RENYI'S ENTROPY FOR GENERALISED DISCRETE AND CONTINUOUS PROBABILITY SCHEMES

J. N. KAPUR & P. N. CHHABRA

Indian Institute of Technology, Kanpur

(*Received* 26 *April* 1968)

Renyi's entropy for a generalised discrete probability scheme is extended to a generalised continuous probability scheme and a number of properties of both Renyi's entropy and its extension are considered.

Let Ω be an arbitrary space, B a σ-algebra of the subsets of Ω and p, a probability measure on B. Renyi[1] defined the concept of a generalised random variable X on $\Omega_1 \in B$; if $p(\Omega_1) = 1$, he called X a complete random variable and if $0 < p(\Omega_1) < 1$, he called X an incomplete random variable. In particular, let X_d be a generalised discrete random variable describing the scheme of events $[x_1, \ldots\ldots, x_n]$ with respective probabilities

$$\left.\begin{array}{l} [p(x_1), \ldots\ldots\ldots\ldots, p(x_n)] \text{ subject to} \\ 0 < \sum_{i=1}^{n} p(x_i) \leqslant 1,\ p(x_i) > 0 \text{ for all } i \end{array}\right\} \tag{1}$$

By using the concept of Kolmogorov-Nagumo, generalised mean of the self-informatory variable $\log_2 \dfrac{1}{p(x)}$ w.r.t. the strictly monotonic and continuous parametric function

$$\underset{\alpha}{g}(x) = 2^{(1-\alpha)x} \qquad \alpha < 0,\ \alpha \neq 1,\ \infty \tag{2}$$

In the same paper Renyi developed the concept of entropy of order α, $\alpha > 0$, $\alpha \neq 1, \infty$. Denoting it by $H_\alpha(X_d)_g$, it is given by

$$H_\alpha(X_d)_g = \frac{1}{1-\alpha} \log_2 \frac{\sum_i p^\alpha(x_i)}{\sum_i p(x_i)} \qquad \alpha > 0,\ \alpha \neq 1,\ \infty \tag{3}$$

In the limit, as $\alpha \to 1$, (3) tends to

$$H_1(X_d)_g = \frac{-\sum_i p(x_i) \log_2 p(x_i)}{\sum_i p(x_i)} \tag{4}$$

In particular, when $\sum_i p(x_i) = 1$, (4) gives Shannon's definition.

Suppose X_c is a generalised continuous random variable and has $f(x)$ as its density function, positive and defined over Ω_1, so that

$$0 < \int_{\Omega_1} f(x)\, dx \leqslant 1$$

Following the procedure adopted for getting (3), if we take the Kolmogorov-Nagumo generalised mean of $\log_2 \frac{1}{f(x)}$ w.r.t. $g_\alpha(x)$, we get α-entropy associated with X_c. Denoting it by $H_\alpha(X_c)_g$, we have

$$H_\alpha(X_c)_g = g_\alpha^{-1} \left[\frac{\int_{\Omega_1} f(x)\, 2^{(1-\alpha)\log_2 \frac{1}{f(x)}}}{\int_{\Omega_1} f(x)\, dx} \right]$$

$$= \frac{1}{1-\alpha} \log_2 \frac{\int_{\Omega_1} f^\alpha(x)\, dx}{\int_{\Omega_1} f(x)\, dx} \qquad \alpha > 0, \alpha \neq 1, \infty \qquad (5)$$

In the limit, as $\alpha \to 1$, (5) tends to

$$H_1(X_c)_g = \frac{-\int_{\Omega_1} f(x) \log_2 f(x)\, dx}{\int_{\Omega_1} f(x)\, dx} \qquad (6)$$

In particular, when $\int_{\Omega_1} f(x)\, dx = 1$, (6) gives the corresponding Shannon's definition.

In this paper, some of the properties of (4) and (5) are considered. In all what follows, Ω_1 will be dropped from the integral sign in (5) but its presence will otherwise be understood, the underlying integrals will be assumed to exist and the base of logarithms will be assumed to be e.

PROPERTIES

(*i*) *Limit as* $\alpha \to \infty$

$$H_\alpha (X_d)_g = \frac{1}{1-\alpha} \log \frac{\sum_i p^\alpha (x_i)}{\sum_i p(x_i)}$$

$$= -\log \left[\frac{\sum_i p(x_i)\, p^{\alpha-1}(x_i)}{\sum_i p(x_i)} \right]^{\frac{1}{\alpha-1}}$$

$$= -\log M_{\alpha-1} \tag{7}$$

where

$$M_{\alpha-1} = \left[\frac{\sum_i p(x_i)\, p^{\alpha-1}(x_i)}{\sum_i p(x_i)} \right]^{\frac{1}{\alpha-1}} \tag{8}$$

Let $p(x_j)$ be the largest or one of the largest values of $p(x_j)$. Since all the terms of $M_{\alpha-1}$ are positive, we have

$$\left[\frac{p(x_j)\, p^{\alpha-1}(x_j)}{\sum_i p(x_i)} \right]^{\frac{1}{\alpha-1}} \leqslant M_{\alpha-1} \leqslant \left(p^{\alpha-1}(x_j) \right)^{\frac{1}{\alpha-1}}$$

or

$$\left[\frac{p(x_j)}{\sum_i p(x_i)} \right]^{\frac{1}{\alpha-1}} p(x_j) \leqslant M_{\alpha-1} \leqslant p(x_j) \tag{9}$$

So as $\alpha \to \infty$, from (9) $M_{\alpha-1}$ tends to $p(x_j)$.

Using this in (7) we have

$$\operatorname*{Lt}_{\alpha \to \infty} H_\alpha (X_d)_g = -\log \max_i \, p(x_i)$$

Similarly

$$H_\alpha (X_c)_g = -\log \left[\frac{\int f(x) f^{\alpha-1}(x)\, dx}{\int f(x)\, dx} \right]^{\frac{1}{\alpha-1}}$$

$$= -\log G_{\alpha-1} \tag{10}$$

where

$$G_{\alpha-1} = \left[\frac{\int f(x) f^{\alpha-1}(x)\, dx}{\int f(x)\, dx} \right]^{\frac{1}{\alpha-1}} \tag{11}$$

So if $G_{\alpha-1}$ is finite for all $\alpha > 0$, then using the above argument we shall have

$$\underset{\alpha \to \infty}{Lt} H_\alpha (X_c)_g = -\log \max f(x)$$

(*ii*) *Continuity*

$M_{\alpha-1}$ given by (8) is a power mean of order $\alpha-1$, $\alpha > 0$ and we know[2] that $M_{\alpha-1}$ is a continuous function of α. Hence from (7), $H_\alpha (X_d)_g$ is also a continuous function of α.

From[3] page 13 of Reference 7, continuity of $G_{\alpha-1}$ is given by (11). Hence by (10) $H_\alpha (X_c)_g$ is also a continuous function of α.

(*iii*) *Monotonicity*

1st method: By[2] Th. 16 of Reference 5, $M_{\alpha-1}$ given by (8) is a monotonically increasing function of α. Hence by (7), $H_\alpha (X_c)_g$ is a monotonically decreasing function of α.

From[2] page 144 of Reference 5, $G_{\alpha-1}$ given by (11) is a monotonically increasing function of α provided it remains finite for $\alpha > 0$. Hence by (10), $H_\alpha (X_c)_g$ is a monotonically decreasing function of α.

2nd method :

$$\frac{d}{d\alpha} H_\alpha (X_c)_g = \frac{1}{(1-\alpha)^2} \log \frac{\int f^\alpha (x)\, dx}{\int f(x)\, dx}$$

$$+ \frac{1}{1-\alpha} \frac{\int f^\alpha (x) \log f(x)\, dx}{\int f^\alpha (x)\, dx}$$

$$= \frac{1}{(1-\alpha)^2}\left[\log \frac{\int f^\alpha(x)\, dx}{\int f(x)\, dx} + \int \frac{f^\alpha(x)}{\int f^\alpha(x)\, dx} \log f^{1-\alpha} (x)\, dx \right] \quad (12)$$

But by Jensen's inequality for integrals we have

$$\int \frac{f^\alpha(x)}{\int f^\alpha(x)\, dx} \log f^{1-\alpha} (x)\, dx \leqslant \log \frac{\int f(x)\, dx}{\int f^\alpha (x)\, dx} = -\log \frac{\int f^\alpha(x)\, dx}{\int f(x)\, dx} \quad (13)$$

Using (13), (12) reduces to

$$\frac{d}{d\alpha} H_\alpha (X_c)_g \leqslant 0 \quad (14)$$

Similarly

$$\frac{d}{d\alpha} H_\alpha (X_d)_g \leqslant 0 \quad (15)$$

By (14) and (15) $H_\alpha (X_c)_g$ and $H_\alpha (X_d)_g$ are monotonically decreasing functions of α.

(*iv*) *Maximality*

$H_\alpha (X_d)_g$ is maximum when all the x_i are equally likely. Kapur[4] proved it by the technique of dynamic programming. It can also be proved by the technique of Legrange's multipliers.

$H_\alpha (X_c)_g$ can be infinitely large.

Consider
$$f(x) = \frac{1}{x^{3/2}} \quad x \geqslant 9$$
$$= 0 \quad x < 9$$

This makes $\int_{-\infty}^{\infty} \frac{dx}{x^{3/2}} = \int_9^\infty \frac{dx}{x^{3/2}} = \frac{2}{3}$. So $f(x)$ is a generalised probability density function. For $\alpha = \frac{1}{2}$

$$H_{\frac{1}{2}} (X_c)_g = 2 \log \left(\int_9^\infty \frac{dx}{x^{3/4}} \Big/ \frac{2}{3} \right) = \infty$$

This makes 'to maximise $H_\alpha (X_c)_g$' itself a problem. When X_c is a complete continuous random variable, this problem of maximising $H_\alpha (X_c)$ (α-entropy for complete continuous distributions) has been solved[5] for three sets of constraints

(a) when X_c is bounded by a finite interval,

(b) when X_c assumes only non-negative values and its first moment is a preassigned number a, $(a > 0)$,

(c) when X_c has a specified standard deviation.

The corresponding results for $H_\alpha (X_c)_g$ for these three sets of constraints can be obtained with slight modifications of the procedures[5] followed for $H_\alpha (X_c)$.

(v) *Successive derivatives*

Let
$$\phi_\alpha (x) = \log \frac{\int f^\alpha (x)\, dx}{\int f(x)\, dx}$$

so that
$$H_\alpha (X_c)_g = \frac{1}{1-\alpha} \phi_\alpha (x) \tag{16}$$

$$\phi'_\alpha (x) = \frac{d}{d\alpha} \phi_\alpha (x) = \frac{\int f^\alpha (x) \log f(x)\, dx}{\int f^\alpha (x)\, dx} \tag{17}$$

Letting
$$S_k = \frac{\int f^\alpha (x) \left(\log f(x) \right)^k dx}{\int f^\alpha (x)\, dx} \tag{18}$$

we have
$$\frac{d}{d\alpha} S_k = S_{k+1} - S_k\, S_1 \tag{19}$$

Using (17), (18) and (19) we have

$$\phi'_\alpha (x) = S_1$$

$$\phi''_\alpha (x) = S_2 - S_1^2$$

$$\phi'''_\alpha (x) = S_3 - 3S_1\, S_2 + 2\, S_1^3$$

. . . .

. . . .

. . . .

Using (16)
$$H'_\alpha (X_c)_g = \frac{1}{1-\alpha} \left[H_\alpha (X_c)_g + \phi'_\alpha (x) \right]$$

$$H''_\alpha (X_c)_g = \frac{1}{1-\alpha} \left[2\, H'_\alpha (X_c)_g + \phi''_\alpha\ (x) \right]$$

Suppose that result is true for kth derivative i.e.,

$$H_\alpha^{(k)} = \frac{1}{1-\alpha} \left[k\ H_\alpha^{(k-1)} + \phi_\alpha^{(k)} \right]$$

$$H_\alpha^{(k+1)} = \frac{1}{(1-\alpha)^2} \phi_\alpha^{(k)} + \frac{1}{1-\alpha} \phi_\alpha^{(k+1)}$$

$$+ \frac{1}{1-\alpha} k\ H_\alpha^{(k)} + \frac{1}{(1-\alpha)^2} k\, H_\alpha^{(k-1)}$$

$$= \frac{1}{1-\alpha} \left[(k+1)\, H_\alpha^{(k)} + \phi_\alpha^{(k+1)} \right]$$

So, in general

$$H_\alpha^{(n)} = \frac{1}{1-\alpha} \left[n\, H_\alpha^{(n-1)} + \phi_\alpha^{(n)} \right] \alpha > 0,\ \alpha \neq 1,\ \infty \tag{20}$$

Multiplying (20) by $(1-\alpha)^n$ and integrating the result w.r.t. α from α_1 to α_2 we have

$$H_{\alpha_2}^{(n-1)} (1-\alpha_2)^n = H_{\alpha_1}^{(n-1)} (1-\alpha_1)^n + \int_{\alpha_1}^{\alpha_2} (1-\alpha)^{n-1} \phi_\alpha^{(n)}\, d\alpha \tag{21}$$

Putting $\alpha_1 = 1$ and $\alpha_2 = \alpha$ in (21) we have

$$H_\alpha^{(n-1)} (1-\alpha)^n = \int_1^{\alpha} (1-\alpha)^{n-1} \phi_\alpha^{(n)}\, d\alpha \tag{22}$$

Putting $n = 1$ in (20) we have

$$H'_\alpha = \frac{1}{1-\alpha} \left(\phi'_\alpha + H_\alpha \right)$$

So

$$\underset{\alpha \to 1}{\mathrm{Lt}}\ H'_\alpha = H'_1 = -\frac{1}{2} \phi''_1$$

Suppose

$$H_1^{(k)} = -\frac{1}{k+1} \phi_1^{(k+1)}$$

From (20) for $n = k+1$ we have

$$H_\alpha^{(k+1)} = \frac{1}{1-\alpha} \left[\phi_\alpha^{(k+1)} + (k+1)\, H_\alpha^{(k)} \right]$$

$$\underset{\alpha \to 1}{\mathrm{Lt}}\ H_\alpha^{(k+1)} = -\underset{\alpha \to 1}{\mathrm{Lt}} \left[\phi_\alpha^{(k+2)} + (k+1)\ H_\alpha^{(k+1)} \right]$$

$$H_1^{(k+1)} = -\frac{1}{k+2}\,\phi_1^{(k+2)}$$

So in general

$$H_1^{(n)} = -\frac{1}{n+1}\,\phi_1^{(n+1)} \tag{23}$$

With the help of (17), (18) and (19) successive derivatives of $\phi_\alpha(x)$ can be determined and then by using any out of (20), (21) and (22), subject to whichever is convenient, successive derivatives of $H_\alpha(X_c)_g$ can be determined. (23) gives the expression for this purpose for $\alpha = 1$.

Similar results for $H_\alpha(X_d)_g$ can be derived in the same manner.

(vi) Bounds for Second Derivative

$$H'_\alpha(X_c)_g = \frac{1}{(1-\alpha)^2}\log\frac{\int f^\alpha(x)dx}{\int f(x)dx} + \frac{1}{1-\alpha}\,\frac{\int f^\alpha(x)\log f(x)\,dx}{\int f^\alpha(x)dx}$$

$$(1-\alpha)^2 H'_\alpha(X_c)_g = \log\frac{\int f^\alpha(x)\,dx}{\int f(x)\,dx} + (1-\alpha)\,\frac{\int f^\alpha(x)\log f(x)\,dx}{\int f^\alpha(x)\,dx}$$

$$\frac{d}{d\alpha}\left[(1-\alpha)^2 H'_\alpha(X_c)_g\right] = (1-\alpha)\left[\frac{\int f^\alpha(x)\left(\log f(x)\right)^2 dx}{\int f^\alpha(x)\,dx} - \frac{\left(\int f^\alpha(x)\log f(x)\,dx\right)^2}{\left(\int f^\alpha(x)\,dx\right)^2}\right]$$

$$\left(\int f^\alpha(x)\,dx\right)^2\frac{d}{d\alpha}\left[(1-\alpha)^2 H'_\alpha(X_c)_g\right] = (1-\alpha)\left[\int f^\alpha(x)\left(\log f(x)\right)^2 dx\int f^\alpha(x)\,dx - \left(\int f^\alpha(x)\log f(x)\,dx\right)^2\right] \tag{24}$$

Performing the mentioned operation on the L.H.S. and applying the Schwarz's inequality to the R.H.S. of (24), we have

$$-2H'_\alpha(X_c)_g + (1-\alpha)\,H''_\alpha(X_c)_g \geqslant 0 \tag{25}$$

For $0 < \alpha < 1$, (25) reduces to

$$H''_\alpha(X_c)_g \geqslant 2\,H'_\alpha(X_c)_g/(1-\alpha) \tag{26}$$

and for $\alpha > 1$, (25) reduces to

$$H''_\alpha(X_c)_g \leqslant -2\,H'_\alpha(X_c)_g/(\alpha-1) \tag{27}$$

(26) gives the lower bound for $H''_{\alpha}(X_c)_g$ for $0 < \alpha < 1$ and (27) its upper bound for $\alpha > 1$. These bounds are precise in the sense that they are attained when $f(x)$ is uniformly distributed. As proved under 'Monotonicity' $H'\alpha\ (X_c)_g \leq 0$, so the lower bound given by (26) is negative and upper bound given by (27) is positive.

Following the above procedure, similar results for $H_{\alpha}\ (X_d)_g$ will be obtained.

COMPARISON

$H_{\alpha}\ (X_d)_g$ is finite, positive and invariant under the transformation of coordinate systems, whereas $H_{\alpha}\ (X_c)_g$ does not have any of these three properties.

'$H_{\alpha}\ (X_c)_g$ may be infinitely large' follows from the example considered in 'Maximality'

$H_{\alpha}\ (X_c)_g$ may be negative.

Consider $f(x) = bx^2 \qquad 0 \leqslant x \leqslant a$

$= 0$ elsewhere

The value of b that makes $f\ (x)$ a generalised probability density function is given by

$$\int_0^a f(x)\, dx = \frac{ba^3}{3} \leqslant 1 \qquad \text{i.e.} \quad b \leqslant \frac{3}{a^3} \tag{28}$$

$$H_{\alpha}(X_c)_g = \frac{1}{1-\alpha} \log \frac{\int_0^a b^{\alpha}\, x^{2\alpha}\, dx}{\int_0^a b\, x^2\, dx} \tag{29}$$

For $0 < \alpha < 1$

$$\frac{\int_0^a b^{\alpha} x^{2\alpha}\, dx}{\int_0^a b\, x^2\, dx} \leqslant 1 \quad \text{if} \qquad a \geqslant \left(\frac{3^{\alpha}}{2\alpha+1}\right)^{\frac{1}{\alpha-1}} \tag{30}$$

and for $\alpha > 1,\ \alpha \neq \infty$

$$\frac{\int_0^a b\ x^{2\alpha}\ dx}{\int_0^a b\, x^2\, dx} \geqslant 1 \quad \text{if} \qquad a \leqslant \left(\frac{3_t^{\alpha}}{2\alpha+1}\right)^{\frac{1}{\alpha-1}} \tag{31}$$

From (29) and (30), for $0 < \alpha < 1$ we have

$$H_\alpha (X_c)_g \leqslant 0 \qquad \text{if} \qquad a \geqslant \left(\frac{3^\alpha}{2\alpha + 1} \right)^{\frac{1}{\alpha - 1}}$$

and from (29) and (31), for $\alpha > 1$, $\alpha \neq \infty$ we have

$$H_\alpha (X_c)_g \leqslant 0 \qquad \text{if} \qquad a \leqslant \left(\frac{3^\alpha}{2\alpha + 1} \right)^{\frac{1}{\alpha - 1}}$$

$H_\alpha (X_c)_g$ is not invariant under the transformation of coördinate systems.

Consider the transformation of the generalised random variable X to another generalised random variable Y by a continuous, monotonic and 1–1 transformation $Y = g(X)$. The CDF of Y in terms of CDF $F(x)$ of x is given by

$$\begin{aligned} G(y) &= P\{g(x) \leqslant y\} \leqslant 1 && \text{if} \quad y \geqslant g(+\infty) \\ &= P\{g(x) \leqslant y\} = F[g^{-1}(y)] && \text{if} \quad g(-\infty) < y < g(+\infty) \\ &= 0 && \text{if} \quad y \leqslant g(-\infty) \end{aligned}$$

and it can be easily verified that the density function $\rho(y)$ of y in terms of $f(x)$ of x is given by

$$\begin{aligned} \rho(y) &= 0 && \text{for} \quad y \leqslant g(-\infty) \\ \rho(y) &= \frac{f(x)}{|dy/dx|} && g(-\infty) < y < g(+\infty) \\ \rho(y) &= 0 && y \geqslant g(+\infty) \end{aligned}$$

Hence

$$H_\alpha (Y_c)_g = \frac{1}{1-\alpha} \log \frac{\int f^\alpha(x) \left| \frac{dy}{dx} \right|^{1-\alpha} dx}{\int f(x)\, dx} \quad \alpha > 0, \alpha \neq 1, \infty \tag{32}$$

In the limit, as $\alpha \to 1$, and $\int f(x)\, dx = 1$, (32) tends to $H_1 (Y_c)_g = H_1 (X_c)_g + E\left(\log \left| \frac{dy}{dx} \right| \right)$ as already known.

In particular, if $Y = AX + B$.

$$H_\alpha (Y_c)_g = H_\alpha (X_c)_g + \log |A|$$

CONCAVITY—CONVEXITY

In this Section, only $H_\alpha (X_d)_g$ will be studied. Throughout this discussion it will be assumed that $p(x_i)$ are distinct and different, at least two in number and are arranged in ascending order, i.e. $p(x_i) < p(x_{i+1})$, $i=1, \ldots, n$. We shall ultimately establish the following :

(*i*) In the neighbourhood of infinity $H_\alpha (X_d)_g$ is convex,

(ii) If $\frac{p(x_i)}{\sum_i p(x_i)} \geq \frac{1}{2}$, then $H_\alpha (X_d)_g$ is convex for $\alpha > 1$

(iii) For $n = 2$, $H_\alpha (X_d)_g$ is convex for $\alpha > 0$

Denoting $\log \frac{\sum_i p^\alpha (x_i)}{\sum_i p(x_i)}$ by $f(\alpha)$,

$\frac{p^\alpha (x_i)}{\sum_i p^\alpha (x_i)}$ by $\beta_i(\alpha)$, $\log p(x_i)$ by λ_i, $\frac{\sum_i p^\alpha (x_i) (\log p(x_i))^k}{\sum_i p^\alpha (x_i)}$

or its equivalent $\sum_i \beta_i (\alpha) \lambda_i^k$ by M_k and following the procedure adopted for getting $\phi_\alpha''' (x)$ in 'Successive Derivatives' we shall have

$$f''' (\alpha) = \sum_i \beta_i \lambda_i^3 - 3 \left(\sum_i \beta_i \lambda_i\right) \left(\sum_i \beta_i \lambda_i^2\right) - 2 \left(\sum_i \beta_i \lambda_i\right)^3 \tag{33}$$

and following the procedure adopted for getting (20), (21), (22) and (23) and denoting the second derivative of $H_\alpha (X_d)_g$ w.r.t. α by H''_α we shall have

$$H''_\alpha = \frac{1}{1-\alpha} \left[M_2 - M_1^2 + \frac{2}{1-\alpha} M_1 + \frac{2}{(1-\alpha)^2} f(\alpha) \right] \tag{34}$$

$$(1-\alpha_2)^3 H''_{\alpha_2} = (1-\alpha_1)^3 H''_{\alpha_1} + \int_{\alpha_1}^{\alpha_2} (1-\alpha)^2 f''' (\alpha)\, d\alpha \tag{35}$$

$$(1-\alpha)^3 H''_\alpha = \int_1^\alpha (1-\alpha)^2 f''' (\alpha)\, d\alpha \tag{36}$$

$$H''_1 = -\frac{1}{3} f''' (1) \tag{37}$$

A comparison of the expressions for H''_α and $f'''(\alpha)$ shows that the expression for H''_α is quite involved and it is much easier to study the behaviour of $f'''(\alpha)$. Since we have relations (35) to (37) connecting H''_α and $f'''(\alpha)$, so we shall study $f'''(\alpha)$ and draw corresponding conclusions for H''_α from these relations.

We observe the following :

(i) λ_i are fixed real numbers with $\lambda_i < \lambda_{i+1}$, for by hypothesis $p(x_i) < p(x_{i+1})$.

(ii) $\beta_i = \frac{p^\alpha (x_i)}{\sum_i p^\alpha (x_i)} > 0$, $\sum_i \beta_i = 1$ and β_i are continuous functions of α for $\alpha > 0$.

(iii) $\beta_1 = \dfrac{p^{\alpha}(x_1)}{\sum_i p^{\alpha}(x_i)}$ is a strictly decreasing function of α , $\alpha > 0$

with $\beta_1 (0) = \dfrac{1}{n}$ and $\underset{\alpha \to \infty}{Lt} \beta_1 (\alpha) = 0$

for

$$\underset{\alpha \to \infty}{Lt} \beta_1 = \underset{\alpha \to \infty}{Lt} \frac{p^{\alpha}(x_1)}{p^{\alpha}(x_1) \left\{ 1 + \dfrac{p^{\alpha}(x_2)}{p^{\alpha}(x_1)} + \ldots\ldots + \dfrac{p^{\alpha}(x_n)}{p^{\alpha}(x_1)} \right\}} = 0$$

(iv) $\beta_n = \dfrac{p^{\alpha}(x_n)}{\sum_i p^{\alpha}(x_i)}$ is a strictly increasing function of α, $\alpha > 0$,

with $\beta_n (0) = \dfrac{1}{n}$ and $\underset{\alpha \to \infty}{Lt} \beta_n (\alpha) = 1$

for

$$\underset{\alpha \to \infty}{Lt} \beta_n = \underset{\alpha \to \infty}{Lt} \frac{p^{\alpha}(x_n)}{p^{\alpha}(x_n) \left\{ \dfrac{p^{\alpha}(x_1)}{p^{\alpha}(x_n)} + \ldots\ldots + \dfrac{p^{\alpha}(x_{n-1})}{p^{\alpha}(x_n)} + 1 \right\}} = 1$$

Lemma 1

The function $f'''(\alpha)$ has the property that if each λ_i is replaced by $\lambda_i + c$ (c independent of i) then $f'''(\alpha)$ remains invariant.

Proof

$$M_k(\alpha,c) = \sum_{i=1}^{n} \beta_i (\lambda_i + c)^k$$

$$\frac{\partial M_k(\alpha,c)}{\partial c} = \sum_i \beta_i \, k (\lambda_i + c)^{k-1} = k \, M_{k-1}(\alpha,c)$$

$$\frac{\partial M_1(\alpha,c)}{\partial c} = \frac{\partial}{\partial c} \sum_i \beta_i (\lambda_i + c) = \sum_i \beta_i = 1$$

So

$$\frac{\partial}{\partial c} f'''(\alpha,c) = 3 M_2(\alpha,c) - 3 M_2(\alpha,c) - 6 M^2_1(\alpha,c) + 6 M^2_1(\alpha,c) = 0$$

whence the result follows.

Theorem 1

If $\beta_n(\alpha_1) = \dfrac{1}{2}$, then $f'''(\alpha) < 0$ for $\alpha > \alpha_1$

Proof

For $n = 2$

Since by the above Lemma $f'''(\alpha)$ remains invariant, so replacing λ by $\lambda_i - \lambda_1$ in it we shall have

$$f'''(\alpha) = \beta_2 (1 - \beta_2)(1 - 2\beta_2)(\lambda_2 - \lambda_1)^3 \qquad (38)$$

The expression $\beta_2 (1 - \beta_2) (\lambda_2 - \lambda_1)^3$ is positive and for $\alpha > \alpha_1$, $\beta_2 > \frac{1}{2}$, whence the result follows.

For $n > 2$

Assume the theorem is valid for all n satisfying $2 \leqslant n \leqslant k$ and consider $n = k + 1$. Let $d_i = \lambda_i - \lambda_k$. Again by the above Lemma replacing λ_i by $\lambda_i - \lambda_k$ in $f'''(\alpha)$ we have

$$f'''(\alpha) = \sum_{i=1}^{k-1} \beta_i d_i^3 - \beta_{k+1} d^3_{k+1} - 3 \left(\sum_{i=1}^{k-1} \beta_i d_i + \beta_{k+1} d_{k+1} \right)$$

$$\left(\sum_{i=1}^{k-1} \beta_i d_i^2 - \beta_{k+1} d^2_{k+1} \right) + 2 \left(\sum_{i=1}^{k-1} \beta_i d_i + \beta_{k+1} d_{k+1} \right)^3$$

$$= \beta_{k+1} d^3_{k+1} - 3 \beta^2_{k+1} d^3_{k+1} + 2 \beta^3_{k+1} d^3_{k+1}$$

$$+ \sum_{i=1}^{k-1} \beta_i d_i^3 - 3 \left(\sum_{i=1}^{k-1} \beta_i d_i \right) \left(\sum_{i=1}^{k-1} \beta_i d_i^2 \right) + 2 \left(\sum_{i=1}^{k-1} \beta_i d_i \right)^3$$

$$+ 3 \beta_{k+1} d^2_{k+1} \left(2 \beta_{k+1} - 1 \right) \sum_{i=1}^{k-1} \beta_i d_i$$

$$+ 3 \beta_{k+1} d_{k+1} \left[2 \left(\sum_{i=1}^{k-1} \beta_i d_i \right)^2 - \sum_{i=1}^{k-1} \beta_i d_i^2 \right]$$

$$= P_1 + P_2 + P_3 + P_4 \tag{39}$$

where

$$(i)\ P_1 = \beta_{k+1} d^3_{k+1} - 3 \beta^2_{k+1} d^3_{k+1} + 2 \beta^3_{k+1} d^3_{k+1}$$

$$= \beta_{k+1} d^3_{k+1} \left(1 - 3 \beta_{k+1} + 2 \beta^2_{k+1} \right)$$

$$= \beta_{k+1} \left(\lambda_{k+1} - \lambda_k \right)^3 \left(1 - 2 \beta_{k+1} \right) \left(1 - \beta_{k+1} \right)$$

Since $\beta_{k+1} \left(1 - \beta_{k+1} \right) \left(\lambda_{k+1} - \lambda_k \right)^3$ is positive, so for $\beta_{k+1} > \frac{1}{2}$, $P_1 \leqslant 0$,

$$(ii)\ P_2 = \sum_{i=1}^{k-1} \beta_i d_i^3 - 3 \left(\sum_{i=1}^{k-1} \beta_i d_i \right) \left(\sum_{i=1}^{k-1} \beta_i d_i^2 \right) + 2 \left(\sum_{i=1}^{k-1} \beta_i d_i \right)^3$$

Regarding the numbers $d_1, d_2, \ldots\ldots, d_{k-1}, 0$ as λ's and the numbers $\beta_1, \beta_2, \ldots\ldots \beta_{k-1}$, $(\beta_k + \beta_{k+1})$ as weights, P_2 becomes $f'''(\alpha)$ with $n = k$; since $\beta_k + \beta_{k+1} > \frac{1}{2}$ whenever $\beta_{k+1} \not> \frac{1}{2}$, the induction hypothesis implies that $P_2 < 0$.

$$(iii)\ P_3 = 3\ \beta_{k+1}\, d^2_{k+1} \left(2\, \beta_{k+1} - 1 \right) \sum_{i=1}^{k-1} \beta_i\, d_i$$

In $\sum_{i=1}^{k-1} \beta_i\, d_i$, each d_i is negative so the sum $\sum_{i=1}^{k-1} \beta_i\, d_i$ is negative. Hence $P_3 < 0$.

$$(iv)\ P_4 = 3\ \beta_{k+1}\ d_{k+1} \left[2 \left(\sum_{i=1}^{k-1} \beta_i\, d_i \right)^2 - \sum_{i=1}^{k-1} \beta_i\, d_i^2 \right]$$

By Cauchy inequality[2]

$$\left(\sum_{i=1}^{k-1} \beta_i\, d_i \right)^2 \leqslant \left(\sum_{i=1}^{k-1} \beta_i\, d_i^2 \right) \left(\sum_{i=1}^{k-1} \beta_i \right) < \frac{1}{2} \sum_{i=1}^{k-1} \beta_i\, d_i^3$$

$$\left[\text{Since } \beta_k + \beta_{k+1} > \frac{1}{2}\ , \text{ so } \sum_{i=1}^{k-1} \beta_i < \frac{1}{2} \right] \tag{40}$$

Relation (40) and the fact that β_{k+1} and d_{k+1} are positive make P_4 negative.

Using (*i*), (*ii*), (*iii*) and (*iv*), $f'''(\alpha)$ given by (39) is negative. This completes the proof.

Theorem 2

There exists an α_1, $0 < \alpha_1 < \infty$, such that $H''_\alpha > 0$ for $\alpha > \alpha_1$.

Proof

Since $\beta_n(\alpha)$ increases continuously from $\frac{1}{n}$ to 1 so there exists a value p of α for which $\beta_n(p) = \frac{1}{2}$ and hence by Theorem 1, $f'''(\alpha) < 0$ for $\alpha > p$.

(*i*) Case $0 < p < 1$

For $p < \alpha < 1$, putting $\alpha_1 = \alpha$, $\alpha_2 = 1$ in (35) we have

$$(1-\alpha)^3 H''_\alpha = -\int_\alpha^1 (1-\alpha)^2 f'''(\alpha)\, d\alpha \tag{41}$$

Since for $\alpha > p$, $f'''(\alpha) < 0$, so from (41) it follows that for α satisfying $p < \alpha < 1$, $H''_\alpha > 0$.

For $\alpha \geqslant 1 > p$, since $f'''(1) < 0$ for $1 > p$ so for $\alpha = 1$, the result follows from (37) and for $\alpha > 1$, the result follows from (36).

(*ii*) Case $p = 1$

The result follows from (36).

(*iii*) Case $p > 1$

If for some $\beta > p$, $H''_\beta < 0$, then from (35) for all α satisfying $p < \alpha < \beta$, $H''_\alpha < 0$; but this is contrary to the fact that H_α is a monotonically decreasing and bounded function of α. Hence for some $\alpha_1 \geqslant p$, $H''_{\alpha_1} > 0$ and from (35) for $\alpha > \alpha_1$, $H''_{\alpha} > 0$.

This completes the proof.

Corollary 1

$H_\alpha (X_d)_g$ is convex in the neighbourhood of ∞.

It follows immediately from the Theorem.

Corollary 2

$$\text{If } \frac{p(x_n)}{\sum_i p(x_i)} \geqslant \frac{1}{2}, \text{ then } H_\alpha (X_d)_g \text{ is convex for } \alpha > 1.$$

Proof

$$\text{Since } \beta_n(\alpha) = \frac{p^\alpha(x_n)}{\sum_i p^\alpha(x_i)}, \text{ so for } \alpha = 1, \beta_n(1) = \frac{p(x_n)}{\sum_i p(x_i)} \geqslant \frac{1}{2}.$$

Hence by Theorem 1 for $\alpha > 1, f'''(\alpha) < 0$ and the proof of Theorem 2 implies that $H''_\alpha > 0$ for $\alpha > 1$.

Lemma 2

$$\text{If } \beta_n(\alpha) = \frac{1}{2}, \text{ then } f'''(\alpha) = 0 \text{ if } n = 2.$$

Proof

If $n > 2$, we have for $\beta_{k+1} = \frac{1}{2}$ from the proof of Theorem 1, $P_1 = 0, P_2 < 0$, $P_3 = 0$ and $P_4 < 0$. This implies that for $n > 2$, $f'''(\alpha) < 0$.

But from (38) for $n = 2$ and $\beta_2 = \frac{1}{2}$, $f'''(\alpha) = 0$.

Lemma 3

For $n = 2$, $\beta_2(\alpha) = \frac{1}{2}$ if $\alpha = 0$.

Proof

This follows from the definition of $\beta_2(\alpha)$ and its strictly increasing character for $\alpha > 0$.

Theorem 3

If $n = 2$, then $H_\alpha (X_d)_g$ is convex for $\alpha > 0$.

Proof

From Lemmas 2 and 3 and Theorem 1, $f'''(\alpha) < 0$ for $\alpha > 0$. Hence by Theorem 2, the result follows.

REMARKS

1. The developments of Successive Derivatives, Bounds for Second Derivatives and Concavity – Convexity are due to the ideas gathered from references 6 and 7.

2. $H_\alpha (X_d)_g$ given by (3) is also defined for $\alpha < 0$, but Renyi restricted to the range $\alpha > 0$ since in proving uniqueness of $H_\alpha (X_d)_g$ Renyi used its property that it is maximum when all $p(x_i)$ are equal and this property is retained by it only for $\alpha > 0$. For $\alpha < 0$, it is in fact monotonically decreasing and minimum when all the $p(x_i)$ are equal.

Shniad[7] showed that the function

$$\log \left(\sum_i p(x_i)\, a_i^{\alpha}\right)^{1/\alpha} \qquad -\infty < \alpha < \infty, \tag{42}$$

where $\sum_i p(x_i) = 1$ and $a_i > 0$, has two horizontal asymptotes and may have more than two points of inflexions, i.e. it is not convexo-concave for $-\infty < \alpha < \infty$.

Considering $H_\alpha \; (X_d)_g$ for $-\infty < \alpha < \infty$, it can be written as

$$H_\alpha (X_d)_g = -\log \left[\frac{\sum_i p(x_i)\, p^{\alpha-1}(x_i)}{\sum_i p(x_i)}\right]^{\frac{1}{\alpha-1}} \tag{43}$$

From a comparison of (42) and (43) it immediately follows that (43) is minus one times of (42) with the further difference that in (43) $p(x_i)$, which play the role of a_i, are restricted to $0 < p(x_i) \leqslant 1$. Hence we can reasonably expect that $H_\alpha(X_d)_g$ for $\alpha > 0$ may not be concavo-convex, i.e. may have more than two points of inflexion but an example to authenticate this is still to be constructed.

REFERENCES

1. RENYI, A., "On Measures of Entropy and Information", *Proc. 4th Berkeley Symp. Math. Stat. and Prob.* pp. 547-561.

2. HARDY, G. H., LITTLEWOOD, J. E., POLYA, G., "Inequalities" (London, Cambridge University Press), 1951.

3. LITTLEWOOD, J. E., "Lectures on the Theory of Functions" (Oxford University Press, Oxford) 1944.

4. KAPUR, J. N., "Application of Dynamic Programming to Information Theory" *Proc. Ind. Acad. Sciences*, LXVII A, No. 1, pp. 1-11, 1968.

5. CHHABRA, P. N., "Entropies for Two-Dimensional Continuous Memoryless Channels", Indian Institute of Technology, Kanpur, Report.

6. BECKENBACK, E. F., "Convexity Properties of Generalised Mean Value Function", *Ann. Math. Statist* **13** (1942), pp. 88-90.

7. SHNIAD, H., "On the Convexity of Mean Value Functions", *Bull. Amer. Math. Soc.*, **54** (1948), pp. 770-776