信息熵相关

Rényi 公设下一般离散信息熵之推导

记号

记$\mathcal{P}$为可测空间$(\Omega, \mathscr{F})$上的测度(或为有符号测度、复测度等), 故$\mathcal{P} \in \prod_{n\geq 1} \mathbb{F}^{(n)}$可视作所有至多可数的有序数组, 即

$$\begin{aligned} \mathcal{P} &= (p_1, p_2, \cdots, p_n, \cdots) && \text{数组无限}, \\ \mathcal{P} &= (p_1, \cdots, p_n, 0, 0 \cdots) && \text{数组有限}. \end{aligned}$$

定义$\cup : (\mathcal{P}_1, \cdots, \mathcal{P}_n, \cdots) \mapsto (\mathcal{P}_1, \cdots, \mathcal{P}_n, \cdots)$为数组之保序无交并.

默认$\mathbb{F}$为某一完备域(如$\mathbb{R}$, $\mathbb{C}$等).

Rényi 公设

1. 实值性：$H(\mathcal{P})$ 为非常值的实函数.
2. 对称性：$H(\mathcal{P})$ 关于$\mathcal{P}$中索引p_i对称.
3. 连续性：$H(\mathcal{P})$ 关于$\mathcal{P}$中索引连续(强调：不是关于$\mathcal{P}$连续).

> *$H \notin C(\{\mathcal{P}\})$. 例：不妨设$\mathcal{P}$是至多可列的, 定义度量*
>
> $$d(\mathcal{P}_1, \mathcal{P}_2) = \sum_{i=1}^{\infty} \frac{|p_i^{(1)} - p_i^{(2)}|}{2^i}$$
>
> *即有反例.*

4. 归一性：$H((1/2)) = 1$.
5. 可加性：$H(\mathcal{P}_1 * \mathcal{P}_2) = H(\mathcal{P}_1) + H(\mathcal{P}_2)$.

> *由于对称性在公设内提及, 故需考虑数组之顺序, 不妨记$\mathcal{P}_1 * \mathcal{P}_2$为矩阵$(p_i^{(1)} p_j^{(2)})$所对应之数组. 若不考虑序关系, 则*
>
> $$\mathcal{P}_1 * \mathcal{P}_2 := \cup_{p_i^{(1)} \in \mathcal{P}_1, p_j^{(2)} \in \mathcal{P}_2} \{p_i^{(1)} p_j^{(2)}\}$$

6. 中值性：存在单调连续函数$g:\mathbb{R}\to\mathbb{R}$, 使得

$$H(\mathcal{P}_1\cup\mathcal{P}_2)=g^{-1}\left[\frac{w(\mathcal{P}_1)g(H(\mathcal{P}_1))+w(\mathcal{P}_2)g(H(\mathcal{P}_2))}{w(\mathcal{P}_1\cup\mathcal{P}_2)}\right]$$

其中$w(\mathcal{P})=|\sum_{p_i\in\mathcal{P}}p_i|$为权.

7. 光滑性：$H((q,1-q))$于$q=0$光滑.

信息熵推导

Lemma 1 于公设5下, 对任意$q\neq 0$均有$H((q))=-c_q\log|q|$, 其中常数$c_q\in\mathbb{R}$满足对任意$q_1/q_2\in\mathbb{Q}$均有$c_{q_1}=c_{q_2}$. 其中, H之可表述性取决于选择公理.

证明：置$f(q):=H((q))$, 则对任意$n\in\mathbb{Z}^*$有$f(nq)=nf(q)$, 进而对任意$r\in\mathbb{Q}^*$均有$f(rq)=rf(q)$. 得证.

Lemma 2 加之公设3下, 对任意$q\neq 0$均有$H((q))=c\log|q|$, 其中常数$c\in\mathbb{F}$.

证明：沿用函数f. 在完备域$\mathbb{F}$中选取收敛于$x\in\mathbb{F}^*$的非零有理数列$\{r_i\}$, 即有$f(r_n)\to f(x)$. 得证.

Lemma 3 加之公设1, 即有$H((q))=-c\log|q|$, 其中$c\in\mathbb{R}^*$.

证明：显然.

Lemma 4 加之公设4, 对任意$q\neq 0$均有$H((q))=-\log_2|q|$.

证明：显然.

Lemma 5 加之公设6, 则g为线性函数或指数函数.

证明：由公设6知

$$\begin{aligned}H(\mathcal{P})=&H((p_1)\cup\cdots\cup(p_n)\cup\cdots)\\=&g^{-1}\left[\frac{\sum_i w((p_i))g(H((p_i)))}{w((p_1)\cup\cdots\cup(p_n)\cup\cdots)}\right]\\=&g^{-1}\left[\frac{\sum_i|p_i|g(-\log_2|p_i|)}{\sum_i|p_i|}\right]\end{aligned}$$

置$f:\mathbb{R}_+\to\mathbb{R},t\mapsto g(-\log_2 t)$, 易知$g$单调连续. 由公设5, 考虑$H(\mathcal{P}*(q))$则有

$$f^{-1}\left[\frac{\sum_i|p_i|f(|p_iq|)}{|\sum_i p_i|}\right]=|q|f^{-1}\left[\frac{\sum_i|p_i|f(|p_i|)}{|\sum_i p_i|}\right]$$

置$h_q(t)=f(|q|t)$, 则

$$h_q^{-1}\left[\frac{\sum_i |p_i| h_q(|p_i|)}{|\sum_i p_i|}\right]=f^{-1}\left[\frac{\sum_i |p_i| f(|p_i|)}{|\sum_i p_i|}\right]$$

可见h_q与f等价平均, 下证明f与h_q为线性关系, 即$h_q=a_q f+b_q$.

首先, 对任意$\mathcal{P}$, $h_q=a_q f+b_q$即说明h_q与f等价平均, 下证明h_q一定为$a_q f+b_q$形式.

Lemma 6 (Hardy, Littlewood, Pólya. *Inequalities*. P55-59) 等价平均互为线性关系.

具体言之: 设$\sum c_n=1, c_n>0$, $\phi(x)$为严格单调且连续的函数, 记加权平均
$\mathfrak{R}_\phi(a):=\phi^{-1}[\sum_i c_i\phi(a_i)]$, 则$\mathfrak{R}_\chi(a)\equiv\mathfrak{R}_\psi(a)$的充要条件为存在常数$\alpha$与$\beta$使得$\chi=\alpha\psi+\beta$.

证明：考虑对任意$t\in[H,K]$

$$\begin{aligned}x&:=\psi^{-1}\left[\frac{K-t}{K-H}\psi(H)+\frac{t-H}{K-H}\psi(K)\right]\\&=\chi^{-1}\left[\frac{K-t}{K-H}\chi(H)+\frac{t-H}{K-H}\chi(K)\right]\end{aligned}$$

则当t遍历(H,K)时, x取遍(H,K)中所有值, 因此

$$\begin{aligned}\chi(x)&=\frac{K-t}{K-H}\chi(H)+\frac{t-H}{K-H}\chi(K)\\&=\frac{\psi(K)-\psi(x)}{\psi(K)-\psi(H)}\chi(H)+\frac{\psi(x)-\psi(H)}{\psi(K)-\psi(H)}\chi(K)\\&=\alpha\psi(x)+\beta\end{aligned}$$

必要性得证. 充分性显然, 故引理得证.

由以上引理得函数方程: $f(|q|t)=s(|q|)f(t)+r(|q|)$, 其中$s(|q|)$与$r(|q|)$与$|q|$关联之常数, 且$s(|q|)>0$. 同时即得$f(|q|)=r(|q|)$. 以$y$代$|q|$, 得

$$\begin{aligned}f(ty)&=s(y)f(t)+f(y)\\f(ty)&=s(t)f(y)+f(t)\end{aligned}$$

故$\dfrac{s(t)-1}{f(t)}=\dfrac{s(y)-1}{f(y)}$. 因此存在常数$c$使得$s(t)=1+cf(t)$, 回代得

$$f(ty)=cf(t)f(y)+f(t)+f(y)$$

1. 当$c=0$时, 解即为$f=C\log t$.
2. 当$c\neq 0$时, 原方程化为$[cf(ty)+1]=[cf(t)+1][cf(y)+1]$, 解为$f=\dfrac{x^\alpha-1}{c}$, 这里$\alpha$为任意常数.

故$g(x)=-ax+b$或$a\cdot 2^{(1-\alpha)x}+b$.

Lemma 7 加之剩余公理, 则$g(x)=-ax+b$或$g(x)=a\cdot 2^{(1-2k)x}$, 其中$k\in\mathbb{N}^+$.

由公理6知$H((q,1-q))=H((q)\cup(1-q))$, 故

当$g(x)=-ax+b$时有

$$-aH((q,1-q))+b=a[q\log_2 q+(1-q)\log_2(1-q)]+b$$

即$H((q,1-q))=-q\log q-(1-q)\log(1-q)$, 显然$H\in C^\infty(Q)$. 此时

$$H(\mathcal{P})=-\frac{\sum_i|p_i|\log_2|p_i|}{|\sum_i p_i|}$$

当$g(x)=a\cdot 2^{(1-\alpha)x}+b$有

$$a\cdot 2^{(1-\alpha)H(\mathcal{P})}+b=\frac{a\sum_i|p_i|^\alpha+b\sum_i|p_i|}{|\sum_i p_i|}$$

取$\mathcal{P}=(q,1-q)$, 由于$\dfrac{|q|+|1-q|}{|q+1-q|}$于$q=0$处不可微, 故$b=0$. 因此

$$H((q,1-q))=\frac{\log_2[|q|^\alpha+|1-q|^\alpha]}{1-\alpha}$$

显然$\alpha<0$时则$H((0,1))$无界, 故舍去. 同时公设1要求$\alpha\neq 0$, 故$\alpha>0$. 若α非整数, 考虑$q\in\mathbb{R}_+$则有

$$\frac{\mathrm{d}H((q,1-q))}{\mathrm{d}q}=\frac{\alpha}{(\alpha-1)\ln 2}\cdot\frac{q^{\alpha-1}+(1-q)^{\alpha-1}}{q^\alpha+(1-q)^\alpha}$$

故容易推得

$$\frac{\mathrm{d}^nH((q,1-q))}{\mathrm{d}q^n}=C_n\cdot\frac{p_1^{(n)}(q)+p_2^{(n)}(q)[q^{\alpha-n}+(1-q)^{\alpha-n}]}{p_2^{(n)}(q)}$$

其中$p_k^{(n)}(q)$为由$q^{\alpha-m}+(1-q)^{\alpha-m},\quad\mathbb{Z}\ni m<n$组成的多项式. 因此取$n$为某一大于$\alpha$之常数, 则有$\dfrac{\mathrm{d}^nH((q,1-q))}{\mathrm{d}q^n}=\infty$, 矛盾. 因此$\alpha$为整数.

同时, α需为偶数, 反之$\dfrac{\mathrm{d}^nH((q,1-q))}{\mathrm{d}q^n}$于$q=0^-$及$q=0^+$时不等.

因此

$$H(\mathcal{P}) = H_{2k}(\mathcal{P}) = -\frac{1}{2k-1}\log_2\left[\frac{\sum_i |p_i|^{2k}}{|\sum_i p_i|}\right]$$

其中$k \in \mathbb{N}^+$.

结论

综上所述

$$H(\mathcal{P}) = H_{2k}(\mathcal{P}) = -\frac{1}{2k-1}\log_2\left[\frac{\sum_i |p_i|^{2k}}{|\sum_i p_i|}\right]$$

或

$$H(\mathcal{P}) = -\frac{\sum_i |p_i| \log_2 |p_i|}{|\sum_i p_i|}$$

均为符合Rényi 公设之信息熵.

概率空间中Rényi熵之若干性质

定义

概率空间不及一般可测空间复杂. 由于概率函数之值域仅$[0, 1]$尔, 相应的Rényi公设7(原点光滑性要求)应当删去. $H((q, 1-q))$于$q \in (0, 1)$时之有界性要求$\alpha > 0$. 因此概率空间中的Rényi熵具有如下一般定义

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2\left(\sum_i p_i^\alpha\right) = \frac{\alpha}{1-\alpha}\log_2 \|P\|_\alpha$$

其中定义范数$\|P\|_\alpha := \sqrt[\alpha]{\sum_i p_i^\alpha}$.

几类特殊的Rényi熵

定义$\alpha = 1$

对$\alpha \to 1$可取极限定义

$$\begin{aligned} H_\alpha(X) &= \frac{1}{1-\alpha} \log_2 \left(\sum_i p_i^\alpha \right) \\ &= \frac{1}{1-\alpha} \log_2 \left(1 + \sum_i p_i (p_i^{\alpha-1} - 1) \right) \\ &\to \frac{1}{(1-\alpha)\log 2} \cdot \sum_i p_i (p_i^{\alpha-1} - 1) \\ &= \frac{1}{\log 2} \sum_i \left(p_i \cdot \frac{p_i^{\alpha-1} - 1}{1-\alpha} \right) \\ &\to \frac{1}{\log 2} \sum_{i=1}^n (-p_i \log p_i) \\ &= -\sum_i (p_i \log_2 p_i) \end{aligned}$$

即一般的Shannon熵.

定义$\alpha = 0$

若$(p_1, p_2, \cdots)$有限, 则$H_0(X) := \log_2 |X|$是自然的.

Collision熵($\alpha = 2$之情形)

$H_2(X) = -\log_2 \sum_i p_i^2 = -\log_2 P(X = Y)$. 其中$Y$于$X$独立且分布相同.

Min-熵

自然定义$H_\infty(X) = \lim_{\alpha \to \infty} \dfrac{\alpha}{1-\alpha} \log_2 \|P\|_\infty = -\log_2 \sup p_i$.

H_α随α之单调性

求导得

$$\frac{\partial H_\alpha(X)}{\partial \alpha} = \frac{1}{(1-\alpha)^2} \sum_i [\theta_i \log(\theta_i / p_i)]$$

其中, 定义$\theta_i := \frac{p_i^\alpha}{\sum_j p_j^\alpha} = \frac{\|p_i\|_\alpha^\alpha}{\|X\|_\alpha^\alpha}$. 容易观察$\sum_i \theta_i = 1$, 因此

$$(1-\alpha)^2 \cdot \frac{\partial H_\alpha(X)}{\partial \alpha} = D(\Theta\|X) \geq 0$$

随之导出H_α在$\alpha > 0$时单调递减.

当$\alpha > 1$时, 对任意$\alpha_1 \neq \alpha_2$, 总存在常数c_1, c_2使得

$$c_1 H_{\alpha_1}(X) \leq H_{\alpha_2}(X) \leq c_2 H_{\alpha_1}(X)$$

对任意X成立. 因为$c_3\|\cdot\|_{\alpha_1} \leq \|\cdot\|_{\alpha_2} \leq c_4\|\cdot\|_{\alpha_1}$.

Rényi熵意义下的信息论

若干公式

离散Rényi熵公式

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2\left(\sum_i p_i^\alpha\right) = \frac{\alpha}{1-\alpha}\log_2\|P\|_\alpha$$

连续Rényi熵公式

$$H_\alpha(X) = \frac{1}{1-\alpha}\log\int_X f_X^\alpha(x)\mathrm{d}x = \frac{\alpha}{1-\alpha}\log_2\|f_X\|_\alpha$$

离散相对Rényi熵公式

$$D(P\|Q) = \frac{1}{1-\alpha}\log_2\left(\sum_{i,j} p_{i,j}(p_{i,j}/q_{i,j})^{\alpha-1}\right)$$

连续相对Rényi熵公式

$$D(f\|g) = \frac{1}{1-\alpha}\log_2\int_{X,Y} f(x,y)(f(x,y)/g(x,y))^{\alpha-1}\mathrm{d}x\mathrm{d}y$$

其他

同样可定义互信息I_α, 条件相对熵等概念, 在此从略.

推荐阅读

[1] 量子信息论第13章

[2] 这里的一张表格, 其他无关紧要.