Presentation

Slides 1-2

* + Purpose of this paper was to explore the analytical side of the KL estimator for entropy
  + There has already been extensive research into the theoretical side and I wanted to see how it actually works by considering samples from distributions with known entropies and comparing this with the computed estimator
  + Talk about the Overview briefly

Introduction to Entropy Slides 3-4

* Introduction to entropy
  + read off quote - mention its from my diss
  + Entropy is the quantity of surprise you should feel upon reading the result of a measurement - this can be seen as the uncertainty of the result.
* Entropy Types
  + Main type, that when we think of entropy in general - we are referring to the Shannon entropy
  + The Tsallis and Renyi entropies are generalisations of the Shannon entropy for q not equal to 1 - where q is some parameter and in the Shannon case we think of this parameter q as being 1.
  + The formulas presented are for when we take the entropy for a sample of continuous random variables. We can also describe entropy for discrete random variables where the integrals over the continuous random variables x in the d dimensional space of real numbers, becomes a sum over the discrete random variable x in the d dimensional space of integers.
  + My paper focuses on continuous 1-dimensional distributions, and uses the Shannon entropy of these distributions as a reference for the exact entropy of that distribution.

Kozachenko-Leonenko Estimator of Entropy Slides 5-7

* Kozachenko-Leonenko Estimator of Entropy
  + Explain the formula in detail on board
  + Introduce a sample
  + d - dimension of samples Xi
  + N - sample size (continuous samples Xi)
  + talk about rho the nearest neighbour thing
  + the volume of the d-dimensional Euclidean unit ball , so when we have
  + the digamma function , Euler Mascheroni constant
    - this is chosen so that
* Development of the K-L Estimator
  + Write up the formula for k=1, say that
  + k=1 - this is no longer the best way - it’s between the other two
  + k fixed
  + k dependent on N
  + Talk about what Asymptotic Unbias and Strong consistency means: so as
    - Asymptotic Unbias
    - Strong Consistency is Asymptotic normality
* Bias of the estimator
  + draw a graph
  + Say

Focus of this paper Slides 8-9

* Main research questions
  + From the investigations into the literature on this estimator, I found common themes arising in small discrepancies between different accounts - so believed these would be interesting to explore
  + firstly the optimal value of k
  + how does the bias act - write up both equations

Methodology Slides 10-14

* Data
  + explain how we find everything - draw a table on board?
  + Explain the values of k
    - say that k=1 didn’t work for the two distributions and why!
* Exact entropies
  + how I found them - the integral
  + why I chose the values of the parameters
    - For exponential when \lambda > 1 we have negative entropy, just easier to consider all positive
* Regression
  + Explain the arrows
  + Draw the resultant dfs
* Regression 2
  + Explain each of the areas of analysis- say why it’s being done and how

Results Slides 15-22

* First 3 graphs (make note of pages that zoom is on in paper)
  + just discuss obv log n against log bias
* Comparison graphs
  + explain what the graph is depicting -
  + overall graph - implies
* k against c graphs
  + first two imply some sort of relationship
  + exponential is inconclusive

Conclusion 23-24

* Research questions
  + Come back to discuss results in terms of them

Questions? 25