Testing theory

an introduction

Testing theory
an introduction

P.J.G. Teunissen

Delft University of Technology

Department of Mathematical Geodesy and Positioning

VSSD

Series on Mathematical Geodesy and Positioning
http://www.vssd.nl/hlf/a030.htm

Adjustment Theory
P.J.G. Teunissen
2003 / 201 p. / ISBN 90-407-1974-8

Dynamic Data Processing
P.J.G. Teunissen
2001 / 241 + x p. / ISBN 90-407-1976-4

Testing Theory
P.J.G. Teunissen
2000 / 147+viii p. / ISBN 90-407-1975-6

Hydrography
C.D. de Jong, G. Lachapelle, S. Skone, I.A. Elema
2003 / x+351 pp. / ISBN 90-407-2359-1 / hardback

First edition 2000-2006

Published by:
Delft Academic Press / VSSD
Leeghwaterstraat 42, 2628 CA Delft, The Netherlands
tel. +31 15 278 2124, e-mail: dap@vssd.nl
URL about this book: www.delftacademicpress.nl

A collection of digital pictures and an elctronic version can be made available for lecturers who adopt this book. Please send a request by e-mail to hlf@vssd.nl

ISBN-10 90-407-1975-6 ISBN-13 978-90-71301-75-2
NUR 930

Keywords: testing, geodesy

Foreword

This book is based on the lecture notes of the course 'Testing theory' (Inleiding Toetsingstheorie) as it has been offered since 1989 by the Department of Mathematical Geodesy and Positioning (MGP) of the Delft University of Technology. This course is a standard requirement and is given in the second year. The prerequisites are a solid knowledge of adjustment theory together with linear algebra, statistics and calculus at the undergraduate level. The theory and application of least-squares adjustments are treated in the lecture notes *Adjustment theory* (Delft University Press, 2000). The material of the present course is a follow up on this course on adjustment theory. Its main goal is to convey the knowledge necessary to be able to judge and validate the outcome of an adjustment. As in other physical sciences, measurements and models are used in Geodesy to describe (parts of) physical reality. It may happen however, that some of the measurements or some parts of the model are biased or in error. The measurements, for instance, may be corrupted by blunders, or the chosen model may fail to give an adequate enough description of physical reality. These mistakes can and will occasionally happen, despite the fact that every geodesist will try his or her best to avoid making such mistakes. It is therefore of importance to have ways of detecting and identifying such mistakes. It is the material of the present lecture notes that provides the necessary statistical theory and testing procedures for resolving situations like these.

Following the *Introduction*, the basic concepts of statistical testing are presented in *Chapter 1*. In *Chapter 2* the necessary theory is developed for testing *simple* hypotheses. As opposed to its *composite* counterpart, a simple hypothesis is one which is completely specified, both in its functional form as well as in the values of its parameters. Although simple hypotheses rarely occur in geodetic practice, the material of this chapter serves as an introduction to the chapters following. In *Chapter 3*, the generalized likelihood ratio principle is used to develop the theory for testing composite hypotheses. This theory is then worked out in detail in *Chapter 4*, for the important case of linear(ized) models. Both the parametric form (observation equations) and the implicit form (condition equations) of linear models are treated. Five different expressions are given for the uniformly, most powerful, invariant teststatistic. As an additional aid in understanding the basic principles involved, a geometric interpretation is given throughout. This chapter also introduces the important concept of reliability. The internal and external reliability measures given, enable a user to determine in advance (i.e. at the designing stage, before the actual measurements are collected) the size of the minimal detectable biases and the size of their potential impact on the estimated parameters of interest.

Many colleagues of the Department of Mathematical Geodesy and Positioning whose assistance made the completion of this book possible are greatly acknowledged. C.C.J.M. Tiberius took care of the editing, while the typing was done by Mrs. J. van der Bijl and Mrs. M.P.M. Scholtes. The drawings were made by Mr. A.B Smits and the statistical tables were generated by Mrs. M. Roselaar. Various lecturers have taught the book's material over the past years. In particular the feedback and valuable recommendations of G.J. Husti, F. Kenselaar and N.F. Jonkman are acknowledged.

P.J.G. Teunissen
June, 2000

Contents

Introduction

The present lecture notes are a follow up on the book *Adjustment theory* (Delft University Press, 2000). Adjustment theory deals with the optimal combination of redundant measurements together with the estimation of unknown parameters. There are two main reasons for performing redundant measurements. First, the wish to increase the accuracy of the results computed. Second, the requirement to be able to check for mistakes or errors. The present book addresses this second topic.

In order to be able to adjust redundant observations, one first needs to choose a mathematical model. This model consists of two parts, the functional model and the stochastic model. The functional model contains the set of functional relations the observables are assumed to obey. For instance, when the three angles of a triangle are observed and when it is assumed that the laws of planar Euclidean geometry apply, the three angles should add up to π. However, since measurements are intrinsically uncertain (perfect measurements do not exist), one should also take the unavoidable variability of the measurements into account. This is done by means of a stochastic model in which the measurement uncertainty is captured through the use of stochastic (or random) variables. In most geodetic applications it is assumed that the results of measurement, the observations, are independent samples drawn from a normal (or Gaussian) distribution.

Once the mathematical model is specified, one can proceed with the adjustment. Although different methods of adjustment exist, one of the leading principles is the principle of least-squares (for a brief account on the early history of adjustment, see Appendix D). Apart from the fact that (linear) least-squares estimators are relatively easy to compute, they also possess two important properties, namely the property of unbiasedness and the property of minimum variance. In layman terms one could say that least-squares solutions coincide with their target value on the average (property of unbiasedness), while the sum of squares of their unavoidable, individual variations about this target value will be the smallest possible on the average (property of minimum variance). These two properties only hold true, however, under the assumption that the mathematical model is correct. They fail to hold in case the mathematical model is misspecified. Errors or misspecifications in the functional model generally result in least-squares estimators that are biased (off target). Similarly, misspecifications in the stochastic model will generally result in least-squares estimators that are less precise (larger variations).

Although one always will try one's best to avoid making mistakes, they can and will occasionally happen. It is therefore of importance to have ways of *detecting* and *identifying* such mistakes. In this book we will restrict ourselves and concentrate only on developing methods for detecting and identifying errors in the functional model. Hence, throughout this book the stochastic model is assumed to be specified correctly. This restriction is a legitimate one for many geodetic applications. From past experience we know that if modelling errors occur, they usually occur in the functional model and not so much in the stochastic model. Putting the exceptions aside, one is usually quite capable of making a justifiable choice for the stochastic model. Moreover, mistakes made in the functional model usually have more serious consequences for the results computed than errors made in the stochastic modelling.

Mistakes or errors in the functional model can come in many different guises. At this point it is of importance to realize, since every model is a caricature of reality, that every model has its shortcomings. Hence, strictly speaking, every model is already in error to begin with. This shows that the notion of a modelling error or a model misspecification has to be considered with some care. In order to understand this notion, it helps if one accepts that the presence of modelling errors can only be felt in the confrontation between data and model. We therefore speak of a modelling error when the discrepancies between the observations and the model are such that they can not be explained by, or attributed to, the unavoidable measurement uncertainty. Such discrepancies can have many different causes. They could be caused by mistakes made by the observer, or by the fact that defective instruments are used, or by wrong assumptions about the functional relations between the observables. For instance, in case of levelling, it could happen that the observer made a mistake when reading off the leveling rod, or in case of direction measurements, it could happen that the observer accidentally aimed the theodolite at the wrong point. These types of mistakes affect individual observations and are usually referred to as blunders or gross errors. Instead of a few individual observations, whole sets of observations may become affected by errors as well. This happens in case defective instruments are used, or when mistakes are made in formulating the functional relations between the observables. Errors with a common cause that affect whole sets of observations are sometimes referred to as systematic errors.

The goal of this book is to convey the necessary knowledge for judging the validity of the model used. Typical questions that will be addressed are: 'How to check the validity of a model? How to search for certain mistakes or errors? How well can errors be traced? How do undetected errors affect the final results?' As to the detection and identification of errors, the general steps involved are as follows:

(i) One starts with a model which is believed to give an adequate enough description of reality. It is usually the simplest model possible which on the basis of past experience has proven itself in similar situations. Since one will ordinarily assume that the measurements and the modelling are done with the utmost care, one is generally not willing, at this stage, to already make allowances for possible mistakes or errors. This is of course an assumption or an hypothesis. This first model is therefore referred to as the *null hypothesis*.

(ii) Since one can never be sure about the absence of mistakes or errors, it is always wise to check the validity of the null hypothesis once it has been selected. Hence, one would like to be able to *detect* an untrustworthy null hypothesis. This is possible in principle, when redundant measurements are available. From the adjustment of the redundant measurements, (least-squares) residuals can be computed. These residuals are a measure of how well the measurements fit the model of the null hypothesis. Large residuals are often indicative for a poor fit, while smaller residuals tend to correspond with a better fit. These residuals are therefore used as input for deciding whether or not one is willing to accept the null hypothesis.

(iii) Would one decide to reject the null hypothesis, one implicitly states that the measurements do not seem to support the assumption that the model under the null hypothesis gives an adequate enough description of reality. One will therefore have to look for an alternative model or an *alternative hypothesis*. It very seldom happens however, that one knows beforehand which alternative to consider. After all, many different errors could have led to the rejection of the null hypothesis. This implies that in practice, instead of considering a single alternative, usually various alternatives will have to be considered. And since different types of errors may occur in different situations, the choice of these alternatives very much depends on the particular situation at hand.

(iv) Once it has been decided which alternatives to consider, one can commence with the process of *identifying* the most likely alternative. This in fact boils down to a search of the alternative hypothesis which best fits the measurements. Since each alternative hypothesis describes a particular mistake or modelling error, the most likely mistake corresponds with the most likely hypothesis. Once one is confident that the modelling errors have been identified, the last step consists of an *adaptation* of the data and/or model. This implies either a re-measurement of the erroneous data or the inclusion of additional parameters in the model such that the modelling errors are accounted for.

It will be intuitively clear that not all errors can be traced equally well. Some errors are better traceable than others. Apart from being able of executing the above steps for the detection and identification of modelling errors, one would therefore also like to know how well these errors can be traced. This depends on the following factors. It depends on the model used (the null hypothesis), on the type and size of the error (the alternative hypothesis), and on the decision procedure used for accepting or rejecting the null hypothesis. Since these decisions are based on uncertain measurements, their outcomes will be to some degree uncertain as well. As a consequence, two kinds of wrong decisions can be made. One can decide to reject the null hypothesis, while in fact it is true (wrong decision of the 1^{st} kind), or one can decide to accept the null hypothesis, although it is false (wrong decision of the 2^{nd} kind). In the first case, one wrongly believes that a mistake or modelling error has been made. This might then lead to an unnecessary re-measurement of the data. In the second case, one wrongly believes that mistakes or modelling errors are absent. As a consequence, one would then obtain biased adjustment results. These issues and how to cope with them, will also be discussed in this book. Once mastered, they will enable one to formulate guidelines for the *reliable* design of measurement set-ups.

1 Basic concepts of hypothesis testing

1.1 Statistical hypotheses

Many social, technical and scientific problems result in the question whether a particular theory or hypothesis is true or false. In order to answer this question one can try to design an experiment such that its outcome can also be predicted by the postulated theory. After performing the experiment one can then confront the experimental outcome with the theoretically predicted value and on the basis of this comparison try to conclude whether the postulated theory or hypothesis should be rejected. That is, if the outcome of the experiment disagrees with the theoretically predicted value, one could conclude that the postulated theory or hypothesis should be rejected. On the other hand, if the experimental outcome is in agreement with the theoretically predicted value, one could conclude that as yet no evidence is available to reject the postulated theory or hypothesis.

Example 1

According to the postulated theory or hypothesis the three points 1, 2 and 3 of Figure 1.1 lie on one straight line. In order to test or verify this hypothesis we need to design an experiment such that its outcome can be compared with the theoretically predicted value.

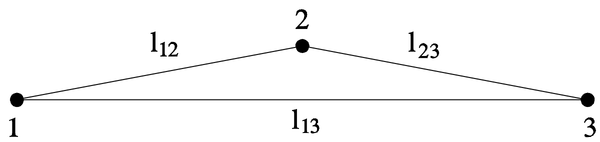

Figure 1.1: Three points on a straight line.

If the postulated hypothesis is correct, the three distances l_{12}, l_{23} and l_{13} should satisfy the relation:

$$l_{13} = l_{12} + l_{23}.$$

Thus, under the assumption that the hypothesis is correct we have:

$$(1) \qquad H: \quad l_{12} + l_{23} - l_{13} = 0.$$

To denote a hypothesis, we will use a capital *H* followed by a colon that in turn is followed by the assertion that specifies the hypothesis. As an experiment we can now measure the three distances l_{12}, l_{23} and l_{13}, compute $l_{12} + l_{23} - l_{13}$ and verify whether this computed value agrees or disagrees with the theoretically predicted value of *H*. If it agrees, we are inclined to accept the hypothesis that the three points lie on one straight line. In case of disagreement we are inclined to reject hypothesis *H*.