Will Pearse and Davies method work for my data?

Susannah Tysor

September 19, 2017

What am I even doing

So I thought I would use the method described in Roberts & Solow (which is from a 1980 Biometrika paper by Cooke) as implemented by Pearse *et al.* (not yet published) to transform my survey data into event data - that is, estimate the start and end dates from yes/no phenology data gathered every few days over a period of ~2 weeks

The survey period sometimes chops off the beginning and/or end of the phenological period for an individual and survey intervals (time between data collection within a survey period) are sometime erratic.

I made some very simple fake data (yeses over a period of 10 days, interval or end censored) and used the Cooke method to estimate a "start date."

This method is inappropriate for my data because the estimate is a function of the interval between observations and it cannot accept negative observations, only positive.

The interval of data collection affects the estimate a lot, but end censoring doesn't. I need it to respond to end censoring but not interval censoring.

Interval censoring

How sensitive is the method described by Pearse and Davies to interval censoring?

I created 4 sequences representing the same phenology period for the same individual. The only difference was that it was "collected at 1, 2, 3, or 4 day intervals. All contain the same initial day. The first sequence (interval 1) represents the full phenological period.

The start day is always estimated to be earlier than the first measurement by more than the length of the sampling interval. **The start day estimate is highly sensitive to interval censoring.** In fact, the estimate *depends on* the interval between observations.

Regular interval sequences with start estimate in red
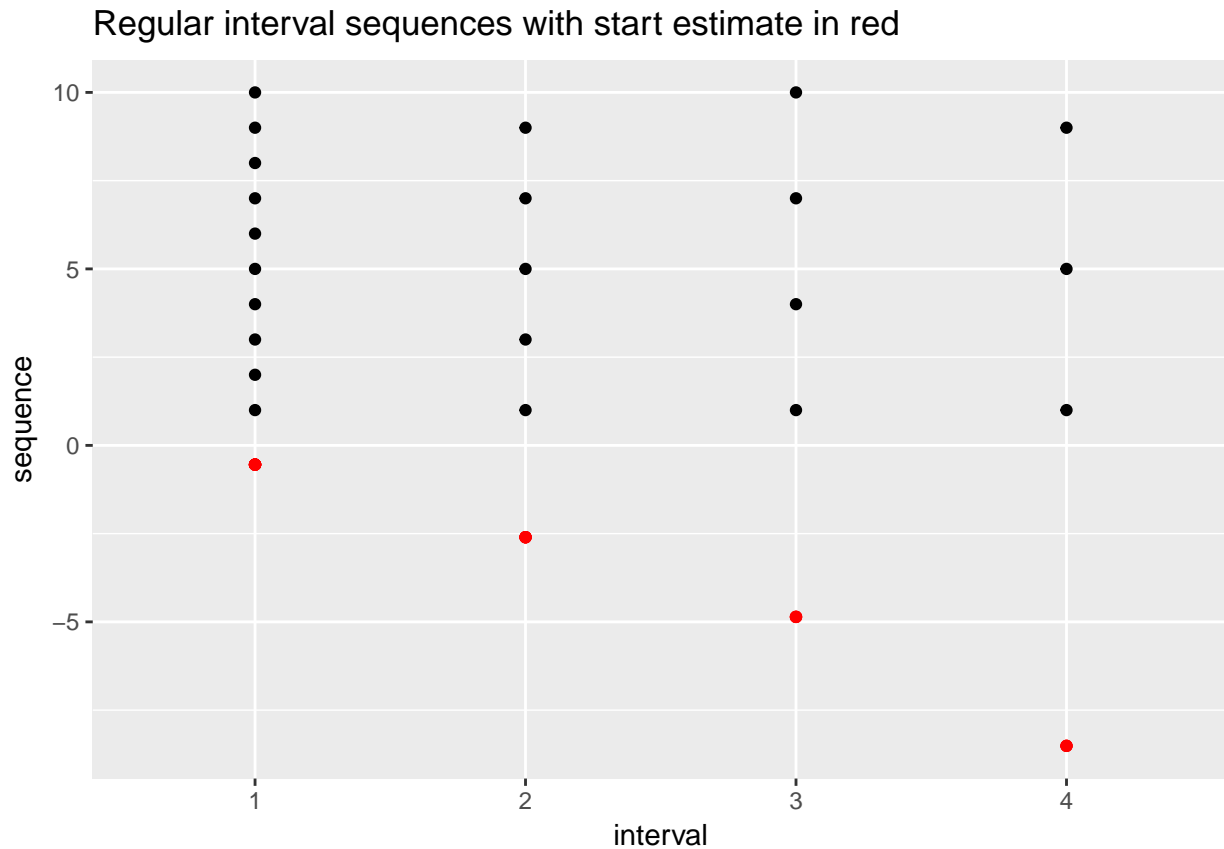
End censoring

How sensitive is this method to end censoring?

I created 8 sequences of observations for 8 different individuals. All 8 individuals have a phenological period that is the same length, but starts on a different day. We do not observe the first 3 to 7 days of 7 of the individuals.

The estimation method produces nearly identical start day estimates for all sequences.

The method is insensitive to end censoring.

End censored sequences with start estimate in red
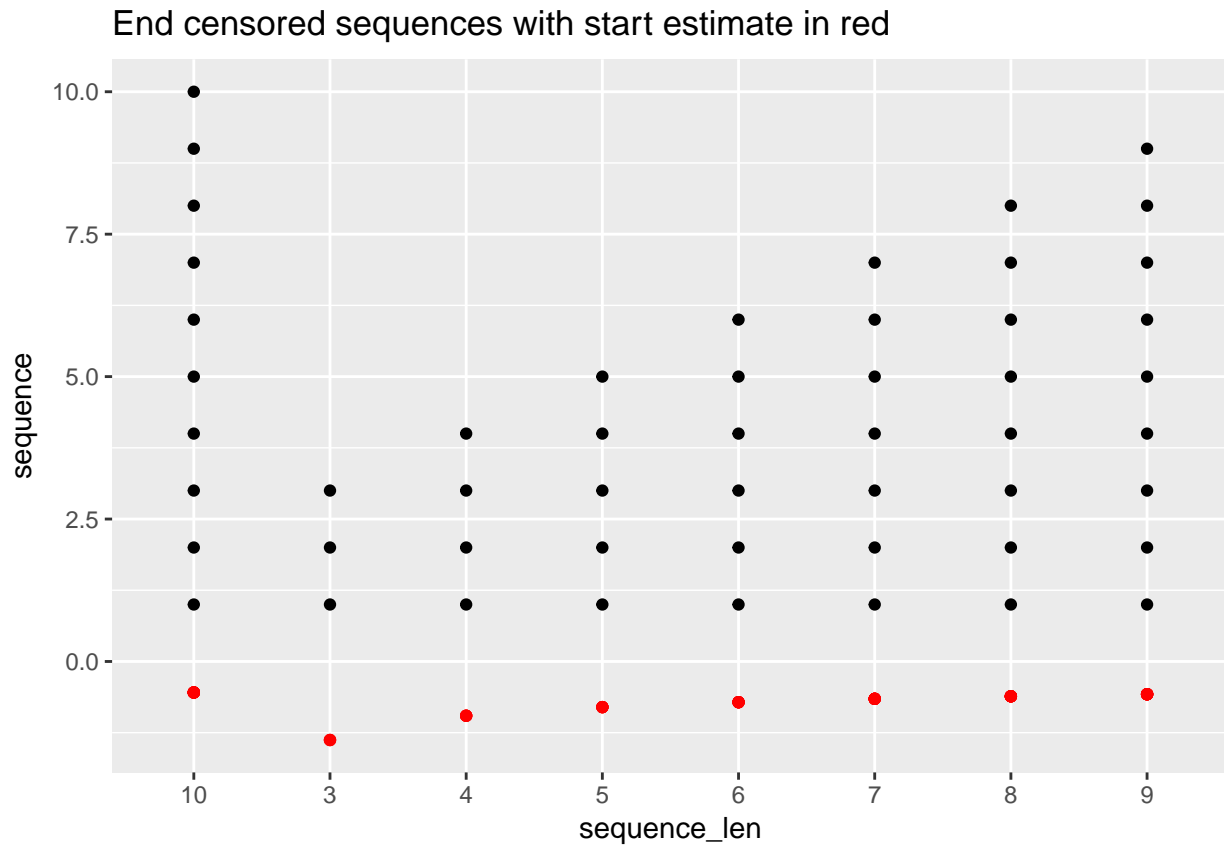
In retrospect, this result is trivially obvious. ><

The end

In my actual data, I'd expect all the individuals in the interval censored example to have similar start dates and those in the end censored graph to have really different start dates. This method doesn't work for my data.