Numerical Implementation Notes for Code for Paper "Epidemiological Dynamics of Ebola Outbreaks"

Thomas House

The numerical methods provided in this repository are typically transparent from the source code, however for the case of large outbreak sizes it is necessary to calculate combinatorial factors in the likelihood appropriately. In particular, the formula for the total size of the outbreak is

$$\begin{aligned}\Pr[K=k|p,q] &= \sum_{a=1}^{k}\Pr[A=a|p]\Pr[Z=(k-a)|A=a,q] \\ &= \sum_{a=1}^{k}(1-p)^{a-1}p\frac{a}{k}\binom{2k-a-1}{k-a}q^k(1-q)^{k-1} \\ &=: \frac{pq}{k}\sum_a a y_{k,a}\,.\end{aligned} \tag{1}$$

To work out the first term $y_{k,1}$ in the limit of large k, we use Stirling's formula,

$$\begin{aligned}y_{k,1} &= \binom{2(k-1)}{k-1}q^{k-1}(1-q)^{k-1} \\ &= \frac{(4q(1-q))^{k-1}}{\sqrt{\pi(k-1)}}\left(1+O\left(k^{-1}\right)\right)\,.\end{aligned} \tag{2}$$

Then subsequent terms can be calculated through simple iteration

$$y_{k,a+1} = \frac{1-p}{1-q}\,\frac{k-a}{2k-a-1}\,y_{k,a}\,. \tag{3}$$

This is the algorithm used in the function `approxC.m`.