Subject: Re: Doing chi square and/or lognormal fits to 1D data?
Posted by Craig Markwardt on Mon, 24 Jul 2006 10:25:27 GMT
View Forum Message <> Reply to Message

swingnut@gmail.com writes:
> 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.

MPFIT does not require an "X" value. That is entirely up to you and your model function. But I'm not sure I get it. If you have a distribution of values, then you can make a histogram and the bin numbers are implicitly "X" values. The chi-square and lognormal probability density distributions -- used as model functions -- are easily found on the web [*]. They are almost trivial to code in IDL, (untested!)

```
function chisqr_density, x, nu
  return, exp(-x/2)*x^(nu/2.-1) / (2^(nu/2.)*gamma(nu/2.))
end

function lognorm_density, x, m, sigma, theta
  return,  exp(-((alog((x-theta)/m))^2/(2.*sigma^2)))/((x-theta)*sigma* sqrt(2*!dpi))
end
```

[*] Example of probability distributions
 http://www.itl.nist.gov/div898/handbook/eda/section3/eda366. htm

> 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?

Are you serious? There are zillions of chi-square fitting routines
for IDL. Half of them are in IDL itself. [And half of a zillion is
still a very large number.] LINFIT, CURVEFIT, MPFIT, SVDFIT, etc.

If you have a model function and data, you can use either CURVEFIT or
MPFIT. I suspect that you are defining chi-square fitting in some
other way...

Craig

--
 -- --------------
Craig B. Markwardt, Ph.D.   EMAIL: craigmnet@REMOVEcow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
 -- --------------

Subject: Re: Doing chi square and/or lognormal fits to 1D data?
Posted by swingnut on Wed, 26 Jul 2006 02:28:56 GMT

View Forum Message <> Reply to Message

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.