1) a) $S = \sum_{i=1}^{n} p_i^2$ Simpson Concentration

Let $S[T] = \sum_{i=1}^{n} p_i^2$ be the Simpson Entropy of text T

" $S[T'] = \sum_{j=1}^{D} p_j^2$ " " " " " " T'

Given: $S[T] = S[T']$

$$= \sum_{j=1}^{D} p_j^2 \; ; \quad p_j = \frac{1}{D} \text{ for } j = 1, 2, 3, \ldots, n'$$

$$= \sum_{j=1}^{D} \left(\frac{1}{D}\right)^2$$

$$S[T] = D\left(\frac{1}{D^2}\right) = D^{-1}$$

$$D^{-1} = \sum_{i=1}^{n} p_i^2 \Rightarrow \boxed{D = \frac{1}{\sum_{i=1} p_i^2}}$$

b) $G \equiv 1 - S = 1 - \sum_{i=1}^{n} p_i^2$ (Gini Index)

$$G[T] = G[T']$$

$$= 1 - \sum_{j=1}^{D} p_j^2$$

$$G[T] = 1 - D^{-1}$$

$$\cancel{1} - D^{-1} = \cancel{1} - \sum_{i=1}^{n} p_i^2$$

$$\Rightarrow \boxed{D = \frac{1}{\sum_{i=1} p_i}}$$

The diversity is the same for $S[T]$ & $G[T]$

c) $H[T] = -\sum_{i=1}^{n} p_i \ln p_i$ Shannon Entropy

$$H[T] = H[T']$$

$$= -\sum_{i=1}^{D} p_j \ln p_j$$

$$= -\sum_{i=1}^{D} \left(\frac{1}{D}\right) \ln\left(\frac{1}{D}\right)$$

$$= -D\left(\frac{1}{D}\right) \ln\left(\frac{1}{D}\right)$$

$$= -\ln\left(\frac{1}{D}\right)$$

$$H[T] = \ln D$$

$$\ln D = -\sum_{i=1}^{n} p_i \ln p_i$$

$$\Rightarrow \boxed{D = e^{-\sum_{i=1}^{n} p_i \ln p_i}}$$

d) $H_q^{(R)} = \frac{1}{q-1}\left[-\ln \sum_{i=1}^{n} p_i^q\right]$ where $q \neq 1$ Rényi Entropy

$$H_q^{(R)}(T] = H_q^{(R)}[T']$$

$$= \frac{1}{q-1}\left[-\ln \sum_{j=1}^{D} p_j^q\right]$$

$$= \frac{1}{1-q} \ln\left[D\left(\frac{1}{D}\right)^q\right]$$

$$= \frac{1}{1-q} \ln\left[D^{1-q}\right]$$

$$H_q^{(R)}[T] = \ln D$$

$$\ln D = \frac{1}{1-q} \ln \sum_{i=1}^{n} p_i^q$$

$$\Rightarrow \boxed{D = e^{\frac{1}{1-q} \ln \sum_{i=1}^{n} p_i^q}} \quad \text{OR} \quad \boxed{D = \left[\sum_{i=1}^{n} p_i^q \right]^{\frac{1}{1-q}}}$$

e) $H_q^{(Ts)} = \frac{1}{q-1} \left[1 - \sum_{i=1}^{n} p_i^q \right]$ where $q \neq 1$ Tsallis Entropy

$$H_q^{(Ts)}[T] = H_q^{(Ts)}[T']$$

$$= \frac{1}{q-1} + \frac{1}{1-q} \sum_{j=1}^{D} p_j^q$$

$$H_q^{(Ts)}[T] = \frac{1}{q-1} + \frac{1}{1-q} D^{1-q}$$

$$\Rightarrow (q-1) H_q^{(Ts)}[T] = 1 - D^{1-q}$$

$$(\cancel{q-1}) \left[\frac{1}{\cancel{q-1}} \left(1 - \sum_{i=1}^{n} p_i^q \right) \right] = 1 - D^{1-q}$$

$$\cancel{1} - \sum_{i=1}^{n} p_i^q = \cancel{1} - D^{1-q}$$

$$\Rightarrow D^{1-q} = \sum_{i=1}^{n} p_i^q$$

$$\boxed{D = \left[\sum_{i=1}^{n} p_i^q \right]^{\frac{1}{1-q}}} \quad \text{OR} \quad \boxed{D = e^{\frac{1}{1-q} \ln \left[\sum_{i=1}^{n} p_i^q \right]}}$$

Rényi & Tsallis diversities are equal!

f) Let $D \equiv D^{(R)} = D^{(T_s)}$

$$\lim_{q\to 1} D = \lim_{q\to 1} e^{\frac{1}{1-q}\ln\left[\sum_{i=1}^{n} p_i^q\right]} = e^{\frac{1}{0}\ln[1]} = e^{\frac{0}{0}} \text{ (indeterminate)}$$

Apply L'Hopitale to the exponent:

$$\lim_{q\to 1} \frac{\ln\left[\sum_{i=1}^{n} p_i^q\right]}{1-q} \overset{L}{=} \lim_{q\to 1} \frac{\frac{d}{dq}\ln\left[\sum_{i=1}^{n} p_i^q\right]}{-1}$$

$$= -\lim_{q\to 1}\left[\ln\left(\frac{\frac{d}{dq}\sum_{i=1}^{n} p_i^q}{\sum_{i=1}^{n} p_i^q}\right)\right]$$

$$= -\lim_{q\to 1}\left[\ln\left(\frac{\sum_{i=1}^{n} q\, p_i^{(q-1)}}{\sum_{i=1}^{n} p_i^q}\right)\right]$$

$$= -\lim\left[\ln q + \ln\left(\frac{\sum_{i=1}^{n} p_i^{(q-1)}}{\sum_{i=1}^{n} p_i^q}\right)\right]$$

$$= -\lim\left(\ln\left(\sum_{i=1}^{n} p_i^{(q-1)}\right) - \ln\left(\sum_{i=1}^{n} p_i^q\right)\right)$$

$$= -\lim_{q\to 1}\left(\ln\left(\sum_{i=1}^{n} p_i^{(q-1)}\right)\right)$$

$$= -\ln \sum_{i=1}^{n} (1)$$

$$= -\ln n \Rightarrow \lim_{q\to 1} D = e^{-\ln n} = \frac{1}{n}$$

f) $D = \left[\sum_{i=1}^{n} p_i^q \right]^{\frac{1}{1-q}}$

$$\Rightarrow \ln D = \frac{1}{1-q} \ln\left[\sum_{i=1}^{n} p_i^q \right]$$

$$\Rightarrow \lim_{q \to 1} \ln D = \lim_{q \to 1} \frac{\ln\left[\sum_{i=1}^{n} p_i^q \right]}{1-q}$$

$$\overset{L}{=} -\lim_{q \to 1} \frac{d}{dq}\left[\ln\left(\sum_{i=1}^{n} p_i^q \right) \right]$$

$$= -\lim_{q \to 1} \ln\left[\frac{\frac{d}{dq} \sum_{i=1}^{n} p_i^q}{\sum_{i=1}^{n} p_i^q} \right] \longrightarrow \text{Recall } \frac{d}{dx}\left[a^x \right] = a^x \ln a, \text{ where } a > 0, \ a \neq 1$$

$$= -\lim_{q \to 1} \ln\left[\frac{\sum_{i=1}^{n} p_i^q \ln p_i}{\underbrace{\sum_{i=1}^{n} p_i^q}_{\to 1 \text{ (normalization)}}} \right]$$

$$\ln D = -\ln \sum_{i=1}^{n} p_i \ln p_i$$

$$\boxed{\therefore D = e^{-\ln \sum_{i=1}^{n} p_i \ln p_i}}$$

In the limit $q \to 1$, the diversity of the Rényi & Tsallis entropies reduce to the diversity of the Shannon Entropy ■

2) $\Psi(p_1, p_2, \ldots, p_n) = F(p_1, p_2, \ldots, p_n) + \lambda G(p_1, p_2, \ldots, p_n)$

$$F(p_1, p_2, \ldots, p_n) = \frac{C}{H} = \frac{\sum_{i=1}^{n} p_i \ln(i+a)}{-g \sum_{i=1}^{n} p_i \ln p_i}$$ "Cost over information function"

$$G(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i - 1 = 0$$

$$p_j = e^{-\left(\frac{1+\lambda H^2}{gC}\right)} (j+a)^{-\frac{H}{gC}}$$ $\rightarrow$ Should get this

Constraint Eq: $\sum_{j=1}^{n} p_j = 1$ (Normalization)

$$p_j = (j+a)^{-\alpha}$$ $\rightarrow$ Want to show, where $\alpha = \frac{H}{gC}$

Want to minimize Ψ:

$$\frac{d\Psi}{dp_i} = 0$$

$$\frac{dF}{di} = \frac{d}{di}\left[\left[\sum_{i=1}^{n} p_i \ln(i+a) \right] \left[-g \sum_{i=1}^{n} p_i \ln p_i \right]^{-1} \right]$$

$$= \sum_{i=1}^{n}\left(\frac{dp_i}{di}\ln(i+a) + p_i\left(\frac{1}{i+a}\right)\right)\left(-g\sum_{i=1}^{n} p_i \ln p_i\right)^{-1} \qquad \frac{dp_i}{di} \equiv p_i'$$

$$+ \left(\sum_{i=1}^{n} p_i \ln(i+a)\right)(-1)\left(-g\sum_{i=1}^{n} p_i \ln p_i\right)^{-2}\left(-g\sum_{i=1}^{n}\left(\frac{dp_i}{di}\ln p_i + p_i\left(\frac{\frac{dp_i}{di}}{p_i}\right)\right)\right)$$

$$\frac{dF}{di} = \frac{-1}{g}\left[\frac{\sum_i\left(p_i'\ln(i+a) + \frac{p_i}{i+a}\right)}{\sum_i p_i \ln p_i}\right] + \frac{-1}{g}\left[\frac{\left(\sum_i p_i' \ln p_i + p_i'\right)\left(\sum_i p_i \ln(i+a)\right)}{\left(\sum_i p_i \ln p_i\right)^2}\right]$$

$$\frac{dG}{di} = \frac{d}{di}\left[\sum_{i=1}^{n} p_i - 1\right] = 0$$

$$\frac{dG}{di} = \underbrace{\sum_i p_i' = 0}$$

↳ Substitute this into $\frac{dF}{di} + \lambda\frac{dG}{di} = \frac{d\varphi}{di} = 0$

$$\frac{d\varphi}{di} = -\frac{1}{g}\left[\frac{\sum p_i'\ln(i+a) + \sum\frac{p_i}{i+a}}{\sum p_i \ln p_i} + \frac{\left(\sum p_i'\ln p_i + \overset{=0}{\cancel{\sum_i p_i'}}\right)\left(\sum p_i \ln(i+a)\right)}{\left(\sum p_i \ln p_i\right)^2}\right] = 0$$

$$= -\frac{1}{g}\left[\frac{\sum p_i'\ln(i+a) + \sum\frac{p_i}{i+a}}{\sum p_i \ln p_i} + \frac{\left(\sum p_i'\ln p_i\right)\overbrace{\left(\sum p_i \ln(i+a)\right)}^{C}}{\left(\sum p_i \ln p_i\right)^2}\right] = 0$$

Recall: $H = -\sum_{i=1}^{n} p_i \log_2 p_i = -g\sum_{i=1}^{n} p_i \ln p_i$ (Shannon Entropy)

$C = \sum_{i=1}^{n} p_i \ln(i+a)$ (cost function)

$$\frac{d\varphi}{di} = \frac{\sum\left(p_i'\ln(i+a) + \frac{p_i}{i+a}\right)}{H} + \frac{\lambda C \sum p_i' \ln p_i}{-\frac{H^2}{g}} = 0$$

$$2)\ \frac{d\Psi}{dp_i} = \frac{dF}{dp_i} + \lambda \frac{dG}{dp_i} = 0$$

<u>p_i - equation</u>

$$= \frac{d}{dp_i}\left[\frac{\sum_{i=1}^{n} p_i \ln(i+a)}{-g \sum_{i=1}^{n} p_i \ln p_i} \right] + \lambda \frac{d}{dp_i}\left[-1 + \sum_{i=1}^{n} p_i \right] \quad ; \sum_{i=1}^{n} \equiv \sum_i$$

$$= \qquad '' \qquad + \lambda \sum \frac{dp_i}{dp_i}$$

$$= \qquad '' \qquad + \lambda \sum_i$$

$$= \qquad '' \qquad + \ ''$$

$$= -\frac{1}{g} \frac{d}{dp_i}\left[\left(\sum_i p_i \ln(i+a)\right)\left(\sum_i p_i \ln p_i\right)^{-1} \right] + \ ''$$

$$= -\frac{1}{g}\left[\left(\frac{d}{dp_i}\left(\sum_i p_i \ln(i+a)\right)\right)\left(\sum_i p_i \ln p_i\right)^{-1} + \left(\sum_i p_i \ln(i+a)\right)\frac{d}{dp_i}\left(\sum_i p_i \ln p_i\right)^{-1} \right] + ''$$

$$= -\frac{1}{g}\left[\left(\sum_i \ln(i+a)\right)\left(\sum_i p_i \ln p_i\right)^{-1} + \left(\sum_i p_i \ln(i+a)\right)(-1)\left(\sum_i p_i \ln p_i\right)^{-2}\left(\sum_i (\ln p_i + 1)\right) \right] + ''$$

$$= -\frac{1}{g}\left[\frac{\sum_i \ln(i+a)}{\sum_i p_i \ln p_i} - \frac{\left(\sum_i p_i \ln(i+a)\right)\left(\sum_i (\ln p_i + 1)\right)}{\left(\sum_i p_i \ln p_i\right)^2} \right] + \lambda \sum_i = 0$$

Recall: $H = -g \sum_i p_i \ln p_i$, $C = \sum_i p_i \ln(i+a)$

$H = -g \sum_i p_i \ln p_i$

$$\frac{d\Psi}{dp_i} = 0 = \frac{\sum_i \ln(i+a)}{H} - \frac{C \sum_i (\ln p_i + 1)}{H \left(\sum_i p_i \ln p_i\right)} + \lambda \sum_i \quad ; \Rightarrow \sum_i p_i \ln p_i = -\frac{H}{g}$$

$$= \frac{\sum_i \ln(i+a)}{H} + \frac{gC \sum (\ln p_i + 1)}{H^2} + \lambda \sum_i$$

$$= \sum_i \left[\frac{1}{H} \ln(i+a) + \frac{gC}{H^2} \ln p_i + \frac{gC}{H^2} + \lambda \right] = 0$$

$$\Rightarrow \frac{1}{H}\ln(i+a) + \frac{gc}{H^2}\ln p_i + \frac{gc}{H^2} + \lambda = 0$$

$$\Rightarrow \frac{gc}{H^2}\ln p_i = -\frac{1}{H}\ln(i+a) - \frac{gc}{H^2} - \lambda$$

$$\Rightarrow \ln p_i = -\frac{H}{gc}\ln(i+a) - 1 - \frac{\lambda H^2}{gc}$$

$$\ln p_i = \ln(i+a)^{-H/gc} - 1 - \frac{\lambda H^2}{gc}$$

$$\Rightarrow p_i = e^{\left[\ln(i+a)^{-H/gc} - 1 - \frac{\lambda H^2}{gc}\right]}$$

$$= e^{\ln(i+a)^{-H/gc}} e^{-1-\frac{\lambda H^2}{gc}}$$

$$\boxed{\therefore p_i = e^{-1-\frac{\lambda H^2}{gc}}(i+a)^{-H/gc}}$$

Let $\alpha \equiv \frac{H}{gc}$

$$\Rightarrow \ln p_i = \left(-1 - \frac{\lambda H^2}{gc}\right) + \ln(i+a)^{-H/gc}$$

$$\ln p_i = \left(-1 - \frac{\lambda H^2}{gc}\right) - \frac{H}{gc}\ln(i+a)$$

Substitute $\ln p_i$ (but no p_i) into H:

$$H = -g \sum p_i\left[-1 - \frac{\lambda H^2}{gc} - \frac{H}{gc}\ln(i+a)\right]$$

$$= g\left[\sum_i p_i + \frac{\lambda H^2}{gc}\sum_i p_i + \frac{H}{gc}\underbrace{\sum p_i \ln(i+a)}_{c}\right]$$

$$= g\left[1 + \frac{\lambda H^2}{gc} + \frac{H}{g}\right]$$

$$\cancel{H} = g + \frac{\lambda H^2}{c} + \cancel{H}$$

$$\Rightarrow \frac{\lambda H^2}{c} = -g$$

$$\Rightarrow \lambda = \frac{-gC}{H^2}$$

Substitute λ into p_i:

$$p_i = e^{-1-\frac{\lambda H^2}{gC}} (i+a)^{-\frac{H}{gC}}$$

$$= e^{-1-\left(\frac{-gC}{H^2}\right)\left(\frac{H^2}{gC}\right)} (i+a)^{\frac{-H}{gC}}$$

$$= e^{-1+1} (i+a)^{\frac{-H}{gC}}$$

$$= (1)(i+a)^{\frac{-H}{gC}}$$

$$\boxed{\therefore\ p_i = (i+a)^{\frac{-H}{gC}}}$$

↳ Inverse Power Law w/ Scaling Exponent $\boxed{\alpha = \frac{H}{gC}}$

3) $p_j = (j+a)^{-\alpha}$ where $j = 1, 2, 3, \dots, n$

a)

Want: Computationally estimate α for $n \to \infty$ & $a = 1$.

Recall: $\sum_{j=1}^{n} p_j = \sum_{j=1}^{n} (j+a)^{-\alpha} = 1$ (Normalization)

Algorithm

$\alpha < 1$

Idea: 1. Initialize random value of α. Set tolerance (use $tol = 10^{-2}$)

2. Calculate $error = \left| \left(\sum_{j=1}^{n} (j+a)^{-\alpha} \right) - 1 \right|$

3. While error > tol:

- If $\left(\sum_{j=1}^{n} (j+a)^{-\alpha} \right) < 1$, decrease α by 10%
- Else, increase α by 10%

$\frac{1}{1^\alpha} + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \dots =$

- Recalculate the error

b) Want: a in terms of n that yields $\alpha = 1$

b) $P_j = (j+a)^{-\alpha}$

$$\Rightarrow \sum_{j=1}^{n} P_j = \sum_{j=1}^{n} (j+a)^{-\alpha} = 1 \qquad j \rightarrow x$$

$$\Rightarrow \int_1^n (x+a)^{-\alpha} dx = 1 \qquad \alpha = 1$$

$$\int_1^n (x+a)^{-1} dx = 1$$

$$\int_1^n \frac{1}{x+a} dx = 1$$

$u = x+a$ $\qquad u_{low} = u\big|_{x=1} = 1+a$

$\frac{du}{dx} = 1 \Rightarrow dx = du \qquad u_{high} = u\big|_{x=n} = n+a$

$$\int_1^n \frac{1}{x+a} dx = \int_{1+a}^{n+a} \frac{1}{u} du$$

$$= \ln u \Big|_{u=1+a}^{u=n+a}$$

$$= \ln(n+a) - \ln(1+a) = 1$$

$$\ln\left(\frac{n+a}{1+a}\right) = 1$$

$$\Rightarrow \frac{n+a}{1+a} = e$$

$$\Rightarrow n+a = e+ea$$

$$\Rightarrow a - ea = e - n$$

$$a(1-e) = e-n$$

$$\Rightarrow \boxed{a = \frac{e-n}{1-e}}$$

As $n \to \infty$, a goes to infinity also. In turn, the probability of each event ($p_j = (j+a)^{-\alpha}$) will tend to zero. This means that each event is very unlikely to happen as the number of possible events increases.