Subject: On errors calculated by curve-fitting routines
Posted by Gernot Hassenpflug on Thu, 06 Mar 2008 03:08:49 GMT
View Forum Message <> Reply to Message

Hello all,

I'm using IDL 6.1, as well as Maple 11, Mathematica 6.0, Matlab 7.5 and the statistical language R. My goal is to calculate the covariance matrix of parameters of a second order polynomial curve fit. To clarify: I refer to this as linear fitting, since the parameters are linear; however, many books, papers and routines refer to this as non-linear fitting.

Matlab and Mathematica do not have built-in functions to do this (Mathematica has an add-on module which my institute has not bought) so I am comparing the parameter covariance matrix from IDL, Maple, R and my own programmed output learned from section 15.4 of Numerical Recipes, 2nd edition, and a paper by Keith Burrell in the American Journal of Physics Vol. 58, No. 2, pp 160--164 (1990) titled "Error analysis for parameters determined in nonlinear least-square fits", both describing the same method which uses the variances of the dependent data combined with the derivatives of the fitting function wrt the fitted parameters; i.e., the dependent data values themselves are not used, apparently.

I find that in IDL the routines POLY_FIT, LMFIT and CURVEFIT can all calculate the parameter covariance matrix and it is documented that LMFIT uses the method of Burrell and Numerical Recipes. I cannot tell what method the other two routines use.

Maple seems to use a different method apparently described on pp 197--198 of David M. Himmelblau's 1970 book titled "Process Analysis by Statistical Methods", which I have ordered used but not yet received.

I am hoping that contributors to this list could give their comments and opinions on what method of parameter variance and covariance is most sound, and which routines are therefore preferred for a polynomial fitting case (possibly over-determined).

Many thanks in advance,
Gernot Hassenpflug
--
BOFH excuse #72:

Satan did it

Subject: Re: On errors calculated by curve-fitting routines
Posted by Craig Markwardt on Fri, 07 Mar 2008 03:53:16 GMT
View Forum Message <> Reply to Message

Gernot Hassenpflug <gernot@nict.go.jp> writes:
> I find that in IDL the routines POLY_FIT, LMFIT and CURVEFIT can all
> calculate the parameter covariance matrix and it is documented that
> LMFIT uses the method of Burrell and Numerical Recipes. I cannot tell
> what method the other two routines use.

Anthony mentioned MPFIT, which is a non-linear fitting engine
translated from MINPACK. As far as I understand, the covariance
matrix is equivalent to that from Numerical recipes.

> I am hoping that contributors to this list could give their comments
> and opinions on what method of parameter variance and covariance is
> most sound, and which routines are therefore preferred for a
> polynomial fitting case (possibly over-determined).

For linear least squares, I think the covariance matrix is reasonably
useful. In my field, it's common to use the delta-chi-square method
described in Numerical Recipes, which usually involves making a
confidence grid for pairs of parameters that are of interest.

Craig

Subject: Re: On errors calculated by curve-fitting routines
Posted by Gernot Hassenpflug on Fri, 07 Mar 2008 05:08:42 GMT
View Forum Message <> Reply to Message

Craig Markwardt <craigmnet@REMOVEcow.physics.wisc.edu> writes:

> Gernot Hassenpflug <gernot@nict.go.jp> writes:
>> I find that in IDL the routines POLY_FIT, LMFIT and CURVEFIT can all
>> calculate the parameter covariance matrix and it is documented that
>> LMFIT uses the method of Burrell and Numerical Recipes. I cannot tell
>> what method the other two routines use.
>
> Anthony mentioned MPFIT, which is a non-linear fitting engine
> translated from MINPACK. As far as I understand, the covariance
> matrix is equivalent to that from Numerical recipes.
>
>> I am hoping that contributors to this list could give their comments
>> and opinions on what method of parameter variance and covariance is
>> most sound, and which routines are therefore preferred for a
>> polynomial fitting case (possibly over-determined).
>

> For linear least squares, I think the covariance matrix is reasonably
> useful. In my field, it's common to use the delta-chi-square method
> described in Numerical Recipes, which usually involves making a
> confidence grid for pairs of parameters that are of interest.

Craig, thank you very much for your response, I will probably spend the entire weekend comparing test cases between the various IDL routines, including your really impressive MPFIT set of packages, Maple, R, Numeric Python, and my own IDL code (using the Numeric Recipies idea).

I understand that multivariate "error" even for only real quantities does not have simple answers, and that covariance is not the only or even best estimator for many cases, as you point out.

I'd just like to ask, since I cannot quite tell if I have grasped the ideas from Numeric Recipes correctly (and so my own IDL code for comparison with the others may be incorrect): the covariance matrix calculation uses the basis functions (e.g., 1, x, x^2) and the variances of the dependent (y) variable, but *not* the dependent variable itself nor any quantitative measures of the goodness of the fitting process (presumably the variances of the dependent variable are supposed to contain all such information in theory).

I ask this because other methods, such as that used by Maple, seem to scale their result by the residual sums of squares, for example. I am still awaiting the book by Bevington (can only get 1st edition from library services, so need to purchase 2nd edition) and the one by Himmelblau from 1970 which is the basis of the Maple method.

Best regards,
Gernot Hassenpflug
--
BOFH excuse #100:

IRQ dropout

Subject: Re: On errors calculated by curve-fitting routines
Posted by Craig Markwardt on Sat, 08 Mar 2008 18:27:16 GMT
View Forum Message <> Reply to Message

Gernot Hassenpflug <gernot@nict.go.jp> writes:
> I'd just like to ask, since I cannot quite tell if I have grasped the
> ideas from Numeric Recipes correctly (and so my own IDL code for
> comparison with the others may be incorrect): the covariance matrix
> calculation uses the basis functions (e.g., 1, x, x^2) and the
> variances of the dependent (y) variable, but *not* the dependent

> variable itself nor any quantitative measures of the goodness of the
> fitting process (presumably the variances of the dependent variable
> are supposed to contain all such information in theory).

That is the formal definition of the covariance matrix, assuming the measurement uncertainties are appropriate.

> I ask this because other methods, such as that used by Maple, seem to
> scale their result by the residual sums of squares, for example. I am
> still awaiting the book by Bevington (can only get 1st edition from
> library services, so need to purchase 2nd edition) and the one by
> Himmelblau from 1970 which is the basis of the Maple method.

This approach *could* be appropriate. The reasoning is that although the fit is formally of bad quality -- indicated by a statistically unacceptable chi-square value -- you *assume* that the fit is good. You do this by multiplying the uncertainties by SQRT(CHI^2 / DOF), which produces a modified reduced chi-square value of 1. That may not always be appropriate, and it depends mostly upon scientific judgement.

Craig

Subject: Re: On errors calculated by curve-fitting routines
Posted by Gernot Hassenpflug on Tue, 11 Mar 2008 09:35:57 GMT
View Forum Message <> Reply to Message

Craig Markwardt <craigmnet@REMOVEcow.physics.wisc.edu> writes:

> Gernot Hassenpflug <gernot@nict.go.jp> writes:
>> I'd just like to ask, since I cannot quite tell if I have grasped the
>> ideas from Numeric Recipes correctly (and so my own IDL code for
>> comparison with the others may be incorrect): the covariance matrix
>> calculation uses the basis functions (e.g., 1, x, x^2) and the
>> variances of the dependent (y) variable, but *not* the dependent
>> variable itself nor any quantitative measures of the goodness of the
>> fitting process (presumably the variances of the dependent variable
>> are supposed to contain all such information in theory).
>
> That is the formal definition of the covariance matrix, assuming the
> measurement uncertainties are appropriate.

Thank you, it seems that as far as that goes, I have understood (formally) the issue well enough that IDL and my hand-done calculations give the same output. Hoorah!

>> I ask this because other methods, such as that used by Maple, seem to

>> scale their result by the residual sums of squares, for example. I am
>> still awaiting the book by Bevington (can only get 1st edition from
>> library services, so need to purchase 2nd edition) and the one by
>> Himmelblau from 1970 which is the basis of the Maple method.
>
> This approach *could* be appropriate. The reasoning is that although
> the fit is formally of bad quality -- indicated by a statistically
> unacceptable chi-square value -- you *assume* that the fit is good.
> You do this by multiplying the uncertainties by SQRT(CHI^2 / DOF),
> which produces a modified reduced chi-square value of 1. That may not
> always be appropriate, and it depends mostly upon scientific
> judgement.

OK, I will have to read up more on that, just received Bevington's
First Edition today, and got the later edition on the weekend to
peruse. Very nice easy-to-follow explanatory chapters on what I need
to fill in the gaps in my understanding.

It is funny how after I left IDL for a while to enjoy the ease of the
GUI and object graphics manipulation in Maple, and the symbolic maths
of Mathematica, I come back to IDL for the features that are either
not in the other programs or only available separately at a further
fee. I guess that is the world of commercial applications, and there
is no excuse for not understanding how the undelrying maths and
statistics works.
--
BOFH excuse #371:

Incorrectly configured static routes on the corerouters.