



DELFT UNIVERSITY OF TECHNOLOGY
Department of Geoscience & Remote Sensing

ADDITIONAL THESIS

Analysing confidence intervals with a strict method

*The difference between the classical and a conditional
testing-estimation*

Pepijn M. van Tol

Supervised by
Dr. Ir. C.C.J.M. (Christian) TIBERIUS
Dr. Marc SCHLEISS

February 3, 2019

Contents

- 1 Introduction** **2**

- 2 Parameter estimation and testing** **3**
 - 2.1 Model 3
 - 2.1.1 Case study 4
 - 2.2 Hypothesis testing 4
 - 2.2.1 Acceptance regions 6
 - 2.2.2 DIA-method 10
 - 2.3 Example parameter estimation case study 10
 - 2.3.1 Example classic method 11
 - 2.4 Strict method 13
 - 2.5 Example strict method 14
 - 2.6 Simulation 15

- 3 Analyzing confidence levels** **17**
 - 3.1 Variable alpha level 17
 - 3.1.1 Absolute values 17
 - 3.1.2 Relative values 18
 - 3.2 Variable bias size 20
 - 3.2.1 Absolute values 20
 - 3.2.2 Relative values 22

- 4 Discussion and conclusion** **23**
 - 4.1 Discussion 23
 - 4.2 Conclusion 23

1 | Introduction

Confidence intervals are often used to describe the quality of an estimator. The confidence interval is centered at the estimate. The confidence level describes the probability that the true value is inside a confidence interval. The length of the interval is based on the probability and driven by the precision. An estimator is called unbiased if the expected value of the estimator is the same as the unknown true value of the estimator. In reality the true value is never known therefore the parameter of interest is estimated.

When determining the length of the confidence interval the probability density function of the estimators have to be known. An approach that is often used is Best Linear Unbiased Estimation. For this estimation method a linear model is used to estimate parameters using observations. Often before the BLUE is applied the observations are tested for different hypotheses to detect outliers. Often the method Detection, Identification and Adaptation (DIA) is used in combination with the BLUE method. The DIA method tests different hypotheses and eventually selects one model. This model is then used in combination with the BLUE method. When the input is normally distributed due to the linearity of the BLUE method the estimator is also normally distributed. This relationship is also used when the DIA method is applied before the BLUE estimation. The problem is that the input is not normally distributed anymore when the input is first tested for outliers. Therefore it can be a better approach to combine testing and estimating using conditional distributions to determine the confidence intervals for the new estimator.

The fact that testing does influence the probability distribution of the input is found in other papers already like [Meeks and D'Agostino, 1983], which noticed the fact that testing should not be seen apart from the final estimation. Further [Gregoire et al., 1992] did also use conditional estimation for least squares estimators and [Chiou and Han, 1999] compared the confidence intervals between conditional estimation and unconditional estimation for a case where the null hypothesis is rejected.

More recent there are some articles about the conditioned estimation like [Teunissen, 2018] and [Zaminpardaz and Teunissen, 2018]. This report starts from [Zaminpardaz et al., 2019] about underestimating the integrity risk.

In comparison with [Chiou and Han, 1999], for this case study not only the alternative hypothesis is used, but also the case that the null hypothesis is accepted. Further more the approach used for this report is the other way around. They compared fixed confidence levels, while in this report the fixed confidence intervals are compared.

The main topic is to describe the quality of the parameter estimator with a strict method by means of a conditional probability density function. After that different confidence intervals are compared to see what the differences are between the conditional and the unconditional methods. The confidence intervals are compared for different alpha levels and different bias sizes under alternative hypotheses.

First the different methods are explained using an example. Using this example the first differences are already visible between the strict and the classical methods. After that a more in depth analysis is given about the comparison of the confidence levels under different scenarios.

2 | Parameter estimation and testing

Parameter estimation is used when an unknown parameter of interest is determined, based on observations. This can for example be something that cannot be determined directly or a parameter that is measured multiple times. An example of something that is often not measured directly are depths. Depths are determined using the travel time of a sound wave and the sound speed. Another example of a parameter that is often measured more than once are distances. In many cases when a distance is measured it is done multiple times to ensure that there are for example no reading errors. When dealing with multiple measurements it is possible to take for example the mean or the median of the dataset.

For parameter estimation the Best Unbiased Linear Estimator (BLUE) is often used. To use this method a linear model is needed. The BLUE method is a linear method which means that the probability distribution of the observables can be used to determine the probability distribution of the estimator. In this case it is assumed that the observables are normally distributed with a known variance. Because the BLUE is a linear method, the estimator also is normally distribution. The BLUE method is often used in combination with the method called Detection, Identification and Adaptation (DIA). In this method testing is involved to determine if there are faults in the observations, and when this is true to identify which measurement contains the error.

2.1 Model

A linear model formulates the observables expectation as a linear combination of parameters. A model is linear when both $f(x + y) = f(x) + f(y)$ and $f(ax) = af(x)$ are true. It is possible to have one or multiple parameters in the model, the amount of parameters determines the amount of columns in the design matrix. A linear model is shown below in equation (2.1). There x is the vector that contains the parameters. The A matrix is called the design matrix of the model.

$$\underline{y} = Ax + \underline{e} \quad (2.1)$$

Where the \underline{y} vector is the vector containing the observables. The i -th element of the \underline{y} vector is a linear combination through the i -th row of the A matrix and the parameters of the model in x . The vector \underline{e} contains the measurement noise.

Another matrix used for the BLUE method is the Q_{yy} matrix of which the inverse is used as the weight matrix. The Q_{yy} matrix is called the variance matrix of the observables. This matrix contain the variances for the different observables on the diagonal. When the observables are correlated, then this is also given in the Q_{yy} matrix. When all observables are uncorrelated then the Q_{yy} matrix is a diagonal matrix, see (2.2). If all observables have the same variance then all the values are the same, but it is also possible to combine different variances in the variance matrix. In this case the variance of the observables are known beforehand.

$$Q_{yy} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_i^2 \end{bmatrix} \quad (2.2)$$

So when the vector x is estimated, the BLUE method is used. This includes also the variance matrix Q_{yy} . The estimator reads as shown below (equation (2.3)):

$$\hat{x} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} \quad (2.3)$$

Since both the probability distribution and the variances of the input are known also the probability distribution of the estimator is known. This is due to the linearity of the BLUE method. Therefore the variance matrix of the estimator \hat{x} , $Q_{\hat{x}\hat{x}}$ is:

$$Q_{\hat{x}\hat{x}} = (A^T Q_{yy}^{-1} A)^{-1} \quad (2.4)$$

The probability function of the estimator is also a normal distribution with the estimated values given in \hat{x} and the variances from the variance matrix $Q_{\hat{x}\hat{x}}$.

The Q_{xx} matrix is used to describe the precision of the estimator. For example to give a length for an interval which need to meet a specified probability. The expected value of the estimator is equal to the true value of the parameter, i.e. $E(\hat{x}) = x$. When there is a bias in the estimation this is not true anymore and thus is the estimation not accurate.

2.1.1 Case study

In this case study a small linear model is used where the different aspects of testing and estimation can be demonstrated. This model can for example concern a distance measurement. It consist of two measurements, both the same distance. There is one unknown, the actual distance.

$$E\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} [x] \quad (2.5)$$

The observations y_1 and y_2 are normally distributed with a known variance.

2.2 Hypothesis testing

Hypothesis testing is used to determine if the model is sufficient. When the parameters for the model are estimated, but there are large residuals then this can be a sign that the model that was used was not good for that specific case.

Least squares residual is the difference between the observed and the estimated value. When multiple observations are used, then the least squares residual is a vector. This vector is given as \hat{e} .

$$\hat{e} = \underline{y} - \hat{\underline{y}} \quad (2.6)$$

Using the sum of the squared residuals in combination with the variance matrix of the observables it is possible to produce a measure of inconsistency between model and observations. The vector is then transformed to one normalized scalar which can directly be used for the hypothesis testing using a critical value.

$$\underline{T} = \|\underline{e}\|_{Q_{yy}}^2 = \hat{e}^T Q_{yy}^{-1} \hat{e} \quad (2.7)$$

Note that the expected value for the least squares residual is equal to zero. Since $E(\underline{y}) = E(\hat{\underline{y}})$ so that $E(\underline{e}) = E(\underline{y}) - E(\hat{\underline{y}}) = 0$.

The misclosure is described using another matrix. This is called matrix is called the B matrix. The transformation from the original model with a design matrix (A) to this model with a B matrix is called the Tienstra transformation.

Tienstra transformation \mathcal{T} is used to describe the model and the misclosure vector in one equation.

$$\mathcal{T} = \begin{bmatrix} [(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}] \\ B^T \end{bmatrix} \quad (2.8)$$

Where B a $[m \times r] = [m \times m - n]$ basis matrix of the null space of A^T [Teunissen, 2018]. m and n are the rows and columns of the A matrix. This means that $B^T A = 0$.

In the case of the used model, the Tienstra transformation looks like:

$$\mathcal{T} = \begin{bmatrix} [(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}] \\ -1 \quad 1 \end{bmatrix} \quad (2.9)$$

This is correct since $B^T A = [-1 \quad 1] [1 \quad 1]^T = 0$. The whole model after the Tienstra transformation looks like as follow:

$$\hat{\underline{z}} = \mathcal{T} \underline{y} \quad (2.10)$$

Where the parameter vector \hat{z} consist of two elements, namely the parameter which was called \hat{x} before and the misclosure \underline{t} . It is possible to normalize the misclosure vector so that the critical value can be used as described before.

$$\underline{T} = \|\underline{t}\|_{Q_{tt}}^2 = \underline{t}^T Q_{tt}^{-1} \underline{t} \quad (2.11)$$

Where Q_{tt} is defined as follow:

$$Q_{tt} = B^T Q_{yy} B \quad (2.12)$$

Critical value is a threshold to reject or not a hypothesis. The test statistic, the T -value, is χ^2 distributed. When the T -value is larger than a specified critical value then the hypothesis is rejected. This critical value is described with as k_α . This critical value is a function of the degree of freedom and the level of significance, also often called the alpha level. The degree of freedom is the amount of observables minus the amount unknown parameters. In this case there are two measured distances and one distance in unknown, thus the degree of freedom is $2 - 1 = 1$.

To get the critical value an alpha level should be selected. For each case when testing is used one can select the desired alpha level. The alpha level is used to describe how strict the test is. A large alpha value results in more false rejections and a small alpha level results in more false acceptances. Since the observables are normally distributed all values from minus infinity till infinity are possible to occur. The critical value with an alpha level of 0.05 and 1 degree of freedom is $= 3.8415$. As said, the observables can be anything from minus infinity till infinity so it is possible that an observable is rejected while it was correct. This occurs because it was not expected to get a value which differ that much from the expected value (the mean).

A missed detection occurs when a random variable does have an outlier but the outlier is not detected in the test. Using a smaller alpha level the probability on a missed detection increase in comparison with a larger alpha level. A false rejection occurs when a random variable does not contain any outlier but it is still rejected. Using a larger alpha level the probability on a false alarm increase in comparison with a smaller alpha level.

As an example the rejection areas are shown in graph, see figure 2.1. For this example the normal distribution $\mathcal{N}(0, 1)$ and the chi-square distribution $X^2(1)$ are used. The alpha level is set to 0.05. The values which are rejected are colored blue for the normal distribution and red for the chi-square distribution.

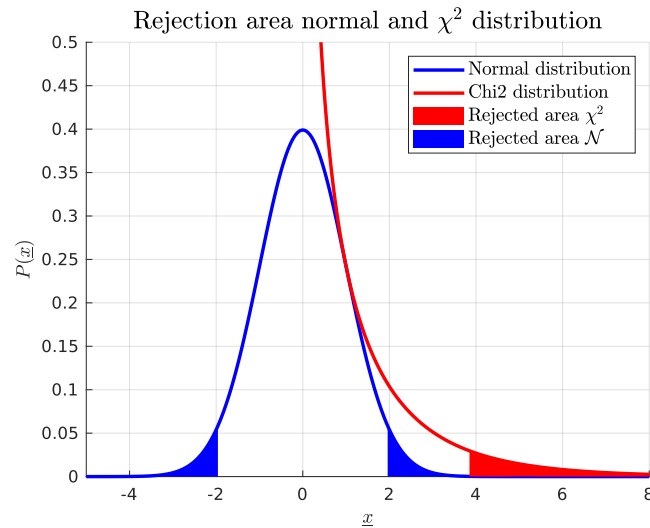


Figure 2.1: Rejection area for both the normal and chi square distribution

In this example the critical value k_α is 3.8415. To calculate if an observable should be rejected the observable is squared, see equation (2.7). Therefore the critical value in terms of misclosure $\frac{\underline{t}}{\sigma_t}$ with a normal distribution is $\pm\sqrt{k_\alpha}$.

Note that in this case only one degree of freedom is used. When the amount of measurements increase but the amount of parameters stay the same then the critical value gets bigger while keeping the alpha the same. The critical value itself is not influenced by the variance of the observation, this is already taken into account when calculating the $\|\hat{e}\|_{Q_{yy}}^2$ see equation (2.7).

Null hypothesis is the hypothesis that is expected to be true by default. The hypothesis is that the working model is correct. The null hypothesis is written as H_0 . For the given example the null hypothesis is thus:

$$H_0 : E \left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} [x] \quad (2.13)$$

When testing is thus included in the procedure of parameter estimation it is possible that the null hypothesis is accepted while in reality (which is not known) the hypothesis was not correct, this is a missed detection (MD). When the null hypothesis is accepted and it was in reality also true then it is called correct acceptance (CA).

When the null hypothesis is not rejected, the model above is assumed to be correct. Therefore the estimate is calculated using this model and equation (2.3). Using the variance propagation of the linear model it is possible to give the PDF for the estimator. This is $f_{\hat{x}_0}$ which is the normal distribution with the expected value equal to $E(\hat{x}_0) = x$ and a variance of $(\sigma_{y1}^{-2} + \sigma_{y2}^{-2})^{-1}$.

Alternative hypothesis is used when the null hypothesis is rejected. In a larger model typically multiple alternative hypotheses are possible, but in the case with two measurements we restrict to is only one alternative hypothesis. The alternative hypothesis is shown in equation (2.14).

$$H_a : E \left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ \nabla \end{bmatrix} \quad (2.14)$$

It is possible that the null hypothesis is rejected but that in reality the null hypothesis was true, this is called a false alarm (FA). When the null hypothesis is rejected and in reality the alternative hypothesis was true it is called a correct detection (CD).

Using the variance propagation law for linear models also the PDF for \hat{x}_a can be given as $f_{\hat{x}_a}$. The expected value for \hat{x}_a is given as $E(\hat{x}_a) = x$ and the variance is the same as the variance of the first measurement, σ_{y1}^2 .

Note that the alternative model is already a consistent model. Therefore a Tienstra transformation is not possible since in this case where the redundancy $r = 1$, the estimated bias $\hat{\nabla} = t$. This means that in the alternative hypothesis there is no misclosure. This is also the reason why it is impossible to test which of the two measurements contains an fault since it is also possible that the first observation contains a bias. For the examples used in this report it is assumed that there can only be a bias in the second measurement and therefore there is only one alternative hypothesis instead of two.

2.2.1 Acceptance regions

As said above the test result is not necessarily correct. This means that there are different possibilities when a test is accepted or rejected. However the total possibility for reality H_0 true or H_a true is equal to one. The area in terms of the misclosure t that accepts the null hypothesis is called \mathcal{P}_0 , and the area where the null hypothesis is rejected is called \mathcal{P}_a . This means that CA and MD both are in \mathcal{P}_0 and both CD and FA are in space \mathcal{P}_a . The space \mathcal{P}_0 is given as the area where $\|t\|_{Q_{tt}}^2 \leq k_\alpha$ and \mathcal{P}_a is $\|t\|_{Q_{tt}}^2 > k_\alpha$.

This means that $P(t \in \mathcal{P}_0|H_0) + P(t \in \mathcal{P}_a|H_0) = 1$. In words, the probability that the test result is either to accept or reject the null hypothesis when the null hypothesis is true is equal to one. The same is true for the case that the alternative hypothesis is true, $P(t \in \mathcal{P}_0|H_a) + P(t \in \mathcal{P}_a|H_a) = 1$.

The same principle is shown in table 2.1. The columns describe either H_0 or H_a true in reality and the row describe the test outcome.

In the same order as table 2.1 figure 2.2 is created. This means that the total area of both the correct acceptance (fig 2.2a) and the false alarm (fig 2.2c) combined give the probability density function (PDF) for the case that there is no bias in both measurements. This total area, where H_0 is accepted and

| | | Reality | |
|------|--|---------|-------|
| | | H_0 | H_a |
| Test | $H_0 (\underline{t} _{Q_{yy}}^2 \leq k_\alpha)$ | CA | MD |
| | $H_a (\underline{t} _{Q_{yy}}^2 > k_\alpha)$ | FA | CD |

Table 2.1: Possibilities between reality and test results

rejected is equal to one. The same applies for the case that there is a bias in one of the measurements. The total area of both the missed detection (fig 2.2b) and the correct detection (fig 2.2d) is equal to one.

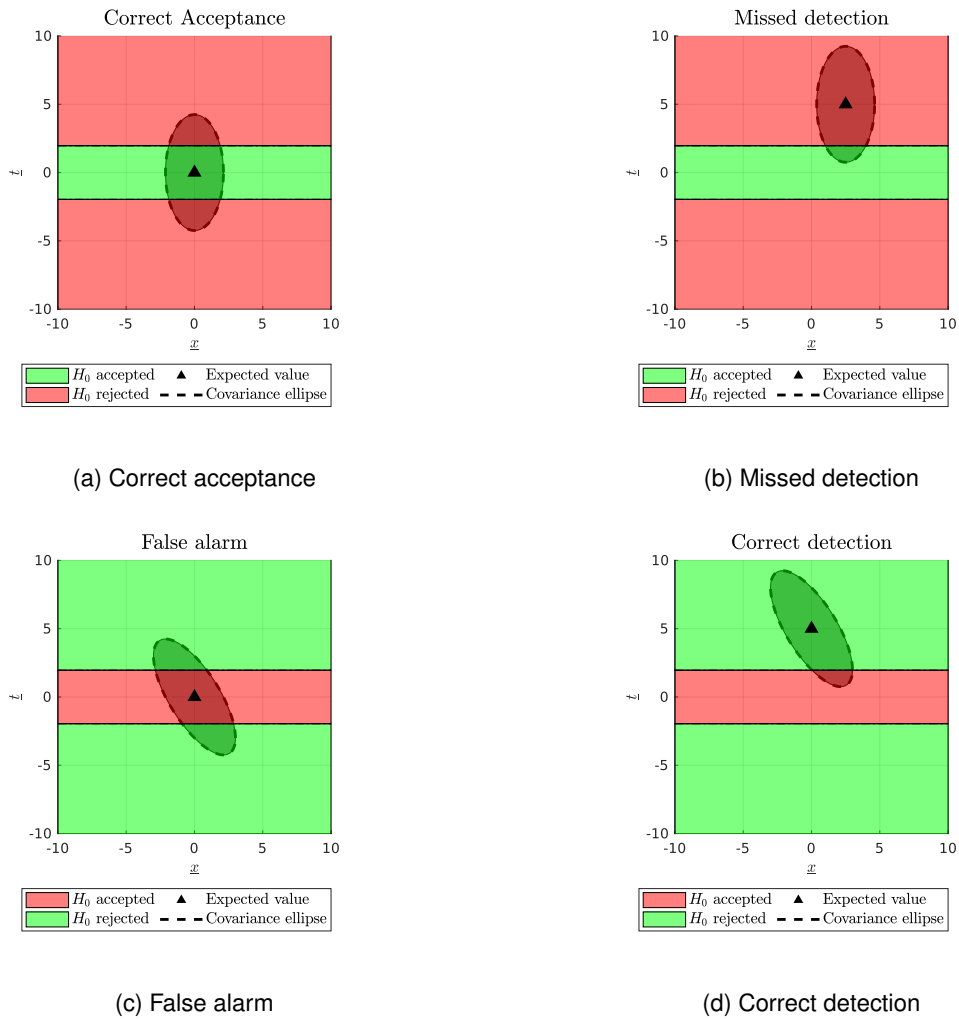


Figure 2.2: The four possibilities, the same order as table 2.1. Approximately 95% of the values lie in the variance ellipse. The expected values for both \underline{x} and \underline{t} are shown with the triangle.

Correct acceptance is the area $\mathcal{P}_0|H_0$. This area is shown in the upper left figure in green (figure 2.2a). Both \underline{t} and \underline{x} are normally distributed. Therefore it is possible to combine both PDF's of the different random variables and calculate the total probability $P(\underline{t} \in \mathcal{P}_0|H_0)$.

Using the multivariate normal distribution in combination with the theorem about marginal and condition distributions, [Teunissen et al., 2008] chapter 4, it is possible to write the probability of two random variables in a one dimensional function. In short the theorem describes the relationship between the marginal distribution and the conditional distribution of two variables, see below.

$$z_1 \sim \mathcal{N}_{n1}(\mu_{z_1}, Q_{11}) \quad (2.15)$$

$$z_2 | z_1 \sim \mathcal{N}_{n2}(\mu_{z_2|1}, Q_{22|1}) \quad (2.16)$$

$$\mu_{z_2|1} = \mu_{z_2} + Q_{21}Q_{11}^{-1}(z_1 - \mu_{z_1}) \quad (2.17)$$

$$Q_{22|1} = Q_{22} - Q_{21}Q_{11}^{-1}Q_{12} \quad (2.18)$$

Where both z_1 and z_2 are normally distributed, in vector notation: $\underline{z} \sim \mathcal{N}(\underline{\bar{z}}, Q_{zz})$. Q is the variance matrix of \underline{z} with Q_{ij} the element of the i -th row and the j -th column. In this case $\underline{z} = [x, t]^T$.

In this case, for correct acceptance, there is no correlation between \hat{x}_0 and \underline{t} . Therefore the variance matrix is a diagonal matrix so $Q_{21} = Q_{12} = 0$. This means that the marginal and conditional distributions are the same as the original distribution of \hat{x}_0 and \underline{t} . The probability part for correct acceptance is given as ca .

$$ca = f_{\hat{x}_0}(x|H_0) \int_{-\sqrt{k_\alpha}}^{\sqrt{k_\alpha}} f_t(t|H_0) dt \quad (2.19)$$

$\int_{-\sqrt{k_\alpha}}^{\sqrt{k_\alpha}} f_t(t|H_0) dt$ is the green area bounded by $t \in \mathcal{P}_0|H_0$. This area is defined with the critical value. All values for t inside the green area are accepted and thus are used to define the probability for correct acceptance. The region of $t \in \mathcal{P}_0$ is described with $\|t\|_{Q_{tt}}^2 \leq k_\alpha$.

False alarm is the area $\mathcal{P}_a|H_0$. This area is the lower left figure shaded in green (figure 2.2c). Here the same method is used before to combine the two random variables. This time there is correlation between \underline{t} and \hat{x}_a therefore Q_{xx} is not a diagonal matrix anymore, and a conditional probability is used.

$$fa = f_{\hat{x}_a}(x|H_0) \left(\int_{-\infty}^{-\sqrt{k_\alpha}} f_{t|\hat{x}_a}(t|x, H_0) dt + \int_{\sqrt{k_\alpha}}^{\infty} f_{t|\hat{x}_a}(t|x, H_0) dt \right) \quad (2.20)$$

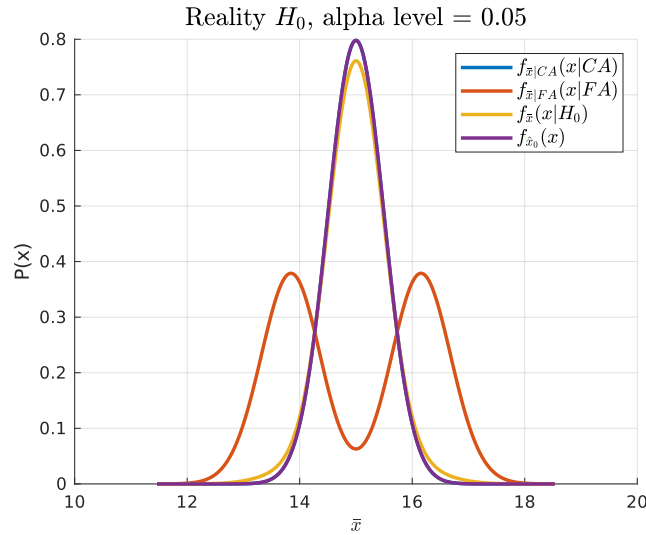
To describe the new estimator both cases, the probability for correct acceptance (ca) and the probability for false alarm (fa), are combined. This means that the equations (2.19) and (2.20) are not the probability density functions since the total integral of both functions is not equal to one.

Using this method another notation has to be used to distinguish between the different methods of estimation. Therefore for the classical method the estimator is denoted as \hat{x} and the notation for the strict method is \bar{x} .

The new probability density function for the estimator for the case that the null hypothesis (H_0) is true is given in (2.21).

$$f_{\bar{x}}(x|H_0) = f_{\hat{x}_0}(x|H_0) \int_{-\sqrt{k_\alpha}}^{\sqrt{k_\alpha}} f_t(t|H_0) dt + f_{\hat{x}_a}(x|H_0) \left(\int_{-\infty}^{-\sqrt{k_\alpha}} f_{t|\hat{x}_a}(t|x, H_0) dt + \int_{\sqrt{k_\alpha}}^{\infty} f_{t|\hat{x}_a}(t|x, H_0) dt \right) \quad (2.21)$$

The new PDF for the estimator when H_0 is true, is shown in the figure 2.3. The probability functions (2.19) and (2.20) of above are scaled so that these functions can be used as probability density functions to plot them.

Figure 2.3: Probability density function when H_0 is true

The functions $f_{\bar{x}|CA}(x|CA)$ and $f_{\hat{x}_0}(x|H_0)$ are the same. This is because when there is only correct acceptance then there is never a false alarm, which means that the original estimator is the same. Also the function for false alarm ($f_{\bar{x}|FA}(x|FA)$) is symmetric around the expected value, but has two peaks. This due to the fact that the rejection is divided over two regions, see figure 2.2c. The final PDF, $f_{\bar{x}}(x|H_0)$ is still symmetric around the expected value, but the PDF is wider and has thus thicker tails compared to $f_{\hat{x}}(x|H_0)$.

Missed detection is the area $\mathcal{P}_0|H_a$. This area is shown in green in figure 2.2b. The probability part for missed detection is given in equation (2.22).

$$md = f_{\hat{x}_0}(x|H_a) \int_{-\sqrt{k_\alpha}}^{\sqrt{k_\alpha}} f_t(t|H_a) dt \quad (2.22)$$

Correct detection is the area $\mathcal{P}_a|H_a$. This area is shown in green in figure 2.2d. The probability part for correct detection is given in equation (2.23).

$$cd = f_{\hat{x}_a}(x|H_a) \left(\int_{-\infty}^{-\sqrt{k_\alpha}} f_{t|\hat{x}_a}(t|H_a) dt + \int_{\sqrt{k_\alpha}}^{\infty} f_{t|\hat{x}_a}(t|H_a) dt \right) \quad (2.23)$$

The new probability density function for the estimator for the case that the alternative hypothesis (H_a) is true is given in (2.24). This PDF is shown in figure 2.4.

$$f_{\bar{x}}(x|H_a) = f_{\hat{x}_0}(x|H_a) \int_{-\sqrt{k_\alpha}}^{\sqrt{k_\alpha}} f_t(t|H_a) dt + f_{\hat{x}_a}(x|H_a) \left(\int_{-\infty}^{-\sqrt{k_\alpha}} f_{t|\hat{x}_a}(t|H_a) dt + \int_{\sqrt{k_\alpha}}^{\infty} f_{t|\hat{x}_a}(t|H_a) dt \right) \quad (2.24)$$

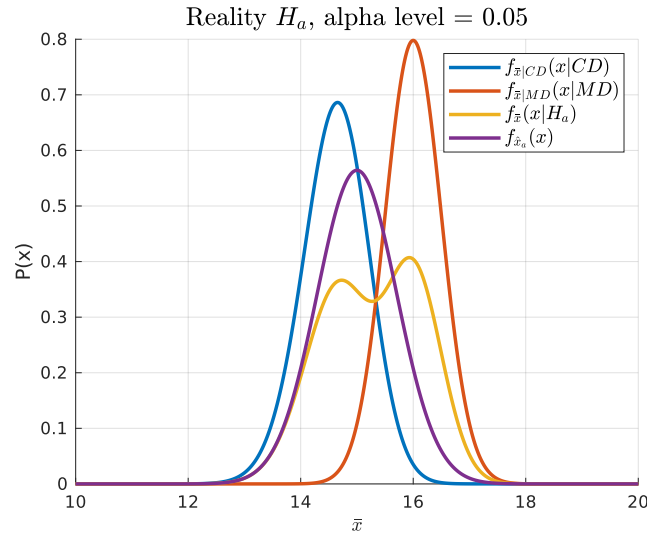


Figure 2.4: Probability density function when H_a is true, and the bias size in the second observation is 2.

In the figure above the probability function of ca and md are scaled so that those are also probability density functions. The purple function $f_{\hat{x}_a}$ is the function using the classic method. The PDF describing the missed detection is narrower compared to the one describing the correct detection. This is because the missed detection used the null hypothesis and thus has a smaller variance compared to the correct detection. The missed detection contains a significant bias, since the expected value is not equal to x anymore. In this case a true value of $x = 15$ was used. The PDF for the missed detection is not centered above 15. The expected value for \bar{x}_a is also not 15 and thus is the estimator not unbiased anymore.

2.2.2 DIA-method

In the Detection, Identification and Adaptation (DIA) method hypothesis testing is included. Normally first the null hypothesis is tested, which is also called the overall model test. This test checks if the overall model is correct. If this test is rejected then the identification is used. In this step alternative hypotheses are tested. Those alternative hypotheses can be anything, but what is often used is to check for each measurement if it contains an outlier. The alternative hypothesis with the smallest test-statistic is used for the parameter estimation. This means that there are as much alternative hypotheses as there are measurements, if this is possible. When a hypothesis is chosen the process starts again. This means that this is an iterative process and it is possible to detect multiple biases.

An example when this is not possible is when $r = 1$. Then there is not enough redundancy to test for each measurement if it contain an outlier, because if the hypothesis assumed an bias in one of the measurement, this means that the solution is unique already. But this is the case for both measurements, thus both alternative hypotheses, thus it is impossible to identify which of the two measurements contain a bias.

This DIA method is also called a two-step approach. This means that first the testing is used to decide which model should be used. The estimation is seen as a separate step and therefore both steps, testing and estimation, are not combined.

2.3 Example parameter estimation case study

In the classical way of parameter estimation the DIA method is often used. This is also called a two-step approach. The reason why this is a also called a two-step approach is because the testing and the parameter estimation are separated. This means that when a hypothesis is chosen, either the null hypothesis is accepted or one of the alternative hypotheses is accepted, it is assumed that the input is normal distributed. Due to the linearity of the BLUE estimation it is therefore also assumed that the estimation of the parameter is normally distributed with $\mathcal{N}(\hat{x}, Q_{\hat{x}\hat{x}})$.

2.3.1 Example classic method

For the example of the classical method the described model is used. There are two cases, namely one where the null hypothesis is true in reality and one where the alternative hypothesis is true in reality.

Null hypothesis true means that there is no bias in one of the measurements. So first define the measurements, the units are in meters:

$$\mathbf{y} = \begin{bmatrix} 14.8 \\ 15.1 \end{bmatrix} \text{ with } : Q_{yy} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \quad (2.25)$$

The "true" value $E(x) = 15$ which means that also the expected value for the observables $E(\mathbf{y}) = 15$ i.e. $\mathbf{y} \sim \mathcal{N}([15, 15]^T, Q_{yy})$. The observation vector \mathbf{y} is thus a realization of the random vector \mathbf{y} . The reason that the observations are not equal is due to the distribution of the observation vector, the measurement noise.

The Tienstra transformation is used to calculate the estimate \hat{x} and the misclosure t . Given the equation for the Tienstra transformation from (2.8)

$$\mathcal{T} = \begin{bmatrix} 0.5 & 0.5 \\ -1 & 1 \end{bmatrix} \quad (2.26)$$

Then using the model described in (2.14) $\hat{\mathbf{z}}$ can be calculated.

$$\hat{\mathbf{z}} = \begin{bmatrix} 0.5 & 0.5 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 14.8 \\ 15.1 \end{bmatrix} \quad (2.27)$$

$$\begin{bmatrix} \hat{x} \\ \hat{t} \end{bmatrix} = \begin{bmatrix} 14.95 \\ 0.3 \end{bmatrix} \quad (2.28)$$

Now the test to check if the null hypothesis is accepted starts. Therefore the t should be normalized to use a standard critical value. Now t is distributed as $t \sim \mathcal{N}(0, Q_{tt})$. Therefore the normalization is used as described in (2.11).

$$T = \|\hat{t}\|_{Q_{tt}}^2 = 0.3 \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} 0.3 = 0.09 \quad (2.29)$$

The critical value is distributed according to a chi-square distribution with 1 degree of freedom, i.e. $k_\alpha \sim \chi^2(1)$. The critical value for an alpha level of 0.05 is equal to 3.8415. The test-statistic T is not bigger than the critical value, thus the null hypothesis is accepted.

The variance matrix of the estimate \hat{x} is:

$$Q_{\hat{\mathbf{z}}\hat{\mathbf{z}}} = \mathcal{T}Q_{yy}\mathcal{T}^T = \begin{bmatrix} 0.25 & 0 \\ 0 & 1 \end{bmatrix} \quad (2.30)$$

This means that the precision of the parameter estimator is equal to $\sigma_{\hat{x}} = \sqrt{0.25} = 0.5$ meter. Another way to present the precision is the so called confidence interval. Three different intervals are given in table 2.2. The interval is described \pm the standard deviation. For example $\hat{x} \pm 2\sigma_{\hat{x}_0} = [13.95, 15.95]$, thus an interval between 14 and 16 meter has approximately a probability of 95% to contain the true x .

| | | | |
|---------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Confidence interval | $\hat{x}_0 \pm 1\sigma_{\hat{x}_0}$ | $\hat{x}_0 \pm 2\sigma_{\hat{x}_0}$ | $\hat{x}_0 \pm 3\sigma_{\hat{x}_0}$ |
| Confidence level | 0.6827 | 0.9545 | 0.9973 |
| Confidence interval | [14.45, 15.45] | [13.95, 15.95] | [13.45, 16.45] |

Table 2.2: Confidence intervals and corresponding confidence level for the estimate \hat{x}_0

Alternative hypothesis true means in this case that there is a bias in the second observation. The expected value of y_1 is still 15, but there is a bias in the second measurement. In this case the bias is 2 thus the second measurement is normally distributed: $y_2 \sim \mathcal{N}(17, \sigma_y^2)$. The second measurement has the same variance as the first measurement.

$$\mathbf{y} = \begin{bmatrix} 14.8 \\ 17.3 \end{bmatrix} \text{ with } : Q_{yy} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \quad (2.31)$$

The difference between the expected value, 15 meter, and the second measurement is 2.3 meter. The standard deviation of the measurement is $\sqrt{0.5} \approx 0.7$ meter. The probability that there is a random realization of an observation that differs at least 2.3 meter is in this case equal to 0.11%. This means that if the observation is done thousand times then there is only one measurement that differs at least 2.3 meter or more from the actual value. Based on ordinary measurement noise under H_0 it is assumed that this measurement contains a bias instead of that it is a realization from the observations under H_0 .

Using the same Tienstra transformation as before, the result is:

$$\hat{\mathbf{z}} = \begin{bmatrix} 0.5 & 0.5 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 14.8 \\ 17.3 \end{bmatrix} \quad (2.32)$$

$$\begin{bmatrix} \hat{x} \\ \hat{t} \end{bmatrix} = \begin{bmatrix} 16.05 \\ 2.5 \end{bmatrix} \quad (2.33)$$

And the normalized misclosure is:

$$T = \|\hat{t}\|_{Q_{tt}}^2 = 2.5 \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} 2.5 = 6.25 \quad (2.34)$$

The normalized misclosure is bigger than the critical value $\|\hat{t}\|_{Q_{tt}}^2 > k_\alpha$ thus the null hypothesis is rejected. Since there is only one alternative hypothesis in this case, it is known which model should be used, namely the alternative model described in (2.14).

$$\begin{bmatrix} \hat{x} \\ \hat{\nabla} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 14.8 \\ 17.3 \end{bmatrix} = \begin{bmatrix} 14.8 \\ 2.5 \end{bmatrix} \quad (2.35)$$

The variance matrix of the estimator \hat{z}_a is:

$$Q_{\hat{z}\hat{z}} = \mathcal{T} Q_{yy} \mathcal{T}^T = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 1 \end{bmatrix} \quad (2.36)$$

This means that the standard deviation of the estimator for the distance is $\sqrt{0.5} = 0.7071$ meter. Also for this estimate \hat{x}_a it is possible to give confidence intervals.

| | | | |
|------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Confidence interval relative | $\hat{x}_a \pm 1\sigma_{\hat{x}_a}$ | $\hat{x}_a \pm 2\sigma_{\hat{x}_a}$ | $\hat{x}_a \pm 3\sigma_{\hat{x}_a}$ |
| Confidence level | 0.6827 | 0.9545 | 0.9973 |
| Confidence interval | [14.09 - 15.51] | [13.39 - 16.21] | [12.67 - 16.92] |

Table 2.3: Confidence intervals and corresponding confidence levels for the estimate \hat{x}_a

In the case of the alternative hypothesis there is correlation between the bias $t = \hat{\nabla}$ and the estimator \hat{x}_a . This was also seen in figure 2.2d where the ellipse axes are not aligned with the coordinate axes. In comparison with the null hypothesis the $\sigma_{\hat{x}_a}$ is bigger, namely equal to σ_{y_1} .

Result classic DIA method To view the difference between the two methods both cases are shown in figure 2.5, below.

Both the theoretical probability density functions are drawn around the expected value (15). The estimates are also shown as dots on the x-axis. What can be seen is that the variance of the estimator with the alternative hypothesis is bigger, hence the graph of the PDF is wider.

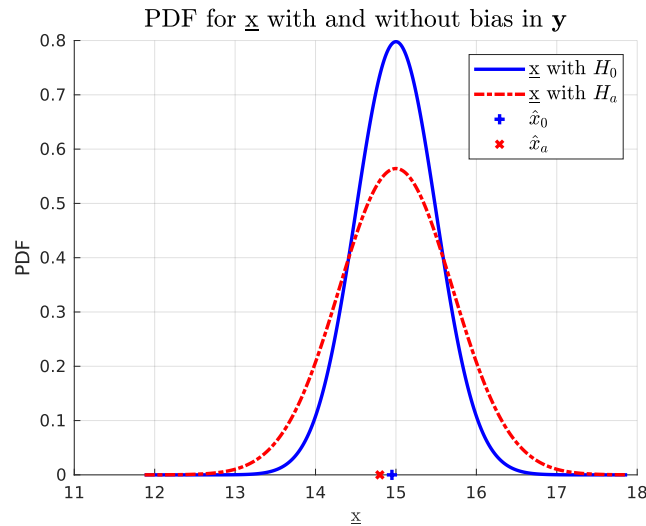


Figure 2.5: Estimator using the null hypothesis (\hat{x}_0), and the alternative hypothesis (\hat{x}_a).

When comparing the different confidence levels from tables 2.2 and 2.3 it is already visible that there is a difference in the width of the PDF. The estimator from the alternative hypothesis result in a wider PDF, thus bigger confidence intervals to describe the same confidence level.

Based on the disclosure one of the models is selected using the testing procedure and that model is used. To give the probability density function of the estimator only that model is used. This means that the linearity of the model is used to propagate the PDF of the observables to the estimators. While when following this procedure it is actual a conditional probability since the condition is that the disclosure in absolute sense is either smaller or bigger than the critical value.

As seen in the figures above, 2.2 it shows that when in reality there is no bias, it is still possible that the test rejects the null hypothesis. For example in the second case, where $y_2 = 17.3$ it is possible that it was not a bias but a realization of the normally distributed observables under H_0 , for which a realization is possible from minus infinity till infinity. However the classic approach does not include the fact that it is possible that there is no bias and only focusses as if the test only returns the truth, i.e. that when the test rejects the null hypothesis then it should be true. The same is true for the first case, it is possible that the second observation ($y_2 = 15.1$) contains a bias. However the test accepts the null hypothesis and it is used like this is in reality true, but it does not include the fact that it is also possible that there is a missed detection.

2.4 Strict method

In the strict method the fact that testing is used to select the model is also used to determine the PDF of the estimator. To do this and to check what the differences are, the new PDF is given again for two cases, one where there is no bias and one with a bias in the observation. The two cases are shown in table 2.1 and figure 2.2 as the two columns. Thus when in reality the null hypothesis is correct then the two possibilities are correct acceptance and false alarm. When in reality the alternative hypothesis is correct the two possibilities are correct detection and missed detection.

In the strict method the probability that the test result is not correct is also used to determine the probability density function of the estimator. This new PDF is then used to describe the precision of the estimator with confidence levels and confidence intervals. The beginning of the strict method is the same as for the classic method since the testing remains the same. The probability distribution of the observables (the input) does not change. After a specific hypothesis is chosen then also the estimate is calculated in the same way. Only the confidence interval of the estimator is different in both methods.

2.5 Example strict method

For the example of the strict method the same cases are used as which were used for the example of the classic method. Since the numbers of observables are the same it is possible to notice the differences between the two methods.

Null hypothesis true means that there is in reality no bias. Both the correct acceptance and the false alarm are used to determine the PDF of the estimator. Therefore the estimate for \bar{x} is equal with \hat{x}_0 , namely 14.95 (2.28). However the confidence level is different for for the same interval when \bar{x} is compared with \hat{x}_0 .

| Confidence interval relative | $\hat{x}_0 \pm 1\sigma_{\hat{x}_0}$ | $\hat{x}_0 \pm 2\sigma_{\hat{x}_0}$ | $\hat{x}_0 \pm 3\sigma_{\hat{x}_0}$ |
|------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Confidence interval | [14.45, 15.45] | [13.95, 15.95] | [13.45, 16.45] |
| Confidence level \bar{x} | 0.6536 | 0.9256 | 0.9843 |
| Confidence level \hat{x}_0 | 0.6827 | 0.9545 | 0.9973 |

Table 2.4: Confidence intervals corresponding with it confidence level for the estimate \bar{x}_0 and the confidence level of \hat{x}_0

In table 2.4 both the confidence levels for the strict and the classic methods are shown. This makes it clear that the same interval does not describe, with the same probability, that the real x is in the interval.

There is thus a difference between the two probability density functions. This is visible in figure 2.6. The area that is colored red is where the PDF of \bar{x} is smaller than the PDF of \hat{x}_0 . The area that is colored green is where the PDF of \bar{x} is bigger than the PDF of \hat{x}_0 .

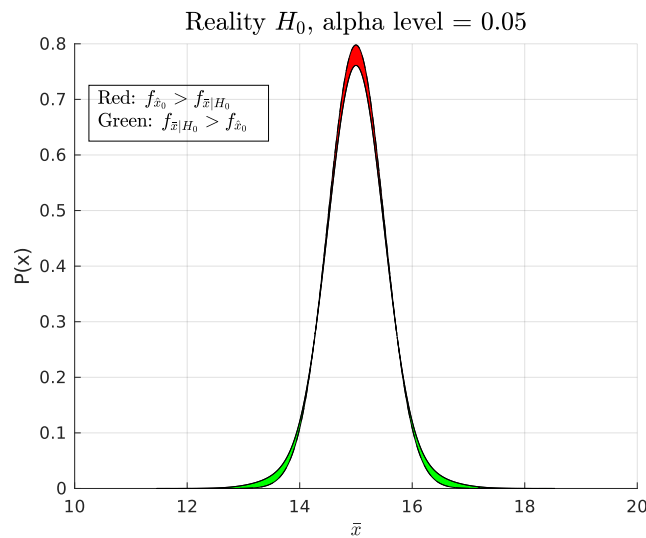


Figure 2.6: Difference between the PDF's for \hat{x}_0 and \bar{x}_{H_0}

Alternative hypothesis true means that there is an bias in the second observation. This is the same as before. The null hypothesis is thus rejected since $\|t\|_{Q_{tt}}^2 > k_\alpha$. The estimate \bar{x} is the same as the estimate \hat{x}_a since the same model is used for the estimation. The estimate is thus 14.8 (2.33).

The probability density functions for both estimators are different, which means that there is also a difference in the confidence intervals.

| | | | |
|------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Confidence interval relative | $\hat{x}_a \pm 1\sigma_{\hat{x}_a}$ | $\hat{x}_a \pm 2\sigma_{\hat{x}_a}$ | $\hat{x}_a \pm 3\sigma_{\hat{x}_a}$ |
| Confidence interval | [14.09, 15.51] | [13.39, 16.21] | [12.67, 16.92] |
| Confidence level \bar{x} | 0.4897 | 0.8785 | 0.9826 |
| Confidence level \hat{x}_a | 0.6827 | 0.9545 | 0.9973 |

Table 2.5: Confidence intervals and corresponding confidence levels for the estimate \bar{x} and the confidence level of \hat{x}

In table 2.5 both the confidence levels for the strict and the classic methods are shown. The difference when a small confidence interval is used is significant. With the classic method it was assumed that the actual value is contained with more than 68% probability while in reality it is less then 49%. The differences between the two probability density functions are shown in figure 2.7. The red area describes where $f_{\hat{x}_a}$ is bigger than $f_{\bar{x}}$, and the green area is where $f_{\hat{x}_a}$ is smaller than $f_{\bar{x}}$.

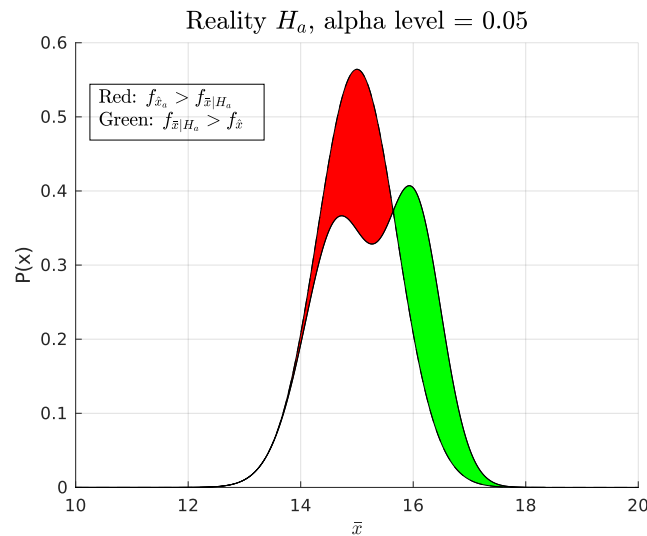


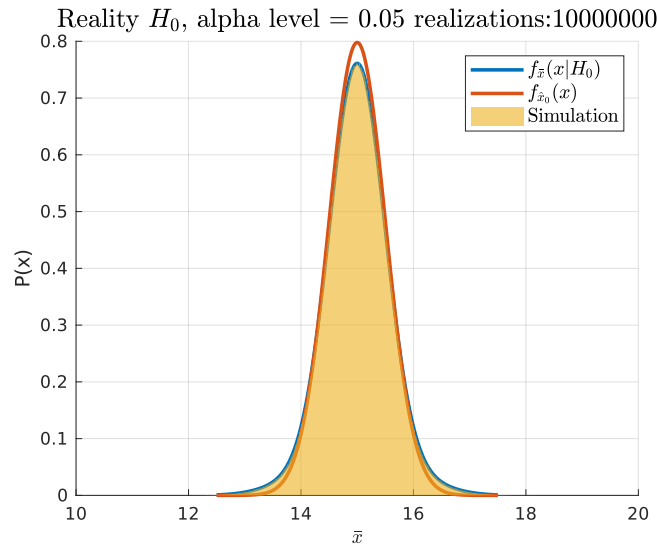
Figure 2.7: Difference between the PDF's for \hat{x}_a and \bar{x}_{H_a}

2.6 Simulation

To check if the theoretical probability density functions are correct multiple realizations are simulated. This means that for N times the whole procedure is simulated once for the case that there is in reality no bias and once for the case that there is a bias. This means that there are N y realization and also N \bar{x} estimates.

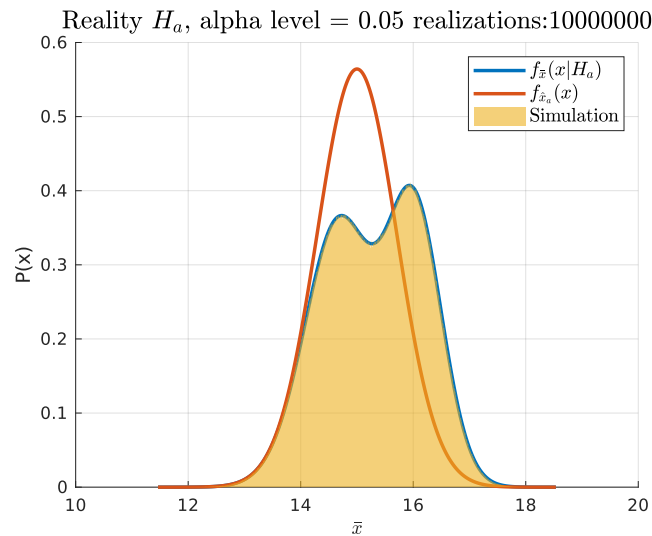
First the case where H_0 is true is simulated, see figure 2.8. Again the same parameters are used, therefore the expected value is still 15 and the used alpha level for the testing is set to 0.05. What can be seen is that the simulation precisely follows $f_{\bar{x}}$. This indicates that the theoretical PDF used above is indeed correct according to this simulation.

Using numerical integration it is possible to check if the total area of the PDF converges to one. This is also the case which means that the function can indeed be a probability density function for \bar{x} .

Figure 2.8: Simulation for reality no bias (H_0 is true)

Using the mean of all those simulations the expected value is simulated. This is also 15, which is the same as the expected value for \hat{x}_0 and the real value for x . Therefore there is no bias in this PDF.

The second case where H_a is true is also simulated, see figure 2.9. Also here the same parameters are used. Thus the expected value is 15, the used alpha level for testing is 0.05 and the bias size is 2. The simulation follows again $f_{\bar{x}}$. Using numerical integration to check the total area under the PDF also converges to one.

Figure 2.9: Simulation for reality with a bias (2) (H_a is true)

The difference with $f_{\bar{x}}$ is that the mean is not 15 anymore. The mean of the simulation is 15.29. Therefore this PDF is biased and $E(\bar{x}|H_a) \neq x$.

3 | Analyzing confidence levels

For the analysis of confidence levels different intervals are used to determine the difference between the strict method compared to the classic method. Therefore different scenarios are used to determine how and when the effects are significant. The two variables considered are the alpha level used for testing and the bias size under H_a . A confidence level is the probability that multiple confidence intervals will contain the expected value of the unknown parameter.

In this chapter the expected value x is zero, the variance of the observables is equal to one.

To analyse the difference in confidence levels the confidence level of both the strict and the classic method are calculated. Therefore multiple intervals are used and the probability that the interval describes is used to determine the difference between the two methods. Note that the probability interval $[x \pm s\sigma_x]$ is used instead of the confidence interval $[\hat{x} \pm s\sigma_{\hat{x}}]$.

3.1 Variable alpha level

When using a test to determine if the null hypothesis should be rejected or not it is needed to set an alpha value. The critical value, which is used to either reject or accept a hypothesis, is a function of the alpha level and the degree of freedom. The degree of freedom stays the same, therefore the alpha level does influence the critical value and thus when the null hypothesis is reject with a small or a large value for the test-statistic.

3.1.1 Absolute values

First the confidence levels are compared in terms of absolute values. This means that the probability from \bar{x} on a given interval with a specified alpha level is compared with the probability for the same interval and the same alpha level for \hat{x}_0 .

$$\Delta = \int_{-s\sigma_y}^{s\sigma_y} f_{\hat{x}}(x|H_0)dx - \int_{-s\sigma_y}^{s\sigma_y} f_{\bar{x}}(x|H_0)dx \tag{3.1}$$

Where Δ is the difference between the two confidence levels. $s\sigma_y$ is the bound of the confidence interval. The interval is a function of the standard deviation of the observable. Thus $s = 3$ means thus 3 times the standard deviation of the observable to either side of the true value for x , which was set to zero here.

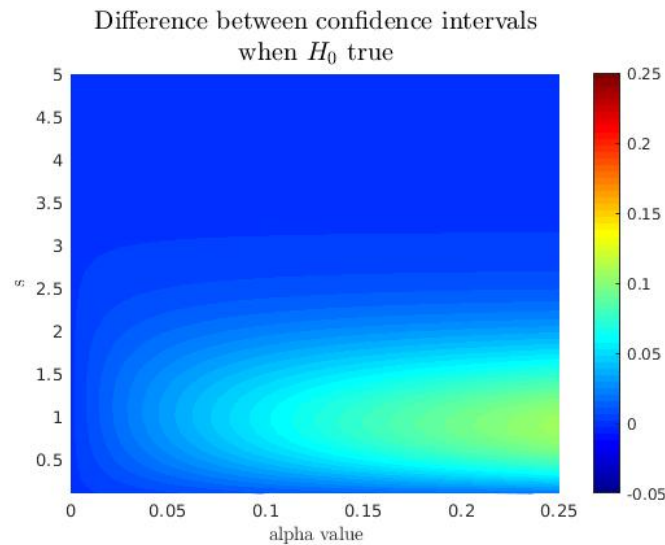


Figure 3.1: The difference between $P(\hat{x}_0)$ and $P(\bar{x})$ on an interval with a given alpha level

Figure 3.1 shows the absolute difference between the probabilities described by the fixed intervals. The alpha level on the horizontal axis is thus used to determine how strict the test is. The s on the vertical axis gives the interval as $x \pm s\sigma_y$ where s is a scale factor and x is the unknown true value.

What can be seen is that the difference is more significant at smaller intervals and higher alpha levels. When either the alpha level is zero or the interval is infinitely large the difference is zero.

The reason why the bigger error occurs when the alpha level increase is because with a higher alpha level there are more false alarms. More false alarms means that the null hypothesis is more often rejected thus the PDF with a bigger variance is used more often. When the interval increase the total probability confidence level converges to one. Both probability density functions have thin tails. Therefore when the intervals are big enough compared to its variance, the difference between the two confidence levels decreases till it is (almost) zero.

When the interval is small the absolute difference is close to zero. This is because the interval is very small and both confidence levels approach zero.

The other case, where the null hypothesis is incorrect, can be used in the same format. Again the intervals (vertical axis) and the alpha levels (horizontal axis) are variable and in the same way the difference between both confidence levels are calculated.

$$\Delta = \int_{-s\sigma_y}^{s\sigma_y} f_{\hat{x}}(x|H_a)dx - \int_{-s\sigma_y}^{s\sigma_y} f_{\bar{x}}(x|H_a)dx \quad (3.2)$$

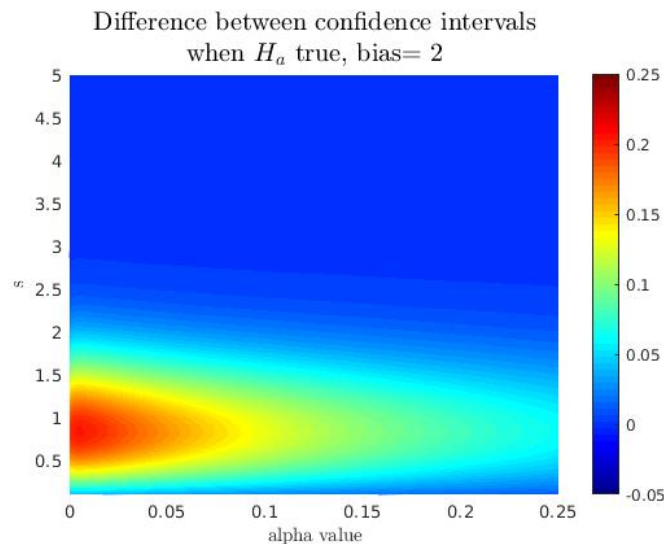


Figure 3.2: The difference between $P(\hat{x}_a)$ and $P(\bar{x})$ on an interval with a given alpha level under H_a

In figure 3.2 the difference is shown for the case that the alternative hypothesis is true. What can be seen is that when the interval increase the differences converge to zero. This is because the total probability converge to one for both probability density functions thus the difference decrease also. Also in this case where the interval is small the difference is almost zero. This is because the differences are shown in the figure thus when the interval is small the difference is also very small, i.e. both confidence levels approach zero.

The biggest difference for the alternative hypothesis can be found for small alpha values. With a small alpha level there is more change on a missed detection and less change on a correct detection. When the alpha level increases the null hypothesis is rejected faster. Since this is the case where the alternative hypothesis is correct, the more probability there is to reject the null hypothesis, the closer will $P(\bar{x})$ be to $P(\hat{x}_a)$.

3.1.2 Relative values

Another possibility is to create the same plots but this time for the relative error. This means the difference between the two different methods like above, but divided by the probability of the classical method.

The difference is then shown as a percentage of the classic confidence level.

The relative difference is calculated according to equation (3.3).

$$\Delta = \frac{\int_{-s\sigma_y}^{s\sigma_y} f_{\hat{x}}(x|H_0)dx - \int_{-s\sigma_y}^{s\sigma_y} f_{\bar{x}}(x|H_0)dx}{\int_{-s\sigma_y}^{s\sigma_y} f_{\hat{x}}(x|H_0)dx} \quad (3.3)$$

This is shown in the figure blow:

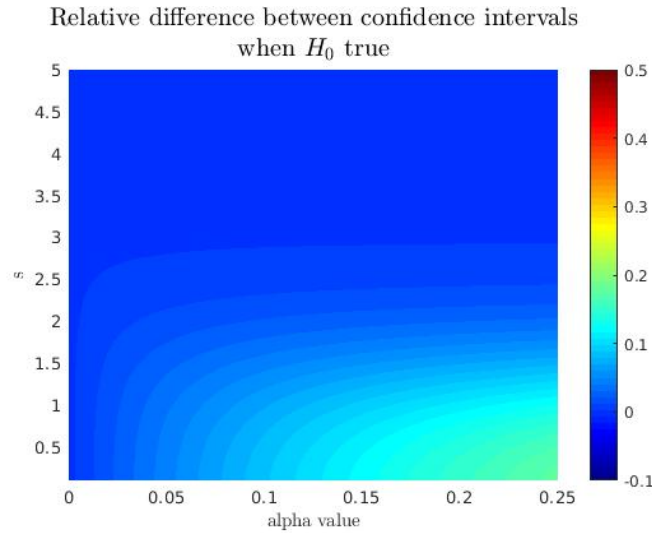


Figure 3.3: The relative difference between $P(\hat{x}_0)$ and $P(\bar{x})$ at an interval with a given alpha level under H_0

What can be seen in figure 3.3 is the same behaviour as before, but this time the relative difference is also significant for small intervals. Still either a big interval or a small alpha value overcomes the problem of different confidence levels and the relative difference is (almost) zero.

A numerical example will show that the relative difference can indeed be that big. Using a variance of 1 for the observables, an alpha level of 0.25 and an interval of $[0 \pm 0.5\sigma_y]$. The interval is then $[-0.5, 0.5]$. The confidence level for $\hat{x}_0 = 0.39$ and the confidence level for $\bar{x}|H_0 = 0.32$. The relative difference is thus $\frac{0.39-0.32}{0.32} = 0.18$.

Also the case that the alternative hypothesis is true can be normalized. This is done using equation (3.4).

$$\Delta = \frac{\int_{-s\sigma_y}^{s\sigma_y} f_{\hat{x}}(x|H_a)dx - \int_{-s\sigma_y}^{s\sigma_y} f_{\bar{x}}(x|H_a)dx}{\int_{-s\sigma_y}^{s\sigma_y} f_{\hat{x}}(x|H_a)dx} \quad (3.4)$$

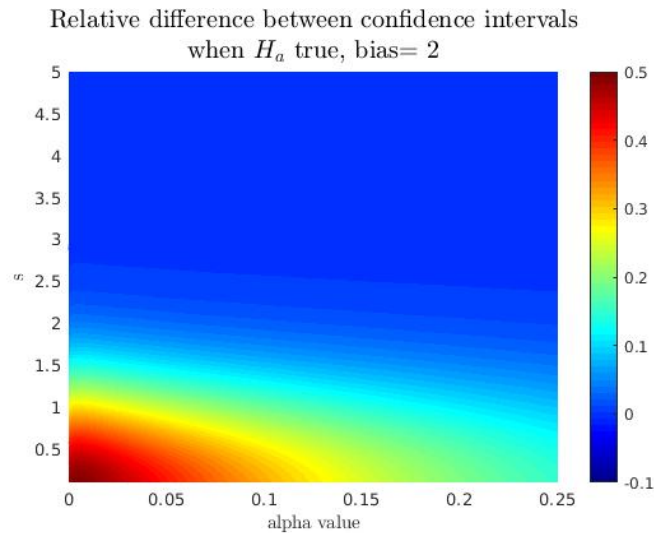


Figure 3.4: The relative difference between $P(\hat{x}_a)$ and $P(\bar{x})$ on an interval with a given alpha level under H_a

Figure 3.4 shows the relative difference between the probability calculated using the two methods. What can be seen in the figure is that the biggest errors occurs at small intervals and with small alpha values.

A numerical example shows that there can be big relative differences between the two different methods. Using a variance of 1 for the observables, an alpha level of 0.005 and an interval of $[0 \pm 0.5\sigma_y]$. The interval is again $[-0.5, 0.5]$. The confidence level for $\hat{x}_a = 0.38$ and the confidence level for $\bar{x}|H_a = 0.22$. The relative difference is thus $\frac{0.38-0.22}{0.38} = 0.42$.

The reason that bigger differences show up when the alternative hypothesis is true is because the shape of $f_{\bar{x}}(x|H_a)$ differs more from the normal distribution than $f_{\bar{x}}(x|H_0)$. Also $f_{\bar{x}}(x|H_0)$ is biased, thus small intervals centered around \hat{x}_a does not contain the biggest densities of $f_{\bar{x}}(x|H_a)$.

3.2 Variable bias size

Now the bias size is used as variable.

The expectation is that in a case with big biases or big intervals the difference is not significant. But whether or not a bias is big depends on the precision of the measurement. Therefore the bias is divided by the standard deviation of the observable. The result is thus a ratio $\frac{bias}{\sigma_y}$, a value without units.

3.2.1 Absolute values

The first scenario shows the differences between the classic and the strict method. Equation (3.2) is used to calculate the differences.

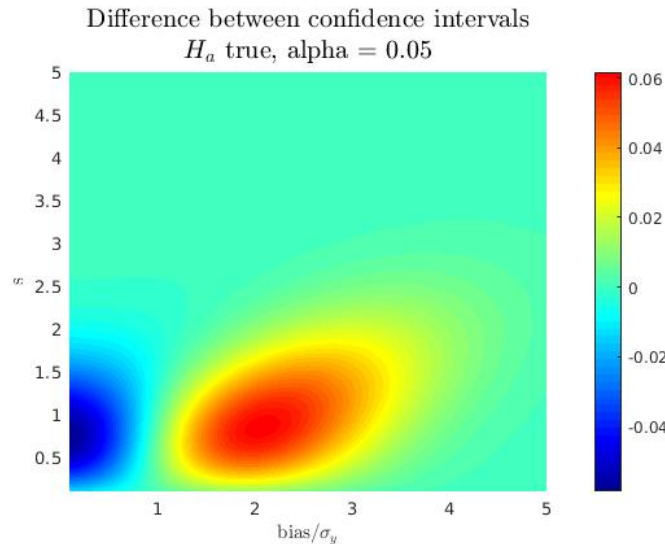


Figure 3.5: The difference between $P(\hat{x})$ and $P(\bar{x}_a)$ on an interval with a given alpha level and changing bias/ σ_y ratio under H_0

What can be seen is that here is a special case. At small intervals and small bias sizes the difference is negative. This means that the confidence level of the interval $\bar{x}|H_0$ is bigger than confidence level under \hat{x} . The reason for this behaviour is that when the bias is small it is not detected. The region of negative values is in the area with small biases and small intervals. When the interval increase the difference between the classic and the strict method decrease. As long as the misclosure is small and the null hypothesis is not rejected the null hypothesis is used, which is thus a missed detection. The null hypothesis has a smaller variance than the alternative hypothesis and therefore the difference between confidence levels at the given interval is negative.

When the bias is around 1.5 times the variance of the observable it is often not detected. The PDF $f_{\bar{x}}(x|H_a)$ is shown in figure 3.6. This is the same case as before, but this time the bias size is $1.5\sigma_y$. It can be seen that the missed detection has a lot of influence on the final PDF. Also the correct detection shows a small bobble at the right. This means that the ellipse from figure 2.2d is detected on both sides of the acceptance area (shown in red).

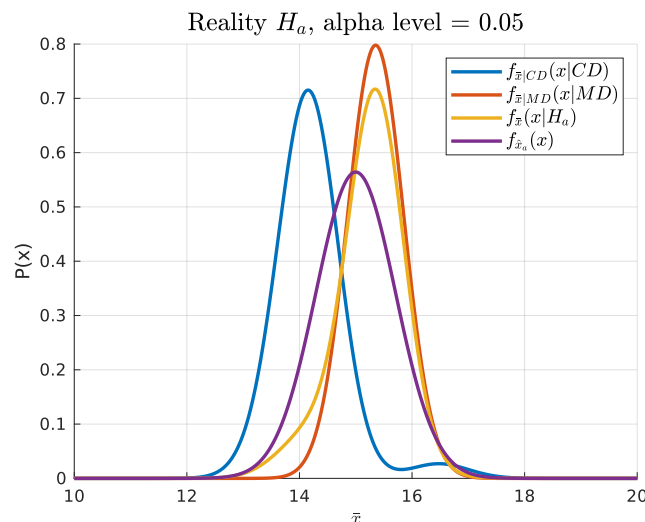


Figure 3.6: The PDF $f_{\bar{x}}(x|H_a)$ for a bias size $1.5\sigma_y$.

3.2.2 Relative values

The relative difference is calculated using equation (3.4). The result is shown in figure 3.7.

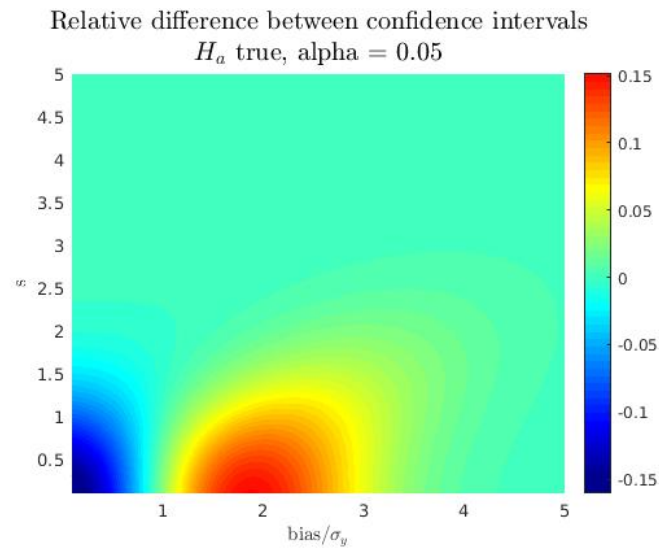


Figure 3.7: The relative difference between $P(\hat{x}_\alpha)$ and $P(\bar{x})$ on an interval with a given alpha level and changing bias/ σ_y ratio

What can be seen is that there is approximately the same behaviour as with the absolute case. However the relative difference remain higher when the bias and the used interval are both increased. This means that there is a positive correlation between the confidence interval and the bias/ σ_y .

4 | Discussion and conclusion

4.1 Discussion

After analyzing the confidence intervals it is clear that the fact that testing is used before parameter estimation should be accounted for when determining the probability density function for the estimator. This is important in particular when in reality there is a bias in one of the measurements.

A simple linear model was used to check if there are indeed differences between the classic and strict method. The next step can be to check how the confidence levels change when multiple measurements are combined. In that case if there is still one measurement containing a bias it would be interesting to see how this influences the confidence levels.

In this report the cases that are studied are cases where the reality is known. When testing is used in practice it is not known whether the parameter is estimated with a reality that H_0 is true or with a reality that H_a is true. The next step will be to understand what the different scenarios means for new parameter estimations. This means that both correct acceptance and missed detection should be combined, since they both are possible outcomes when the null hypothesis is not rejected. Both false alarm and correct detection should be combined when the null hypothesis is rejected, assuming the same model is used and thus that there are only two possible hypotheses. When the PDF for the estimator is determined knowing that the alternative hypothesis is true, it has been shown that this PDF is biased. When the PDF of the estimator is determined based on the test outcome instead on reality, which was the case for this report, then it is most likely that the estimator is biased when the null hypothesis is accepted. To make it possible to give a correct quality description of estimated parameters it is thus necessary to understand the effect of testing depending on the test outcome.

4.2 Conclusion

Confidence intervals and confidence levels are often used to describe the precision of an estimator. Therefore it is important to understand what the influence of testing can be on the quality description of the resulting estimator. In some cases the difference in confidence levels between not accounting and accounting for the impact of testing can be of significant size.

Overall it can be said that when in reality the null hypothesis is true the consequences are less dramatic than in cases where the alternative hypothesis is true. Using the strict estimator in the case that the null hypothesis is true the expected value of $E(\bar{x}|H_0) = x$. The probability density function is wider than the one from the classic method. The significance of the effect is dependent on the alpha level that is used for testing.

Another case is when the alternative hypothesis is true. The difference in confidence level of a given interval can be large. Also the fact that $E(\bar{x}|H_a) \neq x$ can be problematic when one assumes that the estimator is unbiased. The PDF of $\bar{x}|H_a$ has a more complex form and depends on both the alpha level and the bias size.

After analyzing the plots in chapter 3, it is clear that it is not possible to ignore the fact that the strict and the classic method result in different confidence levels. To minimize the false alarm probability the alpha level used for testing should be as small as possible, so that the null hypothesis is never falsely rejected. However to minimize the missed detection probability, which causes the complexity under H_a , the alpha level should be large so that the null hypothesis is rejected already on a small misclosure.

In general it can be said that when the classic method of estimation is used, the confidence interval is set too small for the given confidence level - it is too optimistic on the quality of the estimator when in practice actually \bar{x} is used. The only exception is when there is a small bias and a small interval is used, then the confidence interval is overestimated.

Bibliography

- [Chiou and Han, 1999] Chiou, P. and Han, C.-P. (1999). Conditional interval estimation of the ratio of variance components following rejection of a pre-test.
- [Dekking et al., 2005] Dekking, F., Kraaikamp, C., Lopuhaä, H., and Meester, L. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Springer.
- [Gregoire et al., 1992] Gregoire, T. G., Arabatzis, A. A., and Reynolds, M. R. (1992). Mean squared error performance of simple linear regression conditional upon the outcome of pretesting the intercept. *The American Statistician*, 46(2):89–93.
- [Meeks and D’Agostino, 1983] Meeks, S. L. and D’Agostino, R. B. (1983). A note on the use of confidence limits following rejection of a null hypothesis. *The American Statistician*, 37(2):134–136.
- [Montgomery and Runger, 2007] Montgomery, D. and Runger, G. (2007). *Applied statistics and probability for engineers*. Wiley.
- [Teunissen et al., 2008] Teunissen, P., Simons, D., Tiberius, C., and Delft University of Technology, F. o. A. E. (2008). *Probability and Observation Theory;AE2-E01: Exercises*. TU Delft.
- [Teunissen, 2018] Teunissen, P. J. G. (2018). Distributional theory for the DIA method. *Journal of Geodesy*, 92(1):59–80.
- [Zaminpardaz and Teunissen, 2018] Zaminpardaz, S. and Teunissen, P. J. G. (2018). DIA-datasnooping and identifiability. *Journal of Geodesy*.
- [Zaminpardaz et al., 2019] Zaminpardaz, S., Teunissen, P. J. G., and Tiberius, C. C. J. M. (2019). Risking to underestimate the integrity risk. *GPS Solutions*, 23(2):29.