Subject: Re: Autocorrelation with (LOTS) of missing data.
Posted by jameskuyper on Fri, 14 Mar 2008 15:59:25 GMT
View Forum Message <> Reply to Message

Brian Larsen wrote:

>> I've got a time series 807793 bins long, with missing data in all but
>> 48945 of those bins. Only 7392 of those bins have a non-zero event
>> count. Those bins have a total count of about 1 million events, which
>> tells you that events are highly clustered, at least at the time scale
>> of the bin size (5 minutes).
>>
>> I want to use autocorrelation analysis to investigate the clustering
>> of these events on longer time scales. The large amount of missing
>> data makes such analysis difficult, but the non-missing data is
>> clustered on time spans of 9 bins or so. Therefore, it seems to me
>> that with the right algorithm, it should be possible to estimate the
>> autocorrellation at lags of less than 9 bins. Does anyone know what
>> the right algorithm would be?
>
> Seems to me that this is an issue, I would use normal techniques on
> subsets of the data. There might be other ways but clusters of
> missing data are kinda like small data sets.

The individual clusters are too small to calculculate meaningful autocorrelation values; I would need to know an appropriate way to combine autocorrelation functions calculated from different sets of varying lengths.

I've found an article <http://sankhya.isical.ac.in/search/61a2/61a27036.pdf> which describes three estimators that can be used for this purpose. I was hoping I could use code that had already been written, but it should be pretty straightforward to write a program to calculate those estimators.