

HYGO FREIRE TAGARRO

Uma grande parte das coisas que os cientistas modernos possuem uma escala de tamanho típico. Entretanto, algumas coisas apresentam uma grande variação de escala, com muitos ordens de grandeza de diferença, o que torna difícil a visualização dos dados.

Uma boa parte dos dados que queremos representar possuem uma dispersidade de escala em uma pequena quantidade de medidas, isso faz que ao representar essa informação de forma visual, por exemplo, em um histograma a mesma fique distorcida. Esse caso é conhecido como lei de potência e ocorre na vida real.

O artigo cita maneiras de como realizar uma melhor representação dos dados, cita o uso de uma escala logarítmica, que já representa um ganho na visualização, entretanto ainda temos problemas com essa estratégia pois na cauda do histograma existe uma grande quantidade de ruído. Para resolver isso são propostos algumas alternativas:

- Construção de um histograma usando "logarithmic binning"
- Distribuições cumulativas complementares (CCDF), na CCDF conseguimos evitar a perda de informações que ocorre no caso acima, e ao utilizar a escala logarítmica, a relação original da lei de potência se transforma em uma linha reta, devido a relação exponencial virar uma linha reta.