1 Background

1.1 Entropy

quick intro to entropy

1.2 Applications of Entropy

how we use entropy and why we need to look at the estimation of it

1.3 Estimation of Entropy

There are several estimation methods for the nonparametric estimation of the Shannon entropy of a continuous random sample. The paper *Nonparametric Entropy Estimation: An Overview* (J.Beirlant, E.Dudewicz, L.Gyorfi, E.van der Muelen, 2001) [**?**], gives an overview of the properties of these various methods. I will outline a summary below to the types of estimators, which will lead us to understand why we choose the Kozachenko-Leonenko estimator for entropy.

First, I must set out the types of consistency, so we can see more obviously how it compares to the K-L estimator, for $X_1, ..., X_N$ a i.i.d sample from the distribution $f(X)$, where H_N is the estimator of $H(f)$. Then we have (as $N \to \infty$);

- Weak Consistency

$$H_N \xrightarrow{p} H(f) \tag{1}$$

- Mean Square Consistency

$$\mathbb{E}\{(H_N - H(f))^2\} \to 0 \tag{2}$$

- Strong Consistency

$$\sqrt{N}(\hat{H}_{N,k} - H) \xrightarrow{d} N(0, \sigma^2) \tag{3}$$

$$N\mathbb{E}(\hat{H}_{N,k} - H)^2 \to \sigma^2 \tag{4}$$

This is the type of consistency shown with the K-L estimator in Theorem **??**.

The types of nonparametric estimators can be split into 3 categories; plug-in estimates, estimates based on sample-spacings and estimates based on nearest neighbour distances. The latter is the Kozachenko-Leonenko estimator, which is the main focus of this paper and will be explored in more detail in section TODO.

The plug-in estimates are based upon a consistent density estimate f_N, of density f, which depends on the sample $X_1, ..., X_N$, I will consider two of these;

the most obvious estimator of this type if the integral estimate of entropy. Given by;

$$H_N = -\int_{A_N} f_N(x) log(f_N(x)) dx \qquad (5)$$

where the set A_N excludes the tail values of f_N. When the sample is from a 1-dimensional distribution, Dmitriev and Tarasenko, [?] for $A_N = [-b_N, b_N]$ and f_N the kernel density estimator; proved a strong consistency for this estimator. However, if f_N is not estimated in this form, due to the numeric integration, for dimensions $d \geq 2$, Joe [?] points out that this estimator is not practical and thus proposed the next plug-in estimator for entropy - the resubstitution estimator.

The resubstitution estimate is of the form;

$$H_N = -\frac{1}{N} \sum_{i=1}^{N} log(f_N(X_i)) dx \qquad (6)$$

which was first proposed in 1976, by Ahmad and Lin [?] who showed the mean-square consistency of this estimator, where f_N is a kernel density estimate. Joe [?] then went on to obtain the asymptotic bias and variance, and whilst satisfying certain conditions reduced the mean square error. Moreover, Hall and Morton [?] went on to say that under more restrictive conditions we have strong consistency for 1-dimensional distributions; however, when $d = 2$ the root-n consistent estimator will have significant bias.

There are also two more plug-in estimates discussed in this paper; the splitting data and cross-validation estimates. Where in the first estimator, strong consistency is shown for a general dimension d, under some conditions on f. And in the latter estimator, strong consistency holds for a kernel estimate of f and for other estimates of f under some conditions we have root-n consistency when $1 \leq d \leq 3$.

Thus, so far the estimates for entropy looked at are only consistent whilst under strong conditions on f and f_N and mostly for a 1-dimensional distribution. So it is important to look at the next category of estimates - estimates of entropy based on sample-spacings; namely the m-spacing estimate. Sometimes it in not practical to estimate f_N, so this estimate is found based on spacings between the sample observations.

This estimator is only defined for samples of 1-dimension, where we assume $X_1, ..., X_N$ are an i.i.d sample, and let $X_{N,1} \leq X_{N,2} \leq ... \leq X_{N,N}$ be the corresponding ordered sample, then $X_{N,i+m} - X_{N,m}$ is the m-spacing.

Firstly we look at this estimator of the form, with fixed m;

$$H_{m,N} = \frac{1}{N} \sum_{i=1}^{N-m} log\left(\frac{N}{m}(X_{N,i+m} - X_{n,i})\right) - \Psi(m) + log(m) \qquad (7)$$

where $\Psi(x)$ is the digamma function - more detailed explanation in section ??. For a sample from a uniform distribution this estimator has been shown to be consistent; proved by Tarasenko [?]. Under some conditions on f, on

2

its boundedness, the weak consistency and asymptotic normality was shown by Hall [**?**].

To decrease the asymptotic variance of the estimator, we consider the estimator when $m_N \to \infty$, which is defined slightly differently;

$$H_{m,N} = \frac{1}{N} \sum_{i=1}^{N-m_N} log \left(\frac{N}{m_N} (X_{N,i+m_N} - X_{n,i}) \right) \qquad (8)$$

for this estimator the weak and strong consistencies are proved under the assumption that as $N \to \infty$, $m_N \to \infty$ and $\frac{m_N}{N} \to 0$, for densities with bounded support.

The last category of estimators discussed by Beirlant, Dudewicz, Gyorfi and Muelen are those based on nearest neighbour distances. The main focus of my paper is on the Kozachenko-Leonenko estimator for entropy; which is the estimator covered in this section of their paper. I will not go into detail for this estimator now; however, I will mention that strong consistency holds for dimension $d \leq 3$, but higher dimensions can cause problems. Thus, it is important to note that recently a new estimator has been proposed by Berrerrt, Samworth and Yuan [**?**], formed as a weighted average of k-nearest neighbour estimators for different values of k. This estimator has shown promising results in higher dimensions, where under the same assumptions as for the K-L estimator, the strong consistency condition holds.