Subject: Re: Calculating Error Estimates
Posted by Wayne Landsman on Tue, 14 Jul 1998 07:00:00 GMT
View Forum Message <> Reply to Message

Before giving my answer David Fanning's question about estimating
errors from
CURVEFIT, I first want to post a related complaint about the procedure
SVDFIT. SVDFIT has an output parameter SIGMA which according to the
documentation gives the "1-sigma error estimates of the returned
parameters." What the documentation doesn't say is that unless you
supply
values in the optional WEIGHTS input keyword, then the values returned
by SIGMA
are likely to be complete nonsense.

In order to compute the uncertainties in the derived parameters, SVDFIT
needs to
know the uncertainties in the individual data points. If the user
does not
supply these, then SVDFIT assume that the uncertainties are all equal
(which is
reasonable), and that they are all equal to 1 unit (which is likely to
be very
unreasonable).

What one should do -- and what David F. should do for his example -- is
to
estimate the uncertainties by comparing the data with the best fit.
This is
done correctly by the procedure LINFIT.

SIGMA = SQRT((TOTAL ((Y - YFIT)^2) / (NPTS - NPARAMS)))

where NPTS and NPARAMS are respectively the number of data points and
number of
parameters used in the fit.

As an example of the problem with SVDFIT, consider a linear fit with an
uncertainty of 0.01 in the Y coordinate

STIS> x = indgen(10)
STIS> y = x + 0.01*randomn(seed,10)

First, I try a linear fit with LINFIT and print the derived parameters
and
their sigma values

IDL> param = LINFIT(x, y, sigma = sig)

```
IDL> print,param
   -0.00540641      1.00106
IDL> print, sig
    0.00654651   0.00122627
```

Now try a linear fit (a polynomial with 2 parameters) of the same data
using SVDFIT

```
IDL> param = svdfit(x,y,2,sigma=sig)
IDL> print,param
  -0.00540641      1.00106
IDL> print,sig
   0.587754     0.110096
```

Although SVDFIT gives the right parameter values, the associated sigma
values
are much too large.   This is because SVDFIT assumed individual errors
of 1
unit, whereas LINFIT computed

```
yfit = param(1)*x + param(0)
sigma = total( (yfit-y)^2)/8. ))
```

to estimate individual errors of 0.011 units

Wayne Landsman
landsman@mpb.gsfc.nasa.gov

Subject: Re: Calculating Error Estimates
Posted by   on Tue, 14 Jul 1998 07:00:00 GMT

View Forum Message <> Reply to Message

David Fanning wrote:

> Hi Folks,
>
> It is so hot in Colorado today that I think my brain has
> vapor locked. In any case, I can use some help. :-)
>
> Here is my problem. I have some experimental data. I
> have used CURVEFIT and my roll-your-own function to
> fit a curve through the data. What I want to do is
> display the experimental data on the plot, along with
> the curve. But I want to place error bars through
> the experimental data points. My question is this:
> how do I calculate the errors for the individual
> points so that I can place them on the plot with

> ERRPLOT?
>
> CURVEFIT returns to me a parameter called SIGMA,
> which contains the standard deviations of the returned
> values of the four coefficients in my fitting
> function. What I cannot seem to work out is how
> to use these standard deviations to obtain an
> error estimate for each individual experimental
> point.
>
> I realize this is basic error analysis, but even
> an hour spent refreshing myself with Bevington
> has not successfully stimulated this reptilian brain. :-(
>
> Let's just say I had too much fun on vacation...
>
> Thanks,
>
> David
> --
> David Fanning, Ph.D.
> Fanning Software Consulting
> E-Mail: davidf@dfanning.com
> Phone: 970-221-0438
> Coyote's Guide to IDL Programming: http://www.dfanning.com/

Hi David,

I think you can not obtain error estimates for your data points from
your fit.

The errors of your data points depend on your measurement, i.e. they
have nothing to do with your fit.
For example, when you count different possible events, where you expect
e.g. a gaussian distribution, your data point error is - in first order
- the statistical error, and can be calculated just with:
sigma_y = SQRT(y)
(This is correct only for large values of y, for small values of y you
have to use Poisson-statistics).
Of course you have to take into account all your error sources,
depending on the experimental setup.
But i normally use the above stated SQRT(y), the statisitcal error, for
the errorbars.

For a proper fit, e.g. with CURVEFIT, you need this errors to calculate
the weight w of each datapoint.
There i use w = (1.0 / sigma_y^2.0) .

The standard deviations "sigma_a" for the different parameters in the fit are important for the goodness of your fit (-> chisquared), but completely differ from the data point errors.

By the way, for every user who works with CURVEFIT: The standard deviations, calculated in the routine, are not correct, at least not in any sense known to me.
As i posted some time ago, we have a modified CURVEFIT routine which works proper and also you are able to constrain parameters at your own will.

I hope, i could help you, David,

bye,

Heiko Hï¿½nnefeld

HASYLAB / DESY

Notkestr. 85
22603 Hamburg
Tel.: 040 / 8998-2698
Fax.: 040 / 8998-2787
e-mail: hhuenne@desy.de

Subject: Re: Calculating Error Estimates
Posted by Karl Young on Wed, 15 Jul 1998 07:00:00 GMT
View Forum Message <> Reply to Message

Hi David,

> It is so hot in Colorado today that I think my brain has
> vapor locked. In any case, I can use some help. :-)

I was in Boulder last week for the World Shakuhachi Conference and Imust admit it was a real joy to get back to foggy San Francisco

> Here is my problem. I have some experimental data. I
> have used CURVEFIT and my roll-your-own function to
> fit a curve through the data. What I want to do is
> display the experimental data on the plot, along with
> the curve. But I want to place error bars through
> the experimental data points. My question is this:
> how do I calculate the errors for the individual
> points so that I can place them on the plot with

> ERRPLOT?
>
> CURVEFIT returns to me a parameter called SIGMA,
> which contains the standard deviations of the returned
> values of the four coefficients in my fitting
> function. What I cannot seem to work out is how
> to use these standard deviations to obtain an
> error estimate for each individual experimental
> point.
>
> I realize this is basic error analysis, but even
> an hour spent refreshing myself with Bevington
> has not successfully stimulated this reptilian brain. :-(
>
> Let's just say I had too much fun on vacation...
>

Glad to hear you had a good vacation.

I recently had a similar problem, i.e. getting error estimates for constrained min (not curvefit, but the solution is similar in both cases).
Since I couldn't get information like derivative matrices out of constrained min (like you can out of curvefit, since the source is available) I checked with the folks at Windward Technology (the company that supplies constrained min). They pointed me towards a nice source, the book "Applied Regression Analysis" by Draper and Smith. In chapter 10 of that book they obtain a formula that is quite general for nonlinear optimization problems, and should apply in your case. The formula for the "confidence ellipsoid", i.e. what you can use for error bounds is:

(P - Pfit)' Zfit' Zfit (P - Pfit) < p s^2 F(p,n-p,1-a)

where:

Pfit   is the vector of parameters obtained from your fit
P   is a vector of variables, which you can use to solve for the bounds
Zfit   is the Hessian or second derivative matrix with the fitted values

       of the parameters plugged in (I think you can snag that right out of
       curvefit, i.e. you don't have to actually obtain the second derivatives)
p   is the number of parameters
n   is the number of points fit
F   is the F-distribution

and   s^2 = S(Pfit)/(n-p)

with S(Pfit) just the sum of the squared errors between the data and your function at Pfit.

If you need more details fell free to bug me.

-- KY

--- -------
Karl Young   Phone: (415) 750-2158 lab
UCSF   (415) 750-9463 home
VA Medical Center, MRS Unit (114M) FAX: (415) 668-2864
4150 Clement Street   Email:kyoung@itsa.ucsf.edu
San Francisco, CA 94121
--- -------