Subject: Re: How to determine the WEIGHTS in MPFITFUN for distribution function fit?
Posted by Vince Hradil on Thu, 06 Sep 2007 13:52:50 GMT
View Forum Message <> Reply to Message

On Sep 6, 2:54 am, "dux...@gmail.com" <dux...@gmail.com> wrote:
> I have a set of data and want to use a function to fit its
> distribution function.
> I have downloaded the procedure MPFITFUN and MPFIT.
>
> The data set is divided into many bins by a binsize.
> X is the mean value of the data point in each bin, and Y is the
> statistical count (or normalized) of the data point in each bin.
>
> Then I want to use the function 'myfunt' to fit the X and Y, and
> compute the parameter P for 'myfunt'.
> For example: p = mpfitfun ('myfunt', X, Y, Yerror, start)
> But I do not konw how to determine Yerror.
> If Yerror is set to unity, the 1-sigma errors for the estimated
> parameter P cannot be returned correctly from MPFITFUN.
>
> Best regards,
> jdu

Boy, that's a good question ... Here are some questions for you.
How do you measure the counts? How sure are you of the number of counts?
If they are truly counts, the errors may be Poisson distributed, yerr ~ 1/y. Have you tried this?
Are you really interested in the standard errors of the parameters, or just the parameters? The estimate of the parameters is fairly robust with mpfitfun, so the P values are good estimates regardless of the estimated error.

If you can't estimate the 1-sigma errors for Y you could always use the bootstrap to estimate the P-errors. I've found this very helpful - and really easy to do.

Subject: Re: How to determine the WEIGHTS in MPFITFUN for distribution function fit?
Posted by Craig Markwardt on Thu, 06 Sep 2007 18:26:20 GMT
View Forum Message <> Reply to Message

hradilv <hradilv@yahoo.com> writes:
> Boy, that's a good question ... Here are some questions for you.
> How do you measure the counts? How sure are you of the number of
> counts?

> If they are truly counts, the errors may be Poisson distributed, yerr
> ~ 1/y. Have you tried this?

If Y is measured counts, then the YERR should be ~ SQRT(Y), right?

But overall I agree, the original poster needs to specify how the
numbers were measured in order to estimate the measurement error.

Craig

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

Subject: Re: How to determine the WEIGHTS in MPFITFUN for distribution function fit?
Posted by duxiyu@gmail.com on Fri, 07 Sep 2007 04:43:39 GMT
View Forum Message <> Reply to Message

> Boy, that's a good question ... Here are some questions for you.
> How do you measure the counts? How sure are you of the number of
> counts?
> If they are truly counts, the errors may be Poisson distributed, yerr
> ~ 1/y. Have you tried this?

Yerr ~ 1/y? Do you means that the keyword WEIGHTS should be 1/y?

> Are you really interested in the standard errors of the parameters, or
> just the parameters? The estimate of the parameters is fairly robust
> with mpfitfun, so the P values are good estimates regardless of the
> estimated error.
>
> If you can't estimate the 1-sigma errors for Y you could always use
> the bootstrap to estimate the P-errors. I've found this very helpful
> - and really easy to do.

Could you explain for me what the bootstrap is?
Is it a IDL procedure or a mathematic method to estimate the P-error?
I cannot find it in the IDL Help files.

Thanks for your reply.

jdu

Subject: Re: How to determine the WEIGHTS in MPFITFUN for distribution function fit?
Posted by duxiyu@gmail.com on Fri, 07 Sep 2007 05:26:48 GMT
View Forum Message <> Reply to Message

On 9 7 , 2 26 , Craig Markwardt
<craigm...@REMOVEcow.physics.wisc.edu> wrote:
>
> If Y is measured counts, then the YERR should be ~ SQRT(Y), right?
>
> But overall I agree, the original poster needs to specify how the
> numbers were measured in order to estimate the measurement error.
>
> Craig
>

The effect of Yerr=SQRT(Y) is the same as the effect of WEIGHTS=1/y
for the procedure MPFITFUN, isn't it?

The array named D in IDL means the original mesured data.
Y = histogram (D, locations = X)
I want use a specified function to fit this curve in the x-y plane.
The computed function is the distribution function for the measured
data D, isn't it?
If there are not the measurement errors for the data D, I should set
the Yerr = SQRT(Y) when I use the MPFITFUN to fit the curve.
If there ara the measurement errors Derr which have the same
dimensions as the data D, how do I caculate the Yerr?

The another question is how to determine the START_PARAMS in the
procedure MPFITFUN?
When I use the matlab to fit the curve, the procedure can select the
start_value automatically.
The START_PARAMS in <PFITFUN must be given by user.
The START_PARAMS is set to some values, and the procedure can return
the results without any error.
But when the START_PARAMS change a little, the procedurel gives the
warning "Program caused arithmetic error: Floating illegal operand"
and return the different results.
I do not know the criterion of its selection.

I am looking forward to your reply.

Best regards,
jdu

Subject: Re: How to determine the WEIGHTS in MPFITFUN for distribution function

fit?
Posted by Craig Markwardt on Tue, 11 Sep 2007 05:21:20 GMT
View Forum Message <> Reply to Message

"duxiyu@gmail.com" <duxiyu@gmail.com> writes:

> On 9 7 , 2 26 , Craig Markwardt
> <craigm...@REMOVEcow.physics.wisc.edu> wrote:
>>
>> If Y is measured counts, then the YERR should be ~ SQRT(Y), right?
>>
>> But overall I agree, the original poster needs to specify how the
>> numbers were measured in order to estimate the measurement error.
>>
>> Craig
>>
>
> The effect of Yerr=SQRT(Y) is the same as the effect of WEIGHTS=1/y
> for the procedure MPFITFUN, isn't it?

Yes.

> The array named D in IDL means the original mesured data.
> Y = histogram (D, locations = X)
> I want use a specified function to fit this curve in the x-y plane.
> The computed function is the distribution function for the measured
> data D, isn't it?

Yes, that's one way of thinking of it.

> If there are not the measurement errors for the data D, I should set
> the Yerr = SQRT(Y) when I use the MPFITFUN to fit the curve.
> If there ara the measurement errors Derr which have the same
> dimensions as the data D, how do I caculate the Yerr?

I don't know. For HISTOGRAM() either the point falls within a bin or
it does not. If your "D" has measurement errors, the bin assignment
becomes ambiguous. This is usually solved with Monte Carlo methods
and some kind of probability matrix.

> The another question is how to determine the START_PARAMS in the
> procedure MPFITFUN?

That is up to you. As I point out on my "tutorial" page, this is
somewhat of an art.

> When I use the matlab to fit the curve, the procedure can select the
> start_value automatically.

That sounds interesting, but I would not trust the automatic method.
There are an infinite number of model functions, and for each model,
there are an infinite number of parameterizations. How could 'matlab'
know about every one?

> The START_PARAMS in <PFITFUN must be given by user.
> The START_PARAMS is set to some values, and the procedure can return
> the results without any error.
> But when the START_PARAMS change a little, the procedurel gives the
> warning "Program caused arithmetic error: Floating illegal operand"
> and return the different results.
> I do not know the criterion of its selection.

MPFIT provides diagnostic information (STATUS and ERRMSG keywords),
and numerical errors can be trapped with !EXCEPT. These may help you
solve your problem. One possibility to consider is that there may be
a numerical problem if Y is zero (so YERR is also zero). In that
case, the Gaussian approximation underlying YERR ~ SQRT(Y) no longer
holds.

Good luck,
Craig

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

Subject: Re: How to determine the WEIGHTS in MPFITFUN for distribution function fit?
Posted by Vince Hradil on Tue, 11 Sep 2007 14:15:02 GMT
View Forum Message <> Reply to Message

On Sep 6, 11:43 pm, "dux...@gmail.com" <dux...@gmail.com> wrote:
>> Boy, that's a good question ... Here are some questions for you.
>> How do you measure the counts? How sure are you of the number of
>> counts?
>> If they are truly counts, the errors may be Poisson distributed, yerr
>> ~ 1/y. Have you tried this?
>
> Yerr ~ 1/y? Do you means that the keyword WEIGHTS should be 1/y?
>
>> Are you really interested in the standard errors of the parameters, or
>> just the parameters? The estimate of the parameters is fairly robust
>> with mpfitfun, so the P values are good estimates regardless of the
>> estimated error.

>
>> If you can't estimate the 1-sigma errors for Y you could always use
>> the bootstrap to estimate the P-errors. I've found this very helpful
>> - and really easy to do.
>
> Could you explain for me what the bootstrap is?
> Is it a IDL procedure or a mathematic method to estimate the P-error?
> I cannot find it in the IDL Help files.
>
> Thanks for your reply.
>
> jdu

http://en.wikipedia.org/wiki/Bootstrapping_%28statistics%29

Bootstrapping is a statistical method of resampling. The simples way
(but perhaps not the correct depending on the problem) to apply
bootstrapping for your problem would be:

1- estimate yhat using mpfit
2- create an array of (yhat-y), the residuals.
3- resample the errors (with replacement) to get a bootstrap sample of
residuals.
4- add the bootstrap residuals to yhat -> yboot
5- fit yboot using mpfit -> pboot (the parameters from the fit)
6- repeat 3-5 several (about 200 times should be sufficient)
7- determine confidence limits of the parameters from the distribution
of pboot

It's really quite simple, and has worked really well for me in the
past.

Look at the wikipedia entry and the references for further details.
Especially: http://www.crcpress.com/shopping_cart/products/product_detai
l.asp?sku=C4231&isbn=0412042312&parent_id=&pc=
(a very good read!)
