University of California, los Angeles
Department of Statistics

Metric Entropy

Subtitle

Authors:
Melody Huang
Conor Kresin
Thomas Maierhofer

Stats 200C:
Spring Quarter 2019,
Prof. Arash Amini

Final Project Report

March 22, 2019

Contents

Abstract

TODO

1 Background

1.1 p Norms for Vectors

This section formally introduces p norms with a focus on the special cases $p = 0, 1, 2, \infty$. The p norm of a vector $x \in \mathbb{R}^d$ is denoted as $||x||_p$ and defined as

$$||x||_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{1/p},$$

where $|x_i| = \text{sign}(x_i)x_i$ denotes the absolute value. The most important special cases include the 0 norm is which counting the non-zero entries in x,

$$||x||_0 = \sum_{i=1}^{d} \nVdash\{x_i \neq 0\},$$

the 1 norm, a.k.a. city-block norm,

$$||x||_1 = \sum_{i=1}^{d} |x_i|,$$

the 2 norm, a.k.a. Euclidean norm,

$$||x||_2 = \sqrt{\sum_{i=1}^{d} x_i^2},$$

and the ∞ norm, a.k.a. maximum norm,

$$||x||_\infty = \max_{i=1,\ldots,d} |x_i|.$$

A unit ball for a norm $||\cdot||$ contains all points with distance 1 around the origin, i.e. all points $\{x : ||x|| = 0\}$. The unit balls for the 0, 1, 2, and ∞ norm in $\mathbb{R}^2$ are depicted in Figure 1. Note that the volume of the unit balls increases in p.

1.2 Entropy

Entropy measures how far from deterministic a random variable is, i.e. entropy quantifies a random variables diversity, uncertainty or randomness. A random variable with entropy zero is deterministic, an increasing value means that it is more and more unpredictable.

1.2.1 Shannon Entropy

Let X be a discrete random variable taking values in the set $\{x = x_1, x_2, \ldots, x_n\}$ with probabilities $\{p = p_1, p_2, \ldots, p_n\}$. The Shannon entropy $H(X)$ is defined as

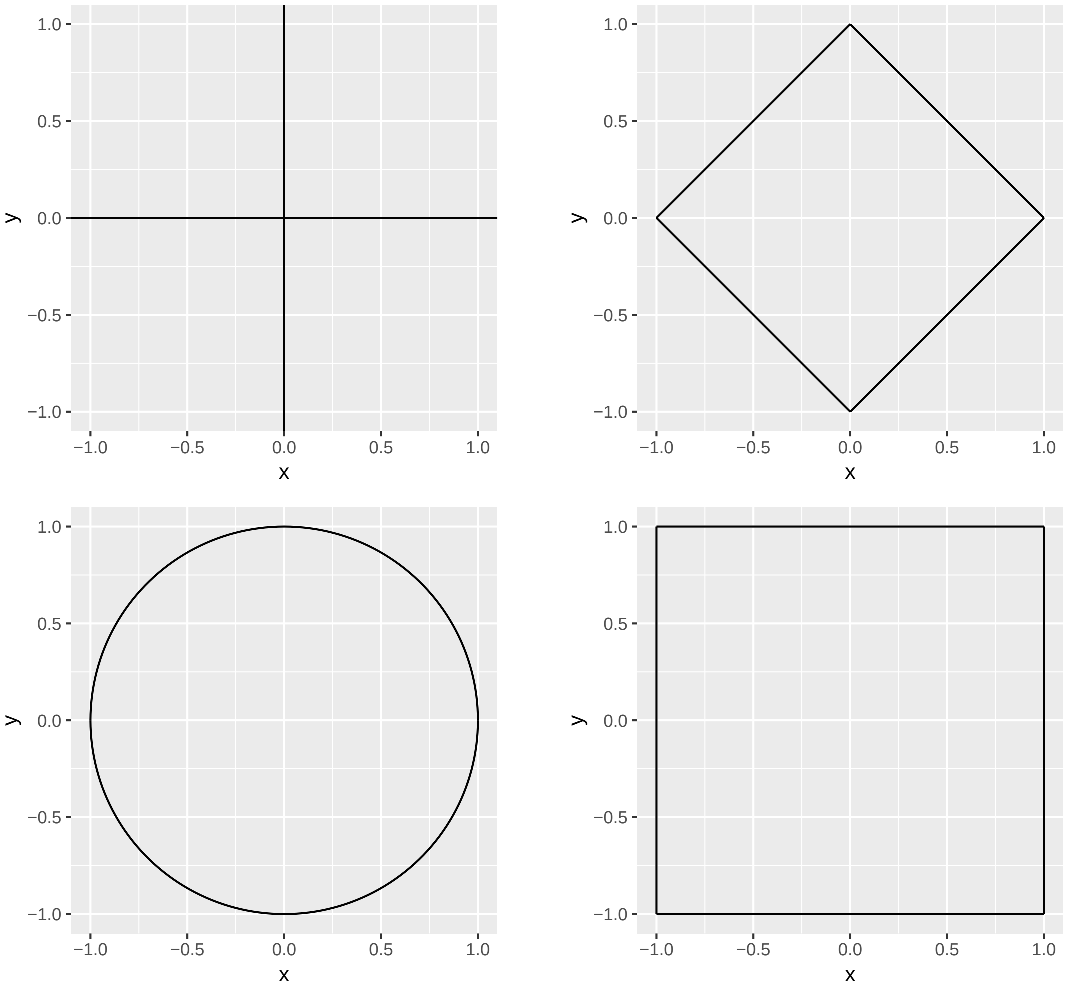

Figure 1: Unit balls in $\mathbb{R}^2$ of the 0 (top left), 1 (top right), 2 (bottom left), an ∞ norm (bottom right).

$$H(X) = H_1(X) = -\sum_{i=1}^{n} p_i \log_2 p_i, \tag{1}$$

and obtains its maximum value for the uniform distribution

$$p_i = \frac{1}{n} \forall i.$$

Considerations of continuity lead to the adoption of the convention, $0 \log 0 = 0$.

1.2.2 Rényi Entropy

The Rényi entropy of order q, $q \geq 0$ and $q \neq 1$, of a discrete random variable X taking n values with probabilities $p_1, \ldots, p_n$, is defined as,

$$H_q(X) = \frac{1}{1-q} \log \left(\sum_{i=1}^{n} p_i^q \right) = \frac{1}{1-q} \log \|p\|_q^q, \tag{2}$$

using the p-norm of order q of $x \in \mathbb{R}^n$ which is defined as

$$\|x\|_q = \left(\sum_{i=1}^{n} |x_i|^q \right)^{1/q}, \text{ for } q \geq 1 \in \mathbb{R}.$$

The most important special cases are $q = 0$ (count norm) for the max-entropy,

$$H_0(X) = \lim_{q \to 0} \frac{1}{1-q} \log \|p\|_q^q = \log |X|,$$

which is the logarithm of the cardinality of X, $q = 1$ (Manhattan norm) for the Shannon entropy [1], see Equation (1), $q = 2$ (Euclidean norm) for the quadratic entropy or collision entropy,

$$H_2(X) = \frac{1}{1-2} \log \|p\|_2^2 = \log \sum_i p_i^2,$$

and $q = \infty$ (maximum norm) used in the Min entropy,

$$H_\infty(X) = \lim_{q \to \infty} \frac{1}{1-q} \log \|p\|_q^q = \log \max_i p_i,$$

This means that the parameter q changes the way the shape of the distribution influences the entropy by designating the norm being used. Different parameters q make the Reényi entropy more or less sensitive to certain characteristics of the probability distributions. Note that the Rényi entropy is monotonically decreasing in q, as the q norms are monotonically decreasing and the log is a monotonic transformation.

2 Metric Entropy

Metric entropy quantifies the size of a set T. This task is superficially unrelated to the concept of entropy, but after closer inspections similarities apparent.

2.1 Covering and Packing

[1] or more precisely $\lim_{q \to 1} H_q(X) = H(X)$

References

Maierhofer, T. (2017). Classification of functional data - interpretable ensemble approaches., *MS thesis. Ludwig-Maximilians-Universität, München* . Available at https://epub.ub.uni-muenchen.de/41008/.