3. Entropy Rate

For random variable X , entropy is $H(X)$, the joint entropy of $X_1, X_2, ..., X_n$ is

$$\begin{aligned} H(X_1, .., X_n) &= \sum_{i=1}^{n} H(X_i|X_{i-1}, .., X_1) \qquad \text{Chain Rule} \\ &\leq \sum_{i=1}^{n} H(X_i) \qquad \text{Since conditioning does not increase entropy} \\ &= nH(X) \qquad \text{if the variables are identically distributed} \end{aligned}$$

If the random variables are also independent, then the joint entropy of n random variables increases with n.

S**tochastic Process**

A **stochastic process** $\{X_i\}$ is an indexed sequence of random variables. They need not be independent nor identically distributed. For each n the distribution on $X^n = (X_1, X_2, \cdots, X_n)$ is characterized by the joint pmf

$$p_{X^n}(x^n) = p(x_1, x_2, .., x_n)$$

Stationary stochastic process:

A stochastic process is stationary if the joint distribution of any subset of the sequence of random variables is invariant with respect to shifts in the time index:

$$p(X_1 = x_1, ..., X_n = x_n) = p(X_{1+l} = x_1, ..., X_{n+l} = x_n) \qquad \forall l, \forall n$$

Theorem: For a stationary process, the limits in $H(X)$ and $H'(X)$ exist and are equal.

For a stationary stochastic process, the following limit always exists

$$H(\mathcal{X}) := \lim_{n\to\infty} \frac{H(X_1, .., X_n)}{n}$$

i.e. the limit of per symbol entropy, and and is equal to

$$H'(\mathcal{X}) := \lim_{n\to\infty} H(X_n|X_{n-1}, ..., X_1)$$

i.e. the limit of the conditional entropy of last random variable given past.

▼ Proof

$$\begin{aligned} H(X_n|X_1, .., X_{n-1}) &\leq H(X_n|X_2, .., X_{n-1}) \\ &= H(X_{n-1}|X_1, .., X_{n-2}) \end{aligned}$$

Therefore, $H(X_n|, X_1, ..., X_{n-1})$ is non-increasing in n, also $H(X) > 0$, then the $H'(\mathcal{X})$ exist.

Applying chain rule of entropy we have:

$$\frac{1}{n}H(X_1, X_2, .., X_n) = \frac{1}{n}\sum_{i=1}^{n} H(X_n|X_1, .., X_{n-1})$$

we already know that

$$H(X_n|X_{n-1}, ..., X_1) \to H'(\mathcal{X})$$

By Cesaro mean we get

$$\frac{1}{n}\sum_{i=1}^{n} H(X_n|X_1, .., X_{n-1}) \to H'(\mathcal{X})$$

Theorem: (AEP for stochastic sources) For a stationary ergodic process

$$-\frac{1}{n}log\ p(X_1, X_2, .., X_n) \to H(\mathcal{X})$$