Autocorrelation irregular time series R

Asked 4 years, 9 months ago Active 3 years, 11 months ago Viewed 994 times

3

1

I have to calculate the autocorrelation of an irregular (i.e. non equi-spaced) time series
`value(time)`

```
time[hr] value

1    3
2    5
3    7
5    9
7    16
12    17
13    19
16    25
19    27
21    30
```

Using the `acf` function it is possible to specify only the lag (i.e. the number of points)

```
acf(df$value,lag.max = 10,type = "correlation",plot = FALSE, na.action =
na.pass)
```

which however in my case does not correspond to a definite time interval.

What I would like is instead to calculate the autocorrelation function specifying the time interval (e.g `3 hr`).

Any help?

Thanks

r

asked Nov 11 '14 at 16:43

user3036416
535 1 6 20

1 This is actually a moderately hard problem. This book gives a couple of possible solutions, but you'd have to code them yourself. – Ben Bolker Sep 18 '15 at 19:19

1

To clarify, the `lag.max` argument determines the maximum time-shift (lag) between the 2 series, not the number of points to use (i.e., you are comparing x_t to x_t-h; `lag.max` is the maximum value for h). In your example, you have 10 data points total, so if you assume you need a minimum of 3 data points to estimate their correlation, then `lag.max=7` at most.

It sounds like perhaps you have samples from a continuous-time function, but the ACF is defined over discrete-time intervals. Therefore, to use `acf()` you would need to first fill in the non-observed times with `NA`, at which point the time lag would be expressed in hours (i.e., what you want).

Assuming your data above were in the data.frame `df`, then

```
df2 <- data.frame(time=seq(min(df$time), max(df$time)), value=NA)
df2[df$time,"value"] <- df$value
acf(df2$value, lag.max=10, type="cor", plot=FALSE, na.action=na.pass)

# Autocorrelations of series ‘df2$value’, by lag
#
#     0      1      2      3      4      5      6      7      8      9     10
# 1.000  0.716  0.665  0.415  0.166  0.194  0.046  0.007  0.029  0.008 -0.041
```

answered Sep 18 '15 at 18:55

Mark S
493 3 7

2 this is slightly dangerous. Naively filling in `NA` values and computing the acf by dropping NA values from the sum can confound the pattern of missingness with the autocorrelation in the underlying process. – Ben Bolker Sep 18 '15 at 19:44

@BenBolker: Agreed -- my hope was that there would be relatively few NA's. Another soln would be to aggregate the data at the lowest common interval (eg, 3 hrs) and then estimate the ACF at that scale, but that would also hide the behavior of the true underlying process. – Mark S Sep 18 '15 at 22:00