



Interuniversity Attraction Pole P6/03
 Statistical Analysis of Association and Dependence in
 Complex Data

Web: www.stat.ucl.ac.be/IAP/



Interuniversity Graduate School of Psychometrics & Sociometrics
 Faculty of Social and Behavioral Sciences, Leiden University
 P.O. Box 9555, 2300 RB Leiden, the Netherlands
 Phone: + 31 (0)71 527 3829, Fax + 31 (0)71 527 3865
 E-mail: iopssecr@iops.nl
 Web: www.iops.nl

Postdoctoral course *Introduction to numerical techniques for statisticians*

Final registration date: April 21, 2008

Instructor(s):

Geert Molenberghs (UHasselt), Geert Verbeke (K.U.Leuven), Francis Tuerlinckx (K.U.Leuven)

Dates:

Monday April 28 (11:00 am – 17:30 pm) and Tuesday April 29 (9:00 am – 16:00 pm), 2008.

Location:

K.U.Leuven, Huis Van 't Sestich (Room 01.10), Naamsestraat 69, 3000 Leuven, Belgium (see map below)

Short description of the course:

The estimation of the parameters of a statistical model requires the solution of an optimization problem. In a few simple cases, closed-form solutions exist but for many probability models the optimal parameter estimates have to be determined by means of an iterative algorithm. The goal of this course is threefold. First, we want to offer the participants an overview of some frequently used optimization algorithms in (applied) statistics. Second, we want to provide a framework for understanding the connections among several optimization algorithms as well as between optimization and aspects of statistical inference. Third, some important related topics such as numerical differentiation and integration will be covered.

In the first part of the course, we start with the presentation of a few motivating examples from statistics that require the optimization of an objective function. Next, we will define the optimization problem in a formal and abstract way and study the necessary and sufficient conditions for its solution. We also discuss some basic tools for tackling optimization problems from multivariate calculus (gradient, Hessian, Taylor series expansion) and from statistics (exponential family). Then it will be shown that in some cases the optimal parameter estimates can be found directly but in other cases an iterative algorithm is needed.

In the second part of the course we will explain some basic methods for iteration-based function optimization. We start by illustrating methods for the optimization of functions of one variable and then we turn to the optimization of functions of more than one variable. Several algorithms are presented (Newton-Raphson, quasi-Newton, conjugate gradient, steepest descent, downhill simplex) which differ

from each other in the amount of information they use about the surface of the objective function. The advantages and disadvantages of these methods will be illustrated. Because analytical differentiation is not always feasible, we show how numerical differentiation works. An introduction to the field of constrained optimization will be covered too.

In a third part we will apply some of the basic algorithms to the important problem of finding maximum likelihood estimates. In this context, an alternative technique, known as Fisher scoring, is available and its use will be illustrated. Next, we discuss the relationship between the maximization of the (log-)likelihood and statistical inference: asymptotical properties of the MLE, estimates of precision, and testing problems. We also consider the iterative proportional fitting algorithm that can be used in multivariate categorical regression settings. In addition, constrained optimization problems and methods to solve them are discussed.

In the fourth part of the course, we present some optimization problems which contain an integral that has to be evaluated (linear and generalized linear mixed models). In some cases the integral can be solved analytically but in other cases it has to be approximated numerically. An overview of techniques for numerical integration will be given.

In the fifth part, we will introduce the expectation-maximization (EM) algorithm which can be applied for incomplete-data problems (where the incompleteness can be genuine or conceptual) to obtain maximum likelihood estimates. By applying the EM algorithm, traditionally difficult estimation problems can become fairly straightforward. The general theory of the EM algorithm will be covered and it will be shown how precision estimates can be obtained from it and how its rate of convergence can be improved upon.

The course will be closed with a discussion of some issues on the practical implementation of the algorithms and the advantages of combining several algorithms for the same problem.

Method of working:

There will be six lectures of 90 minutes. The lectures will include the explanation of the theory which will be illustrated by examples and annotated output of computer programs.

Language: English

Prerequisites:

Participants should have a basic knowledge of the principles of statistical inference at the level of the first three chapters of Welsh (1996). This includes some familiarity with the concept of a likelihood function and likelihood-based inference for linear, binomial, multinomial, and logistic regression models.

Participants should also have a basic understanding of matrix algebra as explained in Chapter 2 of Johnson and Wichern (1998). A working knowledge of the basic elements of univariate calculus is also a prerequisite (the concepts of continuity of a function, derivative and integration).

References:

Welsh, A.H. (1996). *Aspects of statistical inference*. New York: Wiley.

Johnson, R.A., & Wichern, D.W. (1998). *Applied multivariate statistical analysis* (4th ed.). New York: Prentice-Hall.

Literature (books, readers):

Copies of lecturers' overheads will be distributed during the course (cost is about 5 euros).

Target group:

PhD-students and researchers who wish to learn the basic concepts of numerical techniques, with an emphasis on optimization problems, and their applications in statistics.

Minimum number of participants: 10 (maximum 70)

Certificate: yes

Fees

IOPS PhD students	Free
PhD students and post-docs from IAP P6/03, K.U.Leuven, UHasselt	Free, except for course materials (about 5 euros)
University PhD students (appointed as AiO), IOPS staff members, IOPS alumni	€ 200 and € 25 registration fees; additional costs for course materials (about 5 euros)
Other participants	€ 600 and € 25 registration fees; additional costs for course materials (about 5 euros)

Registration is mandatory!**→ For PhD students and post-docs from IAP P6/03 and K.U.Leuven:**

Véronique Limbourg
 Department of Psychology
 Tiensestraat 102
 3000 Leuven – Belgium
 E-mail: Veronique.Limbourg@psy.kuleuven.be

→ For all other participants:

Lidy van Everdingen (IOPS Secretary)
 Faculty of Social and Behavioral Sciences
 P.O. Box 9555
 2300 RB Leiden - The Netherlands
 E-mail: iopssecr@iops.nl
 Voice: +31 (0)71 527 38 29

Final registration date: April 21, 2008

Cancellation: No refund is made for cancellations within one week before the start of the seminar.

Accommodation (rates are only tentative):

Hotel New Damshire, Schapenstraat 1, 3000 Leuven, €100

phone: +32 16 23 21 15; e-mail: reservations@hotelnewdamshire.be; website:

<http://www.hotelnewdamshire.be/>

Bed & Breakfast De Werf, Hogeschoolplein 5, 3000 Leuven; € 23

phone: +32 476 96 42 41; e-mail: dewerf@chello.be; website: www.dewerf-leuven.be

Gasthof De Pastorij, Sint Michielstraat 5, 3000 Leuven; € 55-60-90

phone: +32 16 82 21 09; e-mail: depastorij@yahoo.com; website: www.depastorij.be

Hotel Ibis, Brusselsestraat 52, 3000 Leuven; € 70

phone: +32 16 29 31 11; e-mail: h1457@accor-hotels.com

Hotel La Royale, Martelarenplein 6, 3000 Leuven; € 57

phone +32 16 22 12 52; e-mail: hotel@laroyale.be; website: www.laroyale.be

Paters Oblaten, Pellenbergstraat 160, 3010 Kessel-Lo; € 15

phone: 32 486 40 54 90

Map