Bivariate Models for the Analysis of Internal Nitrogen Use Efficiency: Mixture Models as an Exploratory Tool

Isabel Munoz Santa (Masters of Applied Science in Biometrics)

A thesis submitted for
the degree of Masters of Applied Science in Biometrics
in the University of Adelaide

School of Agriculture Food and Wine

July 2014

Acknowledgments

I would like to express my deepest gratitude to my advisor Dr. Olena Kravchuck for her expertise, guidance, support and encouragement. This thesis would not have been possible without her. Thank you for your efforts to make me a better professional and give me the opportunity to study here!

I would also like to thank my second supervisor Dr. Petra Marschner for her guidance in the biological aspects of my thesis and Dr. Stephan Haefele who kindly provided the data for the case study of this thesis and provided support in the interpretation of the results.

I would like to acknowledge the Faculty of Science for providing the Turner Family Scholarship which supported this research and provided travel assistance to attend the Australian Statistical Conference, July 2014 Adelaide, Young Statistician Conference, February 2013 Melbourne, International Biometrics Society Conference, December 2014 Mandurah and Australian Statistical Conference in conjunction with the Institute of Mathematical Statistics Annual Meeting, July 2014 Sydney.

I would like to thank all the people at the Biometry Hub: Bev, David, Jules, Paul, Stephen and Wayne for their big smiles and for creating a positive working environment as well as the wonderful group of statisticians at Waite, meeting regularly for the professional development discussions.

Thank you to all the friends I have met in Adelaide, where I have had one of the best professional, cultural and personal experiences of my life. Thanks to Negar, Mohsen, Casey, Amanda, Rodrigo, Mariana, Fien, Daniela, Diego, Kanch, Antonio, Maria, Ruben, Alfonso, Lidia, Pablo, Ana, Antonija, Roey, Chris, Konrad, Diana, Luis and and Lorinda.

Finally, with deep love and admiration, I thank my family for their love and support

from thousands of kilometres away and Martin for his support, love and contagious positive vision of life.

Contents

Abstract

Ratios are commonly used among plant and soil scientists, in particular to express the plant nutrient utilisation efficiency of macro- and micro-nutrients. The internal nutrient efficiency can be understood in terms of maximising yield per a unit of nutrient in the plant. At present, IE_N data are usually collected from designed field trials where different treatments are applied (e.g. fertiliser treatments) and analysed by univariate linear mixed models. However, univariate linear models on the ratio do not maintain information on the original traits, including their correlation, which presents a challenge when interpreting the effect of agronomic practices or environmental conditions on the process of nutrient conversion into grain. Moreover, the distributional properties of ratios do not comply with the assumptions of these linear models favoured in the area of soil and plant science research. A more suitable approach is to collect the traits of interest and to use bivariate analyses. These analyses preserve the information on the original traits and avoid issues associated with the ratio distributional properties.

If the data comes from field studies, different experimental and environmental conditions may lead to the presence of patterns (groups) in the data in addition or concurrently with designed treatments. Researchers in plant and soil sciences may be interested in identifying those conditions, for example to understand the nature of genotype-by-environment interactions. The inspection of the groups may reveal the factors defining them, thus gaining insight into the experimental or environmental drivers of the biological traits. Among bivariate analyses, bivariate mixture models of Gaussian distributions are an appropriate methodology for identifying clusters in the nutrient efficiency data, assuming that the traits are jointly normal. Studying this methodology for the analysis of the internal nitrogen use efficiency traits is the focus

of the present thesis.

The application of bivariate mixture models is suggested here as a complementary analysis to bivariate mixed models in designed field trials and for exploratory purposes only. The exploratory and supplementary character of the mixture analysis is due to the potential violation of the independence assumption when the data are collected from designed field trials.

In this project, bivariate mixed and mixture models are applied to a real-life designed field trial on non-irrigated rice in Thailand for the analysis of grain yield (GY) and plant nitrogen uptake (NU) data. The univariate counterparts of these analyses are also applied on the ratio of these two traits (the internal nitrogen use efficiency). The advantages of the bivariate analyses are discussed in comparison to the univariate analyses on the ratio. In this case study, the bivariate mixture approach revealed that soil water availability post-flowering and N supply in soil are the potential factors defining the mixture groups.

The present work can be readily extended to the analysis of other similar traits in agriculture when the objective is to explore potential environmental conditions affecting the traits under study. In order to fully exploit the proposed methodology, field survey is suggested as a more appropriate sampling procedure for the application of mixture models than collecting data from designed field trials.

Declaration of originality

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Common abbreviations in this thesis

BVN	bivariate normal distribution
χ^2	chi-square distribution
CV	coefficient of variation
ρ	correlation coefficient
σ_{xy}	covariance between random variables X and Y
$Cov()$	covariance operator
cdf	cumulative distribution function
CV_x	CV of a random variable X
p	dimension of a random vector
D	Dirichlet distribution
$\sim$	distributed as
EM	Expectation and Maximisation
μ_x	expected value of a random variable X
$\boldsymbol{\mu}$	expected value of a random vector
$\boldsymbol{\mu}_i$	expected value of the i-th component of a mixture of multivariate normal distributions
$E()$	expected value operator
exp{}	exponential function
F	F-distribution
Γ	gamma function
GY	grain yield
$\leftrightarrow \Leftrightarrow$	if only if
G_i	i-th group of a mixture of distributions

i.i.d.	independent and identically distributed
IE_N	internal nitrogen use efficiency
$L()$	log likelihood function
MLE	maximum likelihood estimate
$\boldsymbol{\pi}$	mixing proportions
$\boldsymbol{\psi}$	mixture parameters
$Mult$	multinomial distribution
MVN	multivariate normal distribution
NU	nitrogen uptake
g	number of groups in a mixture
τ_{ij}	posterior probabilities of the j-th observation in G_i
pdf	probability density function
$\propto$	proportional
tr	trace of a matrix
n	sample size
Φ	standard univariate normal cdf
φ	standard univariate normal pdf
S	straw yield
T	student's t-distribution
N	univariate normal
σ	variance
σ_x	variance of a random variable X
$\boldsymbol{\Sigma}$	variance-covariance matrix
$\boldsymbol{\Sigma}_i$	variance-covariance matrix of the i-th component of a multivariate normal distributions
$Var()$	variance operator
$\boldsymbol{\theta}_i$	vector of parameters of the i-th component
W	Wishart distribution

List of Tables

List of Figures