Subject: Re: histogram, how to trasfer from linear bins to logarithmic bin?
Posted by Jeremy Bailin on Thu, 01 Oct 2009 18:53:31 GMT
View Forum Message <> Reply to Message

On Sep 30, 3:29 pm, David Fanning <n...@dfanning.com> wrote:
> JD writes:
>> Very good point. The sort is over the bin vector, which can be (and
>> usually is) much shorter than the data vector. And you will likely
>> setup your bin boundary vector sorted to begin with. That said, for
>> me HISTOGRAM(ALOG10) is still faster than HISTOGRAM(VALUE_LOCATE) (see
>> below). You'll also note some "sky is falling" razors-edge
>> differences between bins if you look closely.
>
>> Hist(log)) 1.9417701
>> Hist(value_locate) 3.7843559
>
> By the way, when I ran your example on my (aging) Windows
> machine, I got these results:
>
> Hist(log)) 6.6090002
> Hist(value_locate) 5.1719999
>
> Hard to say what *that* means. :-)
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming:http://www.dfanning.com/
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

It should also depend on the number of bins... for M bins and an N element data vector, the hist(log) version should work in O(N) while the hist(value_locate) version should work in O(N log M + M) (the final +M is for setting up the bin cutoffs, but in most cases N>>M and it doesn't matter). So it depends how well-optimized log is vs. the log of the number of bins... which I can imagine could vary between architectures and C libraries!

-Jeremy.