Natural Sciences and Engineering Research Council of Canada | Conseil de recherches en sciences naturelles et en génie du Canada

For NSERC office use only

Form 101 - Application for a Grant

Send to NSERC with your attachments, if applicable

Reference Number: 90441842

Applicant: Nancy Reid
Toronto

NSERC PIN: 12406

Program: Discovery Grants - Individual

Application Title: Likelihood inference for complex data

Expected Paper Attachments:

Nancy Reid

Form 101 - Application for a Grant

- ☐ Research Contribution 1 - Preprint on mean likelihood (submitted to Bka)
- ☐ Research Contribution 2 - Preprint on default priors joint with D. Fraser
- ☐ Research Contribution 3 - Preprint on matching priors joint with A.M. Staicu
- ☐ Research Contribution 4 - Chapter 8 of Applied Asymptotics

Natural Sciences and Engineering Research Council of Canada | Conseil de recherches en sciences naturelles et en génie du Canada

FORM 101
Application for a Grant
PART I

Institutional Identifier			
System-ID (for NSERC use only) 90441842			Date 2009/04/28
Family name of applicant: Reid	Given name: Nancy	Initial(s) of all given names: NM	Personal identification no. (PIN): **Valid** 12406

Institution that will administer the grant	Language of application	Time (in hours per month) to be devoted to the proposed research / activity
Toronto	[X] English [ ] French	80

Type of grant applied for	For Strategic Projects, indicate the Target Area and the Research Topic; for Strategic Networks and Strategic Workshops indicate the Target Area.
Discovery Grants - Individual	

Title of proposal
Likelihood inference for complex data

Provide a maximum of 10 key words that describe this proposal. Use commas to separate them.
asymptotic theory, composite likelihood, conditional inference, longitudinal data, multivariate analysis

Research subject code(s)		Area of application code(s)	
Primary	Secondary	Primary	Secondary
3009	3001	1211	1212

CERTIFICATION/REQUIREMENTS

If this proposal involves any of the following, check the box(es) and submit the protocol to the university or college's certification committee.

Research involving : Humans [ ] Human pluripotent stem cells [ ] Animals [ ] Biohazards [ ]

Does any phase of the research described in this proposal a) take place outside an office or laboratory, or b) involve an undertaking as described in Part 1 of Appendix B?

[X] NO [ ] If YES to either question a) or b) – Appendices A and B must be completed

TOTAL AMOUNT REQUESTED FROM NSERC

Year 1	Year 2	Year 3	Year 4	Year 5
101,600	101,600	101,600	101,600	101,600

SIGNATURES (Refer to instructions "What do signatures mean?")

It is agreed that the general conditions governing grants as outlined in the NSERC *Program Guide for Professors* apply to any grant made pursuant to this application and are hereby accepted by the applicant and the applicant's employing institution.

Applicant

Applicant's department, institution, tel. and fax nos., and e-mail
Statistics
Toronto
Tel.: (416) 978 5046
FAX: (416) 978 5133
reid@utstat.utoronto.ca

Head of department

Dean of faculty

President of institution (or representative)

The information collected on this form and appendices will be stored in the Personal Information Bank for the appropriate program.

PROTECTED WHEN COMPLETED

Canada

Date approved by RGO: 2008/10/31

Personal identification no. (PIN)	Family name of applicant
Valid 12406	Reid

SUMMARY OF PROPOSAL FOR PUBLIC RELEASE (Use plain language.)

This plain language summary will be available to the public if your proposal is funded. Although it is not mandatory, you may choose to include your business telephone number and/or your e-mail address to facilitate contact with the public and the media about your research.

Business telephone no. (optional): 001 (416) 978-5046

E-mail address (optional): reid@utstat.utoronto.ca

Modern technology has enabled the collection of large and complex sets of data, and these are being used to answer important research questions in many fields of science and social science. Statistical methods are used to understand both the structure and the noise in this data, and new methods are being rapidly developed by statisticians working in collaboration with biologists, physicists, epidemiologists, social scientists and many others. These new methods typically involve quite complex modeling, and are computationally intensive in their implementation. In many cases these new methods are defined algorithmically in each context. The theory of statistical inference provides us with a set of guidelines for tackling new problems, provides a framework for assessing approaches developed in particular contexts, and searches for the common structure in what may seem to be very diverse problems. Statistical theory has been very successful in finding basic structure and suggesting new solutions, and as the field develops more and more sophisticated approaches to data there is an accompanying need for understanding the basis for the analysis.

The research proposed here is to develop the theory of inference to very complex settings by investigating statistical methods in detail, both mathematically and in practical application. The mathematical analysis provides insight into the process of inference from data, suggesting what information is available and how this information may be extracted. The statistical interpretation of the mathematics can suggest new ways for scientists to analyze their data. This research program emphasizes the development and dissemination of both new theory and new statistical methods.

Second Language Version of Summary (optional).

Personal identification no. (PIN)	Family name of applicant
Valid 12406	Reid

Before completing this section, **read the instructions** and consult the *Use of Grant Funds* section of the NSERC Program Guide for Professors concerning the eligibility of expenditures for the direct costs of research and the regulations governing the use of grant funds.

TOTAL PROPOSED EXPENDITURES (Include cash expenditures only)

	Year 1	Year 2	Year 3	Year 4	Year 5
1) Salaries and benefits					
a) Students	31,600	31,600	31,600	31,600	31,600
b) Postdoctoral fellows	44,000	44,000	44,000	44,000	44,000
c) Technical/professional assistants	0	0	0	0	0
d)	0	0	0	0	0
2) Equipment or facility					
a) Purchase or rental	2,000	2,000	2,000	2,000	2,000
b) Operation and maintenance costs	1,000	1,000	1,000	1,000	1,000
c) User fees	4,000	4,000	4,000	4,000	4,000
3) Materials and supplies	2,500	2,500	2,500	2,500	2,500
4) Travel					
a) Conferences	9,000	9,000	9,000	9,000	9,000
b) Field work	0	0	0	0	0
c) Collaboration/consultation	6,000	6,000	6,000	6,000	6,000
5) Dissemination costs					
a) Publication costs	1,000	1,000	1,000	1,000	1,000
b) web page maintenance	500	500	500	500	500
6) Other (specify)					
a)	0	0	0	0	0
b)	0	0	0	0	0
TOTAL PROPOSED EXPENDITURES	101,600	101,600	101,600	101,600	101,600
Total cash contribution from industry (if applicable)					
Total cash contribution from university (if applicable)					
Total cash contribution from other sources (if applicable)	0	0	0	0	0
TOTAL AMOUNT REQUESTED FROM NSERC (transfer to page 1)	101,600	101,600	101,600	101,600	101,600

Explanation of expenditures

1. Salaries and benefits

   (a) Students

   Support for 2 PhD students at $21,000 per year plus 10% benefits. In 2008-09 these students are Jin Zi and Lequn Zeng. Zi will graduate in 2009, and I will then accept a second PhD student. I have regularly supervised between 2 and 4 PhD students during the past five years and expect this rate to continue. Some students will have external scholarship support. I also expect to co-supervise two students per year and have allocated $7,000 plus 10% benefits for this.

   Support for 1 MSc student at $5,000 per year plus 10% benefits. This is used for research projects for MSc students proceeding to the PhD program, during their summer preceding entry into the program.

   Support for two summer undergraduate students at $3,000 per year. I will apply for USRA awards for these students; if successful each of two will be paid $1,500 per summer as 'top-up'. The students will work on implementation of asymptotic methods on small data sets. In the past I have regularly supported one USRA student; there is currently a large increase in our undergraduate enrollment which suggests there will be a larger pool of qualified applicants for USRAs.

   (b) Postdoctoral Fellow: $44,000 ($40,000 salary and $4,000 benefits). In 2007 I established a research group, consisting of two PDFs, two to four PhD students, occasional faculty visitors, one or more MSc students, and two or three advanced undergraduates. We meet regularly to discuss research in likelihood inference, often jointly with colleagues in the Department of Statistics. This model works extremely well because of the range of expertise available, but also very importantly because the PDFs play a leadership role in organizing the group and presenting material. This has been very effective for my own research program, but it is also clear that PDFs play an increasingly important role in statistical science. Both the students who graduated under my supervision took up PDF positions upon graduation, and this is becoming the norm in our discipline. The PDF gains a valuable opportunity to concentrate on research for two years, and the graduate students learn a great deal from interacting with the PDFs, not only in research, but in how to work in teams and communicate results effectively. As statistical science becomes increasingly interdisciplinary, this type of training is very important. Going forward I plan to have 2 PDFs per year under my supervision, one supported through my Discovery Grant and one supported via other sources of funding.

2. Equipment or facility

   (a) $2,000 on average year for computer hardware: laptop, printer, and printer cartridges. The Macintosh 15" laptop is currently selling for $2,499, and the 20" display for $899. The Mac is a powerful and convenient platform for development of software for R, for development of graphical methods, and for publishing.

   (b) $1,000 for software and software upgrades

(c) Research of students and postdoctoral fellows, and problems (such as large data sets) requiring substantial computing resources are developed on the departmental computing system, a network of Suns. User fees for this system contribute to the infrastructure needs, including the salary support of the system manager, who was in the past supported by an infrastructure grant. The fees have averaged $4,000 per year for the past three years.

3. Materials and supplies

   phone $1,000; photocopy $1,000; miscellaneous materials $500

4. Travel

   (a) Conferences

   - One overseas conference for myself and one for a PhD student or PDF each second year. In 2010 this will be the Annual Meeting of the IMS in Gothenburg, Sweden. $3,000 per year.
   - Two North American conferences annually: $2,000 for myself and $2,000 for a PDF. In 2009 these will be the SSC Annual Meeting in Vancouver and the Joint Statistical Meetings in Washington, DC.
   - four graduate students to the SSC Annual Meeting $2,000.

   (b) Collaboration

   - Anthony Davison to visit Toronto 2009 $3,000
   - Alessandra Brazzale to visit Toronto 2010: $2,000
   - travel to UK once each 2nd year for joint research with David Cox: $1,000

5. Dissemination costs

   (a) Publication costs

   - page charges for self and students: $1,000

   (b) Other

   - web design and maintenance: $500

Other sources

There is a $10,000 pa research grant attached to my University Professorship. This is divided between direct costs for this research program ($8,000) and expenses associated with the position ($2,000).

I am a co-applicant on the MRS proposal for the National Institute of Complex Data Structures. I expect to support one student each year through projects associated with that proposal. This will be in addition to the students supported as outlined above, and the research will be on different problems than outlined in this proposal. The interdisciplinary work of the NICDS projects generates new problems and ideas for statistical methodology and statistical theory, however, and the theoretical research supported by my Discovery Grant certainly benefits from this interaction.

Proposal

Progress report: *(References to papers on Form 100 are given as [n].)*

My main accomplishments during the current grant period have been the publication of a book on higher order approximations with A. Davison and A. Brazzale, the publication of a paper on composite likelihood with David Cox, and new work with Don Fraser, Ana-Maria Staicu, and Ye Sun on various aspects of the theory of higher order asymptotics, with particular emphasis on the study of the overlap between Bayesian and nonBayesian approaches. Two students have completed their PhD theses, two students are currently engaged in PhD research under my direction, and I am supervising an NSERC postdoctoral fellow (PDF) and co-supervising a second PDF.

The book *Applied Asymptotics* [21] presents a concise but detailed account of higher order asymptotic theory for likelihood inference in Chs. 2 and 8, but the main emphasis of the book is the application of these methods to models and problems that arise in practical settings. The goal is to provide enough illustrations on common classes of problems, along with computer code, to make higher order methods accessible to applied statisticians. An second goal is to illustrate both the practicality of higher order methods and their extreme accuracy on a wide variety of problems. The book has been favorably reviewed in *Short Book Reviews*, *J. Appl. Statist.* and *JRSS A*.

The asymptotic theory that my colleagues and I use is based on saddlepoint and Edgeworth expansions, interpreted in a likelihood setting. While the detailed derivations are somewhat technical, the essential point is that an approximation to the p-value function for a scalar parameter of interest, ψ, is completely determined by a pair of functions $\{\ell(\theta;y^0), \varphi(\theta;y^0)\}$, and their derivatives with respect to the parameter θ. Here $\ell(\theta;y)$ is the log-likelihood function for θ based on a response $y = (y_1, \ldots, y_n)$, y^0 is the observed value of the response, and $\varphi(\theta;y^0)$ is a re-parameterization of the model, at the observed data point, that gives an exponential model approximation to the original model, in a neighbourhood of y^0. The re-parameterization $\varphi(\theta;y^0)$ is in turn developed from a location-model approximation to the original model, and incorporates conditioning on an approximately ancillary statistic. The approximation is referred to briefly as the r^* approximation, as an inference quantity r^* can be computed from the pair $\{\ell(\theta;y^0), \varphi(\theta;y^0)\}$. The distribution of r^* is approximately standard normal, with relative error $O(n^{-3/2})$, and the formula for r^* is

$$r^* = r + \frac{1}{r}\log(\frac{q}{r}), \quad r = [2\{\ell(\hat\theta;y^0) - \ell(\hat\theta_\psi;y^0)\}]^{1/2}, \quad q = \frac{|\varphi(\hat\theta) - \varphi(\hat\theta_\psi) \quad (\partial/\partial\lambda')\varphi(\hat\theta_\psi)|}{|(\partial/\partial\theta')\varphi(\hat\theta)|}\frac{|j(\hat\theta)|^{1/2}}{|j_{\lambda\lambda}(\hat\theta_\psi)|^{1/2}}.$$

Here $\theta = (\psi, \lambda)$ is a partition of the model parameter into a scalar parameter of interest and a vector nuisance parameter, $\hat\theta$ and $\hat\theta_\psi = (\psi, \hat\lambda_\psi)$ are the unconstrained and constrained maximum likelihood estimates, $j(\hat\theta) = \partial^2\ell(\theta)/\partial\theta\partial\theta'$ is the observed Fisher information function, and $j_{\lambda\lambda}(\hat\theta_\psi) = \partial^2\ell(\theta)/\partial\lambda\partial\lambda'$ is the observed Fisher information for the nuisance parameter λ, when ψ is fixed.

We have continued to develop and extend this higher order approximation method to more general settings and to study the connections between Bayesian and frequentist inference. In [3] we showed how this approach can be applied to discrete models, where it gives $O(n^{-1})$ accuracy. We are adapting the directional test of Fraser & Massam (1985) to construct approximations for inference for vector parameters in continuous models: first steps were developed in [4] and we are continuing to refine this method. PhD student Staicu undertook a detailed study of approximations for binomial data, obtaining analytical expressions for expansions to $O(n^{-3/2})$ that illuminate the connections between various approximation methods, and explain the accuracy of the method based on r^* (Staicu, 2008). Another chapter of her thesis ([22]) studied matching priors for case-control

studies. In [1], based on Ch. 4 of [22], we showed that the class of probability matching priors developed by Tibshirani (1989) is essentially unique, when used in higher order approximations to Bayesian marginal posterior distributions. PhD student Iglesias-Gonzalez undertook a detailed study of linear and mixed linear models. In [9] Fraser and I developed a theory of strong matching: we defined a data-dependent prior by requiring that the posterior probability intervals equal the likelihood based p-value functions, to a high order of approximation. We have carried this theory much further in [11], where we study the structure of strong matching priors in detail, and propose default priors for scalar and for vector parameters. In [2], which was invited for a *Festschrift* volume for Professor Akahira, we show how higher order approximations can be used to easily assess the sensitivity of posteriors to priors, and also show that using flat priors for α and β in a logistic regression with $\text{logit}(p_i) = \alpha + \beta x_i$ regression leads to poorly calibrated inference for the ED_{50} parameter α/β. This is to be expected in light of theoretical results in [9], but does not seem to be widely appreciated in applications of Bayesian inference. I published two review papers on higher order asymptotics, the Wald lectures [8], and a paper in a *Festschrift* for D.R. Cox [16].

Recently [10] Fraser and I have discovered that an $O(n^{-1})$ approximation suggested by Skovgaard (1996, 2001) can be obtained from a pair $\{\ell(\theta; y^0), \tilde{\varphi}(\theta)\}$ in exactly the same way as the r^* approximation is obtained from $\{\ell(\theta; y^0), \varphi(\theta; y^0)\}$, where the calculation of $\tilde{\varphi}$ involves simply the derivative of the information function $I(\theta_0; \theta) = E_{\theta_0}\{\ell(\theta; y)\}$:

$$\tilde{\varphi}(\theta) = \partial I(\hat{\theta}; \theta)/\partial\hat{\theta}.$$

Since I averages over the distribution of y, $\tilde{\varphi}$ does not depend on an approximately ancillary statistic. This makes it simpler to calculate in many problems: for example in normal theory linear models with fixed and random effects (Lyons & Peters, 2000). Iglesias-Gonzalez ([23]) developed expressions for the more accurate version of r^* for this model, and showed that the third order version was just slightly more accurate, although more complex to implement.

With [6] I began work in a new area of research, that of composite likelihood, and this research is continuing in joint work with PhD students Zi and Zeng, with PDF Plante and with Professor Yun-Yi. A composite likelihood is a product of marginal or conditional likelihoods, sometimes suitably weighted: in its most general form

$$CL(\theta; y) = \prod_{s \in \mathcal{S}} f_s(y_s; \theta)^{w_s},$$

where each component f_s is either a marginal or conditional density for the components of y that fall in the subset s. In [6] we studied the pairwise likelihood $\prod_{r<s} f(y_s, y_r; \theta)$, based on the joint marginal density for each pair of components of y. Composite likelihood is a generalization of Besag's (1974) pseudo-likelihood, and was defined and studied in Lindsay (1988). More recently it has found application in a wide range of applied problems, including both discrete and continuous longitudinal data, frailty models in survival data, generalized linear mixed models, spatial models, and statistical genetics. The maximum composite likelihood estimator is consistent, although not efficient (Lindsay, 1988), as the sample size $n \to \infty$ with the dimension p fixed. In a great many settings it has been verified, usually through simulations, that the efficiency loss is relatively small, and the computational benefits are substantial. An overview of these results is given in Varin (2008). However, as we showed in [6], if $p \to \infty$ with n fixed, as is applicable typically to time series and genetics applications, composite likelihood will not in general give consistent estimates. A workshop held in April 2008 on composite likelihood summarized the current state of understanding and raised a number of open problems, and my research objectives include several problems related to composite likelihood, discussed in the next section.

Short and long term objectives

In the short term my first research objective is to understand the properties of composite likelihood inference. While efficiency of composite likelihood estimators has been well established in a number of applications, it is not at all clear why these estimators are so efficient. They are related to estimators obtained by generalized estimating equations (GEE), and in simple models can be shown to be either more or less efficient than GEE, depending on the context. Likelihood ratio type tests can be formed from composite likelihood, but their distribution is difficult to obtain in practical settings. However, likelihood ratio-based inference in full likelihood methods has better properties in general than inference based on the maximum likelihood estimator and its estimated standard error, and this may well be the case for composite likelihood inference. A particularly important question is whether or not composite likelihood is more robust to model mis-specification than likelihood methods. It seems natural that it would be, since higher order dependencies are not modelled in, for example, pairwise likelihood. My other short term objectives relate to to further development of our recent work on higher order approximations: inference for vector parameters, the interface between Bayesian and likelihood approaches, the connection between default priors and the potential failure of flat priors to be well-calibrated, and semi-parametric models.

My long term objectives are to develop likelihood-based methods for inference in models with very complex structure, and to make likelihood-based inference as broadly accessible as possible. My exposure to models used in environmetrics, survey sampling and genomics, in large part through work as a scientific reviewer, has given me a sense of the practical issues faced when using complex models, as well as the need for some unifying perspectives on approaches to inference. In many respects the underlying problems in a wide range of applications are very similar, although each application has some unique details. Recurring themes include the use of multi-level models; the use of partially parametric models; and the potential for more accurate results by combining information from several sources and of several types, for example observational studies, geographic information systems, and experimental data. I believe it is important to abstract the unifying ideas behind these applications and develop them further, in part to provide a way of understanding the complexity of the models, and in part to provide a basis for approaching new applications. Making new theoretical results accessible to practitioners as quickly and as clearly as possible is also a long-term goal that informs all my research.

Methods and proposed approach

A range of problems associated with composite likelihood will be investigated. Professor Yun-Yi and I are studying the statistical properties of estimators obtained from biased estimating equations, using a new method of adjustment. We were motivated by the problem of longitudinal binary data with missing data and/or measurement error, where the estimating equations based on the observed complete data are biased, although much simpler to use. We are using the asymptotic theory of estimating equations and numerical work to assess the properties of these estimators. The asymptotic theory of estimating equations is quite closely related to that for composite likelihood, as the latter leads to unbiased, but not fully efficient, estimating functions, and our work on biased estimating equations is closely related to the theory of composite likelihood. PhD student Jin Zi is investigating the efficiency of composite likelihood estimators, in settings where they do lead to unbiased estimating equations, in models for discrete and continuous data. In some settings composite likelihood estimators are fully efficient, and we are working on an asymptotic theory to try to explain this. One possible explanation is related to the theory of full exponential family

models, which have minimal sufficient statistics of the same dimension as the parameter. In the multivariate models where composite likelihood is fully efficient the same property holds.

An advantage of composite likelihood over GEE, is the existence of an objective function, which in principle has further information beyond that of the GEE, for example enabling choice between multiple roots. However likelihood ratio-type statistics based on composite likelihood have a complicated asymptotic distribution, a weighted sum of χ^2, with weights depending on the eigenvalues of a matrix related to the variance of the composite likelihood estimator. PDF Ye Sun and I are in the planning stages of a project that will combine this with saddlepoint approximations for quadratic forms, discussed for example in Kuonen (1999). This will need to be assessed numerically in a number of models; if these results are promising then the next step will be to develop asymptotic expansions under model mis-specification, building on work of Viraswami & Reid (1996). This work will be continued with a future PhD student. A quite different approach will be pursued by PhD student Lequn Zeng, who is just beginning his research. He will study higher order asymptotic theory for composite likelihood. It is straightforward to evaluate the information function $I(\theta_0;\theta)$ for composite likelihood, but whether or not this can lead to higher order approximation, as in [10], is an open question. He is beginning by investigating the examples studied in [6], where comparison with full likelihood models is analytically possible.

Tackling the robustness question of composite likelihood is quite challenging; it has proved difficult to come up with the right formulation of the problem. I will focus attention on pairwise likelihood, where the robustness in question would be against making strong assumptions about joint distributions of third and higher order. The difficulty is that it is not clear to me how assumptions about the joint density of pairs constrain the higher order joint distributions. PDF Jean-François Plante and I are considering approaching this through the theory of copulas. Another approach would be to study the results on optimal weighting of marginal distributions, discussed for example in Zhao & Joe (2005) and Kuk & Nott (2000) in the context of pairwise likelihood for clustered familial data, assuming that there is likely to be a trade-off between robustness and efficiency. Plante and I are also investigating the connections between weighted likelihood, the topic of his thesis, and composite likelihood.

I have agreed to be a guest editor for a volume of *Statistica Sinica* on composite likelihood, and as part of this plan to write an overview of research presented at a workshop in Warwick in April, 2007, updating the review paper by Varin (2008).

Methods to be used for new results in higher order approximation are a combination of analytical results based on asymptotic expansions and numerical work. Fraser and I are extending our approach to higher order inference to vector parameters of interest by using the tangent exponential model with both versions of the approximate canonical parameter: the third order version φ based on approximate ancillarity and the simpler version $\tilde{\varphi}$ described above. Extensive numerical work on the comparison of these approaches will be the topic of summer undergraduate projects for 2009. I also plan to have research students examine the agreement between Bayesian and frequentist inference in complex models through simulations and Laplace approximations. For likelihood-based and Bayesian approaches to agree beyond the first order of approximation, it is necessary for the prior to be data-dependent. The approach of empirical Bayes inference also involves a data-dependent prior, although to my knowledge there is no current research that tries to make this connection. The natural place to start to investigate this is in the context of hierarchical models. Some very preliminary work is reported in [2], and the simplification of higher order approximations presented in [10] gives a way to extend this to much more complex hierarchical models. It would be very useful to have some understanding of when flat priors can be used without markedly affecting the posterior

inference. Some progress on this is made in [11], but considering this in the context of applications, of the type treated for example in Gelman et al. (2007), would make this work more practically relevant. This will require substantial computing with large data sets, and would be suitable for an MSc project. In [11] we argued that the parameter of interest has to be flat in an approximate location parameter; my goal is to translate this into an index that could be readily computed by practitioners. An exciting possibility for obtaining higher order approximations for semi-parametric models has emerged that uses [10] in conjunction with the formulation in Crainiceanu et al. (2005) of penalised splines as mixed linear models; joint work on this with Ye Sun is just starting.

In [21] our goal was to make the use of higher order asymptotics easy for practitioners. My goal now is to provide a complete, but concise, account of the theory. The book will emphasize the methods Fraser and I have developed over the past several years, and their connection with the work outlined in the books by Barndorff-Nielsen & Cox (1994), Severini (2000) and Butler (2007). The book will provide an accessible systematic reference for students and researchers, summarize the current state of research in this area, and draw parallels between Bayesian and frequentist methods.

Anticipated significance of the work

Composite likelihood inference is becoming widely used, but there is as yet little work that provides a general understanding of its relation to the more widely used methods. There is also very little work on the asymptotic theory of increasing dimension (p), with fixed sample size (n), beyond that in [6]. The large p small n problem arises in many applications, most prominently in genetics, but increasingly in health and social science applications. Composite likelihood also has the potential to deepen our understanding of multivariate distributions. For example, it is relatively easy to introduce various types of correlation in discrete data by introducing latent random effects. It is much harder to see the structure of the resulting multivariate binary distribution, or even if indeed it is a real distribution. There is empirical evidence (e.g. Liang & Yu, 2003) that the composite likelihood surface can be not only computationally cheaper, but much smoother than the true likelihood surface. An understanding of these results would increase the utility of CL methods and potentially provide warnings about classes of problems that should not be treated in this way.

Improved approximations based on higher order likelihood theory are useful for applications, and are increasingly easier to implement. But a more important benefit of the detailed study of approximations is to deepen our understanding of the nature of model-based inference. For example, the fact that an $O(n^{-1})$ approximation to an arbitrary continuous density gives an $O(n^{-3/2})$ to the p-value relies on a surprisingly simple property of the normal distribution (Andrews et al., 2005). Asymptotics also sheds light on when and why flat priors are highly informative for marginal inference about non-linear parameters, and shows explicitly that frequentist and Bayesian methods diverge at $O(n^{-1})$, unless data-dependent priors are used. Results like these deepen our understanding of statistical inference, which benefits both theoretical and applied statistical science.

Training to take place through the proposal

Training has been described throughout the proposal in the context of specific research problems. Students in my research group range from undergraduate to PDF, and are trained in computational, applied and theoretical areas of statistics. They have opportunities to contribute to all aspects of research, from background work through to publication, and our regular group meetings help to develop their skills in collaboration and communication.

References

Andrews, D.F., Fraser, D.A.S. & Wong, A. (2005). Computation of distribution functions from likelihood information near observed data. *J. Statist. Plann. Inf.* **134**, 180–193.

Barndorff-Nielsen, O.E. & Cox, D.R. (1994). *Inference and Asymptotics.* Chapman & Hall, London.

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. R. Statist. Soc.* B, **34**, 192–236.

Butler, R. (2007). *Saddlepoint Approximations.* Cambridge University Press: Cambridge.

Crainiceanu, C., Ruppert D. & Claeskens, G. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika* **92**, 91–103.

Fraser, D.A.S. & Massam, H. (1985). Conical tests: Observed levels of significance and confidence regions. *Statistische Hefte* **26**, 1-17.

Gelman, A. , Fagan, J. & Kiss, A. (2007). An analysis of the NYPD's stop-and-frisk policy in the context of claims of racial bias. *Journal of the American Statistical Association* **102**, 813–823.

Kuonen, D. (1999). Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika* **86**, 929–935.

Kuk, A.Y.C. & Nott, D.J. (2000). A pairwise likelihood approach to analyzing correlated binary data. *Stat. Prob. Lett.* **47**, 329–35.

Liang, G. & Yu, B. (2003). Maximum pseudo-likelihood estimation in network tomography. *IEEE Trans. Sig. Proc.* **51**, 2043–2053.

Lindsay, B.L. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes*, Ed. N.U. Prabhu, pp. 221–239. Providence: American Mathematical Society.

Lyons, B. & Peters, D. (2000). Applying Skovgaards modied directed likelihood statistics to mixed linear models. *J. Statist. Comp. Sim.* **65**, 225242.

Severini, T.J. (2000). *Likelihood Methods in Statistics.* Oxford University Press, Oxford.

Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli* **2**, 145165.

Skovgaard, I.M. (2001). Likelihood asymptotics. *Scand. J. Statist.* **28**, 3–32.

Staicu, A.-M. (2008). Higher order approximations for interval estimation in binomial settings. in revision for *J. Statist. Plann. Inf.*

Tibshirani, R.J. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604–608.

Varin,C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis* **92**, 1-28.

Viraswami, K. & Reid, N. (1996). Higher order asymptotics under model misspecification. *Canad. J. Statist.* **24**, 263–278.

Zhao, Y. & Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canad. J. Statist.* **33**, 335–356.

Nancy Reid

Form 101 - Application for a Grant

Research Contribution 1

Preprint on mean likelihood (submitted to Bka)

Nancy Reid

Form 101 - Application for a Grant

Research Contribution 2

Preprint on default priors joint with D. Fraser

Nancy Reid

Form 101 - Application for a Grant

Research Contribution 3

Preprint on matching priors joint with A.M. Staicu

Nancy Reid

Form 101 - Application for a Grant

Research Contribution 4

Chapter 8 of Applied Asymptotics

Natural Sciences and Engineering Research Council of Canada | Conseil de recherches en sciences naturelles et en génie du Canada

FORM 100
Personal Data Form
PART I

Date: 2009/04/28

Family name	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Reid	Nancy	NM	12406

☐ I hold a faculty position at an eligible Canadian college (complete Appendices B1 and C)

☐ I do not or will not hold an academic appointment at a Canadian postsecondary institution

Place of employment other than a Canadian postsecondary Institution (give address in Appendix A)

APPOINTMENT AT A POSTSECONDARY INSTITUTION

Title of position: Professor

Department: Statistics

Campus: St. George

Canadian postsecondary institution: Toronto

Tenured or tenure-track academic appointment — Yes [X] No [ ]

Part-time appointment [ ] Full-time appointment [X]

- For all non-tenured or non tenure-track academic appointment and Emeritus Professors, complete Appendices B & C
- For life-time Emeritus Professor and part-time positions, complete Appendix C

ACADEMIC BACKGROUND

Degree	Name of discipline	Institution	Country	Date yyyy/mm
Bachelor's	Statistics	University of Waterloo	Canada	1974 / 06
Master's	Statistics	University of British Columbia	Canada	1976 / 11
Doctorate	Statistics	Stanford University	United States	1979 / 06

TRAINING OF HIGHLY QUALIFIED PERSONNEL

Indicate the number of students, fellows and other research personnel that you:

	Currently		Over the past six years (excluding the current year)		
	Supervised	Co-supervised	Supervised	Co-supervised	Total
Undergraduate	1	1	5		7
Master's	1				1
Doctoral	2	2	4	1	9
Postdoctoral	1	1	2		4
Others					
Total	5	4	11	1	21

Personal information collected on this form and appendices will be stored in the Personal Information Bank for the appropriate program.

PROTECTED WHEN COMPLETED

Version française disponible

Not Completed

Personal identification no. (PIN)	Family name
12406	Reid

ACADEMIC, RESEARCH AND INDUSTRIAL EXPERIENCE (use one additional page if necessary)

Position held (begin with current)	Organization	Department	Period (yyyy/mm to yyyy/mm)
Professor	Toronto	Statistics	1989/07
Professor and Chair	U Toronto	Statistics	1997/07 to 2002/06
Professor	U Toronto	Statistics	1989/07
Associate Professor	U Toronto	Statistics	1986/07 to 1989/06
Assistant Professor	U British Columbia	Statistics	1980/07 to 1986/06
Postdoctoral Fellow	Imperial College, London	Mathematics	1979/09 to 1980/06

PROTECTED WHEN COMPLETED

Personal identification no. (PIN)	Family name
12406	Reid

RESEARCH SUPPORT

List all sources of support **(including NSERC grants and university start-up funds)** held as an applicant or a co-applicant: a) support held in the past four (4) years but now completed; b) support currently held, and c) support applied for. For group grants, indicate the percentage of the funding directly applicable to your research. Use additional pages as required.

Family name and initial(s) of applicant	Title of proposal, funding source and program, and time commitment (hours/month)	Amount per year	Years of tenure (yyyy)
a) Support held in the past 4 years			
Nancy Reid	Statistical Methods for Complex Survey Data National Centres of Excellence MITACS 10 hours/month	11,583	2004
Nancy Reid	Bootstrap and likelihood methods for surveys NSERC National Program on Complex Data Structures 10 hours/month	15,000	2005
b) Support currently held			
Nancy Reid	Inference and Applications NSERC Research Grants 80 hours/month	48,000 48,000 48,000 48,000 48,000	2004 2004 2006 2007 2008
Nancy Reid	Statistical Inference University of Toronto University Professors Research Grant 20 hours/month	10,000 10,000 10,000 10,000 10,000	2004 2005 2006 2007 2008

PROTECTED WHEN COMPLETED

Personal identification no. (PIN)	Family name
12406	Reid

RESEARCH SUPPORT

Family name and initial(s) of applicant	Title of proposal, funding source and program, and time commitment (hours/month)	Amount per year	Years of tenure (yyyy)

List all sources of support **(including NSERC grants and university start-up funds)** held as an applicant or a co-applicant: a) support held in the past four (4) years but now completed; b) support currently held, and c) support applied for. For group grants, indicate the percentage of the funding directly applicable to your research. Use additional pages as required.

b) Support currently held			
Nancy Reid	Statistical theory and applications	10,000	2007
	NSERC	10,000	2008
	Canada Research Chairs Program		
	20 hours/month		
c) Support applied for			
Nancy Reid	Statistical inference for complex data	101,600	2009
	NSERC	101,600	2010
	Discovery Grant	101,600	2011
	80 hours/month	101,600	2012
		101,600	2013
Nancy Reid	National Institute for Complex Data Structures	735,000 (2%)	2009
	NSERC	735,000 (2%)	2010
	Major Research Resources	735,000 (2%)	2011
	30 hours/month	735,000 (2%)	2012
		735,000 (2%)	2013

Highly Qualified Personnel (HQP)

Provide personal data about the HQP that you currently, or over the past six years, have supervised or co-supervised.

Personal identification no. (PIN)	Family name
12406	Reid

Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Lequn Zeng	Doctoral (In Progress)	Supervised 2008 -	Asymptotic theory for composite likelihood	PhD student, statistics, U Toronto
Liu, ZiZhen	Master's (In Progress)	Supervised 2008 -	R programs for Cox and Snell's Applied Statistics	MSc student, U Toronto (statistics)
Elif Acar	Doctoral (In Progress)	Co-supervised 2007 -	Covariate-dependent conditional copula models	PhD student, statistics, U Toronto
Plante, Jean-François	Postdoctoral (In Progress)	Supervised 2007 -	Weighted likelihood methods	Postdoctoral Fellow, U Toronto
Shelley Yun Cao	Doctoral (In Progress)	Co-supervised 2007 -	Bayesian Factor Analysis	PhD student, statistics, U Toronto
Sun, Ye	Postdoctoral (In Progress)	Co-supervised 2007 -	Applications of higher order asymptotics	Postdoctoral Fellow, U Toronto
Jin, Zi	Doctoral (In Progress)	Supervised 2005 -	Composite likelihood	PhD student, statistics, U Toronto
Chouldechova Alexandra	Undergraduate (Completed)	Co-supervised 2008 - 2008	Prior influence on posterior inference	Undergraduate student, U Toronto
Lin, Wei	Undergraduate (Completed)	Supervised 2008 - 2008	Design of experiments	MSc student, U Toronto (statistics)
Romanescu, Razvan	Undergraduate (Completed)	Supervised 2007 - 2007	Bayesian analysis of mixed linear models	MSc student, U Waterloo
Sigfrido Iglesias-Gonza	Postdoctoral (Completed)	Supervised 2007 - 2007	Asymptotic methods for mixed linear models	Postdoctoral Fellow, CIMAT
Zhong, Sheng	Undergraduate (Completed)	Supervised 2007 - 2007	Generalized linear mixed models	PhD student, U Chicago
Kane, Mark	Doctoral (Not Completed)	Co-supervised 2003 - 2007	Asymptotics for quantile regression	PhD student
Staicu, Ana-Maria	Doctoral (Completed)	Supervised 2003 - 2007	Likelihood methods with applications in biostatistics	Postdoctoral Fellow, University of Bristol
Zheng, Zheng	Doctoral (Not Completed)	Supervised 2003 - 2007	Bootstrap and MCMC methods for likelihood inference	PhD student
IglesiasGonza	Doctoral (Completed)	Supervised 2002 - 2007	Highly accurate tests for the mixed linear model	Postdoctoral Fellow, CIMAT
Hong, Zengxin	Doctoral (Not Completed)	Supervised 1999 - 2007	Likelihood inference for semiparametric models	PhD student
Shi, Xia	Undergraduate (Completed)	Supervised 2005 - 2005	Statistical methods for data mining	Statistician, Endurance Re-insurance
Zhu, Lizhen	Undergraduate (Completed)	Supervised 2004 - 2004	Computer implementation of hoa	PhD student, statistics, U Toronto
Brazzale, Alessandra	Postdoctoral (Completed)	Supervised 2003 - 2003	Implementation of Fraser-Reid approach	Associate Professor, U. Reggio Emiliano

Highly Qualified Personnel (HQP)

Provide personal data about the HQP that you currently, or over the past six years, have supervised or co-supervised.

Personal identification no. (PIN)	Family name
12406	Reid

Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Mazumder, Anjali	Undergraduate (Completed)	Supervised 2002 - 2002	Research topics in statistics	PhD student, U Oxford

PART II. RESEARCH CONTRIBUTIONS

1. Most significant contributions (last 6 years)

1. *Applied Asymptotics: Case Studies in Small Sample Statistics*. [17]
This book is the culmination of several years of work on the practical use of higher order asymptotics in applied statistical work, and on making the main results in the theory of higher order asymptotics accessible to the non-specialist. The book consists of a large number of case studies in situations of realistic statistical complexity, detailed computer code in `R`, and a large number of problems and extensions, many involving new research.

2. **The theory of higher order likelihood asymptotics**. I have continued my work in this area in a number of separate, but related directions, including work on the interface between Bayesian and frequentist approaches ([1], [2], [16], [18]), investigation of more complex models (the unpublished theses of Staicu and of Iglesias-Gonzalez), application to vector parameter settings ([2]), discrete data ([3]), and most recently ([10]) a new connection discovered among several versions of higher order asymptotic formulae for tail probabilities.

3. **Composite Likelihood**. My work in this area began with [7], which considered the use of univariate and bivariate marginal distributions, in multivariate settings, to construct a composite likelihood. We considered the limiting properties of the pseudo-score equation, both for increasing sample size and for increasing dimension of the parameter, and show in the latter case that the score equation will not usually lead to a consistent estimator. The case of increasing parameter dimension is relevant to the context where there are a large number of measurements on a small number of individuals. Paper [7] has been widely cited, as there is considerable current interest in using these pseudo or composite likelihoods in models for which the exact computation of the likelihood is infeasible. I co-organized a workshop at Warwick University in spring 2008, and have been invited to organize a session on composite likelihood for the 2009 meeting of the Institute of Mathematical Statistics. PhD student Ji Zin is completing her thesis work on the efficiency of composite likelihood, and with Professor Grace Yun-Yi we are working on several related problems. PhD student Lequn Zeng is beginning his PhD research with a study of asymptotics connected with composite likelihood.

4. **Complex data structures**. As part of my involvement with the National Institute on Complex Data Structures, I have been investigating likelihood and composite likelihood methods for data from areas closely related to applications in the health and social sciences. A project with PhD student Zheng Zheng compared bootstrap methods with higher order approximations in the context of survey data, and ongoing work with PDF Jean-François Plante is studying the relationship between weighted likelihood and composite likelihood for settings relevant to meta-analysis.

5. **Statistical methods in high energy physics**. I have been involved since 2003 with the series of *Phystat* conferences, in which a small-ish number of statisticians meets in a workshop setting with particle physicists to discuss methods of statistical inference of use for data obtained from high energy particle colliders. I have publications in the proceedings of *Phystat03, Phystat05* and *Phystat07*, and *Phys. Rev. D*. I was the principal organizer for a 5 day workshop at the Banff International Research Station in July 2006, which set a number of research directions targetted on analysis of data from the Large Hadron Collider at CERN.

2. List of Further Research Contributions (last 6 years)

Note: Authorship is alphabetical except in [1], which is based on this student's thesis and in [9], [18] and [19], which I wrote based on our joint work. In all other cases contributions of joint authors are equal. All research supported by NSERC.

2.1. Articles in refereed publications

[1] **Staicu, A.-M.** and Reid, N. (2008). On probability matching priors. *Canad. J. Statist.* **36**, to appear (December).

[2] Reid, N. and **Sun, Y.** (2008). Assessing sensitivity to priors using higher order asymptotics. *Communications in Statistics: Theory and Methods*, to appear.

[3] Davison, A.C., Fraser, D.A.S. and Reid, N. (2006). Likelihood inference for categorical data. *J. R. Statist. Soc* B **68**, 495–508.

[4] Fraser, D.A.S. and Reid, N. (2006) Assessing a vector parameter. *Student* **5**, 247–256.

[5] Fraser, D.A.S., Reid, N. and Wong, A.C.M. (2004). Inference for bounded parameters: a different perspective. *Phys. Rev. D* **69**, 033002.

[6] Cox, D.R. and Reid, N. (2004). A note on pseudo-likelihood constructed from marginal densities. *Biometrika* **91**, 729–737.

[7] Fraser, D.A.S., Reid, N., Li, R., and Wong, A. (2003). p-value formulas from likelihood asymptotics: Bridging the singularities. *J. Statist. Research.* **37**, 1-15.

[8] Reid, N. (2003). Asymptotics and the theory of inference. *Ann. Statist.*, **31**, 1695–1731.

[9] Fraser, D.A.S. and Reid, N. (2002). Strong matching of frequentist and Bayesian parametric inference. *J. Statist. Plann. Inf.* **103**, 263–285.

Submitted

[10] Reid, N. and Fraser, D.A.S. (2008). Mean likelihood and higher order approximations. submitted to *Biometrika*. 15 ms pages.

[11] Fraser, D.A.S., Reid, N., Marras, E. and Yun-Yi, G. (2008). Default priors for Bayesian and frequentist inference. submitted to *J. R. Statist. Soc.* B. 28 ms pages.

[12] Ghosh, M., Fraser, D.A.S. and Reid, N. (2008) Ancillary statistics: an overview. submitted to *Statistica Sinica*. 10 ms pages.

2.2. Other refereed contributions

[13] Reid, N. (2008). Some aspects of design of experiments. in Proceedings of PHYSTAT Workshop On Statistical Issues for LHC Physics, eds. H.B. Prosper, L. Lyons and A. DeRoeck, 99-110.

[14] Cox, D.R. and Reid, N. (2008). The wish-list: some comments. In Proceedings of PHYSTAT Workshop On Statistical Issues for LHC Physics, eds. H.B. Prosper, L. Lyons and A. DeRoeck, 120-124.

[15] Reid, N. (2006). Summary of statistical issues arising in PhyStat2005. in *Statistical Problems in Particle Physics, Astrophysics and Cosmology: Proceedings of PHYSTAT2005*, L. Lyons and M. Ünel, eds. World Scientific, London. 279–282

[16] Reid, N. (2005). Asymptotics and the theory of statistics. in *Celebrating Statistics: Papers in Honour of D.R. Cox*, eds. A.C. Davison, Y.Dodge, N.Wermuth. Oxford University Press, Oxford, 73–88.

[17] Fraser, D.A.S., Reid, N. and Yun-Yi, G. (2003). Direct Bayes for interest parameters. in *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (eds) 529–534. Oxford University Press, Oxford.

[18] Reid, N., Mukerjee, R. and Fraser, D.A.S. (2003) Some aspects of matching priors. in *Mathematical Statistics and Applications: Festschrift for C. VanEeden* (M. Moore, S. Froda, C. Léger,

eds.) 31–44. Lecture notes Monograph Series 42, Institute of Mathematical Statistics, Hayward.
[19] Reid, N. and Fraser, D.A.S. (2003). Likelihood inference in the presence of nuisance parameters. in *Proceedings of PHYSTAT2003*, L. Lyons, R. Mount, R. Reitmeyer, eds. SLAC e-Conf C030908, 265–271.

2.3. Non-refereed contributions

[20] Reid, N. (2008). Introduction to "Using specially designed exponential families for density estimation" by B. Efron and R. Tibshirani, in *The Science of Bradley Efron*, C. Morris and R. Tibshirani, eds., Springer-Verlag, New York. 302–304.
[21] *Applied Asymptotics: Case Studies in Higher Order Asymptotics.* (2007). A.R. Brazzale, A.C. Davison and N. Reid. Cambridge University Press: Cambridge.
[22] *On Some Aspects of Likelihood Methods with Application to Biostatistics.* **Ana-Maria Staicu** (2007). PhD Dissertation, University of Toronto.
[23] *Highly Accurate Tests for the Mixed Linear Model.* **Sigfrido Iglesias-Gonzalez** (2007). PhD Dissertation, University of Toronto.
[24] Reid, N. (2006). Discussion of "Treatment of nuisance parameters in high energy physics" by R.D. Cousins, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology: Proceedings of PHYSTAT2005*, L. Lyons and M. Ünel, eds. World Scientific, London. 86–87.
[25] Reid, N. (2005). Contribution to the discussion of "Model choice in time series studies of air pollution and mortality by R. D. Peng, F. Dominici and T. A. Louis, *J. R. Statist. Soc.* A, 16, p.200.
[26] Reid, N. (2003). Contribution to the discussion of Berger (2003): "Could Neyman, Fisher and Jeffreys have agreed on testing?". *Statist. Sci.* **18**, 27.
[27] Reid, N. (2006). Orthogonal parameters. *Encyclopedia of Statistical Sciences.* DOI: 10.1002/0471667196.ess6059.
[28] D.A.S. Fraser and Reid, N. (2002). Contribution to the discussion of McCullagh (2002) "What is a statistical model?" *Ann. Statist.*, **30**, 1283–1286.
[29] Reid, N. (2002). Contribution to the discussion of Chen, Lockhart and Stephens (2002) "Asymptotic theory for the transformation model." *Canad. J. Statist.*, **30**, 211–212.

2.4. Contributions to practical applications of knowledge

Since 2003 I have served on the Scientific Advisory Committee for the National Program on Complex Data Structures, and in 2007-2009 was a co-applicant and served as a member of the writing team for the MRS proposal for its successor, the National Institute for Complex Data Structures, submitted in October 2007, the appeal of the decision, submitted in May 2008, and the re-application, submitted in October 2008. The NICDS develops and funds interdisciplinary projects with statistical leadership to address scientific problems involving the analysis of complex data.

From 2006-2008 I served on the advisory board for the Genome Canada funded project on gene-environment interactions in Type I diabetes; the PI is Dr. Jayne Danska of the University Health Network.

I have been involved since 2003 in a collaboration with high energy physicists on the analysis of data from particle accelerator experiments. This has taken the form of review lectures given at the PhyStat series of workshops, in 2003 in Stanford, 2005 in Oxford, and 2007 in Geneva. I also co-organized a workshop at BIRS in July 2006 which set a number of research directions targetted on analysis of data from the Large Hadron Collider at CERN.

From 2000-2008 I served on the review committee of the Health Effects Institute (HEI), Boston. HEI is an independent, nonprofit corporation which funds and publishes research findings on the health effects of pollution; each major publication includes a commentary by the review committee. To date I have contributed statistical reviews on approximately 40 studies, and contributed to commentaries on approximately 15 studies. I also served on the review panel for a special report on Revised Analyses of Time-Series Studies of Air Pollution and Health.

3. Other Evidence of Impact and Contributions

Prestigious Invited Lectures

Kuwait Foundation Lecture, Cambridge University, April 2009
17th Ralph Bradley Lectures, University of Georgia, April, 2009
Parzen Prize Lecture, Texas A&M University, May, 2008
Craig Lectures, University of Iowa, October, 2007
10th Anniversary Lecture Series, Pacific Institute for Mathematical Sciences, April, 2007
ADVANCE Distinguished Lecturer, Case Western Reserve University, September, 2004

Honours

Parzen Prize for Statistical Innovation, 2008 (awarded July 2007)
Canada Research Chair, Tier I, University of Toronto, 2007
Elected Member, Sigma Xi, 2006
University Professor, University of Toronto, 2003
Distinguished Alumni Achievement Award, University of Waterloo, 2003
Fellow, Fields Institute for Research in the Mathematical Sciences, 2003
Fellow, American Association for the Advancement of Science, 2002

Editorial Work

Associate Editor: *Statistical Science* 2008 –
Associate Editor: *Bernoulli* 2007 –
Associate Editor: *Metrika* 2008 –
Associate Editor: *J. Royal Statist. Soc.* B 2003–2007
Associate Editor: *Annals of Statistics* 1998 – 2003
Associate Editor: Chapman & Hall/CRC Monograph Series 1990–2004

Major committee/society contributions

Chair: Statistical Society of Canada NSERC Liaison Committee, 2007–2009
Chair: External Review Committee, Department of Statistics, UBC, 2008
Member: Program Committee, Workshop on Composite Likelihood Methods, April 2008
Chair: Committee for Fields Institute Distinguished Lecture Series on Statistical Science, 2008–
Member: Scientific Program Committee, Int'l Congress of Industrial and Applied Math 2011
Member: External Review Committee, Comp. & Math. Sci., U Tor. at Scarborough, April 2007
Chair: Awards Committee, Statistical Society of Canada, 2006–2007
Chair: Organizing Committee, BIRS Workshop on Statistics and Physics, Banff July 2006
President: Statistical Society of Canada, 2004–2005
Member: Scientific Advisory Board, Gene-Environment Interactions in Type 1 Diabetes, PI Dr. J. Danska, 2006 –2008
Member: PIMS Scientific Review Panel 2005–
Panellist: NSF Review Panel for SAMSI, 2005
Member: NSF Grant Review Panel, December 2005
Review Committee: Health Effects Institute, 2000–2008

Scientific Advisory Panel: Fields Institute, 1999–2003
Research Management Committee: Mitacs NCE, 1999–2004

Additional Invited Lectures

Conferences: Invited speaker: Annual Meeting of the SSC, Ottawa, May 2008.
Workshop on Composite Likelihiood Methods, Warwick, April 2008.
Gregynog Conference on Statistics, Newtown, Wales, April 2008.
Phystat07, CERN, June 2007.
First Annual Canada-Mexico Statistics Meeting, Guanajuato, February 2008.
University of Waterloo Department of Statistics 40th Anniversary Conference. Waterloo, May 2007.
Canadian Association of Specialised Library and Information Services, February 2007.
Workshop on Frontiers of Statistics: in honour of Peter Bickel, Princeton University, May 2006.
Southern Ontario Graduate Students Statistics days, May 2006.
Workshop on Statistics and Physics, SAMSI, March 2006.
Conference in honour of O.E. Barndorff-Nielsen. Guanajuato, MX, March 2006.
Winter Workshop on Frontiers in Statistics, University of Florida, January 2006.
Statistical Society of Montreal, Montreal, Quebec, November 2004.
Symposium in honour of D.R. Cox, Neuchatel, July, 2004.
Universities: University of Georgia, April 2009; University of California Santa Cruz, October 2008; Texas A& M, May 2008; University of Iowa, October 2007; Institute for Clinical and Evaluative Sciences, Sunnybrook Health Sciences Centre, May 2006; University of Windsor, March 2006; York University, February 2006; University of Michigan, November 2005; ADVANCE Distinguished Lecturer, Case Western Reserve University, September 2004; University of Padua, November 2004; University of Wisconsin, April 2004; McMaster University, Nov 03; EPFL Lausanne, May 03; University of Geneva, April 03; ETH Zurich, April 03; U. Oxford, Feb 03; U. Bern, Nov 02

5. Contributions to the Training of Highly Qualified Personnel

I am currently supervising postdoctoral fellow Jean-François Plante, 2007–2009 and co-supervising postdoctoral fellow Ye Sun, 2007–2009. I supervised postdoctoral fellow Sigfrido Iglesias-Gonzalez in 2007 (May to December),

I am currently supervising two PhD students. Jin Zi is completing her dissertation on efficiency of pseudo- and composite likelihood. Lequn Zeng is beginning his research on higher order asymptotics for composite likelihood. From 2003-2007 I supervised 4 PhD students: Ana-Maria Staicu and Sigfrido Iglesias-Gonzalez graduated in 2007, and very unfortunately students Zengxin Hong and Zheng Zheng have both lapsed their PhD programs due to personal circumstances. I am currently co-supervising two PhD students: Elif Acar and Shelley Yun Cao.

In summer 2008 I supervised two USRA award holders: Alexandra Chouldocheva and Wei Lin, in projects on Bayesian inference and Design of Experiments, respectively. I supervised a research associate, ZiZhen Liu, who is continuing his research through the academic year 2008-09, in a project to prepare at R library for the book *Applied Statistics* by Cox and Snell. In summer 2007 I supervised an USRA award holder (Razvan Romanescu) and a UTEA award holder (Sheng Zhong). (A UTEA is an internal U Toronto summer award for non NSERC-eligible students). In summer 2005 I supervised USRA award holder Lizhen Zhu, on a project in higher order asymptotics. In summer 2004 I supervised two undergraduate research assistants on projects in data mining and higher order asymptotics.

Natural Sciences and Engineering Research Council of Canada | Conseil de recherches en sciences naturelles et en génie du Canada

SEND ONE ORIGINAL ONLY DO NOT PHOTOCOPY

APPENDIX A
Personal Data
(Form 100)

Complete this appendix (i) if you are an applicant or co-applicant applying for the first time; (ii) if you need to update information submitted with a previous application; or (iii) if you do not hold an appointment at a Canadian postsecondary institution. For updates, include only the revised information in addition to the date, your name and your PIN.

This information will be used by NSERC primarily to contact applicants and award holders. It may also be used to identify prospective reviewers and committee members, and to generate statistics. It will not be seen or used in the adjudication process.

Date: 2009/04/28

Family name	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Reid	Nancy	NM	12406

Position and complete mailing address if your primary place of employment is not a Canadian postsecondary institution or if your current mailing address is temporary

100 St. George St.

Toronto ON M5S3G3
CANADA

If address is temporary, indicate:

Starting date

Leaving date

Telephone number	Facsimile number	E-mail address
(416) 978 5046	(416) 978 5133	reid@utstat.utoronto.ca

Telephone number (alternate)

Give an alternate telephone number only if you can be reached at that number during business hours.

Gender (completion optional): [ ] Male [X] Female

LANGUAGE CAPABILITY

English	Read [X]	Write [X]	Speak [X]
French	Read [ ]	Write [ ]	Speak [ ]

I wish to receive my correspondence: in English [X] in French [ ]

AREA(S) OF EXPERTISE

Provide a maximum of 10 key words that describe your area(s) of expertise. Use commas to separate them. If you have expertise with particular instruments and techniques, specify which one(s).

asymptotic theory, conditional inference, likelihood, saddlepoint approximations, design of experiments, Bayesian inference, robust estimation

Research subject code(s)

Primary: 3001

Secondary: 3004

Appendix D (Form 100)
Consent to Provide Limited Personal Information About Highly Qualified Personnel (HQP) to NSERC

NSERC applicants are required to describe their contributions to the training or supervision of highly qualified personnel (HQP) by providing certain details about the individuals they have trained or supervised during the six years prior to their current application. HQP information must be entered on the Personal Data Form (Form 100). This information includes the trainee's name, type of HQP training (e.g., undergraduate, master's, technical etc.) and status (completed, in-progress, incomplete), years supervised or co-supervised, title of the project or thesis, and the individual's present position.

Based on the federal *Privacy Act* rules governing the collection of personal information, applicants are asked to obtain consent from the individuals they have supervised before providing personal data about them to NSERC. In seeking this consent, the NSERC applicant must inform these individuals what data will be supplied, and assure them that it will only be used by NSERC for the purpose of assessing the applicant's contribution to HQP training. To reduce seeking consent for multiple applications, applicants will only need to seek consent one time for a six-year period. If the trainee provides consent by e-mail, the response must include confirmation that they have read and agree to the text of the consent form.

When consent cannot be obtained, applicants are asked to not provide names, or other combinations of data, that would identify those supervised. However, they may still provide the type of HQP training and status, years supervised or co-supervised, a general description of the project or thesis, and a general indication of the individual's present position if known.

An example of entering HQP information on Form 100 (with and without consent):

Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Consent Received from Marie Roy				
Roy, Marie	Undergraduate (Completed)	Supervised 1994 - 1997	Isotope geochemistry in petroleum engineering	V-P (Research), Earth Analytics Inc., Calgary, Alberta
Consent Not Obtained from Marie Roy				
(name withheld)	Undergraduate (Completed)	Supervised 1994 - 1997	Isotope geochemistry	research executive in petroleum industry - western Canada

Consent Form

Name of Trainee

Applicant Information

Name Reid, Nancy NM

Department	Postsecondary Institution
Statistics	Toronto

I hereby allow the above-named applicant to include limited personal data about me in grant applications submitted for consideration to NSERC for the next six years. This limited data will only include my name, type of HQP training and status, years supervised or co-supervised, title of the project or thesis and, to the best of the applicant's knowledge, my position title and company or organization at the time the application is submitted. I understand that NSERC will protect this data in accordance with the *Privacy Act*, and that it will only be used in processes that assess the applicant's contributions to the training of highly qualified personnel (HQP), including confidential peer review.

Trainee's signature _____ Date _____

Note: This form must be retained by the applicant and made available to NSERC upon request.