institute of Science Education and Research Bhopal,
Bhopal Bypass Road, 462066, India
E-mail: `ali.saeb@gmail.com`