Subject: Doing chi square and/or lognormal fits to 1D data?
Posted by swingnut on Sun, 23 Jul 2006 21:55:46 GMT

View Forum Message <> Reply to Message

I'm trying to analyze several collections of power law fits. Previous work implies that the constants and coefficients of these power laws are lognormal and that the exponents are chi square with 2 degrees of freedom. We haven't been able to get ahold of the person who did that previous work for over a year, but the new data I have looks like it follows the same pattern. It is possible that he did his analysis in Matlab, but really we have no idea what he used.

I've searched the web and combed through lots of libaries, usenet posts, webpages, etc, but as far as I can tell, no one has built what I need: drop-in IDL routines that would let me do lognormal and/or chi square fits to data. mpfit (and PAN) looked promising, but according to the documentation they require 2D data to fit to (i.e., they require X-Y pairs), whereas I only have 1D data (the Y half of each pair). I'm not trying to find a dependence on some value; rather, I am trying to find an approximation of the distribution these values could have been drawn from.

Do you all have any suggestions? I could kludge the lognormal analyses in SASS and just overplot a histogram of the data with a lognormal using the parameters it spits out. I'm ok with that for my work, but I'm trying to set up a system that is mostly automated for future students (e.g., my advisor's new student, who made it clear she is not a coder of any sort).

The chi square fit, well, there's plenty of routines to do a goodness-of-fit test, but I didn't find any at all, not even any references that this project or that project has code to do it. Has anyone heard of an IDL routine for this?

Thanks for the help.

Subject: Re: Doing chi square and/or lognormal fits to 1D data?
Posted by Paolo Grigis on Wed, 26 Jul 2006 07:39:02 GMT

View Forum Message <> Reply to Message

Well, you mentioned earlier that one of the distributions you wanted was lognormal... then the logarithms of your data should be normally distributed --> find the mean & standard deviation of that distribution (with errors if you like): no binning required there! Then I guess it should not be too hard to figure out how to transform the normal distribution in log-space back to lognormal distribution in linear space...

Ciao,
Paolo

swingnut@gmail.com wrote:
> Thanks for the info. Between the webpages for mpfit and PAN, the
> documentation looked like it wouldn't work with "univariate data".
>
> Yes, you are right, I wasn't particularly clear about what I was trying
> to describe. I've been thinking about this for three days, and you just
> can't reliably use (bin counts,bin centers/edges) as (x,y) and then
> fit. The problem is that bin counts are entirely too sensitive to bin
> width. See e.g,
>
> http://arxiv.org/abs/physics/0605197
> http://www.mathworks.com/products/statistics/demos.html?file
=/products/demos/shipping/stats/cfitdfitdemo.html.
>
> What I want to do is fit for the parameters of the probability
> distribution that would reasonably represent a single column of data,
> without any errors availalbe. I'm thinking that bootstrapping to get
> error estimates is fine, since I have no idea how to generate them. (I
> didn't do the original algorithm, and my advisor has literally no clue
> about the statistics of it -- she drops numbers into a black box and
> applies the standard rules of thumb to interpret the output from the
> black box.) I'll keep cranking away til I figure it out.
>

Subject: Re: Doing chi square and/or lognormal fits to 1D data?
Posted by swingnut on Wed, 26 Jul 2006 10:38:12 GMT
View Forum Message <> Reply to Message

Paolo,

Funny you should suggest that. I went back and reviewed the previous
analysis, and sure enough that's what had been done. He didn't even
bother to transform back, just left it in log-space all nice, pretty,
and uncomplicated for publication...

Thank you for the help.

Subject: Re: Doing chi square and/or lognormal fits to 1D data?
Posted by swingnut on Sun, 30 Jul 2006 05:05:51 GMT
View Forum Message <> Reply to Message

So, after enough reading, I've determined that my predecessor was less than accurate in describing his work. First, I misinterpreted what he meant in describing his chi square fits -- he calls it a "2 parameter" chi square distribution. After realizing this, I set off searching for this mystical beast I had never heard of, given that your standard chi square distribution has only one parameter, the degrees of freedom. There is indeed a two paramter chi square distribution, the noncentral chi square. This is generated by adding up squares of normal random variables that are not standard, i.e., the variances of each RV is 1, but the means are not required to be zero. (Info at
 http://en.wikipedia.org/wiki/Noncentral_chi-square_distribut ion.)

At this point, I stopped to review my predecessor's thesis again. The equation he provided is not the pdf or the probability function for the noncentral -- it has a scale parameter, not an offset -- so I went searching for generalizations and related distributions. Turns out the mystery distribution is the gamma distribution (http://en.wikipedia.org/wiki/Gamma_distribution), so I'll probably just use Matlab's library fuction to fit that chunk of data.

[Ed. note: Written for the sake of helping future google searchers.]