# "Research Proposal Title"
# by Student, T.
# for a PhD in Statistics

Student ID: 87654321
Email: tstu007@aucklanduni.ac.nz

Supervisor: Professor R. A. Fisher
Co-supervisor: Dr C. R. Rao
Advisory Committee: Drs B. Efron, D. R. Cox, K. Pearson.[1]

July 22, 2010

This document represents the student's research proposal after one year of provisional PhD registration. Confirmed PhD registration is now sought.

# 1 Introduction

This document is a template that can be edited by the student. Fill in the details as you go along and delete these instructions. Do have your research proposal scrutinized by your supervisor before submitting it to the department.

Give a few paragraphs here on the general background to your research proposal; put the full details in Section 2. You might want to cite a few previous important results/papers. Describe any theory briefly here but put most of the details in Section 2. Describe any data in general with a few lines here but put the full details in Section 4.

Motivate this section well ... why is the topic important? What use will the thesis have? What is unknown that will be known upon completion?

This section should be understandable by an intelligent person who is not familiar with the technical details involved. The reader should be convinced that your work is worthwhile.

Some of our PhD students have published in the primary literature directly using material from their thesis, e.g., Worsley (1979), Yee and Wild (1996), Wild and Yee (1996), Murrell (1999), Jiang et al. (2006), Roever et al. (2007), Millar and Stewart (2007), or submitted such as Lee and Hirose (2008). Ideally you should aim at least one of your publications towards a mainstream statistics journal, preferably your first one. All publications should be in a peer-reviewed international journal. It should be in the primary literature. Statistical theory and methodology is to be regarded as the highest form of quality output. Accompanying high quality software that can easily be used by others (such as a R package) is also good. It is best to submit to journals before the completion of a thesis because an acceptance validates the quality of that research.

A word about writing up your thesis. In general, the Department of Statistics recommends LaTeX rather than Microsoft Word. LaTeX is easily run on the Department's Linux machines. A website that converts pdf into Word is www.pdftoword.com. The University of Auckland offers a half-day workshop on LaTeX to postgraduate students. There is Sweave (Leisch, 2003) which allows R (Ihaka and Gentleman, 1996) to be embedded into a LaTeX document, and so the output does not have to be manually cut-and-pasted in. Sweave is useful for presentations too, in conjunction with Beamer.

There is an editor called Emacs which is useful for all types of editing, including LaTeX, C, and R. In Windows there is Tinn-R for R files and MiKTeX for LaTeX.

Note that BibTeX items are available from MathSciNet which is part of the "Library Databases" in the UoA electronic library system. There is a close relationship between EndNote and BibTeX.

---

[1]Specify the affiliation if outside the UoA Statistics Department.

# 2 Background

Background knowledge is crucial, therefore describe it fully here. Specify the background to the problems you are trying to solve. Define anything technical that most statisticians will not be familiar with.

What previous work has been done in the area? Cite references relating to such work, e.g., in the form of Reid (2003) and Efron and Tibshirani (1993). Has anybody described any open problems in the literature?

Make sure that problems arising from the background knowledge are clearly described. There should be a logical flow of thought in this section starting from the background and leading to the problem on hand.

# 3 So What's New?

List *specific* and *new* results that your thesis plans to obtain, especially with regard to new statistical theory and methodology. It is essential that you have provided comprehensive background (in Section 2) so that it is clear what new contributions will be made (here), and how they will extend what is currently known. Tie these with Appendix A(i), and a list such as below is a good idea.

For example, all of the points below are new (to my knowledge) and have not been done yet.

1. Develop a new method for quantile regression based on the asymmetric Laplace distribution and the concept of link functions and parallelism assumptions (via constraint matrices).

2. Derive the asymptotic properties of the above method, both as $n \to \infty$ and as $p \to \infty$.

3. Conduct a small simulation to check the mean squared error and other properties. To be done in R with some sections written in C calling LINPACK (Dongarra et al., 1979) for additional speed.

4. Compare my results to those of Wei and Carroll (2009), especially relating to boundary effects and small-sample properties.

5. Tackle an important dataset or problem that no one has satisfactorily resolved before. It is essential that new methodologies are investigated and that novel insights are obtained. Note that a conventional analysis, even of a very large dataset, is *not* PhD-level research.

# 4 Data and Special Needs

Almost all Statistics PhD theses should analyse at least one data set, even if it is a toy data set.

State their source fully. If new data is important to the thesis describe what's new. Why is this data set important? What is novel about it? Specify any ethics or confidentiality issues.

A plot of some of the data may be useful; see Figure 1.

Do you have any special needs such as IT needs? These include hardware (supercomputing facilities), and expertise/technical knowledge. Often extra large data sets have special needs associated with its processing. How do you invisage those needs to be met?

Funding: anything to report here? If your thesis work involves experiments, travel or special resources in order to collect data, or essential collaboration with experts in your field, then you should describe these requirements here along with how you, or your supervisor, will fund these requirements. Departments work on tightly planned budgets and generally do not have the facilities to pay for unplanned PhD related expenses.
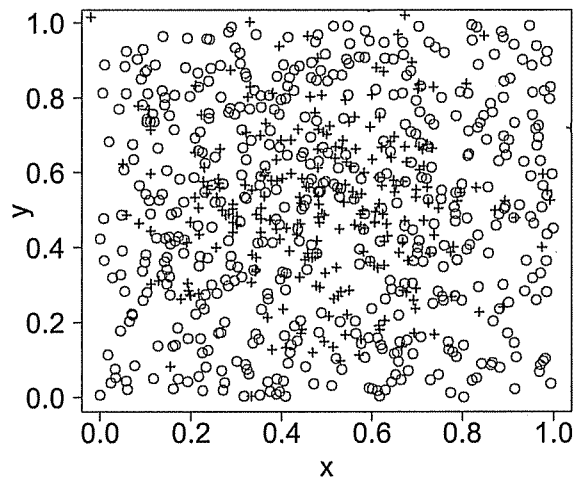
Figure 1: Data for a two-class classification problem. The red "+" is for class 1, and $n = 250$. The blue "o" is for class 2, and $n = 500$.

If ethics approval(s)/permissions is needed then put any details here.

# 5    Goals

List *specific* goals you wish to achieve in your PhD.

1. I wish to derive the first three moments of the Slash distribution. This will involve working out the characteristic function first.

    a. Then work out the observed information matrix based on Bickel et al. (2009).

    b. Then work out the expected information matrix (Scott et al., 2007).

    c. Apply the EM algorithm (Dempster et al., 1977) to estimate $\lambda$.

    d. Write a R package to implement my method. It will be written in S4 and use object oriented methods (Chambers, 1998).

    e. Apply the method to the radiation data set (Holmes, 2009).

    f. Extend biplots (Gower, 1966) for my data.

    g. Release the package on CRAN (cf. Murrell and Ihaka (2000)).

    h. Publish my results in at least two papers, earmarked for JASA and JRSS-B. Also an applications paper in Biometrics. See Section 6.

2. Solve the Fisher-Behrens problem (Efron, 2009, 2008; Kojadinovic and Holmes, 2009).

# 6    Timeline and Activities

I have fulfulled all my first year requirements. These are:

1. All EoI provisional goals:

Table 1: Timeline for my thesis. Here are *some* activities—fill in more.

| Date | Activity |
|---|---|
| 2010-04-01 | Provisional PhD registration (PhD in Statistics). |
| 2010-05-01 | Updated my personal webpage at www.stat.auckland.ac.nz/~myStudentName. |
| 2010-05-20 | Attended one of the Doctoral Skills Programme's Induction Days. |
| 2010-09-01 | Gave my first talk at NZSA conference. Won first prize. |
| 2010-11-01 | Gave a talk to PhD Talks Day. |
| 2011-01-15 | Submit my first paper to *Annals of Statistics* (co-authored with supervisor). |
| 2011-05-01 | Presented my research progress to a departmental seminar. |
| 2012-01-20 | Achieve Goal 1. |
| 2012-08-20 | Achieve Goal 2. |
| 2012-10-23 | Submit my second paper to *Annals of Applied Statistics* (co-authored with supervisor). |
| 2013-01-15 | Submit my third paper to *Biometrics* (sole authorship). |
| 2013-02-01 | Submit my PhD thesis. |

(a) *Full thesis proposal normally completed within 6 months*: Here it is!

(b) *Completion of one substantial piece of written work within 12 months*: Wrote 90 percent of Chapter 1 of my thesis, submitted one journal article, wrote 50 percent of Chapter 2 of my thesis.

(c) *Presentation of research progress to a departmental seminar*: 2011-05-01.

(d) *Approval of the full thesis proposal by the appropriate departmental/faculty postgraduate committee*: Being done.

(e) *Ethics approval(s)/permissions obtained for the research (if required)*: Not necessary.

(f) *Attendance at one of the Doctoral Skills Programme's Induction Days*: 2010-05-20.

2. I have updated my webpage (www.stat.auckland.ac.nz/~myStudentName) giving details of my thesis, links to other research resources in my topic, and some personal stuff to make it interesting.

3. I have diligently attend as many Statistics Department seminars as I could. They are given in Table 2. This is much more than the minimum quota set by the department. Consequently there should be no problem due to this when getting my annual report signed off[2]. The 2010-04-08 seminar was particularly useful because it gave me an idea on how to solve one of my problems.

4. I did STATS 730 in Semester 1 of 2010 and obtained an A+.

5. I did STATS 782 in Semester 2 of 2010 and obtained an A.


# 7   Summary

Give a few sentence summary of your thesis, especially what's new.

---

[2]If insufficient seminars have been attended then sign off will occur *after* the minimum number is reached.

Table 2: Departmental seminars I have attended (top part of the table). Talks from another UoA department are in the middle tier. Conferences and workshops are in the bottom tier, e.g., 2010-05-29 event was an all-day workshop.

| Date | Speaker | Title |
|---|---|---|
| 2010-04-08 | David Brillinger | Random trajectories, some theory and applications |
| 2010-02-30 | David Cox | Frequentist statistics as a theory of inductive inference |
| 2009-10-08 | David Matthews | Estimating diagnostic test likelihood ratios |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| 2010-06-16 | James B. Conant | "Geography is not a university subject" (geo Department) |
| yyyy-mm-dd | Speaker Name | Title of Talk |
| 2010-05-29 | David Siegmund | Workshop on 'Genetic Mapping' at UoA. |

# Appendix A

Delete or replace the contents of this section with any appendices you may have.

The *University of Auckland Statute and Guidelines for the Degree of Doctor of Philosophy (PhD)* (2008) reads[3]:

"The PhD degree is awarded for a formal and systematic exposition of a coherent programme of advanced research work carried out over the period of registration for the degree which in the opinion of the examiners and the Board of Graduate Studies satisfies all of the following criteria:

(i) to be an original contribution to knowledge or understanding in its field, and

(ii) to meet internationally recognised standards for such work, and

(iii) to demonstrate a knowledge of the literature relevant to the subject and the field or fields to which the subject belongs, and the ability to exercise critical and analytical judgement of it, and

(iv) to be satisfactory in its methodology, in the quality and coherence of its written expression, and in its scholarly presentation and format."

---

[3]Clause 1, Preamble, item (d).

# References

Bickel, P. J., Ritov, Y., Tsybakov, A. B., 2009. Simultaneous analysis of lasso and Dantzig selector. Ann. Statist. 37 (4), 1705–1732.

Chambers, J. M., 1998. Programming with Data: A Guide to the S Language. Springer, New York, USA.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the *EM* algorithm (with discussion). J. Roy. Statist. Soc. Ser. B 39, 1–38.

Dongarra, J. J., Bunch, J. R., Moler, C. B., Stewart, G. W., 1979. LINPACK User's Guide. SIAM Publications, Philadelphia.

Efron, B., 2008. Simultaneous inference: when should hypothesis testing problems be combined? Ann. Appl. Stat. 2 (1), 197–223.

Efron, B., 2009. Empirical Bayes estimates for large-scale prediction problems. J. Amer. Statist. Assoc. 104 (487), 1015–1028.

Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York, USA.

Gower, J. C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53 (3–4), 325–328.

Holmes, M. H., 2009. Introduction to the foundations of applied mathematics. Vol. 56 of Texts in Applied Mathematics. Springer, New York.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5 (3), 299–314.

Jiang, Y., Scott, A. J., Wild, C. J., 2006. Secondary analyses of case-control sampled data. Statistics in Medicine 25, 1323–1339.

Kojadinovic, I., Holmes, M., 2009. Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. J. Multivariate Anal. 100 (6), 1137–1154.

Lee, A. J., Hirose, Y., 2008. Semi-parametric efficiency bounds for regression models under generalised case-control sampling: the profile likelihood approach. Submitted for publication , .

Leisch, F., October 2003. Sweave, part II: Package vignettes. R News 3 (2), 21–24.
URL http://CRAN.R-project.org/doc/Rnews/

Millar, R. B., Stewart, W. S., 2007. Assessment of locally influential observations in Bayesian models. Bayesian Anal. 2 (2), 365–383.

Murrell, P., Ihaka, R., 2000. An approach to providing mathematical annotation in plots. J. Comput. Graph. Statist. 9 (3), 582–599.

Murrell, P. R., 1999. Layouts: A mechanism for arranging plots on a page. J. Comput. Graph. Statist. 8 (1), 121–134.

Reid, N., 2003. Asymptotics and the theory of inference. Ann. Statist. 31 (6), 1695–1731.

Roever, C., Meyer, R., Guidi, G., Vicere, A., Christensen, N., 2007. Coherent Bayesian analysis of inspiral signals. Classical and Quantum Gravity 24, S607–616.

Scott, A. J., Lee, A. J., Wild, C. J., 2007. On the Breslow-Holubkov estimator. Lifetime Data Anal. 13 (4), 545–563.

Wei, Y., Carroll, R. J., 2009. Quantile regression with measurement error. J. Amer. Statist. Assoc. 104 (487), 1129–1143.

Wild, C. J., Yee, T. W., 1996. Additive extensions to generalized estimating equation methods. J. Roy. Statist. Soc. Ser. B 58 (4), 711–725.

Worsley, K. J., 1979. On the likelihood ratio test for a shift in location of normal populations. J. Amer. Statist. Assoc. 74 (366, part 1), 365–367.

Yee, T. W., Wild, C. J., 1996. Vector generalized additive models. J. Roy. Statist. Soc. Ser. B 58 (3), 481–493.