

Re: Identifying the distribution of a data set

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2006-08/msg00437.html>

- *From:* john2 <john2@xxxxxxxxxxxxxxxxxxxx>
 - *Date:* Fri, 11 Aug 2006 08:16:23 +0100
-

ali wrote:

Dear all

I am creating a software that is reading tcp packets from a link. I have the following information available:

size of the packet: 1 4 7 9.....
frequency of the packet: 12 6 9 1.....

This is just an example. In reality I have thousands of these values.

Now I want to check what distribution fits the packet sizes best for e.g whether the distribution is Poisson or hyper exponential or Pareto or Gamma etc.

One way I guess is to plot a histogram and then study the shape. But I want the task to be fully automated and performed implicitly by the tool that I am developing.

While I hope you find something, it may be a futile exercise. Packet sizes sometimes approximate a self-similar alpha stable distribution which is similar to a Pareto, (search for "packet size distribution" with Google), but can also be heavily modified by routing, TCP flow control and queue management algorithms when different data streams are combined so, in practice, almost anything can appear. Trying to fit an actual name to such distributions is probably not possible in general.

john2

.