MODULE

ANALYSIS OF CELLULOSE SYNTHASE COMPLEX (CSC)

May 19, 2018

Student ID: 71244370
University of British Columbia
Department of Statistics

SUMMARY

In this report, we seek to analyze the effects of the mor1-1 gene on CSC velocities. Observations of these velocities come from different seedlings. This provides us with an interesting challenge. We discuss the applications of models that account for differences between seedlings and ones that do not. Finally, we conclude with our choice between the two.

INTRODUCTION

Cellulose Synthase Complex (CSC) is responsible for the synthesis of cellulose. CSC converts glucose into cellulose. It is then deposited along the cell's wall, as the CSC moves along 'micro tubule' tracks which lie beneath the cell's plasma. In this report we analyze the dynamics of CSCs in a genome sample from the Arabidopsis thaliana species (a small flowering plant).

Our primary interest lies in the effects of the mutant mor1-1 gene on CSC velocities. This is because it is hypothesized that this gene disrupts the micro tubule networks the CSC moves along. This could affect the speed CSCs travel within the plant cell at. The formation of cellulose along a plant cell's walls would also be affected as a result. Thus it could alter the direction in which the seedling grows as it breaks through the soil, in search of light and water. We test the aforementioned hypothesis by answering two questions. They are:

- Are CSC velocities different in particles with the mor1-1 gene?
- Are CSC velocities different inside a micro tubule versus outside a micro tubule?

PROPOSED STATISTICAL METHODS

The next section proceeds as follows:

- We describe the data creation process, its collection and its structure.
- We then propose models we feel appropriately answer the questions raised.

Data Description

Cell samples from a seedling are treated with fluorescent proteins. This gives micro tubules a red color and CSCs a green color when seen under a microscope. The seedlings are then studied under a TIRF- microscope to visualize the movement of CSCs. Each particle's movement is represented by a line, traced by hand. The gradient of the line gives the particle's velocity. It's speed can then be determined through a formula relating a particle's speed to it's velocity.

The data collected is hierarchical in nature. Observations of particles expressing the Wild Type gene come from 3 different seedlings, while those expressing the mutant Mor 1-1 gene come from 4 different seedlings. The number of observations are not the same across seedlings. This experimental design is therefore unbalanced. Additionally, a record is made of whether CSCs moves inside a micro tubule or not.

The table below summarizes the variables in the data set. They are:

Variable	**Type**	**Description**
Velocity	Continuous	Response variable that describes the velocity of CSC.
MT	Binary	Indicates if a particle traveled along a micro tubule domain.
Type	Binary	Gene expressed (Wild Type or Mor1-1).
Seedling	Categorical	Indicates the seedling the CSC was observed from.

Proposed Statistical Analysis

In this sub section we propose two statistical models. The first approach treats seedlings as a random draw from a population of seedlings. The second **'pools'** their observations together.

Generalized Linear Mixed Models (GLMM)

GLMMs are able to capture variations in CSC velocities within a seedling and between seedlings. Under this specification, the variables Type, MT and their interaction MT $*$ Type are treated as fixed effects (a variable whose levels we care about). They are factors with 2, 2 and 4 levels respectively. These variables capture the effects of interest. For instance, the coefficient of MT represents the change in CSC velocities when a particle moves outside the tubule compared to the baseline (set to moving inside)

Seedlings are treated as a random effect here (a variable whose specific levels we do not care about). They form random intercepts in the model. That is each of the 7 seedlings are fit with a different linear regression line.

Another flexible feature of a GLMM is that it does not require CSC velocities be normally distributed to compute parameter estimates. However, with only 7 seedlings, tests under this model may lack degrees of freedom. As a result only large differences in velocities will lead to rejection of the null whilst small, subtle differences will not. Thus these tests will be severely under powered. A test's power is defined as its ability to **'correctly'** reject the null hypothesis.

Linear Regression

Under this setup, we assume all seedlings are identical to one another. Hence we do not need to account for any variation between them. This allows us to pool together our observations. In this model Velocity is the response variable. MT, Type and MT$*$Type (an interaction term) are the explanatory variables, with 2, 2 and 4 levels respectively. A nice feature of this model is that its parameter estimates are easily interpretable.

This model yields a more powerful test as compared with the GLMM. This is because pooling observations together gives 982 degrees of freedom. However, pooling in the presence of significant variation between the seedlings will inflate type 1 errors. That is tests under this model will often find significant differences when there are none.

CONCLUSION

In this report we have discussed two models that appropriately capture the complexities in the data. We feel the most appropriate of them is the GLMM. To overcome its lack of power, we suggest sampling from a larger (ideally 20 or more if feasible) number of seedlings and repeating the experiment. This would allow subtle differences in velocities to be detected and provide more confidence in the results obtained.

REFERENCES

A Practical Guide to Mixed Models in R [online] Available at:<https://ase.tufts.edu/gsc/gradresources/guidetomixedmodelsinr/mixed%20model%20guide.html>

Power Analysis for Generalized Mixed Models in R [online] Available at: <http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12504/full>

Mixed Models [online] Available at: <http://www.stat.cmu.edu/~hseltman/309/Book/chapter15.pdf>

CSC Data Description [online] Available at: <https://connect.ubc.ca/bbcswebdav/pid-4656008-dt-c
1/courses/SIS.UBC.STAT.550.201.2017W2.95325/CSC_activity1.pdf>

CSC Project Proposal [online] Available at: <https://connect.ubc.ca/bbcswebdav/pid-4651239-dt-c
1/courses/SIS.UBC.STAT.550.201.2017W2.95325/CSC_case_submission.pdf>

Cellulose synthase complexes: composition and regulation [online] Available at: <https://www.frontiersin.org/articles/10.3389/fpls.2012.00075/full>