Subject: Re: Simultaneous fitting in IDL
Posted by Allan Whiteford on Fri, 15 May 2009 11:15:36 GMT
View Forum Message <> Reply to Message

Gianluca Li Causi wrote:
> Hi all,
> I need to *simultaneously* fit two sets of data, Y1 and Y2, with two
> functions, F1(X,P) and F2(X,P), having the same set of parameters P,
> by using the routine LMFIT.
>
> In order to do this I join the two datasets in a uniq vector Y=[Y1,Y2]
> and do the same for the model F=[F1,F2], as suggested also by Craig
> Markwardt in a past thread in this group:
> http://groups.google.com/group/comp.lang.idl-pvwave/browse_t
hread/thread/ba341b5169e7557f/498234b82c67279f?hl=en&lnk =gst&q=lmfit#498234b82c67279f
>
> But my datasets contains a different number of points, N1 and N2, with
> N2<<N1, so that the smaller one is almost neglected because its weight
> in the uniq vector Y is proportional to N2 / N1.
>
> SO THAT: I've tried to multiply each error vector by the SQRT of its
> fraction:
>
>   Y1_err *= SQRT(N1 / (N1+N2))
>   Y2_err *= SQRT(N2 / (N1+N2))
>
> and now I get a nice fit for the two dataset simultaneously!
>
> BUT: when I compute the Reduced ChiSquare = ChiSquare / (N1 + N2 -
> N_parameters) I get a completely wrong result (vey high, far from
> 1.0), due to my modification of the true error vector !
>
> THUS: what I can do ?
> One idea is to force the Y1 and Y2 vectors to have the same number of
> elements M, which must be the minimum multiple of both N1 and N2, but
> it could be a very large number.....
>
>
> Could anybody help me ?!
> Thanks a lot!
> Gianluca
>
>
>

Gianluca,

If you really believe your data and, more importantly, really believe

the error bars associated with them then it's formally correct that all of your Y2 data points are essentially being ignored. Bet you're sorry you went to the trouble of measuring them now? :)

The simple scenario when they wouldn't be ignored would be when one (or more) of the parameters only has a significant effect on the modelling of Y2 data. This doesn't seem to be the case here since when you mess with the errors you do get a change in your fit.

Are the errors on your points (particularly your Y1 points) truly random? Chances are your Y1 error estimates don't contain only random errors but also have some form of systematic error - hence correlation between the points which is a whole new can of worms.

In the case of correlated errors, your weight vector needs to become a matrix with the diagonal elements being your weight as before and the off-diagonals being the correlation between points (you'll need to invert this matrix along the way - hopefully it's not too big). When you include this (assuming it has off-diagonal elements which are significant) one of the results will be that more attention is paid to Y2 data making you glad you went to the trouble of measuring them :).

Scientists typically compensate for this whole complicated correlation problem by messing around with weighting of the errors in a similar way to what you have been trying - it can often work quite well. Sometimes they realise what they are compensating for but mostly they just do it because they get curves which look nicer and they have a gut feeling that some measurements have to count for something.

After you've messed with your errors/weights though, don't expect formal statistical tests to have much meaning. Your problem in this case is that your chi2 statistic is still behaving correctly and saying "most of these points are further away than they should be" because you've mangled the fit so that it moves Y1 points away to compensate for Y2 points. Who knows what the errors in your measured parameters now represent... but they won't be correct!

If you measured how long it took for a ball to drop 100 times (N1=100) and the guy next door measured how long it takes something to roll down a slope once (N2=1) would you really want his single data point to count as much as all of your points put together just because be did it in a different room with a different piece of equipment? (Note: if his experiment was better then it would come with a lower error bar so would already be given a higher weight in a fit - you don't need to compensate for this twice.) Maybe you would want his to count a bit higher if you thought your error was partly based on a systematic problem with, e.g., your stopwatch or general setup meaning all of your points might be skewed in the same way... i.e. correlated error between your data points.

Don't go near a statistics department for a few days, they may shoot you
on sight :). Also, guys responsible for detectors love being asked for
the correlation between every point with every other measured point,
have fun with that conversation :).

Thanks,

Allan