Bijective tranformation with Information Volume change

Author Franc Jarnovich

Retired computer programmer
Parizlje 64 d

Braslovce, 3314, Slovenia
franc.jarnovic@siol.net

Abstract*: The primary goal of this document is to show that order exist in every binary sequence beyond some length. According to Shannon Information Law , sequences ordered in any way can be compressed. Presented is one of the possible transformations that allows compression at least one bit in one pass until finite final length is achieved.*

1. Introduction

First i would like to mention simple transformations that don't make no change in Shannon Information volume. Simple example is to convert byte stream to two hexadecimal streams with the same probability distribution as input. (Fig. 1) Ideal lossless compression will produce the same result in both ways. Actually , Information Volume is different, but this difference is allways less than 1 bit.

But there are also transformations that change probability distributions. Let's split input stream into two or more fractions and ideal compression of input and transformed one will produce either the same size.Example is splitting byte stream into 257 parts , in wich the first part is Euclidian distance between ordinal values of two elements and 256 corresponding parts according to this distance. Now probaity distribution of the first stream differs due to input. (Fig 2.). Reverse procecess exist and it is obvious.

Major transformations causes increase of information in this manner, some leave information of the same size - but very few of them decrease information. This transformations is of my interest to introduce them to you.

Information decrease

Suppose that we have input not compressible by ideal compression methods. Now we split input in 4 bit format into sequence with probability distribution shown in (Fig3.) and 2 sequences with uniform distribution. After merging this three inputs in specified way that ensure bijection and split them into transformed first and 16 dependent fractions, we get desired result – increased Information in first and decreased Information in other 16 parts. And final, sum of output Information is significal less or equal to sum of Information of said three inputs. Shown are probability distributions after Transformation.(Fig. 4). Processing time consumption is linear due to data size.

Math used in entropy calculation and Information Volume

Instead of classical **ent** program, I choose different approach to estimate Information Volume. It is a number of bits required to encode number of all possible combinations of input n elements in sequence. Formula is $I = \log_2 (n!/(ef1! * ef2! * ef3! \dots efn!))$ where sum of elements frequencies is n. To avoid huge numbers produced with this formula, we convert it to I = (sum(1..n,i)log(i)- sum(1..nf,i)(sum(1..efi,j)log(j))/log(2). For Markov chain when n goes to infinity it is quite good approximation for Shannon Information of channel. This formula is used to detect and compute Information change.

There is hypothesis that is not formally proven, but practical simulations confirmes it.

I0 = const. , I(N+1)=f(I(N)) and I(N)=g(I(N+1)), f(g(I(N)))=I(N) where N is iteration number. With words: All binary sequences beyond I0 is representation of I0 and time where I0 is an irreducible pattern. These irreducible patterns seems to be no longer than several megabytes accordig to siumulations.

Conclusion.

There is a wide range of this Transformation usage, complete code and demo available on Github.com/brainmover/Cpball

2. Figures and Tables

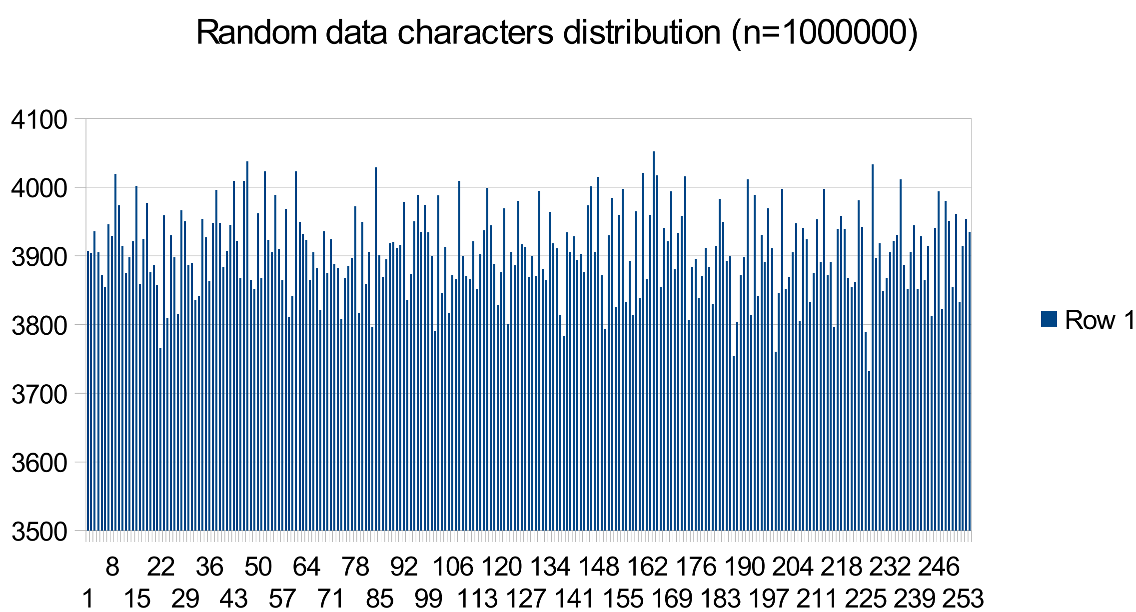


Figure 1

Figure 2

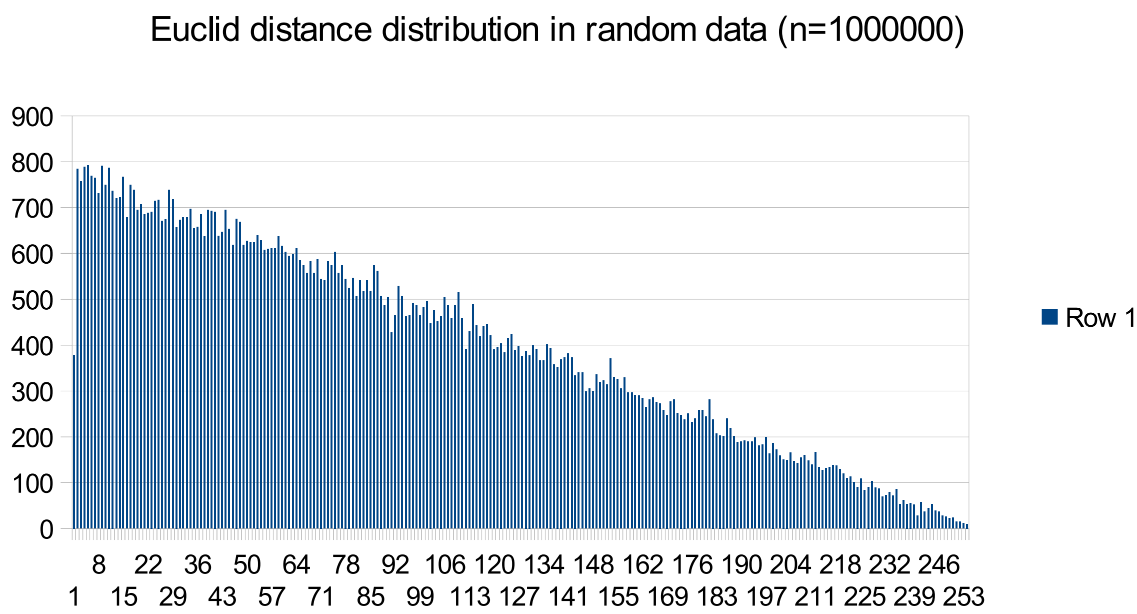


Figure 3

It is obvious that this fractions are compressibe with entropy coder

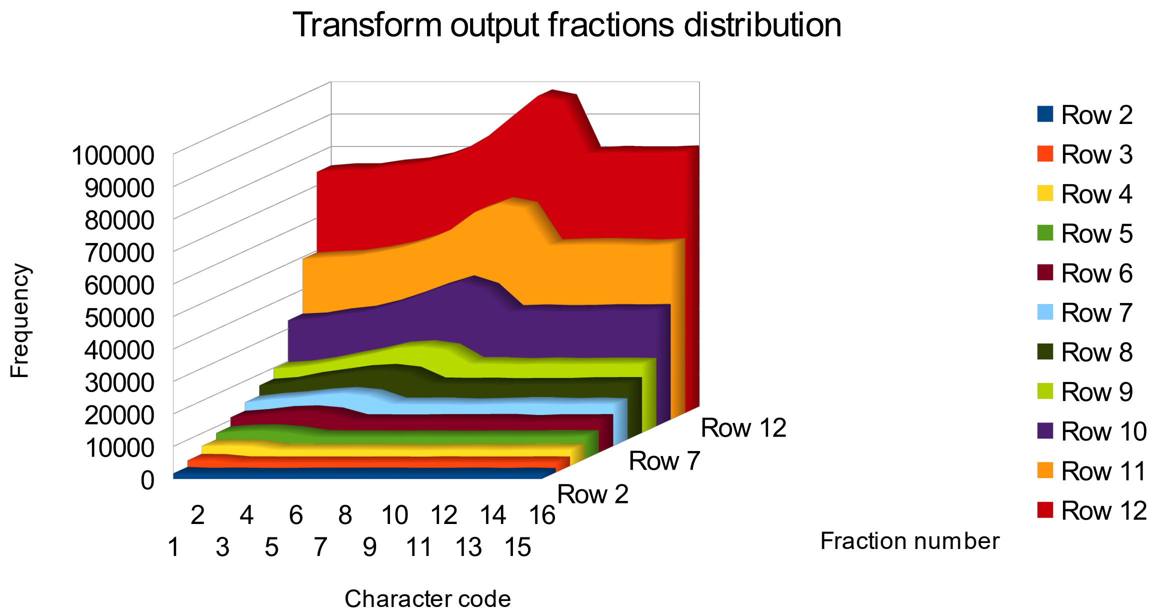
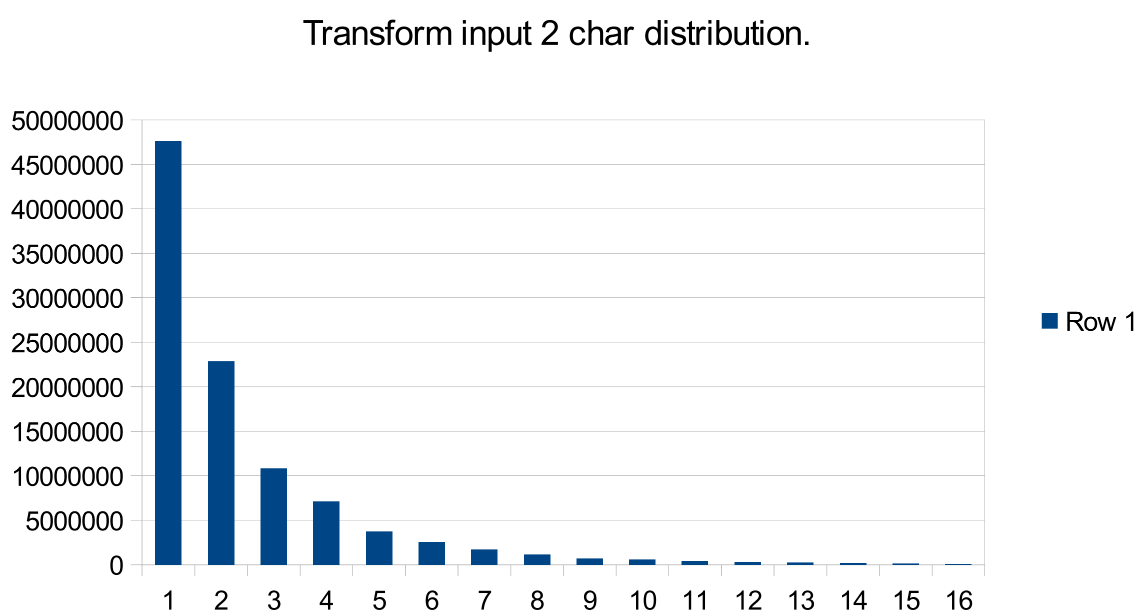


Figure 4



Franc Jarnovič

Digitalno podpisal Franc Jarnovič
Datum: 2023.01.18 23:14:06 +01'00'

References

[1] C. E. Shannon, “A mathematical theory of communication”