

Statistical Inference for Entropy

Karina Marks

October 25, 2016

1 Introduction

2 Entropies and Properties

Entropy can be thought of as a representation of the average information content of an observation; sometimes referred to as a measure of unpredictability or disorder.

2.1 Shannon Entropy

The Shannon entropy of a random vector X with density function f is given by;

$$H = -\mathbb{E}\{\log(f(x))\} = - \int_{x:f(x)>0} f(x)\log(f(x))dx \quad (1)$$

2.2 Rényi and Tsallis Entropy

These entropies are for the order $q \neq 1$ and the construction of them relies upon the generalisation of the Shannon entropy 1. For a random vector $X \in \mathbb{R}^d$ with density function f , we define;

Rényi entropy

$$H_q^* = \frac{1}{1-q} \log \left(\int_{\mathbb{R}^d} f^q(x) dx \right) \quad (q \neq 1) \quad (2)$$

Tsallis entropy

$$H_q = \frac{1}{q-1} \left(1 - \int_{\mathbb{R}^d} f^q(x) dx \right) \quad (q \neq 1) \quad (3)$$

When the order of the entropy $q \rightarrow 1$, both the Rényi, (2), and Tsallis, (3), entropies tend to the Shannon entropy, (1).

3 Estimation of Entropy

3.1 Kozachenko-Leonenko Estimator

We now wish to introduce the Kozachenko-Leonenko estimator of the entropy H . Let X_1, X_2, \dots, X_N , $N \geq 1$ be independent and identically distributed random vectors in \mathbb{R}^d , and denote $\|\cdot\|$ the Euclidean norm on \mathbb{R}^d .

- For $i = 1, 2, \dots, N$, let $X_{(1),i}, X_{(2),i}, \dots, X_{(N-1),i}$ denote an order of the X_k for $k = \{1, 2, \dots, N\} \setminus \{i\}$, such that $\|X_{(1),i} - X_i\| \leq \dots \leq \|X_{(N-1),i} - X_i\|$. Let the metric ρ , defined as;

$$\rho_{(k),i} = \|X_{(k),i} - X_i\| \quad (4)$$

denote the k th nearest neighbour of X_i .

- For dimension d , the volume of the unit d -dimensional Euclidean ball is defined as;

$$V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(1 + \frac{d}{2})} \quad (5)$$

- For the k th nearest neighbour, the digamma function is defined as;

$$\Psi(k) = -\gamma + \sum_{j=1}^{k-1} \frac{1}{j} \quad (6)$$

where $\gamma = 0.577216$ is the Euler-Mascheroni constant (where the digamma function is chosen so that $\frac{e^{\Psi(k)}}{k} \rightarrow 1$ as $k \rightarrow \infty$).

Then the Kozachenko-Leonenko estimator for entropy, H , is given by;

$$\hat{H}_{N,k} = \frac{1}{N} \sum_{i=1}^N \log \left[\frac{\rho_{(k),i}^d V_d (N-1)}{e^{\Psi(k)}} \right] \quad (7)$$

This estimator for entropy, when $d \leq 3$, under a wide range of k and some regularity conditions, this estimator satisfies;

$$N\mathbb{E}(\hat{H}_{N,k} - H)^2 \rightarrow 0 \quad (N \rightarrow \infty) \quad (8)$$

so $\hat{H}_{N,k}$ is efficient in the sense that the asymptotic variance is the best attainable; $N^{\frac{1}{2}}(\hat{H}_{N,k} - H) \xrightarrow{d} N(0, \text{Var}[\log(f(x))])$, the normal distribution with 0 mean and variance as shown.

Later, I will further discuss this estimator for the specific dimensions $d = 1$ and $d = 2$; however, it is important to note that for larger dimensions this estimator is not accurate. When $d = 4$, equation (8) no longer holds but the estimator $\hat{H}_{N,k}$, defined by (7), is still root- N consistent, provided k is bounded. Also, when $d \geq 5$ there is a non trivial bias, regardless of the choice of k .

There is a new proposed estimator, formed as a weighted average of $\hat{H}_{N,k}$ for different values of k , explored in ...SOMEONE... . Moreover, this will not be examined here as this paper focuses only on the 1-dimensional samples;

$$\hat{H}_{N,k} = \frac{1}{N} \sum_{i=1}^N \log \left[\frac{\rho_{(k),i} V_1(N-1)}{e^{\Psi(k)}} \right] \quad (9)$$

and the 2-dimensional;

$$\hat{H}_{N,k} = \frac{1}{N} \sum_{i=1}^N \log \left[\frac{\rho_{(k),i}^2 V_2(N-1)}{e^{\Psi(k)}} \right] \quad (10)$$

3.2 Bias of the K-L estimator

$\hat{H}_{N,k}$ is approximately an unbiased estimator for H ; we wish to explore how approximate this is, by considering the bias of the estimator for entropy;

$$Bias(\hat{H}_{N,k}) = \mathbb{E}(\hat{H}_{N,k}) - H = \mathbb{E}(\hat{H}_{N,k} - H) \quad (11)$$

To do this we consider the consistency and asymptotic bias of the estimator $\hat{H}_{N,k}$, ...SOMEONE... has explored this in detail thus the following theorems hold.

4 Monte-Carlo Simulations

In this section I will explore simulations of the bias of estimator (7) in comparison to the size of the sample estimated from, with respect to different values of k ; firstly exploring 1-dimensional distributions and then progressing onto 2-dimensional.

I will begin by exploring entropy of samples from the normal distribution $N(0, \sigma^2)$, where without loss of generality we can use the mean $\mu = 0$ and change the variance σ^2 as needed. The normal distribution has an exact formula to work out the entropy, given the variance σ^2 . Using equation (1) and the density function for the normal distribution $f(x) = \frac{1}{\sqrt{(2\pi)\sigma}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$ for $x \in \mathbb{R}$, given $\mu = 0$, we can write the exact entropy for the normal distribution;

$$H = \log(\sqrt{(2\pi e)\sigma}) \quad (12)$$

The motivation for these simulations is to explore the consistency of this estimator for different values of k ; the relationship between the size of the bias of the estimator $\hat{H}_{N,k}$, $Bias(\hat{H}_{N,k})$, and the sample size, N . We believe this relationship is of the form;

$$|Bias(\hat{H}_{N,k})| = \frac{c}{N^a} \quad (13)$$

for $a, c > 0$. By taking the logarithm of this, we can see that this relationship is in fact linear;

$$\log|Bias(\hat{H}_{N,k})| = \log(c) - a[\log(N)] \quad (14)$$

We will investigate the consistency of this estimator for a sample from the normal distribution, dependent on the value of k .

4.1 1-dimensional Normal Distribution

4.1.1 k=1

4.1.2 k=2