Statistical Inference for Entropy

Karina Marks

October 22, 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.

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

$$H = -\mathbb{E}\{log(f(x))\} \tag{1}$$
$$= -\int_{x:f(x)>0} f(x)log(f(x))dx \tag{2}$$

3 Estimation of Entropy

3.1 Kozachenko-Leonenko Estimator

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

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

  $$\rho_{(k),i} = \|X_{(k),i} - X_i\| \tag{3}$$

  denote the kth nearest neighbour or X_i.

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

  $$V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(1+\frac{d}{2})} \tag{4}$$

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

$$\Psi(k) = -\gamma + \sum_{j=1}^{k-1} \frac{1}{j} \tag{5}$$

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

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

$$\hat{H}_n = \frac{1}{n} \sum_{i=1}^{n} log \left[\frac{\rho_{(k),i}^{d} V_d (n-1)}{e^{\Psi(k)}} \right] \tag{6}$$

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

$$n\mathbb{E}(\hat{H}_n - H)^2 \to 0 \quad (n \to \infty) \tag{7}$$

so $\hat{H}_n$ is efficient in the sense that the asymptotic variance is the best attainable; $n^{\frac{1}{2}}(\hat{H}_n - H) \xrightarrow{d} N(0, Var[log(f(x))])$.

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 7, no longer holds but the estimator $\hat{H}_n$, defined by 6, 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 $\hat{H}_n$ for different values of k, explored in ... . Moreover, as this paper only considers $d = 1, 2$, this will not be examined here.

4 Monte-Carlo Simulations