

A Simple Derivation of the Linearization of the Regression Estimator

KEES ZEELENBERG¹

ABSTRACT

We show how the use of matrix calculus can simplify the derivation of the linearization of the regression coefficient estimator and the regression estimator.

KEY WORDS: Matrix calculus; Regression estimator; Taylor expansion.

1. INTRODUCTION

Design-based sampling variances of non-linear statistics are often calculated by means of a linear approximation obtained by a Taylor expansion; examples are the variances of the general regression coefficient estimator and the regression estimator. The linearizations usually need some complicated differentiations. The purpose of this paper is to show how matrix calculus can simplify these derivations, to the extent that even the Taylor expansion of the regression coefficient estimator can be derived in one line, which should be compared with the nearly one page that Särndal *et al.* (1992, p. 205-206) need. To be honest, the use of matrix calculus requires some more machinery to be set up, which is not needed for traditional methods. However this set-up can be regarded as an investment; once it has been learned, it can be used fruitfully in many other applications. After this paper had been written, Binder (1996) appeared, in which similar techniques are used to derive variances by means of linearization. The present paper can be seen as a pedagogical note, in which the use of differentials is exposed.

2. MATRIX DIFFERENTIALS

2.1 Introduction

We will use the matrix calculus by means of differentials, as set out by Magnus and Neudecker (1988); this calculus differs somewhat from the usual methods, which focus on derivatives instead of differentials. Therefore in this section we will briefly describe the definitions and properties of differentials (see Zeelenberg 1993, for a more extensive survey). We first define differentials for vector functions, and then generalize to matrix functions.

2.2 Vector Functions

Let f be a function from an open set $S \subset \mathbb{R}^m$ to \mathbb{R}^n ; let x_0 be a point in S . The function f is *differentiable* at x_0 if there

exists a real $n \times m$ -matrix A , depending on x_0 , such that for any $u \in \mathbb{R}^m$ for which $x_0 + u \in S$, there holds

$$f(x_0 + u) = f(x_0) + A_{x_0} u + o(u), \quad (1)$$

where $o(u)$ is a function such that $\lim_{|u| \rightarrow 0} |o(u)|/|u| = 0$; the matrix A is called the *first derivative* of f at x_0 ; it is denoted as $Df(x_0)$ or $\partial f / \partial (x')|_{x=x_0}$. The derivative Df is equal to the matrix of partial derivatives, i.e., $Df(x)_{ij} = \partial f_i / \partial x_j$. The linear function $df_{x_0}: \mathbb{R}^m \rightarrow \mathbb{R}^n$ defined by $df_{x_0}: u \mapsto A_{x_0} u$ is called the *differential* of f at x_0 . Usually we write dx instead of u so that $df_{x_0}(dx) = A_{x_0} dx$. From (1) we see that the differential corresponds to the linear part of the function, which can also be written as

$$y - y_0 = A_{x_0}(x - x_0),$$

where $y_0 = f(x_0)$. Therefore the differential of a function is the linearization of the function: it is the equation of the hyperplane through the origin that is parallel to the hyperplane tangent to the graph of f at x_0 ; so the linearized function can be written as

$$f(x) \doteq f(x_0) + A_{x_0}(x - x_0). \quad (2)$$

Alternatively, if B is a matrix such that $df_{x_0}(dx) = B dx$, then B is the derivative of f at x_0 and contains the partial derivatives of f at x_0 . This one-to-one relationship between differentials and derivatives is very useful, since differentials are easy to manipulate.

Finally, we usually omit the subscript 0 in x_0 , so that we write $df = A_x dx$.

2.3 Matrix Functions

A matrix function F from an open set $S \subset \mathbb{R}^{m \times n}$ to $\mathbb{R}^{p \times q}$ is differentiable if $\text{vec } F$ is differentiable. The derivative DF is the derivative of $\text{vec } F$ with respect to $\text{vec } X$, and is also denoted by $\partial \text{vec } F / \partial (\text{vec } X)'$. The differential dF is the matrix function defined by $\text{vec } dF_{x_0}(U) = A_{x_0} \text{vec } U$.

¹ Kees Zeelenberg, Department of Statistical Methods, Statistics Netherlands, P.O. Box 4000, 2270 JM Voorburg, The Netherlands.

2.4 Properties of Differentials

Let A be a matrix of constants, F and G differentiable matrix functions, and α a real scalar. Then the following properties are easily proved:

$$dA = 0, \quad (3)$$

$$d(\alpha F) = \alpha dF, \quad (4)$$

$$d(F + G) = dF + dG, \quad (5)$$

$$d(FG) = (dF)G + F(dG), \quad (6)$$

$$dF^{-1} = -F^{-1}(dF)F^{-1}. \quad (7)$$

The last property can be proved by taking the differential of $FF^{-1} = I$ and rearranging.

3. LINEARIZATION OF THE REGRESSION COEFFICIENT ESTIMATOR

The π -estimator (Horvitz-Thompson estimator) of the finite population regression coefficient (cf. Särndal *et al.* 1992, section 5.10) is

$$\hat{B} = \hat{T}^{-1} \hat{t}, \quad (8)$$

where

$$\hat{T} = \sum_{k \in s} \frac{x_k x_k'}{\pi_k},$$

$$\hat{t} = \sum_{k \in s} \frac{x_k y_k}{\pi_k},$$

y_k is the variable of interest for individual k , x_k is the vector with the auxiliary variables for individual k , π_k is the inclusion probability for individual k , and s denotes the sample.

Taking the total differential of (8), using properties (6) and (7), and evaluating at the point where $\hat{T} = T$, $\hat{t} = t$, we get

$$d\hat{B} = -T^{-1}(d\hat{T})T^{-1}t + T^{-1}(d\hat{t}). \quad (9)$$

Because of the connection between differentials and linear approximation, as given in equation (2), it immediately follows that (9) corresponds to the linearization of the regression coefficient estimator:

$$\hat{B} \doteq B - T^{-1}(\hat{T} - T)T^{-1}t + T^{-1}(\hat{t} - t) = B + T^{-1}(\hat{t} - \hat{T}B),$$

where $B = T^{-1}t$.

4. LINEARIZATION OF THE REGRESSION ESTIMATOR

The regression estimator of a population total is (cf. Särndal *et al.* 1992, section 6.6)

$$\hat{t}_{yr} = \hat{t}_{y\pi} + (t_x - \hat{t}_{x\pi})' \hat{B}, \quad (10)$$

where $\hat{t}_{y\pi}$ is the π -estimator of the variable of interest, t_x is the vector with the population totals of the auxiliary variables, $\hat{t}_{x\pi}$ is the vector with the π -estimators of the auxiliary variables, and \hat{B} is the estimator of the regression coefficient of the auxiliary variables on the variable of interest. Taking the total differential of (10), using properties (3) and (6), and evaluating at the point where $\hat{t}_{y\pi} = t_y$, $\hat{t}_{x\pi} = t_x$, and $\hat{B} = B$, we get the linear approximation of the regression estimator

$$d\hat{t}_{yr} = d\hat{t}_{y\pi} - (d\hat{t}_{x\pi})' B,$$

so that

$$\hat{t}_{yr} \doteq t_y + \hat{t}_{y\pi} - t_y + (t_x - \hat{t}_{x\pi})' B = \hat{t}_{y\pi} + (t_x - \hat{t}_{x\pi})' B.$$

Note that for the linearization of the regression estimator we do not need that of the regression coefficient estimator B .

ACKNOWLEDGEMENTS

I wish to thank Jeroen Pannekoek, Jos de Ree, Robbert Renssen, two referees, and an Associate Editor for their comments. The views expressed in this article are those of the author and do not necessarily reflect the policy of Statistics Netherlands.

REFERENCES

- BINDER, D.A. (1996). Linearization methods for single phase and two-phase samples: a cookbook approach. *Survey Methodology*, 22, 17-22.
- MAGNUS, J.R., and NEUDECKER, H. (1988). *Matrix Differential Calculus*. New York: Wiley.
- SÄRNDAL, C.-E., SWENSSON, B., and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.
- ZEELNBERG, C. (1993). *A Survey of Matrix Differentiation*. Research Paper, Department of Statistical Methods, Statistics Netherlands, Voorburg.

CONTENTS

TABLE DES MATIÈRES

Volume 25, No. 1, March/mars 1997

- J.N.K. RAO
Developments in sample survey theory: an appraisal
- T.M. Fred SMITH
Social surveys and social science
- Feifang HU
The asymptotic properties of the maximum relevance weighted likelihood estimators
- R.R. SITTER and J.N.K. RAO
Imputation for missing values and corresponding variance estimation
- Patrick J. FARRELL, Brenda MacGIBBON and Thomas J. TOMBERLIN
Bootstrap adjustments for empirical Bayes interval estimates of small area proportions
- D.A.S. FRASER, N. REID and A. WONG
Simple and accurate inference for the mean of the gamma model
- Jianguo SUN and David E. MATTHEWS
A random-effect regression model for medical follow-up studies
- Philippe CAPÉRAA and Ana Isabel Garralda GUILLEM
Taux de résistance des tests de rang d'indépendance

Volume 25, No. 2, June/juin 1997

- X. Joan HU and Jerald F. LAWLESS
Pseudolikelihood estimation in a class of problems with response-related missing covariates
- Irwin GUTTMAN and George D. PAPANDONATOS
A Bayesian approach to a reliability problem: theory, analysis and interesting numerics
- R.J. OHARA HINES
Fitting generalized linear models to retrospectively sampled clusters with categorical responses
- R.R. SITTER and I. FAINARU
Optimal designs for the logit and probit models for binary data
- Boxin TANG and C.F.J. WU
A method for constructing supersaturated designs and its $E(s^2)$ optimality
- Shu YAMADA and Dennis K.J. LIN
Supersaturated design including an orthogonal base
- A.G. BENN and R.J. KULPBERGER
Integrated marked Poisson processes with application to image correlation spectroscopy
- Khalid El HIMDI and Roch ROY
Tests for the non-correlation of two multivariate ARMA time series
- John J. SPINELLI and Michael A. STEPHENS
Cramér-von Mises tests of fit for the Poisson distribution
- Thomas W. O'GORMAN
An adaptive test for the one-way layout

JOURNAL OF OFFICIAL STATISTICS

An International Review Published by Statistics Sweden

JOS is a scholarly quarterly that specializes in statistical methodology and applications. Survey methodology and other issues pertinent to the production of statistics at national offices and other statistical organizations are emphasized. All manuscripts are rigorously reviewed by independent referees and members of the Editorial Board.

Contents

Volume 12, Number 4, 1996

Derivation and Properties of the X11ARIMA and Census X11 Linear Filters <i>Estela Bee Dagum, Norma Chhab, and Kim Chiu</i>	329
Correcting Unit Nonresponse via Response Modeling and Raking in the California Tobacco Survey <i>Charles C. Berry, Shirley W. Cavin, and John P. Pierce</i>	349
Multiple Workloads per Stratum Designs <i>Lynn Weidmann and Lawrence R. Ernst</i>	365
Neural Network Imputation Applied to the Norwegian 1990 Population Census Data <i>Svein Nordbotten</i>	385
Modeling Income in the U.S. Consumer Expenditure Survey <i>Geoffrey D. Paulin and Elizabeth M. Sweet</i>	403
The Survey Reinterview: Respondent Perceptions and Response Strategies <i>Johnny Blair and Seymour Sudman</i>	421
Corrigendum	427
Book Reviews	429
Editorial Collaborators	441
Index to Volume 12, 1996	445

Volume 13, Number 1, 1997

Who Lives Here? Survey Undercoverage and Household Roster Questions <i>Roger Tourangeau, Gary Shapiro, Anne Kearney, and Lawrence Ernst</i>	1
Suggestive Interviewer Behaviour in Surveys: An Experimental Study <i>Johannes H. Smit, Wil Dijkstra, and Johannes van der Zouwen</i>	19
Effects of Post-Stratification on the Estimates of the Finnish Labour Force Survey <i>Kari Djerf</i>	29
Variance Estimation for Measures of Income Inequality and Polarization - The Estimating Equations Approach <i>Milorad S. Kovačević and David A. Binder</i>	41
Issues in the Use of a Plant-Capture Method for Estimating the Size of the Street Dwelling Population <i>Elizabeth Martin, Eugene Laska, Kim Hopper, Morris Meisner, and Joe Wanderling</i>	59
A Bayesian Approach to Data Disclosure: Optimal Intruder Behavior for Continuous Data <i>Stephen E. Fienberg, Udi E. Makov, and Ashish P. Sanil</i>	75
Book Review	91
In Other Journals	101

Volume 13, Number 2, 1997

Evaluation of a Reconstruction of the Adjusted 1990 Census for Florida <i>Michael M. Meyer and Joseph B. Kadane</i>	103
Individual Diaries and Expense Documents in the Italian Consumer Expenditure Survey <i>Carlo Filippucci and Maria Rosaria Ferrante</i>	113
Testing of Distribution Functions from Complex Sample Surveys <i>Abba M. Krieger and Danny Pfeffermann</i>	123
Estimating Consumer Price Indices for Small Reference Populations <i>Martin Boon and Jan de Haan</i>	143
Cognitive Dynamics of Proxy Responding: The Diverging Perspectives of Actors and Observers <i>Norbert Schwarz and Tracy Wellens</i>	159
Question Difficulty and Respondents' Cognitive Ability: The Effect on Data Quality <i>Bärbel Knäuper, Robert F. Belli, Daniel H. Hill, and A. Regula Herzog</i>	181

All inquiries about submissions and subscriptions should be directed to the Chief Editor:
Lars Lyberg, R&D Department, Statistics Sweden, Box 24 300, S - 104 51 Stockholm, Sweden.