
Subject: Re: fitting many linear eqs simultaneously with outliers

Posted by [Matt Francis](#) on Fri, 14 Oct 2011 05:37:12 GMT

[View Forum Message](#) <> [Reply to Message](#)

I have had reasonable success with noisy outlier full data using AMOEBA (from IDL Astronomy User's Library) to minimise the following fitness function derived in "Data Analysis: A Bayesian Tutorial" (D.S. Sivia) section 8.3.1 (second edition):

$$F = \text{SUM}_i \{ 1/(\sigma * \sqrt{2 \pi}) * [(1 - \exp(-R_i^2/2))/R_i^2] \}$$

where $R_i = (\text{model} - \text{measured})/\sigma$

and σ is your best a priori guess of the measurement error.

The fitness function above works remarkably well at toning down the influence of outliers. The down side is that this doesn't use the linearity of the equations at all, but AMOEBA works pretty efficiently and your parameter space probably won't be too multi-modal so a downhill solver should work fine.

Subject: Re: fitting many linear eqs simultaneously with outliers

Posted by [Bringfried Stecklum](#) on Fri, 14 Oct 2011 09:32:02 GMT

[View Forum Message](#) <> [Reply to Message](#)

Jeremy Bailin wrote:

- > So I have a large number of very simple linear equations, which all look
- > like:
- >
- > $a_i + b_i x_{ij} = a_k + b_k x_{kj}$
- >
- > for $i, k = 1..N$ (all unique combinations of i and k), $j = 1..M$. The data
- > values x_{ij} and x_{kj} are measurements of the brightness of object j in
- > images i and k respectively, where the images have an unknown zero point
- > and scalings that I am trying to determine.
- >
- > (aside to astronomers: this may sound suspiciously like re-implementing
- > mscimatch)
- >
- > Not all objects appear in all images, so there are different numbers of
- > equations relating each pair of images. In principle, any number of
- > linear solvers should work... BUT: it needs to be extremely robust to
- > outliers. I know for a fact that there are many many many outliers, and
- > there are some pairs of images where it looks like pure scatter. So I
- > need some sort of solver that will do sigma clipping. Essentially, I
- > want a sigma-clipping linear least squares solver that can solve more
- > than one line at once.

>
> Does anyone know of such a beast already existing? Or something that's
> vaguely similar enough that I can use it as a basis?
>
> -Jeremy.

Perhaps fitting a line in 3D may serve as a starting point. From looking at the equations a generalization to higher dimensions seems feasible.

<http://www.scribd.com/doc/31477970/Regressions-et-trajectoir-es-3D>

cheers, Bringfried

Subject: Re: fitting many linear eqs simultaneously with outliers
Posted by [Craig Markwardt](#) on Sat, 15 Oct 2011 01:46:56 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Oct 13, 1:01 am, Jeremy Bailin <astroco...@gmail.com> wrote:

> So I have a large number of very simple linear equations, which all look
> like:
>
> $a_i + b_i x_{ij} = a_k + b_k x_{kj}$
>
> for $i, k = 1..N$ (all unique combinations of i and k), $j = 1..M$. The data
> values x_{ij} and x_{kj} are measurements of the brightness of object j in
> images i and k respectively, where the images have an unknown zero point
> and scalings that I am trying to determine.
>
> (aside to astronomers: this may sound suspiciously like re-implementing
> mscimatch)
>
> Not all objects appear in all images, so there are different numbers of
> equations relating each pair of images. In principle, any number of
> linear solvers should work... BUT: it needs to be extremely robust to
> outliers. I know for a fact that there are many many many outliers, and
> there are some pairs of images where it looks like pure scatter. So I
> need some sort of solver that will do sigma clipping. Essentially, I
> want a sigma-clipping linear least squares solver that can solve more
> than one line at once.
>
> Does anyone know of such a beast already existing? Or something that's
> vaguely similar enough that I can use it as a basis?

I'm pretty sure you can use MPFIT to solve this set of equations (MPFIT, not MPFITFUN). You need to rephrase the equations trivially as $LEFT - RIGHT = 0$, and then MPFIT will happily solve all of these

equations in a least squares sense. Your user function computes LEFT-RIGHT for each equation. Presumably you would want to scale by the uncertainties in each equation as well so that (LEFT-RIGHT)/SCALE has an expected variance of 1.

As for outliers, I have used a TANH() filter in the past. In other words, solve this slightly modified problem,

$$\text{NSIGMA} * \text{TANH}((\text{LEFT} - \text{RIGHT}) / (\text{NSIGMA} * \text{SCALE})) = 0$$

TANH() has the property of being linear near the origin and "truncating" smoothly values much greater than 1. The NSIGMA part is a N-sigma filter, i.e. if NSIGMA=3 then 1- and 2-sigma variations should pass through relatively unscathed, but 3 to 100 sigma outliers would be stomped down to about 3 sigma.

If it were my preference, I would re-express the problem so that your model function predicts the measured intensity of each source in each plate. For example, using MPFITFUN and this model function,

$$X_IJ_MODEL = C_I + D_I * F_J$$

where C_I and D_I are slightly different formulations of your zero-point and scale for the Ith image, and F_J is the "true" relative flux of the Jth source. The F_J are nuisance parameters (and you need to set F_0 = 1 to prevent degeneracy). I see this as better because you probably have measurement uncertainties of X_IJ so the problem is likely to be more linear and stable when expressed this way.

Good luck,
Craig

Subject: Re: fitting many linear eqs simultaneously with outliers

Posted by [Jeremy Bailin](#) on Tue, 18 Oct 2011 06:56:16 GMT

[View Forum Message](#) <> [Reply to Message](#)

On 10/13/11 1:01 AM, Jeremy Bailin wrote:

> So I have a large number of very simple linear equations, which all look
> like:
>
> $a_i + b_i x_{ij} = a_k + b_k x_{kj}$
>
> for $i, k = 1..N$ (all unique combinations of i and k), $j = 1..M$. The data
> values x_{ij} and x_{kj} are measurements of the brightness of object j in
> images i and k respectively, where the images have an unknown zero point
> and scalings that I am trying to determine.
>

> (aside to astronomers: this may sound suspiciously like re-implementing
> mscimatch)
>
> Not all objects appear in all images, so there are different numbers of
> equations relating each pair of images. In principle, any number of
> linear solvers should work... BUT: it needs to be extremely robust to
> outliers. I know for a fact that there are many many many outliers, and
> there are some pairs of images where it looks like pure scatter. So I
> need some sort of solver that will do sigma clipping. Essentially, I
> want a sigma-clipping linear least squares solver that can solve more
> than one line at once.
>
> Does anyone know of such a beast already existing? Or something that's
> vaguely similar enough that I can use it as a basis?
>
> -Jeremy.

In case anyone cares, my solution was to get the user's brain (in this case, my undergrad) to do the hard work. I plotted up all the points for each possible pair of image combinations, got them to draw the best fit line on there, finessed it a little with a linear fit to only those points very close to the drawn line, and then solved for the best global solution for all images using singular value decomposition. Certainly not the most automated solution, but the results are beautiful. And my first ever IDL widget program...

-Jeremy.
