Testing theory

An introduction (3rd edition)

Teunissen, P.J.G.

DOI
10.59490/tb.96

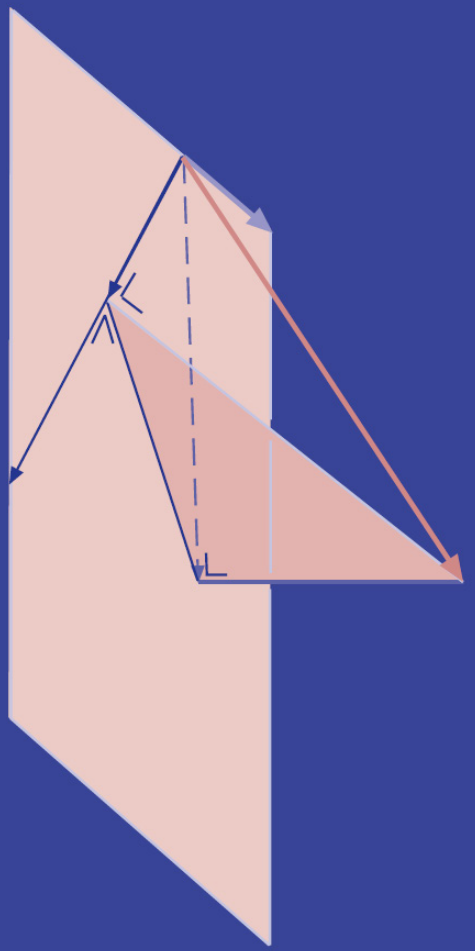
Publication date
2024

Document Version
Final published version

Citation (APA)
Teunissen, P. J. G. (2024). *Testing theory: An introduction (3rd edition)*. (3 ed.) TU Delft OPEN Publishing. https://doi.org/10.59490/tb.96

Testing theory: an introduction

Peter J.G. Teunissen



TU Delft OPEN

Testing theory

an introduction

Testing theory
an introduction

P.J.G. Teunissen

Delft University of Technology

Department of Mathematical Geodesy and Positioning

Series on Mathematical Geodesy and Positioning

ISBN: 978-94-6366-892-7 (Ebook)
ISBN: 978-94-6366-891-0 (Paperback/softback)
Doi: https://doi.org/10.59490/tb.96

Keywords: testing theory, quality control, geodesy, surveying, remote sensing

Preface 2nd Edition

Statistical validation of data and model remains an important topic when dealing with the confrontation of measured data with linked mathematical models. When passed unnoticed, observational errors, outcomes of defective instruments or erroneous assumptions about the underlying models, may seriously deteriorate the final results of any parameter estimation process. Testing theory provides the necessary knowledge for the detection, identification and adaptation of such errors.

The substance of this book has a vast range of exciting applications. As examples we mention the remote sensing and the positioning, navigation and timing (PNT) domains, in which Global Navigation Satellite Systems, such as GPS and Galileo, play a prominent role. In this 2nd edition, we have corrected misprints and other errors, which kindly were brought to our attention by students and lecturers who used the book in their courses.

P.J.G. Teunissen
December, 2006

Preface 3rd Edition

To promote open access, this new edition of Testing Theory is published by TU Delft Open Publishing instead of Delft Academic Press. Appendix D of the 2nd edition, describing the historical context of adjustment theory, has now been logically placed as appendix G in the book *Adjustment Theory* (TU Delft Open Publishing, 2024).

July, 2024
Peter J.G. Teunissen

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* (TU Delft Open Publishing, 2024). 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 G of the book *Adjustment theory*). Apart from the fact that properly weighted (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.

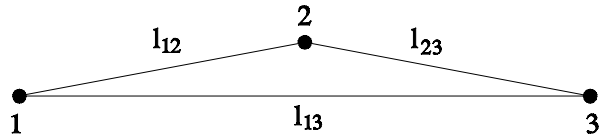


Figure 1.1: Three points on a straight line.

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

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.

It will be clear that in practice the testing of hypotheses is complicated by the fact that experiments (in particular experiments where measurements are involved) in general do not give outcomes that are exact. That is, experimental outcomes are usually affected by an amount of uncertainty, due for instance to measurement errors. In order to take care of this uncertainty, we will, in analogy with our derivation of estimation theory in "Adjustment theory", model the uncertainty by making use of the results from the theory of random variables. The verification or testing of postulated hypotheses will therefore be based on the testing of hypotheses of random variables of which the probability distribution depends on the theory or hypothesis postulated. From now on we will therefore consider statistical hypotheses.

A *statistical hypothesis* is an assertion or conjecture about the probability distribution of one or more random variables, for which it is assumed that a random sample (mostly through measurements) is available.

The structure of a statistical hypothesis H is in general the following:

(2) $$H: \quad \underline{y} \sim p_{\underline{y}}(y|x) \text{ with } x \text{ fully or partially specified.}$$

This statistical hypothesis should be read as follows: According to H the scalar or vector observable random variable $\underline{y}$ has a probability density function given by $p_{\underline{y}}(y|x)$. The scalar, vector or matrix parameter x used in the notation of $p_{\underline{y}}(y|x)$ indicates that the probability density function of $\underline{y}$ is known except for the unknown parameter x. Thus, by specifying (either fully or partially) the parameter x, an assertion or conjecture about the density function of $\underline{y}$ is made. In order to see how a statistical hypothesis for a particular problem can be formulated, let us continue with our Example 1.

Example 1 (continued)

We know from experience that in many cases the uncertainty in geodetic measurements can be adequately modelled by the normal distribution. We therefore model the three distances between the three points 1, 2 and 3 as normally distributed random variables [1]. If we also assume that the three distances are uncorrelated and all have the same known variance $\frac{1}{3}\sigma^2$, the simultaneous probability density function of the three distance observables becomes:

[1] Note that strictly speaking distances can never be normally distributed. A distance is always nonnegative, whereas the normal distribution, due to its infinite tails, admits negative sample values.

$$(3) \qquad \begin{pmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{pmatrix} \sim N\left(\begin{pmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{pmatrix}, Q \right) \text{ with } \quad Q = \tfrac{1}{3}\sigma^2 I_3 .$$

Statement (3) could already be considered a statistical hypothesis, since it has the same structure as (2). Statement (3) asserts that the three distance observables are indeed normally distributed with unknown mean, but with known variancematrix Q. Statement (3) is however not yet the statistical hypothesis we are looking for. What we are looking for is a statistical hypothesis of which the probability density function depends on the theory or hypothesis postulated. For our case this means that we have to incorporate in some way the hypothesis that the three points lie on one straight line. We know mathematically that this assertion implies that:

$$(4) \qquad l_{12} + l_{23} - l_{13} = 0.$$

However, we cannot make this relation hold for the random variables $\underline{l}_{12}$, $\underline{l}_{23}$ and $\underline{l}_{13}$. This is simply because of the fact that random variables cannot be equal to a constant. Thus, a statement like: $\underline{l}_{12} + \underline{l}_{23} - \underline{l}_{13} = 0$ is nonsensical. What we can do is assume that relation (4) holds for the expected values of the random variables $\underline{l}_{12}$, $\underline{l}_{23}$ and $\underline{l}_{13}$:

$$(5) \qquad E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0.$$

For the hypothesis considered this relation makes sense. It can namely be interpreted as stating that if the measurement experiment were to be repeated a great number of times, then on the average the measurements will satisfy (5). With (3) and (5) we can now state our statistical hypothesis as:

$$(6) \qquad H : \begin{pmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{pmatrix} \sim N\left(\begin{pmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{pmatrix}, \tfrac{1}{3}\sigma^2 I_3 \right) \text{ with } \quad E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0.$$

This hypothesis has the same structure of (2) with the three means playing the role of the parameter x.

In many hypothesis-testing problems two hypotheses are discussed: The first, the hypothesis being tested, is called the *null hypothesis* and is denoted by H_0. The second is called the *alternative hypothesis* and is denoted by H_A. The thinking is that if the null hypothesis H_0 is false, then the alternative hypothesis H_A is true, and vice versa. We often say that H_0 is tested against, or versus, H_A. In studying hypotheses it is also convenient to classify them into one of two types by means of the following definition: if a hypothesis completely specifies the distribution, that is, if it specifies its functional form as well as the values of its parameters, it

is called a *simple hypothesis* (enkelvoudige hypothese); otherwise it is called a *composite hypothesis* (samengestelde hypothese).

Example 1 (continued)

In our example (6) is the hypothesis to be tested. Thus, the null hypothesis reads in our case:

$$(7) \qquad H_0 : \begin{pmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{pmatrix} \sim N\left(\begin{pmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{pmatrix}, \tfrac{1}{3}\sigma^2 I_3 \right) \text{ with } \quad E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0.$$

Since we want to find out whether $E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} = 0$ or not, we could take as alternative the inequality $E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} \neq 0$. However, we know from the geometry of our problem that the left hand side of the inequality can never be negative. The alternative should therefore read: $E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} > 0$. Our alternative hypothesis takes therefore the form:

$$(8) \qquad H_A : \begin{pmatrix} \underline{l}_{13} \\ \underline{l}_{12} \\ \underline{l}_{23} \end{pmatrix} \sim N\left(\begin{pmatrix} E\{\underline{l}_{13}\} \\ E\{\underline{l}_{12}\} \\ E\{\underline{l}_{23}\} \end{pmatrix}, \tfrac{1}{3}\sigma^2 I_3 \right) \text{ with } \quad E\{\underline{l}_{12}\} + E\{\underline{l}_{23}\} - E\{\underline{l}_{13}\} > 0.$$

When comparing (7) and (8) we see that the type of the distribution of the observables and their variance matrix are not in question. They are assumed to be known and identical under both H_0 and H_A. Both of the above hypotheses, H_0 and H_A, are examples of composite hypotheses. The above null hypothesis H_0 would become a simple hypothesis if the individual expectations of the observables were assumed known.

1.2 Test of statistical hypotheses

After the statistical hypotheses H_0 and H_A have been formulated, one would like to test them in order to find out whether H_0 should be rejected or not.

A *test* of a statistical hypothesis:

$$H_0 : \underline{y} \sim p_{\underline{y}}(y|x) \text{ with } x \text{ fully or partially specified}$$

is a rule or procedure, in which a random sample of $\underline{y}$ is used for deciding whether to reject or not reject H_0. A test of a statistical hypothesis is completely specified by the so-called *critical region* (kritiek gebied), which will be denoted by K.

The *critical region K* of a test is the set of sample values of $\underline{y}$ for which H_0 is to be rejected. Thus, H_0 is rejected if $y \in K$.

It will be obvious that we would like to choose a critical region so as to obtain a test with desirable properties, that is, a test that is "best" in a certain sense. Criteria for comparing tests and the theory for obtaining "best" tests will be developed in the next and following sections. But let us first have a look at a simple testing problem for which, on more or less intuitive grounds, an acceptable critical region can be found.

Example 2

Let us assume that a geodesist measures a scalar variable, and that this measurement can be modelled as a random variable $\underline{y}$ with density function:

$$\underline{y} \sim \frac{1}{\sqrt{2\pi}} \exp[-\tfrac{1}{2}(y - E\{\underline{y}\})^2]. \tag{9}$$

Thus, it is assumed that $\underline{y}$ has a normal distribution with unit variance. Although this assumption constitutes a statistical hypothesis, it will not be tested here because the geodesist is quite certain of the validity of this assumption. The geodesist is however not certain about the value of the expectation of $\underline{y}$. His assumption is that the value of $E\{\underline{y}\}$ is x_0. This assumption is the statistical hypothesis to be tested. Denote this hypothesis by H_0. Then:

$$H_0 : \quad E\{\underline{y}\} = x_0. \tag{10}$$

Let H_A denote the alternative hypothesis that $E\{\underline{y}\} \neq x_0$. Then:

$$H_A : \quad E\{\underline{y}\} \neq x_0. \tag{11}$$

Thus the problem is one of testing the simple hypothesis H_0 against the composite hypothesis H_A. To test H_0, a single observation on the random variable $\underline{y}$ is made. In real-life problems one usually takes several observations, but to avoid complicating the discussion at this stage only one observation is taken here. On the basis of the value of $\underline{y}$ obtained, denoted by y, a decision will be made either to accept H_0 or reject it. The latter decision, of course, is equivalent to accepting H_A. The problem then is to determine what values of $\underline{y}$ should be selected for accepting H_0 and what values for rejecting H_0. If a choice has been made of the values of $\underline{y}$ that will correspond to rejection, then the remaining values of $\underline{y}$ will necessarily correspond to acceptance. As defined above, the rejection values of $\underline{y}$ constitute the critical region K of the test. Figure 1.2 shows the distribution of $\underline{y}$ under H_0 and under two possible alternatives H_{A_1} and H_{A_2}.

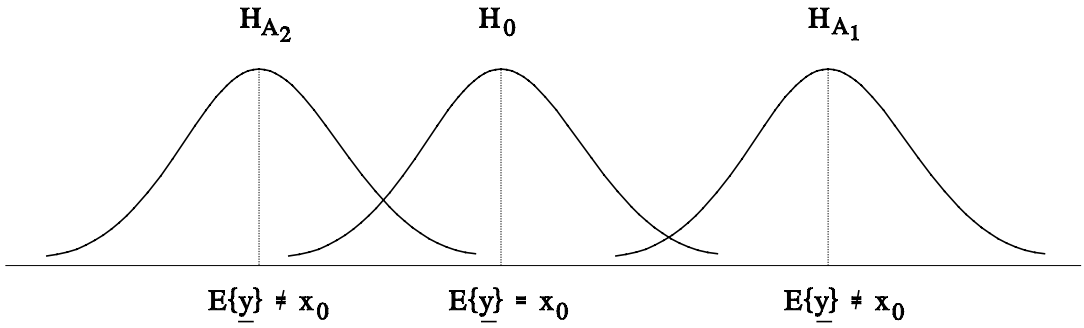


Figure 1.2: $H_0 : E\{\underline{y}\} = x_0$ versus $H_A : E\{\underline{y}\} \neq x_0$.

Looking at this figure, it seems reasonable to reject H_0 if the observation y is remote enough from $E\{\underline{y}\} = x_0$. If H_0 is true, the probability of a sample of $\underline{y}$ falling in a region remote from $E\{\underline{y}\} = x_0$ is namely small. And if H_A is true, this probability may be large. Thus the critical region K should contain those sample values of $\underline{y}$ that are remote enough from $E\{\underline{y}\} = x_0$. Also, since the alternative hypothesis can be located on either side of $E\{\underline{y}\} = x_0$, it seems obvious to have one portion of K located in the left tail of H_0 and one portion of K located in the right tail of H_0. Finally, one can argue that since the distribution is symmetric about its mean value, also the critical region K should be symmetric about $E\{\underline{y}\} = x_0$. This as a result gives the form of the critical region K as shown in Figure 1.3. Although this critical region has been found on more or less intuitive grounds, it can be shown that it possesses some desirable properties. We will return to this matter in a later section.

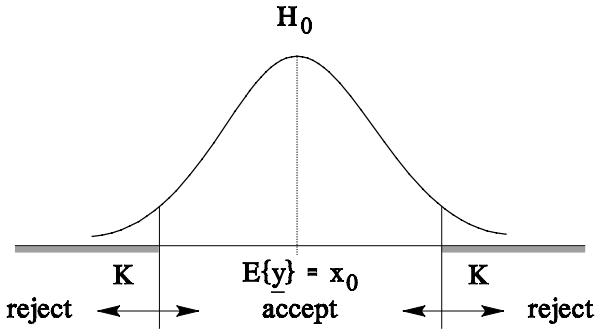


Figure 1.3: Critical region K for testing $H_0 : E\{\underline{y}\} = x_0$ versus $H_A : E\{\underline{y}\} \neq x_0$.

1.3 Two types of errors

We have seen that a test of a statistical hypothesis is completely specified once the critical region K of the test is given. The null hypothesis H_0 is rejected if the sample value or observation of $\underline{y}$ falls in the critical region, i.e. if $y \in K$. Otherwise the null hypothesis H_0 is accepted, i.e. if $y \notin K$. With this kind of thinking two types of errors can be made:

Type I error: Rejection of H_0 when in fact H_0 is true.

Type II error: Acceptance of H_0 when in fact H_0 is false.

Table 1.1 shows the decision table with the type I and II errors.

	H_0 true	H_0 false
Reject H_0 $y \in K$	Wrong Type I error	Correct
Accept H_0 $y \notin K$	Correct	Wrong Type II error

Table 1.1: Decision table with type I and type II error.

The *size* of a type I error is defined as the probability that a sample value of $\underline{y}$ falls in the critical region when in fact H_0 is true. This probability is denoted by α and is called the size of the test or the *level of significance* of the test (onbetrouwbaarheid van de test). Thus:

$$\alpha = P(\text{type I error}) = P(\text{rejection of } H_0 \text{ when } H_0 \text{ true})$$

or

$$\alpha = P(\underline{y} \in K | H_0) = \int_K p_{\underline{y}}(y | H_0) \mathrm{d}y \quad . \tag{12}$$

The size of the test, α, can be computed once the critical region K and the probability density function of $\underline{y}$ is known under H_0. The size of a type II error is defined as the probability that a sample value of $\underline{y}$ falls outside the critical region when in fact H_0 is false. This probability is denoted by β. Thus:

$$\beta = P(\text{type II error}) = P(\text{acceptance of } H_0 \text{ when } H_0 \text{ is false})$$

or

$$\beta = P(\underline{y} \notin K | H_A) = 1 - \int_K p_{\underline{y}}(y | H_A) \mathrm{d}y \quad . \tag{13}$$

The size of a type II error, β, can be computed once the critical region K and the probability density function of $\underline{y}$ is known under H_A.

Example 3

Assume that $\underline{y}$ is distributed as:

$$\underline{y} \sim N(E\{\underline{y}\}, \sigma^2) \tag{14}$$

with known variance σ^2.

The following two simple hypotheses are considered:

(15) $$H_0 : \; E\{\underline{y}\} = x_0$$

and

(16) $$H_A : \; E\{\underline{y}\} = x_A > x_0.$$

The situation is sketched in Figure 1.4.

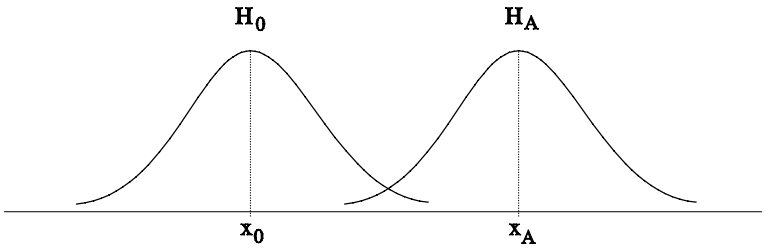


Figure 1.4: The two simple hypotheses: $H_0 : E\{\underline{y}\} = x_0$ and $H_A : E\{\underline{y}\} = x_A > x_0$.

Since the alternative hypothesis H_A is located on the right of the null hypothesis H_0, it seems intuitively appealing to choose the critical region K right-sided. Figure 1.5a and 1.5b show two possible right-sided critical regions K.

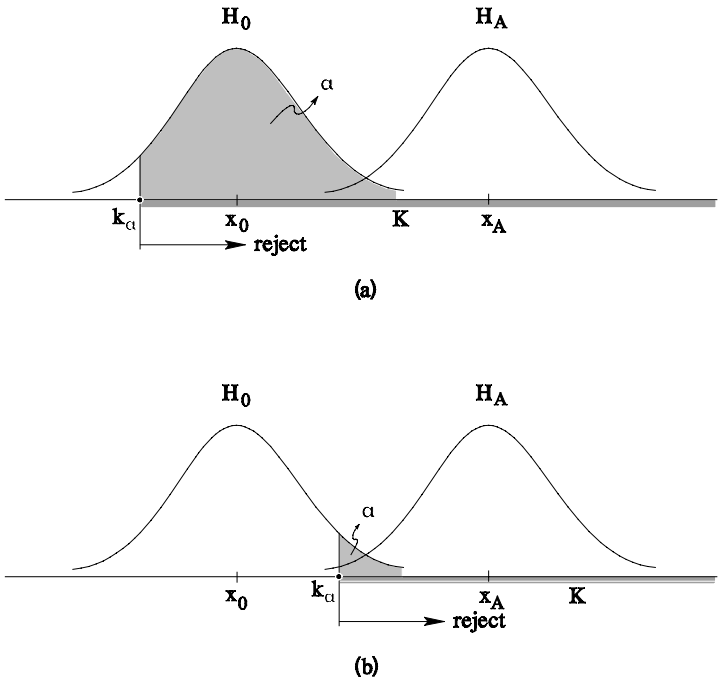


Figure 1.5: Critical region K and size of test, α.

They also show the size of the test, α, which corresponds to the area under the graph of the distribution of $\underline{y}$ under H_0 for the interval of the critical region K.

The size of the test, α, can be computed once the probability density function of $\underline{y}$ under H_0 is known and the form and location of the critical region K is known. In the present example the form of the critical region has been chosen right-sided. Its location is determined by the value of k_α, the so-called *critical value* (kritieke waarde) of the test. Thus, for the present example the size of the test can be computed as:

$$\alpha = \int_{k_\alpha}^{\infty} p_{\underline{y}}(y|x_0)\mathrm{d}y$$

or, since:

$$p_{\underline{y}}(y|x_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_0)^2]$$

as:

$$\alpha = \int_{k_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_0)^2]\mathrm{d}y. \tag{17}$$

When one is dealing with one-dimensional normally distributed random variables, one can usually compute the size of the test, α, from tables given for the standard normal distribution (see appendix B). In order to compute (17) with the help of such a table, we first have to apply a transformation of variables. Since $\underline{y}$ is normally distributed under H_0 with mean x_0 and variance σ^2, it follows that the random variable $\underline{z}$, defined as:

$$\underline{z} = \frac{\underline{y}-x_0}{\sigma} \tag{18}$$

is standard normally distributed under H_0. And since:

$$\alpha = P(\underline{y} > k_\alpha|H_0) = P(\underline{z} > \frac{k_\alpha - x_0}{\sigma}|H_0) \tag{19}$$

we can use the last expression of (19) for computing α. Application of the change of variables (18) to (17) gives:

$$\alpha = \int_{\frac{k_\alpha - x_0}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp[-\tfrac{1}{2}z^2]\mathrm{d}z \ . \tag{20}$$

We can now make use of the table of the standard normal distribution. Table 1.2 shows some typical values of the α and k_α for the case that $x_0 = 1$ and $\sigma = 2$.

α	$\frac{k_\alpha - x_0}{\sigma}$	k_α
0.1	1.28	3.56
0.05	1.65	4.29
0.01	2.33	5.65
0.001	3.09	7.18

Table 1.2: Test size α, critical value k_α for $x_0 = 1$ and $\sigma = 2$.

As we have seen the location of the critical region K is determined by the value chosen for k_α, the critical value of the test. But what value should we choose for k_α? Here the geodesist should base his judgement on his experience. Usually one first makes a choice for the size of the test, α, and then by using (20) or Table 1.2 determines the corresponding critical value k_α. For instance, if one fixes α at $\alpha = 0.01$, the corresponding critical value k_α (for the present example with $x_0 = 1$ and $\sigma = 2$) reads $k_\alpha = 5.65$. The choice of α is based on the probability of a type I error one is willing to accept. For instance, if one chooses α as $\alpha = 0.01$, one is willing to accept that 1 out of a 100 experiments leads to rejection of H_0 when in fact H_0 is true.

Let us now consider the size of a type II error, β. Figure 1.6 shows for the present example the size of a type II error, β. It corresponds to the area under the graph of the distribution of $\underline{y}$ under H_A for the interval complementary to the critical region K.

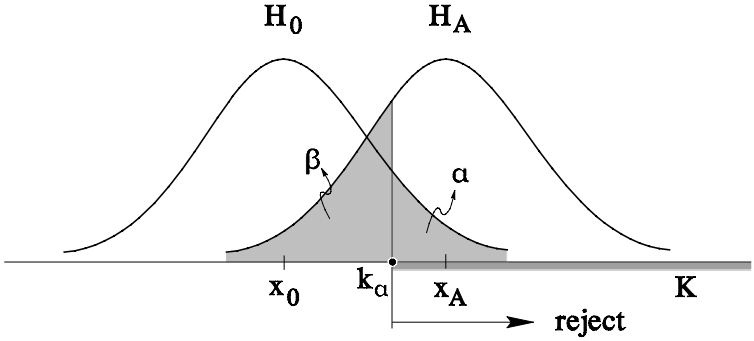


Figure 1.6: The sizes of type I and type II error, α and β, for testing H_0: $E\{\underline{y}\} = x_0$ versus H_A: $E\{\underline{y}\} = x_A > x_0$.

The size of a type II error, β, can be computed once the probability density function of $\underline{y}$ under H_A is known and the critical region K is known. Thus, for the present example the size of the type II error can be computed as:

$$\beta = \int_{-\infty}^{k_\alpha} p_{\underline{y}}(y|x_A)\mathrm{d}y$$

or since:

$$p_{\underline{y}}(y|x_A) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]$$

as:

$$\beta = \int_{-\infty}^{k_\alpha} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]\,dy. \tag{21}$$

Also this value can be computed with the help of the table of the standard normal distribution. But first some transformations are needed. It will be clear that the probability that a sample or observation of $\underline{y}$ falls in the critical region K when H_A is true, is identical to 1 minus the probability that the sample does not fall in the critical region when H_A is true. Thus:

$$\beta = P(\underline{y} \notin K|H_A) = 1 - P(\underline{y} \in K|H_A). \tag{22}$$

Since for the present example:

$$P(\underline{y} \in K|H_A) = \int_{k_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]\,dy$$

substitution into (22) gives:

$$1 - \beta = \int_{k_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_A)^2]\,dy. \tag{23}$$

This formula has the same structure as (17). The value $1-\beta$ can therefore be computed in exactly the same manner as the size of the test, α, was computed. And from $1-\beta$ it is trivial to compute β, the size of the type II error.

Figure 1.7 gives the probability $1-\beta$ of rejecting H_0, when indeed H_A is true, as function of the unknown mean x_A under H_A. When this probability is requested to be at least $1-\beta = 0.80$, the unknown mean under H_A has to be at least x_A=7.34. We return to the probability $\gamma = 1-\beta$, the power, in Section 4.5 on reliability. The size of the test was fixed to $\alpha = 0.01$.

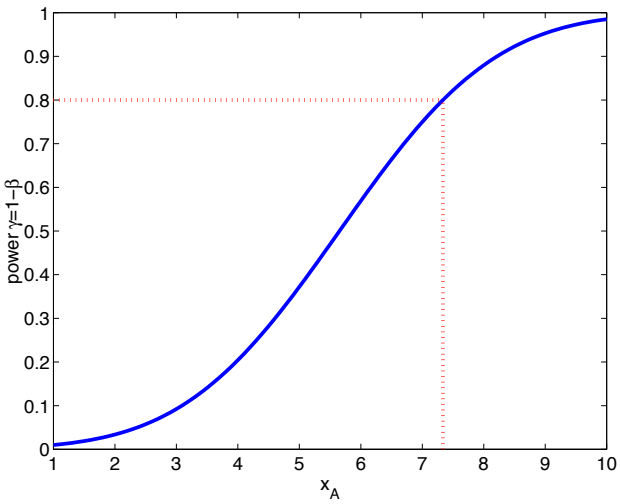


Fig. 1.7: Probability $\gamma = 1-\beta$ as function of x_A, for testing H_0: $E\{\underline{y}\} = x_0$ versus H_A: $E\{\underline{y}\} = x_A > x_0$, with $x_0 = 1$ and $\sigma = 2$.

1.4 A testing principle

We have seen that two types of errors are involved when testing a null hypothesis H_0 against an alternative hypothesis H_A: (1) The rejection of H_0 when in fact H_0 is true (type I error); (2) the acceptance of H_0 when in fact H_0 is false (type II error). One might reasonably use the sizes of the two types of errors, α and β, to set up criteria for defining a best test. If this is possible, it would automatically give us a method of choosing a critical region K. A good test should be a test for which α is small (ideally 0) and β is small (ideally 0). It would therefore be nice if we could define a test, i.e. define a critical region K, that simultaneously minimizes both α and β. Unfortunately this is not possible. As we decrease α, we tend to increase β, and vice versa. The *Neyman-Pearson* principle provides a workable solution to this situation. This principle says that we should fix the size of the type I error, α, and minimize the size of the type II error, β. Thus:

A testing principle (Neyman et al., 1933): Among all tests or critical regions possessing the same size type I error, α, choose one for which the size of the type II error, β, is as small as possible.

The justification for fixing the size of the type I error to be α, (usually small and often taken as 0.05 or 0.01) seems to arise from those testing situations where the two hypotheses, H_0 and H_A, are formulated in such a way that one type of error is more serious than the other. The hypotheses are stated so that the type I error is the more serious, and hence one wants to be certain that it is small. Testing principles other than the above given one can of course easily be suggested: for example, minimizing the sum of sizes of the two types of error, $\alpha + \beta$. However, the Neyman-Pearson principle has proved to be very useful in practice. In this book we will therefore base our method of finding tests on this principle. Now let us consider a testing problem from the point of view of the Neyman-Pearson principle.

Example 4

Assume that $\underline{y}$ has the following probability density function:

(24) $$p_{\underline{y}}(y|x) = xe^{-yx} \ , \ x > 0 \ , \ y \geq 0 \ .^2$$

The following two simple hypotheses are considered:

(25) $$H_0 : x = 2 \qquad \text{and} \qquad H_A : x = 1 \, .$$

Figure 1.8 shows the density function of $\underline{y}$ under H_0 and H_A.

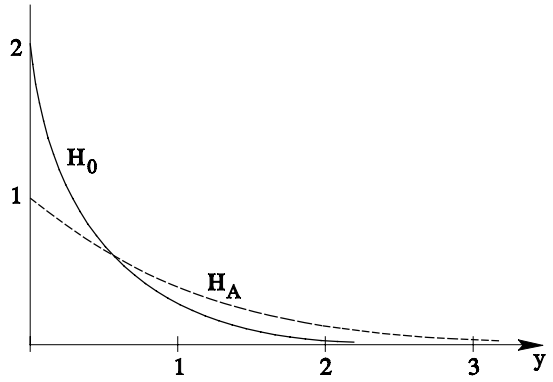


Figure 1.8: The function xe^{-yx}, $x>0$, $y\geq 0$ for $x = 2$ and $x = 1$.

Contrary to our Example 3, it is now not that obvious how to choose the form of the critical region K. Let us first consider the case of a right-sided critical region K. Thus:

(26) $$K = \{y \in \mathbb{R}_0^+ \mid y > k_\alpha\} \, .$$

In order to compute α and β we need to evaluate an integral of the type:

(27) $$\int_a^b xe^{-yx} dy = e^{-ax} - e^{-bx} .$$

For the right-sided critical region (26) this gives for the size of the type I error:

(28) $$\alpha = \int_{k_\alpha}^{\infty} 2e^{-y2} dy = e^{-2k_\alpha} .$$

The corresponding size of the type II error is:

---

[2] Prove yourself that this function is indeed a probability density function.

(29) $$\beta = 1 - \int_{k_\alpha}^{\infty} 1e^{-y1}dy = 1 - e^{-k_\alpha}.$$

Now let us consider a left-sided critical region K^* as alternative. Thus:

(30) $$K^* = \{y \in \mathbb{R}_0^+ \mid 0 \leq y < k_\alpha^*\}.$$

For this critical region the size of the type I error becomes:

(31) $$\alpha^* = \int_0^{k_\alpha^*} 2e^{-y2}dy = 1 - e^{-2k_\alpha^*}.$$

And the corresponding size of the type II error is given by:

(32) $$\beta^* = 1 - \int_0^{k_\alpha^*} 1e^{-y1}dy = e^{-k_\alpha^*}.$$

Let us now compare the two tests, that is, the one with the right-sided critical region K with the one with the left-sided critical region K^*. We will base this comparison on the Neyman-Pearson principle. According to this principle, both tests have the same size of type I error. Thus:

(33) $$\alpha = \alpha^*.$$

With (28) and (31) this gives $e^{-2k_\alpha^*} = 1 - e^{-2k_\alpha}$ or:

(34) $$(e^{-k_\alpha^*})^2 = (1 - e^{-k_\alpha})(1 + e^{-k_\alpha}).$$

Using (29) and (32) this equation can be expressed in terms of β^* and β as:

$$(\beta^*)^2 = \beta(2 - \beta).$$

Hence:

(35) $$\beta^* = (2\beta - \beta^2)^{\frac{1}{2}}.$$

Figure 1.9 shows the graph of this function. It clearly shows that:

(36) $$\beta < \beta^*.$$

The conclusion reads therefore that of the two tests the one having the right-sided critical region K is the best in the sense of the Neyman-Pearson principle.

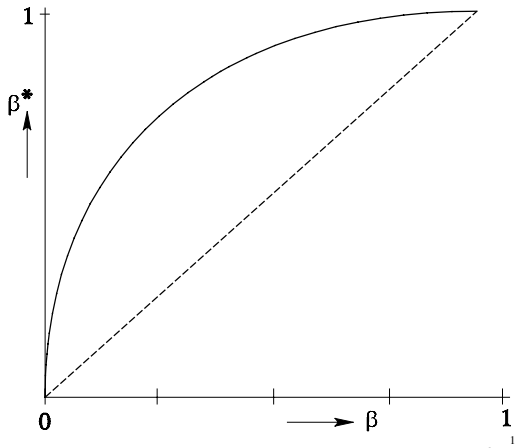


Figure 1.9: The function $\beta^* = (2\beta - \beta^2)^{\frac{1}{2}}$.

1.5 General steps in testing hypotheses

Thus far we have discussed the basic concepts underlying most of the hypothesis-testing problems. The same concept and guidelines will provide the basis for solving more complicated hypothesis-testing problems as treated in the next chapters. Here we summarize the main steps on testing hypotheses about a general probability model.

(a) From the nature of the experimental data and the consideration of the assertions that are to be examined, identify the appropriate null hypothesis and alternative hypothesis:

$$H_0\colon \underline{y} \sim p_{\underline{y}}(y|x_0) \quad \text{versus} \quad H_A\colon \underline{y} \sim p_{\underline{y}}(y|x_A)\,.$$

(b) Choose the form of the critical region K that is likely to give the best test. Use the Neyman-Pearson principle to make this choice.

(c) Specify the size of the type I error, α, that one wishes to assign to the testing process. Use tables to determine the location of the critical region K from:

$$\alpha = P(\underline{y} \in K|H_0) = \int_K p_{\underline{y}}(y|x_0)\,dy.$$

(d) Compute the size of the type II error:

$$\beta = P(\underline{y} \notin K|H_A) = 1 - \int_K p_{\underline{y}}(y|x_A)dy$$

to ensure that there exists a reasonable protection against type II errors.

(e) After the test has been explicitly formulated, determine whether the sample or observation y of $\underline{y}$ falls in the critical region K or not. Reject H_0 if $y \in K$, and accept H_0 if $y \notin K$. Never claim however that the hypotheses have been proved false or true by the testing.

2 Testing of simple hypotheses

2.1 The simple likelihood ratio test

In this chapter we consider testing a simple null hypothesis H_0 against a simple alternative hypothesis H_A. This case is actually not very useful in practical applications, but it will serve the purpose of developing some theory of testing hypotheses. We will assume that the $m \times 1$ vector random variable $\underline{y}$ is distributed as:

$$\underset{m\times 1}{\underline{y}} \sim p_{\underline{y}}(y|x). \tag{1}$$

The following two simple hypotheses are considered:

$$H_0: \quad x = x_0 \quad \text{versus} \quad H_A: \quad x = x_A. \tag{2}$$

Our objective is, given an observation y on $\underline{y}$, to determine from which distribution the observation came from; from $p_{\underline{y}}(y|x_0)$ or from $p_{\underline{y}}(y|x_A)$? In this section we will give a general method for solving this testing problem. The method is closely related to the *maximum likelihood principle* as discussed in Adjustment theory.

For a fixed value of x the function $p_{\underline{y}}(y|x)$ is a function of y and for different values of x the function $p_{\underline{y}}(y|x)$ may take different forms (see Figure 2.1).

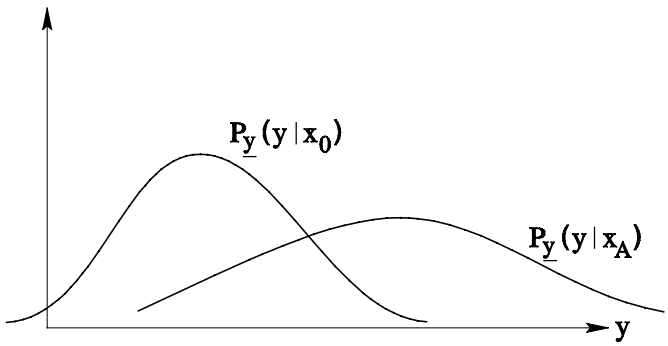


Figure 2.1: The density function of $\underline{y}$ under H_0 and H_A.

In the context of estimation theory the objective was to determine or estimate the unknown parameter x on the basis of the observation vector y. In the present context of hypothesis testing, the objective is to decide between H_0 and H_A. In both cases, that is, in the case of estimation theory and in the case of hypothesis testing, one could say that one would like to determine the correct value of the parameter x that produced the observed y. This suggests considering for each possible x how probable the observed y would be if x were the true value. The higher this probability, the more one is attracted to the explanation that the x in question produced y, and the more likely the value of x appears. In estimation theory, where no constraints were put on x, this principle resulted in the maximum likelihood method. This method chooses as an estimate of x that value which maximizes $p_{\underline{y}}(y|x)$ for the given observed y. For the problem of testing the two simple hypotheses H_0 and H_A we can now apply the same

principle. But instead of maximizing $p_{\underline{y}}(y|x)$ as function of x, we only need to compare the two likelihood values of $p_{\underline{y}}(y|x_0)$ and $p_{\underline{y}}(y|x_A)$. We decide that the observation y came from H_0 if $p_{\underline{y}}(y|x_0) > p_{\underline{y}}(y|x_A)$ and, conversely, decide that the observation y came from H_A if $p_{\underline{y}}(y|x_0) < p_{\underline{y}}(y|x_A)$. This simple method of obtaining a test for testing H_0 against H_A can be expanded into a family of tests that, as we will see, will contain some good tests.

The simple likelihood ratio test is defined by:

(3)

$$\text{reject } H_0 \text{ if } \frac{p_{\underline{y}}(y|x_0)}{p_{\underline{y}}(y|x_A)} < a$$

$$\text{accept } H_0 \text{ if } \frac{p_{\underline{y}}(y|x_0)}{p_{\underline{y}}(y|x_A)} > a$$

where a is a positive constant.

For each different value of a we have different tests. For a fixed value of a the test says to reject H_0 if the ratio of likelihoods is small; that is, reject H_0 if it is more likely that the observation came from $p_{\underline{y}}(y|x_A)$ than from $p_{\underline{y}}(y|x_0)$. Let us consider some examples to see how the simple likelihood ratio test works.

Example 1

We assume that the $m\times 1$ random vector $\underline{y}$ is normally distributed as:

(4)
$$\underline{y} \sim N(0, \sigma^2 I_m).$$

The two simple hypotheses considered are:

(5)
$$H_0: \quad \sigma^2 = \sigma_0^2 \quad \text{versus} \quad H_A: \quad \sigma^2 = \sigma_A^2 > \sigma_0^2.$$

Figure 2.2 shows the distribution of $\underline{y}$ under H_0 and H_A for $m = 1$. For $m = 1$, it seems intuitively appealing to reject H_0 if the observation y is remote from the zero-mean value. Due to the symmetry of the distribution of $\underline{y}$, it also seems intuitively appealing to choose the critical region K symmetric about 0. Thus, based on these two intuitive arguments we would choose to reject H_0 if (see Figure 2.2):

(6)
$$y^2 > k_\alpha.$$

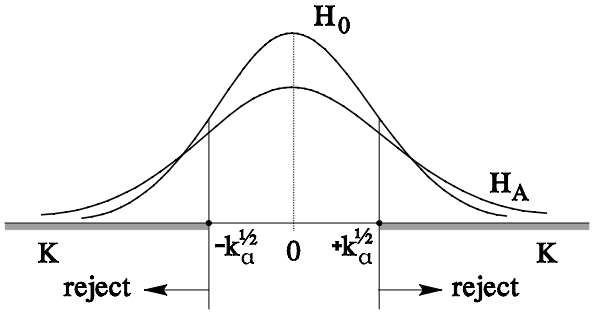


Figure 2.2: The distributions $N(0,\sigma_0^2)$ and $N(0,\sigma_A^2)$, $\sigma_A^2 > \sigma_0^2$.

Figure 2.3 shows the contourlines of equal density of the distribution of $\underline{y}$ under H_0 and H_A for $m>1$. As a generalization of (6), it seems in this case intuitively appealing to reject H_0 if (see Figure 2.3):

(7) $$y^*y = \sum_{i=1}^{m} y_i^2 > k_\alpha .$$

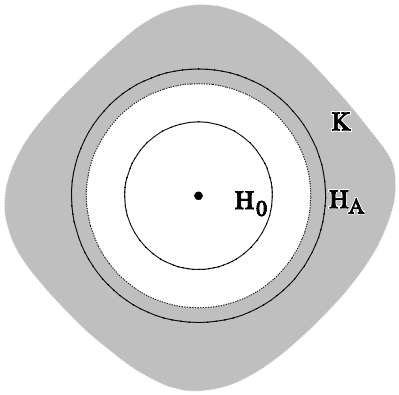


Figure 2.3: Contourlines of equal density of $N(0,\sigma_0^2 I_m)$ and $N(0,\sigma_A^2 I_m)$, with $\sigma_A^2 > \sigma_0^2$.

Now let us apply the simple likelihood ratio test for this particular example, and see how it compares with (6) and (7) respectively. With:

$$p_{\underline{y}}(y|x_0) = (2\pi)^{-m/2}(\sigma_0^2)^{-m/2} \exp[-\tfrac{1}{2}\tfrac{1}{\sigma_0^2}y^*y]$$

and

$$p_{\underline{y}}(y|x_A) = (2\pi)^{-m/2}(\sigma_A^2)^{-m/2} \exp[-\tfrac{1}{2}\tfrac{1}{\sigma_A^2}y^*y]$$

it follows that:

(8) $$\frac{p_{\underline{y}}(y|x_0)}{p_{\underline{y}}(y|x_A)} = \left(\frac{\sigma_A}{\sigma_0}\right)^m \exp[-\tfrac{1}{2}y^*y\ (\tfrac{1}{\sigma_0^2} - \tfrac{1}{\sigma_A^2})].$$

From (8) and (3) we get:

(9) $$\text{reject } H_0 \text{ if:} \quad \left(\frac{\sigma_A}{\sigma_0}\right)^m \exp[-\tfrac{1}{2}y^*y\ (\tfrac{1}{\sigma_0^2}-\tfrac{1}{\sigma_A^2})] < a.$$

In order to compare (9) with (7), we first transform (9) into a simpler inequality. The inequality of (9) can also be written as:

$$\exp[-\tfrac{1}{2}y^*y\ (\tfrac{1}{\sigma_0^2}-\tfrac{1}{\sigma_A^2})] < \left(\frac{\sigma_0}{\sigma_A}\right)^m a.$$

Taking the logarithm gives:

$$-\tfrac{1}{2}y^*y\ (\tfrac{1}{\sigma_0^2}-\tfrac{1}{\sigma_A^2}) < \ln[\left(\frac{\sigma_0}{\sigma_A}\right)^m a].$$

Since $\sigma_A^2 > \sigma_0^2$, division by $\frac{1}{\sigma_0^2}-\frac{1}{\sigma_A^2}$ gives:

$$-\tfrac{1}{2}y^*y < \ln[\left(\frac{\sigma_0}{\sigma_A}\right)^m a] \Big/ \left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_A^2}\right).$$

Finally, multiplication with -2 gives:

(10) $$y^*y > \ln[\left(\frac{\sigma_0}{\sigma_A}\right)^m a]^{-2} \Big/ \left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_A^2}\right).$$

If we denote the right-hand side of this inequality by k_α, we see that the simple likelihood ratio test gives a critical region K which is identical to the one chosen earlier (see (7)) on intuitive grounds. Thus, for this particular example the simple likelihood ratio test is:

(11) $$\text{reject } H_0 \text{ if} \quad y^*y > k_\alpha.$$

In order to perform or execute this test, we still need to choose a particular value for the critical value k_α. The critical value k_α can be computed once the size of the type I error, α, has been fixed, and once the distribution of $\underline{y}^*\underline{y}$ is known under H_0. Since $\underline{y}$ is distributed as $N(0,\sigma_0^2 I_m)$ under H_0, it follows (see appendix A) that $\underline{y}^*\underline{y}$ is distributed under H_0 as a central $\sigma_0^2\chi^2$-distribution with m degrees of freedom. In this case, there are no unknown parameters, $n = 0$ and hence $m - n = m$. Thus:

(12) $$H_0: \quad \underline{y}^*\underline{y} \sim \sigma_0^2\,\chi^2(m,0).$$

Similarly we have for the distribution of $\underline{y}^*\underline{y}$ under H_A:

(13) $$H_A: \quad \underline{y}^*\underline{y} \sim \sigma_A^2\,\chi^2(m,0).$$

Since:

(14) $$\alpha = P(\underline{y}^*\underline{y} > k_\alpha | H_0) = P\left(\frac{\underline{y}^*\underline{y}}{\sigma_0^2} > \frac{k_\alpha}{\sigma_0^2} | H_0\right)$$

and since $\underline{y}^*\underline{y}/\sigma_0^2$ is distributed as $\chi^2(m,0)$ under H_0, we can use a table of the χ^2-distribution (see appendix B) to compute the critical value k_α from the chosen size of type I error, α. Table 2.1 shows some typical values of α and k_α for the case $\sigma_0^2 = 2$ and $m = 1$ (on the left).

α	$m = 1$		$m = 4$	
	$\frac{k_\alpha}{\sigma_0^2} = \frac{1}{2}k_\alpha$	k_α	$\frac{1}{2}k_\alpha$	k_α
0.1	2.71	5.41	7.78	15.56
0.05	3.84	7.68	9.49	18.98
0.01	6.63	13.27	13.28	26.55
0.001	10.83	21.66	18.47	36.93

Table 2.1: α and k_α for the distribution $\sigma_0^2\chi(m,0)$ with $\sigma_0^2 = 2$, $m = 1$ and $m = 4$.

From k_α and the distribution of $\underline{y}^*\underline{y}$ under H_A, we can also compute the size of the type II error, β. Since:

$$\beta = P(\underline{y}^*\underline{y} < k_\alpha | H_A) = 1 - P(\underline{y}^*\underline{y} > k_\alpha | H_A)$$

$$= 1 - P\left(\frac{\underline{y}^*\underline{y}}{\sigma_A^2} > \frac{k_\alpha}{\sigma_A^2} | H_A\right)$$

we may use:

(15) $$1 - \beta = P\left(\frac{\underline{y}^*\underline{y}}{\sigma_A^2} > \frac{k_\alpha}{\sigma_A^2} | H_A\right)$$

and the table of the χ^2-distribution to compute β from k_α. Table 2.2 shows some typical values of k_α and β for the case $m = 1$ and $\sigma_A^2 = 4$.

k_α	$\frac{k_\alpha}{\sigma_A^2} = \frac{1}{4}k_\alpha$	$1-\beta$	β
5.41	1.35	0.24	0.76
7.68	1.92	0.17	0.83
13.27	3.32	0.07	0.93
21.66	5.41	0.02	0.98

Table 2.2: k_α and β for the distribution $\sigma_A^2\chi^2(m,0)$ with $\sigma_A^2=4$ and $m=1$.

Table 2.3 shows some typical values of k_α and β for the case $m=4$ and $\sigma_A^2=4$.

k_α	$\frac{k_\alpha}{\sigma_A^2} = \frac{1}{4}k_\alpha$	$1-\beta$	β
15.56	3.89	0.42	0.58
18.98	4.74	0.31	0.69
26.55	6.64	0.16	0.84
36.93	9.23	0.06	0.94

Table 2.3: k_α and β for the distribution $\sigma_A^2\chi^2(m,0)$ with $\sigma_A^2=4$ and $m=4$.

Upon comparing Table 2.2 and Table 2.3 we note that at the same size of type I error and thus at the same critical value k_α, the β for the case $m=4$ is less than the β for the case $m=1$. This is also what one would expect, since by increasing the number of observations one would expect to have a higher probability of correctly accepting H_A. Show for yourself that the β-values of table 3 will increase if instead of $H_A : \sigma_A^2=4$ we have the alternative $H_A : \sigma_A^2=3$.

Example 2

Assume that $\underline{y}$ is distributed as:

(16) $$\underline{y} \sim N(E\{\underline{y}\},\sigma^2)$$

with known variance σ^2. The following two simple hypotheses are considered:

(17) $$H_0 : \quad E\{\underline{y}\} = x_0 \quad \text{versus} \quad H_A : \quad E\{\underline{y}\} = x_A > x_0.$$

With:

$$p_{\underline{y}}(y|x_0) = (2\pi)^{-1/2}(\sigma^2)^{-1/2}\ \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y-x_0)^2]$$

and

$$p_{\underline{y}}(y|x_A) = (2\pi)^{-1/2}(\sigma^2)^{-1/2} \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y-x_A)^2]$$

it follows that:

$$\frac{p_{\underline{y}}(y|x_0)}{p_{\underline{y}}(y|x_A)} = \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}\{(y-x_0)^2-(y-x_A)^2\}]. \tag{18}$$

With (3) this gives:

$$\text{reject } H_0 \text{ if } \quad \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}\{(y-x_0)^2-(y-x_A)^2\}] < a. \tag{19}$$

Taking the logarithm of the inequality in (19) gives:

$$-\tfrac{1}{2}\tfrac{1}{\sigma^2}\{(y-x_0)^2-(y-x_A)^2\} < \ln a.$$

Multiplication with $-2\sigma^2$ gives:

$$(y-x_0)^2-(y-x_A)^2 > \sigma^2 \ln a^{-2}$$

or

$$2y(x_A-x_0)+(x_0^2-x_A^2) > \sigma^2 \ln a^{-2}$$

or

$$2y(x_A-x_0) > \sigma^2 \ln a^{-2}+(x_A^2-x_0^2).$$

Since $x_A > x_0$, division by $2(x_A-x_0)$ finally gives:

$$y > \frac{\sigma^2 \ln a^{-1}+\frac{1}{2}(x_A^2-x_0^2)}{x_A-x_0}.$$

If we denote the right-hand side of this inequality by k_α, we see that the simple likelihood ratio test for this particular example reduces to:

$$\text{reject } H_0 \text{ if } \quad y > k_\alpha. \tag{20}$$

The corresponding critical region K of this test is shown in Figure 2.4. Note that it is identical to the critical region of Example 3 of the previous chapter, the one which was chosen on more or less intuitive grounds. In Example 3 of the previous chapter we noted that a transformation of $\underline{y}$ to the standard normal distribution was useful for computing the sizes α and β. We might therefore just as well write test (20) in terms of this transformed random variable. This gives:

$$\text{reject } H_0 \text{ if } \quad \frac{y-x_0}{\sigma} > \frac{k_\alpha-x_0}{\sigma} = \text{new critical value} \tag{21}$$

with $(\underline{y}-x_0)/\sigma$ standard normally distributed under H_0.

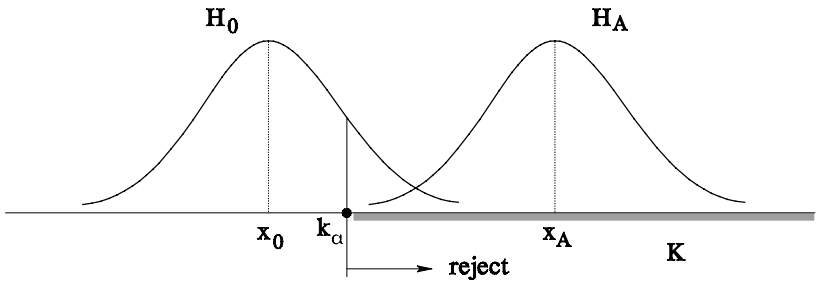


Figure 2.4: Critical region K for $H_0 : E\{\underline{y}\} = x_0$ versus $H_A : E\{\underline{y}\} = x_A > x_0$.

Example 3

Assume that $\underline{y}$ is distributed as:

(22) $$\underline{y} \sim xe^{-yx}, \quad x > 0, \quad y \geq 0.$$

The following two simple hypotheses are considered:

(23) $$H_0 : \quad x = x_0 \qquad \text{versus} \qquad H_A : \quad x = x_A < x_0.$$

With:

$$p_{\underline{y}}(y|x_0) = x_0 e^{-yx_0}$$

and

$$p_{\underline{y}}(y|x_A) = x_A e^{-yx_A}$$

it follows that the simple likelihood ratio reads:

(24) $$\frac{p_{\underline{y}}(y|x_0)}{p_{\underline{y}}(y|x_A)} = \frac{x_0}{x_A}\exp[-y(x_0 - x_A)].$$

With (3) this gives:

(25) $$\text{reject } H_0 \text{ if} \qquad \frac{x_0}{x_A}\exp[-y(x_0 - x_A)] < a.$$

Simplification of the inequality gives:

(26) $$y > \frac{\ln[\frac{x_A}{x_0}a]^{-1}}{x_0 - x_A}.$$

If we denote the right-hand side of this inequality by k_α, we see that the simple likelihood ratio test for this particular example is given by:

(27) $$\text{reject } H_0 \text{ if } \quad y > k_\alpha .$$

Compare this result with Example 4 of the previous chapter.

2.2 Most powerful tests

In Section 1.4 of the previous chapter we presented the *Neyman-Pearson* testing principle. This principle says to choose among all tests possessing the same size α, the one for which the size of the type II error, β, is as small as possible. This statement is expressed in terms of β, the probability that the sample will fall in the non critical region when in fact H_A is true. It is usually, however, more convenient to work exclusively with the critical region K. It is therefore customary to calculate $1-\beta$, which is the probability that the sample will fall in the critical region K when in fact H_A is true. The probability $1-\beta$ is called the power of the test and it is denoted by γ. Thus:

The *power* γ of a test is the probability of correctly rejecting H_0. The power can be calculated as:

(28) $$\boxed{\gamma = P(\underline{y} \in K \,|\, H_A) = \int_K p_{\underline{y}}(y \,|\, x_A)\,dy}\ .$$

We can now rephrase the Neyman-Pearson testing principle in terms of the power γ. This gives the following definition of a most powerful test.

A test of $H_0 : x = x_0$ versus $H_A : x = x_A$, with a critical region K and a size α is defined to be a *most powerful* test of size α if and only if:

(i) $\alpha = P(\underline{y} \in K \,|\, H_0)$
and
(ii) $\gamma = P(\underline{y} \in K \,|\, H_A) \geq P(\underline{y} \in K^* \,|\, H_A)$

for any other test with critical region K^* and size $\alpha = P(\underline{y} \in K^* \,|\, H_0)$.

So far we have seen in our example that the simple likelihood ratio test produces critical regions that are indeed intuitively appealing. We have however not yet considered the question of optimality of the simple likelihood ratio test. The following important theorem, by Neyman and Pearson, shows that the simple likelihood ratio test is a most powerful test.

Neyman-Pearson theorem: Let y be a sample or observation from $p_{\underline{y}}(y \,|\, x)$ where x is one of two known values x_0 and x_A, and let $0 < \alpha < 1$ be fixed. Let a be a positive constant and K be a subset of the sample space which satisfies:

(i) $\alpha = P(\underline{y} \in K \mid H_0)$

(ii) $$\frac{p_{\underline{y}}(y \mid x_0)}{p_{\underline{y}}(y \mid x_A)} \begin{cases} < a & \text{if } y \in K \\ > a & \text{if } y \notin K. \end{cases}$$

Then the test corresponding to the critical region K, that is, the simple likelihood ratio test, is a most powerful test of size α for testing $H_0 : x = x_0$ versus $H_A : x = x_A$.

Proof

To prove the Neyman-Pearson theorem, let K^* be any other critical region of size α. The regions K and K^* may be represented geometrically as the regions interior to the indicated closed surfaces in Figure 2.5.

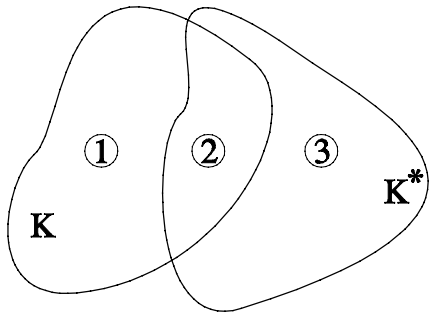


Figure 2.5: Critical regions K and K^* of size α.

Since K and K^* are both critical regions of size α:

$$\alpha = P(\underline{y} \in K \mid H_0) = P(\underline{y} \in K^* \mid H_0)$$

or

(29) $$\int_K p_{\underline{y}}(y \mid x_0)\mathrm{d}y = \int_{K^*} p_{\underline{y}}(y \mid x_0)\mathrm{d}y.$$

But, from Figure 2.5 it is clear that the integral over ② which is the common part of K and K^*, will cancel from both sides of (29) and reduce it to the form:

(30) $$\int_{①} p_{\underline{y}}(y \mid x_0)\mathrm{d}y = \int_{③} p_{\underline{y}}(y \mid x_0)\mathrm{d}y.$$

Since the power of a test is given by the probability that the sample will fall inside the critical region when H_A is true, we have for the two critical regions K and K^*:

$$\gamma = \int_K p_{\underline{y}}(y|x_A)\mathrm{d}y \qquad \text{and} \qquad \gamma^* = \int_{K^*} p_{\underline{y}}(y|x_A)\mathrm{d}y.$$

Consequently we get for the difference in power:

$$\gamma - \gamma^* = \int_K p_{\underline{y}}(y|x_A)\mathrm{d}y - \int_{K^*} p_{\underline{y}}(y|x_A)\mathrm{d}y.$$

Since the integral over the common part cancels, this difference reduces to:

$$\gamma - \gamma^* = \int_{①} p_{\underline{y}}(y|x_A)\mathrm{d}y - \int_{③} p_{\underline{y}}(y|x_A)\mathrm{d}y. \tag{31}$$

Since region ① lies in K, it follows from (ii) of the theorem that every point y of ① satisfies the inequality:

$$p_{\underline{y}}(y|x_0) < a p_{\underline{y}}(y|x_A) \qquad \text{for } y \in ①.$$

Hence:

$$\int_{①} p_{\underline{y}}(y|x_A)\mathrm{d}y > \frac{1}{a}\int_{①} p_{\underline{y}}(y|x_0)\mathrm{d}y. \tag{32}$$

Similarly, since ③ lies outside K, it follows from (ii) of the theorem that every point y of ③ satisfies the inequality:

$$p_{\underline{y}}(y|x_0) > a p_{\underline{y}}(y|x_A) \qquad \text{for } y \in ③.$$

Hence:

$$\int_{③} p_{\underline{y}}(y|x_A)\mathrm{d}y < \frac{1}{a}\int_{③} p_{\underline{y}}(y|x_0)\mathrm{d}y. \tag{33}$$

When the results (32) and (33) are used in (31), it follows that:

$$\gamma - \gamma^* > \frac{1}{a}\int_{①} p_{\underline{y}}(y|x_0)\mathrm{d}y - \frac{1}{a}\int_{③} p_{\underline{y}}(y|x_0)\mathrm{d}y.$$

But from (30), the right side of this inequality must be equal to zero, hence:

$$\gamma > \gamma^*.$$

Since γ^* is the power of the test using any other critical region K^* of size α, the preceding analysis proves that the test corresponding to the critical region K is indeed a most powerful test of size α.

End of proof.

Although the theorem does not explicitly say how to find the constant a and the region K, implicitly it does since the form of the test, that is, the critical region K, is given by (ii) of the theorem. In practice it is often, as shown in previous examples, not necessary to find a. Instead the inequality of (ii) of the theorem for $y \in K$ is manipulated into an equivalent form that is easier to work with, and the actual test is then expressed in terms of the new inequality. The following example should make this clear.

Example 4

We will now consider the multi-dimensional generalization of Example 2. Assume therefore that $\underset{m\times 1}{\underline{y}}$ is an $m\times 1$ random vector which is distributed as:

(34) $$\underset{m\times 1}{\underline{y}} \sim N(\underset{m\times 1}{E\{\underline{y}\}}, \underset{m\times m}{\sigma^2 I_m})$$

with known variance σ^2. The following two simple hypotheses are considered:

(35) $$H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times 1}{x_0} \quad \text{versus } H_A : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times 1}{x_A}.$$

The situation is sketched in Figure 2.6. Figure 2.6 shows the location of the two simple hypotheses H_0 and H_A in the sample space $\mathbb{R}^m$. It also shows the contours of constant density of the distribution of $\underline{y}$, and it shows the location of the sample point y.

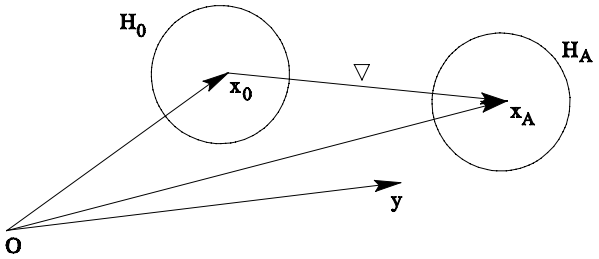


Figure 2.6: The geometry of $H_0 : E\{\underline{y}\} = x_0$ and $H_A : E\{\underline{y}\} = x_A$.

In order to apply the simple likelihood ratio test we need to know the density functions $p_{\underline{y}}(y|x_0)$ and $p_{\underline{y}}(y|x_A)$. They read:

$$p_{\underline{y}}(y|x_0) = (2\pi)^{-m/2}(\sigma^2)^{-m/2}\exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y-x_0)^*(y-x_0)]$$

and

$$p_{\underline{y}}(y|x_A) = (2\pi)^{-m/2}(\sigma^2)^{-m/2}\exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y-x_A)^*(y-x_A)].$$

Hence:

(36) $$\frac{p_{\underline{y}}(y|x_0)}{p_{\underline{y}}(y|x_A)} = \exp[-\frac{1}{2}\frac{1}{\sigma^2}\{(y-x_0)^*(y-x_0) - (y-x_A)^*(y-x_A)\}].$$

With (3) this gives:

(37) $$\text{reject } H_0 \text{ if } \quad \exp[-\frac{1}{2}\frac{1}{\sigma^2}\{(y-x_0)^*(y-x_0) - (y-x_A)^*(y-x_A)\}] < a.$$

We will now transform this inequality into an inequality that can be considered as the multi-dimensional generalization of the inequality of (21). After taking the logarithm and multiplying with $-2\sigma^2$, the inequality of (37) takes the form:

$$(y-x_0)^*(y-x_0) - (y-x_A)^*(y-x_A) > \sigma^2 \ln a^{-2}$$

or

$$2(x_A-x_0)^*y + x_0^*x_0 - x_A^*x_A > \sigma^2 \ln a^{-2}.$$

By adding and subtracting $2(x_A-x_0)^*x_0$, this can also be written as:

$$2(x_A-x_0)^*(y-x_0) - (x_A^*x_A - 2x_A^*x_0 + x_0^*x_0) > \sigma^2 \ln a^{-2}$$

or as:

(38) $$2(x_A-x_0)^*(y-x_0) > \sigma^2 \ln a^{-2} + (x_A-x_0)^*(x_A-x_0).$$

By denoting the length of the vector $x_A - x_0$ by ∇:

$$\|x_A - x_0\| = \nabla$$

and the unit vector in the direction of $x_A - x_0$ by c:

$$\frac{x_A - x_0}{\|x_A - x_0\|} = c$$

it follows from substituting:

$$x_A - x_0 = c\nabla$$

into (38) that:

$$2\nabla c^*(y-x_0) > \sigma^2 \ln a^{-2} + \nabla^2.$$

Division by $2\nabla\sigma$ then finally gives:

(39) $$\frac{c^*(y-x_0)}{\sigma} > \frac{\sigma \ln a^{-1}}{\nabla} + \frac{1}{2\sigma}\nabla.$$

If we denote the right-hand side of this inequality by k_α, we see that the simple likelihood ratio test for this particular example reduces to:

(40) $$\text{reject } H_0 \text{ if } \quad \frac{c^*(y-x_0)}{\sigma} > k_\alpha, \qquad \text{with} \qquad c = \frac{x_A - x_0}{\|x_A - x_0\|}.$$

This test can be considered the multi-dimensional generalization of test (21) of Example 2. The form of the critical region K corresponding to test (40) is shown in Figure 2.7. For the case shown we have $y \in K$, implying that H_0 is rejected.

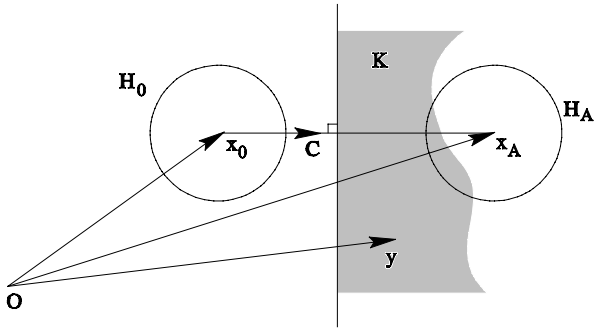


Figure 2.7: Critical region K for testing $H_0 : E\{\underline{y}\} = x_0$ versus $H_A : E\{\underline{y}\} = x_A$.

Note that the scalar random variable:

(41) $$\underline{z} = \frac{c^*(\underline{y}-x_0)}{\sigma}$$

has a standard normal distribution under H_0. It is therefore rather straightforward to compute from a table of the standard normal distribution the critical value k_α and power γ of the test for a fixed size α. The critical value k_α follows from:

(42) $$\alpha = \int_{k_\alpha}^{\infty} (2\pi)^{-1/2} \exp\left(-\tfrac{1}{2}\tfrac{1}{\sigma^2} z^2\right) dz.$$

Since the power γ is given by:

$$\gamma = P(\underline{y} \in K \,|\, H_A) = P(\underline{z} > k_\alpha \,|\, H_A)$$

and since $\underline{z}$ is distributed as $\underline{z} \sim N(\frac{\nabla}{\sigma},1)$ under H_A, it follows that:

$$\gamma = \int_{k_\alpha}^{\infty} (2\pi)^{-1/2} \exp\left[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(z - \tfrac{\nabla}{\sigma})^2\right] dz$$

which can be transformed into an integral of the standard normal distribution as:

(43) $$\gamma = \int_{k_\alpha - \nabla/\sigma}^{\infty} (2\pi)^{-1/2} \exp(-\tfrac{1}{2}z^2)\,dz.$$

Note that the power γ, for a fixed critical value k_α, is a monotone increasing function of ∇/σ. Thus γ gets larger if ∇ gets larger. This is what one would expect. The further H_0 and H_A are apart (see Figure 2.7) the higher one would expect the power γ to be. The power γ also gets larger if the standard deviation σ gets smaller. This is also what one would expect. The better the precision of the observations, the higher one would expect the power γ to be.

2.3 The $\underline{w}$-teststatistic

Recall from Adjustment theory the *linear model of observation equations*:

(44) $$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{\boldsymbol{A}}\,\underset{n\times 1}{x} \quad ; \quad \underset{m\times m}{D\{\underline{y}\}} = \underset{m\times m}{Q_y}, \qquad \text{with} \qquad \operatorname{rank}\boldsymbol{A} = n,\ \operatorname{rank}Q_y = m.$$

Let us now try to find out if and how the theory of hypothesis testing, as developed in the previous sections, can be applied for testing a model like (44). First of all we have to assume a probability distribution for $\underline{y}$. Since the normal distribution is adequate for most of the geodetic applications, we assume that the $m\times 1$ random vector $\underline{y}$ is normally distributed with mean $E\{\underline{y}\} = Ax$ and variance matrix $D\{\underline{y}\} = Q_y$. Our null hypothesis H_0 reads therefore:

(45) $$H_0 : \underset{m\times 1}{\underline{y}} \sim N(\underset{m\times n}{\boldsymbol{A}}\,\underset{n\times 1}{x},\ \underset{m\times m}{Q_y}).$$

Note that the $n\times 1$ parameter vector x in (45) is unspecified. Hence, the above null hypothesis H_0 is a composite hypothesis. It seems therefore that our theory which so far only holds for simple hypotheses, cannot be applied. The theory can be applied, however, if we are able to transform (45) into a simple hypothesis. Recall from Adjustment theory the *linear model of condition equations*:

(46) $$\underset{b\times m}{\boldsymbol{B}^*}\,\underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0} \quad ; \quad \underset{m\times m}{D\{\underline{y}\}} = \underset{m\times m}{Q_y}, \qquad \text{with} \qquad \operatorname{rank}\boldsymbol{B} = b,\ \operatorname{rank}Q_y = m.$$

As we know, this model is completely equivalent to (44). We also know that the two matrices A and B respectively of (44) and (46) satisfy the relation:

(47) $$\underset{b\times m}{\boldsymbol{B}^*}\,\underset{m\times n}{\boldsymbol{A}} = \underset{b\times n}{0}.$$

Using this relation, we may also write (45) as:

(48) $$H_0 : \underset{m\times 1}{\underline{y}} \sim N(\underset{m\times 1}{E\{\underline{y}\}}, \underset{m\times m}{Q_y}) \qquad \text{with} \qquad \underset{b\times m}{\boldsymbol{B}^*}\,\underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0}.$$

This hypothesis is equivalent to the null hypotheses of (45), just like (46) is equivalent to (44). H_0 of (48) is of course still a composite hypothesis. This follows since only $b < m$ linear

independent functions of $E\{\underline{y}\}$ are specified in (48). This leaves $m-b$ linear independent functions of $E\{\underline{y}\}$ unspecified. And $m-b$ is equal to n, the number of unspecified parameters in (45). Thus, since (48) is composite as well, no direct application of our theory is possible. As was mentioned above however, the theory can be applied if we are able to transform (48) into a simple hypothesis. This is therefore the approach we will take in this section. We define the $b\times 1$ random vector $\underline{t}$ as:

$$\underset{b\times 1}{\underline{t}} = \underset{b\times m}{\boldsymbol{B}^*}\ \underset{m\times 1}{\underline{y}}\ . \tag{49}$$

Recall from Adjustment theory that this is the vector of misclosures (tegenspraken). Under the null hypothesis of (48) the random vector $\underline{t}$ is normally distributed with mean $E\{\underline{t}\} = 0$ and variance-matrix $Q_t = B^*Q_yB$. Thus under H_0 we have:

$$\boxed{H_0' : \underset{b\times 1}{\underline{t}} \sim N(\underset{b\times 1}{0}, \underset{b\times b}{Q_t = \boldsymbol{B}^*Q_y\boldsymbol{B}})}\ . \tag{50}$$

Note that this null hypothesis H_0' is a simple hypothesis. But, also note that (50) is not equivalent to (48). That is, the hypothesis H_0' follows from H_0, but H_0 does not follow from H_0'. This is due to the fact that the matrix B of (49) is not invertible. Although the simple hypothesis H_0' is not equivalent to the composite hypothesis H_0, we will settle with H_0' and try to test it against an alternative hypothesis. Then, if H_0' gets rejected, H_0 should be rejected too. This is because H_0 cannot be true while H_0' is false. On the other hand, if H_0' gets accepted one should be very careful in accepting H_0. H_0 can namely be false while H_0' is true. The following example makes this clear.

Example 5

Assume that the true hypothesis is:

$$H_{\text{true}} : \underset{m\times 1}{\underline{y}} \sim N(\underset{m\times n}{\boldsymbol{A}}\underset{n\times 1}{x} + \underset{m\times 1}{\nabla y},\ \underset{m\times m}{Q_y} + \underset{m\times m}{\nabla Q_y})\,. \tag{51}$$

If $\nabla y \neq 0$ and/or $\nabla Q_y \neq 0$, this hypothesis is clearly different from H_0 of (45). Now consider the effect of H_{true} on the distribution of $\underline{t} = B^*\underline{y}$. For the mean of $\underline{t}$ under H_{true} we have:

$$E\{\underline{t}\,|\,H_{\text{true}}\} = \boldsymbol{B}^*E\{\boldsymbol{A}x + \nabla y\} = \boldsymbol{B}^*\nabla y.$$

And for the variance matrix of $\underline{t}$ under H_{true} we have:

$$Q_{t|H_{\text{true}}} = \boldsymbol{B}^*(Q_y + \nabla Q_y)\boldsymbol{B} = Q_t + \boldsymbol{B}^*\nabla Q_y\boldsymbol{B}.$$

Hence, the distribution of $\underline{t}$ under H_{true} reads:

(52) $$H'_{\text{true}} : \quad \underset{b\times 1}{\underline{t}} \sim N(\underset{b\times 1}{B^{*}\nabla y},\ \underset{b\times b}{Q_t + B^{*}\nabla Q_y B}).$$

In general this hypothesis differs from H'_0 of (50). But if the vector ∇y and the columns of matrix ∇Q_y lie in the nullspace of B^*, that is, $\nabla y \in N(B^*)$ and $R(\nabla Q_y) \subset N(B^*)$, then $B^*\nabla y = 0$ and $B^*\nabla Q_y B = 0$. In this case H'_{true} of (52) becomes identical to H'_0 of (50), while H_{true} of (51) still differs from H_0 of (45). This shows that H_0 can be false while H'_0 is true.

Now let us have a look at an alternative hypothesis for H_0. Many different types of alternative hypotheses may be considered. For instance, the alternative hypothesis may specify that $\underline{y}$ has a mean Ax, a variance matrix Q_y, but a distribution that differs from the normal distribution. Or, the alternative hypothesis may specify that $\underline{y}$ is normally distributed with mean Ax, but with a variance matrix that differs from Q_y. In these lecture notes however we will primarily be concerned with alternative hypotheses that differ from the null hypothesis in the mean of $\underline{y}$. The reason is that in most geodetic applications the alternative hypotheses are used to model errors or blunders in the observations. For instance, if we want to find out whether the ith-observation is erroneous or not we may model the alternative hypothesis as:

(53) $$H_A : \underset{m\times 1}{\underline{y}} \sim N(\underset{m\times n}{A}\underset{n\times 1}{x} + \underset{m\times 1}{\nabla y},\ \underset{m\times m}{Q_y})$$

with

(54) $$\underset{m\times 1}{\nabla y} = \underset{m\times 1}{c_y}\ \underset{1\times 1}{\nabla}, \qquad c_y = (0 \cdots 0\ \underset{\underset{i\text{th}}{\uparrow}}{1}\ 0 \cdots 0)^*, \qquad \nabla \in \mathbb{R}.$$

In this case, the scalar ∇ is the error or blunder in the observation and the vector c_y models the error ∇ to be in the ith-observation.

The vector ∇y in (53) may also be used to model systematic errors in the observations. For instance, if all observations contain a systematic error of ∇, the vector ∇y of (53) takes instead of (54) the form:

(55) $$\underset{m\times 1}{\nabla y} = \underset{m\times 1}{c_y}\ \underset{1\times 1}{\nabla}, \qquad c_y = (1\ 1 \cdots 1\ 1)^*.$$

These two examples show that one can model different types of errors in the observations through an appropriate choice of the vector c_y. Now let us consider the effect of H_A on the distribution of $\underline{t} = B^*\underline{y}$. It follows that the distribution of $\underline{t}$ under H_A is given by:

(56) $$H'_A : \quad \underline{t} \sim N(\nabla t = B^*\nabla y\ ,\ Q_t).$$

With the definitions:

(57) $$\underset{m\times 1}{\nabla y} = \underset{m\times 1}{c_y}\,\nabla_{1\times 1} \qquad \text{and} \qquad \underset{b\times 1}{c_t} = \underset{b\times m}{\boldsymbol{B}^*}\,\underset{m\times 1}{c_y}$$

we can write (56) also as:

(58) $$\boxed{H_A' : \quad \underset{b\times 1}{\underline{t}} \sim N(\underset{b\times 1}{c_t}\,\nabla_{1\times 1}, \underset{b\times b}{Q_t})}\ .$$

This hypothesis can be considered the alternative of H_0' of (50). The hypothesis H_A' is a composite hypothesis if the parameter ∇ remains unspecified. In order to make it into a simple hypothesis we will therefore assume that beside c_t also ∇ is known. For most geodetic applications this is not very realistic, because one will hardly ever know a priori how large an error in an observation will be if it occurs. In the next chapter we will therefore relax this assumption and assume ∇ unknown.

For the present application of the theory we have to assume however that (58) is a simple hypothesis and therefore that ∇ is known and positive. Now that we formulated the two simple hypotheses H_0' and H_A', we are in the position to apply our theory of hypothesis-testing. In order to apply the simple likelihood ratio test we need to know the probability density functions of $\underline{t}$ under H_0' and H_A' respectively. They read:

$$p_{\underline{t}}(t|H_0') = (2\pi)^{-b/2}|Q_t|^{-1/2}\ \exp[-\tfrac{1}{2}t^*Q_t^{-1}t]$$

and

$$p_{\underline{t}}(t|H_A') = (2\pi)^{-b/2}|Q_t|^{-1/2}\ \exp[-\tfrac{1}{2}(t-c_t\nabla)^*Q_t^{-1}(t-c_t\nabla)]\,.$$

Hence:

(59) $$\frac{p_{\underline{t}}(t|H_0')}{p_{\underline{t}}(t|H_A')} = \exp[-\tfrac{1}{2}\{t^*Q_t^{-1}t-(t-c_t\nabla)^*Q_t^{-1}(t-c_t\nabla)\}]\,.$$

With (3) this gives:

(60) $$\text{reject } H_0'(H_0) \text{ if} \qquad \exp[-\tfrac{1}{2}\{t^*Q_t^{-1}t - (t-c_t\nabla)^*Q_t^{-1}(t-c_t\nabla)\}] < \boldsymbol{a}\ .$$

The inequality can be simplified to:

(61) $$c_t^*Q_t^{-1}t > \frac{\ln\boldsymbol{a}^{-1}}{\nabla} + \tfrac{1}{2}c_t^*Q_t^{-1}c_t\nabla.$$

Since $c_t^*Q_t^{-1}\underline{t}$ is distributed under H_0' as $N(0, c_t^*Q_t^{-1}c_t)$, we may bring (61) into the standard normal form by dividing by $(c_t^*Q_t^{-1}c_t)^{1/2}$. This gives:

(62) $$\text{reject } H_0'(H_0) \text{ if } \quad \frac{c_t^* Q_t^{-1} t}{\sqrt{c_t^* Q^{-1} c_t}} > \frac{\ln a^{-1}}{\sqrt{c_t^* Q_t^{-1} c_t}\nabla} + \frac{1}{2}\sqrt{c_t^* Q_t^{-1} c_t}\nabla.$$

If we denote the right-hand side of this inequality by k_α, and define the random variable $\underline{w}$ as:

(63) $$\boxed{\underline{w} = \frac{c_t^* Q_t^{-1} \underline{t}}{\sqrt{c_t^* Q_t^{-1} c_t}}}$$

the simple likelihood ratio test reduces to:

(64) $$\text{reject } H_0'(H_0) \text{ if } \quad w > k_\alpha .$$

Note that the random variable $\underline{w}$ is distributed under $H_0'(H_0)$ and $H_A'(H_A)$ as:

(65) $$H_0'(H_0): \quad \underline{w} \sim N(0,1) \; ; \quad H_A'(H_A): \quad \underline{w} \sim N(\sqrt{c_t^* Q_t^{-1} c_t}\nabla, 1).$$

The random variable $\underline{w}$ is called the *w-teststatistic* (w-toetsgrootheid) and, as we will see in later chapters, it plays a very important role in hypothesis-testing for geodetic applications.

It is very illustrative if we interpret the simple likelihood ratio test (64) and the $\underline{w}$-teststatistic (63) geometrically. In order to do so we define the following innerproduct in the space $\mathbb{R}^b$:

(66) $$\text{innerproduct between } u,v \in \mathbb{R}^b: \quad (u,v)_{Q_t} = u^* Q_t^{-1} v.$$

The norm (or length) of a vector in $\mathbb{R}^b$ and the innerproduct of two vectors in $\mathbb{R}^b$ can be written as:

(67) $$\|u\|_{Q_t} = \sqrt{u^* Q_t^{-1} u} \; ; \quad (u,v)_{Q_t} = \|u\|_{Q_t}\|v\|_{Q_t}\cos\alpha \qquad \alpha = \sphericalangle u,v.$$

With these definitions we may write (63) also as:

$$\underline{w} = \frac{(c_t,\underline{t})_{Q_t}}{\sqrt{(c_t,c_t)_{Q_t}}} = \frac{\|c_t\|_{Q_t}\|\underline{t}\|_{Q_t}\cos\alpha}{\|c_t\|_{Q_t}} = \|\underline{t}\|_{Q_t}\cos\alpha.$$

This shows that $\underline{w}$ is the orthogonal projection of $\underline{t}$ onto the line with direction vector c_t (see Figure 2.8).

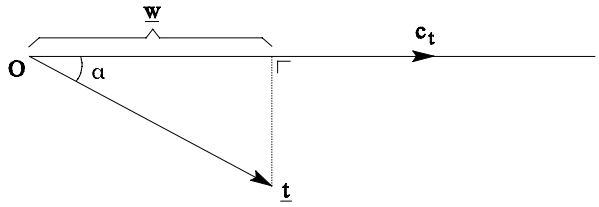


Figure 2.8: $\underline{w}$ is the orthogonal projection of $\underline{t}$ onto c_t.

In a similar way we may now also illustrate test (64) geometrically. This is done in Figure 2.9. For the case shown we have $t \notin K$, implying that H_0' gets accepted.

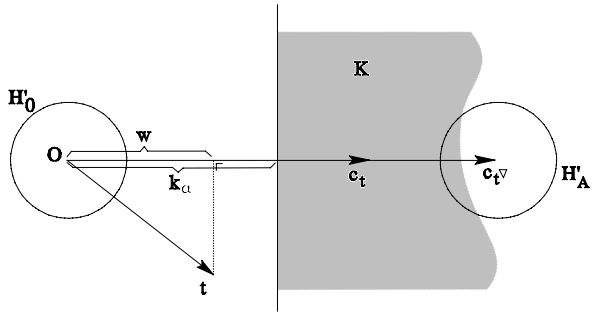


Figure 2.9: Critical region K for testing H_0' versus H_A'.

The $\underline{w}$-teststatisic (63) has been formulated in terms of $\underline{t}, Q_t$ and c_t. We may however also express $\underline{w}$ in terms of the original quantities $\underline{y}, Q_y$ and c_y. Substitution of:

$$\underline{t} = \boldsymbol{B}^* \underline{y} \quad , \quad Q_t = \boldsymbol{B}^* Q_y \boldsymbol{B} \text{ and } c_t = \boldsymbol{B}^* c_y$$

into (63) gives:

(68)
$$\underline{w} = \frac{c_y^* \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y}}{\sqrt{c_y^* \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* c_y}} .$$

Now, recall from Adjustment theory that the least-squares residual vector $\underline{\hat{e}}$ and its variance matrix $Q_{\hat{e}}$, expressed in quantities belonging to the model of condition equations, read:

$$\underline{\hat{e}} = Q_y \boldsymbol{B} (\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* \underline{y}$$

and

$$Q_{\hat{e}} = Q_y \boldsymbol{B} (\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* Q_y \ .$$

If we substitute this into (68) we get:

(69)
$$\boxed{\underline{w} = \frac{c_y^* Q_y^{-1} \underline{\hat{e}}}{\sqrt{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y}}} .$$

This shows that the $\underline{w}$-teststatistic can be computed directly form the results of the least- squares adjustment of either the model of observation equations (44) or the model of condition equations (46). Also expression (69) can be interpreted geometrically. Recall from Adjustment theory that:

$$(70)\qquad \begin{cases} \hat{\underline{e}} &= (I-P_A)\underline{y} &= P_A^{\perp}\underline{y}\ , \\ Q_{\hat{e}} &= P_A^{\perp}Q_y P_A^{\perp *} &= Q_y P_A^{\perp *} Q_y^{-1} P_A^{\perp} Q_y\ , \\ Q_y^{-1}P_A^{\perp} &= P_A^{\perp *} Q_y^{-1} P_A^{\perp}\ . \end{cases}$$

Using these results in (69) gives:

$$(71)\qquad \underline{w} = \frac{(P_A^{\perp}c_y)^* Q_y^{-1}(P_A^{\perp}\underline{y})}{\sqrt{(P_A^{\perp}c_y)^* Q_y^{-1}(P_A^{\perp}c_y)}}\ .$$

Note that this expression has the same structure as (63). The geometric interpretation is therefore very similar to the previous given one. We define the following innerproduct in the sample space $\mathbb{R}^m$:

$$(72)\qquad \text{innerproduct between } u,v,\in\mathbb{R}^m\text{: } (u,v)_{Q_y} = u^* Q_y^{-1} v.$$

With this innerproduct we may write (71) as:

$$\underline{w} = \frac{(P_A^{\perp}c_y, P_A^{\perp}\underline{y})_{Q_y}}{\sqrt{(P_A^{\perp}c_y, P_A^{\perp}c_y)_{Q_y}}} = \frac{\|P_A^{\perp}c_y\|_{Q_y}\|P_A^{\perp}\underline{y}\|_{Q_y}\cos\alpha}{\|P_A^{\perp}c_y\|_{Q_y}} = \|P_A^{\perp}\underline{y}\|_{Q_y}\cos\alpha\ .$$

This shows that $\underline{w}$ is the orthogonal projection of $P_A^{\perp}\underline{y}$ onto the line with direction vector $P_A^{\perp}c_y$. Note that $P_A^{\perp}\underline{y}$ and $P_A^{\perp}c_y$ are both the orthogonal projections of $\underline{y}$ and c_y respectively on $R(A)^{\perp}$, the orthogonal complement of the rangespace, $R(A)$, of A. Figure 2.10 gives a sketch of test (64) in terms of quantities that are located in the sample space $\mathbb{R}^m$.

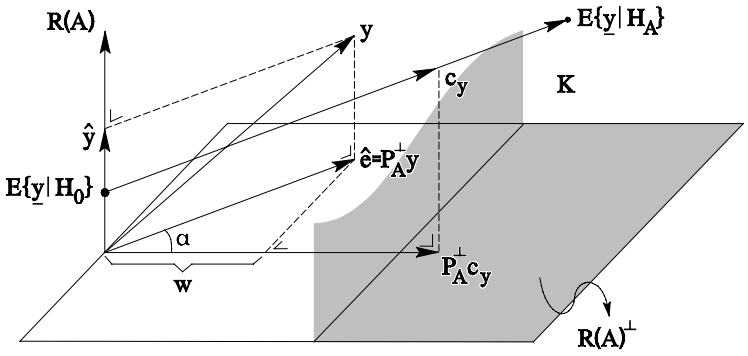


Figure 2.10: Critical region $K\subset\mathbb{R}^m$ for test (64).

In order to see the theory at work we now will consider a typical geodetic example.

Example 6

Figure 2.11 shows a typical levelling network of four points with two loops.

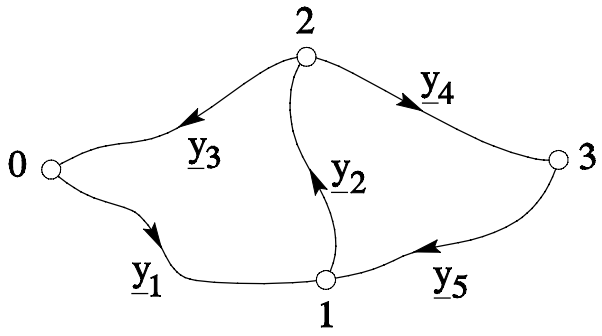


Figure 2.11: A levelling network.

If we assume that the height x_0 of point 0 is known and equal to zero, the linear model of observation equations reads:

$$H_0: \quad E\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\right\} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad ; \quad Q_y = \sigma^2 I_5. \tag{73}$$

Note that we have assumed that the variancematrix of the observables is equal to a scaled identity matrix. We will also assume that the observables are normally distributed. The linear model can of course also be expressed in terms of condition equations. In terms of condition equations we get:

$$H_0: \quad \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} E\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\right\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad ; \quad Q_y = \sigma^2 I_5. \tag{74}$$

The models (73) or (74) together with the assumption of normally distributed observables, constitute our null hypothesis H_0. Let us now consider the alternative hypothesis H_A. For this particular example we assume to know that if H_0 is false, then an error in observation y_2 has been made of a known amount ∇. The alternative hypothesis in terms of observation equations reads therefore:

(75) $$H_A: \quad E\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\right\} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \nabla \quad ; \quad Q_y = \sigma^2 I_5.$$

And in terms of condition equations this hypothesis reads:

(76) $$H_A: \quad \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} E\left\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\right\} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \nabla \quad ; \quad Q_y = \sigma^2 I_5.$$

With (74) and (76) we are now in the position to compute the quantities which are needed in the $\underline{w}$-teststatistic (63). The vector of misclosures, $\underline{t} = B^* \underline{y}$, and its variancematrix, $Q_t = B^* Q_y B$ follow from (74) as:

(77) $$\underline{t} = \begin{pmatrix} \underline{t}_1 \\ \underline{t}_2 \end{pmatrix} = \begin{pmatrix} \underline{y}_1 + \underline{y}_2 + \underline{y}_3 \\ \underline{y}_2 + \underline{y}_4 + \underline{y}_5 \end{pmatrix} \quad \text{and}$$

(78) $$Q_t = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \sigma^2 I_5 \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = \sigma^2 \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

The vector $c_t = B^* c_y$ follows from (76) as:

(79) $$c_t = (1 \;\; 1)^*.$$

Substitution of (77), (78) and (79) into (63) results in:

$$\underline{w} = \frac{(1\ \ 1)\sigma^{-2}\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}^{-1}\begin{pmatrix}\underline{y}_1 + \underline{y}_2 + \underline{y}_3\\ \underline{y}_2 + \underline{y}_4 + \underline{y}_5\end{pmatrix}}{\sqrt{(1\ \ 1)\sigma^{-2}\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}^{-1}\begin{pmatrix}1\\ 1\end{pmatrix}}}$$

which gives:

$$\underline{w} = \frac{\underline{y}_1 + 2\underline{y}_2 + \underline{y}_3 + \underline{y}_4 + \underline{y}_5}{\sigma 2\sqrt{2}} = \frac{\underline{t}_1 + \underline{t}_2}{\sigma 2\sqrt{2}}. \tag{80}$$

With this result and a computed value for k_α we are now able to execute the simple likelihood ratio test. The power of this test follows from:

$$\gamma = P(\underline{w} > k_\alpha \,|\, H_A)$$

or, since $\underline{w}$ is distributed under H_A as $\underline{w} \sim N(\frac{\nabla}{\sigma\sqrt{2}},1)$, from:

$$\begin{aligned}\gamma &= \int_{k_\alpha}^{\infty} (2\pi)^{-1/2}\ \exp[-\tfrac{1}{2}(w - \frac{\nabla}{\sigma\sqrt{2}})^2]\,dw\\ &= \int_{k_\alpha - \frac{\nabla}{\sigma\sqrt{2}}}^{\infty} (2\pi)^{-1/2}\ \exp[-\tfrac{1}{2}w^2]\,dw.\end{aligned} \tag{81}$$

Again we note that the power γ gets larger if ∇ gets larger or σ gets smaller. Thus the probability of detecting an error of size ∇ in the observable $\underline{y}_2$ gets larger if the size of the error gets larger or when the precision of the observables gets better. But, apart from these two effects, the power γ can also be shown to depend on the *design* or *structure* of the levelling network. In the case of Figure 2.11 the observable $\underline{y}_2$ occurs in both levelling loops. Hence we have two linearly independent condition equations with which a possible error in the observation can be detected. One would expect that the power decreases if $\underline{y}_2$ would occur in only one condition equation. In order to verify this we consider the situation as sketched in Figure 2.12. In this case $\underline{y}_2$ occurs only in one levelling loop.

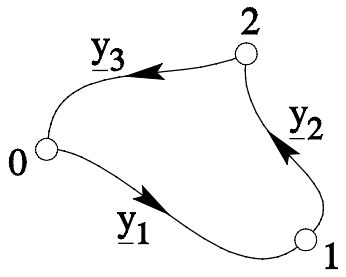


Figure 2.12: A levelling network with one loop.

Following the same kind of derivation as above one can show that the $\underline{w}$-teststatistic for detecting an error of size ∇ in the observation y_2 reads:

(82) $$\underline{w}^* = \frac{\underline{y}_1 + \underline{y}_2 + \underline{y}_3}{\sigma\sqrt{3}} . \qquad \text{(prove this yourself)}$$

Since $\underline{w}^*$ is distributed under the alternative hypothesis as $\underline{w}^* \sim N(\frac{\nabla}{\sigma\sqrt{3}},1)$, the power of the test becomes:

(83) $$\gamma^* = \int_{k_\alpha - \frac{\nabla}{\sigma\sqrt{3}}}^{\infty} (2\pi)^{-1/2} \exp[-\tfrac{1}{2}w^2]\,dw .$$

A comparison of (83) with (81) clearly shows that $\gamma > \gamma^*$. Thus a simple likelihood ratio test of size α based on the configuration of Figure 2.11 has a higher probability of detecting an error of size ∇ in the observation y_2, than a simple likelihood ratio test of size α based on the configuration of Figure 2.12. This conclusion shows how important it is when designing geodetic networks to make sure that an observation occurs in enough condition equations.

In the previous example we have seen that the power of the simple likelihood ratio test of size α depends on:

1) ∇, the size of the error;
2) σ, the precision of the observations; and
3) the design or structure of the network.

It is important to realize that this is not only valid for the case considered in the previous example, but that it is also valid for the likelihood ratio test (64) in general. This can be seen as follows. We know that the power of the simple likelihood ratio test of size α (64) can be computed as:

$$\gamma = \boldsymbol{P}(\underline{y} \in K \,|\, H_A) = \boldsymbol{P}(\underline{w} > k_\alpha \,|\, H_A).$$

Since $\underline{w}$ is distributed under H_A as $\underline{w} \sim N(\sqrt{c_t^{\,*} Q_t^{-1} c_t}\,\nabla, 1)$, it follows that:

$$\gamma = \int_{k_\alpha}^{\infty} (2\pi)^{-1/2} \exp[-\tfrac{1}{2}(w - \sqrt{c_t^{\,*} Q_t^{-1} c_t}\,\nabla)^2] \mathrm{d}w$$

or that:

(84) $$\gamma = \int_{k_\alpha - \sqrt{c_t^{\,*} Q_t^{-1} c_t}\,\nabla}^{\infty} (2\pi)^{-1/2} \exp[-\tfrac{1}{2} w^2] \,\mathrm{d}w.$$

This shows that γ, for a fixed size α, decreases if ∇ decreases or $c_t^{\,*} Q_t^{-1} c_t$ decreases. The precision of the observations and the structure of the network are contained in the scalar $c_t^{\,*} Q_t^{-1} c_t$. This can be seen if we write $c_t^{\,*} Q_t^{-1} c_t$ as:

(85) $$c_t^{\,*} Q_t^{-1} c_t = c_y^{\,*} B (B^{\,*} Q_y B)^{-1} B^{\,*} c_y.$$

The structure of the network is reflected in matrix B and the precision of the observations in matrix Q_y.

To conclude this section we have summarized in Table 2.4 the various steps which were followed in deriving the $\underline{w}$-teststatistic.

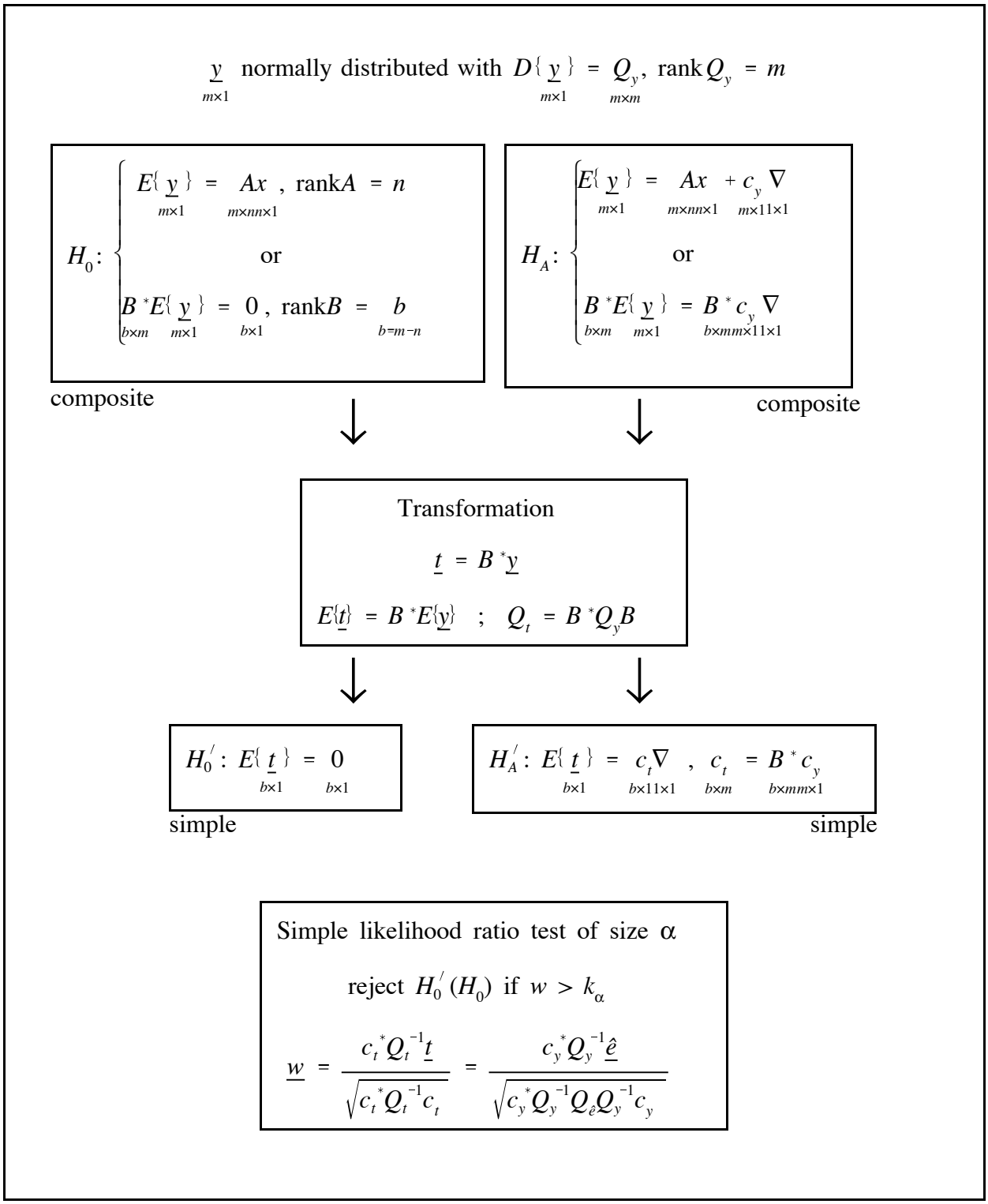


Table 2.4: The simple likelihood ratio test by the $\underline{w}$-teststatistic.

2.4 The $\underline{v}$-teststatistic

In the previous section it was shown that the simple likelihood ratio test of size α for testing:

(86) $$H_0': E\{\underline{t}\} = 0 \quad , \quad D\{\underline{t}\} = Q_t \quad \text{versus} \quad H_A': E\{\underline{t}\} = B^* c_y \nabla = c_t \nabla \quad , \quad D\{\underline{t}\} = Q_t$$

was given by:

(87) $$\text{reject } H_0' \text{ if } w > k_\alpha$$

with

(88) $$\underline{w} = \frac{c_t^* Q_t^{-1} \underline{t}}{\sqrt{c_t^* Q_t^{-1} c_t}} \ .$$

The simple hypotheses of (86) were obtained from the composite hypotheses:

(89) $$H_0: E\{\underline{y}\} = Ax, \ D\{\underline{y}\} = Q_y \quad \text{versus} \quad H_A: E\{\underline{y}\} = Ax + c_y \nabla, \ D\{\underline{y}\} = Q_y$$

through the transformation $\underline{t} = B^* \underline{y}$ with $B^* A = 0$. It was also shown that the $\underline{w}$-teststatistic (88) could be expressed in terms of quantities located in the sample space $\mathbb{R}^n$ as:

(90) $$\underline{w} = \frac{c_y^* Q_y^{-1} \underline{\hat{e}}}{\sqrt{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y}} \ .$$

Furthermore it was pointed out that rejection of H_0' implies rejection of H_0, but that acceptance of H_0' not necessarily implies that one should accept H_0. Finally an example was given, showing how the theory could be applied for detecting errors of known size in the observations.

In this section we consider a testing problem that, although mathematically equivalent to the above given testing problem, occurs when one wants to test the *significance of parameters*. We will derive the appropriate simple likelihood ratio test of size α and the corresponding $\underline{v}$-teststatistic. Let us assume as before that the $m \times 1$ random vector $\underline{y}$ is normally distributed with full rank variancematrix Q_y. The following two hypotheses are considered:

(91) $$H_0: \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{Ax} \ ; \ \underset{1\times nn\times 1}{b^* x} = \underset{1\times 1}{0} \qquad \text{versus} \qquad H_A: \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{Ax} \ ; \ \underset{1\times nn\times 1}{b^* x} = \underset{1\times 1}{\nabla} \ .$$

The two hypotheses H_0 and H_A differ in the sense that under H_0 it is assumed that the linear function of x, $b^* x$, is identical to zero, whereas under H_A it is assumed that this function is identical to the known scalar $\nabla \neq 0$ Thus, what we would like to find out is whether

b *x 0 or b *x = ∇. Note that H of (91) is of the mixed model type which was discussed in Chapter 5.3 of Adjustment theory. In order to be able to apply the theory of the previous section

we will first show how to rewrite the above H_0 and H_A in such a form that their structure is equivalent to the hypotheses H_0 and H_A of (89).

Consider the inhomogeneous equation:

(92) $$\underset{1\times nn\times 1}{\boldsymbol{b}^{*}}\underset{nn\times 1}{x} = \underset{1\times 1}{\nabla}.$$

We know that its solution is given by the sum of a particular solution and the homogeneous solution. A particular solution of (92) is:

(93) $$x_{\text{part.}} = \boldsymbol{b}(\boldsymbol{b}^{*}\boldsymbol{b})^{-1}\nabla.$$

In order to find the solution of the homogeneous equation:

(94) $$\underset{1\times nn\times 1}{\boldsymbol{b}^{*}}\underset{nn\times 1}{x} = \underset{1\times 1}{0}$$

we denote the $n\times(n-1)$ matrix of which the columnvectors are orthogonal to b by $b^{\perp}$. Then:

(95) $$\underset{1\times nn\times(n-1)}{\boldsymbol{b}^{*}}\ \underset{nn\times(n-1)}{\boldsymbol{b}^{\perp}} = \underset{1\times(n-1)}{0}\ .$$

With (95) the parametric representation of the homogeneous equation (94) becomes:

(96) $$\underset{n\times 1}{x} = \underset{n\times(n-1)(n-1)\times 1}{\boldsymbol{b}^{\perp}}\ \underset{}{\lambda}\ .$$

The general solution of the inhomogeneous equation (92) is therefore given by the sum of (93) and (96):

(97) $$x = \boldsymbol{b}^{\perp}\lambda + \boldsymbol{b}(\boldsymbol{b}^{*}\boldsymbol{b})^{-1}\nabla\ .$$

Now, since (96) is equivalent to (94) and (97) is equivalent to (92), the hypotheses of (91) may also be written as:

(98) $$H_0\colon\ E\{\underline{y}\} = \boldsymbol{A}\boldsymbol{b}^{\perp}\lambda \quad \text{versus} \quad H_{\boldsymbol{A}}\colon\ E\{\underline{y}\} = \boldsymbol{A}\boldsymbol{b}^{\perp}\lambda + \boldsymbol{A}\boldsymbol{b}(\boldsymbol{b}^{*}\boldsymbol{b})^{-1}\nabla.$$

Comparison of (98) with (89) shows the equivalence in structure. That is, the matrix $Ab^{\perp}$ of (98) plays the role of the matrix A in (89), and the vector $Ab(b^{*}b)^{-1}$ of (98) plays the role of the vector c_y in (89). Because of this equivalence in structure of the hypotheses, the simple likelihood ratio test for the present testing problem will have the same structure as the test developed in the previous section. The corresponding teststatistic, which will be denoted by $\underline{v}$, follows then if we replace c_y in (90) by $Ab(b^{*}b)^{-1}$:

(99) $$\underline{v} = \frac{(\boldsymbol{b}^{*}\boldsymbol{b})^{-1}\boldsymbol{b}^{*}\boldsymbol{A}^{*}\boldsymbol{Q}_{y}^{-1}\underline{\hat{\boldsymbol{e}}}}{\sqrt{(\boldsymbol{b}^{*}\boldsymbol{b})^{-1}\boldsymbol{b}^{*}\boldsymbol{A}^{*}\boldsymbol{Q}_{y}^{-1}\boldsymbol{Q}_{\hat{\boldsymbol{e}}}\boldsymbol{Q}_{y}^{-1}\boldsymbol{A}\boldsymbol{b}(\boldsymbol{b}^{*}\boldsymbol{b})^{-1}}}\ .$$

The least-squares residual vector $\underline{\hat{e}}$ and its variancematrix $Q_{\hat{e}}$ in formula (99) correspond to the least-squares solution of model H_0 in (91). Recall from Chapter 5.3 of Adjustment theory that the least-squares solution of the mixed model:

(100) $$E\{\underline{y}\} = Ax \quad , \quad b^* x = 0$$

reads:

(101) $$\begin{cases} \underline{\hat{x}}_A = Q_{\hat{x}_A} A^* Q_y^{-1} \underline{y} \quad , \quad Q_{\hat{x}_A} = (A^* Q_y^{-1} A)^{-1} \\ \underline{\hat{x}} = [I - Q_{\hat{x}_A} b (b^* Q_{\hat{x}_A} b)^{-1} b^*] \underline{\hat{x}}_A \\ \underline{\hat{y}} = A \underline{\hat{x}} \\ \underline{\hat{e}} = \underline{y} - \underline{\hat{y}}. \end{cases}$$

From this follows that:

(102) $$\begin{cases} \underline{\hat{e}} = \underline{y} - A\underline{\hat{x}}_A + AQ_{\hat{x}_A} b (b^* Q_{\hat{x}_A} b)^{-1} b^* \underline{\hat{x}}_A \text{ , and} \\ Q_{\hat{e}} = Q_y - AQ_{\hat{x}_A} A^* + AQ_{\hat{x}_A} b (b^* Q_{\hat{x}_A} b)^{-1} b^* Q_{\hat{x}_A} A^* \text{ .} \end{cases}$$

In (99) we need $A^* Q_y^{-1} \underline{\hat{e}}$ and $A^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} A$. With (102) this gives:

(103) $$\begin{cases} A^* Q_y^{-1} \underline{\hat{e}} = b (b^* Q_{\hat{x}_A} b)^{-1} b^* \underline{\hat{x}}_A \quad , \text{ and} \\ A^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} A = b (b^* Q_{\hat{x}_A} b)^{-1} b^* \text{ .} \end{cases}$$

Substitution of these results into (99) gives the following simple expression for $\underline{v}$-teststatistic:

(104) $$\boxed{\underline{v} = \frac{b^* \underline{\hat{x}}_A}{\sqrt{b^* Q_{\hat{x}_A} b}}} \text{ .}$$

The corresponding simple likelihood ratio test of size α for the testing problem (91) reads therefore:

(105) $$\text{reject } H_0 \text{ if } v > k_\alpha .$$

Note that this test is also intuitively appealing. For instance, if $b = (0 \ldots \underset{ith}{1} \ 0 \ldots 0)^*$ then $\underline{\hat{x}}_A = \underline{\hat{x}}_{i_A}$ and $b^* Q_{\hat{x}_A} b = \sigma^2_{i_A}$, and the $\underline{v}$-teststatistic reduces to $\underline{v} = \underline{\hat{x}}_{i_A} / \sigma_{i_A}$, implying that H_0 gets rejected if $\hat{x}_{i_A}$ is larger than k_α times σ_{i_A}. Since $\underline{\hat{x}}_A$ is distributed under H_A as $\underline{\hat{x}}_A \sim N(b^\perp \lambda + b(b^* b)^{-1} \nabla, Q_{\hat{x}_A})$ it follows that $\underline{v}$ is distributed under H_A as $\underline{v} \sim N(\frac{\nabla}{\sqrt{b^* Q_{\hat{x}_A} b}}, 1)$. The power of the test (105) reads therefore:

(106)
$$\gamma = \int_{k_a - \frac{\nabla}{\sqrt{b^* Q_{\hat{x}_A} b}}}^{\infty} (2\pi)^{-1/2} \exp[-\tfrac{1}{2}v^2]\mathrm{d}v.$$

Example 7

We assume to have measured the levelling network as shown in Figure 2.13.

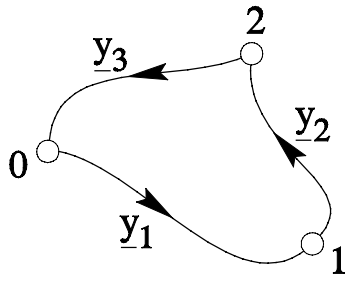


Figure 2.13: A levelling network.

We assume that the observables are normally distributed, uncorrelated and have equal variance σ^2. The following two hypotheses are considered:

(107)
$$H_0 : E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{pmatrix}\} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad (1 \ \ -1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$$

versus:

$$H_A : E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{pmatrix}\} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad (1 \ \ -1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \nabla.$$

Thus, what we would like to find out is whether the height difference between the points 1 and 2 equals zero or equals ∇. In order to compute the teststatistic $\underline{v}$ of (104) we need the vectors b and $\underline{\hat{x}}_A$, and the variancematrix $Q_{\hat{x}_A}$. According to (107) vector b reads $b = (1 \ \ -1)^*$. The vector $\underline{\hat{x}}_A$ and matrix $Q_{\hat{x}_A}$ follow from a least-squares adjustment of the model:

$$E\left\{\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \underline{y}_3\end{pmatrix}\right\} = \begin{pmatrix}1 & 0\\ -1 & 1\\ 0 & -1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}.$$

this gives:

$$Q_{\hat{x}_A} = (\sigma^{-2}\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix})^{-1} = \frac{1}{3}\sigma^2\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}$$

and

$$\underline{\hat{x}}_A = \begin{pmatrix}\underline{\hat{x}}_1\\ \underline{\hat{x}}_2\end{pmatrix}_A = Q_{\hat{x}_A}\begin{pmatrix}1 & -1 & 0\\ 0 & 1 & -1\end{pmatrix}\sigma^{-2}\begin{pmatrix}\underline{y}_1\\ \underline{y}_2\\ \underline{y}_3\end{pmatrix} = \frac{1}{3}\begin{pmatrix}2\underline{y}_1 - \underline{y}_2 - \underline{y}_3\\ \underline{y}_1 + \underline{y}_2 - 2\underline{y}_3\end{pmatrix}.$$

Substituting these results into (104) finally gives:

$$\underline{v} = \frac{\underline{y}_1 - 2\underline{y}_2 + \underline{y}_3}{\sigma\sqrt{6}}.$$

3 Testing of composite hypotheses

3.1 The generalized likelihood ratio test

In the previous chapter we considered testing a simple hypothesis against a simple alternative. We return now to the more general hypotheses-testing problem, that of testing composite hypotheses. We will assume that the $m\times 1$ vector random variable $\underset{\sim}{y}$ is distributed as:

(1) $$\underset{m\times 1}{\underset{\sim}{y}} \sim p_{\underset{\sim}{y}}(y|x) \text{ with } \underset{n\times 1}{x} \in \Phi$$

where Φ is a set of possible values the $n\times 1$ vector x may take. The following two composite hypotheses are considered:

(2) $$H_0 : \underset{n\times 1}{x} \in \Phi_0 \quad \text{versus} \quad H_A : \underset{n\times 1}{x} \in \Phi \setminus \Phi_0.$$

Where $\Phi\setminus\Phi_0$ is the subset of Φ that is complementary to Φ_0. Thus $\Phi\setminus\Phi_0 = \{x\in\Phi \,|\, x\notin\Phi_0\}$. We begin by discussing a general method of constructing a test for testing H_0 against H_A.

The generalized likelihood ratio test is defined by:

(3)
$$\text{reject } H_0 \text{ if } \frac{\max_{x\in\Phi_0} p_{\underset{\sim}{y}}(y|x)}{\max_{x\in\Phi} p_{\underset{\sim}{y}}(y|x)} < a$$

$$\text{accept } H_0 \text{ if } \frac{\max_{x\in\Phi_0} p_{\underset{\sim}{y}}(y|x)}{\max_{x\in\Phi} p_{\underset{\sim}{y}}(y|x)} > a$$

where a is a nonnegative constant.

Note that the ratio of (3) lies in the closed interval [0,1]. The ratio is larger or equal to zero since we have a ratio of nonnegative quantities, and the ratio is less than or equal to one since the maximum taken in the denominator is over a larger set of parameter values than that in the numerator; hence the denominator cannot be smaller than the numerator. Also note that although (3) resembles the simple likelihood ratio test (see (3) of Chapter 2), it does not reduce to the simple likelihood ratio test for $\Phi = \{x_0, x_A\}$. The simple likelihood ratio is namely not restricted to the closed interval [0,1]. The nonnegative constant a is taken to lie in the open interval (0,1). The value $a = 0$ is excluded, since we would like to reject H_0 if the ratio in (3) equals zero. And the value $a = 1$ is excluded, since would like to accept H_0 if the ratio in (3) equals one. The generalized likelihood ratio test makes good intuitive sense since the ratio in (3) will tend to be small when H_0 is not true, since then the denominator of the ratio tends to be larger than the numerator. In general (but not always), a generalized likelihood ratio test will be a good test. One possible drawback of the test is that it is sometimes difficult to find $\max p_{\underset{\sim}{y}}(y|x)$; another is that it can be difficult to find the probability distribution of the ratio which is required to evaluate the size α and the power γ of the test.

Example 1

Assume that the scalar random variable $\underline{y}$ has the following probability density function:

$$p_{\underline{y}}(y|x) = xe^{-yx},\ x > 0,\ y \geq 0. \tag{4}$$

The following two hypotheses are considered:

$$H_0 : x = x_0 \quad \text{versus} \quad H_A : x < x_0. \tag{5}$$

Thus in this case $\Phi = \{x \in \mathbb{R} | 0 < x \leq x_0\}$, $\Phi_0 = \{x_0\}$ and $\Phi \backslash \Phi_0 = \{x \in \mathbb{R} | 0 < x < x_0\}$. Note that H_0 is a simple hypothesis, whereas H_A is a composite hypothesis. In order to perform the generalized likelihood ratio test we need:

$$\max_{x = x_0} p_{\underline{y}}(y|x)$$

and

$$\max_{0 < x \leq x_0} p_{\underline{y}}(y|x).$$

The first maximum is trivial and reads:

$$\max_{x = x_0} p_{\underline{y}}(y|x) = x_0 e^{-yx_0}. \tag{6}$$

The second maximum is a bit more complicated to derive. Let us first consider the maximum problem without the restrictions on x:

$$\max_{x} p_{\underline{y}}(y|x) = \max_{x} xe^{-yx}. \tag{7}$$

From Elementary calculus you know that a necessary condition for $x_{\max}$ to be a solution of (7) is:

$$\frac{\mathrm{d}}{\mathrm{d}x} p_{\underline{y}}(y|x)_{x = x_{\max}} = e^{-yx_{\max}}(1 - yx_{\max}) = 0.$$

From this follows that:

$$x_{\max} = \frac{1}{y}. \tag{8}$$

From Elementary calculus you also know that $x_{\max}$ corresponds to a maximum if:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} p_{\underline{y}}(y|x)_{x = x_{\max}} = -ye^{-yx_{\max}}(2 - yx_{\max}) < 0.$$

Substitution of (8) shows that the inequality is indeed fulfilled. Thus, $x_{\max} = \frac{1}{y}$ maximizes: xe^{-yx} and

$$\max_x \; p_{\underline{y}}(y|x) \;=\; \frac{1}{y}e^{-1}.$$

Let us now consider the maximization problem with the restrictions on x:

$$\max_{0 < x \le x_0} \; p_{\underline{y}}(y|x) \;=\; \max_{0 < x \le x_0} \; xe^{-yx}.$$

We know that $x_{\max} = 1/y$ produces the maximum for the case without restrictions. But if $x_{\max} = 1/y \le x_0$, it will also produce the maximum for the case with the restrictions. Hence:

(9) $$\max_{0 < x \le x_0} \; p_{\underline{y}}(y|x) \;=\; \frac{1}{y}e^{-1} \quad \text{if} \quad \frac{1}{y} \le x_0.$$

Let us now consider what happens if $\frac{1}{y} > x_0$. Figure 3.1 shows a sketch of xe^{-yx} with its maximum at $x_{\max} = 1/y$:

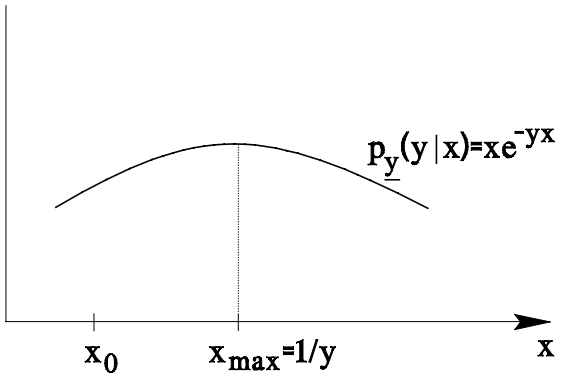


Figure 3.1 : Sketch of graph of xe^{-yx}.

This shows that for the case $0 < x \le x_0$ and $x_{\max} = \frac{1}{y} > x_0$ the maximum of $p_{\underline{y}}(y|x)$ is reached at x_0. Thus:

(10) $$\max_{0 < x \le x_0} \; p_{\underline{y}}(y|x) \;=\; x_0 e^{-yx_0} \quad \text{if} \quad \frac{1}{y} > x_0.$$

From (6), (9) and (10) follows therefore:

(11) $$\frac{\max\limits_{x = x_0} p_{\underline{y}}(y|x)}{\max\limits_{0 < x \le x_0} p_{\underline{y}}(y|x)} \;=\; \begin{cases} 1 & \text{if} \quad yx_0 < 1 \\ yx_0 e^{-(yx_0 - 1)} & \text{if} \quad yx_0 \ge 1. \end{cases}$$

Since $a \in (0,1)$ we may restrict ourselves to the second equation of (11). This gives with (3) the generalized likelihood ratio test:

(12) $$\text{reject } H_0 \text{ if} \quad yx_0 \ge 1 \quad \text{and} \quad yx_0 e^{-(yx_0 - 1)} < a.$$

Write:

(13) $$z = yx_0,$$

and note that the function $ze^{-(z-1)}$ has its maximum at $z = 1$ (prove this yourself). Hence $z \geq 1$ and $ze^{-(z-1)} < a$ if and only if $z > k$, where k is a constant satisfying $k > 1$ (see Figure 3.2). We see therefore that the generalized likelihood ratio test reduces to:

$$\text{reject } H_0 \text{ if } \quad yx_0 > k, \text{ where } k > 1$$

or

(14) $$\text{reject } H_0 \text{ if } \quad y > k_\alpha, \text{ where } k_\alpha > x_0^{-1}.$$

Compare this with Example 4 of Chapter 1.

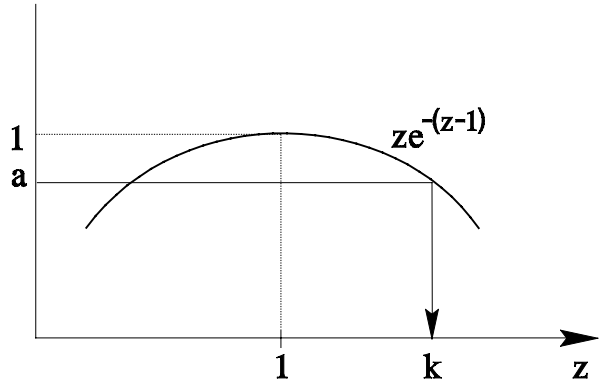


Figure 3.2 : Graph of $ze^{-(z-1)}$.

Example 2

Assume that the scalar random variable $\underline{y}$ is normally distributed with variance σ^2. The following two hypotheses are considered:

(15) $$H_0 : E\{\underline{y}\} = x_0 \quad \text{versus} \quad H_A : E\{\underline{y}\} > x_0.$$

The numerator of the likelihood ratio reads:

(16a) $$\max_{x = x_0} p_{\underline{y}}(y|x) = p_{\underline{y}}(y|x_0) = (2\pi)^{-1/2}(\sigma^2)^{-1/2}\exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y - x_0)^2].$$

The denominator of the likelihood ratio is given as:

(16b) $$\max_{x \geq x_0} p_{\underline{y}}(y|x) = \max_{x \geq x_0} (2\pi)^{-1/2}(\sigma^2)^{-1/2}\exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y - x)^2].$$

The solution to this maximization problem is given by (see Figure 3.3):

(17)
$$\max_{x \geq x_0} p_{\underline{y}}(y|x) = \begin{cases} (2\pi)^{-1/2}(\sigma^2)^{-1/2} & \text{if} \quad y \geq x_0 \\ (2\pi)^{-1/2}(\sigma^2)^{-1/2} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_0)^2] & \text{if} \quad y < x_0. \end{cases}$$

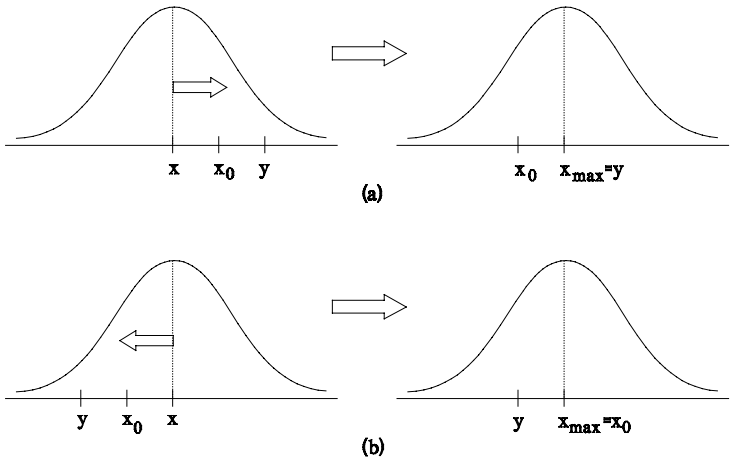


Figure 3.3: (a) Maximization of $p_{\underline{y}}(y|x)$ for $y \geq x_0$, $x \geq x_0$
(b) Maximization of $p_{\underline{y}}(y|x)$ for $y < x_0$, $x \geq x_0$.

From (16a) and (17) follows:

(18)
$$\frac{\max_{x = x_0} p_{\underline{y}}(y|x)}{\max_{x \geq x_0} p_{\underline{y}}(y|x)} = \begin{cases} \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_0)^2] & \text{if} \quad y \geq x_0 \\ 1 & \text{if} \quad y < x_0. \end{cases}$$

This gives with (3) the generalized likelihood ratio test:

(19)
$$\text{reject } H_0 \text{ if} \quad y \geq x_0 \quad \text{and} \quad \exp[-\frac{1}{2}\frac{1}{\sigma^2}(y-x_0)^2] < a.$$

The second inequality can be rewritten as:

$$\left(\frac{y-x_0}{\sigma}\right)^2 > \ln a^{-2}.$$

This, together with the first inequality of (19) gives:

$$\frac{y-x_0}{\sigma} \geq 0 \quad \text{and} \quad \frac{y-x_0}{\sigma} > (\ln a^{-2})^{1/2}$$

or simply $(y-x_0)/\sigma > k_\alpha$ with $k_\alpha > 0$. The generalized likelihood ratio test reduces therefore to:

$$\text{reject } H_0 \text{ if } \quad \frac{y - x_0}{\sigma} > k_\alpha \text{ with } k_\alpha > 0.$$

Note that $(y - x_0)/\sigma$ has a standard normal distribution under H_0. Compare the above result with Example 3 of Chapter 1.

Example 3

Again it is assumed that the scalar random variable $\underline{y}$ is normally distributed with variance σ^2. The following two hypotheses are considered:

(21) $$H_0 : E\{\underline{y}\} = x_0 \quad \text{versus} \quad H_A : E\{\underline{y}\} \neq x_0.$$

Again the numerator of the likelihood ratio reads:

(22) $$\max_{x = x_0} p_{\underline{y}}(y|x) = (2\pi)^{-1/2}(\sigma^2)^{-1/2} \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y - x_0)^2].$$

The denominator follows as:

(23) $$\max_{x} p_{\underline{y}}(y|x) = \max_{x} (2\pi)^{-1/2}(\sigma^2)^{-1/2} \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y - x)^2] = (2\pi)^{-1/2}(\sigma^2)^{-1/2}.$$

From (22) and (23) follows therefore that:

(24) $$\frac{\max_{x = x_0} p_{\underline{y}}(y|x)}{\max_{x} p_{\underline{y}}(y|x)} = \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y - x_0)^2].$$

With (3) this gives:

$$\text{reject } H_0 \text{ if } \quad \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y - x_0)^2] < a.$$

If we denote $\ln a^{-2}$ by k_α, this reduces to:

$$\text{reject } H_0 \text{ if } \quad \left(\frac{y - x_0}{\sigma}\right)^2 > k_\alpha \text{ with } k_\alpha > 0$$

or to:

(25) $$\text{reject } H_0 \text{ if } \quad \frac{y - x_0}{\sigma} < -k_\alpha^{1/2} \quad \text{or} \quad \frac{y - x_0}{\sigma} > +k_\alpha^{1/2}.$$

Compare this result with Example 2 of Chapter 1.

Example 4

It is assumed that the $mx1$ vector random variable $\underline{y}$ has a probability density function:

(26) $$p_{\underline{y}}(y|x,\sigma^2) = (2\pi)^{-m/2}(\sigma^2)^{-m/2}\ \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y-ex)^*(y-ex)]$$

with $\underset{m\times 1}{e} = (1,1,\cdots,1,1)^*$. The following two hypotheses are considered:

(27) $$H_0\text{:}\quad x = x_0\ ,\ \sigma^2 = \sigma_0^2 \quad \text{versus} \quad H_A\text{:}\quad x = x_0\ ,\ \sigma^2 > \sigma_0^2.$$

The numerator of the likelihood ratio reads:

(28) $$\max_{x = x_0,\,\sigma^2 = \sigma_0^2} p_{\underline{y}}(y|x,\sigma^2) = (2\pi)^{-m/2}(\sigma_0^2)^{-m/2}\ \exp[-\tfrac{1}{2}\tfrac{1}{\sigma_0^2}(y-ex_0)^*(y-ex_0)].$$

The denominator follows from:

(29) $$\max_{x = x_0,\,\sigma^2 \geq \sigma_0^2} p_{\underline{y}}(y|x,\sigma^2) = \max_{\sigma^2 \geq \sigma_0^2} (2\pi)^{-m/2}(\sigma^2)^{-m/2}\ \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2}(y-ex_0)^*(y-ex_0)].$$

Let us first consider the unrestricted maximum of $p_{\underline{y}}(y|x_0,\sigma^2)$. The following holds (prove yourself):

(30) $$\frac{\mathrm{d}}{\mathrm{d}\sigma^2}\ p_{\underline{y}}(y|x_0,\sigma^2) = p_{\underline{y}}(y|x_0,\sigma^2)\ \cdot\ \tfrac{1}{2}\sigma^{-4}\ \cdot\ \{-m\sigma^2 + (y-ex_0)^*(y-ex_0)\}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}\sigma^{2^2}}\ p_{\underline{y}}(y|x_0,\sigma^2) = p_{\underline{y}}(y|x_0,\sigma^2)\ \cdot\ \sigma^{-8}\ \cdot\ \{[(\frac{m}{2}+1)\sigma^2 - \tfrac{1}{2}(y-ex_0)^*(y-ex_0)]^2 - (\frac{m}{2}+1)\sigma^4\}.$$

(31)

Setting (30) equal to zero gives:

(32) $$\sigma^2_{\max} = \tfrac{1}{m}(y-ex_0)^*(y-ex_0).$$

Substitution of (32) into (31) shows that the second derivative of $p_{\underline{y}}(y|x_0,\sigma^2)$ at $\sigma^2 = \sigma^2_{\max}$ is negative; hence $\sigma^2_{\max}$ maximizes $p_{\underline{y}}(y|x_0,\sigma^2)$. This shows that:

(33) $$\max_{x = x_0,\,\sigma^2 \geq \sigma_0^2} p_{\underline{y}}(y|x,\sigma^2) = (2\pi)^{-m/2}(\sigma^2_{\max})^{-m/2}\ \exp[-\tfrac{1}{2}\tfrac{1}{\sigma^2_{\max}}(y-ex_0)^*(y-ex_0)] \quad \text{if} \quad \sigma^2_{\max} \geq \sigma_0^2.$$

For the case that $\sigma^2_{\max} \leq \sigma_0^2$ it follows from Figure 3.4 that:

$$(34)\qquad \max_{x=x_0,\,\sigma^2\geq\sigma_0^2} p_{\underline{y}}(y\,|\,x,\sigma^2) \;=\; (2\pi)^{-m/2}(\sigma_0^2)^{-m/2}\,\exp[-\tfrac{1}{2}\tfrac{1}{\sigma_0^2}(y-ex_0)^*(y-ex_0)] \quad \text{if} \quad \sigma_{\max}^2 \leq \sigma_0^2.$$

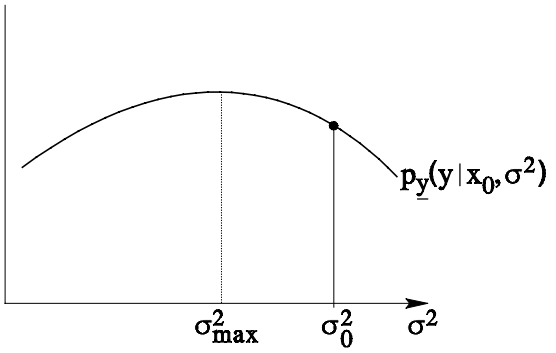


Figure 3.4: Graph of $p_{\underline{y}}(y\,|\,x_0,\sigma^2)$.

We may now collect our results. From (28), (32), (33) and (34) follows then that:

$$(35)\qquad \frac{\max\limits_{x=x_0,\,\sigma^2=\sigma_0^2} p_{\underline{y}}(y\,|\,x,\sigma^2)}{\max\limits_{x=x_0,\,\sigma^2\geq\sigma_0^2} p_{\underline{y}}(y\,|\,x,\sigma^2)} \;=\; \begin{cases} \left(\dfrac{\sigma_{\max}^2}{\sigma_0^2}\right)^{\frac{m}{2}} \exp[-\tfrac{1}{2}m\left(\dfrac{\sigma_{\max}^2}{\sigma_0^2}-1\right)] & \text{if} \quad \sigma_{\max}^2 \geq \sigma_0^2 \\ 1 & \text{if} \quad \sigma_{\max}^2 \leq \sigma_0^2 \end{cases}.$$

Hence, the generalized likelihood ratio test becomes:

$$(36)\qquad \text{reject } H_0 \text{ if} \quad \sigma_{\max}^2 \geq \sigma_0^2 \quad \text{and} \quad \left(\frac{\sigma_{\max}^2}{\sigma_0^2}\right)^{\frac{m}{2}} \exp[-\tfrac{1}{2}m\left(\frac{\sigma_{\max}^2}{\sigma_0^2}-1\right)] < a\,.$$

Write:

$$(37)\qquad z \;=\; \frac{\sigma_{\max}^2}{\sigma_0^2}$$

and note that the function $z^{\frac{m}{2}}\exp[-\tfrac{1}{2}m(z-1)]$ has its maximum at $z=1$ (prove this yourself). Hence, $z\geq 1$ and $z^{\frac{m}{2}}\exp[-\tfrac{1}{2}m(z-1)]<a$ if and only if $z>k_\alpha$, where k_α is a constant satisfying $k_\alpha\geq 1$ (see Figure 3.5). We see therefore that the generalized likelihood ratio test (36) reduces to:

$$\text{reject } H_0 \text{ if} \quad \frac{\sigma_{\max}^2}{\sigma_0^2} > k_\alpha, \text{ where } k_\alpha \geq 1$$

or with (32) to:

$$(38)\qquad \text{reject } H_0 \text{ if} \quad \frac{(y-ex_0)^*(y-ex_0)}{m\sigma_0^2} > k_\alpha, \text{ where } k_\alpha \geq 1.$$

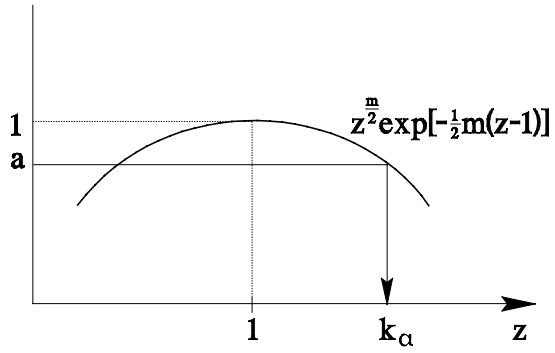


Figure 3.5: The graph of $z^{\frac{m}{2}}\exp[-\frac{1}{2}m(z-1)]$.

Compare this result with Example 1 of Chapter 2. Note that $(\underline{y}-ex_0)^*(\underline{y}-ex_0)/\sigma_0^2$ is distributed under H_0 as a central χ^2-distribution with m degrees of freedom.

Example 5

It is assumed that $\underline{y}$ has the same probalility density function as in the previous example. The following two hypotheses are considered:

(39) $$H_0: \; x = x_0 \;,\; \sigma^2 = \sigma_0^2 \quad \text{versus} \quad H_A: \; x = x_0 \;,\; \sigma^2 \neq \sigma_0^2.$$

The numerator of the likelihood ratio is identical to (28) and reads:

(40) $$\max_{x = x_0,\, \sigma^2 = \sigma_0^2} p_{\underline{y}}(y\,|\,x,\sigma^2) = (2\pi)^{-m/2}(\sigma_0^2)^{-m/2}\,\exp[-\tfrac{1}{2}\tfrac{1}{\sigma_0^2}(y-ex_0)^*(y-ex_0)].$$

The denominator of the likelihood ratio is given by the unrestricted maximum of $p_{\underline{y}}(y\,|\,x_0,\sigma^2)$. From the previous example we know that:

(41) $$\sigma_{\max}^2 = \frac{(y-ex_0)^*(y-ex_0)}{m}$$

maximizes $p_{\underline{y}}(y\,|\,x_0,\sigma^2)$: Hence:

(42) $$\max_{x = x_0,\, \sigma^2} p_{\underline{y}}(y\,|\,x,\sigma^2) = (2\pi)^{-m/2}(\sigma_{\max}^2)^{-m/2}\,\exp[-\tfrac{1}{2}\tfrac{1}{\sigma_{\max}^2}(y-ex_0)^*(y-ex_0)].$$

From (40), (41) and (42) follows then:

(43) $$\frac{\max_{x = x_0,\, \sigma^2 = \sigma_0^2} p_{\underline{y}}(y\,|\,x,\sigma^2)}{\max_{x = x_0,\, \sigma^2} p_{\underline{y}}(y\,|\,x,\sigma^2)} = \left(\frac{\sigma_{\max}^2}{\sigma_0^2}\right)^{\frac{m}{2}} \exp\left[-\frac{1}{2}m\left(\frac{\sigma_{\max}^2}{\sigma_0^2}-1\right)\right].$$

The generalized likelihood ratio test reads therefore:

(44) $$\text{reject } H_0 \text{ if } \quad z^{\frac{m}{2}} \exp[-\tfrac{1}{2}m(z-1)] < a$$

with

(45) $$z = \frac{\sigma^2_{\max}}{\sigma^2_0}.$$

Since the function in (44) has a maximum of 1 at $z = 1$, it follows that the generalized likelihood ratio test may also be written as:

$$\text{reject } H_0 \text{ if } \quad 0 \le z \le k_1 \quad \text{or} \quad z > k_2, \text{ where } k_1 \le 1 \text{ and } k_2 \ge 1.$$

With (41) and (45) this gives:

(46) $$\text{reject } H_0 \text{ if } \quad 0 \le \frac{(y-ex_0)^*(y-ex_0)}{m\sigma^2_0} < k_1 \quad \text{or} \quad \frac{(y-ex_0)^*(y-ex_0)}{m\sigma^2_0} > k_2$$

where $k_1 \le 1$ and $k_2 \ge 1$.

A sketch of the critical region K of this test is given in Figure 3.6. Compare this figure with Figure 2.3 of Chapter 2.

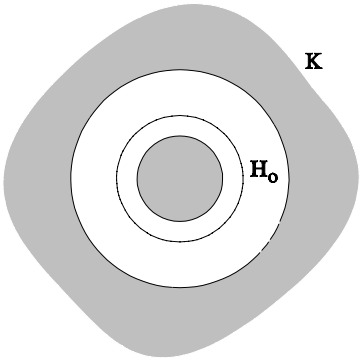


Figure 3.6: Critical region K for test (46).

3.2 Uniformly most powerful tests

Recall that the power γ of a test is defined as the probability of correctly rejecting H_0. In case of a simple alternative hypothesis H_A the power can be calculated as (see (28) of Section 2.2):

(47) $$\gamma = p(\underline{y} \in K | H_A) = \int_K p_{\underline{y}}(y | x_A)\mathrm{d}y.$$

For more general classes of alternative hypotheses, the power will depend on the particular alternative value of the parameter x being considered. In order to determine how good the chosen

test may be, compared to a competing test, it is therefore necessary to compare the power for all possible alternative values of x rather than for just one alternative value as in (47). For this purpose, it is necessary to consider the calculation of the power as a function of x. This leads to the concept of the powerfunction $\gamma(x)$.

The *powerfunction* $\gamma(x)$ of a test is the function of the parameter x that gives the probability that the sample or observation will fall in the critical region of the test when x is the true value of the parameter.

The powerfunction $\gamma(x)$ can be calculated as:

(48) $$\boxed{\gamma(x) = \int_K p_{\underline{y}}(y|x)\mathrm{d}y} \ .$$

In terms of the powerfunction we now define an optimum property that a test may possess. Let $\gamma(x)$ be the powerfunction corresponding to the critical region K, and let $\gamma^*(x)$ be the powerfunction corresponding to the critical region K^*. A test of $H_0: x\in\Phi_0$ versus $H_A: x\in\Phi\backslash\Phi_0$, with critical region K, is defined to be a *uniformly most powerful test of size* α if and only if:

(i) $\alpha = \max_{x\in\Phi_0} \gamma(x)$, and

(ii) $\gamma(x) \geq \gamma^*(x)$, $\forall x\in\Phi\backslash\Phi_0$ and for any test with critical region K^* and size $\alpha = \max_{x\in\Phi_0} \gamma^*(x)$.

The adverb *uniformly* in the above definition refers to *all* alternative x values. As we will see, a uniformly most powerful test does not exist for all testing problems, but when one does exist, we can see that it is quite a nice test since among all test of size α it has the greatest chance of rejecting H_0 whenever it should.

In some cases when H_0 is simple and H_A is composite it is possible to find a uniformly most powerful test with the help of the Neyman-Pearson theorem. Assume:

(49) $$H_0: \quad x = x_0 \qquad \text{versus} \qquad H_A: \quad x\in\Phi\backslash\{x_0\}.$$

Now choose a particular x say x_1 from $\Phi\backslash\{x_0\}$. Then according to the Neyman-Pearson theorem the simple likelihood ratio test:

(50) $$\text{reject } H_0 \text{ if } \quad \frac{p_{\underline{y}}(y|x_0)}{p_{\underline{y}}(y|x_1)} < a$$

is a most powerful test for testing:

(51) $$H_0: \quad x = x_0 \quad \text{versus} \quad H_A': \quad x = x_1 .$$

Now, if it is possible to show that the same test (50) follows when $x_1 \in \Phi \setminus \{x_0\}$ is replaced by another arbitrary parameter from $\Phi \setminus \{x_0\}$, then this test is a uniformly most powerful test. If this is not possible, then no uniformly most powerful test for testing (49) exists.

Example 6

Let the probability density function of $\underline{y}$ be given as:

(52) $$p_{\underline{y}}(y|x) = xe^{-yx} , \; x > 0 , \; y \geq 0 .$$

From Example 3 of Chapter 2 we know that the simple likelihood ratio test for testing:

(53) $$H_0: \quad x = x_0 \quad \text{versus} \quad H_A': \quad x = x_A < x_0$$

reads:

(54) $$\text{reject } H_0 \text{ if } \quad \frac{x_0}{x_A} \exp[-y(x_0 - x_A)] < a .$$

The inequality can also be written as:

(55) $$y(x_0 - x_A) > \ln[\frac{x_A}{x_0} a]^{-1} .$$

And this inequality reduces for all $x_A < x_0$ to:

(56) $$y > \frac{\ln[\frac{x_A}{x_0} a]^{-1}}{x_0 - x_A} .$$

Thus the test:

(57) $$\text{reject } H_0 \text{ if } \quad y > k_\alpha$$

is a uniformly most powerful test for testing:

(58) $$H_0: \; x = x_0 \quad \text{versus} \quad H_A: \; x < x_0 .$$

On the other hand, inequality (55) reduces for all $x_A > x_0$ to:

(59)
$$y < \frac{\ln[\frac{x_A}{x_0}a]^{-1}}{x_0 - x_A}.$$

This shows that the test:

(60)
$$\text{reject } H_0 \text{ if } \quad y < k_\alpha$$

is a uniformly most powerful test for testing:

(61)
$$H_0: \quad x = x_0 \qquad \text{versus} \qquad H_A: \quad x > x_0.$$

Since the two tests (57) and (60) which correspond to (58) and (61) respectively, are not identical, it follows that *no* uniformly most powerful test exists for testing:

(62)
$$H_0: \quad x = x_0 \qquad \text{versus} \qquad H_A: \quad x \neq x_0.$$

Thus the generalized likelihood ratio test for testing (62) cannot be a uniformly most powerful test.

Example 7

Assume that the scalar random variable $\underline{y}$ has a $\chi^2(m,\lambda)$ distribution. Its probability density function reads then:

(63)
$$p_{\underline{y}}(y|\lambda) = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{(\frac{\lambda}{2})^j y^{\frac{m}{2}+j-1} \exp[-\frac{1}{2}y]}{j!\, 2^{\frac{m}{2}+j}\, \Gamma(\frac{m}{2}+j)}, \quad 0 < y < \infty.$$

The following two hypotheses are considered:

(64)
$$H_0: \quad \lambda = 0 \qquad \text{versus} \qquad H_A: \quad \lambda > 0.$$

In order to derive a uniformly most powerful test, we first consider the following two simple hypotheses:

(65)
$$H_0: \quad \lambda = 0 \qquad \text{versus} \qquad H_A': \quad \lambda = \lambda_A > 0.$$

The simple likelihood ratio reads then:

$$\frac{p_{\underline{y}}(y|0)}{p_{\underline{y}}(y|\lambda_A)} = \frac{\dfrac{y^{\frac{m}{2}-1}\exp[-\frac{1}{2}y]}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}}{e^{-\frac{\lambda_A}{2}}\sum_{j=0}^{\infty}\dfrac{(\frac{\lambda_A}{2})^j y^{\frac{m}{2}+j-1}\exp[-\frac{1}{2}y]}{j!\,2^{\frac{m}{2}+j}\Gamma(\frac{m}{2}+j)}}$$

or

$$\frac{p_{\underline{y}}(y|0)}{p_{\underline{y}}(y|\lambda_A)} = \frac{1}{e^{-\frac{\lambda_A}{2}}\sum_{j=0}^{\infty}\dfrac{(\frac{\lambda_A}{2})^j y^j\,\Gamma(\frac{m}{2})}{j!\,2^j\Gamma(\frac{m}{2}+j)}}.$$

This function is clearly a monotone decreasing function for y. From this follows that $p_{\underline{y}}(y|0)\,/\,p_{\underline{y}}(y|\lambda_A) < a$ if and only if $y > k_\alpha$, where k_α is some positive constant. Hence, the most powerful test for testing (65) is:

(66) $$\text{reject } H_0 \text{ if } \quad y > k_\alpha \quad , \quad \text{where} \quad k_\alpha > 0.$$

Since the inequality $y > k_\alpha$ is independent of $\lambda_A > 0$, it follows that (66) is the uniformly most powerful test for testing (64).

Example 8

Assume that the scalar random variable $\underline{y}$ has an $F(m, n, \lambda)$ distribution. Its probability density function reads then:

(67) $$p_{\underline{y}}(y|\lambda) = e^{-\frac{\lambda}{2}}\sum_{j=0}^{\infty}\frac{(\frac{\lambda}{2})^j y^{\frac{m}{2}+j-1}(\frac{m}{n})^{\frac{m}{2}+j}\Gamma(\frac{m+n}{2}+j)}{j!(1+\frac{m}{n}y)^{\frac{1}{2}(m+n)+j}\Gamma(\frac{n}{2})\Gamma(\frac{m}{2}+j)} \quad 0 < y < \infty.$$

The following two hypotheses are considered:

(68) $$H_0: \quad \lambda = 0 \quad \text{versus} \quad H_A: \quad \lambda > 0.$$

In order to derive a uniformly most powerful test, we first consider the following two simple hypotheses:

(69) $$H_0: \quad \lambda = 0 \quad \text{versus} \quad H_A': \quad \lambda = \lambda_A > 0.$$

The simple likelihood ratio reads then:

$$\frac{p_{\underline{y}}(y|0)}{p_{\underline{y}}(y|\lambda_A)} = \frac{\dfrac{y^{\frac{m}{2}-1}(\frac{m}{n})^{\frac{m}{2}}\Gamma(\frac{m+n}{2})}{(1+\frac{m}{n}y)^{\frac{1}{2}(m+n)}\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}}{e^{-\frac{\lambda_A}{2}}\sum_{j=0}^{\infty}\dfrac{(\frac{\lambda_A}{2})^j y^{\frac{m}{2}+j-1}(\frac{m}{n})^{\frac{m}{2}+j}\Gamma(\frac{m+n}{2}+j)}{j!(1+\frac{m}{n}y)^{\frac{1}{2}(m+n)+j}\Gamma(\frac{n}{2})\Gamma(\frac{m}{2}+j)}}$$

or

$$\frac{p_{\underline{y}}(y|0)}{p_{\underline{y}}(y|\lambda_A)} = \frac{1}{e^{-\frac{\lambda_A}{2}}\sum_{j=0}^{\infty}\dfrac{(\frac{\lambda_A}{2})^j y^j(\frac{m}{2})^j\,\Gamma(\frac{m}{2})\Gamma(\frac{m+n}{2}+j)}{j!(1+\frac{m}{n}y)^j\,\Gamma(\frac{m+n}{2})\Gamma(\frac{m}{2}+j)}}$$

or

$$\frac{p_{\underline{y}}(y|0)}{p_{\underline{y}}(y|\lambda_A)} = \frac{1}{e^{-\frac{\lambda_A}{2}}\sum_{j=0}^{\infty}\dfrac{(\frac{\lambda_A}{2})^j(\frac{m}{n})^j\Gamma(\frac{m}{2})\Gamma(\frac{m+n}{2}+j)}{j!(y^{-1}+\frac{m}{n})^j\Gamma(\frac{m+n}{2})\Gamma(\frac{m}{2}+j)}}.$$

This function is clearly a monotone decreasing function of y. From this follows that $p_{\underline{y}}(y|0)\,/\,p_{\underline{y}}(y|\lambda_A) < a$ if and only if $y > k_\alpha$, where k_α is some positive constant. Hence the most powerful test for testing (69) is:

(70) $$\text{reject} \quad H_0 \text{ if} \quad y > k_\alpha \quad , \quad \text{where } k_\alpha > 0.$$

Since the inequality $y > k_\alpha$ is independent of $\lambda_A > 0$, it follows that (70) is the uniformly most powerful test for testing (68).

In the above Example 6 we discussed a testing problem for which no uniformly most powerful test exists. Unfortunately there are many such hypothesis-testing problems for which no uniformly most powerful test exists. In fact, this is the case for all testing problems that will be considered in the remaining part of these lecture notes. The reason why a uniformly most powerful test does not exists for a particular testing problem is usually due to the fact that one considers a class of critical regions which is too large. The idea is therefore to restrict the class of critical regions and to search for a uniformly most powerful test in this restricted class. One

way to restrict the class of critical regions is based on the *principle of invariance*. The following example should make this idea clear.

Example 9

It is assumed that the $m\times 1$ random vector $\underset{m\times 1}{\underline{y}}$ is distributed under H_0 and H_A as:

$$H_0: \quad \underset{m\times 1}{\underline{y}} \sim N(\underset{m\times 1}{0}, \underset{m\times m}{I_m}) \qquad \text{versus} \qquad H_A: \quad \underset{m\times 1}{\underline{y}} \sim N(\underset{m\times 1}{x\neq 0}, \underset{m\times m}{I_m}). \tag{71}$$

Now consider the invertible linear transformation:

$$\underset{m\times 1}{\underline{v}} = \underset{m\times m}{\boldsymbol{R}} \; \underset{m\times 1}{\underline{y}} \qquad (\boldsymbol{R}^{-1} \text{ exists}). \tag{72}$$

Then in terms of $\underline{v}$ the hypotheses H_0 and H_A become:

$$H_0: \quad \underset{m\times 1}{\underline{v}} \sim N(\underset{m\times 1}{0}, \underset{m\times m}{\boldsymbol{R}\boldsymbol{R}^*}) \qquad \text{versus} \qquad H_A: \quad \underline{v} \sim N(\underset{m\times 1}{\boldsymbol{R}x\neq 0}, \underset{m\times m}{\boldsymbol{R}\boldsymbol{R}^*}). \tag{73}$$

Now we note that if $RR^* = I_m$ (R is an orthogonal matrix), then (73) can be written as:

$$H_0: \quad \underset{m\times 1}{\underline{v}} \sim N(\underset{m\times 1}{0}, \underset{m\times m}{I_m}) \qquad \text{versus} \qquad H_A: \quad \underset{m\times 1}{\underline{v}} \sim N(\underset{m\times 1}{\boldsymbol{R}x\neq 0}, \underset{m\times m}{I_m}). \tag{74}$$

Comparison of (74) with (71) shows the equivalence of the two testing problems. We say that the testing problem (71) is invariant under the transformation (72) if matrix R is orthogonal ($RR^* = I_m$). Because of the equivalence of (71) and (74), we would of course also like to have the same test for both problems. This implies that if K is the critical region for the test of (71), K should also be the critical region for the test of (74). Thus, if $y \in K$ then also $v \in K$ and if $y \notin K$ then also $v \notin K$. But since $v = Ry$, this implies that K should be invariant for this transformation. From (72) follows with $RR^* = I$ or $R^* = R^{-1}$ that:

$$v^*v = y^*\boldsymbol{R}^*\boldsymbol{R}y = y^*\boldsymbol{R}^{-1}\boldsymbol{R}y = y^*y. \tag{75}$$

From this follows that the critical region K must have a (hyper) spherical shape with its centre at 0. Hence this leaves us with the following two possibilities:

$$\begin{cases} K_1 = \{y \in \mathbb{R}^m \mid y^*y < c_1\}, \text{ or} \\ K_2 = \{y \in \mathbb{R}^m \mid y^*y > c_2\}. \end{cases} \tag{76}$$

Within this restricted class of critical regions we may now try to find a uniformly most powerful test. If it exists, it is called the uniformly most powerful invariant test. The scalar random variable $\underline{y}^*\underline{y}$ has a χ^2-distribution and is distributed under H_0 and H_A as:

(77) $$H_0: \quad \underline{y}^* \underline{y} \sim \chi^2(m, 0) \quad \text{versus} \quad H_A: \quad \underline{y}^* \underline{y} \sim \chi^2(m, \lambda).$$

From Example 7 we know that the critical region K_2 of (76) gives the most power. Hence, the uniformly most powerful invariant test of (71) reads:

(78) $$\text{reject } H_0 \text{ if } \quad y^* y > k_\alpha.$$

Now let us have a look at the generalized likelihood ratio test of (71). It is given as:

$$\frac{\max\limits_{x=0} p_{\underline{y}}(y|x)}{\max\limits_{x} p_{\underline{y}}(y|x)} = \frac{(2\pi)^{-\frac{m}{2}} \exp[-\frac{1}{2} y^* y]}{(2\pi)^{-\frac{m}{2}}} < a.$$

But this inequality reduces to the same inequality of (78). We have reached therefore the important conclusion that the generalized likelihood ratio test of (71) is a uniformly most powerful invariant test.

Without proof we now state that all *generalized likelihood ratio tests* of the next chapters are in fact *uniformly most powerful invariant tests* (for a proof see (Arnold, 1981)).

4 Hypothesis testing in linear models

4.1 The models of condition and observation equations

In this chapter we will derive and discuss the generalized likelihood ratio test for the important case of linear models. In this section we consider the linear models of both *condition equations* and *observation equations*.

We assume that the $m\times1$ vector of observables $\underline{y}$ is normally distributed with known variancematrix Q_y:

$$\underset{m\times1}{\underline{y}} \sim N(\underset{m\times1}{E\{\underline{y}\}}, \underset{m\times m}{Q_y}). \tag{1}$$

It is assumed that matrix Q_y is of full rank. The hypotheses that will be considered in this chapter are all hypotheses on the mean, $E\{\underline{y}\}$, of $\underline{y}$. The following two hypotheses are considered:

$$\boxed{H_0 : \underset{b\times m}{B^T}\underset{m\times1}{E\{\underline{y}\}} = \underset{b\times1}{0} \quad \text{versus} \quad H_A : \underset{b\times m}{B^T}\underset{m\times1}{E\{\underline{y}\}} = \underset{b\times q}{C_t}\underset{q\times1}{\nabla},\ \nabla \neq 0} \ . \tag{2}$$

It is assumed that rank$B = b$, rank$C_t = q$ and that the $q\times1$ vector ∇ is unknown under H_A. Note that both the hypotheses H_0 and H_A are composite if $b < m$. If $b = m$, then the hypothesis H_0 reduces to a simple hypothesis. The hypotheses of (2) are formulated in terms of condition equations. As we know a completely equivalent formulation is possible in terms of observation equations. In order to transform (2) into observation equations we consider the inhomogeneous system of linear equations:

$$\underset{b\times m}{B^T}\ \underset{m\times1}{E\{\underline{y}\}} = \underset{b\times q}{C_t}\underset{q\times1}{\nabla}. \tag{3}$$

We know that the solution of this inhomogeneous system is given by the sum of a particular solution and the solution of the homogeneous system. If the $m\times q$ matrix C_y is defined such that it satisfies $B^T C_y = C_t$, the particular solution of (3) is given by:

$$\underset{m\times1}{E\{\underline{y}\}_{\text{part}}} = \underset{m\times q}{C_y}\underset{q\times1}{\nabla}, \quad \text{rank}C_y = q. \tag{4}$$

The homogeneous solution is the solution of $B^T E\{\underline{y}\} = 0$. If we denote the $m\times(m - b)$ matrix of which the columnvectors are orthogonal to the columnvectors of matrix B by A, then $B^T A = 0$ and the homogeneous solution becomes:

$$\underset{m\times1}{E\{\underline{y}\}_{\text{hom}}} = \underset{m\times n}{A}\underset{n\times1}{x}, \quad \text{with } m - n = b. \tag{5}$$

Taking the sum of (4) and (5) given the solution to (3):

(6) $$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{\boldsymbol{A}x} + \underset{m\times qq\times 1}{C_y\nabla}.$$

In terms of observation equations the two hypotheses of (2) read therefore:

(7) $$\boxed{H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{\boldsymbol{A}x} \quad \text{versus} \quad H_A : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{\boldsymbol{A}x} + \underset{m\times q}{C_y}\nabla, \ \underset{q\times 1}{\nabla} \neq \mathbf{0}}.$$

Both the matrices A and C_y are of full rank. Thus rank $A = n$ and rank $C_y = q$. Furthermore rank $(A : C_y) = n+q$. In practical applications the formulation of the hypotheses of (2) and (7) is usually achieved in the following way. In geodetic practice one generally has a good idea of how to model a particular problem in terms of either condition equations or observation equations. This then results in the null hypothesis H_0. However, while formulating the null hypothesis H_0, usually a number of assumptions are made. For instance, one assumes that the data are free from blunders, or that the effect of refraction is negligible, or that the points of a geodetic network lie in a two dimensional Euclidean plane etc. In order to find out whether these assumptions are valid or not, one opposes the null hypothesis H_0 to a more relaxed alternative hypothesis H_A in which more explanatory variables, namely ∇ in (2) and (7), are introduced. The explanatory variables ∇ are then supposed to model those effects which were assumed absent in H_0. For instance, through ∇ one may model the presence of one or more blunders in the data, or the presence of refraction, etc. The test of H_0 versus H_A informs us then on whether or not the additional explanatory variables ∇ should be taken into account. That is, the test should then inform us on whether for instance blunders in the data are absent or not. However, referring to the two types of errors one can make in testing, the type I and the type II error, and to the fact that every model is only an approximation, one should never forget that the result of a test is only indicative and never a proof of the correctness of one model over another!

Now let us derive the *generalized likelihood ratio test* for testing H_0 against H_A. From the previous chapter (see (3) in Section 3.1) we know that this test can be computed from the probability density function of $\underline{y}$ under H_0 and H_A. The probability density function of $\underline{y}$ under H_0 reads:

(8) $$p_{\underline{y}}(y|x) = (2\pi)^{-\frac{m}{2}} |Q_y|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(y-\boldsymbol{A}x)^* Q_y^{-1}(y-\boldsymbol{A}x)].$$

And under H_A it reads:

(9) $$p_{\underline{y}}(y|x,\nabla) = (2\pi)^{-\frac{m}{2}} |Q_y|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(y-\boldsymbol{A}x-C_y\nabla)^* Q_y^{-1}(y-\boldsymbol{A}x-C_y\nabla)].$$

The numerator of the generalized likelihood ratio test is given by $\max_{x\in\mathbb{R}^n} p_{\underline{y}}(y|x)$. Let us denote the value of x that maximizes $p_{\underline{y}}(y|x)$ by $\hat{x}_0$. The index "0" is used to indicate that the density function of $\underline{y}$ under H_0 is taken. Since $\hat{x}_0$ maximizes $p_{\underline{y}}(y|x)$, we have:

(10) $$\max_{x \in \mathbb{R}^n} p_{\underline{y}}(y|x) = p_{\underline{y}}(y|\hat{x}_0).$$

Recall from Adjustment theory that $\hat{x}_0$ is the *maximum likelihood estimate* of x and that the maximum likelihood estimate, in case of a normal distribution, is identical to the least-squares estimate of x. Since the least-squares residual vector is given by:

(11) $$\hat{e}_0 = y - A\hat{x}_0$$

it follows from (8) and (10) that the *numerator* of the generalized likelihood ratio test is given by:

(12) $$\max_{x \in \mathbb{R}^n} p_{\underline{y}}(y|x) = (2\pi)^{-\frac{m}{2}} |Q_y|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\hat{e}_0^* Q_y^{-1} \hat{e}_0].$$

Now let us have a look at the denominator of the generalized likelihood ratio test. It is given by $\max_{x \in \mathbb{R}^n, \nabla \in \mathbb{R}^q} p_{\underline{y}}(y|x,\nabla)$. Let us denote the value of x and the value of ∇ that maximize $p_{\underline{y}}(y|x,\nabla)$ by $\hat{x}_A$ and $\hat{\nabla}$ respectively. Then:

(13) $$\max_{x \in \mathbb{R}^n, \nabla \in \mathbb{R}^q} p_{\underline{y}}(y|x) = p_{\underline{y}}(y|\hat{x}_A, \hat{\nabla}).$$

In this case $\hat{x}_A$ and $\hat{\nabla}$ are the maximum likelihood estimates of x and ∇ under H_A respectively. They are therefore also the least-squares estimates of x and ∇ under H_A respectively. Since the least-squares residual vector under H_A is given by:

(14) $$\hat{e}_A = y - (A \vdots C_y)\begin{pmatrix} \hat{x}_A \\ \hat{\nabla} \end{pmatrix}$$

it follows from (9) and (13) that the *denominator* of the generalized likelihood ratio test is given by:

(15) $$\max_{x \in \mathbb{R}^n, \nabla \in \mathbb{R}^q} p_{\underline{y}}(y|x) = (2\pi)^{-\frac{m}{2}} |Q_y|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\hat{e}_A^* Q_y^{-1} \hat{e}_A].$$

From (12) and (15) it follows that:

$$\frac{\max_{x \in \mathbb{R}^n} p_{\underline{y}}(y|x)}{\max_{x \in \mathbb{R}^n, \nabla \in \mathbb{R}^q} p_{\underline{y}}(y|x,\nabla)} = \exp[-\tfrac{1}{2}(\hat{e}_0^* Q_y^{-1} \hat{e}_0 - \hat{e}_A^* Q_y^{-1} \hat{e}_A)].$$

Since this ratio is less than a positive constant if and only if the term within the brackets [...] is larger than a positive constant, it follows that the generalized likelihood ratio test for testing H_0 against H_A reads:

(16) $$\boxed{\text{reject } H_0 \text{ if } \quad \hat{e}_0^* Q_y^{-1} \hat{e}_0 - \hat{e}_A^* Q_y^{-1} \hat{e}_A > k_\alpha}.$$

The left-hand side of the inequality in (16) is expressed in terms of $\hat{e}_0$ and $\hat{e}_A$. It is also possible however to express the left-hand side of the inequality in (16) solely in terms of:

(17) $$\hat{y}_0 = A\hat{x}_0 \quad \text{and} \quad \hat{y}_A = A\hat{x}_A + C_y\hat{\nabla}.$$

In order to see this, note that:

$$\hat{e}_0^* Q_y^{-1}\hat{e}_0 - \hat{e}_A^* Q_y^{-1}\hat{e}_A = (\hat{e}_0 - \hat{e}_A)^* Q_y^{-1}(\hat{e}_0 - \hat{e}_A) + 2(\hat{e}_0 - \hat{e}_A)^* Q_y^{-1}\hat{e}_A$$

or that:

(18) $$\hat{e}_0^* Q_y^{-1}\hat{e}_0 - \hat{e}_A^* Q_y^{-1}\hat{e}_A = (\hat{y}_0 - \hat{y}_A)^* Q_y^{-1}(\hat{y}_0 - \hat{y}_A) - 2(\hat{y}_0 - \hat{y}_A)^* Q_y^{-1}\hat{e}_A.$$

The second term on the right-hand side of (18) can be written with the help of (17) as:

(19) $$-2(\hat{y}_0 - \hat{y}_A)^* Q_y^{-1}\hat{e}_A = -2[A(\hat{x}_0 - \hat{x}_A) - C_y\hat{\nabla}]^* Q_y^{-1}\hat{e}_A.$$

But this term is identical to zero. Thus:

(20) $$-2(\hat{y}_0 - \hat{y}_A)^* Q_y^{-1}\hat{e}_A = 0.$$

In order to see this, recall from Adjustment theory that one of the properties of the least-squares method is that the least-squares residual vector is orthogonal to the columns of the designmatrix. In the present context this means that $\hat{e}_A$ is orthogonal to the columnvectors of the matrix $(A : C_y)$, where orthogonal is "measured" with respect to the Q_y^{-1}-metric. This implies that:

(21) $$\begin{pmatrix} A^* \\ C_y^* \end{pmatrix} Q_y^{-1}\hat{e}_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

With (21), equation (20) follows from (19). Substitution of (20) into (18) gives with (16):

(22) $$\boxed{\text{reject } H_0 \text{ if } \quad (\hat{y}_0 - \hat{y}_A)^* Q_y^{-1}(\hat{y}_0 - \hat{y}_A) > k_\alpha}\ .$$

Note that intuitively this test makes sense. One would expect to reject H_0 if $\hat{y}_A$ differs considerably from $\hat{y}_0$, that is, one would expect to reject H_0 if $(\hat{y}_0 - \hat{y}_A)^* Q_y^{-1}(\hat{y}_0 - \hat{y}_A)$ is large. Also note that since the left-hand side of the inequality in (22) is always non-negative, $\hat{e}_0^* Q_y^{-1}\hat{e}_0$ must always be larger than or equal to $\hat{e}_A^* Q_y^{-1}\hat{e}_A$. This corresponds with our earlier remark in the previous chapter that the denominator of the generalized likelihood ratio is always larger than or equal to the numerator. It seems, the way in which (16) and (22) are formulated, that we need both $\hat{e}_0$ and $\hat{e}_A$ or $\hat{y}_0$ and $\hat{y}_A$ in order to perform the test. This would imply that a least-squares computation under both H_0 and H_A is needed. Fortunately this is not the case. We will show that $\hat{e}_A$ in (16) or $\hat{y}_A$ in (22) are not explicitly needed in order to perform the test. In order to show this, we will first write $\hat{y}_A$ in terms of $\hat{y}_0$ and $\hat{\nabla}$. Consider therefore the two systems of *normal equations* that correspond to H_0 and H_A:

(23) $$A^{*}Q_y^{-1}A\ \hat{x}_0 = A^{*}Q_y^{-1}y$$

and

(24) $$\begin{pmatrix} A^{*}Q_y^{-1}A & A^{*}Q_y^{-1}C_y \\ C_y^{*}Q_y^{-1}A & C_y^{*}Q_y^{-1}C_y \end{pmatrix}\begin{pmatrix} \hat{x}_A \\ \hat{\nabla} \end{pmatrix} = \begin{pmatrix} A^{*}Q_y^{-1}y \\ C_y^{*}Q_y^{-1}y \end{pmatrix}.$$

These systems of equations have a unique solution since rank $A = n$, rank $(A \vdots C_y) = n+q$. Substitution of (23) into (24) gives:

(25) $$\begin{pmatrix} A^{*}Q_y^{-1}A & A^{*}Q_y^{-1}C_y \\ C_y^{*}Q_y^{-1}A & C_y^{*}Q_y^{-1}C_y \end{pmatrix}\begin{pmatrix} \hat{x}_A \\ \hat{\nabla} \end{pmatrix} = \begin{pmatrix} A^{*}Q_y^{-1}A\hat{x}_0 \\ C_y^{*}Q_y^{-1}y \end{pmatrix}.$$

Pre-multiplication of this system of equations with the square and full rank matrix:

$$\begin{pmatrix} I & 0 \\ -C_y^{*}Q_y^{-1}A(A^{*}Q_y^{-1}A)^{-1} & I \end{pmatrix}$$

gives:

(26) $$\begin{pmatrix} A^{*}Q_y^{-1}A & A^{*}Q_y^{-1}C_y \\ 0 & C_y^{*}Q_y^{-1}[Q_y - A(A^{*}Q_y^{-1}A)^{-1}A^{*}]Q_y^{-1}C_y \end{pmatrix}\begin{pmatrix} \hat{x}_A \\ \hat{\nabla} \end{pmatrix} = \begin{pmatrix} A^{*}Q_y^{-1}A\hat{x}_0 \\ C_y^{*}Q_y^{-1}[I - A(A^{*}Q_y^{-1}A)^{-1}A^{*}Q_y^{-1}]y \end{pmatrix}.$$

Now recall from Adjustment theory that:

(27) $$\begin{cases} \hat{e}_0 = [I - A(A^{*}Q_y^{-1}A)^{-1}A^{*}Q_y^{-1}]y \text{ and} \\ Q_{\hat{e}_0} = Q_y - A(A^{*}Q_y^{-1}A)^{-1}A^{*}. \end{cases}$$

Substitution of (27) into (26) gives:

(28) $$\begin{pmatrix} A^{*}Q_y^{-1}A & A^{*}Q_y^{-1}C_y \\ 0 & C_y^{*}Q_y^{-1}Q_{\hat{e}_o}Q_y^{-1}C_y \end{pmatrix}\begin{pmatrix} \hat{x}_A \\ \hat{\nabla} \end{pmatrix} = \begin{pmatrix} A^{*}Q_y^{-1}A\hat{x}_0 \\ C_y^{*}Q_y^{-1}\hat{e}_0 \end{pmatrix}.$$

From the first equation of (28) follows that:

$$\hat{x}_A = \hat{x}_0 - (A^{*}Q_y^{-1}A)^{-1}A^{*}Q_y^{-1}C_y\hat{\nabla}$$

or that:

$$A\hat{x}_A + C_y\hat{\nabla} = A\hat{x}_0 + [I - A(A^*Q_y^{-1}A)^{-1}A^*Q_y^{-1}]C_y\hat{\nabla}$$

or with the second equation of (27) that:

$$\hat{y}_A = \hat{y}_0 + Q_{\hat{e}_0}Q_y^{-1}C_y\hat{\nabla}. \tag{29}$$

This formula expresses $\hat{y}_A$ in terms of $\hat{y}_0$ and $\hat{\nabla}$. In (22) we need $(\hat{y}_0 - \hat{y}_A)^*Q_y^{-1}(\hat{y}_0 - \hat{y}_A)$. With (29) this gives:

$$(\hat{y}_0 - \hat{y}_A)^*Q_y^{-1}(\hat{y}_0 - \hat{y}_A) = \hat{\nabla}^*C_y^*Q_y^{-1}Q_{\hat{e}_0}Q_y^{-1}Q_{\hat{e}_0}Q_y^{-1}C_y\hat{\nabla}. \tag{30}$$

Now we know from Adjustment theory that:

$$Q_{\hat{e}_0}Q_y^{-1}Q_{\hat{e}_0} = Q_{\hat{e}_0} \tag{31}$$

(verify this yourself). Substitution of (31) into (30) gives together with (22) the test:

$$\boxed{\text{reject } H_0 \text{ if } \quad \hat{\nabla}^*C_y^*Q_y^{-1}Q_{\hat{e}_0}Q_y^{-1}C_y\hat{\nabla} > k_\alpha} \; . \tag{32}$$

Again note that this test makes sense. It says to reject H_0 if $\hat{\nabla}$, which is supposed to be zero under H_0, is large. From the second equation of (28) follows that the inverse of the error estimator's variance matrix reads $Q_{\hat{\nabla}}^{-1} = C_y^*Q_y^{-1}Q_{\hat{e}_0}Q_y^{-1}C_y$, which is the central term in (32). The test (32) is formulated in terms of $\hat{\nabla}$ and therefore still does not show that an explicit least-squares computation under H_A is not needed. We will now express $\hat{\nabla}$ in terms of $\hat{e}_0$. From the second equation of (28) also follows that:

$$\hat{\nabla} = (C_y^*Q_y^{-1}Q_{\hat{e}_0}Q_y^{-1}C_y)^{-1}C_y^*Q_y^{-1}\hat{e}. \tag{33}$$

Substitution of (33) into (32) then finally gives:

$$\boxed{\text{reject } H_0 \text{ if } \quad \hat{e}_0^*Q_y^{-1}C_y(C_y^*Q_y^{-1}Q_{\hat{e}_0}Q_y^{-1}C_y)^{-1}C_y^*Q_y^{-1}\hat{e}_0 > k_\alpha} \; . \tag{34}$$

This result shows that $\hat{e}_A, \hat{y}_A$ and $\hat{\nabla}$ are not explicitly needed to perform the generalized likelihood ratio test for testing H_0 against H_A. So far we have seen four different expressions for the generalized likelihood ratio test, namely (16), (22), (32) and (34). There is however also a fifth useful expression. This expression is in particular useful if the hypotheses are formulated in terms of condition equations like in (2). The expression is formulated in terms of $\underline{t}$, the vector of misclosures. Recall that $\hat{e}_0$ and $Q_{\hat{e}_0}$ may be written is terms of $t = B^*y$ and $Q_t = B^*Q_yB$ as:

$$\begin{cases} \hat{e}_0 = Q_yB(B^*Q_yB)^{-1}B^*y = Q_yBQ_t^{-1}t, \text{ and} \\ Q_{\hat{e}_0} = Q_yB(B^*Q_yB)^{-1}B^*Q_y = Q_yBQ_t^{-1}B^*Q_y. \end{cases} \tag{35}$$

Substitution of (35) into (34) gives with $C_t = B^*C_y$:

$$\boxed{\text{reject } H_0 \text{ if } \quad t^*Q_t^{-1}C_t(C_t^*Q_t^{-1}C_t)^{-1}C_t^*Q_t^{-1}t > k_\alpha} \; . \tag{36}$$

The random variable defined by the left-hand side of the inequalities in (16), (22), (32), (34) and (36) will be denoted by $\underline{T}_q$. Thus in terms of the expression in (36) we have:

$$\underline{T}_q = \underline{t}^* Q_t^{-1} C_t (C_t^* Q_t^{-1} C_t)^{-1} C_t^* Q_t^{-1} \underline{t}. \tag{37}$$

Now in order to compute the critical value k_α from the size α of the test, we need the distribution of $\underline{T}_q$. Substitution of:

$$\underline{z} = C_t^* Q_t^{-1} \underline{t} \quad \text{and} \quad Q_z = C_t^* Q_t^{-1} C_t \tag{38}$$

into (37) gives:

$$\underline{T}_q = \underline{z}^* Q_z^{-1} \underline{z}. \tag{39}$$

Since the random variable $\underline{z}$ of (38) is distributed under H_0 and H_A as:

$$H_0 : \underset{q\times 1}{\underline{z}} \sim N(\underset{q\times 1}{0}, \underset{q\times q}{Q_z}) \quad ; \quad H_A : \underset{q\times 1}{\underline{z}} \sim N(\underset{q\times 1}{Q_z \nabla}, \underset{q\times q}{Q_z})$$

it follows from Appendix A that $\underline{T}_q$ is distributed under H_0 and H_A as:

$$H_0 : \underline{T}_q \sim \chi^2(q,0) \quad ; \quad H_A : \underline{T}_q \sim \chi^2(q,\lambda), \ \lambda = \nabla^* Q_z \nabla \tag{40}$$

with

$$Q_z = C_t^* Q_t^{-1} C_t = C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \tag{41}$$

this finally gives for the distribution of $\underline{T}_q$:

$$\boxed{\begin{array}{l} H_0 : \underline{T}_q \sim \chi^2(q,0) \quad ; \quad H_A : \underline{T}_q \sim \chi^2(q,\lambda), \\ \text{with noncentrality parameter } \lambda = \nabla^* C_t^* Q_t^{-1} C_t \nabla = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \nabla \end{array}} \tag{42}$$

To conclude this section a summary of the important results is given in Table 4.1.

Hypotheses

$$H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{\boldsymbol{A}x} \;,\; \text{rank}\boldsymbol{A} = n$$

versus

$$H_A : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{\boldsymbol{A}x} + \underset{m\times qq\times 1}{C_y\nabla} \;,\; \text{rank}(\boldsymbol{A} : C_y) = n+q$$

$$H_0 : \underset{b\times m}{\boldsymbol{B}^*}\underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0} \;,\; \text{rank}\boldsymbol{B} = b$$

versus

$$H_A : \underset{b\times m}{\boldsymbol{B}^*}\underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times qq\times 1}{C_t\nabla} \;,\; \text{rank}C_t = q$$

$$\underset{b\times mm\times n}{\boldsymbol{B}^*\boldsymbol{A}} = \underset{b\times n}{0},$$

$$m-n = b,$$

$$\underset{b\times mm\times q}{\boldsymbol{B}^*C_y} = \underset{b\times q}{C_t}$$

Teststatistic $\underline{T}_q$:

$$\begin{aligned}
\underline{T}_q &= \underline{\hat{e}}_0^* Q_y^{-1}\underline{\hat{e}}_0 - \underline{\hat{e}}_A^* Q_y^{-1}\underline{\hat{e}}_A \\
&= (\underline{\hat{y}}_0 - \underline{\hat{y}}_A)^* Q_y^{-1}(\underline{\hat{y}}_0 - \underline{\hat{y}}_A) \\
&= \underline{\hat{\nabla}}^* C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \underline{\hat{\nabla}} \\
&= \underline{\hat{e}}_0^* Q_y^{-1} C_y (C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y)^{-1} C_y^* Q_y^{-1}\underline{\hat{e}}_0 \\
&= \underline{t}^* Q_t^{-1} C_t (C_t^* Q_t^{-1} C_t)^{-1} C_t^* Q_t^{-1}\underline{t}
\end{aligned}$$

Distribution of $\underline{T}_q$:

$$H_0 : \underline{T}_q \sim \chi^2(q,0) \;\; ; \;\; H_A : \underline{T}_q \sim \chi^2(q,\lambda)$$
$$\lambda = \nabla^* C_t^* Q_t^{-1} C_t \nabla = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \nabla$$

Generalized likelihood ratio test :

reject H_0 if $T_q > k_\alpha$

Table 4.1: Overview of the teststatistic $\underline{T}_q$.

4.2 A geometric interpretation of $\underline{T}_q$

In the previous section it was shown *algebraically* that:

$$
\begin{aligned}
\underline{T}_q &= \underline{\hat{e}}_0^* Q_y^{-1} \underline{\hat{e}}_0 - \underline{\hat{e}}_A^* Q_y^{-1} \underline{\hat{e}}_A \\
&= (\underline{\hat{y}}_0 - \underline{\hat{y}}_A)^* Q_y^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_A) \\
&= \underline{\hat{\nabla}}^* C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \underline{\hat{\nabla}} \\
&= \underline{\hat{e}}_0^* Q_y^{-1} C_y (C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y)^{-1} C_y^* Q_y^{-1} \underline{\hat{e}}_0 \\
&= \underline{t}^* Q_t^{-1} C_t (C_t^* Q_t^{-1} C_t)^{-1} C_t^* Q_t^{-1} \underline{t}.
\end{aligned}
\tag{43}
$$

In the present section the equality of these expressions will be shown *geometrically*. Let us first consider the hypotheses H_0 and $H_{A:}$

$$
H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{Ax} \quad \text{versus} \quad H_A : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{Ax} + \underset{m\times qq\times 1}{C_y \nabla}, \ \nabla \neq 0. \tag{44}
$$

Since it was assumed that rank $A = n$ and rank $(A \vdots C_y) = n+q$, the dimensions of the range spaces of A and $(A \vdots C_y)$ are respectively: $\dim R(A) = n$ and $\dim R(A \vdots C_y) = n+q$. Since the matrices A and $(A \vdots C_y)$ have m-number of rows, it follows that the columnvectors of these matrices are elements of $\mathbb{R}^m$. Thus $R(A) \subset \mathbb{R}^m$ and $R(A \vdots C_y) \subset \mathbb{R}^m$. Since the columns of matrix A can be written as linear combinations of the columns of matrix $(A \vdots C_y)$ it follows that $R(A) \subset R(A \vdots C_y)$. Thus the rangespace of A is a linear subspace of the rangespace of $(A \vdots C_y)$. The equation of H_0 in (44) states that $E\{\underline{y}\}$ under H_0 can be written as a linear combination of the columnvectors of matrix A. This implies that $E\{\underline{y}|H_0\} \in R(A)$. Similarly the equation of H_A in (44) can be translated into $E\{\underline{y}|H_A\} \in R(A \vdots C_y)$. The above results can be summarized as:

$$
\begin{cases}
E\{\underline{y}|H_0\} \in R(A) \subset R(A \vdots C_y) \subset \mathbb{R}^m \ ; & E\{\underline{y}|H_A\} \in R(A \vdots C_y) \subset \mathbb{R}^m \\
\dim R(A) = n \ ; & \dim R(A \vdots C_y) = n+q.
\end{cases}
\tag{45}
$$

A sketch of (45) is given in Figure 4.1.

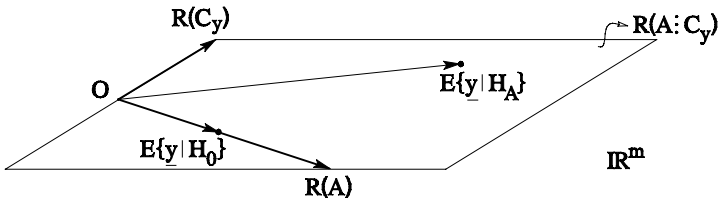


Figure 4.1: The hypotheses $H_0 : E\{\underline{y}|H_0\}$ and $H_A : E\{\underline{y}|H_A\}$.

Recall from Adjustment theory that the method of least-squares can be interpreted geometrically as a method of *orthogonal projection*. That is $\underline{\hat{y}}_0$ follows from the orthogonal projection of $\underline{y}$ onto the rangespace of A, and $\underline{\hat{y}}_A$ follows from the orthogonal projection of $\underline{y}$ onto the rangespace of $(A \vdots C_y)$. Thus:

(45) $$\underline{\hat{y}}_0 = P_A\underline{y} \quad \text{and} \quad \underline{\hat{y}}_A = P_{(A:C_y)}\underline{y}.$$

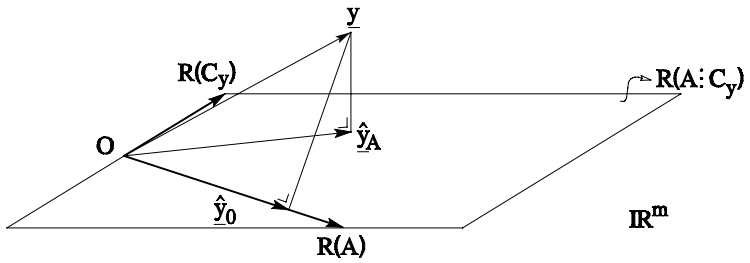


Figure 4.2: $\underline{y}\in\mathbb{R}^m$, $\underline{\hat{y}}_0 = P_A\underline{y}\in R(A)$, $\underline{\hat{y}}_A = P_{(A:C_y)}\underline{y}\in R(A:C_y)$.

This is shown in Figure 4.2. Recall that orthogonality is "measured" with respect to the Q_y^{-1}-metric. This means that the innerproduct and norm in $\mathbb{R}^m$ are defined as:

(46) $$\begin{cases} \text{innerproduct:} & (u,v) = u^*Q_y^{-1}v \ , \ u,v\in\mathbb{R}^m \\ \text{norm} & : \ \|u\| = (u^*Q_y^{-1}u)^{1/2} \ , \ u,\in\mathbb{R}^m. \end{cases}$$

Since $\underline{\hat{y}}_0$ is the orthogonal projection of $\underline{y}$ onto $R(A)$, it follows that $\underline{y}-\underline{\hat{y}}_0$ is orthogonal to $\underline{\hat{y}}_0$. Thus $(\underline{y}-\underline{\hat{y}}_0,\underline{\hat{y}}_0) = 0$, see also Figure 4.2. Since $\underline{\hat{y}}_A$ is the orthogonal projection of $\underline{y}$ onto $R(A:C_y)$ it follows that $\underline{y}-\underline{\hat{y}}_A$ is orthogonal to $R(A:C_y)$ and thus also orthogonal to $\underline{\hat{y}}_A\in R(A:C_y)$. Thus: $(\underline{y}-\underline{\hat{y}}_A,\underline{\hat{y}}_A) = 0$, see also Figure 4.2. Since $\underline{y}-\underline{\hat{y}}_A$ is orthogonal to $R(A:C_y)$, it is also orthogonal to $R(A) \subset R(A:C_y)$. But $\underline{\hat{y}}_0\in R(A)$. Hence, $\underline{y}-\underline{\hat{y}}_A$ is also orthogonal to $\underline{\hat{y}}_0$. Thus: $(\underline{y}-\underline{\hat{y}}_A,\underline{\hat{y}}_0) = 0$. Summarizing we have:

(47) $$(i) \quad (\underline{y}-\underline{\hat{y}}_0,\underline{\hat{y}}_0) = 0, \quad (ii) \quad (\underline{y}-\underline{\hat{y}}_A,\underline{\hat{y}}_A) = 0, \quad (iii) \quad (\underline{y}-\underline{\hat{y}}_A,\underline{\hat{y}}_0) = 0.$$

If we substract (47ii) from (47iii) we get:

(48) $$(\underline{y}-\underline{\hat{y}}_A,\underline{\hat{y}}_0-\underline{\hat{y}}_A) = 0.$$

The four orthogonality relations of (47) and (48) are shown in Figure 4.3.

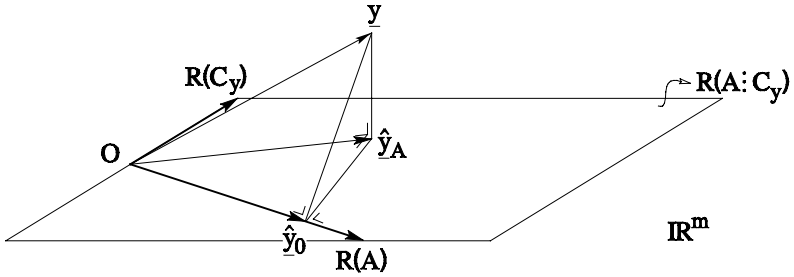


Figure 4.3: The right-angled triangle $\underline{y},\underline{\hat{y}}_0,\underline{\hat{y}}_A$.

The right-angled triangle $\underline{y}, \underline{\hat{y}}_0, \underline{\hat{y}}_A$ of Figure 4.3 has been shown again in Figure 4.4.

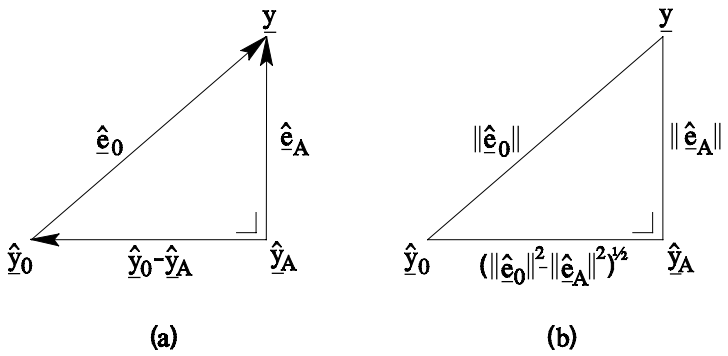


Figure 4.4: The right-angled triangle $\underline{y}, \underline{\hat{y}}_0, \underline{\hat{y}}_A$.

From Figure 4.4 and the *Pythagoras theorem* we learn that:

(49) $$\boxed{\|\underline{\hat{e}}_0\|^2 - \|\underline{\hat{e}}_A\|^2 = \|\underline{\hat{y}}_0 - \underline{\hat{y}}_A\|^2}\ .$$

In terms of the matrix Q_y^{-1} this can be written as:

(50) $$\underline{\hat{e}}_0^* Q_y^{-1} \underline{\hat{e}}_0 - \underline{\hat{e}}_A^* Q_y^{-1} \underline{\hat{e}}_A = (\underline{\hat{y}}_0 - \underline{\hat{y}}_A)^* Q_y^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_A).$$

Compare this result with the first two equations of (43).

Let us now consider the third equation of (43). We know that $\underline{\hat{y}}_A$ can be written as:

(51) $$\underline{\hat{y}}_A = A\underline{\hat{x}}_A + C_y \underline{\hat{\nabla}}.$$

This decomposition of $\underline{\hat{y}}_A$ is shown in Figure 4.5.

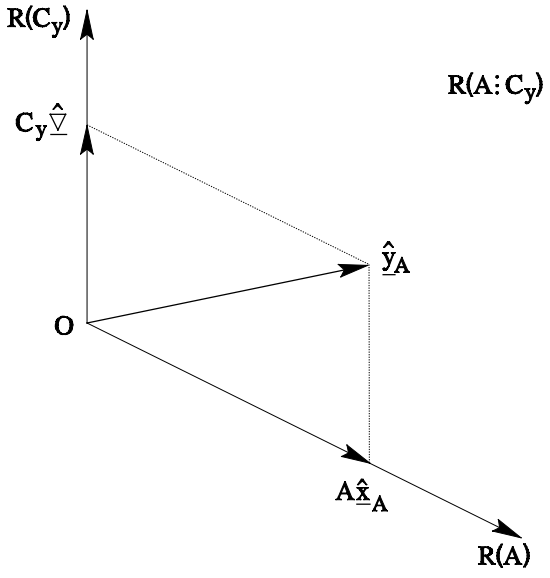


Figure 4.5: Decomposition of $\underline{\hat{y}}_A$ into $A\underline{\hat{x}}_A$ and $C_y \underline{\hat{\nabla}}$.

The vector $C_y\hat{\underline{\nabla}}$ can be further decomposed in a part that lies in the range space of A, $R(A)$, and in a part that lies in the orthogonal complement of $R(A)$, $R(A)^{\perp}$. This gives:

(52) $$C_y\hat{\underline{\nabla}} = \boldsymbol{P_A}C_y\hat{\underline{\nabla}} + \boldsymbol{P_A^{\perp}}C_y\hat{\underline{\nabla}}.$$

This orthogonal decomposition of $C_y\hat{\underline{\nabla}}$ is shown in Figure 4.6. Substitution of (52) into (51) gives:

(53) $$\hat{\underline{y}}_A = \boldsymbol{A}\hat{\underline{x}}_A + \boldsymbol{P_A}C_y\hat{\underline{\nabla}} + \boldsymbol{P_A^{\perp}}C_y\hat{\underline{\nabla}}.$$

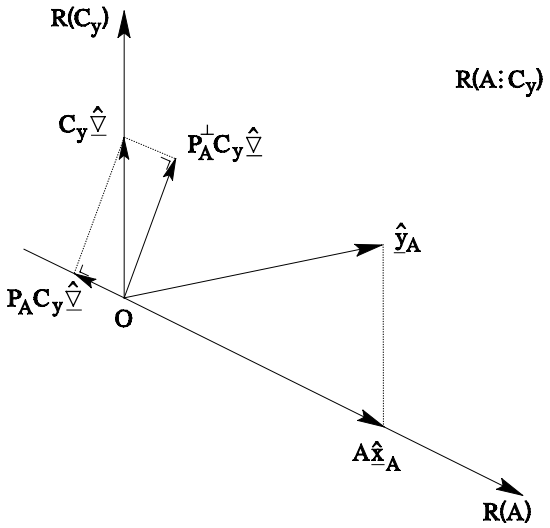


Figure 4.6: Orthogonal decomposition of $C_y\hat{\underline{\nabla}}$ into $P_AC_y\hat{\underline{\nabla}}$ and $P_A^{\perp}C_y\hat{\underline{\nabla}}$.

We know that $\hat{\underline{y}}_A - \hat{\underline{y}}_0 \in R(A)^{\perp}$ and $\hat{\underline{y}}_0 \in R(A)$. From this follows that $P_A^{\perp}(\hat{\underline{y}}_A - \hat{\underline{y}}_0) = \hat{\underline{y}}_A - \hat{\underline{y}}_0$ and $P_A^{\perp}(\hat{\underline{y}}_A - \hat{\underline{y}}_0) = P_A^{\perp}\hat{\underline{y}}_A$, and thus that:

(54) $$\hat{\underline{y}}_A - \hat{\underline{y}}_0 = \boldsymbol{P_A^{\perp}}\hat{\underline{y}}_A.$$

Substitution of (53) into the right-hand side of (54) gives:

$$\hat{\underline{y}}_A - \hat{\underline{y}}_0 = \boldsymbol{P_A^{\perp}}(\boldsymbol{A}\hat{\underline{x}}_A + \boldsymbol{P_A}C_y\hat{\underline{\nabla}} + \boldsymbol{P_A^{\perp}}C_y\hat{\underline{\nabla}}) = \boldsymbol{P_A^{\perp}}\boldsymbol{P_A^{\perp}}C_y\hat{\underline{\nabla}}$$

or:

(55) $$\hat{\underline{y}}_A - \hat{\underline{y}}_0 = \boldsymbol{P_A^{\perp}}C_y\hat{\underline{\nabla}}.$$

This is shown in Figure 4.7. If we take the norm of (55) we get:

(56) $$\boxed{\|\hat{\underline{y}}_0 - \hat{\underline{y}}_A\| = \|P_A^{\perp}C_y\hat{\underline{\nabla}}\|}\ .$$

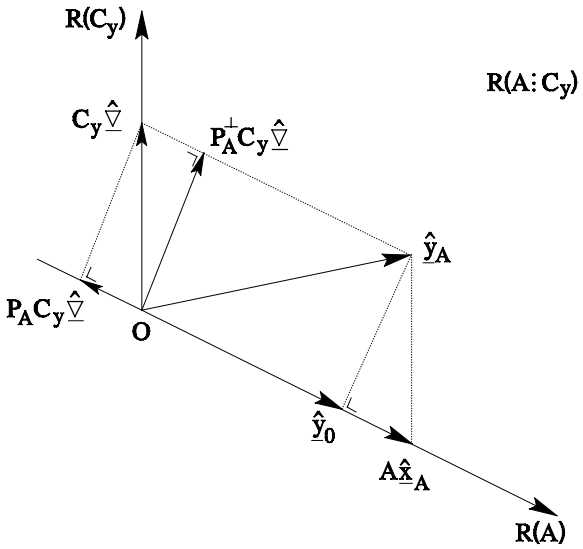


Figure 4.7: $\underline{\hat{y}}_A - \underline{\hat{y}}_0 = \boldsymbol{P}_A^{\perp} C_y \underline{\hat{\nabla}}$.

In terms of matrices this can be written as:

$$(\underline{\hat{y}}_0 - \underline{\hat{y}}_A)^* Q_y^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_A) = (\boldsymbol{P}_A^{\perp} C_y \underline{\hat{\nabla}})^* Q_y^{-1} (\boldsymbol{P}_A^{\perp} C_y \underline{\hat{\nabla}})$$

or as:

(57) $$(\underline{\hat{y}}_0 - \underline{\hat{y}}_A)^* Q_y^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_A) = \underline{\hat{\nabla}}^* C_y^* \boldsymbol{P}_A^{\perp *} Q_y^{-1} \boldsymbol{P}_A^{\perp} C_y \underline{\hat{\nabla}}.$$

Now recall from Adjustment theory that:

(58) $$\boldsymbol{P}_A^{\perp *} Q_y^{-1} \boldsymbol{P}_A^{\perp} = Q_y^{-1} \boldsymbol{P}_A^{\perp} = Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1}.$$

Substitution of (58) into (57) gives:

(59) $$(\underline{\hat{y}}_0 - \underline{\hat{y}}_A)^* Q_y^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_A) = \underline{\hat{\nabla}}^* C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \underline{\hat{\nabla}}.$$

Compare this result with the second and third equation of (43).

Let us now consider the fourth equation of (43). According to (55) $\underline{\hat{y}}_A - \underline{\hat{y}}_0 = P_A^{\perp} C_y \underline{\hat{\nabla}}$. From this follows that:

(60) $$\underline{\hat{e}}_0 - \underline{\hat{e}}_A = \boldsymbol{P}_A^{\perp} C_y \underline{\hat{\nabla}} \quad \text{(see Figure 4.8)}.$$

Since $\underline{\hat{e}}_A \in R(A : C_y)^{\perp}$ and $R(P_A^{\perp} C_y) \subset R(A : C_y)$ it follows that:

(61) $$\boldsymbol{P}_{\boldsymbol{P}_A^{\perp} C_y} \underline{\hat{\boldsymbol{e}}}_A = \boldsymbol{0}.$$

Premultiplication of (60) with $P_{P_A^{\perp}C_y}$ gives therefore with (61):

$$\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0 = \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y\underline{\hat{\nabla}} = \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y\underline{\hat{\nabla}}. \tag{62}$$

If we take the norm of (62) we get:

$$\boxed{\|\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y\underline{\hat{\nabla}}\| = \|\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0\|}\,. \tag{63}$$

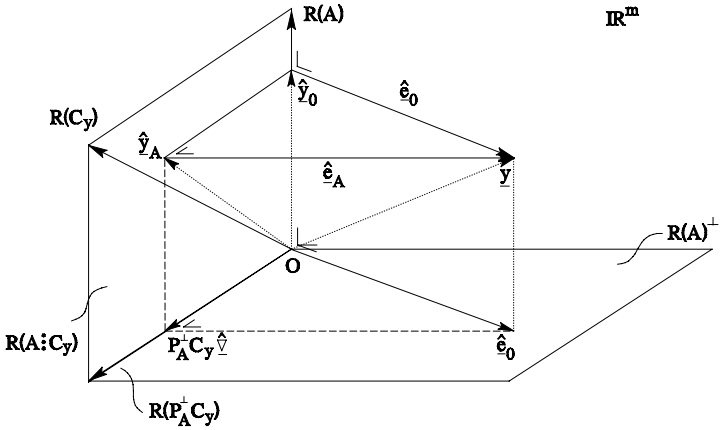


Figure 4.8: $\underline{\hat{e}}_0 = P_A^{\perp}C_y\hat{\nabla} + \underline{\hat{e}}_A$; $P_{P_A^{\perp}C_y}\underline{\hat{e}}_0 = P_A^{\perp}C_y\underline{\hat{\nabla}}$.

This is shown in Figure 4.8. The right-hand side of (63) can be written in terms of matrices as:

$$\|\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0\|^2 = (\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0)^*\boldsymbol{Q}_y^{-1}(\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0) \tag{64}$$

or as:

$$\|\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0\|^2 = \underline{\hat{\boldsymbol{e}}}_0^*\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}^*\boldsymbol{Q}_y^{-1}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0 = \underline{\hat{\boldsymbol{e}}}_0^*\boldsymbol{Q}_y^{-1}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0. \tag{65}$$

The matrix $P_{P_A^{\perp}C_y}$ is given as:

$$\begin{aligned}\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y} &= (\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y)[(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y)^*\boldsymbol{Q}_y^{-1}(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y)]^{-1}(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y)^*\boldsymbol{Q}_y^{-1} \\ &= \boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y(\boldsymbol{C}_y^*\boldsymbol{P}_{\boldsymbol{A}}^{\perp^*}\boldsymbol{Q}_y^{-1}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y)^{-1}\boldsymbol{C}_y^*\boldsymbol{P}_{\boldsymbol{A}}^{\perp^*}\boldsymbol{Q}_y^{-1} \\ &= \boldsymbol{Q}_y\boldsymbol{P}_{\boldsymbol{A}}^{\perp^*}\boldsymbol{Q}_y^{-1}\boldsymbol{C}_y(\boldsymbol{C}_y^*\boldsymbol{Q}_y^{-1}\boldsymbol{Q}_{\hat{\boldsymbol{e}}_0}\boldsymbol{Q}_y^{-1}\boldsymbol{C}_y)^{-1}\boldsymbol{C}_y^*\boldsymbol{Q}_y^{-1}\boldsymbol{P}_{\boldsymbol{A}}^{\perp}.\end{aligned} \tag{66}$$

Substitution of (66) into (65) gives:

$$\|\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0\|^2 = (\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{\hat{\boldsymbol{e}}}_0)^*\boldsymbol{Q}_y^{-1}\boldsymbol{C}_y(\boldsymbol{C}_y^*\boldsymbol{Q}_y^{-1}\boldsymbol{Q}_{\hat{\boldsymbol{e}}_0}\boldsymbol{Q}_y^{-1}\boldsymbol{C}_y)^{-1}\boldsymbol{C}_y^*\boldsymbol{Q}_y^{-1}(\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\underline{\hat{\boldsymbol{e}}}_0). \tag{67}$$

Since $P_A^{\perp}\underline{\hat{e}}_0 = \underline{\hat{e}}_0$ this finally gives:

$$\|\boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{A}}^{\perp}\boldsymbol{C}_y}\underline{\hat{\boldsymbol{e}}}_0\| = \underline{\hat{\boldsymbol{e}}}_0^*\boldsymbol{Q}_y^{-1}\boldsymbol{C}_y(\boldsymbol{C}_y^*\boldsymbol{Q}_y^{-1}\boldsymbol{Q}_{\hat{\boldsymbol{e}}_0}\boldsymbol{Q}_y^{-1}\boldsymbol{C}_y)^{-1}\boldsymbol{C}_y^*\boldsymbol{Q}_y^{-1}\underline{\hat{\boldsymbol{e}}}_0. \tag{68}$$

Compare this result with the fourth equation of (43).

Let us now consider the fifth and last equation of (43). The geometry of this equation is quite different from the geometry of the previous four equations. Note namely that the first four quadratic forms of (43) are all expressed in terms of vectors that are elements of $\mathbb{R}^n$. That is: $\underline{\hat{e}}_0 \in \mathbb{R}^n$, $\underline{\hat{e}}_A \in \mathbb{R}^n$, $\underline{\hat{y}}_0 \in \mathbb{R}^n$, $\underline{\hat{y}}_A \in \mathbb{R}^n$, and $C_y\hat{\nabla} \in \mathbb{R}^n$. The fifth quadratic form of (43) is expressed however in the vector of misclosures, $\underline{t}$, which is an element of $\mathbb{R}^b$, Thus $\underline{t} \in \mathbb{R}^b$ and $\underline{t} \notin \mathbb{R}^n$. If we consider $\mathbb{R}^b$ to have an innerproduct defined by the Q_t^{-1}-matrix, it is still possible to interpret the fifth quadratic form of (43) geometrically. In fact:

(69) $$\boxed{\|P_{C_t}\underline{t}\|^2 = \underline{t}^* Q_t^{-1} C_t (C_t^* Q_t^{-1} C_t)^{-1} C_t^* Q_t^{-1} \underline{t}} \; .$$

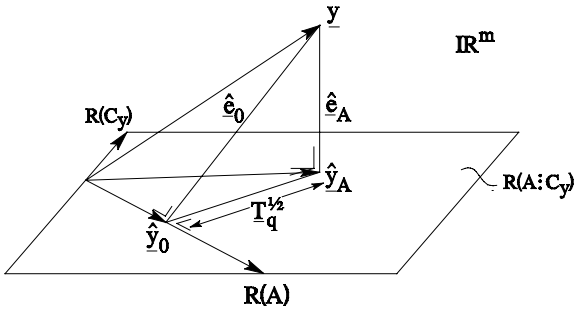
This follows since:

$$\|P_{C_t}\underline{t}\|^2 = (P_{C_t}\underline{t})^* Q_t^{-1} (P_{C_t}\underline{t})$$

and

$$P_{C_t}\underline{t} = C_t (C_t^* Q_t^{-1} C_t)^{-1} C_t^* Q_t^{-1} \underline{t}.$$

A summary of the results of this section is given in Table 4.2.



"Algebra" of $\underline{T}_q$	"Geometry" of $\underline{T}_q$:
$\underline{\hat{e}}_0^* Q_y^{-1} \underline{\hat{e}}_0 - \underline{\hat{e}}_A^* Q_y^{-1} \underline{\hat{e}}_A$ =	$\|\underline{\hat{e}}_0\|^2 - \|\underline{\hat{e}}_A\|^2$ =
$(\underline{\hat{y}}_0 - \underline{\hat{y}}_A)^* Q_y^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_A)$ =	$\|\underline{\hat{y}}_0 - \underline{\hat{y}}_A\|^2$ =
$\hat{\nabla}^* C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y \hat{\nabla}$ =	$\|P_A^{\perp} C_y \hat{\nabla}\|^2$ =
$\underline{\hat{e}}_0^* Q_y^{-1} C_y (C_y^* Q_y^{-1} Q_{\hat{e}_0} Q_y^{-1} C_y)^{-1} C_y^* Q_y^{-1} \underline{\hat{e}}_0$ =	$\|P_{P_A^{\perp} C_y} \underline{\hat{e}}_0\|^2$ =
$\underline{t}^* Q_t^{-1} C_t (C_t^* Q_t^{-1} C_t^{-1})^{-1} C_t^* Q_t^{-1} \underline{t}$ =	$\|P_{C_t}\underline{t}\|^2$.

Table 4.2: The geometry of $\underline{T}_q$.

4.3 The case q=1: $\underline{w}$-teststatistic

In the previous two sections we have seen that the generalized likelihood ratio test for testing:

$$H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{Ax} \quad \text{versus} \quad H_A : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nm\times q}{(A \vdots C_y)} \underset{(n+q)\times 1}{\begin{pmatrix} x \\ \nabla \end{pmatrix}} , \ \nabla \neq 0 \tag{70}$$

or

$$H_0 : \underset{b\times m}{B^*} \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0} \quad \text{versus} \quad H_A : \underset{b\times m}{B^*} \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times qq\times 1}{C_t \nabla} , \quad \nabla \neq 0 \tag{71}$$

is given by (see Figure 4.9):

$$\text{reject } H_0 \text{ if } T_q > k_\alpha. \tag{72}$$

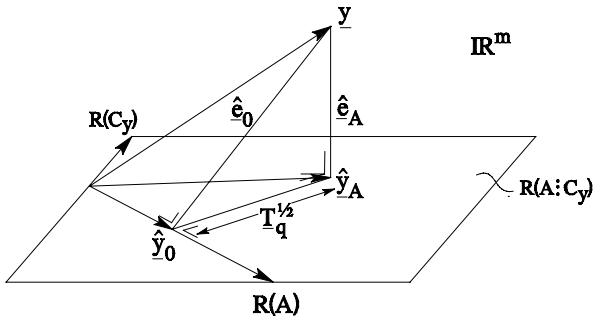


Figure 4.9: The teststatistic $\underline{T}_q$.

Because it was assumed that rank$(A \vdots C_y) = n+q$, it follows that q can never be larger than m-n. If q would be larger than m-n, then rank $(A \vdots C_y)$ would be larger than m, which is impossible since the matrix $(A \vdots C_y)$ only has m rows. The value of q can also not be chosen equal to zero. If $q = 0$, then the matrix C_y would not exist and the two hypotheses H_0 and H_A would be identical. Thus we may conclude that the range of q is given by:

$$\boxed{1 \leq q \leq m-n.} \tag{73}$$

In this section we consider the case q=1. For this case the following three expressions of $\underline{T}_q$ are of interest:

$$\begin{aligned} \underline{T}_q &= \underline{t}^* Q_t^{-1} C_t (C_t^* Q_t^{-1} C_t)^{-1} C_t^* Q_t^{-1} \underline{t} \\ &= \underline{\hat{e}}^* Q_y^{-1} C_y (C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y)^{-1} C_y^* Q_y^{-1} \underline{\hat{e}} \\ &= \underline{\hat{\nabla}}^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \underline{\hat{\nabla}}. \end{aligned} \tag{74}$$

We have dropped the index "0", because it will be clear by now that the least-squares residual vector $\underline{\hat{e}}$ belongs to model H_0. If $q = 1$, the $b\times q$ matrix C_t and the $m\times q$ matrix C_y reduce to $b\times 1$

and $m\times 1$ vectors respectively. In order to accentuate this, we will replace the capitals "C_y" and "C_t" by the small letters "c_y" and "c_t". In this case the first expression of (74) can be written as:

$$\underline{T}_{q=1} = \underline{t}^* Q_t^{-1} c_t (c_t^* Q_t^{-1} c_t)^{-1} c_t^* Q_t^{-1} \underline{t}$$

or as:

(75)
$$\boxed{\underline{T}_{q=1} = \frac{(c_t^* Q_t^{-1} \underline{t})^2}{c_t^* Q_t^{-1} c_t}} .$$

Remembering the $\underline{w}$-teststatistic of Section 2.3 we note that $\underline{T}_{q=1} = (\underline{w})^2$. Hence, the test (72) may also be written as:

(76)
$$\text{reject } H_0 \text{ if } \quad w < -k_\alpha^{1/2} \quad \text{or} \quad w > k_\alpha^{1/2} .$$

In a similar way we find for the second expression of (74):

(77)
$$\boxed{\underline{T}_{q=1} = \frac{(c_y^* Q_y^{-1} \underline{\hat{e}})^2}{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y}}$$ (see (69) of Section 2.3).

For the third expression of (74) we find:

(78)
$$\underline{T}_{q=1} = \frac{\underline{\hat{\nabla}}^2}{(c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y)^{-1}} .$$

The estimator for the model error (33) reduces for q=1 to:

$$\underline{\hat{\nabla}} = \frac{c_y^* Q_y^{-1} \underline{\hat{e}}}{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y} ,$$

the denominator of (78) equals the variance of $\underline{\hat{\nabla}}$, $\sigma^2_{\hat{\nabla}}$ (see also equation (28) of Section 4.1) and we may write (78) as:

(79)
$$\boxed{\underline{T}_{q=1} = \frac{\underline{\hat{\nabla}}^2}{\sigma^2_{\hat{\nabla}}}} .$$

With (75), (77) and (79) we have three expressions for the 1-dimensional $\underline{T}$-teststatistic or the square of the $\underline{w}$-teststatistic. The first two of them are the more useful ones because they do not need explicitly the results of least-squares computation under H_A. The first expression (75) is useful when the hypotheses are formulated in terms of condition equations. The second expression is however the most commonly used expression in practice.

An important application in geodetic practice of the 1-dimensional $\underline{T}$-teststatistic or the $\underline{w}$-teststatistic is for blunder detection in the observations. From experience we know that in geodetic applications misspecifications in the null hypothesis H_0 are very often caused by blunders or gross errors in the observations. Of course one never knows whether blunders are present, or how many of them are present, or in which observations they are present. In order to test for the presence of blunders we will therefore follow the convention that only one blunder is assumed to be present at a time. In this way only one additional explanatory variable is needed in the corresponding conventional alternative hypothesis H_{A_i}. For instance, if we want to test for the presence of a blunder in the *i*th- observation, the hypotheses take the form:

$$H_0\colon E\{\underline{y}\} = Ax \quad \text{versus} \quad H_{A_i}\colon E\{\underline{y}\} = (A \vdots c_{y_i}) \begin{pmatrix} x \\ \nabla_i \end{pmatrix}, \quad \nabla_i \neq 0. \tag{80}$$

with

$$\underset{m\times 1}{c_{y_i}} = (0,\ldots,\underset{i\text{th}}{1},0,\ldots)^*. \tag{81}$$

The corresponding test reads then:

$$\text{reject } H_0 \text{ if} \quad w_i < -k_\alpha^{1/2} \quad \text{or} \quad w_i > k_\alpha^{1/2} \tag{82}$$

with

$$\underline{w}_i = \frac{c_{y_i}^* Q_y^{-1} \underline{\hat{e}}}{\sqrt{c_{y_i}^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_{y_i}}}. \tag{83}$$

If test (82) comes to reject H_0, a blunder or gross error in the *i*th observation is suspected. Checking and/or remeasurement will then be necessary. By taking *i* in the above test to be successively 1,...,*m* the whole observations vector can be screened for observational blunders. This procedure is called *datasnooping*. Generally the observation with the largest value of (83), in absolute sense, should be rejected.

In many applications of datasnooping the variancematrix Q_y is a diagonal matrix. If this is the case, then (83) simplifies to:

$$\underline{w}_i = \frac{\underline{\hat{e}}_i}{\sigma_{\hat{e}_i}}. \tag{84}$$

A summary of the results of this section is given in Table 4.3.

Hypotheses

$$H_0\colon\ E\{\underline{y}\} = \boldsymbol{A}x \quad \text{versus} \quad H_{\boldsymbol{A}}\colon\ E\{\underline{y}\} = (\boldsymbol{A} \vdots c_y)\begin{pmatrix} x \\ \nabla \end{pmatrix}$$

or

$$H_0\colon\ \boldsymbol{B}^* E\{\underline{y}\} = 0 \quad \text{versus} \quad H_{\boldsymbol{A}}\colon\ \boldsymbol{B}^* E\{\underline{y}\} = c_t \nabla$$

$$c_t = \boldsymbol{B}^* c_y$$

Datasnooping

$$c_y\colon\ c_{y_i} = (0,0,\ldots,\underset{ith}{1},0,\ldots)^*$$
$$i = 1,\ldots,m$$

$\underline{w}$-Teststatistic

$$\underline{w} = \frac{c_t^* Q_t^{-1} \underline{t}}{\sqrt{c_t^* Q_t^{-1} c_t}} = \frac{c_y^* Q_y^{-1} \underline{\hat{e}}}{\sqrt{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y}} = \frac{\underline{\hat{\nabla}}}{\sigma_{\hat{\nabla}}}$$

Distribution of $\underline{w}$

$$H_0\colon\ \underline{w} \sim N(0,1) \quad ; \quad H_{\boldsymbol{A}}\colon\ \underline{w} \sim N(\nabla w, 1)$$
$$\nabla w = \sqrt{c_t^* Q_t^{-1} c_t}\,\nabla = \sqrt{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y}\,\nabla$$

Generalized likelihood ratio test

reject H_0 if $w < -k_\alpha^{1/2}$ or $w > +k_\alpha^{1/2}$

Table 4.3: Overview of $\underline{w}$-teststatistic.

4.4 The case $q = m\text{-}n$: $\underline{\hat{\sigma}}^2$-teststatistic

Consider again testing the hypotheses:

$$(85) \qquad H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{Ax} \quad \text{versus} \quad H_A : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{(A} \vdots \underset{m\times q}{C_y)} \begin{pmatrix} x \\ \nabla \end{pmatrix}_{(n+q)\times 1}, \ \nabla \neq 0$$

or

$$(86) \qquad H_0 : \underset{b\times m}{B^*} \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0} \quad \text{versus} \quad H_A : \underset{b\times m}{B^*} \underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times qq\times 1}{C_t \nabla}, \ \nabla \neq 0.$$

We know that the corresponding generalized likelihood ratio test reads:

$$(87) \qquad \text{reject } H_0 \text{ if } \quad T_q > k_\alpha. \quad \text{(see Figure 4.10).}$$

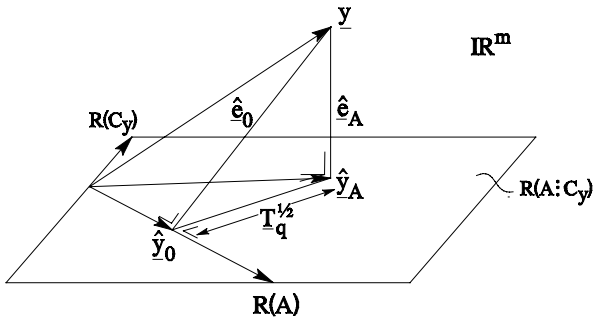


Figure 4.10: The teststatistic $\underline{T}_q$.

In the previous section we considered the case $q = 1$. In this section we consider the other extreme, namely $q = m\text{-}n$. For this case the following two expressions of $\underline{T}_q$ are of interest:

$$(88) \qquad \begin{aligned} \underline{T}_q &= \underline{\hat{e}}_0^* Q_y^{-1} \underline{\hat{e}}_0 - \underline{\hat{e}}_A^* Q_y^{-1} \underline{\hat{e}}_A \\ &= \underline{t}^* Q_t^{-1} C_t (C_t^* Q_t^{-1} C_t)^{-1} C_t^* Q_t^{-1} \underline{t}. \end{aligned}$$

Since it was assumed that $\text{rank}(A \vdots C_y) = n + q$, it follows that if $q = m\text{-}n$ then $\text{rank}(A \vdots C_y) = m$. In this case matrix C_y is chosen such that the matrix $(A \vdots C_y)$ is square and of full rank. But this means that no restrictions are placed on $E\{\underline{y}\}$ under H_A. That is, since $R(A \vdots C_y) = \mathbb{R}^n$ if $q=m\text{-}n$ we have $E\{\underline{y} | H_A\} \in \mathbb{R}^n$. In other words, by choosing $q=m\text{-}n$, the number of explanatory variables that are added to H_0 in order to form H_A are such that the *redundancy* (overtalligheid) of the linear model under H_A equals zero! But this implies that:

$$(89) \qquad \underline{\hat{y}}_A \equiv \underline{y} \quad \text{and} \quad \underline{\hat{e}}_A \equiv 0.$$

This is shown is Figure 4.11.

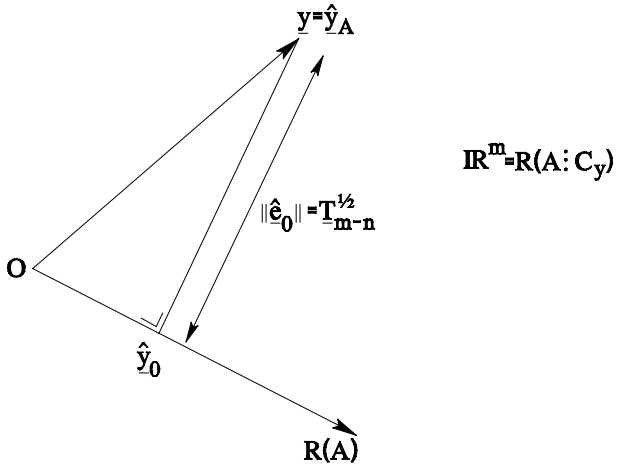


Figure 4.11: If $q = m\text{-}n \Rightarrow \mathbb{R}^m = \boldsymbol{R}(\boldsymbol{A} : C_y) \Rightarrow \underline{\hat{y}}_{\boldsymbol{A}} \equiv \underline{y}$, $\underline{\hat{e}}_{\boldsymbol{A}} \equiv 0$.

With (89) the first expression of (88) becomes:

(90) $$\boxed{\underline{T}_{\boldsymbol{q}=m-n} = \underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}}} \ .$$

We have again dropped the index "0", because it will be clear that the least-squares residual vector $\underline{\hat{e}}$ of (90) belongs to model H_0. Now let us see what happens with the second expression of (88) if $q = m–n$. If $q = m–n$, then the full rank matrix C_t of (86) has b-number of rows and $(m–n)$-number of columns. But we know that $b = m–n$. Hence, in case $q = m–n$ the matrix C_t is square and of full rank. But this means that the matrix C_t is invertible and therefore gets eliminated from the second expression of (88). Thus if $q = m–n$, then:

(91) $$\boxed{\underline{T}_{\boldsymbol{q}=m-n} = \underline{t}^* Q_t^{-1} \underline{t}} \ .$$

With (90) or (91) the generalized likelihood ratio test for testing the hypotheses:

(92) $$H_0 : E\{\underline{y}\} = \boldsymbol{A}x \quad \text{versus} \quad H_{\boldsymbol{A}} : E\{\underline{y}\} \in \mathbb{R}^m$$

or

(93) $$H_0 : \boldsymbol{B}^* E\{\underline{y}\} = 0 \quad \text{versus} \quad H_{\boldsymbol{A}} : E\{\underline{y}\} \in \mathbb{R}^m$$

reads:

(94) $$\text{reject } H_0 \text{ if } \quad T_{\boldsymbol{q}=m-n} > k_\alpha .$$

The distribution of $\underline{T}_{q=m-n}$ under H_0 and H_A follows from (42) as:

(95) $$\begin{cases} H_0 : \underline{T}_{\boldsymbol{q}=m-n} \sim \chi^2(m-n,0) \ ; \ H_{\boldsymbol{A}} : \underline{T}_{\boldsymbol{q}=m-n} \sim \chi^2(m-n,\lambda) \\ \text{with } \lambda = \nabla^* C_t^* Q_t^{-1} C_t \nabla = \nabla^* C_y^* Q_y^{-1} Q_{\hat{\boldsymbol{e}}} Q_y^{-1} C_y \nabla . \end{cases}$$

In many publications where the generalized likelihood ratio test for testing (92) or (93) is described one will see that not the teststatistic $\underline{T}_{q=m-n}$ is used, but instead of $\underline{T}_{q=m-n}$ the teststatistic $\underline{T}_{q=m-n}/(m-n)$. This teststatistic is denoted by $\underline{\hat{\sigma}}^2$. Thus:

(96)
$$\boxed{\underline{\hat{\sigma}}^2 = \frac{\underline{T}_{q=m-n}}{m-n} = \frac{\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}}}{m-n} = \frac{\underline{t}^* Q_t^{-1} \underline{t}}{m-n}}.$$

The test reads then instead of (94) as:

(97)
$$\text{reject } H_0 \text{ if } \quad \hat{\sigma}^2 > k'_\alpha, \quad k'_\alpha = k_\alpha/(m-n).$$

The distribution of $\underline{\hat{\sigma}}^2$ under H_0 and H_A is given as (see also appendix A):

(98)
$$\begin{cases} H_0 : \underline{\hat{\sigma}}^2 \sim F(m-n,\infty,0) \quad ; \quad H_A : \underline{\hat{\sigma}}^2 \sim F(m-n,\infty,\lambda) \\ \\ \text{with } \lambda = \nabla^* C_t^* Q_t^{-1} C_t \nabla = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \nabla. \end{cases}$$

It will be clear of course that test (97) is completely identical to test (94). Hence there is no special reason why the teststatistic $\underline{\hat{\sigma}}^2$ should be used instead if $\underline{T}_{q=m-n}$. However, there does exist a special reason why the notation "$\underline{\hat{\sigma}}^2$" is used in (96). Recall from Adjustment theory (Section 2.4) that:

(99)
$$E\{\underline{\hat{e}}^* Q_y^{-1} \underline{\hat{e}}\} = m-n.$$

This implies that if the variancematrix of $\underline{y}$ is given as $D\{\underline{y}\} = Q_y = \sigma^2 Q$, where σ^2 is the variance factor of unit weight, and thus $E\{\underline{\hat{e}}^* D\{\underline{y}\}^{-1} \underline{\hat{e}}\} = (m-n)$, we have:

$$E\{\underline{\hat{e}}^* Q^{-1} \underline{\hat{e}}\} = \sigma^2 (m-n)$$

or with: (96)

(100)
$$E\{\underline{\hat{\sigma}}^2\} = \sigma^2.$$

Hence, $\underline{\hat{\sigma}}^2$ can be considered an *unbiased estimator* of the *variance factor of unit weight* σ^2. This is the reason why the notation "$\underline{\hat{\sigma}}^2$" is used in (96).

The practical importance of the above given test ((94) or (97)) for testing (92) or (93) lies in the fact that no restrictions are imposed in the mean of $\underline{y}$ under H_A, that is $E\{\underline{y}|H_A\} \in \mathbb{R}^m$. In other words, for the case $q = m\text{-}n$ no matrix C_y or matrix C_t needs to be specified. This in contrast to all those cases for which $q < m\text{-}n$. For all those cases for which $q < m\text{-}n$ one needs to specify C_y or C_t, and therefore one has to have some idea of what kind of misspecifications to expect in H_0. In some cases this is possible. For instance, experience has learned that the class of conventional alternative hypotheses used in datasnooping is one class that should always be taken into account in geodetic network applications. But still this class may not cover the totality of

misspecifications in H_0 that occur in a particular application. In fact, one will never be able to completely specify the class of alternative hypotheses for a particular problem, simply because one never knows beforehand what misspecification has occured in H_0. In this light one should see test (94) or (97) as an important safeguard. The test gives an indication of the validity of H_0 without the need to specify the alternative hypothesis through C_y or C_t. As such it can be considered an overall model test. Appendix C elaborates on the relation between the overall model test and the *w*-test of the previous section. A summary of the results of this section is given in Table 4.4.

Hypotheses

$$H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{\boldsymbol{A}x} \quad \text{versus} \quad H_A : E\{\underline{y}\}\in\mathbb{R}^m$$

or

$$H_0 : \underset{b\times m}{\boldsymbol{B}^*}\underset{m\times 1}{E\{\underline{y}\}} = \underset{b\times 1}{0} \quad \text{versus} \quad H_A : E\{\underline{y}\}\in\mathbb{R}^m$$

$\underline{\hat{\sigma}}^2$ -Teststatistic

$$\underline{\hat{\sigma}}^2 = \frac{\underline{T}_{q=m-n}}{m-n} = \frac{\underline{\hat{e}}^* Q_y^{-1}\underline{\hat{e}}}{m-n} = \frac{\underline{t}^* Q_t^{-1}\underline{t}}{m-n}$$

Distribution of $\underline{\hat{\sigma}}^2$

$$H_0 : \underline{\hat{\sigma}}^2 \sim F(m-n,\infty,0) \;;\; H_A : \underline{\hat{\sigma}}^2 \sim F(m-n,\infty,\lambda)$$
$$\lambda = \nabla^* C_t^* Q_t^{-1} C_t \nabla = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \nabla$$

Generalized likelihood ratio test
overall model test

reject H_0 if $\hat{\sigma}^2 > k_\alpha$

Table 4.4: Overview of overall model test.

4.5 Internal reliability

From section 4.1 we know that the generalized likelihood ratio test for testing:

$$(101) \qquad H_0\colon \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times nn\times 1}{Ax} \quad \text{versus} \quad H_A\colon \underset{m\times 1}{E\{\underline{y}\}} = (\underset{m\times n}{A} \vdots \underset{m\times q}{C_y}) \underbrace{\begin{pmatrix} x \\ \nabla \end{pmatrix}}_{(n+q)\times 1}, \quad \nabla \neq 0$$

is given by:

$$(102) \qquad \text{reject } H_0 \text{ if } \quad T_q > \chi^2_\alpha(q,0)$$

where:

$$(103) \qquad \underline{T}_q = \underline{\hat{e}}^* Q_y^{-1} C_y (C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y)^{-1} C_y^* Q_y^{-1} \underline{\hat{e}}$$

and

$$(104) \qquad H_0 : \underline{T}_q \sim \chi^2(q,0) \;\; ; \;\; H_A : \underline{T}_q \sim \chi^2(q,\lambda)$$

with

$$(105) \qquad \lambda = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \nabla.$$

In (102) we have used the notation " $\chi^2_\alpha(q,0)$ " for the critical value instead of the notation " k_α ". The notation " $\chi^2_\alpha(q,0)$ " makes it clearer that the critical value should be computed from the central χ^2 -distribution with q degrees of freedom. Instead of test (102) we may also write:

$$(106) \qquad \text{reject } H_0 \text{ if } \quad \frac{T_q}{q} > F_\alpha(q,\infty,0)$$

where:

$$(107) \qquad H_0 : \frac{\underline{T}_q}{q} \sim F(q,\infty,0) \;\; ; \;\; H_A : \frac{\underline{T}_q}{q} \sim F(q,\infty,\lambda).$$

The two tests (102) and (106) are of course identical. In order to perform the generalized likelihood ratio test (102) or (106), one needs to compute the critical value, $\chi^2_\alpha(q,0)$ or $F_\alpha(q,\infty,0)$, for a chosen size α and a fixed number q of degrees of freedom. Let us denote the probability density distributions of $\chi^2(q,\lambda)$ and $F(q,\infty,\lambda)$ respectively by $p_{\underline{\chi^2}}(\chi^2|q,\lambda)$ and $p_{\underline{F}}(F|q,\infty,\lambda)$. Then:

$$(108) \qquad \begin{aligned} \alpha &= \int_{\chi^2_\alpha(q,0)}^{\infty} p_{\underline{\chi^2}}(\chi^2|q,0)\,\mathrm{d}\chi^2 \\ &= \int_{F_\alpha(q,\infty,0)}^{\infty} p_{\underline{F}}(F|q,\infty,0)\,\mathrm{d}F. \end{aligned}$$

These relations can be used to compute the critical values $\chi^2_\alpha(q,0)$ or $F_\alpha(q,\infty,0)$ from α and q. Standard tables exist that give $\chi^2_\alpha(q,0)$ or $F_\alpha(q,\infty,0)$ for various values α and q (see appendix B). Some typical values of $\chi^2_\alpha(q,0)$ or $F_\alpha(q,\infty,0)$ are given in Table 4.5 and Table 4.6 respectively. Note from these tables that for a fixed number of degrees of freedom, the critical values $\chi^2_\alpha(q,0)$ or $F_\alpha(q,\infty,0)$ get smaller for larger α. This is also what one would expect. One would expect that if H_0 is true, the occurence of large values of T_q in (102) is less frequent than the occurence of smaller values of T_q.

	q=1	q=10	q=20	q=30
α = 0.001	10.83	29.59	45.31	59.70
α = 0.005	7.88	25.19	40.00	53.67
α = 0.01	6.63	23.21	37.57	50.89
α = 0.05	3.84	18.31	31.41	43.77
α = 0.1	2.71	15.99	28.41	40.26

Table 4.5: Critical values $\chi^2_\alpha(q,0)$.

	q=1	q=10	q=20	q=30
α = 0.001	10.83	2.96	2.27	1.99
α = 0.005	7.88	2.52	2.00	1.79
α = 0.01	6.63	2.32	1.88	1.70
α = 0.05	3.84	1.83	1.57	1.46
α = 0.1	2.71	1.60	1.42	1.34

Table 4.6: Critical values $F_\alpha(q,\infty,0)$.

Also note from Table 4.5 that for a fixed α, the critical values $\chi^2_\alpha(q,0)$ get larger for larger q. This is also what one would expect. Since the χ^2-distribution is defined as a sum of squares of independent standard normal random variables, one would expect that the right tail of the χ^2-distribution gets thicker for larger sums (see Figure 4.12). Note on the other hand from Table 4.6 that for a fixed size α, the critical values $F_\alpha(q,\infty,0)$ get smaller for larger q. This is of course due to the division by q in (106).

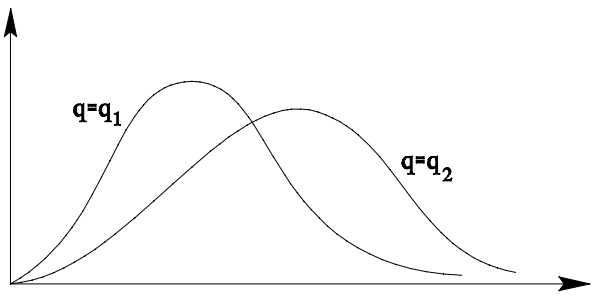


Figure 4.12: Density of $\chi^2(q,0)$ for $q_2 > q_1$.

In Section 1.5 where the general steps for testing hypotheses were outlined, it was pointed out that one should compute the size of the type II error in order to ensure that a reasonable protection exists against type II errors. Since the size of a type II error equals one minus the power of the test, we might as well compute the power γ. The power of test (102) or (106) follows as:

$$\begin{aligned}\gamma &= \int_{\chi^2_\alpha(q,0)}^{\infty} p_{\chi^2}(\chi^2|q,\lambda)\mathrm{d}\chi^2 \\ &= \int_{F_\alpha(q,\infty,o)}^{\infty} p_F(F|q,\infty,\lambda)\mathrm{d}F.\end{aligned} \tag{109}$$

Note that the power γ depends on: 1) the chosen size α; 2) the number of degrees of freedom q; 3) the non-centrality parameter λ. In Table 4.7 some typical values of γ are given. Table 4.7 shows that the power γ gets larger if the size α of the test is chosen larger. This is also what one would expect. A larger size α implies a smaller critical value $\chi^2_\alpha(q,0)$ or $F_\alpha(q,\infty,0)$, and therefore with (109) a larger power γ. Table 4.7 also shows that the power γ gets smaller for larger q.

$\alpha = 0.01$	$q = 1$	$q = 7$
$\lambda = 2$	0.1227	0.0415
$\lambda = 8$	0.5997	0.2710
$\lambda = 18$	0.9522	0.7430

$\alpha = 0.05$	$q = 1$	$q = 7$
$\lambda = 2$	0.2930	0.1378
$\lambda = 8$	0.8074	0.5017
$\lambda = 18$	0.9888	0.8946

$\alpha = 0.1$	$q = 1$	$q = 7$
$\lambda = 2$	0.4099	0.2272
$\lambda = 8$	0.8817	0.6287
$\lambda = 18$	0.9953	0.9413

Table 4.7: The power of test (102) or (106) for different values of α, q and λ.

This is understandable if one thinks of q as the number of additional parameters in H_A. The smaller q is, the less additional parameters are used in H_A and therefore the more "information" is used in formulating H_A. For such an alternative hypothesis one would expect that if H_A is true the probability of accepting it is higher. Finally note that Table 4.7 shows that the power gets larger if the *non-centrality parameter* λ gets larger. This is understandable if one looks at the geometry of the testing problem. Substitution of:

(110) $$Q_y^{-1} Q_{\hat{e}} Q_y^{-1} = P_A^{\perp *} Q_y^{-1} P_A^{\perp}$$

into (105) shows that:

(111) $$\lambda = \|P_A^{\perp} C_y \nabla\|^2 .$$

Since $E\{\underline{y}|H_A\} = E\{\underline{y}|H_0\} + C_y\nabla$ (see (101)), it follows that:

(112) $$\|E\{\underline{y}|H_A\} - E\{\underline{y}|H_0\}\|^2 = \|C_y\nabla\|^2 .$$

Thus $\|C_y\nabla\|$ is the separation or distance between H_0 and H_A (see Figure 4.13). Now, one would expect that the power of the test increases if the distance between H_0 and H_A, thus $\|C_y\nabla\|$, increases. But $\|C_y\nabla\|$ gets larger if λ of (111) gets larger. Hence, one would indeed expect that the power gets larger if λ gets larger.

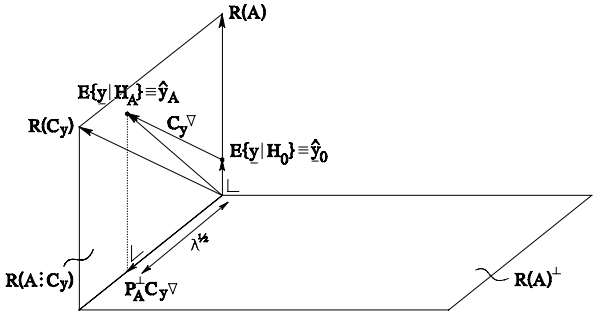


Figure 4.13: $E\{\underline{y}|H_A\} = E\{\underline{y}|H_0\} + C_y\nabla$; $\|C_y\nabla\|^2$ larger if $\|P_A^{\perp}C_y\nabla\|^2 = \lambda$ larger.

We may summarize the above discussion as [Ghosh, 1973]:

(i) The power γ of test (102) or (106) is monotonic increasing in α for fixed q and λ
(ii) The power γ of test (102) or (106) is monotonic decreasing in q for fixed α and λ
(iii) The power γ of test (102) or (106) is monotonic increasing in λ for fixed α and q.

Since the power γ of the test (102) or (106) depends on α, q and λ, it seems that we have three possibilities to construct a test which has a reasonable protection against type II errors. We could increase α. But increasing α implies increasing the probability of a type I error. The size α is therefore usually chosen at a fixed value. We could also decrease q. But usually we are not free in choosing q. The value of q depends on the particular alternative hypothesis against which one wants to test H_0. Finally one could try to increase the non-centrality parameter λ. What possibilities do we have to increase λ? With:

$$Q_{\hat{e}} = Q_y - A(A^* Q_y^{-1} A)^{-1} A^*$$

it follows from (105) that:

(113) $$\lambda = (C_y\nabla)^*[Q_y^{-1} - Q_y^{-1}A(A^*Q_y^{-1}A)^{-1}A^*Q_y^{-1}](C_y\nabla).$$

This formula shows that the non-centrality parameter λ depends on:

(114)
$$\begin{cases} \text{(i)} & Q_y & \text{the precision of the observables} \\ \text{(ii)} & A & \text{the designmatrix} \\ \text{(iii)} & C_y\nabla & \text{the difference of } E\{\underline{y}|H_A\} \text{ and } E\{\underline{y}|H_0\}. \end{cases}$$

Let us now investigate what the effect on λ is when either Q_y, A or $C_y\nabla$ are changed.

ad i:

It will intuitively be clear that one can increase λ by increasing the precision of the observables. For instance if one uses μQ_y, where μ is a positive scalar, instead if Q_y, then the non-centrality parameter λ_μ becomes (see (113)):

$$\lambda_\mu = \mu^{-1}\lambda.$$

This shows that the non-centrality parameter increases if μ decreases, that is if the observables have a higher precision. Compare this with example 4 of Section 2.2. The dependence of λ on Q_y and therefore the dependence of the power γ of the test on Q_y, makes it possible to obtain a test with sufficient power if the variance matrix Q_y is appropriately chosen. Since Q_y depends on the precision of the measurement equipment, an appropriate choice of measurement equipment enables one to obtain a test with sufficient power.

ad ii:

In geodetic network applications matrix A depends on the structure of the network. Hence by changing the structure of the network one changes A and therefore also changes λ. This is an important result, because it shows that one can look for a design or structure of a network that is optimal in the sense that it gives a test with sufficient power. It will intuitively be clear that one can increase λ and therefore also increase the power of the test, by increasing the number of observables. In order to prove this, let us consider the following two situations. We have a network for which the following model holds:

(115) $$\underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{A}\ \underset{n\times 1}{x} \quad , \quad \underset{m\times m}{D\{\underline{y}\}} = \underset{m\times m}{Q_y}$$

and we have a network for which holds:

$$E\{\underset{(m+1)\times 1}{\begin{pmatrix}\underline{y}\\ \underline{z}\end{pmatrix}}\} = \underset{(m+1)\times n \quad n\times 1}{\begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{a}^*\end{pmatrix}x} \quad , \quad \underset{(m+1)\times(m+1)}{\boldsymbol{D}\{\begin{pmatrix}\underline{y}\\ \underline{z}\end{pmatrix}\}} = \underset{(m+1)\times(m+1)}{\begin{pmatrix}Q_y & 0\\ 0 & \sigma_z^2\end{pmatrix}}. \tag{116}$$

The two models, that is the two networks, differ in the sense that the second model consists of the first model plus one additional observation equation, namely $E\{\underline{z}\} = a^*x$. In terms of condition equations the two models can be written as:

$$\underset{(m-n)\times m \; m\times 1}{\boldsymbol{B}^*E\{\underline{y}\}} = \underset{(m-n)\times 1}{0} \quad , \quad \underset{m\times m}{\boldsymbol{D}\{\underline{y}\}} = \underset{m\times m}{Q_y} \tag{117}$$

and

$$\underset{(m-n+1)\times(m+1)}{\begin{pmatrix}\boldsymbol{B}^* & 0\\ b_1^* & b_2\end{pmatrix}} \; \underset{(m+1)\times 1}{E\{\begin{pmatrix}\underline{y}\\ \underline{z}\end{pmatrix}\}} = \underset{(m-n+1)\times 1}{\begin{pmatrix}0\\ 0\end{pmatrix}} \quad , \quad \underset{(m+1)\times(m+1)}{\boldsymbol{D}\{\begin{pmatrix}\underline{y}\\ \underline{z}\end{pmatrix}\}} = \underset{(m+1)\times(m+1)}{\begin{pmatrix}Q_y & 0\\ 0 & \sigma_z^2\end{pmatrix}} \tag{118}$$

with $B^*A = 0$ and $b_1^*A + b_2a^* = 0$. Note that the additional observation equation in (116) implies an additional condition equation in (118). We will now show that the non-centrality parameter of model (118), denoted by λ_b, is always larger than the non-centrality parameter of model (116) denoted by λ, The non-centrality parameter of model (116) reads (see (42)):

$$\lambda = (C_y\nabla)^*\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{B}^*(C_y\nabla). \tag{119}$$

Similarly we find for the non-centrality parameter of model (118):

$$\lambda_b = (\begin{pmatrix}C_y\\ c_z\end{pmatrix}\nabla)^*\begin{pmatrix}\boldsymbol{B} & b_1\\ 0 & b_2\end{pmatrix}\begin{pmatrix}\boldsymbol{B}^*Q_y\boldsymbol{B} & \boldsymbol{B}^*Q_yb_1\\ b_1^*Q_y\boldsymbol{B} & (b_1^*Q_yb_1 + b_2^2\sigma_z^2)\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{B}^* & 0\\ b_1^* & b_2\end{pmatrix}(\begin{pmatrix}C_y\\ c_z\end{pmatrix}\nabla). \tag{120}$$

Since:

$$\begin{pmatrix}\boldsymbol{B}^*Q_y\boldsymbol{B} & \boldsymbol{B}^*Q_yb_1\\ b_1^*Q_y\boldsymbol{B} & (b_1^*Q_yb_1 + b_2^2\sigma_z^2)\end{pmatrix} =$$

$$\begin{pmatrix}I & 0\\ b_1^*Q_y\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1} & I\end{pmatrix}\begin{pmatrix}\boldsymbol{B}^*Q_y\boldsymbol{B} & 0\\ 0 & b_1^*[Q_y - Q_y\boldsymbol{B}(\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{B}^*Q_y]b_1 + b_2^2\sigma_z^2\end{pmatrix}\begin{pmatrix}I & (\boldsymbol{B}^*Q_y\boldsymbol{B})^{-1}\boldsymbol{B}^*Q_yb_1)\\ 0 & I\end{pmatrix}$$

(verify this yourself), inversion gives:

$$\begin{pmatrix} \boldsymbol{B}^* Q_y \boldsymbol{B} & \boldsymbol{B}^* Q_y b_1 \\ b_1^* Q_y \boldsymbol{B} & (b_1^* Q_y b_1 + b_2^2 \sigma_z^2) \end{pmatrix}^{-1} =$$

$$\begin{pmatrix} I & -(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* Q_y b_1 \\ 0 & I \end{pmatrix} \begin{pmatrix} (\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} & 0 \\ 0 & \{b_1^* [Q_y - Q_y \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* Q_y] b_1 \\ & + b_2^2 \sigma_z^2\}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -b_1^* Q_y \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} & I \end{pmatrix}.$$

From this follows with (120) that:

$$\begin{pmatrix} \boldsymbol{B} & b_1 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{B}^* Q_y \boldsymbol{B} & \boldsymbol{B}^* Q_y b_1 \\ b_1^* Q_y \boldsymbol{B} & (b_1^* Q_y b_1 + b_2^2 \sigma_z^2) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{B}^* & 0 \\ b_1^* & b_2 \end{pmatrix} =$$

$$\begin{pmatrix} \boldsymbol{B} & [I - \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* Q_y] b_1 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} (\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} & 0 \\ 0 & \{b_1^* [Q_y - Q_y \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \\ & \boldsymbol{B}^* Q_y] b_1 + \sigma_z^2\}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}^* & 0 \\ b_1^* [I - Q_y \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^*] & b_2 \end{pmatrix}.$$

(121)

If we abbreviate $[I - B(B^* Q_y B)^{-1} B^* Q_y] b_1$ as:

(122) $$\bar{b}_1 = [I - \boldsymbol{B}(\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} \boldsymbol{B}^* Q_y] b_1$$

we can write (121) as:

(123)
$$\begin{pmatrix} \boldsymbol{B} & b_1 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{B}^* Q_y \boldsymbol{B} & \boldsymbol{B}^* Q_y b_1 \\ b_1^* Q_y \boldsymbol{B} & (b_1^* Q_y b_1 + b_2^2 \sigma_z^2) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{B}^* & 0 \\ b_1^* & b_2 \end{pmatrix} =$$
$$\begin{pmatrix} \boldsymbol{B} & \bar{b}_1 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} (\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} & 0 \\ 0 & (\bar{b}_1^* Q_y \bar{b}_1 + b_2^2 \sigma_z^2)^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}^* & 0 \\ \bar{b}_1^* & b_2 \end{pmatrix}.$$

Substitution of (123) into (120) gives:

(124) $$\lambda_b = \left(\begin{pmatrix} \boldsymbol{B}^* C_y \\ \bar{b}_1^* C_y + b_2 c_z \end{pmatrix} \nabla \right)^* \begin{pmatrix} (\boldsymbol{B}^* Q_y \boldsymbol{B})^{-1} & 0 \\ 0 & (\bar{b}_1^* Q_y \bar{b}_1 + b_2^2 \sigma_z^2)^{-1} \end{pmatrix} \left(\begin{pmatrix} \boldsymbol{B}^* C_y \\ \bar{b}_1^* C_y + b_2 c_z \end{pmatrix} \nabla \right).$$

From this follows with (119) that:

(125) $$\lambda_b = \lambda + [(\bar{b}_1^* C_y + b_2 c_z) \nabla]^* (\bar{b}_1^* Q_y \bar{b}_1 + b_2^2 \sigma_z^2)^{-1} [(\bar{b}_1^* C_y + b_2 c_z) \nabla].$$

Since the quadratic form on the right-hand side of (125) is always non negative, equation (125) shows that:

(126) $$\lambda_b \geq \lambda .$$

This shows that the power of the test indeed gets larger if the number of observations or the number of condition equations gets larger. Compare this with Example 6 of Section 2.3.

ad iii:

Equation (113) shows that λ and therefore the power of the test can be changed by changing $C_y\nabla = E\{\underline{y}|H_A\} - E\{\underline{y}|H_0\}$. From Figure 4.13 we learn that in general λ gets larger if the separation between $E\{\underline{y}|H_A\}$ and $E\{\underline{y}|H_0\}$ is increased. Note however that the component of $C_y\nabla$ which lies in $R(A)$ has no effect on λ. In practice of course $C_y\nabla$ is unknown. Hence one will never be able to compute the actual power of the test. Still, by choosing some representative values for the separation, $C_y\nabla = E\{\underline{y}|H_A\} - E\{\underline{y}|H_0\}$, between H_A and H_0, one can compute what the power of the test would be if $C_y\nabla$ were the "true" separation. In this way one can find out how well the test can detect a particular misspecification $C_y\nabla$ in H_0. For instance in blunder detection the scalar ∇ models the size of the blunder. By choosing a representative value for the blunder, one can compute through λ the probability that the test will detect the blunder with the chosen size ∇. If one considers this probability too low, one has two possibilities to increase this probability, either by changing Q_y or by changing A.

So far we have been concentrating on the power γ of the test, that is, on the probability of rejecting H_0 when in fact H_A is true. We have seen that the power γ can be computed from the size of the test, α, from the degrees of freedom, q, and from the non-centrality parameter λ. Symbolically this may be written as:

(127) $$\gamma = \gamma(\alpha, q, \lambda).$$

In geodetic practice one is however not so much interested in the power of the test. One is much more interested in the misspecification or modelerror $C_y\nabla$ that generates γ. That is, one is much more interested in the model error that can be detected with a certain probability γ. The approach taken in geodetic practice is therefore to fix γ at a reference value γ_0, for instance $\gamma_0 = 50\%$, or 60%, or 70%, but usually 80%. From α, q and the chosen reference value $\gamma = \gamma_0$ one can then compute the corresponding value for the non-centrality parameter, symbolically:

(128) $$\lambda_0 = \lambda(\alpha, q, \gamma = \gamma_0).$$

The non-centrality parameter plays an important role in linking the overall model test and the *w*-test in Appendix C. From $\lambda = \lambda_0$ one can now compute the corresponding modelerror $C_y\nabla$. This is done by solving the quadratic form (see (105)):

(129) $$\boxed{\lambda_0 = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \nabla}$$

for ∇. Once ∇ is known, the modelerror $\nabla y = E\{\underline{y}|H_A\} - E\{\underline{y}|H_0\}$ follows as:

(130)
$$\underset{m\times 1}{\nabla y} = \underset{m\times q}{C_y}\ \underset{q\times 1}{\nabla}.$$

The $m\times 1$ vector ∇y is said to describe the *internal reliability* (inwendige betrouwbaarheid) of H_0 with respect to H_A. One should not confuse the geodetic usage of the word "betrouwbaarheid" with its usage in mathematical statistics. The internal reliability as described by ∇y is thus a measure of the model error that can be detected with a probability $\gamma = \gamma_0$ by test (102) or (106). How can we compute the $q\times 1$ vector ∇ from (129)? Unfortunately (129) has no unique solution for ∇. We will consider the following two cases: $q = 1$ and $1 < q \leq m-n$.

The case $q=1$: If $q=1$, then the $m\times q$ matrix C_y reduces to the $m\times 1$ vector c_y, and the $q\times 1$ vector ∇ reduces to the scalar ∇. For this case equation (129) can also be written as:

$$\lambda_0 = c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y \nabla^2.$$

The solution in terms of ∇ reads therefore:

(131)
$$|\nabla| = \left(\frac{\lambda_0}{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y} \right)^{1/2}$$

($|\nabla|$ is called a *minimal detectable bias* (grenswaarde)).

Note that one is only able to determine the size of ∇, but not its sign. In order to give a geometric interpretation to (131), recall that:

$$Q_y^{-1} Q_{\hat{e}} Q_y^{-1} = P_A^{\perp *} Q_y^{-1} P_A^{\perp} = Q_y^{-1} P_A^{\perp}.$$

Hence:

(132)
$$c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y = (P_A^{\perp} c_y)^* Q_y^{-1} (P_A^{\perp} c_y) = \|P_A^{\perp} c_y\|^2$$

and

(133)
$$c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y = c_y^* Q_y^{-1} (P_A^{\perp} c_y) = \|c_y\|.\|P_A^{\perp} c_y\| \cos\theta$$

where use is made of the cosine rule. From (132) and (133) follows that:

(134)
$$c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y = \|c_y\|^2 \cos^2\theta.$$

A sketch is given in Figure 4.14.

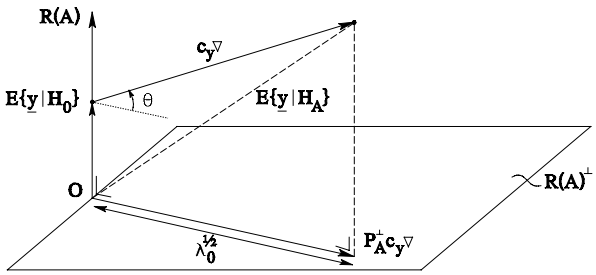


Figure 4.14: $E\{\underline{y}|H_A\} = E\{\underline{y}|H_0\} + c_y\nabla$.

Formula (134) shows that the denominator of (131) is small, and thus $|\nabla|$ is large, if the angle θ is close to $\frac{1}{2}\pi$. Thus $|\nabla|$ gets smaller and the internal reliability improves, the smaller the angle θ between c_y and $R(A)^\perp$ gets. If $\theta = \frac{1}{2}\pi$, then $c_y \in R(A)$ and $|\nabla| = \infty$. This implies that the corresponding model error can never be detected by the test. The internal reliability is then said to be infinitely poor. Since $0 \leq \cos^2\theta \leq 1$, it follows from (134) and (131) that:

(135) $$\left(\frac{\lambda_0}{c_y^* Q_y^{-1} c_y}\right)^{1/2} \leq |\nabla| \leq \infty \quad .$$

In case of *datasnooping* we have:

(136) $$c_y := c_{y_i} = (0,\cdots,\underset{\uparrow}{1},0,\cdots)^* .$$
$$i\text{th}$$

For this case the bound of (135) can be written as:

(137) $$|\nabla_i| \geq \sigma_{y_i}(\lambda_0)^{1/2} .$$

In many practical applications the variance matrix Q_y is a *diagonal matrix* (see also (84)). If Q_y is diagonal, it follows with the choice (136) that:

$$c_{y_i}^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_{y_i} = \sigma_{y_i}^{-4} c_{y_i}^* Q_{\hat{e}} c_{y_i}$$

or with $Q_{\hat{e}} = Q_y - Q_{\hat{y}}$ that:

(138) $$c_{y_i}^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_{y_i} = \sigma_{y_i}^{-4}(\sigma_{y_i}^2 - \sigma_{\hat{y}_i}^2) .$$

Substitution of (138) into (131) gives then for the minimal detectable bias:

(139) $$|\nabla_i| = \sigma_{y_i}\left(\frac{\lambda_0}{(1 - \sigma_{\hat{y}_i}^2/\sigma_{y_i}^2)}\right)^{1/2} .$$

This shows that $|\nabla_i|$ is large if $\sigma_{\hat{y}_i}^2$ is close to $\sigma_{y_i}^2$, and that $|\nabla_i|$ is small if $\sigma_{\hat{y}_i}^2$ is small compared to $\sigma_{y_i}^2$. The dimensionless number:

(140) $$r_i = 1 - \sigma^2_{\hat{y}_i} / \sigma^2_{y_i}$$

is called the *i*th *local redundancy number*. Note that since $0 \leq \sigma^2_{\hat{y}_i} \leq \sigma^2_{y_i}$, the *i*th local redundancy number r_i always lies in the closed interval:

(141) $$0 \leq r_i \leq 1.$$

The reason why r_i is called the *i*th local redundancy number follows from the fact that:

(142) $$\sum_{i=1}^{m} r_i = \sum_{i=1}^{m} (1 - \sigma^2_{\hat{y}_i} / \sigma^2_{y_i}) = m - n.$$

Thus the sum of the local redundancy numbers equals the total redundancy. The proof of (142) goes as follows. From (140) follows that:

$$r_i = c^*_{y_i}(I - Q_{\hat{y}} Q_y^{-1}) c_{y_i} = c^*_{y_i}(I - P_A) c_{y_i} = c^*_{y_i} P_A^{\perp} c_{y_i}.$$

Hence:

(143) $$\sum_{i=1}^{m} r_i = \sum_{i=1}^{m} c^*_{y_i} P_A^{\perp} c_{y_i} = \text{trace } P_A^{\perp}.$$

From Linear algebra you know that the trace of a matrix equals the sum of its eigenvalues. Thus:

(144) $$\text{trace } P_A^{\perp} = \sum_{i=1}^{m} \lambda_i$$

where $\lambda_i, i = 1,...,m,$ are the m eigenvalues of $P_A^{\perp}$. We know that $P_A^{\perp}$ is an orthogonal projector with the properties:

(145) $$\begin{cases} P_A^{\perp} z = z \text{ for } z \in R(A)^{\perp} \\ P_A^{\perp} z = 0 \text{ for } z \in R(A). \end{cases}$$

Since $\dim R(A) = n$ and $\dim R(A)^{\perp} = m - n$, it follows from (145) that $P_A^{\perp}$ has $(m - n)$ number of eigenvalues that equal 1 and n number of eigenvalues that equal 0. This together with (144) and (143) shows that (142) must hold. Since the sum of the local redundancy numbers equals the total redundancy $m - n$, we may define the average redundancy $\bar{r}$ as:

(146) $$\bar{r} = \frac{\sum_{i=1}^{m} r_i}{m} = \frac{m - n}{m}.$$

If we replace the local redundancy numbers in (139) by the average redundancy, we get the following rough approximation of $|\nabla_i|$:

(147) $$|\nabla_i| \approx \sigma_{y_i}\left(\frac{\lambda_0}{(m-n)/m}\right)^{1/2}.$$

The case $1 < q \leq m-n$: For the case $q = 2$ the quadratic form:

(148) $$\lambda_0 = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \nabla$$

describes an ellipse, for the case $q = 3$ it describes an ellipsoid and for the case $q > 3$ it describes a hyperellipsoid. In order to get a form that resembles formula (131) we parametrize the vector ∇ as:

(149) $$\underset{q\times 1}{\nabla} = \|\nabla\| \underset{q\times 1}{d}, \quad \text{with } \underset{q\times 1}{d} \text{ a unit vector}.$$

Substitution of (149) into (148) gives:

$$\lambda_0 = d^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y d \, \|\nabla\|^2.$$

This together with (149) shows that ∇ may be written as:

(150) $$\boxed{\underset{q\times 1}{\nabla} = \left(\frac{\lambda_0}{d^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y d}\right)^{1/2} \underset{q\times 1}{d} \quad \text{with } d \text{ a unit vector}}.$$

By letting the vector d scan the unit sphere in $\mathbb{R}^q$, the vector ∇ of (150) scans the ellipsoid as described by (148). If one is interested in the principle axes of the ellipsoid (148) one should choose d as one of the q number of eigenvectors of the matrix $C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y$:

$$C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y d_k = \lambda_k d_k \quad k = 1,2,...,q.$$

For the principal axes, expression (150) reduces then to:

(151) $$\underset{q\times 1}{\nabla_k} = \left(\frac{\lambda_0}{\lambda_k}\right)^{1/2} \underset{q\times 1}{d_k}, \quad k = 1,2,...,q$$

where d_k is a normalized eigenvector and λ_k is the corresponding eigenvalue.

We have seen that the model error that can be detected with a probability $\gamma = \gamma_0$ is given by the $m\times 1$ vector $\nabla y = C_y \nabla$. In some practical applications however it can be rather cumbersome to evaluate ∇y. Note namely that the number of vectors ∇y that need to be evaluated equals the number of alternative hypotheses H_A considered. This implies that one has to evaluate the m elements of ∇y for every alternative hypothesis considered. This amounts to a lot of evaluations and may therefore not be very practical. A notable exception occurs in case of datasnooping, where the vector ∇y has only one non-zero element. In order to reduce the number of

evaluations one could try to replace the vectorial measure ∇y by a scalar measure. The measure λ_y defined below is a scalar measure that can be used as such. If we consider ∇y as a possibly non-detected "bias" in $\underline{y}$ and the variance matrix Q_y as a description of the "noise" in $\underline{y}$, we may define a scalar squared *bias-to-noise* ratio [Papoulis, 1985] for $\underline{y}$ as:

(152) $$\boxed{\lambda_y = \nabla y^* Q_y^{-1} \nabla y} .$$

A large value of λ_y indicates that the model error ∇y is significant, and a small value of λ_y indicates that the model error is insignificant. Note that $\lambda_y = \|C_y \nabla\|^2$. Thus λ_y is the separation squared between $E\{\underline{y}|H_0\}$ and $E\{\underline{y}|H_A\}$ (see Figure 4.14).

Substitution of $\nabla y = C_y \nabla$ with (150) for the case $1 < q \leq m\text{-}n$ into (152) gives:

(153) $$\lambda_y = \frac{d^* C_y^* Q_y^{-1} C_y d}{d^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y d} \lambda_0.$$

In case of datasnooping the case $q=1$ with a diagonal variancematrix Q_y, formula (153) simplifies to:

(154) $$\lambda_{y_i} = (1 - \sigma_{\hat{y}_i}^2 / \sigma_{y_i}^2)^{-1} \lambda_0.$$

Let us denote the maximum value of the ratio in (153) by λ_{max}. Thus:

$$\max_{d \in \mathbb{R}^q} \frac{d^* C_y^* Q_y^{-1} C_y d}{d^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y d} = \lambda_{max}.$$

With λ_{max} we have the following upperbound for λ_y:

$$\lambda_y \leq \lambda_{max} \lambda_0.$$

Recall from Linear algebra that λ_{max} equals the largest eigenvalue of the generalized eigenvalue problem:

(155) $$|C_y^* Q_y^{-1} C_y - \lambda C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y| = 0.$$

A summary of the results of this section is given in Table 4.8.

Example 1

Figure 4.15 shows a typical levelling network of four points with two loops.

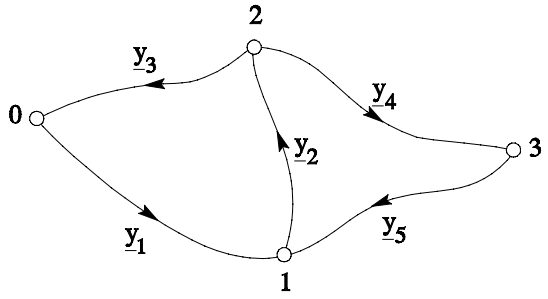


Figure 4.15: A levelling network.

We assume that the variance matrix of the normally distributed observables is equal to a scaled identity matrix. The linear model of condition equations reads then:

(156)
$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} E\{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \\ \underline{y}_5 \end{pmatrix}\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad Q_y = \sigma^2 I_5 .$$

We are interested in the minimal detectable bias $|\nabla_2|$ of $\underline{y}_2$. Since $c_{y_2} = (01000)^*$ and Q_y is diagonal we may use formula (139). Computation of $\sigma^2_{\hat{y}_2}$ according to:

(157)
$$\sigma^2_{\hat{y}_2} = c^*[Q_y - Q_y B(B^* Q_y B)^{-1} B^* Q_y] c_{y_2}$$

gives with (156):

$$\sigma^2_{\hat{y}_2} = \tfrac{1}{2}\sigma^2 .$$

With $\sigma^2_{y_2} = \sigma^2$, this gives for (139) and for (152):

(158)
$$|\nabla_2| = \sigma(2\lambda_0)^{1/2}$$
$$\sqrt{\lambda_{y_2}} = (2\lambda_0)^{1/2} .$$

Now consider the network of Figure 4.16.

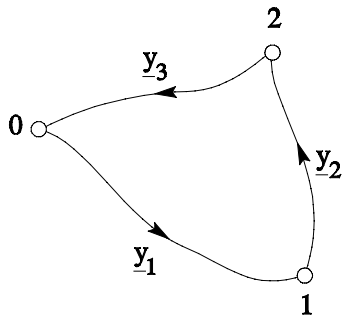


Figure 4.16: A levelling network of one loop.

Its linear model of condition equations reads:

$$(159) \qquad (1 \quad 1 \quad 1)\, E\left\{\begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \\ \underset{\sim}{y}_3 \end{pmatrix}\right\} = 0\ ; \quad Q_y = \sigma^2 I_3 .$$

Again we are interested in the minimal detectable bias $|\nabla_2|$ of $\underset{\sim}{y}_2$. Computation of $\sigma^2_{\hat{y}_2}$ according to (157) gives for the model (159):

$$\sigma^2_{\hat{y}_2} = \tfrac{2}{3}\sigma^2 .$$

With $\sigma^2_{y_2} = \sigma^2$, this gives for (139) and for (152):

$$(160) \qquad |\nabla_2| = \sigma(3\lambda_0)^{1/2}$$

$$\sqrt{\lambda_2} = (3\lambda_0)^{1/2} \ .$$

Comparison of (158) with (160) shows that a blunder in the second observation is better detectable with the two loop network than with the one loop network. Compare this with our discussion in Example 6 of Section 2.3.

Hypotheses

$$H_0 : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{\boldsymbol{A}}\,\underset{n\times 1}{x}$$

versus

$$H_{\boldsymbol{A}} : \underset{m\times 1}{E\{\underline{y}\}} = \underset{m\times n}{(\boldsymbol{A}}\,\vdots\,\underset{m\times \boldsymbol{q}}{C_y})\begin{pmatrix} x \\ \nabla \end{pmatrix}_{(n+\boldsymbol{q})\times 1}$$

or

$$H_0 : \underset{\boldsymbol{b}\times m}{\boldsymbol{B}^*}\,\underset{m\times 1}{E\{\underline{y}\}} = \underset{\boldsymbol{b}\times 1}{0}$$

versus

$$H_{\boldsymbol{A}} : \underset{\boldsymbol{b}\times m}{\boldsymbol{B}^*}\,\underset{\boldsymbol{b}\times m}{E\{\underline{y}\}} = \underset{\boldsymbol{b}\times \boldsymbol{q}\boldsymbol{q}\times 1}{C_t\,\nabla}$$

$$C_t = \boldsymbol{B}^* C_y$$

Generalized likelihood ratio test

reject H_0 if $T_{\boldsymbol{q}} > \chi^2_{\alpha}(\boldsymbol{q},0)$

$$H_0 : \underline{T}_{\boldsymbol{q}} \sim \chi^2(\boldsymbol{q},0) \;;\quad H_{\boldsymbol{A}} : \underline{T}_{\boldsymbol{q}} \sim \chi^2(\boldsymbol{q},\lambda)$$

$$\lambda = \nabla^* C_y^* Q_y^{-1} Q_{\hat{\boldsymbol{e}}} Q_y^{-1} C_y \nabla = \nabla^* C_t^* Q_t^{-1} C_t \nabla = \nabla^* Q_{\hat{\nabla}}^{-1} \nabla$$

Internal reliability

$$\lambda_0 = \lambda(\alpha,\boldsymbol{q},\gamma = \gamma_0)$$

$$\underset{m\times 1}{\nabla y} = \underset{m\times \boldsymbol{q}}{C_y}\,\underset{\boldsymbol{q}\times 1}{\nabla} \quad \text{and} \quad \lambda_y = \nabla y^* Q_y^{-1} \nabla y$$

with:

for $\boldsymbol{q} = 1$: $$|\nabla| = \left(\frac{\lambda_0}{c_y^* Q_y^{-1} Q_{\hat{\boldsymbol{e}}} Q_y^{-1} c_y}\right)^{1/2} = \left(\frac{\lambda_0}{c_t^* Q_t^{-1} c_t}\right)^{1/2}$$

for $1 < \boldsymbol{q} \le m-n$: $$\underset{\boldsymbol{q}\times 1}{\nabla} = \left(\frac{\lambda_0}{\boldsymbol{d}^* C_y^* Q_y^{-1} Q_{\hat{\boldsymbol{e}}} Q_y^{-1} C_y \boldsymbol{d}}\right)^{1/2} \underset{\boldsymbol{q}\times 1}{\boldsymbol{d}} \;,\; \boldsymbol{d} = \text{unit vector}$$

Table 4.8: Overview of the internal reliability.

4.6 External reliability

In the previous section internal reliability was defined as the model error that can be detected with the generalized likelihood ratio test with a probability $\gamma = \gamma_0$. It is described by the $m \times 1$ vector:

$$\nabla y = C_y \nabla = E\{\underline{y} \mid H_A\} - E\{\underline{y} \mid H_0\}$$

with the $q \times 1$ vector ∇ satisfying:

$$\lambda_0 = \nabla^* C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y \nabla.$$

In this section we consider the external reliability (uitwendige betrouwbaarheid). External reliability is defined as the influence of the model error ∇y on the final results of a geodetic computation or adjustment. The importance of external reliability stems from the fact that the final results of a geodetic computation are usually not the adjusted observations, but instead derived quantities such as coordinates. It is therefore of importance to know how the final results are influenced by possibly non-detected model errors ∇y. Let $\underline{\hat{x}}$ be the least-squares estimator of x under H_0. The following three cases will be considered in this section:

(i) The influence of ∇y on $\underline{\hat{x}}$

(ii) The influence of ∇y on a part of $\underline{\hat{x}}$, namely $\underline{\hat{x}}_1$

(iii) The influence of ∇y on a linear function of $\underline{\hat{x}}$ namely, $\underset{1\times 1}{\underline{\hat{\theta}}} = \underset{1\times n}{a^*}\ \underset{n\times 1}{\underline{\hat{x}}}$.

ad (i):

The least-squares estimator of x under H_0 is given by:

$$\underline{\hat{x}} = (A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} \underline{y}.$$

From this follows that:

$$E\{\underline{\hat{x}} \mid H_A\} = (A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} E\{\underline{y} \mid H_A\}.$$

Substitution of $E\{\underline{y} \mid H_A\} = Ax + C_y \nabla$ and $E\{\underline{\hat{x}} \mid H_0\} = x$ gives:

$$E\{\underline{\hat{x}} \mid H_A\} = E\{\underline{\hat{x}} \mid H_0\} + (A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1} C_y \nabla. \tag{161}$$

If we use the abbreviations $\nabla y = C_y \nabla = E\{\underline{y} \mid H_A\} - E\{\underline{y} \mid H_0\}$ and $\nabla \hat{x} = E\{\underline{\hat{x}} \mid H_A\} - E\{\underline{\hat{x}} \mid H_0\}$, we may write (161) as:

$$\underset{n\times 1}{\nabla \hat{x}} = \underset{n\times m}{(A^* Q_y^{-1} A)^{-1} A^* Q_y^{-1}} \underset{m\times 1}{\nabla y}. \tag{162}$$

This vector describes the influence of the model error ∇y on $\underline{\hat{x}}$. From (162) follows that $A\nabla\hat{x} = P_A\nabla y$. Therefore:

$$\begin{aligned}\nabla y &= \boldsymbol{P_A}\nabla y + \boldsymbol{P_A^{\perp}}\nabla y \\ &= \boldsymbol{A}\nabla\hat{x} + \boldsymbol{P_A^{\perp}}\nabla y.\end{aligned} \tag{163}$$

This orthogonal decomposition of the model error ∇y into $R(A)$ and $R(A)^{\perp}$ is shown in Figure 4.17.

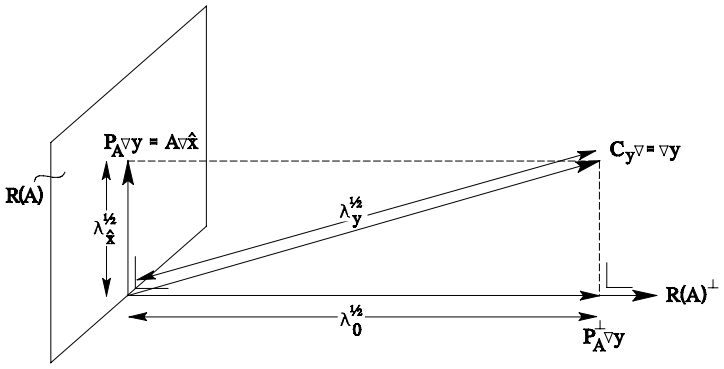


Figure 4.17: $\nabla y = A\nabla\hat{x} + P_A^{\perp}\nabla y$.

If we consider $\nabla\hat{x}$ of (162) as the possibly non-detected "bias" in $\underline{\hat{x}}$ and $Q_{\hat{x}}$ as a description of the "noise" in $\underline{\hat{x}}$, we may define a scalar squared *bias-to-noise* ratio for $\underline{\hat{x}}$ as:

$$\boxed{\lambda_{\hat{x}} = \nabla\hat{x}^{\,*}\boldsymbol{Q}_{\hat{x}}^{-1}\nabla\hat{x}} \,. \tag{164}$$

A large value of $\lambda_{\hat{x}}$ indicates that the influence of the model error ∇y on $\underline{\hat{x}}$ is significant, and a small value of $\lambda_{\hat{x}}$ indicates that this influence is insignificant. Since $Q_{\hat{x}}^{-1} = A^{\,*}Q_y^{-1}A$, it follows from (164) that:

$$\lambda_{\hat{x}} = \nabla\hat{x}^{\,*}\boldsymbol{A}^{\,*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}\nabla\hat{x} = (\boldsymbol{P_A}\nabla y)^{*}\boldsymbol{Q}_y^{-1}(\boldsymbol{P_A}\nabla y) = \|\boldsymbol{P_A}\nabla y\|^2. \tag{165}$$

This is also shown in Figure 4.17. Using the Pythagoras' theorem we may now relate $\lambda_{\hat{x}}$ to λ_0. Application of the Pythagoras' theorem to (163) gives:

$$\|\nabla y\|^2 = \|\boldsymbol{P_A}\nabla y\|^2 + \|\boldsymbol{P_A^{\perp}}\nabla y\|^2. \tag{166}$$

Since $\lambda_0 = \|P_A^{\perp}\nabla y\|^2$, $\lambda_{\hat{x}} = \|P_A\nabla y\|^2$ (see (165)), and $\lambda_y = \|\nabla y\|^2$ (see (157)), it follows from (166) that (see Figure 4.17):

$$\boxed{\lambda_{\hat{x}} = \lambda_y - \lambda_0} \,. \tag{167}$$

With (164) and (167) we have two ways of computing $\lambda_{\hat{x}}$: either via $\nabla\hat{x}$ as in (164) or via λ_y as in (167). Since the computation of λ_y is rather straightforward (especially if the variance matrix Q_y is diagonal), one usually uses (167) for computing $\lambda_{\hat{x}}$. The scalar $\lambda_{\hat{x}}$ may be used for

constructing an upperbound of an individual element of $\nabla\hat{x}$. Let us assume that we are interested in the ith element $\nabla\hat{x}_i$ of $\nabla\hat{x}$. Then:

$$\underset{1\times 1}{\nabla\hat{x}_i} = \underset{1\times n}{c_i^*}\,\underset{n\times 1}{\nabla\hat{x}}, \text{ with } \underset{n\times 1}{c_i} = (0,\cdots,\underset{\uparrow}{1},0,\cdots,0)^*. \quad (168)$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i\text{th}$$

Substitution of (162) into (168) gives:

$$\begin{aligned}\nabla\hat{x}_i &= c_i^*(A^*Q_y^{-1}A)^{-1}A^*Q_y^{-1}\nabla y\\ &= c_i^*(A^*Q_y^{-1}A)^{-1/2}\cdot(A^*Q_y^{-1}A)^{-1/2}A^*Q_y^{-1}\nabla y.\end{aligned}$$

This is an inner product which can be written with the help of the cosine rule as:

$$\nabla\hat{x}_i = [c_i^*(A^*Q_y^{-1}A)^{-1}c_i]^{1/2}\cdot[\nabla y^*Q_y^{-1}A(A^*Q_y^{-1}A)^{-1}A^*Q_y^{-1}\nabla y]^{1/2}\cdot\cos\theta_i. \quad (169)$$

In this expression we recognize $[c_i^*(A^*Q_y^{-1}A)^{-1}c_i]^{1/2}$ as $\sigma_{\hat{x}_i}$ and $[\nabla y^*Q_y^{-1}A(A^*Q_y^{-1}A)^{-1}A^*Q_y^{-1}\nabla y]^{1/2}$ as $\|P_A\nabla y\| = \lambda_{\hat{x}}^{1/2}$. Since $|\cos\theta_i| \le 1$, the upperbound follows therefore from (169):

$$\boxed{\left|\frac{\nabla\hat{x}_i}{\sigma_{\hat{x}_i}}\right| \le \lambda_{\hat{x}}^{1/2}}. \quad (170)$$

In the previous section for the case $1 < q \le m\text{-}n$ the expression for ∇ of (150) was substituted into the expression for λ_y. Similarly we can substitute ∇ into the expression for $\lambda_{\hat{x}}$. Since:

$$\begin{aligned}\lambda_{\hat{x}} &= \nabla\hat{x}^*Q_{\hat{x}}^{-1}\nabla\hat{x}\\ &= \nabla^*C_y^*Q_y^{-1}A(A^*Q_y^{-1}A)^{-1}A^*Q_y^{-1}C_y\nabla\\ &= \nabla^*C_y^*Q_y^{-1}Q_{\hat{y}}Q_y^{-1}C_y\nabla\end{aligned}$$

substitution of (150) gives:

$$\lambda_{\hat{x}} = \frac{d^*C_y^*Q_y^{-1}Q_{\hat{y}}Q_y^{-1}C_yd}{d^*C_y^*Q_y^{-1}Q_{\hat{e}}Q_y^{-1}C_yd}\,\lambda_0. \quad (171)$$

Substitution of $Q_{\hat{y}} = Q_y - Q_{\hat{e}}$ gives:

$$\lambda_{\hat{x}} = \left(\frac{d^*C_y^*Q_y^{-1}C_yd}{d^*C_y^*Q_y^{-1}Q_{\hat{e}}Q_y^{-1}C_yd} - 1\right)\lambda_0$$

which shows once again with (153) that (167) holds. In case of data snooping with a diagonal variance matrix Q_y, formula (171) simplifies to:

$$\lambda_{\hat{x}} = (\sigma_{y_i}^2/\sigma_{\hat{y}_i}^2 - 1)^{-1}\,\lambda_0. \quad (172)$$

This can be written in terms of the local redundancy number r_i as:

$$\lambda_{\hat{x}} = [(1-r_i)/r_i]\ \lambda_0.$$

ad (ii):

Let us partition $\underline{\hat{x}}$ as $\underline{\hat{x}} = (\underline{\hat{x}}_1^*, \underline{\hat{x}}_2^*)^*$. The partitioned system of normal equations reads then:

$$\begin{pmatrix} A_1^* Q_y^{-1} A_1 & A_1^* Q_y^{-1} A_2 \\ A_2^* Q_y^{-1} A_1 & A_2^* Q_y^{-1} A_2 \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \end{pmatrix} = \begin{pmatrix} A_1^* Q_y^{-1} \underline{y} \\ A_2^* Q_y^{-1} \underline{y} \end{pmatrix}. \tag{173}$$

This system corresponds to the partitioned linear model:

$$\underset{m\times 1}{H_0 : E\{\underline{y}\}} = \underset{m\times n_1}{(A_1} \vdots \underset{m\times n_2}{A_2)} \underset{(n_1+n_2)\times 1}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}.$$

In order to find the solution for $\underline{\hat{x}}_1$, we premultiply (173) with the square and regular matrix:

$$\begin{pmatrix} I & -A_1^* Q_y^{-1} A_2 (A_2^* Q_y^{-1} A_2)^{-1} \\ 0 & I \end{pmatrix}.$$

This gives:

$$\begin{pmatrix} A_1^* Q_y^{-1} [I - A_2 (A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1}] A_1 & 0 \\ A_2^* Q_y^{-1} A_1 & A_2^* Q_y^{-1} A_2 \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \end{pmatrix} = \begin{pmatrix} A_1^* Q_y^{-1} [I - A_2 (A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1}] \underline{y} \\ A_2^* Q_y^{-1} \underline{y} \end{pmatrix}. \tag{174}$$

In this expression we recognize the orthogonal projector:

$$P_{A_2}^{\perp} = I - A_2 (A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1}. \tag{175}$$

Using the abbreviation:

$$\bar{A}_1 = P_{A_2}^{\perp} A_1 \tag{176}$$

and noting that:

$$Q_y^{-1} P_{A_2}^{\perp} = P_{A_2}^{\perp *} Q_y^{-1} = P_{A_2}^{\perp *} Q_y^{-1} P_{A_2}^{\perp}$$

we may write (174) also as:

$$\begin{pmatrix} \bar{A}_1^* Q_y^{-1} \bar{A}_1 & 0 \\ A_2^* Q_y^{-1} A_1 & A_2^* Q_y^{-1} A_2 \end{pmatrix} \begin{pmatrix} \underline{\hat{x}}_1 \\ \underline{\hat{x}}_2 \end{pmatrix} = \begin{pmatrix} \bar{A}_1^* Q_y^{-1} \underline{y} \\ A_2^* Q_y^{-1} \underline{y} \end{pmatrix}.$$

From this result follows that we may write the least-squares estimator of x_1 under H_0 as:

$$\underline{\hat{x}}_1 = (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1} \bar{A}_1^* Q_y^{-1} \underline{y}. \tag{177}$$

From this follows that:

$$E\{\underline{\hat{x}}_1 | H_A\} = (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1} \bar{A}_1^* Q_y^{-1} E\{\underline{y} | H_A\}.$$

Substitution of $E\{\underline{y}|H_A\} = A_1 x_1 + A_2 x_2 + C_y \nabla$ gives:

$$E\{\underline{\hat{x}}_1 | H_A\} = (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1} \bar{A}_1^* Q_y^{-1} (A_1 x_1 + A_2 x_2 + C_y \nabla). \tag{178}$$

Since:

$$\bar{A}_1^* Q_y^{-1} A_1 = \bar{A}_1^* Q_y^{-1} P_{A_2}^{\perp} A_1 = \bar{A}_1^* Q_y^{-1} \bar{A}_1$$

and

$$\bar{A}_1^* Q_y^{-1} A_2 = A_1^* P_{A_2}^{\perp *} Q_y^{-1} A_2 = A_1^* Q_y^{-1} P_{A_2}^{\perp} A_2 = 0$$

equation (178) simplifies to:

$$E\{\underline{\hat{x}}_1 | H_A\} = x_1 + (\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1} \bar{A}_1^* Q_y^{-1} C_y \nabla. \tag{179}$$

If we use the abbreviations $\nabla y = C_y \nabla$ and $\nabla \hat{x}_1 = E\{\underline{\hat{x}}_1 | H_A\} - E\{\underline{\hat{x}}_1 | H_0\}$ and $x_1 = E\{\underline{\hat{x}}_1 | H_0\}$, we may write (179) as:

$$\underset{n_1 \times 1}{\nabla \hat{x}_1} = \underset{n_1 \times m}{(\bar{A}_1^* Q_y^{-1} \bar{A}_1)^{-1} \bar{A}_1^* Q_y^{-1}} \underset{m \times 1}{\nabla y}. \tag{180}$$

This vector describes the influence of the model error ∇y on $\underline{\hat{x}}_1$. Compare this result with (162). From (180) follows that $\bar{A}_1 \nabla \hat{x}_1 = P_{\bar{A}_1} \nabla y$. What is the relation between $\bar{A}_1 \nabla \hat{x}_1$ and $A \nabla \hat{x} = P_A \nabla y$? Since $\bar{A}_1 = P_{A_2}^{\perp} A$ it follows that:

$$R(\bar{A}_1 : A_2) = R(A_1 : A_2).$$

Therefore:

$$P_A = P_{(A_1 : A_2)} = P_{(\bar{A}_1 : A_2)}. \tag{181}$$

We also know that:

$$\bar{A}_1^* Q_y^{-1} A_2 = A_1^* Q_y^{-1} P_{A_2}^{\perp} A_2 = 0.$$

This implies that $R(\bar{A}_1) \perp R(A_2)$, and therefore that:

(182) $$\boldsymbol{P}_{(\bar{\boldsymbol{A}}_1 : \boldsymbol{A}_2)} = \boldsymbol{P}_{\bar{\boldsymbol{A}}_1} + \boldsymbol{P}_{\boldsymbol{A}_2}.$$

Hence, it follows from (181) and (182) that:

(183) $$\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\bar{\boldsymbol{A}}_1} + \boldsymbol{P}_{\boldsymbol{A}_2}.$$

But this implies that $P_A \nabla y = P_{\bar{A}_1} \nabla y + P_{A_2} \nabla y$ or that:

(184) $$\boldsymbol{A} \nabla \hat{x} = \bar{\boldsymbol{A}}_1 \nabla \hat{x}_1 + \boldsymbol{P}_{\boldsymbol{A}_2} \nabla y.$$

This orthogonal decomposition of $A \nabla \hat{x} = P_A \nabla y$ into $R(\bar{A}_1)$ and $R(A_2)$ is shown in Figure 4.18. Compare this with Figure 4.17.

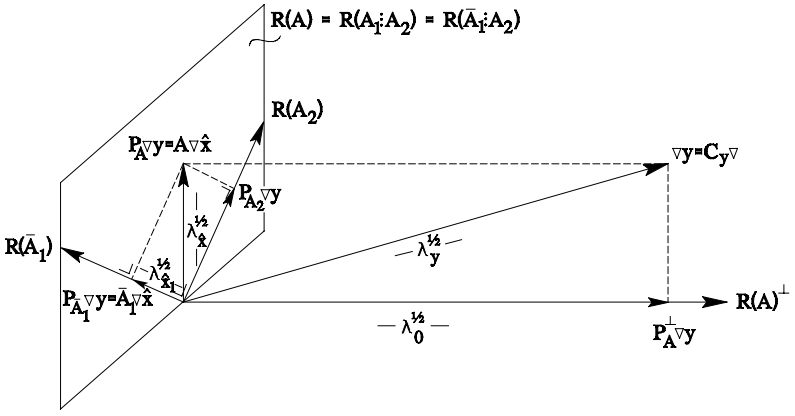


Figure 4.18: $P_A \nabla y = \bar{A}_1 \nabla \hat{x}_1 + P_{A_2} \nabla y$.

If we consider $\nabla \hat{x}_1$ of (180) as the possibly non-detected "bias" in $\underline{\hat{x}}_1$ and $Q_{\hat{x}_1}$ as a description of the "noise" in $\underline{\hat{x}}_1$, we may define analogous to (164) the squared "bias-to-noise" ratio for $\underline{\hat{x}}_1$ as:

(185) $$\boxed{\lambda_{\hat{x}_1} = \nabla \hat{x}_1^* \boldsymbol{Q}_{\hat{x}_1}^{-1} \nabla \hat{x}_1}.$$

Since $Q_{\hat{x}_1}^{-1} = \bar{A}_1^* Q_y^{-1} \bar{A}_1$ (see (177)), it follows that:

(186) $$\lambda_{\hat{x}_1} = \nabla \hat{x}_1^* \bar{\boldsymbol{A}}_1^* \boldsymbol{Q}_y^{-1} \bar{\boldsymbol{A}}_1 \nabla \hat{x}_1 = (\boldsymbol{P}_{\bar{\boldsymbol{A}}_1} \nabla y)^* \boldsymbol{Q}_y^{-1} (\boldsymbol{P}_{\bar{\boldsymbol{A}}_1} \nabla y) = \|\boldsymbol{P}_{\bar{\boldsymbol{A}}_1} \nabla y\|^2.$$

This is shown in Figure 4.18. Using the Pythagoras' theorem we may now relate $\lambda_{\hat{x}_1}$ to $\lambda_{\hat{x}}$. Application to (184) gives:

(187) $$\|\boldsymbol{A} \nabla \hat{x}\|^2 = \|\bar{\boldsymbol{A}}_1 \nabla \hat{x}_1\|^2 + \|\boldsymbol{P}_{\boldsymbol{A}_2} \nabla y\|^2.$$

Since $\|A \nabla \hat{x}\|^2 = \|P_A \nabla y\|^2 = \lambda_{\hat{x}}$ and $\|\bar{A}_1 \nabla \hat{x}_1\|^2 = \|P_{\bar{A}_1} \nabla y\|^2 = \lambda_{\hat{x}_1}$, it follows from (187) that (see Figure 4.18):

$$\lambda_{\hat{x}_1} = \lambda_{\hat{x}} - \|\boldsymbol{P}_{\boldsymbol{A}_2} \nabla y\|^2$$

or that:

(188) $$\boxed{\lambda_{\hat{x}_1} = \lambda_{\hat{x}} - \nabla y^{*} Q_y^{-1} A_2 (A_2^{*} Q_y^{-1} A_2)^{-1} A_2^{*} Q_y^{-1} \nabla y}\ .$$

Substitution of (167) into (188) gives:

(189) $$\boxed{\lambda_{\hat{x}_1} = \lambda_y - \lambda_0 - \nabla y^{*} Q_y^{-1} A_2 (A_2^{*} Q_y^{-1} A_2)^{-1} A_2^{*} Q_y^{-1} \nabla y}\ .$$

Formula (170) gives an upperbound for the "bias-to-noise" ratio of an individual element of $\underline{\hat{x}}$. In a completely analogous way one can derive the following upperbound for the "bias-to-noise" ratio of an individual element $\underline{\hat{x}}_{1_i}$, of $\underline{\hat{x}}_1$:

(190) $$\boxed{\left|\frac{\nabla \hat{x}_{1_i}}{\sigma_{\hat{x}_{1_i}}}\right| \leq \lambda_{\hat{x}_1}^{1/2}}\ .$$

Since $\lambda_{\hat{x}_1} \leq \lambda_{\hat{x}}$, the bound of (190) is sharper than the bound of (170).

ad (iii):

Now consider an arbitrary linear function of $\underline{\hat{x}}$:

(191) $$\underset{1\times 1}{\underline{\hat{\theta}}} = \underset{1\times n}{a^{*}}\ \underset{n\times 1}{\underline{\hat{x}}}\ .$$

Then:

(192) $$E\{\underline{\hat{\theta}}\,|\,H_A\} = a^{*} E\{\underline{\hat{x}}\,|\,H_A\} = a^{*}(E\{\underline{\hat{x}}\,|\,H_0\} + \nabla \hat{x}) = E\{\underline{\hat{\theta}}\,|\,H_0\} + a^{*} \nabla \hat{x}.$$

If we use the abbreviation $\nabla \hat{\theta} = E\{\underline{\hat{\theta}}\,|\,H_A\} - E\{\underline{\hat{\theta}}\,|\,H_0\}$, we may write (192) as:

(193) $$\boxed{\underset{1\times 1}{\nabla \hat{\theta}} = \underset{1\times n}{a^{*}}\ \underset{n\times 1}{\nabla \hat{x}}}\ .$$

This shows how an arbitrary linear function of $\underline{\hat{x}}$ is influenced by model errors. If we write (193) as $\nabla \hat{\theta} = a^{*} Q_{\hat{x}}^{1/2} Q_{\hat{x}}^{-1/2} \nabla \hat{x}$, application of the cosine rule gives:

(194) $$\nabla \hat{\theta} = (a^{*} Q_{\hat{x}} a)^{1/2} \cdot (\nabla \hat{x}^{*} Q_{\hat{x}}^{-1} \nabla \hat{x})^{1/2} \cdot \cos\varphi\ .$$

In this expression we recognize $\sigma_{\hat{\theta}} = (a^{*} Q_{\hat{x}} a)^{1/2}$ and $\lambda_{\hat{x}}^{1/2} = (\nabla \hat{x}^{*} Q_{\hat{x}}^{-1} \nabla \hat{x})^{1/2}$. The upperbound follows therefore from (194):

(195) $$\boxed{\left|\frac{\nabla \hat{\theta}}{\sigma_{\hat{\theta}}}\right| \leq \lambda_{\hat{x}}^{1/2}}\ .$$

This result shows that $\lambda_{\hat{x}}^{1/2}$ gives an upperbound for the "bias-to-noise" ratio of every arbitrary function of $\underline{\hat{x}}$.

A summary of the results of this section is given in Table 4.9.

Influence on $\underline{\hat{x}}$

$$\nabla\hat{x} = (A^{*}Q_y^{-1}A)^{-1}A^{*}Q_y^{-1}\nabla y$$

$$\lambda_{\hat{x}} = \nabla\hat{x}^{*}Q_{\hat{x}}^{-1}\nabla\hat{x} = \lambda_y - \lambda_0$$

$$|\nabla\hat{x}_i/\sigma_{\hat{x}_i}| \leq \lambda_{\hat{x}}^{1/2}$$

Influence on $\underline{\hat{x}}_1$

$$\nabla\hat{x}_1 = (\bar{A}_1^{*}Q_y^{-1}\bar{A}_1)^{-1}\bar{A}_1^{*}Q_y^{-1}\nabla y$$

$$\lambda_{\hat{x}_1} = \nabla\hat{x}_1^{*}Q_{\hat{x}_1}^{-1}\nabla\hat{x}_1 = \lambda_y - \lambda_0 - \nabla^{*}C_y^{*}Q_y^{-1}A_2(A_2^{*}Q_y^{-1}A_2)^{-1}A_2^{*}Q_y^{-1}C_y\nabla$$

$$|\nabla\hat{x}_{1_i}/\sigma_{\hat{x}_{1_i}}| \leq \lambda_{\hat{x}_1}^{1/2}$$

Influence on $\underline{\hat{\theta}} = a^{*}\underline{\hat{x}}$

$$\nabla\hat{\theta} = a^{*}\nabla\hat{x}$$

$$|\nabla\hat{\theta}/\sigma_{\hat{\theta}}| \leq \lambda_{\hat{x}}^{1/2}$$

Datasnooping & Q_y = diagonal

$$|\nabla_i| = \sigma_{y_i}[\lambda_0/(1 - \sigma_{\hat{y}_i}^2/\sigma_{y_i}^2)]^{1/2}$$

$$\lambda_{y_i} = [1 - \sigma_{\hat{y}_i}^2/\sigma_{y_i}^2]^{-1}\lambda_0$$

$$\lambda_{\hat{x}} = [\sigma_{y_i}^2/\sigma_{\hat{y}_i}^2 - 1]^{-1}\lambda_0$$

Table 4.9: Overview of the external reliability.

4.7 Reliability: an example

In this section we will give an example of a linear model of observation equations:

$$H_0 : E\{\underline{y}\} = (\boldsymbol{A}_1 \vdots \boldsymbol{A}_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ; \quad D\{\underline{y}\} = Q_y \tag{196}$$

for which the variance matrix Q_y is assumed to be diagonal. This means that in case of datasnooping the following formulae of internal and external reliability may be applied:

$$\begin{cases} \text{(a)} \quad |\nabla_i| = \sigma_{y_i}[\lambda_0/(1-\sigma^2_{\hat{y}_i}/\sigma^2_{y_i})]^{1/2} \\ \text{(b)} \quad \lambda_{y_i} = |\nabla_i|^2/\sigma^2_{y_i} \\ \text{(c)} \quad \lambda_{\hat{x}} = \lambda_{y_i} - \lambda_0 = (\sigma^2_{y_i}/\sigma^2_{\hat{y}_i} - 1)^{-1}\lambda_0 \\ \text{(d)} \quad \lambda_{\hat{x}_1} = \lambda_{y_i} - \lambda_0 - c^*_{y_i} Q_y^{-1} \boldsymbol{A}_2 (\boldsymbol{A}_2^* Q_y^{-1} \boldsymbol{A}_2)^{-1} \boldsymbol{A}_2^* Q_y^{-1} c_{y_i} \ |\nabla_i|^2 \\ \text{(e)} \quad \lambda_{\hat{x}_2} = \lambda_{y_i} - \lambda_0 - c^*_{y_i} Q_y^{-1} \boldsymbol{A}_1 (\boldsymbol{A}_1^* Q_y^{-1} \boldsymbol{A}_1)^{-1} \boldsymbol{A}_1^* Q_y^{-1} c_{y_i} \ |\nabla_i|^2. \end{cases} \tag{197}$$

The model that will be considered is given as:

$$E\{\underset{m\times 1}{\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_m \end{pmatrix}}\} = \underset{m\times 2}{\begin{pmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_m \end{pmatrix}} \underset{2\times 1}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} ; \quad D\{\underline{y}\} = \sigma^2 I_m. \tag{198}$$

The observables are assumed to be normally distributed. Since the observation equations are of the form $E\{\underline{y}_i\} = x_1 + a_i x_2$, they describe the equation of a straight line with intercept x_1 and slope x_2. This is shown in Figure 4.19.

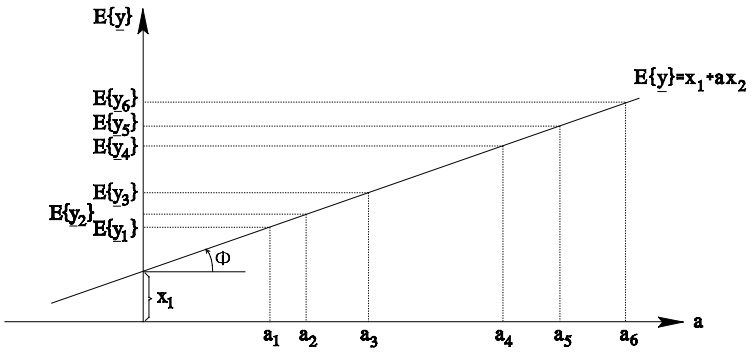


Figure 4.19: The line $E\{\underline{y}\} = x_1 + ax_2$ with intercept x_1 and slope $\tan\varphi = x_2$.

The least-squares estimates of x_1 and x_2 follow from the minimization problem:

$$\min_{x_1,x_2} \frac{1}{\sigma^2} \sum_{i=1}^{m} (y_i - x_1 - a_i x_2)^2 . \tag{199}$$

Since $|y_i - x_1 - a_i x_2|$ is the vertical distance from the point (a_i, y_i) to the straight line $E\{\underline{y}\} = x_1 + a x_2$, the least-squares estimates $\hat{x}_1$ and $\hat{x}_2$ follow from a minimization of the sum of the squares of these vertical distances (see Figure 4.20).

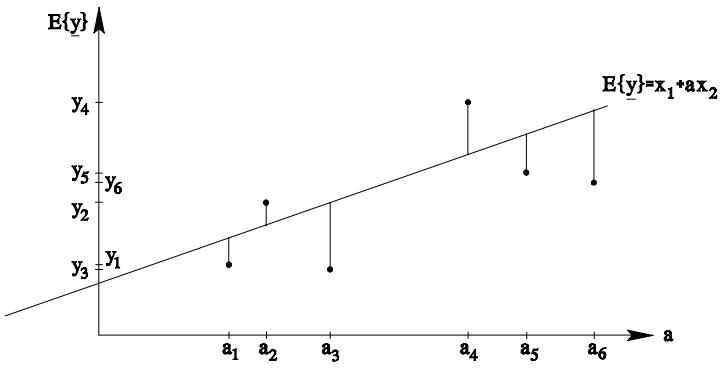


Figure 4.20: $\hat{x}_1$ and $\hat{x}_2$ follow from $\min_{x_1,x_2} \frac{1}{\sigma^2} \sum_{i=1}^{m} (y_i - x_1 - a_i x_2)^2$.

Let us first derive the minimal detectable bias $|\nabla_i|$ of the ith observable. According to (197a) one can compute $|\nabla_i|$ from λ_0, $\sigma^2_{y_i}$ and $\sigma^2_{\hat{y}_i}$. Since λ_0 is fixed and $\sigma^2_{y_i} = \sigma^2$, we only need to compute:

$$\sigma^2_{\hat{y}_i} = c^*_{y_i} Q_{\hat{y}} c_{y_i} = c^*_{y_i} A (A^* Q_y^{-1} A)^{-1} A^* c_{y_i} . \tag{200}$$

With $Q_y = \sigma^2 I_m$ and

$$\underset{m\times 2}{A} = \begin{pmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_m \end{pmatrix} \tag{201}$$

it follows that:

$$\underset{2\times 2}{(A^* Q_y^{-1} A)} = \frac{1}{\sigma^2} \begin{pmatrix} m & \sum_{j=1}^{m} a_j \\ \sum_{j=1}^{m} a_j & \sum_{j=1}^{m} a_j^2 \end{pmatrix} . \tag{202}$$

The inverse of this matrix reads therefore:

$$(203)\qquad (\underset{2\times 2}{\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}})^{-1} = \sigma^2[m\sum_{j=1}^{m} a_j^2 - (\sum_{j=1}^{m} a_j)^2]^{-1}\begin{pmatrix} \sum_{j=1}^{m} a_j^2 & -\sum_{j=1}^{m} a_j \\ -\sum_{j=1}^{m} a_j & m \end{pmatrix}.$$

In order to simplify expression (203) somewhat we define:

$$(204)\qquad \begin{cases} a_c \triangleq \dfrac{1}{m}\sum_{j=1}^{m} a_j & \text{(average value of the } a_j\text{'s)} \\ \bar{a}_j \triangleq a_j - a_c & \text{(centred with respect to } a_c\text{)}. \end{cases}$$

Then:

$$(205)\qquad \begin{aligned} \sum_{j=1}^{m} \bar{a}_j^2 &= \sum_{j=1}^{m} (a_j^2 - 2a_j a_c + a_c^2) \\ &= \sum_{j=1}^{m} a_j^2 - m a_c^2 \\ &= \frac{1}{m}[m\sum_{j=1}^{m} a_j^2 - (\sum_{j=1}^{m} a_j)^2]. \end{aligned}$$

We may write (203) therefore also as:

$$(206)\qquad (\underset{2\times 2}{\boldsymbol{A}^{*}\boldsymbol{Q}_y^{-1}\boldsymbol{A}})^{-1} = \frac{\sigma^2}{\sum_{j=1}^{m} \bar{a}_j^2}\begin{pmatrix} \frac{1}{m}\sum_{j=1}^{m} a_j^2 & -a_c \\ -a_c & 1 \end{pmatrix}.$$

From the structure of the variance matrix of (206) three conclusions can be drawn:

1. The least-squares estimators $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$ are uncorrelated if and only if $a_c = 0$, that is, if the coordinates a_i, $i = 1,\ldots,m$ are distributed symmetrically about $a = 0$.

2. The covariance between $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$ is negative if and only if a_c is positive, that is, if the cluster of points (a_i,y_i) is situated in the first or fourth quadrant. This means that if a_c is positive, an increase in x_1 implies a decrease in x_2 for an optimal fit (see Figure 4.21).

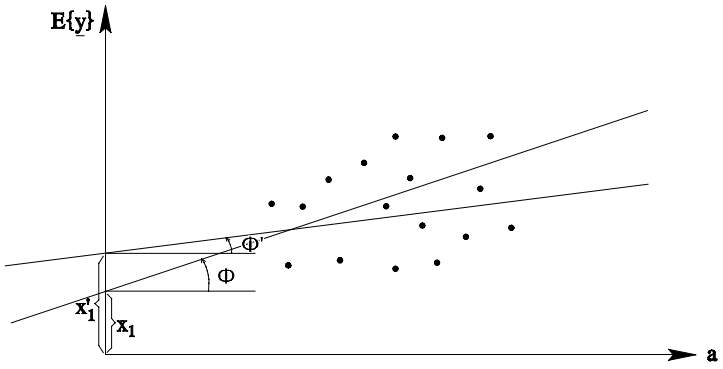


Figure 4.21: If $x_1' > x_1$ then $x_2' = \tan\varphi' < x_2 = \tan\varphi$.

3. The closer the coordinates a_j, $j = 1,...,m$, are to a_c, the larger the variances of $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$ get. In the extreme case that $a_j = a_c \ \forall \ j = 1,...,m$, the two columns of matrix A of (201) are linearly dependent and the variances of $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$ are infinite. Thus the closer the coordinates a_j, $j = 1,...,m$, are to a_c, the more difficult it becomes to estimate x_1 and x_2 (see Figure 4.22).

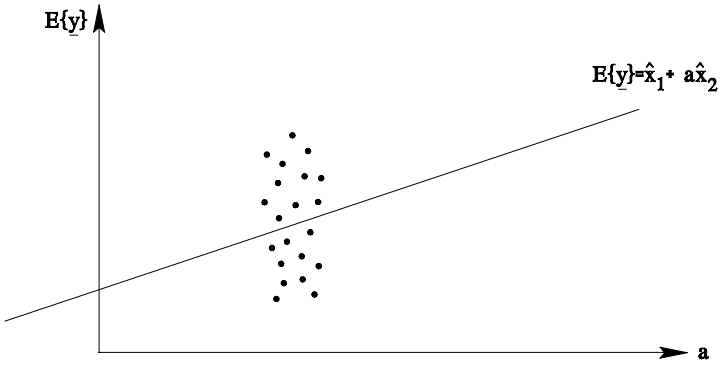


Figure 4.22: The line $E\{\underline{y}\} = x_1 + ax_2$ is poorly determinable.

In case of datasnooping the $m \times 1$ vector c_y takes the form:

$$c_{y_i} = (0,\cdots,\underset{\substack{\uparrow \\ i\text{th}}}{1},0,\cdots)^*.$$

With (201) this gives:

(207)
$$c_{y_i}^* A = (1 \ a_i).$$

Substitution of (206) and (207) into (200) gives:

(208) $$\sigma^2_{\hat{y}_i} = \frac{\frac{\sigma^2}{m}(\sum_{j=1}^{m} a_j^2 - 2ma_i a_c + ma_i^2)}{\sum_{j=1}^{m} \bar{a}_j^2}.$$

With (204) and (205) this can also be written as:

(209) $$\sigma^2_{\hat{y}_i} = \sigma^2\left(\frac{1}{m} + \frac{\bar{a}_i^2}{\sum_{j=1}^{m} \bar{a}_j^2}\right).$$

Substitution of (209) into (197a) gives with $\sigma^2_{y_i} = \sigma^2$ for the minimal detectable bias of the *i*th-observable:

(210) $$\boxed{|\nabla_i| = \sigma\left(\frac{\lambda_0}{1 - [\frac{1}{m} + \frac{\bar{a}_i^2}{\sum_{j=1}^{m} \bar{a}_j^2}]}\right)^{1/2}}.$$

Note that the rough approximation given in (147) of Section 4.5, corresponds for the present case to the approximation:

$$\frac{1}{m} \doteq \frac{\bar{a}_i^2}{\sum_{j=1}^{m} \bar{a}_j^2}.$$

It follows from (210) that $|\nabla_i|$ is smaller for points that have coordinates a_i closer to a_c. Hence, a blunder in the *i*th observable is better detectable if the corresponding point (a_i,y_i) lies near the centre of the cluster (a_j,y_i) $j = 1,...,m$, than when it would be near the left or right edges of the cluster. A similar effect can be seen for $\lambda_{\hat{x}}$. Substitution of (209) into (197c) gives namely:

(211) $$\boxed{\lambda_{\hat{x}} = \left(\frac{1}{[\frac{1}{m} + \frac{\bar{a}_i^2}{\sum_{j=1}^{m} \bar{a}_j^2}]} - 1\right)^{-1} \lambda_0}.$$

Let us now consider the "bias-to-noise" ratios of the individual estimators $\underline{\hat{x}}_1$ and $\underline{\hat{x}}_2$. First we will compute $\lambda_{\hat{x}_1}$. With $Q_y = \sigma^2 I_m$ and $A_2 = (a_1,\cdots,a_m)^*$ it follows that:

(212) $$c_{y_i}^* Q_y^{-1} A_2 (A_2^* Q_y^{-1} A_2)^{-1} A_2^* Q_y^{-1} c_{y_i} = \frac{1}{\sigma^2} \frac{a_i^2}{\sum_{j=1}^{m} a_j^2}.$$

Substitution of (210), (211) and (212) into (197d) gives:

(213)
$$\lambda_{\hat{x}_1} = \left(\frac{1 - a_i^2 / \sum_{j=1}^{m} a_j^2}{1 - \bar{a}_i^2 / \sum_{j=1}^{m} \bar{a}_j^2 - 1/m} - 1 \right) \lambda_0 \; .$$

This shows that $\lambda_{\hat{x}_1} = (\nabla \hat{x}_1 / \sigma_{\hat{x}_1})^2$ is small if a_i is large and/or $\bar{a}_i$ is small. Thus the effect of a possibly non-detected blunder in the ith observable on the intercept estimator $\hat{\underline{x}}_1$, is less significant for points with large coordinates a_i than for points with smaller coordinates a_i. And it is even less significant if also a_i is close to a_c. With $Q_y = \sigma^2 I_m$ and $A_1 = (1,\cdots,1)^*$ it follows that:

(214)
$$c_{y_i}^* Q_y^{-1} A_1 (A_1^* Q_y^{-1} A_1)^{-1} A_1^* Q_y^{-1} c_{y_i} = \frac{1}{\sigma^2} \frac{1}{m} .$$

Substitution of (210), (211) and (214) into (197e) gives:

(215)
$$\lambda_{\hat{x}_2} = \left(\frac{1}{[1 - \frac{m\bar{a}_i^2}{(m-1)\sum_{j=1}^{m} \bar{a}_j^2}]} - 1 \right) \lambda_0 \; .$$

This result shows that $\lambda_{\hat{x}_2} = 0$ if $\bar{a}_i = 0$. Hence the effect of a possibly non-detected blunder in the ith observable on the slope estimator $\hat{\underline{x}}_2$ is insignificant if a_i is close enough to a_c. This effect increases however the more a_i differs from a_c.

Appendix A

Some standard distributions

The normal distribution

Definition: An $n\times 1$ random vector $\underline{x}$ is said to be normally distributed if its probability density function, $p_{\underline{x}}(x)$, is given as:

(1) $$p_{\underline{x}}(x) = (2\pi)^{-\frac{n}{2}}|Q|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}(x-\mu)^{*}Q^{-1}(x-\mu)]$$

with Q an $n\times n$ positive definite matrix, and μ an $n\times 1$ vector. Note that a normal distribution is completely specified once Q and μ are given. The following notation will be used for an $n\times 1$ normally distributed vector $\underline{x}$:

(2) $$\underset{n\times 1}{\underline{x}} \sim \underset{n\times 1}{N}(\underset{n\times 1}{\mu}, \underset{n\times n}{Q}).$$

***Theorem*:** The expectation, $E\{\underline{x}\}$, and dispersion (or variancematrix), $D\{\underline{x}\}$, of $\underline{x} \sim N(\mu,Q)$ are:

(3) $$E\{\underline{x}\} = \mu \quad \text{and} \quad D\{\underline{x}\} = Q.$$

***Theorem*:** Let the expectation and dispersion of the random $n\times 1$ vector $\underline{x}$ be given as: $E\{\underline{x}\} = x$ and $D\{\underline{x}\} = Q_x$. Let the random $m\times 1$ vector $\underline{y}$ be defined by $\underset{m\times 1}{\underline{y}} = \underset{m\times n}{A}\,\underset{n\times 1}{\underline{x}} + \underset{m\times 1}{a}$. Then:

(4) $$E\{\underline{y}\} = Ax + a \quad \text{and} \quad D\{\underline{y}\} = AQ_xA^{*}.$$

***Theorem*:** If $\underline{x} \sim N(x,Q_x)$ and $\underline{y} = A\underline{x} + a$, then:

(5) $$\underline{y} \sim N(Ax+a,\ AQ_xA^{*}).$$

The non-central χ^2-distribution

Definition: A scalar random variable $\underline{x}$ is said to have a noncentral Chi-square distribution with n degrees of freedom and non-centrality parameter λ if its probability density function, $p_{\underline{x}}(x)$, is given as:

(6) $$p_{\underline{x}}(x) = \begin{cases} e^{-\frac{\lambda}{2}}\sum_{j=0}^{\infty}\dfrac{(\frac{\lambda}{2})^{j}\,x^{\frac{n}{2}+j-1}\exp[-\frac{x}{2}]}{j!\;2^{\frac{n}{2}+j}\Gamma(\frac{n}{2}+j)} & \text{for } 0 < x < \infty \\ 0 & \text{for } x \le 0 \end{cases}$$

with

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0.$$

The following notation will be used for a Chi-square random variable $\underline{x}$ with n degrees of freedom and non-centrality parameter λ:

(7) $$\underline{x} \sim \chi^2(n,\lambda).$$

We speak of a central Chi-square distribution if $\lambda = 0$.

Theorem: The expectation, $E\{\underline{x}\}$, and dispersion, $D\{\underline{x}\}$, of $\underline{x} \sim \chi^2(n,\lambda)$ are:

(8) $$E\{\underline{x}\} = n + \lambda \quad \text{and} \quad D\{\underline{x}\} = 2n + 4\lambda.$$

Theorem: If $\underset{n\times 1}{\underline{x}} \sim \underset{n\times 1}{N}(x,\underset{n\times n}{Q_x})$ and $\underline{y} = \underline{x}^* Q_x^{-1} \underline{x}$, then:

(9) $$\underline{y} \sim \chi^2(n,\lambda) \quad \text{with} \quad \lambda = x^* Q_x^{-1} x.$$

The non-central *F*-distribution

Definition: A scalar random variable $\underline{x}$ is said to have a non-central F-distribution with m and n degrees of freedom and non-centrality parameter λ if its probability density function, $p_{\underline{x}}(x)$, is given as:

(10) $$p_{\underline{x}}(x) = \begin{cases} e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \dfrac{(\frac{\lambda}{2})^j \, x^{\frac{m}{2}+j-1} \, m^{\frac{m}{2}+j} \, n^{\frac{n}{2}} \, \Gamma(\frac{m}{2}+\frac{n}{2}+j)}{j! \; \Gamma(\frac{m}{2}+j)\Gamma(\frac{n}{2})(n+mx)^{\frac{m}{2}+\frac{n}{2}+j}}, & \text{for } 0 < x < \infty \\ 0 & \text{for } x \le 0 \end{cases}.$$

The following notation will be used for an F-distribution with m and n degrees of freedom and non-centrality parameter λ:

(11) $$\underline{x} \sim F(m,n,\lambda).$$

We speak of a central F-distribution if $\lambda = 0$.

Theorem: If $\underset{m\times 1}{\underline{u}} \sim \underset{m\times 1}{N}(u,\underset{m\times m}{Q_u})$, $\underset{n\times 1}{\underline{v}} \sim \underset{n\times 1}{N}(0,\underset{n\times n}{Q_v})$ and $\underline{u}$ and $\underline{v}$ are uncorrelated, then $\underline{x} = (\underline{u}^* Q_u^{-1} \underline{u} \,/m) \,/\, (\underline{v}^* Q_v^{-1} \underline{v} \,/n)$ is distributed as:

(12) $$\underline{x} \sim F(m,n,\lambda) \quad \text{with} \quad \lambda = u^* Q_u^{-1} u.$$

Remark: The distribution of $\underline{z} = (\underline{u}^* Q_u^{-1} \underline{u} \,/m)$ is sometimes noted down as: $\underline{z} \sim F(m,\infty,\lambda)$.

Appendix B

Statistical tables

Normal distribution: computation of one-sided level of significance

k	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002

Table B.1: Standard normal distribution. N(0, 1); given is α, probability in right-hand tail, for critical values k, e.g. k=1.96 yields α= 0.0250.

calculation in Matlab:
alpha = (1-normcdf (critical value, mu, sigma))

Chi-square distribution: computation of critical value

q\α	0.500	0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	0.455	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	1.386	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	2.366	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	3.357	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	4.351	6.626	9.236	11.07	12.83	15.09	16.75	20.52
6	5.348	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	6.346	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	7.344	10.22	13.36	15.51	17.53	20.09	21.95	26.12
9	8.343	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	9.342	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.31
21	20.34	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	22.34	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	26.34	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	28.34	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	49.33	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	59.33	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	69.33	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	79.33	88.13	96.58	101.9	106.6	112.3	116.3	124.8
90	89.33	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	99.33	109.1	118.5	124.3	129.6	135.8	140.2	149.4

Table B.2: Chi-square distribution. χ^2 (q, 0); given is k, critical value, for α, probability in right-hand tail, and q, degrees of freedom, e.g. α=0.010 and q=10 yield k = 23.21; $k = \chi^2_\alpha(q,0)$ for test (102) in Section 4.5.

calculation in Matlab:
critical value = chi2inv (1-alpha, degrees of freedom)

Central *F*-distribution: computation of critical value

$\alpha = 0.10$

$q_2 \backslash q_1$	1	2	3	4	5	6	8	10	20	100
1	39.86	49.50	53.59	55.83	57.24	58.20	59.44	60.19	61.74	63.01
2	8.526	9.000	9.162	9.243	9.293	9.326	9.367	9.392	9.441	9.481
3	5.538	5.462	5.391	5.343	5.309	5.285	5.252	5.230	5.184	5.144
4	4.545	4.325	4.191	4.107	4.051	4.010	3.955	3.920	3.844	3.778
5	4.060	3.780	3.619	3.520	3.453	3.405	3.339	3.297	3.207	3.126
6	3.776	3.463	3.289	3.181	3.108	3.055	2.983	2.937	2.836	2.746
8	3.458	3.113	2.924	2.806	2.726	2.668	2.589	2.538	2.425	2.321
10	3.285	2.924	2.728	2.605	2.522	2.461	2.377	2.323	2.201	2.087
20	2.975	2.589	2.380	2.249	2.158	2.091	1.999	1.937	1.794	1.650
100	2.756	2.356	2.139	2.002	1.906	1.834	1.732	1.663	1.494	1.293
∞	2.706	2.303	2.084	1.945	1.847	1.774	1.670	1.599	1.421	1.185

$\alpha = 0.05$

$q_2 \backslash q_1$	1	2	3	4	5	6	8	10	20	100
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	241.9	248.0	253.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.40	19.45	19.49
3	10.13	9.552	9.277	9.117	9.013	8.941	8.845	8.786	8.660	8.554
4	7.709	6.944	6.591	6.388	6.256	6.163	6.041	5.964	5.803	5.664
5	6.608	5.786	5.409	5.192	5.050	4.950	4.818	4.735	4.558	4.405
6	5.987	5.143	4.757	4.534	4.387	4.284	4.147	4.060	3.874	3.712
8	5.318	4.459	4.066	3.838	3.687	3.581	3.438	3.347	3.150	2.975
10	4.965	4.103	3.708	3.478	3.326	3.217	3.072	2.978	2.774	2.588
20	4.351	3.493	3.098	2.866	2.711	2.599	2.447	2.348	2.124	1.907
100	3.936	3.087	2.696	2.463	2.305	2.191	2.032	1.927	1.676	1.392
∞	3.841	2.996	2.605	2.372	2.214	2.099	1.938	1.831	1.571	1.243

$\alpha = 0.01$

$q_2 \backslash q_1$	1	2	3	4	5	6	8	10	20	100
1	4052.	5000.	5403.	5625.	5764.	5859.	5981.	6056.	6209.	6334.
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.40	99.45	99.49
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.23	26.69	26.24
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.55	14.02	13.58
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	10.05	9.553	9.130
6	13.75	10.92	9.780	9.148	8.746	8.466	8.102	7.874	7.396	6.987
8	11.26	8.649	7.591	7.006	6.632	6.371	6.029	5.814	5.359	4.963
10	10.04	7.559	6.552	5.994	5.636	5.386	5.057	4.849	4.405	4.014
20	8.096	5.849	4.938	4.431	4.103	3.871	3.564	3.368	2.938	2.535
100	6.895	4.824	3.984	3.513	3.206	2.988	2.694	2.503	2.067	1.598
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.511	2.321	1.878	1.358

Table B.3: Central *F*-distribution. F (q_1, q_2, 0); given is k, critical value, for q_1 and q_2, degrees of freedom, for some values of α, probability in right-hand tail, e.g. α=0.01, q_1=10, q_2=∞ yield k = 2.321; k=F_α (q, ∞, 0) for test (106) in Section 4.5.

calculation in Matlab:
critical value = finv (1-alpha, degrees of freedom q1, q2)

Non-central *F*-distribution: computation of type II error probability

$\alpha = 0.10$ and $q_1=1$ | $\alpha = 0.10$ and $q_1=2$

$q_2\backslash\lambda$	1	2	6	10	15	21	1	2	6	10	15	21
1	0.85	0.82	0.70	0.62	0.54	0.47	0.88	0.86	0.79	0.74	0.69	0.64
2	0.82	0.74	0.51	0.35	0.22	0.12	0.86	0.81	0.67	0.55	0.43	0.31
3	0.80	0.70	0.41	0.23	0.11	0.04	0.84	0.78	0.58	0.42	0.27	0.16
4	0.78	0.68	0.35	0.17	0.07	0.02	0.83	0.76	0.52	0.34	0.19	0.09
5	0.78	0.66	0.32	0.15	0.05	0.01	0.82	0.75	0.48	0.29	0.14	0.06
6	0.77	0.65	0.30	0.13	0.04	0.01	0.82	0.74	0.45	0.26	0.12	0.04
8	0.76	0.64	0.28	0.11	0.03	0.01	0.81	0.72	0.42	0.22	0.09	0.03
10	0.76	0.63	0.26	0.10	0.03	0.00	0.81	0.71	0.39	0.19	0.07	0.02
20	0.75	0.61	0.24	0.08	0.02	0.00	0.80	0.69	0.35	0.15	0.05	0.01
100	0.74	0.59	0.22	0.07	0.01	0.00	0.79	0.67	0.31	0.12	0.03	0.01
∞	0.74	0.59	0.21	0.06	0.01	0.00	0.78	0.67	0.30	0.11	0.03	0.00

$\alpha = 0.10$ and $q_1=6$ | $\alpha = 0.10$ and $q_1=10$

$q_2\backslash\lambda$	1	2	6	10	15	21	1	2	6	10	15	21
1	0.89	0.88	0.86	0.84	0.81	0.78	0.90	0.89	0.87	0.86	0.84	0.82
2	0.88	0.87	0.81	0.76	0.69	0.63	0.89	0.88	0.85	0.81	0.77	0.72
3	0.88	0.86	0.77	0.69	0.59	0.49	0.89	0.87	0.82	0.77	0.70	0.63
4	0.87	0.85	0.74	0.63	0.51	0.38	0.88	0.87	0.80	0.73	0.64	0.55
5	0.87	0.84	0.71	0.58	0.44	0.31	0.88	0.86	0.78	0.70	0.59	0.48
6	0.87	0.83	0.69	0.55	0.39	0.25	0.88	0.86	0.76	0.67	0.55	0.42
8	0.86	0.82	0.65	0.49	0.32	0.18	0.88	0.85	0.74	0.62	0.48	0.34
10	0.86	0.81	0.63	0.45	0.28	0.14	0.87	0.84	0.72	0.59	0.43	0.28
20	0.85	0.79	0.56	0.36	0.18	0.07	0.87	0.83	0.66	0.49	0.31	0.16
100	0.84	0.77	0.49	0.27	0.11	0.03	0.85	0.80	0.58	0.37	0.18	0.07
∞	0.83	0.76	0.47	0.25	0.09	0.02	0.85	0.80	0.55	0.33	0.15	0.05

Table B.4: Non-central *F*-distribution. $F\ (q_1, q_2, \lambda)$; given is β, probability in left-hand tail of $F(q_1, q_2, \lambda)$, for λ, non-centrality parameter, and q_2, degrees of freedom, for α=0.10, probability in right-hand tail of $F(q_1, q_2, 0)$, for some values of q_1, degrees of freedom, e.g. q_1=1, q_2=∞ and α=0.10 yield, with λ=2, β=0.59 and hence γ=0.41, see also Table 4.7 in Section 4.5;

$$\alpha = \int_{F_\alpha(q,\infty,0)}^{\infty} p_{\underline{F}}(F|q,\infty,0)\mathrm{d}F.$$

and

$$\beta = 1 - \int_{F_\alpha(q,\infty,0)}^{\infty} p_{\underline{F}}(F|q,\infty,\lambda)\mathrm{d}F$$

for test (106).

calculation in Matlab:
critical value = finv (1-alpha, degrees of freedom q1, q2)
beta = ncfcdf (critical value, degrees of freedom q1, q2, lambda)

Non-central *F*-distribution: computation of type II error probability

$\alpha = 0.05$ and $q_1=1$

$q_2 \backslash \lambda$	1	2	6	10	15	21
1	0.93	0.91	0.85	0.80	0.76	0.72
2	0.90	0.86	0.71	0.58	0.46	0.34
3	0.89	0.83	0.60	0.43	0.27	0.15
4	0.88	0.80	0.54	0.34	0.18	0.08
5	0.87	0.79	0.49	0.28	0.13	0.05
6	0.86	0.78	0.46	0.25	0.11	0.03
8	0.86	0.76	0.42	0.21	0.08	0.02
10	0.85	0.75	0.40	0.19	0.06	0.02
20	0.84	0.73	0.36	0.15	0.04	0.01
100	0.83	0.71	0.32	0.12	0.03	0.00
∞	0.83	0.71	0.31	0.11	0.03	0.00

$\alpha = 0.05$ and $q_1=2$

$q_2 \backslash \lambda$	1	2	6	10	15	21
1	0.94	0.93	0.89	0.87	0.84	0.81
2	0.93	0.90	0.82	0.74	0.65	0.56
3	0.92	0.88	0.75	0.62	0.48	0.35
4	0.91	0.87	0.69	0.53	0.37	0.23
5	0.90	0.86	0.65	0.47	0.29	0.16
6	0.90	0.85	0.62	0.42	0.24	0.12
8	0.89	0.83	0.57	0.36	0.18	0.07
10	0.89	0.82	0.55	0.32	0.15	0.05
20	0.88	0.80	0.48	0.25	0.10	0.03
100	0.87	0.78	0.43	0.20	0.06	0.01
∞	0.87	0.77	0.42	0.18	0.06	0.01

$\alpha = 0.05$ and $q_1=6$

$q_2 \backslash \lambda$	1	2	6	10	15	21
1	0.95	0.94	0.93	0.92	0.90	0.89
2	0.94	0.93	0.90	0.87	0.84	0.80
3	0.94	0.93	0.88	0.83	0.76	0.69
4	0.94	0.92	0.85	0.78	0.69	0.58
5	0.93	0.91	0.83	0.74	0.63	0.50
6	0.93	0.91	0.81	0.71	0.58	0.43
8	0.93	0.90	0.78	0.65	0.49	0.33
10	0.93	0.90	0.76	0.61	0.44	0.27
20	0.92	0.88	0.70	0.50	0.30	0.14
100	0.91	0.86	0.62	0.39	0.18	0.06
∞	0.91	0.85	0.60	0.36	0.16	0.05

$\alpha = 0.05$ and $q_1=10$

$q_2 \backslash \lambda$	1	2	6	10	15	21
1	0.95	0.95	0.94	0.93	0.92	0.91
2	0.95	0.94	0.92	0.90	0.88	0.85
3	0.94	0.94	0.91	0.88	0.84	0.79
4	0.94	0.93	0.89	0.85	0.79	0.72
5	0.94	0.93	0.88	0.82	0.75	0.66
6	0.94	0.93	0.87	0.80	0.71	0.61
8	0.94	0.92	0.85	0.76	0.65	0.52
10	0.93	0.92	0.83	0.73	0.60	0.45
20	0.93	0.90	0.78	0.63	0.45	0.28
100	0.92	0.89	0.71	0.50	0.29	0.13
∞	0.92	0.88	0.68	0.46	0.24	0.09

Table B.5: Non-central *F*-distribution. $F\ (q_1, q_2, \lambda)$ given is β, probability in left-hand tail of $F(q_1, q_2, \lambda)$, for λ, non-centrality parameter, and q_2, degrees of freedom, for α=0.05, probability in right-hand tail of $F(q_1, q_2, 0)$, for some values of q_1, degrees of freedom.

calculation in Matlab:
critical value = finv (1-alpha, degrees of freedom q1, q2)
beta = ncfcdf (critical value, degrees of freedom q1, q2, lambda)

Non-central *F*-distribution: computation of type II error probability

$\alpha = 0.01$ and $q_1=1$

$q_2\backslash\lambda$	1	2	6	10	15	21
1	0.99	0.98	0.97	0.96	0.95	0.94
2	0.98	0.97	0.93	0.90	0.85	0.80
3	0.98	0.96	0.89	0.81	0.71	0.60
4	0.97	0.95	0.84	0.72	0.58	0.43
5	0.97	0.94	0.81	0.65	0.48	0.31
6	0.96	0.93	0.78	0.60	0.41	0.24
8	0.96	0.92	0.73	0.52	0.31	0.15
10	0.96	0.92	0.70	0.47	0.26	0.11
20	0.95	0.90	0.63	0.37	0.17	0.05
100	0.94	0.88	0.57	0.30	0.11	0.03
∞	0.94	0.88	0.55	0.28	0.10	0.02

$\alpha = 0.01$ and $q_1=2$

$q_2\backslash\lambda$	1	2	6	10	15	21
1	0.99	0.99	0.98	0.97	0.97	0.96
2	0.99	0.98	0.96	0.94	0.92	0.89
3	0.98	0.97	0.94	0.90	0.84	0.77
4	0.98	0.97	0.91	0.85	0.75	0.64
5	0.98	0.96	0.89	0.80	0.67	0.53
6	0.98	0.96	0.87	0.76	0.60	0.44
8	0.97	0.95	0.84	0.69	0.50	0.32
10	0.97	0.95	0.81	0.64	0.43	0.24
20	0.97	0.93	0.74	0.52	0.29	0.12
100	0.96	0.92	0.67	0.41	0.18	0.06
∞	0.96	0.92	0.65	0.39	0.16	0.05

$\alpha = 0.01$ and $q_1=6$

$q_2\backslash\lambda$	1	2	6	10	15	21
1	0.99	0.99	0.99	0.98	0.98	0.98
2	0.99	0.99	0.98	0.97	0.97	0.96
3	0.99	0.98	0.97	0.96	0.94	0.92
4	0.99	0.98	0.97	0.95	0.92	0.88
5	0.99	0.98	0.96	0.93	0.89	0.83
6	0.99	0.98	0.95	0.91	0.86	0.78
8	0.98	0.98	0.94	0.88	0.80	0.68
10	0.98	0.98	0.93	0.86	0.75	0.60
20	0.98	0.97	0.89	0.77	0.59	0.39
100	0.98	0.96	0.83	0.64	0.40	0.19
∞	0.98	0.95	0.81	0.59	0.34	0.14

$\alpha = 0.01$ and $q_1=10$

$q_2\backslash\lambda$	1	2	6	10	15	21
1	0.99	0.99	0.99	0.99	0.98	0.98
2	0.99	0.99	0.98	0.98	0.98	0.97
3	0.99	0.99	0.98	0.97	0.96	0.95
4	0.99	0.99	0.98	0.97	0.95	0.93
5	0.99	0.99	0.97	0.96	0.93	0.90
6	0.99	0.98	0.97	0.95	0.92	0.87
8	0.99	0.98	0.96	0.93	0.89	0.82
10	0.99	0.98	0.96	0.92	0.85	0.76
20	0.98	0.98	0.93	0.86	0.74	0.57
100	0.98	0.97	0.88	0.74	0.53	0.31
∞	0.98	0.97	0.86	0.69	0.46	0.23

Table B.6: Non-central *F*-distribution. $F\ (q_1, q_2, \lambda)$; given is β, probability in left-hand tail of $F\ (q_1, q_2, \lambda)$ for λ, non-centrality parameter, and q_2, degrees of freedom, for α=0.01, probability in right-hand tail of $F\ (q_1, q_2, 0)$, for some values of q_1, degrees of freedom.

calculation in Matlab:
critical value = finv (1-alpha, degrees of freedom q1, q2)
beta = ncfcdf (critical value, degrees of freedom q1, q2, lambda)

Appendix C

Detection, identification and adaptation

We have given the teststatistic for testing the null hypothesis H_0 against a particular alternative hypothesis H_A. In most practical applications however, it is usually not only one model error one is concerned about, but quite often many more than one. This implies that one needs a *testing procedure* for handling the various alternative hypotheses. In this appendix we will discuss a way of structuring such a testing procedure. It will consist of the following three steps: detection, identification and adaptation.

Detection

Since one usually first wants to know whether one can have any confidence in the assumed null hypothesis without the need to specify any particular alternative hypothesis, the first step consists of a check on the *overall* validity of H_0. This implies that one opposes the null hypothesis to the most relaxed alternative hypothesis possible (see Section 4.4). The most relaxed alternative hypothesis is the one that leaves the observables completely free. Hence, under this alternative hypothesis no restrictions at all are imposed on the observables. We therefore have the situation:

(1) $$H_o: E\{\underline{y}\} = \boldsymbol{A}x \text{ versus } H_{\boldsymbol{A}}: E\{\underline{y}\} \in \mathbb{R}^m.$$

It can be shown that in this case, the appropriate teststatistic reads:

(2) $$\underline{T}_{\boldsymbol{q}=m-n} = \underline{\hat{\boldsymbol{e}}}^* \boldsymbol{Q}_y^{-1} \underline{\hat{\boldsymbol{e}}}.$$

The appropriate teststatistic for testing the null hypothesis against the most relaxed alternative hypothesis, is thus equal to the weighted sum-of-squares of the least-squares residuals. The null hypothesis will then be rejected when:

(3) $$T_{\boldsymbol{q}=m-n} > \chi^2_{\alpha}(m-n,0)$$

with α, the chosen level of significance, and m-n, the redundancy.

The $\hat{\sigma}^2$ *test*: In the literature one often sees the above *overall model test* also formulated in a slightly different way. Let us use the factorization $D\{\underline{y}\}=Q_y=\sigma^2 Q$, where σ^2 is the variance factor of unit weight and Q is the corresponding cofactor matrix. It can be shown that:

$$\underline{\hat{\sigma}}^2 = \frac{\underline{\hat{\boldsymbol{e}}}^* Q^{-1} \underline{\hat{\boldsymbol{e}}}}{m-n}$$

is an *unbiased* estimator of σ^2 (see also (100) in Section 4.4). Thus $E\{\underline{\hat{\sigma}}^2\}=\sigma^2$. The test (3) can now also be formulated as:

$$\frac{\hat{\sigma}^2}{\sigma^2} > \frac{\chi^2_\alpha(m-n,0)}{m-n} = F_\alpha(m-n,\infty,0)$$

where $F(m-n,\infty,0)$ is the central F-distribution having m-n and ∞ degrees of freedom.

Identification

In the detection phase, one tests the overall validity of the null hypothesis. If this leads to a rejection of the null hypothesis, one has to search for possible model misspecifications. That is, one will have to try to identify the model error which caused the rejection of the null hypothesis. This implies that one will have to specify the type of likely model errors. This specification of possible alternative hypotheses is application dependent and is one of the more difficult tasks in hypothesis testing. It depends very much on ones experience.

The *1-dimensional case*: In case the model error can be represented by a scalar, the alternative hypothesis takes the form:

(4) $$H_A : E\{\underline{y}\} = Ax + c_y\nabla.$$

The alternative hypothesis is specified, once the vector c_y is specified (see Section 4.3). The appropriate teststatistic for testing the null hypothesis against the above alternative hypothesis H_A is given as:

$$\underline{T}_{q=1} = \underline{w}^2 = \frac{[c_y^* Q_y^{-1}\underline{\hat{e}}]^2}{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y}$$

or when the square-root is taken:

(5) $$\underline{w} = \frac{c_y^* Q_y^{-1}\underline{\hat{e}}}{\sqrt{c_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_y}}.$$

This teststatistic has a standard normal distribution $N(0,1)$ under H_0. The evidence on whether the model error as specified by (4) did or did not occur, is based on the two-sided test:

(6) $$|w| > N_{\frac{1}{2}\alpha_1}(0,1)$$

with α_1, the chosen level of significance.

Data snooping: Apart from the possibility of having a one dimensional test as (6), it is standard practice in geodesy to always first check the individual observations for potentional blunders. This implies that the alternative hypotheses take the form:

(7) $$H_{A_i} : E\{\underline{y}\} = Ax + c_{y_i}\nabla_i \qquad i=1,...,m$$

with

$$c_{y_i} = (0,...,0,1,0,...,0)^*.$$

Thus c_{y_i} is a unit vector having the 1 as its *i*th entry. The additional term $c_{y_i}\nabla_i$ models the presence of a blunder in the *i*th observation. The appropriate teststatistic for testing the null hypothesis against the above alternative hypothesis H_A, is again of the general form of (5), but now with the *c*-vector chosen as c_{y_i}, see also (83) in Section 4.3:

$$\underline{w}_i = \frac{c_{y_i}^* Q_y^{-1} \hat{\underline{e}}}{\sqrt{c_{y_i}^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} c_{y_i}}} . \tag{8}$$

This teststatistic has of course also a standard normal distribution $N(0,1)$ under H_0. By letting *i* run from 1 up to and including *m*, one can screen the whole data set on the presence of potential blunders in the individual observations. The teststatistic $\underline{w}_i$ which returns the in absolute value largest value, then pinpoints the observation which is most likely corrupted with a blunder. Its significance is measured by comparing the value of the teststatistic with the critical value. Thus the *j*th observation is suspected to have a blunder, when:

$$|w_j| \geq |w_i| \forall i \text{ and } |w_j| > N_{\frac{1}{2}\alpha_1}(0,1) \tag{9}$$

This procedure of screening each individual observation for the presence of a blunder, is known as *data snooping*.

In many applications in practice, the variance matrix Q_y is diagonal. If that is the case, the expression of the above teststatistic simplifies considerably. With a diagonal Q_y-matrix, we have:

$$\underline{w}_i = \frac{\hat{\underline{e}}_i}{\sigma_{\hat{e}_i}}.$$

The appropriate teststatistic is then thus equal to the least-squares residual of the *i*th observation divided by the standard deviation of the residual.

The *higher dimensional case*: It may happen that a particular model error can not be represented by a single scalar. In that case $q > 1$ and ∇ becomes a vector. The appropriate teststatistic is then the one we met earlier, namely:

$$\underline{T}_q = \hat{\underline{e}}^* Q_y^{-1} C_y (C_y^* Q_y^{-1} Q_{\hat{e}} Q_y^{-1} C_y)^{-1} C_y^* Q_y^{-1} \hat{\underline{e}}. \tag{10}$$

It is through the matrix C_y that one specifies the type of model error.

Adaptation

Once one or more likely model errors have been identified, a corrective action needs to be undertaken in order to get the null hypothesis accepted. Here, one of the two following approaches can be used in principle. Either one replaces the data or part of the data with new data such that the null hypothesis does get accepted, or, one replaces the original null hypothesis with a new hypothesis that does take the identified model errors into account. The first approach amounts to a remeasurement of (part of) the data. This approach is feasible for instance, when in case of datasnooping some individual observations are identified as being potentially corrupted

by blunders. These are then the observations which get remeasured. In the second approach no remeasurement is undertaken. Instead the model of the null hypothesis is enlarged by adding additional parameters such that all identified model errors are taken care of. Thus with this approach, the identified alternative hypothesis becomes the new null hypothesis.

Once the adaptation step is completed, one of course still has to make sure whether the newly created situation is acceptable or not. This at least implies a repetition of the detection step. It is possible that a gross error in one observation masks the gross error in another observation. This may have as consequence that the gross error which is masked, fails to have a large enough effect on its *w*-teststatistic; in other words, this *w*-test is not rejected. It is therefore good practice, once an observation is rejected, to repeat the adjustment without the rejected observation and again apply to this result the datasnooping procedure. In this way, one can infer whether it is likely that any gross errors remained undetected in the first step. Of course, if redundancy permits, one can repeat this again after the second step. This procedure is called *iterative datasnooping*.

When adaptation is applied, one also has to be aware of the fact that since the model may have changed, also the 'strength of the model' may have changed. In fact, when the model is adapted through the addition of more explanatory parameters, the model has become weaker in the sense that the teststatistics will now have less detection and identification power. That is, the reliability has become poorer. It depends on the particular application at hand, whether this is considered acceptable or not.

On the choice of testing parameters

When executing the above tests, choices need to be made about the testing parameters so as to control the errors of the first and second type. Although various approaches are possible, we only present briefly one such approach, namely the *B-method of testing* (Baarda, 1968). For a more detailed discussion on this topic, including the possible pittfalls involved, we refer to (Miller, 1966 and Arnold, 1981).

In the B-method of testing, the $T_{q=m-n}$-test of the detection step and the *w*-test of the identification step are related to each other by a special choice of their testing parameters:

$$\lambda_0 = \lambda(\alpha, q=m-n, \gamma=\gamma_0) = \lambda(\alpha_1, q=1, \gamma=\gamma_0). \tag{11}$$

The procedure is then to make a choice for α_1 and γ_0 and compute λ_0 and α from the above relation. This choice of equal values for the non-centrality parameter $\lambda = \lambda_0$ and power $\gamma = \gamma_0$ in both tests, implies that a certain model error can be found with the same probability by both the $T_{q=m-n}$ and the *w*-test. Both tests will therefore have the same reliability, i.e. the same values for the minimal detectable biases (MDB). Thus if the null hypothesis is accepted in the detection step, no further testing is necessary and the reliability for any 1-dimensional alternative hypothesis is given by its corresponding MDB computed on the basis of the value λ_0.

One consequence of the above coupling that one should be aware of, is the dependence of α on the redundancy *m-n*. Due to this coupling the value of α will increase when the redundancy

increases (see Figure C.1). For a large redundancy this may lead to a too large value of α, so that the null hypothesis gets too often falsely rejected. For such situations, Baarda proposes to carry out the adjustment and testing in steps.

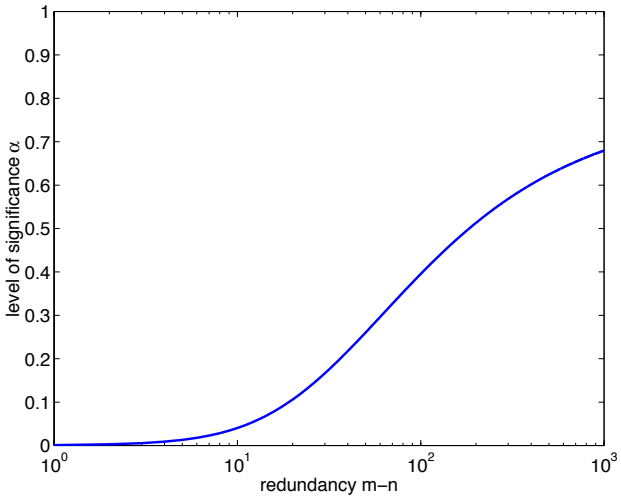


Figure C.1: Level of significance α versus redundancy *m-n* according to the B-method of testing (11); $\alpha_1 = 0.001$, $\gamma_0 = 0.80$.

Literature

1. Calculus and linear algebra

Almering, J.H.J. et al., revised by H. Bavinck and R.W. Goldbach; *Analyse*
Delft University Press, 6th ed. (1996)

Braber, C.A. den, H. van Yperen and M.A. Vlietgever; *Matrixrekening*
Delft University Press, 2nd ed. (1989)

Decnop, G.W., H. van Iperen and R. Martini; *Lineaire Algebra*
Delft University Press, 4th ed. (1986)

Lay, D.C.; *Linear algebra and its applications*
Reading, Addison-Wesley, 2nd ed. (1997)

Ortega, J.M.; *Matrix theory, a second course*
New York, Plenum Press, (1987)

Strang, G.; *Linear algebra and its applications*
San Diego, Harcourt Brace Jovanovich Publishers, 3rd ed. (1988)

Teunissen, P.J.G.; *Generalized inverses, adjustment, the datum problem and S-transformations*
In: Optimization and design of geodetic networks, Ed. E. Grafarend and F. Sansò, Springer Verlag, (1985), pp. 11-54

2. Probability theory

Breiman, L.; *Probability and stochastic processes: with a view toward applications*
Boston, Houghton Mifflin Company, (1969)

Grimmett, G. and D. Welsh; *Probability, an introduction*
Oxford, Clarendon Press, (1986)

Lopuhaä, H.P.; *Lecture notes: Statistiek voor geodeten*
Delft University of Technology, Faculty of Information Technology and Systems, (1994)

Miller, R.G.; *Simultaneous statistical inference*
New York, McGraw-Hill Book Company, (1966)

Papoulis, A.; *Probability, random variables and stochastic processes*
New York, McGraw-Hill Book Company, (1991)

Peebles, P.Z.; *Probability, random variables and random signal principles*
New York, McGraw-Hill Book Company, (1987)

Soest, J. van; *Elementaire statistiek*
Delft University Press, 7th ed. (1997)

3. Estimation theory and theory of hypothesis testing

Abramowitz, M. and I.A. Stegun; *Handbook of mathematical functions, graphs and mathematical tables*
New York, Dover, 7th ed. (1970)

Arnold, S.F.; *The theory of linear models and multivariate analysis*
New York, John Wiley & Sons, (1981)

Baarda, W.; *Statistical concepts in geodesy*
Netherlands Geodetic Commisssion, Publ. on Geodesy, New Series 2(4), (1967)

Baarda, W.; *A testing procedure for use in geodetic networks*
Netherlands Geodetic Commission, Publ. on Geodesy, New Series 2(5), (1968)

Ghosh, B.K.; *Some monotonicity theorems for χ^2, F and t-distributions with applications*
J.R. Statist. Soc. 35, (1973), pp. 480-492

Graybill, F.A.; *Theory and application of the linear model*
North Scituate, Duxbury Press, (1976)

Hays, W.L.; *Statistics*
New York, Holt, Rinehart and Winston, 4th ed. (1988)

Koch, K.R.; *Parameter estimation and hypothesis testing in linear models*
Berlin, Springer Verlag, 2nd ed. (1999)

Melsa, J. and D. Cohn; *Decision and estimation theory*
New York, McGraw-Hill Book Company, (1978)

Mikhail, E.; *Observations and least-squares*
New York, IEP-Dun-Donnelley, (1976)

Mood, A.M. and F.A. Graybill; *Introduction to the theory of statistics*
New York, McGraw-Hill Book Company, 3rd ed. (1974)

Neyman, J. and E.S. Pearson; *On the problem of the most efficient tests of statistical hypotheses*
Philos. Trans. Roy. Soc., London Series A231, (1933), pp. 289-337

Teunissen, P.J.G.; *Adjustment theory, an introduction*
Delft University Press, 1st ed. (2000)

Teunissen, P.J.G.; *Quality control in geodetic networks*
In: Optimization and design of geodetic networks, E. Grafarend and F. Sansò (editors), Springer Verlag, (1985), pp. 526-546

Teunissen, P.J.G.; *Adjusting and testing with the models of the affine and similarity transformation*
Manuscripta Geodaetica, 11(3), (1986), pp. 214-225

Index

Testing theory:
an introduction

Peter J.G. Teunissen

These lecture notes are a follow up on Adjustment theory. 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. 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. Mistakes or errors can come in many different guises. 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. When passed unnoticed, these errors will deteriorate the final results.

The goal of this introductory course on testing theory is therefore to convey the necessary knowledge for testing the validity of both the measurements and the mathematical model. Typical questions that will be addressed are: 'How to check the validity of the mathematical model? How to search for certain mistakes or errors? How well can errors be traced? And how do undetected errors affect the final results?' The theory is worked out in detail 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. As an additional aid in understanding the basic principles involved, a geometric interpretation is given throughout. Attention is also paid to the performance of the testing procedures. The closely related concept of reliability is introduced and diagnostic measures are given to determine the size of the minimal detectable biases. In this introductory text the methodology of testing is emphasized, although various examples are given to illustrate the theory. The methods discussed form the basis for geodetic quality control and they provide the ingredients for the formulation of guidelines for the reliable design of measurement set-ups.

TU Delft

P.J.G. Teunissen
Delft University of Technology
Faculty of Civil Engineering and Geosciences

Dr (Peter) Teunissen is Professor of Geodesy at Delft University of Technology (DUT) and an elected member of the Royal Netherlands Academy of Arts and Sciences. He is research-active in various fields of Geodesy, with current research focused on the development of theory, models, and algorithms for high-accuracy applications of satellite navigation and remote sensing systems. His past DUT positions include Head of the Delft Earth Observation Institute, Education Director of Geomatics Engineering and Vice-Dean of Civil Engineering and Geosciences. His books at TU Delft OPEN Publishing are Adjustment Theory, Testing Theory, Dynamic Data Processing and Network Quality Control.

ISBN 978-94-6366-892-7
DOI https://doi.org/10.59490/tb.96
textbooks.open.tudelft.nl

Cover image: P.J.G. Teunissen

Series on Mathematical Geodesy and Positioning