University of Granada

Department of Statistics and Operations Research

Doctoral Thesis

Doctoral Program in Mathematical and Applied Statistics

Inferences on the Weighted Kappa Coefficient of Binary Diagnostic Tests

Raid Amro

Granada

May, 2017
Inferences on the Weighted Kappa Coefficient of Binary Diagnostic Tests

Report on the DOCTORAL THESIS written under the supervision of José Antonio Roldán Nofuentes, Associate professor of the Department of Statistics and Operations Research of the University of Granada, presented by the Graduate Raid Amro to obtain a PhD in Statistics

Signed: Raid Amro

Approval by the Supervisor

Signed: José Antonio Roldán Nofuentes
The doctoral candidate Raid Amro and the thesis supervisor José Antonio Roldán Nofuentes

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Place and date:

Granada. May, 2017

Thesis supervisor: 

Signed: José Antonio Roldán Nofuentes

Doctoral candidate:

Signed: Raid Amro
Acknowledgements

To the memory of my late Father Mohammad.

I would like to place on record my deepest gratitude to my family, especially my mother, my wife Ikhlas and my children Mohammad, Ashraf, Gwan and Mays, who tolerated my absence while away from them in Spain for several months. Indeed, their moral support is highly appreciated.

I also would like to express sincere thanks to my advisor Professor Jose Antonio Roldán Nofuentes, for his constant and continuous support throughout my entire study for the Ph.D. His immense knowledge, guidance and inspiration have been invaluable. I could not have imagined having a better mentor.
Index

Summary
1. Binary Diagnostic Test and its Parameters................................. 1
   1.1. Introduction........................................................................ 1
   1.2. Sensitivity and specificity.................................................. 4
   1.3. Likelihood ratios.............................................................. 5
   1.4. Predictive values.............................................................. 7
   1.5. Weighted kappa coefficient................................................. 9
2. Estimation of the Parameters of a Binary Diagnostic Test............ 17
   2.1. Introduction....................................................................... 17
   2.2. Estimation subject to a cross-sectional study....................... 18
      2.2.1. Sensitivity and specificity.......................................... 19
         2.2.1.1. Clopper-Pearson CI............................................ 20
         2.2.1.2. Wilson score CI................................................ 20
         2.2.1.3. Agresti-Coull CI.............................................. 21
         2.2.1.4. Modified score CI............................................ 21
         2.2.1.5. Arcsine CI....................................................... 22
      2.2.2. Likelihood ratios...................................................... 23
         2.2.2.1. Gart-Nam CI................................................... 23
         2.2.2.2. Logarithmic CI................................................ 24
         2.2.2.3. Martín-Andrés and Álvarez-Hernández CI............. 25
      2.2.3. Predictive Values..................................................... 26
         2.2.3.1. Clopper-Pearson CI......................................... 27
         2.2.3.2. Wilson score CI............................................. 28
         2.2.3.3. Agresti-Coull CI.......................................... 28
         2.2.3.4. Modified score CI.......................................... 29
         2.2.3.5. Arcsine CI..................................................... 29
2.2.4. Weighted kappa coefficient................................. 30
  2.2.4.1. Wald CI............................................. 30
  2.2.4.2. Logit CI........................................... 31
  2.2.4.3. Bootstrap CI....................................... 32
2.3. Estimation subject to a case-control study.................... 32
  2.3.1. Sensitivity and specificity............................... 34
    2.3.1.1. Clopper-Pearson CI.................................. 34
    2.3.1.2. Wilson score CI.................................... 35
    2.3.1.3. Agresti-Coull CI................................... 35
    2.3.1.4. Modified score CI.................................. 36
    2.3.1.5. Arcsine CI......................................... 36
  2.3.2. Likelihood ratios.......................................... 37
    2.3.2.1. Gart-Nam CI....................................... 37
    2.3.2.2. Logarithmic CI..................................... 38
    2.3.2.3. Martín-Andrés and Álvarez-Hernández CI.......... 38
  2.3.3. Predictive values.......................................... 40
    2.3.3.1. Study of Mercaldo et al.................................. 41
    2.3.3.2. Study of Stamey and Holt.............................. 44
  2.3.4. Weighted kappa coefficient................................ 45
    2.3.4.1. Approximate CIs................................... 46
      2.3.4.1.1. Wald CI........................................ 47
      2.3.4.1.2. Logit CI........................................ 48
      2.3.4.1.3. Arcsine CI..................................... 49
      2.3.4.1.4. Bootstrap CI................................... 50
      2.3.4.1.5. Bayesian CI.................................... 51
    2.3.4.2. Simulation experiments................................ 54
    2.3.4.3. Sample sizes........................................ 63
    2.3.4.4. Example............................................ 66
    2.3.4.5. Discussion.......................................... 69
3. Comparison and Combination of Two Binary Diagnostic Tests... 73
  3.1 Introduction.................................................. 73
  3.2. Comparison of two BDTs...................................... 74
    3.2.1. Sensitivities and specificities............................ 76
    3.2.2. Likelihood ratios...................................... 78
3.2.3. Predictive values……………………………………… 81
  3.2.3.1. Study of Leisenring et al………………………… 83
  3.2.3.2. Study of Wang et al……………………………… 84
  3.2.3.3. Study of Kosinski………………………………… 84
  3.2.3.4. Study of Roldán-Nofuentes et al…………………. 86
  3.2.4. Weighted kappa coefficients……………………… 87
3.3. Combination of two BDTs……………………………… 89
  3.3.1. Study of Marshal……………………………………… 90
  3.3.2. Study of Lin………………………………………… 92
  3.3.3. Study of Macaskill et al…………………………… 92
  3.3.4. Combination of two weighted kappa coefficients…… 94
    3.3.4.1. Combined weighted kappa coefficient…………… 95
      3.3.4.1.1. Increase with the AND rule…………………… 98
      3.3.4.1.2. Increase with the OR rule…………………… 99
    3.3.4.2. Estimation and CIs………………………………… 101
    3.3.4.3. Simulation experiments…………………………… 104
    3.3.4.4. Program in R……………………………………… 109
    3.3.4.5. Example…………………………………………… 110
      3.3.4.5.1. AND rule………………………………………. 110
      3.3.4.5.2. OR rule………………………………………… 113
    3.3.4.6. Discussion………………………………………… 116
Conclusions……………………………………………………… 119
Appendices………………………………………………………… 123
  Appendix 2.1: Estimation of the variance………………….. 123
  Appendix 2.2: Comparison of the performance of the CIs… 125
  Appendix 3.1: Combined weighted kappa coefficient……… 126
  Appendix 3.2: The AND rule………………………………… 127
  Appendix 3.3: The OR rule…………………………………… 134
  Appendix 3.4: The program “cwkc”………………………… 141
References………………………………………………………… 157
Summary

Diagnostic Methods are fundamental in Clinical Medicine and in Epidemiology. Therefore, part of the discipline of Statistics has focused on the development of new methods to solve the problems that have been posed in this field, leading to what are known as Statistical Methods for Diagnosis in Medicine. This doctoral thesis seeks to contribute to research into new methods of estimation of parameters of binary diagnostic tests. It focuses on the study of binary diagnostic tests, whose assessment in relation to a gold standard gives rise to a $2 \times 2$ table when there is a single diagnostic test, or a $2 \times 4$ table when there are two binary diagnostic tests. In all the situations analysed in this Thesis, it is assumed that the disease status of all the individuals in the sample, or samples, is known. This doctoral thesis is structured in three Chapters.

In Chapter 1, the main parameters of a binary diagnostic test are defined and studied: sensitivity and specificity, likelihood ratios, predictive values and the weighted kappa coefficient.

Chapter 2 studies the estimations of the parameters presented in Chapter 1 when the study is cross-sectional and when it is case-control. The cross-sectional study consists of the application of the binary diagnostic test and the gold standard to all the individuals in a random sample; and the case-control study consists of applying the binary diagnostic test to all of the individuals in two samples, one of individuals with the disease (case sample) and another of individuals without the disease (control sample). The contribution made by this Chapter is the estimation of the weighted kappa coefficient subject to case-control sampling. Several confidence intervals are studied for this parameter, Monte Carlo simulation experiments are carried out to study the
asymptotic coverage of these intervals and a method is proposed to calculate the size of each sample. The results obtained are applied to real example.

Chapter 3 studies two different problems: the comparison of parameters of two binary diagnostic tests subject to a paired design and the combination of parameters of two binary diagnostic tests. On the one hand, we present the hypothesis tests and confidence intervals to compare the parameters of two binary diagnostic tests, and on the other, we study the combination of parameters of two binary diagnostic tests. The contribution of this Chapter is the combination of the weighted kappa coefficients of two binary diagnostic tests in parallel testing, defining the weighted kappa coefficient of the combination of the two diagnostic tests and studying its properties. We have studied the conditions in which the combination of the two diagnostic tests produces an increase in the weighted kappa coefficient of the combination. Fieller’s method is applied to obtain a confidence interval for the ratio between the weighted kappa coefficient of the combination and each weighted kappa coefficient, and Monte Carlo simulation experiments are carried out to study the asymptotic behaviour of this interval. An R program is written to solve the problem posed and the results were applied to a real example.

Granada

May, 2017
Chapter 1

Binary Diagnostic Test and its Parameters

1.1. Introduction

The diagnosis of diseases is fundamental in the practice of Medicine and the study of Statistical Methods for Diagnosis in Medicine is an important topic in Biostatistics (Zhou et al, 2002; Pepe, 2003). A diagnostic test (DT) is a medical test that is applied to an individual in order to determine the presence or absence of a certain disease. The level of PSA for the diagnosis of prostate cancer and a stress test for the diagnosis of coronary artery disease are two examples of DTs. A DT can be helpful for screening, for diagnosis and for individual management:
a) Screening: to detect a disease in asymptomatic individuals, and therefore reduce disease morbidity and mortality through early treatment.

b) Diagnosis: establish or exclude the presence of a disease in symptomatic individuals.

c) Individual management: evaluate the severity of a disease, estimate prognosis, monitor the course of a disease (progression, stability, or resolution), detect disease recurrence, select drugs and therapy and adjust them.

The application of a DT for the assessment of diseases has different purposes (Zhou et al, 2002):

a) To provide reliable information about the disease status of an individual.

b) To influence the planning of the treatment of an individual.

c) To understand the mechanism and the nature of the disease through research.

The interpretation of a DT depends on several factors (Zhou et al, 2002):

a) The intrinsic ability of the DT in order to distinguish between diseased and non-diseased individuals (discriminatory accuracy).

b) The particular characteristics of each individual.

c) The environment in which the DT is applied.

A DT may make a mistake in the diagnosis of the disease status of an individual, and therefore the accuracy of a DT is measured in terms of probabilities (or functions of them). When the result of a DT is positive (indicating the presence of the disease) or negative (indicating the absence of the disease), the DT is called a binary diagnostic test (BDT) and its accuracy is measured in terms of two fundamental parameters: sensitivity
and specificity. Other parameters to assess the performance of a BDT are the likelihood ratios (LRs), the predictive values (PVs) and the weighted kappa coefficient $\left( \kappa(c) \right)$. In order to obtain an unbiased estimators of the sensitivity and the specificity of the BDT, and therefore for the other parameters, it is necessary to know the true disease status (present or absent) of each individual in the random sample. The medical test through which the true disease status of each individual is known is called the gold standard (GS). A biopsy for the diagnosis of prostate cancer and an angiography for the diagnosis of coronary disease are two examples of GS. Therefore, there are two methods to diagnose the disease: DT and GS. The DT may make a mistake in the diagnosis of the disease while the GS does not. Why not always use the GS? There are several reasons to use a DT instead of a GS:

a) The DT is usually less expensive than the GS.

b) The use of a GS may lead to some risk for the individual. For example, a coronary angiography may cause a stroke, thrombosis or even death of the individual.

c) There is not always a GS. For example, in the case of some psychiatric diseases.

The objective is to estimate the accuracy of the BDT, not to know if the individual has the disease or not. In the following sections of this chapter we will study each of the parameters of a BDT:

a) Sensitivity and specificity.

b) Likelihood ratios.

c) Predictive values.

d) Weighted kappa coefficient.
1.2. Sensitivity and specificity

Sensitivity and specificity are the fundamental measures of accuracy of a BDT. Let $D$ be the random variable which models the result of the GS, so that $D = 1$ when an individual does have the disease and $D = 0$ when an individual does not have the disease. Let $T$ be the random variable which models the result of the BDT, in such a way that $T = 1$ when the result is positive (indicating the presence of the disease) and $T = 0$ when it is negative (indicating the absence of the disease). The probability of a randomly chosen individual from among all of the individuals in the population having the disease, denoted as $p = P(D = 1)$, is known as the disease prevalence. Sensitivity ($Se$) is the probability that the result of the BDT will be positive when an individual has the disease, i.e.

$$Se = P(T = 1|D = 1).$$

(1.1)

The probability $P(T = 0|D = 1) = 1 - Se$ is called the probability of a false negative. Specificity ($Sp$) is the probability that the result of the BDT will be negative when the individual does not have the disease, i.e.

$$Sp = P(T = 0|D = 0).$$

(1.2)

The probability $P(T = 1|D = 0) = 1 - Sp$ is called the probability of a false positive. Sensitivity and specificity only depend on the intrinsic ability of the BDT to distinguish between individuals who have the disease and those who do not, i.e. these parameters depend on the physical, chemical and biological bases upon which the BDT has been developed, and they are not affected by the prevalence of the disease. A BDT with high
Se is useful to exclude a diagnosis because a highly sensitive test will render few results that are falsely negative. A BDT with high Sp is useful to confirm a diagnosis, because a highly specific test will have few results that are falsely positive. Cicchetti et al. (1995) have classified the Se and the Sp at the following intervals: <70% “Poor”, 70%–79% “Fair”, 80%–89% “Good” and 90%–100% “Excellent”.

The parameter $Y = Se + Sp - 1$ is called Youden index (Youden, 1950). The Youden index is a summary measure of accuracy of a BDT. The Youden index does not depend on the prevalence of disease and it indicates the likelihood of a positive result among individuals with the disease versus those without the disease. The Youden index has the following property: if the BDT and the disease are independent, then $Se$ and $Sp$ are complementary ($Se + Sp = 1$). This aspect is not desirable in a BDT, and therefore it is demanded that the Youden index of a BDT be bigger than zero ($Y > 0$ or $Se + Sp > 1$). If $Y < 0$ the results of the diagnosis are interchanged, $T = 1$ should be a negative result and $T = 0$ should be a positive result, and the analysis should be limited only to the positive values of the Youden index.

1.3. Likelihood ratios

Other parameters for evaluating the performance of a BDT are the likelihood ratios ($LR$s). When the result of the BDT is positive, the $LR$ (called the positive likelihood ratio, $LR^+$) is the quotient between the probability of a positive result for the BDT when the individual has the disease and the probability of a positive result for the BDT when the individual does not have the disease, i.e.
\[ LR^+ = \frac{Se}{1 - Sp}. \] (1.3)

When the result of the BDT is negative, the \( LR \) (called the negative likelihood ratio, \( LR^- \)) is the quotient between the probability of a negative result of the BDT when the individual has the disease and the probability of a negative result of the BDT when the individual does not have the disease, i.e.

\[ LR^- = \frac{1 - Se}{Sp}. \] (1.4)

The \( LRs \) depend on the \( Se \) and \( Sp \) of the diagnostic test and do not depend on the disease prevalence, and their values vary between zero and infinite. When the result of the BDT and the disease status are independent, then \( LR^+ = LR^- = 1 \). When the BDT correctly classifies all of the individuals (diseased and non-diseased), then \( LR^+ = \infty \) and \( LR^- = 0 \). A value \( LR^+ > 1 \) indicates that a positive result of the BDT is more probable in a diseased individual than in a non-diseased individual, and a value \( LR^- < 1 \) indicates that a negative result of the BDT is more probable in a non-diseased individual than in a diseased individual. The \( LRs \) quantify the increase in knowledge of the presence of the disease through the application of the BDT. Before applying the BDT, the odds of a individual being diseased is

\[ \text{pre-test odds} = \frac{p}{1 - p}, \]

where \( p \) is the disease prevalence. After applying the BDT, the odds of disease are

\[ \text{post-test odds} = \frac{P(D = 1|T = j)}{P(D = 0|T = j)}, \quad j = 0,1. \]

The \( LRs \) relate the pre-test odds and post-test odds.
\[
\text{Post test odds (} T = 1 \text{)} = L R^+ \times \text{pre test odds}
\]
\[
\text{Post test odds (} T = 0 \text{)} = L R^- \times \text{pre test odds},
\]

Therefore, the likelihood ratios quantify the change in the odds of disease obtained by knowledge of the application of the BDT.

### 1.4. Predictive values

Predictive values (PVs) are the measures of the clinical accuracy of a BDT. When the result of the BDT is positive, the PV (called the positive predictive value, PPV) is the probability of an individual being diseased when the test result is positive, i.e.

\[
PPV = P(D = 1|T = 1) = \frac{p \times Se}{p \times Se + q \times (1 - Sp)}, \quad (1.5)
\]

where \( q = 1 - p \). When the result of the BDT is negative, the PV (called the negative predictive value, NPV) is the probability of an individual not being diseased when the test result is negative, i.e.

\[
NPV = P(D = 0|T = 0) = \frac{q \times Sp}{p \times (1 - Se) + q \times Sp}. \quad (1.6)
\]

While sensitivity and specificity quantify how well the BDT reflects the true disease status (present or absent), the PVs quantify the clinical value of the BDT, since both the clinic and the individual are more interested in knowing how probable it is to have the disease given a diagnostic test result. Cicchetti et al (1995) have classified the PPV and the NPV at the following intervals: \(<70\%\) “Poor”, \(70\%–79\%\) “Fair”, \(80\%–89\%\) “Good” and \(90\%–100\%\) “Excellent”. In Figures 1.1 and 1.2 we show how the PVs varies according to the prevalence \( p \) for different values of \( Se \) and \( Sp \).
Figure 1.1. Predictive values for $Se = 0.95$ and $Sp = 0.90$.

Figure 1.2. Predictive values for $Se = 0.80$ and $Sp = 0.95$. 

0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
NPV
PPV

0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
NPV
PPV
1.5. Weighted kappa coefficient

The weighted kappa coefficient (Kraemer, 1992; Kraemer et al, 2002) of a BDT is defined as a measure of the beyond-chance agreement between the BDT and the GS, and it is a parameter that considers the losses associated with an erroneous classification with the BDT. Let $L$ be the loss or the cost that occurs when wrongly classifying a diseased individual with the BDT, and $L'$ the loss or cost that occurs when wrongly classifying a non-diseased individual with the BDT. It is assumed that the losses $L$ and $L'$ are 0 when an individual (diseased or non-diseased) is classified correctly with the BDT. Table 1.1 shows the probabilities and the losses associated to the assessment of the BDT, where the random variables $T$ and $D$ are defined in Section 1.2. In terms of the probabilities of Table 1.1, the expected loss is

$$p(1-Se)L + q(1-Sp)L', \quad (1.7)$$

and the random loss is

$$p\{p(1-Se) + qSp\} L + q\{pSe + q(1-Sp)\} L'. \quad (1.8)$$

The expected loss given by equation (1.7) is described by Bloch (1998) as the risk of error and is interpreted as the average loss which occurs when the BDT erroneously classifies an individual (diseased or non-diseased) and its values vary between zero and infinite. The random loss given by equation (1.8) is the expected loss when the BDT and the GS are independent, i.e. it is the expected loss when $P(T = i\mid D = j) = P(T = i)$ with $i, j = 0,1$. The weighted kappa coefficient of a BDT is defined as

$$\kappa = \frac{\text{Random loss} - \text{Expected loss}}{\text{Random loss}} \quad (1.9)$$
Substituting in equation (1.9) each loss with its expression it holds that the weighted kappa coefficient of the BDT is

\[
\kappa(c) = \frac{pqy}{p(1-Q)c + qQ(1-c)},
\]

(1.10)

where \( Q = pSe + (1-p)(1-SP) \) is the probability that the result of the BDT will be positive, \( 1-Q = p(1-Se) + qSp \) is the probability that the result of the BDT will be negative, and \( c = L/(L+L') \) is the weighting index.

<table>
<thead>
<tr>
<th>Probabilities</th>
<th>( T = 1 )</th>
<th>( T = 0 )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D = 1 )</td>
<td>( p \times Se )</td>
<td>( p \times (1-Se) )</td>
<td>( p )</td>
</tr>
<tr>
<td>( D = 0 )</td>
<td>( q \times (1-Sp) )</td>
<td>( q \times Sp )</td>
<td>( q )</td>
</tr>
<tr>
<td>Total</td>
<td>( Q = p \times Se + q \times (1-Sp) )</td>
<td>( 1-Q = p \times (1-Se) + q \times Sp )</td>
<td>( 1 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Losses</th>
<th>( T = 1 )</th>
<th>( T = 0 )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D = 1 )</td>
<td>0</td>
<td>( L )</td>
<td>( L )</td>
</tr>
<tr>
<td>( D = 0 )</td>
<td>( L' )</td>
<td>0</td>
<td>( L' )</td>
</tr>
<tr>
<td>Total</td>
<td>( L' )</td>
<td>( L )</td>
<td>( L + L' )</td>
</tr>
</tbody>
</table>

When the loss \( L \) is zero then \( c = 0 \) and the weighted kappa coefficient is

\[
\kappa(0) = \frac{SP - (1-Q)}{Q} = \frac{PPV - p}{q},
\]

(1.11)
and when the loss $L'$ is zero then $c = 1$ and the weighted kappa coefficient is

$$\kappa(1) = \frac{Se - Q}{1 - Q} = \frac{NPV - q}{p}.$$  \hspace{1cm} (1.12)

The weighted kappa coefficient $\kappa(c)$ can be written in terms of $p$, $Q$, $\kappa(0)$ and $\kappa(1)$ as

$$\kappa(c) = \frac{p(1 - Q)c\kappa(1) + qQ(1 - c)\kappa(0)}{p(1 - Q)c + qQ(1 - c)},$$ \hspace{1cm} (1.13)

with $0 \leq c \leq 1$, and therefore the weighted kappa coefficient is a weighted average of $\kappa(0)$ and $\kappa(1)$. Index $c$ is between 0 and 1, and it is a clinician’s judgment of the relative clinical cost of false positives and false negatives. For example, let us consider a diagnosis of colon cancer using a colonography as BDT. If the colonography is positive for an individual who does not have the cancer (false positive), a colonoscopy (GS) will be performed on the individual, and this will give a final negative diagnosis. The loss $L'$ is determined from the economic costs of the diagnosis and also of factors such as risk, stress and anxiety caused to the individual. If the colonography is negative for an individual with the cancer (false negative), the individual may be diagnosed later. In this case the cancer may have spread and the possibility of the treatment being successful will be reduced. The loss $L$ is determined from these considerations. Hence, the losses $L$ and $L'$ are not only measured in economic terms but also by factors such as risk and stress, for which reason the value of these losses cannot be determined in clinical practice. This is why the relative discrepancy between the losses $L$ and $L'$ is replaced by the relative discrepancy between the false positives and the false negatives. The value of the weighting index $c = L/(L' + L)$ may be supposed, depending on the considerations
taken into account by the clinician about the false positives and the false negatives. If the clinician is going to use the BDT as a screening test, then there is more concern about false negatives and the $c$ index is greater than 0.5 ($0.5 < c \leq 1$). If the clinician is going to use the BDT as a first step towards an risk treatment, then there is more concern about false positives and the $c$ index is less than 0.5 ($0 \leq c < 0.5$). Therefore, the value of the $c$ index will depend on what are the clinical objectives for using the BDT. If in equation (1.13) $L = L'$, then $c = 0.5$ and $\kappa(0.5)$ is known as Cohen’s kappa coefficient; if $L > L'$ then $0.5 < c \leq 1$, and if $L' > L$ then $0 \leq c < 0.5$. For example, if the clinician considers that the false positives are three times more important than the false negatives, then $c = 1/(3+1) = 0.25$; and if the clinician considers that the false negatives are four times more important than the false positives then $c = 4/(1+4) = 0.8$.

The weighted kappa coefficient has the following properties:

a) When the agreement between the BDT and the GS is perfect ($Se = Sp = 1$), the expected loss is zero and then $\kappa(c) = 1$.

b) When $Se$ and $Sp$ are complementary ($Se = 1 - Sp$), i.e. when the diagnosis of the disease is made randomly, then $\kappa(c) = 0$.

c) If the random loss is greater than expected loss then $\kappa(c) > 0$, and if the expected loss is greater than random loss then $\kappa(c) < 0$. If $\kappa(c) < 0$ the results of the diagnosis are interchanged, $T = 1$ should be a negative result and $T = 0$ should be a positive result, and the analysis should be limited only to the positive values of the weighted kappa coefficient ($0 \leq \kappa(c) \leq 1$).
d) The weighted kappa coefficient is a function of the $c$ index which may be
increasing (if $Q > p$) or decreasing (if $Q < p$), or it can be a constant function
which is equal to the Youden index $(Se + Sp - 1)$ if $Q = p$.

Figures 1.3 and 1.4 show how the weighted kappa coefficient varies according to the
weighting index $c$ and the disease prevalence $p$, for high values of $Se$ and $Sp$.
Coefficient $\kappa(c)$ is an increasing or decreasing function, depending on $p$, in the
weighting index $c$. 
Figure 1.3. Weighted kappa coefficients for $Se = 0.95$, $Sp = 0.90$ and different values of prevalence $p$. 

$0.95 Se = 0$, $0.90 Sp = 0$ and different values of prevalence $p$. 

- $p = 0.05$
- $p = 0.10$
- $p = 0.30$
- $p = 0.50$
- $p = 0.70$
- $p = 0.90$
Figure 1.4. Weighted kappa coefficients for $Se = 0.90$, $Sp = 0.95$ and different values of prevalence $p$. 
Chapter 2

Estimation of the Parameters of a Binary Diagnostic Test

2.1. Introduction

In this Chapter we study the estimation of the parameters of a BDT studied in Chapter 2 subject to two types of study: cross-sectional study and case-control study. The estimation of the weighted kappa coefficient subject to a case-control study is the first contribution to this Doctoral Thesis. For the other parameters, different methods of estimation are reviewed. The advantages and disadvantages of case-control design over the cross-sectional can be seen in the book of Pepe (2003). Summarizing, case-control study has some advantages over the cross-sectional study:
a) Case-control study is more efficient in terms of sample size requirements.

b) Case-control study allow for the exploration of subject-related characteristics on the test.

Nevertheless, the case-control study has the disadvantage that by using it we cannot estimate the prevalence of the disease $p$. Therefore, if in a case-control study we wish to estimate parameters that depend on the prevalence, it is necessary to have a value for the prevalence, e.g. an estimation obtained from health surveys, other studies, etc.

First, we study the estimation subject to a cross-sectional study, and secondly we study the estimation subject to a case-control study.

### 2.2. Estimation subject to a cross-sectional study

The estimation of the parameters of a $BDT$ in relation to a $GS$ subject to a cross-sectional study consists of applying the $BDT$ and the $GS$ to all of the individuals in a random sample sized $n$, giving rise to Table 2.1.

<table>
<thead>
<tr>
<th></th>
<th>$T = 1$</th>
<th>$T = 0$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 1$</td>
<td>$s_1$</td>
<td>$s_0$</td>
<td>$s$</td>
</tr>
<tr>
<td>$D = 0$</td>
<td>$r_1$</td>
<td>$r_0$</td>
<td>$r$</td>
</tr>
<tr>
<td>Total</td>
<td>$s_1 + r_1$</td>
<td>$s_0 + r_0$</td>
<td>$n$</td>
</tr>
</tbody>
</table>
2.2.1. Sensitivity and specificity

Conditioning in variable $D$, the samples $(s_i, s_0)$ and $(r_i, r_0)$ are two independent samples, and it is verified that the observed frequency $s_i$ is the product of a binomial distribution $B(s, S_e)$, and the frequency $r_0$ is the product of a binomial distribution $B(r, S_p)$. Therefore, the estimators of $S_e$ and $S_p$ are the estimators of binomial proportions, i.e.

$$\hat{S}_e = \frac{s_i}{s} \quad \text{and} \quad \hat{S}_p = \frac{r_0}{r},$$

(2.1)

and the estimators of their variances are

$$\hat{\text{Var}}(\hat{S}_e) = \frac{\hat{S}_e(1 - \hat{S}_e)}{s} \quad \text{and} \quad \hat{\text{Var}}(\hat{S}_p) = \frac{\hat{S}_p(1 - \hat{S}_p)}{r}.$$  

(2.2)

Subject to a cross-sectional study, the estimator of the prevalence of the disease $p$ is

$$\hat{p} = \frac{s}{n}.$$  

(2.3)

The estimation of a binomial proportion has been the object of many studies. We present five confidence intervals (CIs) for $S_e$ and $S_p$: Clopper-Pearsson, Wilson, Agresti-Coull, Yu et al, and the arcsine CI of Martín-Andrés and Álvarez-Hernández. The first CI is an exact interval and the other CIs are approximate intervals that have a good asymptotic behaviour.
Chapter 2. Estimation of the parameters of a BDT

2.2.1.1. Clopper-Pearson CI

Clopper and Pearson (1934) studied an exact CI for a binomial proportion. The 100(1−α)% Clopper-Pearson CI for Se is

\[
\left( \frac{s_1}{s_1 + (s_0 + 1) F_{\alpha/2} \left( 2 \left( s_0 + 1 \right) ; 2 s_1 \right) }, \quad \frac{(s_1 + 1) F_{\alpha/2} \left( 2 \left( s_1 + 1 \right) ; 2 s_0 \right)}{s_0 + (s_1 + 1) F_{\alpha/2} \left( 2 \left( s_1 + 1 \right) ; 2 s_0 \right)} \right),
\]

(2.4)

and for Sp the CI is

\[
\left( \frac{r_0}{r_0 + (r_1 + 1) F_{\alpha/2} \left( 2 \left( r_1 + 1 \right), 2 r_0 \right) }, \quad \frac{(r_0 + 1) F_{\alpha/2} \left( 2 \left( r_0 + 1 \right), 2 r_1 \right)}{r_1 + (r_0 + 1) F_{\alpha/2} \left( 2 \left( r_0 + 1 \right), 2 r_1 \right)} \right),
\]

(2.5)

where \( F_{\alpha/2} (v_1, v_2) \) is the \( \alpha/2 \) quantile from an \( F \)-distribution with \( v_1 \) and \( v_2 \) degrees of freedom.

2.2.1.2. Wilson score CI

Wilson (1927) proposed, using the approximation to the normal distribution, an approximate CI for a binomial proportion. The Wilson CI is sometimes called the Wilson score CI. The 100(1−α)% Wilson CI for Se is

\[
\frac{s}{s + z_{1-\alpha/2}^2} \left( \hat{Se} + \frac{z_{1-\alpha/2}^2}{2s} \sqrt{\frac{\hat{Se} \left( 1 - \hat{Se} \right)}{s} + \frac{z_{1-\alpha/2}^2}{4s^2}} \right),
\]

(2.6)

and for Sp is

\[
\frac{r}{r + z_{1-\alpha/2}^2} \left( \hat{Sp} + \frac{z_{1-\alpha/2}^2}{2r} \sqrt{\frac{\hat{Sp} \left( 1 - \hat{Sp} \right)}{r} + \frac{z_{1-\alpha/2}^2}{4r^2}} \right),
\]

(2.7)
where \( z_{1-\alpha/2} \) is the \( 100(1-\alpha/2)\% \) percentile of the standard normal distribution.

### 2.2.1.3. Agresti-Coull CI

The CI of Agresti and Coull (1998) is another approximate CI for a binomial proportion. The \( 100(1-\alpha)\% \) Agresti and Coull CI for \( Se \) is

\[
\hat{Se} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Se}(1-\hat{Se})}{s + z_{1-\alpha/2}^2}},
\]

(2.8)

where \( \hat{Se} = \frac{s + z_{1-\alpha/2}^2/2}{s + z_{1-\alpha/2}^2} \) is the adjusted estimator of \( Se \). The \( 100(1-\alpha)\% \) Agresti and Coull CI for \( Sp \) is

\[
\hat{Sp} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Sp}(1-\hat{Sp})}{r + z_{1-\alpha/2}^2}},
\]

(2.9)

where \( \hat{Sp} = \frac{r + z_{1-\alpha/2}^2/2}{r + z_{1-\alpha/2}^2} \) is the adjusted estimator of \( Sp \).

### 2.2.1.4. Modified score CI

Yu et al (2014) have proposed a CI, called modified score interval, based on a modification of the midpoint of the Wilson CI. The \( 100(1-\alpha)\% \) modified score CI for \( Se \) is
\begin{equation}
0.5 + \frac{53}{s + z_{1-\alpha/2}} \left( \hat{S}e - 0.5 \right) \pm \frac{z_{1-\alpha/2}}{s + z_{1-\alpha/2}} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{s_1s_0}{s}}.
\end{equation}

(2.10)

Similarly, the \(100(1 - \alpha)\)% modified score CI for \(S_p\) is

\begin{equation}
0.5 + \frac{53}{r + z_{1-\alpha/2}} \left( \hat{S}p - 0.5 \right) \pm \frac{z_{1-\alpha/2}}{r + z_{1-\alpha/2}} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{r_0r_1}{r}}.
\end{equation}

(2.11)

\subsection{2.2.1.5. Arcsine CI}

Martín-Andrés and Álvarez-Hernández (2014a) evaluated 29 approximate CIs (excluding the Yu et al CI) for a binomial proportions, recommending using the arcsine CI with continuity correction. The \(100(1 - \alpha)\)% arcsine CI with continuity correction for \(S_e\) is

\begin{equation}
\sin^2 \left( \sin^{-1} \sqrt{\frac{s_1 + 0.5}{s + 1}} \pm \frac{z_{1-\alpha/2}}{\sqrt{4(s+1)}} \right).
\end{equation}

(2.12)

and for \(S_p\)

\begin{equation}
\sin^2 \left( \sin^{-1} \sqrt{\frac{r_0 + 0.5}{r + 1}} \pm \frac{z_{1-\alpha/2}}{\sqrt{4(r+1)}} \right).
\end{equation}

(2.13)

Martín-Andrés and Álvarez-Hernández (2016) have compared the performance of the 29 CIs studied, recommending:

a) For \(s \leq 80\) \((r \leq 80)\) and \(\alpha = 1\%\) or \(\alpha = 5\%\), to use the modified score CI of Yu et al (2014).
b) For \( s \geq 100 \ (r \geq 100) \) and \( \alpha = 10\% \), to use the arcsine CI with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a).

c) In other situations, to use the CI of Agresti and Coull (1998).

### 2.2.2. Likelihood ratios

The estimators of the likelihood ratios are

\[
\hat{LR}^+ = \frac{\hat{Se}}{1 - \hat{Sp}} = \frac{s_1 \cdot r}{r_1 \cdot s} \quad \text{and} \quad \hat{LR}^- = \frac{1 - \hat{Se}}{\hat{Sp}} = \frac{s_0 \cdot r}{r_0 \cdot s}.
\]  

(2.14)

Applying the delta method, the estimated asymptotic variances are

\[
\hat{\text{Var}}(\hat{LR}^+) = \frac{\hat{LR}^+}{(1 - \hat{Sp})} \left( \frac{1 - \hat{Se}}{s} + \frac{\hat{Sp}}{r} \hat{LR}^+ \right) \quad \text{and} \quad \hat{\text{Var}}(\hat{LR}^-) = \frac{\hat{LR}^-}{\hat{Sp}} \left( \frac{\hat{Se}}{s} + \frac{1 - \hat{Sp}}{r} \hat{LR}^- \right).
\]  

(2.15)

The LR*s are the ratio of the two independent binomial proportions, so that the LR*s can be estimated applying methods to estimate the ratio of two independent binomial proportions. We now present several of these methods.

#### 2.2.2.1. Gart-Nam CI

The Gart and Nam (1988) CI is a classic method to estimate the ratio of two independent binomial proportions. The \( 100(1 - \alpha)\% \) CI of Gart and Nam for \( LR^* \) is obtained by solving the equation
\[
\begin{align*}
    s_i - LR^* \left(1 - \tilde{S}p\right) s & = \pm z_{1-a/2}, \\
    \left\{1 - LR^* \left(1 - \tilde{S}p\right)\right\} & \sqrt{\frac{1}{\tilde{S}p} \frac{1 - LR^* \left(1 - \tilde{S}p\right)}{r \left(1 - \tilde{S}p\right) + \frac{1 - LR^* \tilde{S}p}{s LR^* \tilde{S}p}}}
\end{align*}
\]

where \( \tilde{S}p \) is the appropriate solution for the quadratic equation

\[
LR^* \left(1 - \tilde{S}p\right)^2 n - \{(r_s + s) LR^* + s_i + r_i\} + s_i + r_i = 0.
\]

In a similar way, the \( 100(1-\alpha)\% \) CI for \( LR^- \) is obtained by solving the equation

\[
\begin{align*}
    s_0 - s LR^- \tilde{S}p & = \pm z_{1-a/2}, \\
    \left\{1 - LR^- \tilde{S}p\right\} & \sqrt{\frac{1}{\tilde{S}p} \frac{1 - LR^- \tilde{S}p}{r \tilde{S}p + \frac{1 - LR^- \tilde{S}p}{s LR^- \tilde{S}p}}}
\end{align*}
\]

where \( \tilde{S}p \) is the appropriate solution for the quadratic equation

\[
nLR^- \tilde{S}p^2 - \{(r_0 + s) LR^- + s_0 + r_0\} \tilde{S}p + s_0 + r_0 = 0.
\]

### 2.2.2.2. Logarithmic CI

The \( LR \)s have a non-symmetrical distribution and therefore the napierian logarithm of the \( LR \)s has a more symmetrical distribution which is closer to the normal distribution.

Thus, based on the asymptotic normality of the napierian logarithms of the \( \hat{LR} \)s, Simel et al (1991) proposed the following CIs for \( LR \)s,

\[
LR^* \in \exp\left\{ \ln(\hat{LR}^*) \pm z_{1-a/2} \sqrt{\frac{s_0 + r_0}{s \hat{S}r}} \right\}
\]

and

\[
\begin{align*}
    s_0 - s LR^- \tilde{S}p & = \pm z_{1-a/2}, \\
    \left\{1 - LR^- \tilde{S}p\right\} & \sqrt{\frac{1}{\tilde{S}p} \frac{1 - LR^- \tilde{S}p}{r \tilde{S}p + \frac{1 - LR^- \tilde{S}p}{s LR^- \tilde{S}p}}}
\end{align*}
\]
\[ LR^- \in \exp \left\{ \ln(\hat{LR}^-) \pm \frac{z_{1-\alpha/2}}{2} \sqrt{\frac{s_1 + r_1}{s_0 r_0}} \right\}. \]  
(2.19)

2.2.2.3. Martín-Andrés and Álvarez-Hernández CI

Martín-Andrés and Álvarez-Hernández (2014b) studied CIs for the ratio of two independent binomial proportions. The 100(1 − α)% CI for the LRs are

\[ LR^+ \in \frac{n's'_i r'_i + \frac{z_{1-\alpha/2}^2}{2} \left(s's'_i + r'r'_i - 2s'_i r'_i\right)}{r'_i \left(n's'_i - \frac{z_{1-\alpha/2}^2}{4} \left(s's'_i - r'r'_i\right)\right)} \pm \frac{z_{1-\alpha/2}}{2} \sqrt{\frac{n^2 s'_i r'_i \left(s'_i + r'_i - n'\hat{p}'_1 \hat{p}'_2\right) + \frac{z_{1-\alpha/2}^2}{4} \left(s'_i r'_i + r'_i - r'_i\right) \left(s'_i - r'_i\right)\right)^2}{r'_i \left(n's'_i - \frac{z_{1-\alpha/2}^2}{4} \left(s'_i - r'_i\right)\right)}}, \]  
(2.20)

and

\[ LR^- \in \frac{n's'_i r'_i + \frac{z_{1-\alpha/2}^2}{2} \left(2s'_i r'_i - 2s'_i r'_i\right)}{r'_i \left(n's'_i - \frac{z_{1-\alpha/2}^2}{4} \left(s'_i - r'_i\right)\right)} \pm \frac{z_{1-\alpha/2}}{2} \sqrt{\frac{2n^2 s'_i r'_i \left(s'_i + r'_i - n'\hat{p}'_1 \hat{p}'_2\right) + \frac{z_{1-\alpha/2}^2}{4} \left(s'_i r'_i + r'_i - r'_i\right) \left(s'_i - r'_i\right)\right)^2}{r'_i \left(n's'_i - \frac{z_{1-\alpha/2}^2}{4} \left(s'_i - r'_i\right)\right)}}, \]  
(2.21)

where \( s'_i = s_i + 0.5 \), \( r'_i = r_i + 0.5 \), \( s' = s'_i + s'_0 \), \( r' = r'_i + r'_0 \), \( n' = s' + r' \), \( \hat{p}'_i = \frac{r'_i}{s'_i} \), \( \hat{p}'_2 = s'_i / s' \), \( \hat{p}'_3 = r'_i / r' \) and \( \hat{p}'_3 = s'_0 / s' \). If the upper limit of the interval for the \( LR^+ \) is lower than \( s'_i / (n' - r'_i) \) or higher than \( \hat{LR}^+ \), then the upper limit of the interval is

\[ \frac{1}{s' \left(\hat{p}'_i\right)^2 + \frac{z_{1-\alpha/2}^2}{2} - \frac{z_{1-\alpha/2}^2}{4} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{s'_i s'_i}{s'_i} \left(\hat{p}'_1 - \hat{p}'_2\right)\right)}}, \]

and the upper limit of this interval is higher than \( \left(n' - s'_i\right) / r'_i \) or lower than \( \hat{LR}^+ \), then the upper limit of the CI is
Chapter 2. Estimation of the parameters of a BDT

\[
\frac{1}{r'(\hat{p}_i)} \left\{ r'_i \hat{p}_i^2 + \frac{z_{1-a/2}^2}{2} + z_{1-a/2} \sqrt{\frac{z_{1-a/2}^2}{4} + r'_i (\hat{p}_i - \hat{p}_i')} \right\}.
\]

Regarding the CI for \( LR^- \), if the upper limit of this CI is lower than \( s_0'/(n' - r_0') \) or higher than \( \hat{LR}^- \), then the lower limit is

\[
\frac{1}{s' (\hat{p}_i')^2 + z_{1-a/2}^2} \left\{ s_0' \hat{p}_i' + \frac{z_{1-a/2}^2}{2} - z_{1-a/2} \sqrt{\frac{z_{1-a/2}^2}{4} + s_0' (\hat{p}_i' - \hat{p}_i')} \right\},
\]

and if the upper limit of this interval is higher than \( (n' - s_0')/r_0' \) or lower than \( \hat{LR}^- \), then the upper limit is

\[
\frac{1}{r'(\hat{p}_i')^2} \left\{ r'_0 \hat{p}_i' + \frac{z_{1-a/2}^2}{2} + z_{1-a/2} \sqrt{\frac{z_{1-a/2}^2}{4} + r'_0 (\hat{p}_i' - \hat{p}_i')} \right\}.
\]

The CI proposed by Martín-Andrés and Álvarez-Hernández is the interval that has a better asymptotic coverage.

2.2.3. Predictive Values

Conditioning in variable \( T \), the samples \((s, r)\) and \((s_0, r_0)\) are two independent samples. It is verified that the frequency \( s_1 \) is the product of a binomial distribution \( B(s_1 + r_1, PPV) \) and that the observed frequency \( r_0 \) is the realization of a binomial distribution \( B(s_0 + r_0, NPV) \). Therefore, the estimators of the \( PVs \) are the estimators of binomial proportions, i.e.
Chapter 2. Estimation of the parameters of a BDT

\[ \hat{PPV} = \frac{s_1}{s_1 + r_1} \] and \[ \hat{NPV} = \frac{r_0}{s_0 + r_0}, \]

(2.22)

and the estimators of their variances are

\[ \hat{\text{Var}}(\hat{PPV}) = \frac{\hat{PPV}(1 - \hat{PPV})}{s_1 + r_1} \] and \[ \hat{\text{Var}}(\hat{NPV}) = \frac{\hat{NPV}(1 - \hat{NPV})}{s_0 + r_0}. \]

(2.23)

The estimation through CIs of the PVs can be carried out by applying the same CIs that for a binomial proportion.

2.2.3.1. Clopper-Pearson CI

The 100(1−α)% Clopper-Pearson CI for PPV is

\[ \left\{ \frac{s_1}{s_1 + (r_1 + 1) F_{\alpha/2}(2(s_1 + 1); 2r_1)}, \frac{(s_1 + 1) F_{\alpha/2}(2(s_1 + 1); 2r_1)}{r_1 + (s_1 + 1) F_{\alpha/2}(2(s_1 + 1); 2r_1)} \right\}, \]

(2.24)

and for NPV the CI is

\[ \left\{ \frac{r_0}{r_0 + (s_0 + 1) F_{\alpha/2}(2(s_0 + 1); 2r_0)}, \frac{(r_0 + 1) F_{\alpha/2}(2(s_0 + 1); 2r_0)}{s_0 + (r_0 + 1) F_{\alpha/2}(2(r_0 + 1); 2s_0)} \right\}, \]

(2.25)

where \( F_{\alpha/2}(v_1, v_2) \) is the \( \alpha/2 \) quantile from an F-distribution with \( v_1 \) and \( v_2 \) degrees of freedom.
2.2.3.2. Wilson score CI

The $100(1-\alpha)\%$ Wilson CI for PPV is

$$\frac{s_1 + r_1}{s_1 + r_1 + z_{1-\alpha/2}^2} \left( \hat{P}_{PV} + \frac{z_{1-\alpha/2}^2}{2(s_1 + r_1)} \pm \frac{z_{1-\alpha/2}}{4(s_1 + r_1)^2} \sqrt{\frac{\hat{P}_{PV}(1 - \hat{P}_{PV})}{s_1 + r_1}} \right)$$

and for NPV is

$$\frac{s_0 + r_0}{s_0 + r_0 + z_{1-\alpha/2}^2} \left( \hat{N}_{PV} + \frac{z_{1-\alpha/2}^2}{2(s_0 + r_0)} \pm \frac{z_{1-\alpha/2}}{4(s_0 + r_0)^2} \sqrt{\frac{\hat{N}_{PV}(1 - \hat{N}_{PV})}{s_0 + r_0}} \right)$$

2.2.3.3. Agresti-Coull CI

The $100(1-\alpha)\%$ Agresti and Coull CI for the PPV

$$\hat{P}_{PV} \pm \frac{1}{2} \left( 1 - \hat{P}_{PV} \right) \left( \frac{z_{1-\alpha/2}}{s_1 + r_1 + z_{1-\alpha/2}^2} \right)$$

where $\hat{P}_{PV} = \frac{s_1 + z_{1-\alpha/2}^2/2}{s_1 + r_1 + z_{1-\alpha/2}^2}$ is the adjusted estimator of PPV. The $100(1-\alpha)\%$ Agresti and Coull CI for the NPV

$$\hat{N}_{PV} \pm \frac{1}{2} \left( 1 - \hat{N}_{PV} \right) \left( \frac{z_{1-\alpha/2}}{s_0 + r_0 + z_{1-\alpha/2}^2} \right)$$

where $\hat{N}_{PV} = \frac{r_0 + z_{1-\alpha/2}^2/2}{s_0 + r_0 + z_{1-\alpha/2}^2}$ is the adjusted estimator of NPV.
2.2.3.4 Modified score CI

The \(100(1-\alpha)\%\) modified score CI of Yu et al for PPV is

\[
0.5 + \frac{s_i + r_i}{s_i + r_i + z_{1-\alpha/2}^2} \left( \frac{s_i + r_i}{s_i + r_i + z_{1-\alpha/2}^2} \right) \pm \frac{z_{1-\alpha/2}}{s_i + r_i + z_{1-\alpha/2}^2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{s_i r_i}{s_i + r_i}},
\]

and the \(100(1-\alpha)\%\) modified score CI for NPV is

\[
0.5 + \frac{s_0 + r_0}{s_0 + r_0 + z_{1-\alpha/2}^2} \left( \frac{s_0 + r_0}{s_0 + r_0 + z_{1-\alpha/2}^2} \right) \pm \frac{z_{1-\alpha/2}}{s_0 + r_0 + z_{1-\alpha/2}^2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{s_0 r_0}{s_0 + r_0}}.
\]

2.2.3.5 Arcsine CI

The \(100(1-\alpha)\%\) arcsine CI of Martín-Andrés and Álvarez-Hernández (2014a) for the PPV is

\[
\sin^2 \left( \sin^{-1} \left( \frac{s_i + 0.5}{s_i + r_i + 1} \pm \frac{z_{1-\alpha/2}}{\sqrt{4(s_i + r_i + 1)}} \right) \right),
\]

and for the NPV

\[
\sin^2 \left( \sin^{-1} \left( \frac{r_0 + 0.5}{r_0 + 1} \pm \frac{z_{1-\alpha/2}}{\sqrt{4(r_0 + 1)}} \right) \right).
\]

Martín-Andrés and Álvarez-Hernández (2016) recommend:

a) For \(s_i + r_i \leq 80\) \(s_0 + r_0 \leq 80\) and \(\alpha = 1\%\) or \(\alpha = 5\%\), to use the modified score CI of Yu et al (2014).
b) For \( s_i + r_i \geq 100 \ (s_o + r_o \geq 100) \) and \( \alpha = 10\% \), to use the arcsine CI with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a).

c) In other situations, to use the CI of Agresti and Coull (1998).

2.2.4. Weighted kappa coefficient

Substituting in equation (1.10), or (1.13), the parameters \( Se, Sp \) and \( p \) with their estimators given by the equations (2.1) and (2.3) respectively, the estimator of \( \kappa(c) \) is

\[
\hat{\kappa}(c) = \frac{s_o r_o - s_o r_i}{r(s_i + r_i)(1 - c) + s(s_o + r_o)c},
\]

with \( 0 \leq c \leq 1 \). Applying the delta method, the estimated asymptotic variance of \( \hat{\kappa}(c) \) is

\[
\hat{\text{Var}}(\hat{\kappa}(c)) = \frac{n r}{s \left[ n^2 (1 - c) r_i + n (c r_o - 2(1 - c) r_i) s_o + n (r_o - (1 - c) r_i) s_i + s (s_o r_i - s_o r_o) \right]^4} \]

\[
\left\{ (s_o r_i - s_o r_o) \left[ 2(1 - c) r_i n s - (1 - c) r_i n^2 + s (c r_o + 2s_o r_i + s_i r_i) - r_i s \right]^2 + s_i s_o n r^3 [ (1 - c) r_i n + s (c r_i) ]^3 + r_i n^3 s r^3 \left[ s_i r + c (s^2 - s_i n) \right]^2 \right\}.
\]

Roldán-Nofuentes et al (2009) have studied different CIs for \( \kappa(c) \): Wald CI, Logit CI and bootstrap CI.

2.2.4.1. Wald CI

Assuming the asymptotic normality of \( \hat{\kappa}(c) \), the \( 100(1 - \alpha)\% \) CI for \( \kappa(c) \) is

\[
\hat{\kappa}(c) \pm z_{1-\alpha/2} \sqrt{\text{Var}\left[ \hat{\kappa}(c) \right]}.
\]

(2.36)
This CI performs well for relatively small samples \((n = 100)\).

### 2.2.4.2. Logit CI

Assuming the asymptotic normality of \(\hat{\kappa}(c)\), the logit transformation of \(\hat{\kappa}(c)\),
\[
\text{logit}[\hat{\kappa}(c)] = \log[\hat{\kappa}(c)/(1-\hat{\kappa}(c))],
\]
is closer to a normal distribution with mean \(\text{logit}[\kappa(c)]\). The \(100(1-\alpha)\%\) CI for the \(\text{logit}[\kappa(c)]\) is
\[
\text{logit}[\hat{\kappa}(c)] \pm z_{1-\alpha/2} \sqrt{\text{Var}[\text{logit}(\hat{\kappa}(c))]},
\]
where
\[
\hat{\text{Var}}[\text{logit}(\hat{\kappa}(c))] = \frac{1}{s(\hat{r}_s - c r_n - c(s_0 \hat{r}_0 + 2s_1 \hat{r}_1 + s_1 \hat{r}_2))^2} \times
\]
\[
\left\{ s_1 s_2 r_s^2 \left[ (1-c) r_n + (c r - r_s) s \right]^2 + r_s r_s \left[ s \mu - s \kappa + c(s^2 - s_n) \right]^2 \right\} + \frac{2(1-c) r_n s - (1-c) r_n^2 + s \left( c(s_0 \hat{r}_0 + 2s_1 \hat{r}_1 + s_1 \hat{r}_2) - r_s \right)}{n s r}.
\]

Finally, the logit CI for \(\kappa(c)\) is
\[
\text{expit}\left\{\text{logit}[\hat{\kappa}(c)] \pm z_{1-\alpha/2} \sqrt{\text{Var}[\text{logit}(\hat{\kappa}(c))]},\right\},
\]
where \(\text{expit}(\cdot) = \exp(\cdot) / [1 + \exp(\cdot)]\) is the inverse of logit. This CI performs well for samples of 200 or more.
2.2.4.3. Bootstrap CI

The bootstrap CI is calculated generating $K$ samples with replacement from the sample. The coefficient $\kappa(c)$ is estimate from each sample with replacement, and the bootstrap estimator of $\kappa(c)$ is estimated as the average of the $K$ estimated $\hat{\kappa}_i(c)$, i.e.

$$\hat{\kappa}_b(c) = \frac{1}{K} \sum_{i=1}^{K} \hat{\kappa}_i(c),$$

and its variance is estimated as

$$\frac{\sum_{k=1}^{K} [\hat{\kappa}_i(c) - \hat{\kappa}_b(c)]^2}{(K-1)}. $$

Then, the bias-corrected bootstrap CI (Efron and Tibshirani, 1993) is calculated. Let $A = \#(\hat{\kappa}_i(c) < \hat{\kappa}(c))$ be the number of bootstrap estimators that are lower than the estimator, and let $\hat{z}_0 = \Phi^{-1}(A/K)$, where $\Phi^{-1}(\cdot)$ is the inverse function of the standard normal cumulative distribution function. Let $\alpha_1 = \Phi\left(2\hat{z}_0 - z_{1-\alpha/2}\right)$ and $\alpha_2 = \Phi\left(2\hat{z}_0 + z_{1-\alpha/2}\right)$, where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the normal standard distribution, then the bias-corrected bootstrap CI is

$$\left(\hat{\kappa}^{(\alpha_1)}(c), \hat{\kappa}^{(\alpha_2)}(c)\right), \quad (2.39)$$

where $\hat{\kappa}^{(\alpha_j)}(c)$ is the $j$th quantile of the distribution of the $K$ bootstrap estimations of $\kappa(c)$. In general, the performance of the bootstrap CI is similar to that of the Wald CI.

2.3. Estimation subject to a case-control study

The estimation of the parameters of a $BDT$ subject to a case-control study consists of applying the $BDT$ to two random independent samples, one of $n_1$ individuals who have the disease (sample of case) and another of $n_0$ individuals who do not have the disease
(sample of control). Let us suppose that of the $n_1$ individuals who have the disease, in $n_{10}$ the BDT gives a negative result and in the rest ($n_{11} = n_1 - n_{10}$) the BDT gives a positive result. In the same way, let us suppose that of the $n_0$ individuals without the disease, in $n_{00}$ the BDT gives a negative result and in the rest ($n_{01} = n_0 - n_{00}$) the BDT gives a positive result. The results are summarized in Table 2.2. The sample of individuals that have the disease is extracted from a population of individuals that have the disease (e.g. registers of diseases), and the control sample is extracted from a population of individuals who are known not to have the disease.

<table>
<thead>
<tr>
<th></th>
<th>$T = 1$</th>
<th>$T = 0$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>$n_{11}$</td>
<td>$n_{10}$</td>
<td>$n_1$</td>
</tr>
<tr>
<td>Control</td>
<td>$n_{01}$</td>
<td>$n_{00}$</td>
<td>$n_0$</td>
</tr>
</tbody>
</table>

In this situation, it is verified that the observed frequency $n_{11}$ is the product of a binomial distribution $B(n_1, Se)$ and the observed frequency $n_{00}$ is the product of a binomial distribution $B(n_0, Sp)$. In a case-control study, the quotient $n_1/(n_1 + n_0)$ is not an estimator of the disease prevalence because the sample sizes $n_1$ and $n_0$ are set by the researcher. Therefore, it is not possible to estimate the prevalence of the disease from a case-control study.
2.3.1. Sensitivity and specificity

The estimators of $Se$ and $Sp$ are the estimators of binomial proportions, i.e.

$$\hat{Se} = \frac{n_{11}}{n_1} \quad \text{and} \quad \hat{Sp} = \frac{n_{00}}{n_0}, \quad (2.40)$$

and the estimators of their variances are

$$\hat{Var}(\hat{Se}) = \frac{\hat{Se}(1-\hat{Se})}{n_1} \quad \text{and} \quad \hat{Var}(\hat{Sp}) = \frac{\hat{Sp}(1-\hat{Sp})}{n_0}. \quad (2.41)$$

The CIs for $Se$ and $Sp$ are the same that in Section 2.2.1 but replacing $s_1$ by $n_{11}$, $s_0$ by $n_{10}$, $r_1$ by $n_{01}$, $r_0$ by $n_{00}$, $s$ by $n_1$ and $r$ by $n_0$.

2.3.1.1. Clopper-Pearson CI

The $100(1-\alpha)\%$ Clopper-Pearson CI (1934) for $Se$ is

$$\left( \frac{n_{11}}{n_{11} + (n_{10} + 1) F_{a/2} \left( 2(n_{10} + 1); 2n_{11} \right) }, \quad \frac{(n_{11} + 1) F_{a/2} \left( 2(n_{11} + 1); 2n_{10} \right) }{n_{10} + (n_{11} + 1) F_{a/2} \left( 2(n_{11} + 1); 2n_{10} \right) } \right), \quad (2.42)$$

and for $Sp$ the CI is

$$\left( \frac{n_{00}}{n_{00} + (n_{01} + 1) F_{a/2} \left( 2(n_{01} + 1); 2n_{00} \right) }, \quad \frac{(n_{00} + 1) F_{a/2} \left( 2(n_{00} + 1); 2n_{00} \right) }{n_{01} + (n_{00} + 1) F_{a/2} \left( 2(n_{00} + 1); 2n_{00} \right) } \right). \quad (2.43)$$
2.3.1.2. Wilson score CI

The $100(1-\alpha)\%$ Wilson CI (1927) for $Se$ is

$$\frac{n_1}{n_1 + z_{1-\alpha/2}^2} \left( \hat{Se} + \frac{z_{1-\alpha/2}^2}{2n_1} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Se} (1-\hat{Se})}{n_1} + \frac{z_{1-\alpha/2}^2}{4n_1^2}} \right), \quad (2.44)$$

and for $Sp$

$$\frac{n_0}{n_0 + z_{1-\alpha/2}^2} \left( \hat{Sp} + \frac{z_{1-\alpha/2}^2}{2n_0} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Sp} (1-\hat{Sp})}{n_0} + \frac{z_{1-\alpha/2}^2}{4n_0^2}} \right), \quad (2.45)$$

2.3.1.3. Agresti-Coull CI

The $100(1-\alpha)\%$ Agresti and Coull (1998) CI for $Se$ is

$$\hat{Se} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Se} (1-\hat{Se})}{n_1 + z_{1-\alpha/2}^2}}, \quad (2.46)$$

where $\hat{Se} = \frac{n_1 + z_{1-\alpha/2}^2/2}{n_1 + z_{1-\alpha/2}^2}$ is the adjusted estimator of $Se$. The CI for $Sp$ is

$$\hat{Sp} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{Sp} (1-\hat{Sp})}{n_0 + z_{1-\alpha/2}^2}}, \quad (2.47)$$

where $\hat{Sp} = \frac{n_0 + z_{1-\alpha/2}^2/2}{n_0 + z_{1-\alpha/2}^2}$ is the adjusted estimator of $Sp$. 
2.3.1.4. Modified score CI

The $100(1-\alpha)\%$ modified score CI of Yu et al (2014) for $Se$ is

$$0.5 + \frac{53}{n_1 + z_{1-\alpha/2}^2} \left( \hat{Se} - 0.5 \right) \pm \frac{z_{1-\alpha/2}}{n_1 + z_{1-\alpha/2}^2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{n_1 n_{10}}{n_1}},$$

(2.48)

and for $Sp$

$$0.5 + \frac{53}{n_0 + z_{1-\alpha/2}^2} \left( \hat{Sp} - 0.5 \right) \pm \frac{z_{1-\alpha/2}}{n_0 + z_{1-\alpha/2}^2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + \frac{n_0 n_{00}}{n_0}}.$$

(2.49)

2.3.1.5. Arcsine CI

The $100(1-\alpha)\%$ arcsine CI with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a) for $Se$ is

$$\sin^2 \left\{ \sin^{-1} \left( \frac{n_{11} + 0.5}{\sqrt{n_1 + 1}} \right) \pm \frac{z_{1-\alpha/2}}{\sqrt{4(n_1 + 1)}} \right\},$$

(2.50)

and for $Sp$

$$\sin^2 \left\{ \sin^{-1} \left( \frac{n_{00} + 0.5}{\sqrt{n_0 + 1}} \right) \pm \frac{z_{1-\alpha/2}}{\sqrt{4(n_0 + 1)}} \right\}.$$

(2.51)

The recommendations are similar to the previous case (Martín-Andrés and Álvarez-Hernández, 2016):

d) For $n_1 \leq 80 \, (n_0 \leq 80)$ and $\alpha = 1\% \, \text{or} \, \alpha = 5\%$, to use the modified score CI of Yu et al (2014).
Chapter 2. Estimation of the parameters of a BDT

e) For \( n_1 \geq 100 \) (\( n_0 \geq 100 \)) and \( \alpha = 10\% \), to use the arcsine CI with continuity correction of Martín-Andrés and Álvarez-Hernández (2014a).

f) In other situations, to use the CI of Agresti and Coull (1998).

2.3.2. Likelihood ratios

The estimators of the likelihood ratios are

\[ \hat{LR}^+ = \frac{\hat{Se}}{1-\hat{Sp}} = \frac{n_1n_{11}}{n_{11}n_{11}} \quad \text{and} \quad \hat{LR}^- = \frac{1-\hat{Se}}{\hat{Sp}} = \frac{n_0n_{00}}{n_{00}n_{00}}, \quad (2.52) \]

and applying the delta method the estimated asymptotic variances are

\[ \hat{Var}(\hat{LR}^+) = \frac{\hat{LR}^+}{(1-\hat{Sp})}\left(1-\frac{1}{n_1}\hat{Sp} + \frac{\hat{Sp}}{n_0}LR^+\right) \quad \text{and} \quad \hat{Var}(\hat{LR}^-) = \frac{\hat{LR}^-}{\hat{Sp}}\left(1-\frac{1}{n_1}\hat{Sp} - \frac{1}{n_0}\hat{Sp}\hat{LR}^-\right). \quad (2.53) \]

The CIs for LRs are the same that in Section 2.2.2.

2.3.2.1. Gart-Nam CI

The \( 100(1-\alpha)\% \) CI of Gart and Nam (1988) for \( LR^+ \) is obtained by solving the equation

\[ \frac{n_{11} - LR^+(1-\hat{Sp})n_1}{\left\{1-LR^+(1-\hat{Sp})\right\}} \sqrt{\frac{1}{\hat{Sp}} + \frac{1-\hat{Sp}}{LR^+(1-\hat{Sp})n_1}} = \pm \hat{z}_{1-\alpha/2}, \quad (2.54) \]

where \( \hat{Sp} \) is the appropriate solution for the quadratic equation

37
Chapter 2. Estimation of the parameters of a BDT

\[
LR^* \left(1 - \tilde{Sp} \right)^2 \left( n_i + n_0 \right) - \left\{ \left( n_{00} + n_i \right) LR^* + n_{10} + n_0 \right\} \left(1 - \tilde{Sp} \right) + n_{11} + n_{01} = 0.
\]

The 100(1 - \alpha)\% CI for \( LR^- \) is obtained by solving the equation

\[
\frac{n_{10} - n_i LR^- \tilde{Sp}}{\left(1 - LR^- \tilde{Sp}\right) \sqrt{\frac{1}{n_0 \tilde{Sp}} + \frac{1}{n_i LR^- \tilde{Sp}}}} = \pm z_{1-\alpha/2}, \quad (2.55)
\]

where \( \tilde{Sp} \) is the appropriate solution for the quadratic equation

\[
(n_i + n_0) LR^- \tilde{Sp}^2 - \left\{ \left( n_{00} + n_i \right) LR^- + n_{10} + n_0 \right\} \tilde{Sp} + n_{11} + n_{00} = 0.
\]

2.3.2.2. Logarithmic CI

The 100(1 - \alpha)\% logarithmic CI (Simel et al, 1991) for LRs are

\[
LR^* \in \exp \left\{ \ln \left( \hat{LR}^* \right) \pm z_{1-\alpha/2} \sqrt{n_{10} + n_{00}} \right\} \quad (2.56)
\]

and

\[
LR^- \in \exp \left\{ \ln \left( \hat{LR}^- \right) \pm z_{1-\alpha/2} \sqrt{n_{10} + n_{00}} \right\}. \quad (2.57)
\]

2.3.2.3. Martín-Andrés and Álvarez-Hernández CI

The 100(1 - \alpha)\% CI of Martín-Andrés and Álvarez-Hernández (2014b) for the LRs are

\[
LR^* \in \frac{n_i' n_o' + \frac{z_{1-\alpha/2}^2}{2} \left( n_i' n_o' + n_i' n_o' - 2 n_i' n_o' \right) \pm z_{1-\alpha/2} \sqrt{n_i'^2 n_o'^2 \left(n_i' + n_o'ight) \left(n_i' - n_o' \right) \left(n_i' - n_o' \right)}}{n_i' \left[n_i' n_o' - \frac{z_{1-\alpha/2}^2}{4} \left(n_i' - n_o' \right)^2 \right]} \quad (2.58)
\]
and
\[
LR^+ = \frac{n'n_0n_0' - \frac{z_{1-a/2}^2}{2}(n'n_0 + n_0'n_0' - 2n'n_0') \pm z_{1-a/2} \sqrt{n'n_0n_0'(n_0 + n_0' - n'n_0')} + \frac{z_{1-a/2}^2}{4}(n'n_0 + n_0'n_0')}{n_0\{n'n'_1 - z_{1-a/2} (n_n' - n_0')\}}, \quad (2.59)
\]

where \( n'_i = n_i + 0.5 \), \( n'_0 = n_0 + 0.5 \), \( n'_1 = n'_{11} + n'_{10} \), \( n'_0 = n'_{01} + n'_{00} \), \( n' = n'_1 + n'_0 \), \( \hat{p}'_i = n'_0/n'_0 \), \( \hat{p}'_2 = n'_{11}/n'_1 \), \( \hat{p}'_3 = n'_{00}/n'_0 \) and \( \hat{p}'_4 = n'_{10}/n'_i \). If the upper limit of the interval for the \( LR^+ \) is lower than \( n'_{11}/(n' - n'_{00}) \) or higher than \( \hat{L}R^+ \), then the upper limit of the interval is
\[
\frac{1}{n'_i(\hat{p}'_i)^2 + z_{1-a/2}^2} \left\{ n'_i\hat{p}'_i + \frac{z_{1-a/2}^2}{2} - z_{1-a/2} \sqrt{\frac{z_{1-a/2}^2}{4} + n'_i(\hat{p}'_i - \hat{p}'_2)} \right\},
\]
and the upper limit of this interval is higher than \( (n' - n'_{11})/n'_{01} \) or lower than \( \hat{L}R^+ \), then the upper limit of the \( CI \) is
\[
\frac{1}{n'_0(\hat{p}'_2)^2 + z_{1-a/2}^2} \left\{ n'_0\hat{p}'_2 + \frac{z_{1-a/2}^2}{2} + z_{1-a/2} \sqrt{\frac{z_{1-a/2}^2}{4} + n'_0(\hat{p}'_2 - \hat{p}'_1)} \right\}.
\]

Regarding the \( CI \) for \( LR^- \), if the upper limit of this \( CI \) is lower than \( n'_{10}/(n' - n'_{00}) \) or higher than \( \hat{L}R^- \), then the lower limit is
\[
\frac{1}{n'_i(\hat{p}'_3)^2 + z_{1-a/2}^2} \left\{ n'_i\hat{p}'_3 + \frac{z_{1-a/2}^2}{2} - z_{1-a/2} \sqrt{\frac{z_{1-a/2}^2}{4} + n'_i(\hat{p}'_3 - \hat{p}'_4)} \right\},
\]
and if the upper limit of this interval is higher than \( (n' - n'_{10})/n'_{00} \) or lower than \( \hat{L}R^- \), then the upper limit is
Chapter 2. Estimation of the parameters of a BDT

\[
\frac{1}{n_0' \hat{p}_i} \left\{ n_0' \hat{p}_i^2 + \frac{z^2_{1-\alpha/2}}{2} + z_{1-\alpha/2} \sqrt{\frac{z^2_{1-\alpha/2}}{4} + n_0' \left( \hat{p}_i - \hat{p}_0 \right)} \right\}.
\]

The CI proposed by Martín-Andrés and Álvarez-Hernández is the interval that has a better asymptotic coverage.

2.3.3. Predictive values

The PVs of a BDT depend on the Se, Sp and disease prevalence \( p \). As the prevalence cannot be estimated from a case-control study, since the quotient \( n_i \left( n_i + n_0 \right) \) is not an estimator of the prevalence, the PVs cannot be estimated form a case-control study. Therefore, in order to estimate the PVs it is necessary to know an estimation of the prevalence of the disease. From now on it is assumed that we have an estimation of the prevalence of the disease. This value can be obtained from other clinical studies, health surveys, etc. If \( \hat{p} \) is an estimation of the disease prevalence, then the estimators of the PVs are

\[
\hat{PPV} = \frac{p \hat{p}_i n_i n_0}{p \hat{p}_i n_i n_0 + q n_0 n_1^2} \quad \text{and} \quad \hat{NPV} = \frac{q n_0 n_i n_0}{p \hat{p}_i n_i n_0 + q n_0 n_1^2},
\]

(2.60)

where \( q = 1 - p \). Different CIs have been studied for the PVs subject to a case-control study. Mercaldo et al (2007) have studied four frequentist CIs for the PVs and Stamey and Holt (2010) have studied two Bayesian CIs.
2.3.3.1. Study of Mercaldo et al

Mercaldo et al. (2007) have studied four CIs for the PVs subject to a case-control study, a Wald type CI and a logit CI, in both cases using the classical estimators and the adjusted estimators of Se and Sp. The $100(1-\alpha)\%$ Wald CIs for the PVs are

$$PPV \in P\hat{PV} \pm z_{1-\alpha/2} \sqrt{\hat{Var}(P\hat{PV})} \quad \text{and} \quad NPV \in N\hat{PV} \pm z_{1-\alpha/2} \sqrt{\hat{Var}(N\hat{PV})}, \quad (2.61)$$

where

$$\hat{Var}(P\hat{PV}) = \left[ pq\gamma^2 \hat{Se}(1-\hat{Sp}) \right] \frac{n_1}{n_0} + \left[ pq\hat{Sp}(1-\hat{Sp}) \right] \frac{n_0}{n_0} \left[ q\hat{Se} + p(1-\hat{Sp}) \right]^4$$

and

$$\hat{Var}(N\hat{PV}) = \left[ pq\hat{Sp}\gamma^2 \hat{Se}(1-\hat{Sp}) \right] \frac{n_1}{n_0} + \left[ pq(1-\hat{Se}) \right] \frac{n_0}{n_0} \left[ p(1-\hat{Se}) + q\hat{Sp} \right]^4. \quad (2.63)$$

The $100(1-\alpha)\%$ Wald CIs for the PVs with the adjusted estimates of Se and Sp are

$$PPV \in P\hat{PV} \pm z_{1-\alpha/2} \sqrt{\hat{Var}(P\hat{PV})} \quad \text{and} \quad NPV \in N\hat{PV} \pm z_{1-\alpha/2} \sqrt{\hat{Var}(N\hat{PV})}, \quad (2.64)$$

where

$$P\hat{PV} = \frac{p\hat{Se}}{p\hat{Se} + q(1-\hat{Sp})} \quad \text{and} \quad N\hat{PV} = \frac{q\hat{Sp}}{p(1-\hat{Se}) + q\hat{Sp}}, \quad (2.65)$$

$$\hat{Se} = \frac{n_1\hat{Se} + z_{1-\alpha/2}^2}{\hat{n}_1} \quad \text{and} \quad \hat{Sp} = \frac{n_2\hat{Sp} + z_{1-\alpha/2}^2}{\hat{n}_2}, \quad (2.66)$$
with \( \bar{n}_i = n_i + z_{1-\alpha/2}^2 \), and

\[
\hat{\text{Var}}(P\hat{PV}) = \frac{\left[ pq(1 - \hat{S}p) \right]^2 \hat{S}e(1 - \hat{S}e) + \left( pq\hat{S}e \right)^2 \hat{S}p(1 - \hat{S}p)}{[p\hat{S}e + q(1 - \hat{S}p)]^4} \tag{2.67}
\]

and

\[
\hat{\text{Var}}(N\hat{PV}) = \frac{\left( pq\hat{S}p \right)^2 \hat{S}e(1 - \hat{S}e) + \left[ pq(1 - \hat{S}e) \right]^2 \hat{S}p(1 - \hat{S}p)}{[p(1 - \hat{S}e) + q\hat{S}p]^4}. \tag{2.68}
\]

On the other hand, the \( 100(1-\alpha)\% \) CI for the logit of the PVs are

\[
\text{logit}(PPV) \in \text{logit}(P\hat{PV}) \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}[\text{logit}(P\hat{PV})]}
\]

\[
(2.69)
\]

and

\[
\text{logit}(NPV) \in \text{logit}(N\hat{PV}) \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}[\text{logit}(N\hat{PV})]},
\]

\[
(2.70)
\]

where

\[
\text{logit}(PPV) = \log \left( \frac{PPV}{1 - PPV} \right) = \log \left( \frac{p\hat{S}e}{q(1 - \hat{S}p)} \right)
\]

and

\[
\text{logit}(NPV) = \log \left( \frac{NPV}{1 - NPV} \right) = \log \left( \frac{q\hat{S}p}{p(1 - \hat{S}e)} \right).
\]

Finally, the \( 100(1-\alpha)\% \) logit CI for the PVs are
Chapter 2. Estimation of the parameters of a BDT

\[ VPP \in \expit\left(\logit\left(\hat{P}_{PV}\right) \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}\left[\logit\left(\hat{P}_{PV}\right)\right]} \right) \]  \hspace{1cm} (2.71)

and

\[ NPV \in \expit\left(\logit\left(\hat{N}_{PV}\right) \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}\left[\logit\left(\hat{N}_{PV}\right)\right]} \right), \]  \hspace{1cm} (2.72)

where \( \expit(\cdot) = \exp(\cdot)/(1 + \exp(\cdot)) \) is the inverse of logit. The variances, obtained applying the delta method, are

\[ \hat{\text{Var}}\left[\logit\left(\hat{P}_{PV}\right)\right] = \frac{1-\hat{S}_e}{n_1 \hat{S}_e} + \frac{\hat{S}_p}{n_0 (1-\hat{S}_p)} \]  \hspace{1cm} (2.73)

and

\[ \hat{\text{Var}}\left[\logit\left(\hat{N}_{PV}\right)\right] = \frac{\hat{S}_e}{n_1 (1-\hat{S}_e)} + \frac{1-\hat{S}_p}{n_0 \hat{S}_p}. \]  \hspace{1cm} (2.74)

With the adjusted estimates of \( Se \) and \( Sp \), the logit CIs are

\[ VPP \in \expit\left(\logit\left(\hat{P}_{PV}\right) \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}\left[\logit\left(\hat{P}_{PV}\right)\right]} \right) \]  \hspace{1cm} (2.75)

and

\[ NPV \in \expit\left(\logit\left(\hat{N}_{PV}\right) \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}\left[\logit\left(\hat{N}_{PV}\right)\right]} \right), \]  \hspace{1cm} (2.76)

and where the variances are

\[ \hat{\text{Var}}\left[\logit\left(\hat{P}_{PV}\right)\right] = \frac{1-\hat{S}_e}{\tilde{n}_1 \hat{S}_e} + \frac{\hat{S}_p}{\tilde{n}_0 (1-\hat{S}_p)} \]  \hspace{1cm} (2.77)

and
\[
\hat{\text{Var}}[\logit(NPV)] = \frac{\hat{\text{Se}}}{\hat{n}_1 (1 - \hat{\text{Se}})} + \frac{1 - \hat{\text{Sp}}}{\hat{n}_0 \hat{\text{Sp}}}.
\] (2.78)

Mercaldo et al have compared the asymptotic coverage of these four CIs, and in general terms, they recommend to apply the logit CI not adjusted, except when the estimator of a PV is equal to 1 in which case they recommend to apply the adjusted logit CI.

### 2.3.3.2. Study of Stamey and Holt

Stamey and Holt (2010) have studied two Bayesian CIs for the PVs subject to a case-control study: a CI without assuming distribution for the prevalence, and another CI assuming a probability distribution for the prevalence. Stamey and Holt have also shown that the second CI has a better asymptotic behaviour than the first CI. A summary of this second CI is presented below.

The observed frequencies \(n_{11}\) and \(n_{00}\) are the product of binomial distributions, i.e.

\[
n_{11} \to B(n,Se) \text{ and } n_{00} \to B(n,Sp).
\] (2.79)

For \(Se\) and \(Sp\), conjugate beta prior distributions are proposed, i.e.

\[
\hat{\text{Se}} \to \text{Beta}(\alpha_1, \beta_1) \text{ and } \hat{\text{Sp}} \to \text{Beta}(\alpha_2, \beta_2),
\] (2.80)

and therefore, their posterior distributions are

\[
\hat{\text{Se}}|d \to \text{Beta}(x_{11} + \alpha_1, n_1 - x_{11} + \beta_1) \text{ and } \hat{\text{Sp}}|d \to \text{Beta}(x_{00} + \alpha_2, n_0 - x_{00} + \beta_2),
\] (2.81)

And for the prevalence, the model is

\[
p \to \text{Beta}(\alpha_3, \beta_3) \text{ and } y \to B(n,p),
\] (2.82)
with posterior distribution

\[ p \mid d \rightarrow \text{Beta}(y + \alpha_3, n - y + \beta_3), \quad (2.83) \]

where \( d = (x_{11}, x_{00}, y, n_1, n_0, n) \), and the data \((y, n)\) are obtained from a survey or other study. Then, posterior distributions for \( PVs \) is approximated via Monte Carlo sampling by simulating \( B \) values from the posteriors distributions (2.81) and (2.83). In each iteration, the generated values of \( Se, Sp \) and \( p \) are plugged in to the equations

\[
PPV^{(j)} = \frac{p^{(j)}Se^{(j)}}{p^{(j)}Se^{(j)} + q^{(j)}(1 - Sp^{(j)})} \quad \text{and} \quad NPV^{(j)} = \frac{q^{(j)}Sp^{(j)}}{p^{(j)}(1 - Se^{(j)}) + q^{(j)}Sp^{(j)}},
\]

with \( q^{(j)} = 1 - p^{(j)} \). Finally, from the \( B \) values of \( PVs \), \( CIs \) based on the quantiles are calculated for \( PPV \) and \( NPV \). Simulation experiments have shown that the Bayesian \( CI \) has a better asymptotic behaviour than the Mercaldo et al \( CI \).

### 2.3.4. Weighted kappa coefficient

Subject to case-control study, Jannarone et al (1987) and Kraemer and Bloch (1990) have studied the point estimation of the weighted kappa coefficient, assuming that we have an estimation of the prevalence obtained from another study. Roldán-Nofuentes et al (2009) have studied various \( CIs \) (Wald, logit and bootstrap) for the weighted kappa coefficient subject to a cross-sectional study. The results obtained by Roldán-Nofuentes et al (2009), summarized in Section 2.2.4, are not valid in a case-control study, because, as has been pointed out previously, the quotient \( n_i / (n_i + n_o) \) it is not an estimator of the prevalence of the disease. The studies by Jannarone et al (1987) and Kraemer and Bloch (1990) focused on studying the point estimation of the weighted kappa coefficient and
Chapter 2. Estimation of the parameters of a BDT

deducing the variance of the estimator. Therefore, it is necessary to explore this topic in more depth, study the estimation by CIs and study a method to determine the sample sizes necessary to carry out the study.

In this Section, two objectives are studied: firstly, different approximate CIs for the weighted kappa coefficient subject to case-control study, and secondly we propose a method to determine the sample sizes (case sample and control sample) necessary to estimate the weighted kappa coefficient. For both objectives, it is necessary to know a value of the prevalence of the disease. In Section 2.3.4.1, different CIs are proposed to estimate this parameter subject to a case-control study. In Section 2.3.4.2, simulation experiments are carried out to study the asymptotic coverage of these CIs, giving some general rules of application. In Section 2.3.4.3, we present a method to determine the (case and control) sample sizes to estimate the weighted kappa coefficient with the precision required. In Section 2.3.4.4, the results are applied to a real example, and in Section 2.3.5.5 the results obtained are discussed.

2.3.4.1. Approximate CIs

The weighted kappa coefficient of a BDT is

\[ \kappa(c) = \frac{pq(Se + Sp - 1)}{p[p(1-Se) + qSp]c + p[pSe + q(1-Sp)](1-c)}, \]  

(2.85)

where the only unknown parameters are the \(Se\) and the \(Sp\), since it is assumed that we have an estimator \(p\) of the disease prevalence. Substituting in this equation (2.85) each parameter with its estimator, the estimator of the weighted kappa coefficient is
\[ \hat{\kappa}(c) = \frac{pq(n_{11}n_{00} - n_{10}n_{01})}{p(p_{10}n_0 + qn_{00})c + q(p_{11}n_0 + qn_{01})}(1-c), \]  

(2.86)

where \( q = 1 - p \). Applying the delta method the estimation of the variance of \( \hat{\kappa}(c) \) is

\[ \hat{\text{Var}}[\hat{\kappa}(c)] = \left( \frac{\hat{\kappa}(c)}{pq} \right)^2 \times \left[ \frac{pq - \hat{\kappa}(c)\{p(q-c)\}}{n_{11}} + \frac{pq + \hat{\kappa}(c)\{q(c-q)\}}{n_{00}} \right]. \]  

(2.87)

The demonstration can be seen in Appendix 2.1. We will now propose several CIs for the weighted kappa coefficient.

2.3.4.1.1. Wald CI

The Wald CI for a parameter is well known. Assuming that we know an estimation of the prevalence of the disease and that the researcher has set a value of the weighting index \( c \), the estimator \( \hat{\kappa}(c) \) is a function of \( \hat{Se} \) and of \( \hat{Sp} \). This situation is a particular case which is analysed by Koch et al (1977). Applying the results of Koch et al (1977), the distribution of \( \hat{\kappa}(c) \) is asymptotically normal, i.e. \( \hat{\kappa}(c) \overset{d}{\rightarrow} N(\kappa(c), \text{Var}[\hat{\kappa}(c)]) \).

Based on the asymptotic normality, the 100(1-\( \alpha \))% CI for \( \kappa(c) \) is

\[ \hat{\kappa}(c) \pm z_{1-\alpha/2} \sqrt{\text{Var}[\hat{\kappa}(c)]} \]  

(2.88)

where \( z_{1-\alpha/2} \) is the 100(1-\( \alpha /2 \))th percentile of the standard normal distribution.

This CI can also be calculated from the observed frequencies increased by a certain quantity, such as 0.5, 2 or \( z_{1-\alpha/2}^2 / 2 \). This procedure is very frequent in the analysis of
2×2 tables and in the estimation of binomial proportions (for example, Anscombe, 1956; Agresti and Coull, 1998; Martín-Andrés and Álvarez-Hernández, 2014a) or combinations of them (Martín-Andrés and Álvarez-Hernández, 2014b). These corrections aim to improve the coverages of the CIs, especially when the samples are small. Let us \( m_{ij} = n_{ij} + h \) and \( m_{0j} = n_{0j} + h \), with \( h \) equal to 0.5, 2 or \( z_{1-\alpha/2}^2/2 \), and \( j = 0,1 \). In this situation, the adjusted estimators of the Se and the Sp are \( \tilde{Se} = m_{11}/m_i \) and \( \tilde{Sp} = m_{00}/m_0 \), and the estimated variances are \( \hat{Var}(\tilde{Se}) = m_{00} m_{11}/m_i^3 \) and \( \hat{Var}(\tilde{Sp}) = m_{00} m_{01}/m_0^3 \), with \( m_i = n_i + 2h \). Substituting in equations (2.86) and (2.87) each parameter with its adjusted estimator and each frequency \( n_{ij} \) and \( n_i \) with \( m_{ij} \) and \( m_i \) we obtain the adjusted estimators of the weighted kappa coefficient \( \hat{\kappa}(c) \) and the adjusted Wald CIs. For \( \alpha = 5\% \) it holds that \( h = 2 \) is approximately equal to the case \( h = z_{1-\alpha/2}^2/2 = 1.96^2/2 \). For \( h = 0 \), the expressions (2.86), (2.87) and (2.88) are obtained.

2.3.4.1.2. Logit CI

As the value of \( \kappa(c) \) is between 0 and 1, the logit transformation can be applied. Assuming the asymptotic normality of the logit of \( \hat{\kappa}(c) \), i.e.

\[
\logit[\hat{\kappa}(c)] \xrightarrow{d} N\left(\logit[\kappa(c)], \text{Var}\left[\logit[\hat{\kappa}(c)]\right]\right),
\]

the 100(1−\( \alpha \))% CI for \( \logit[\kappa(c)] \) is

\[
\logit[\hat{\kappa}(c)] \pm z_{1-\alpha/2} \sqrt{\text{Var}\left[\logit[\hat{\kappa}(c)]\right]},
\]

Applying the delta method it is obtained that
The arcsine transformation is a transformation that has been used to estimate a binomial proportion (Brown et al., 2001; Martín-Andrés and Álvarez-Hernández, 2014a) and is a transformation that stabilizes the variance. Performing the transformation $F[\kappa(c)] = \arcsin(\sqrt{\kappa(c)})$ and assuming the asymptotic normality of this transformation, i.e. $\arcsin(\sqrt{\hat{\kappa}(c)}) \xrightarrow{d} N\left(\arcsin(\sqrt{\kappa(c)}), \text{Var}\left[\arcsin(\sqrt{\kappa(c)})\right]\right)$, the approximate CI for $F[\kappa(c)]$ is

$$F[\hat{\kappa}(c)] \pm z_{1-\alpha/2} \sqrt{\text{Var}\left[F[\hat{\kappa}(c)]\right]},$$

(2.92)

where the variance is estimated applying the delta method and whose expression is
\[
\hat{\text{Var}} \{ F[\hat{\kappa}(c)] \} = \\
\frac{-pq}{4\hat{\mathcal{Y}} \left\{ \left( q + p\hat{\mathcal{Y}} - \hat{S}p \right) - q \left( 1 - \hat{S}p \right) \right\} \left[ c \left( 1 - \hat{S}p + p(\hat{\mathcal{Y}} - 1) \right) - q \left( 1 - \hat{S}p + p\hat{\mathcal{Y}} \right) \right]^2} \\
\left[ \frac{n_{00}n_{11} \left\{ q \left( 1 - \hat{S}p \right) - c \left( \hat{S}p - q \right) \right\}^2}{n_i^3} + \frac{n_{00}n_{01} \left\{ c \left( p - \hat{S}e \right) + q\hat{S}e \right\}^2}{n_0^3} \right]. 
\] (2.93)

As the inverse function of \( \arcsin(\kappa(c)) \) is \( \sin^2 \), then the CI for \( \kappa(c) \) is
\[
\sin^2 \left[ F[\hat{\kappa}(c)] \pm z_{\alpha/2} \sqrt{\hat{\text{Var}} \{ F[\hat{\kappa}(c)] \}} \right]. \tag{2.94}
\]
As in the case of the logit CI, the arcsine CI can be calculated adding to the observed frequencies the quantity \( h \), and the process is similar to that of the previous case.

### 2.3.4.1.4. Bootstrap CI

The bootstrap CI is calculated generating \( K \) samples with replacement from the case sample and another \( K \) samples with replacement based on the control sample. From each case sample with replacement we estimate \( \hat{S}e \); and from each control sample with replacement we estimate \( \hat{S}p \); then we estimate the weighted kappa coefficient as
\[
\hat{\kappa}_i(c) = \frac{pq \left( \hat{S}e + \hat{S}p - 1 \right)}{p \left( 1 - \hat{Q}_i \right) c + q\hat{Q}_i \left( 1 - c \right)}, \tag{2.95}
\]
with \( i = 1, \ldots, K \). The bootstrap estimator of the weighted kappa coefficient is estimated as the average of the \( K \) estimated weighted kappa coefficients, i.e.
\[
\hat{K}_b(c) = \frac{\sum_{i=1}^{K} \hat{\kappa}_i(c)}{K},
\]
(2.96)
and its variance is estimated as 
\[
\left(\sum_{i=1}^{K} \left[ \hat{\kappa}_i(c) - \hat{K}_b(c) \right]^2 \right) /(K-1).
\]
We then calculate the bias-corrected bootstrap CI (Efron and Tibshirani, 1993) as will now be shown. Let
\[
A = \# \left( \hat{\kappa}_{ci} < \hat{\kappa}_c \right)
\]
be the number of bootstrap estimators that are lower than the estimator, and let \( \hat{z}_0 = \Phi^{-1} \left( A/K \right) \), where \( \Phi^{-1} \) is the inverse function of the standard normal cumulative distribution function. Let \( \alpha_1 = \Phi \left( 2\hat{z}_0 - z_{1-\alpha/2} \right) \) and \( \alpha_2 = \Phi \left( 2\hat{z}_0 + z_{1-\alpha/2} \right) \), where \( z_{1-\alpha/2} \) is the \( 100(1-\alpha/2) \)th percentile of the normal standard distribution, then the bias-corrected bootstrap CI is
\[
\left( \hat{\kappa}_i^{[\alpha_1]}(c), \hat{\kappa}_i^{[\alpha_2]}(c) \right),
\]
(2.97)
where \( \hat{\kappa}_i^{[\alpha]}(c) \) is the \( j \)th quantile of the distribution of the \( K \) bootstrap estimations of \( \kappa(c) \).

2.3.4.1.5. Bayesian CI

The CIs proposed in the previous sections are frequentist and are based on the asymptotic normality of the maximum likelihood estimator of \( \kappa(c) \) (and of its logit and arcsine transformations). The problem is now approached from a Bayesian perspective. From the case sample it is obtained that \( n_{11} \sim Binomial(n_{1}, Se) \), and from the control sample \( n_{00} \sim Binomial(n_{0}, Sp) \). For sensitivity and specificity, conjugate beta prior
Chapter 2. Estimation of the parameters of a BDT

distributions are proposed, which are appropriate distributions for binomial proportions (as are \( \hat{Se} \) and \( \hat{Sp} \)), i.e.

\[
\hat{Se} \rightarrow \text{Beta}\left(\alpha_{Se}, \beta_{Se}\right), \quad (2.98)
\]

and

\[
\hat{Sp} \rightarrow \text{Beta}\left(\alpha_{Sp}, \beta_{Sp}\right). \quad (2.99)
\]

For the prevalence of the disease (which is also a binomial proportion) we propose

\[
p \rightarrow \text{Beta}\left(\alpha_{p}, \beta_{p}\right) \quad (2.100)
\]

and

\[
t \rightarrow \text{Binomial}\left(n^*, p\right), \quad (2.101)
\]

where \( n^* \) is the sample size starting from which the prevalence has been estimated, and \( t \) is a binomial variable that represents the number of individuals with the disease among the \( n^* \). The a posteriori distributions for \( \hat{Se} \) and the \( \hat{Sp} \) are

\[
\hat{Se}\left|\left(n_{10}, n_1, n_{00}, n_0\right)\right. \rightarrow \text{Beta}\left(n_{11} + \alpha_{Se}, n_{10} + \beta_{Se}\right) \quad (2.102)
\]

and

\[
\hat{Sp}\left|\left(n_{10}, n_1, n_{00}, n_0\right)\right. \rightarrow \text{Beta}\left(n_{00} + \alpha_{Sp}, n_{01} + \beta_{Sp}\right). \quad (2.103)
\]

Considering the a priori distribution (2.100) and the data \( n^* \) and \( t \), the a posteriori distribution for the prevalence is

\[
p\left|\left(n_{10}, n_1, n_{00}, n_0, t, n^*\right)\right. \rightarrow \text{Beta}\left(t + \alpha_{p}, n^* - t + \beta_{p}\right). \quad (2.104)
\]
Once all of the distributions are defined, the estimation of the weighted kappa coefficient is carried out applying the Monte Carlo method and calculating a CI based on quantiles. The algorithm for the Monte Carlo method is the following:

**Step 1.** Set the values of the parameters of the a priori distributions of $\hat{Se}$, $\hat{Sp}$ and $p$, i.e., set the values of $\alpha_{Se}$, $\beta_{Se}$, $\alpha_{Sp}$, $\beta_{Sp}$, $\alpha_p$ and $\beta_p$. For example, a non-informative distribution can be used, such as the distribution $Beta(1,1)$, for each a priori distribution.

**Step 2.** Generate the random number of each one of distributions (2.98), (2.99) and (2.100).

**Step 3.** Generate a random number of distribution (2.101) from the random number generated with distribution (2.100).

**Step 4.** Generate a random number of distributions (2.102), (2.103) and (2.104), using for this purpose the random values obtained in Steps 2 and 3.

**Step 5.** Calculate $\hat{\kappa}(c)$ with equation (2.85) using the random numbers obtained in Step 4. If it is verified that $\hat{Y} = \hat{Se} + \hat{Sp} - 1 \leq 0$, then we must go back to Step 1 (since for all BDTs we must demand that their Youden index be greater than 0 and, therefore, that $\hat{\kappa}(c) > 0$).

**Step 6.** Repeat $M$ times Steps 1 to 5, obtaining $M$ estimations of $\kappa(c)$.

Once this algorithm has been applied, from the $M$ estimations of $\kappa(c)$ a CI based on the quantiles is calculated, i.e. the $100 \times (1 - \alpha)\%$ CI for $\kappa(c)$ is
where $q_\gamma$ is the $\gamma$th quantile of the distribution of the $M$ estimations of $\kappa(c)$. As the estimator of $\kappa(c)$, the average of the $M$ estimations is calculated.

### 2.3.4.2. Simulation experiments

Simulation experiments were carried out to study the asymptotic coverage of the CIs proposed in previous Section 2.3.4.1. For this purpose, 10000 random samples were generated, both case samples and control samples, with sizes 25, 50, 75, 100 and 200, for case and control samples. The nine possible combinations with $\{n_i, n_o\} = \{25, 50, 75\}$ and the four possible combinations with $\{n_i, n_o\} = \{100, 200\}$ have been considered, and therefore 13 pairs of sample sizes have been considered. The case (control) samples were generated from binomial distributions with parameters $n_i$ and $Se(Sp)$. These random samples were generated from values of weighted kappa coefficients, and not setting the values of $Se$ and of $Sp$, through the following method:

1) For the weighted kappa coefficients $\kappa(0)$ and $\kappa(1)$ we set the values $\{0.01, 0.02, ..., 0.98, 0.99\}$.

2) For the weighting index $c$ ($0 < c < 1$) we set the values 0.1 and 0.9.

3) As the disease prevalence we took the values 10%, 25% and 50%.
4) Once the values of \( \kappa(0) \), \( \kappa(1) \), \( c \), and of \( p \) were set, the values of \( Se \) and of \( Sp \) were calculated solving (through the Newton-Raphson method) the system formed by equations (1.11) and (1.12), i.e.

\[
\kappa(0) = \frac{p \times Sp - p \times (1 - Se)}{p \times Se + q \times (1 - Sp)} \quad \text{and} \quad \kappa(1) = \frac{q \times Se - q \times (1 - Sp)}{p \times (1 - Se) + q \times Sp}.
\]

5) Finally, the value of the weighted coefficient was calculated with equation (1.13), i.e.

\[
\kappa(c) = \frac{p(1 - Q) c \kappa(1) + qQ(1 - c) \kappa(0)}{p(1 - Q)c + qQ(1 - c)}.
\]

Therefore, through this method the random samples were generated setting values for weighted kappa coefficients. We then calculated all of the CIs proposed in Section 2.3.4.1 to 95\% (\( \alpha = 5\% \)), calculating the coverage probability and the average length of each one of them. Regarding the bootstrap CI, for each one of the 10000 case and control samples 5000 samples with replacement were generated. As for the Bayesian CI, for each one of the 10000 samples, another 10000 random samples were generated considering several scenarios. As a priori distributions for \( Se \) and \( Sp \), non-informative distributions were considered, i.e. \( \text{Beta}(1,1) \) distributions, for both parameters. For the prevalence of the disease, two a priori distributions were considered: \( \text{Beta}(1,1) \) and \( \text{Beta}(pn^*, qn^*) \). The first one corresponds to a non-informative distribution, and the second distribution is a beta distribution with an average \( p \) and one which uses information provided by the sample from which the estimation of the prevalence was
obtained. As a value of $n^*$, we took 500 and 5000, and thus we studied the effect of this sample size on the asymptotic behaviour of this CI.

The comparison of the performance of the CIs was made following criteria similar to those of Agreti and Coull (1998), Price and Bonnett (2004), and Martín-Andrés and Álvarez-Hernández (2014a, 2014b). These criteria consists of determining if the method “fails” for a confidence at 95%, which happens if the CI has a probability of coverage lower than or equal to 93% (in Appendix 2.2 this method is justified). The selection of the optimum CI was made with the following steps:

1) Choosing the CIs with the fewest failures (probability of coverage $>93\%$)

2) From the CIs with the fewest failures, choose those with the lowest average width (more accurate).

In Tables 2.3 and 2.4 we show the probabilities of coverage and the average lengths of the CIs when $\kappa(c)$ is equal to 0.1 and 0.9 (which are values close to the extremes) and for some sample sizes, indicating in each Table the values of $Se$, $Sp$, $\kappa(0)$, $\kappa(1)$, $c$ and $p$ with those with which we have calculated the value of $\kappa(c)$. Therefore, considering the classification of the values of the weighted kappa coefficient given by Cicchetti (2001), values of the weighted kappa coefficient were considered with different levels of clinical significance (if $0 < \kappa_c < 0.40$ the level of clinical significance is poor; if $0.40 \leq \kappa_c \leq 0.59$ the level of clinical significance is fair; if $0.60 \leq \kappa_c < 0.74$ the level of clinical significance is good; and if $0.75 \leq \kappa_c < 1$ the level of clinical significance is excellent). In these Tables, we indicate the failures (coverage probability $\leq 93\%$) in bold, and we also indicate the coverage probabilities and lengths for the 13
pairs of sample sizes considered. The results obtained for $h = 2$ are not shown since they are very similar to those obtained for $h = z_{a/2}^2/2$. For the Bayesian CI we only show the results obtained for $n' = 500$. From the results obtained in the scenarios considered and for the sample sizes studied, the following conclusions are obtained for each CI:

a) Wald CI. For $h = 0$ and $h = 0.5$ this interval does not fail, even when the value of $\kappa(c)$ is near to 0 or to 1. When the samples are small ($n_i \leq 75$) the Wald CI $h = 0.5$ shows, in general terms, a better performance than when $h = 0$; and for larger sample sizes ($n_i \geq 100$) the value of $h$ (0 or 0.5) has practically no effect on the asymptotic behaviour of this CI. The Wald CI with $h = z_{a/2}^2/2$ (or $h = 2$) has failures, especially when the samples are small ($n_i \leq 75$).

b) Logit CI. For $h = 0$ this interval fails when $\kappa(c)$ is near to 0 ($\kappa(c) = 0.1$) and does not fail when $\kappa(c) \geq 0.2$; for $h = 0.5$ the interval does not fail when $\kappa(c)$ is not near to 0 or 1 ($0.2 \leq \kappa(c) \leq 0.8$) and the sample sizes are large ($n_i \geq 100$). The logit CI with $h = z_{a/2}^2/2$ (or $h = 2$) has failures especially when the sample sizes are small ($n_i \leq 75$).

c) Arcsine CI. The behaviour of the arcsine CI with $h = 0$ is very similar to that of the logit CI with $h = 0$. For $h = 0.5$ this interval does not fail, and for $h = z_{a/2}^2/2$ (or $h = 2$) has failures when the samples are small ($n_i \leq 75$).
Chapter 2. Estimation of the parameters of a BDT

**d) Bootstrap CI.** The bias-corrected bootstrap CI fails when the value of $\kappa(c)$ is near to the extremes ($\kappa(c) \leq 0.2$ or $\kappa(c) \geq 0.8$) and the samples are small ($n \leq 75$); for the rest of the values, in general terms this CI has a similar behaviour to that of Wald CI.

**e) Bayesian CI.** When the a priori distribution of $p$ is a $Beta(1,1)$ this interval fails for moderate or very high values of $\kappa(c)$. When the a priori distribution of $p$ is a $Beta(pn^*,qn^*)$, the Bayesian CI fails when the samples are small ($n \leq 75$) and the value of $\kappa(c)$ is nearer to 1. For the rest of the values of $\kappa(c)$ the behaviour of this is generally similar to that of Wald CI. The performance of the Bayesian CI is better when we use the informative beta distribution, $p \rightarrow Beta(pn^*,qn^*)$, as when we use the non-informative one, $p \rightarrow Beta(1,1)$. This result was predictable, since the informative distribution introduces into the Bayesian model the information provided by the sample from which we estimate the disease prevalence; whereas the non-informative beta distribution is flat for all possible values of $p$ and this distribution has a minimal impact on the a posteriori distribution of $p$. The performance of the Bayesian CI when $n^* = 5000$ is very similar to the case of $n^* = 500$, so that the sample size with which we can estimate the prevalence has practically no effect on the asymptotic behaviour of the Bayesian CI.

Only in terms of the sample sizes, we can give the following rules for the application of the CIs:

**a) When** the sample sizes are small ($n \leq 75$), use the Wald CI with $h = 0.5$, as it is an interval that does not fail for the values of $\kappa(c)$ considered.
b) For other sample sizes \((n_i \geq 100)\) use the Wald CI with \(h = 0\), also because it is the only interval that does not fail.

For sample sizes \(n_i \geq 100\) we can also use the logit and arcsine CIs (with \(h = 0\)), bootstrap and Bayesian with \(p \rightarrow \text{Beta}(pn^*, qn^*)\), although these last two require a greater computational effort than the rest, and the Wald one is the easiest to calculate.
Chapter 2. Estimation of the parameters of a \( BDT \)

Table 2.3. Coverage probabilities and lengths of the CIs for \( \kappa(c) = 0.1 \).

\( Se = 0.55 \  Sp = 0.55 \  p = 50\% \  c = 0.1 \  \kappa_1 = 0.1 \  \kappa_1 = 0.1 \)

|               |               |               |               |               |               |
|---------------|---------------|---------------|---------------|---------------|
|               | Wald CI       |               |               |               |
|               |               |               |               |               |
|               |               |               |               |               |
| \( h = 0 \)   | \( h = 0.5 \) | \( h = \frac{z_{1 - a/2}^2}{2} \) |               |
|               |               |               |               |               |
| \( n_1 \)     | \( n_0 \)     | Cov.          | Leng.         | Cov.          | Leng.         |
| 25            | 25            | 0.959         | 0.434         | 0.966         | 0.424         | 0.981         | 0.397         |
| 25            | 50            | 0.953         | 0.357         | 0.959         | 0.351         | 0.977         | 0.337         |
| 50            | 25            | 0.970         | 0.394         | 0.975         | 0.385         | 0.983         | 0.363         |
| 50            | 50            | 0.961         | 0.319         | 0.964         | 0.315         | 0.974         | 0.305         |
| 75            | 75            | 0.973         | 0.264         | 0.973         | 0.262         | 0.978         | 0.256         |
| 100           | 100           | 0.968         | 0.235         | 0.969         | 0.233         | 0.972         | 0.229         |
| 200           | 200           | 0.981         | 0.179         | 0.981         | 0.179         | 0.983         | 0.178         |
|               |               |               |               |               |               |
|               | Bootstrap CI   |               |               |               |
|               |               |               |               |               |
|               |               |               |               |               |
| \( n_1 \)     | \( n_0 \)     | Cov.          | Leng.         |               |               |               |               |
| 25            | 25            | 0.797         | 0.480         |               |               |               |               |
| 25            | 50            | 0.764         | 0.401         |               |               |               |               |
| 50            | 25            | 0.793         | 0.439         |               |               |               |               |
| 50            | 50            | 0.873         | 0.360         |               |               |               |               |
| 75            | 75            | 0.866         | 0.298         |               |               |               |               |
| 100           | 100           | 0.895         | 0.262         |               |               |               |               |
| 200           | 200           | 0.957         | 0.191         |               |               |               |               |
|               |               |               |               |               |               |               |               |
|               | Bayesian CI    |               |               |               |
|               |               |               |               |               |
|               |               |               |               |               |
|               | \( p \rightarrow Beta(1,1) \) | \( p \rightarrow Beta(p^*, q^n^*) \) |               |
|               |               |               |               |               |
| \( n_1 \)     | \( n_0 \)     | Cov.          | Leng.         |               |               |               |               |
| 25            | 25            | 0.719         | 0.476         |               |               |               |               |
| 25            | 50            | 0.796         | 0.365         |               |               |               |               |
| 50            | 25            | 0.803         | 0.452         |               |               |               |               |
| 50            | 50            | 0.881         | 0.340         |               |               |               |               |
| 75            | 75            | 0.923         | 0.277         |               |               |               |               |
| 100           | 100           | 0.932         | 0.244         |               |               |               |               |
| 200           | 200           | 0.977         | 0.180         |               |               |               |               |

Cov.: coverage probability. Leng.: length.
Table 2.4. Coverage probabilities and lengths of the CIs for $\kappa(c) = 0.9$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Wald CI</th>
<th></th>
<th></th>
<th>Logit CI</th>
<th></th>
<th></th>
<th>Arcsine CI</th>
<th></th>
<th></th>
<th>Bootstrap CI</th>
<th>Bayesian CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>h = 0</td>
<td>h = 0.5</td>
<td>h = $z_{1-a/2}^2/2$</td>
<td>h = 0</td>
<td>h = 0.5</td>
<td>h = $z_{1-a/2}^2/2$</td>
<td>h = 0</td>
<td>h = 0.5</td>
<td>h = $z_{1-a/2}^2/2$</td>
<td>p $\rightarrow$ Beta(1,1)</td>
<td>p $\rightarrow$ Beta($pn^<em>$,$qn^</em>$)</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>--------------</td>
<td>--------</td>
<td>---------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.990</td>
<td>0.251</td>
<td>0.982</td>
<td>0.269</td>
<td>0.868</td>
<td>0.328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.996</td>
<td>0.236</td>
<td>0.989</td>
<td>0.251</td>
<td>0.914</td>
<td>0.303</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.934</td>
<td>0.198</td>
<td>0.975</td>
<td>0.201</td>
<td>0.881</td>
<td>0.234</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.932</td>
<td>0.174</td>
<td>0.986</td>
<td>0.181</td>
<td>0.912</td>
<td>0.205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>0.933</td>
<td>0.142</td>
<td>0.947</td>
<td>0.147</td>
<td>0.934</td>
<td>0.160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.946</td>
<td>0.123</td>
<td>0.956</td>
<td>0.126</td>
<td>0.948</td>
<td>0.135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.947</td>
<td>0.088</td>
<td>0.949</td>
<td>0.089</td>
<td>0.951</td>
<td>0.092</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cov.: coverage probability. Leng.: length.
In Figure 2.1, we show the probabilities of coverage of the Wald CIs (Wald with $h = 0.5$ when $n_i \leq 75$ and Wald with $h = 0$ when $n_i \geq 100$) obtained for values of $\kappa(c)$ equal to 0.2, 0.4, 0.6 and 0.8, and it can be observed that, for the sample sizes considered, these intervals do not fail.

Regarding the corrections $h = 2$ and $h = \frac{z_{1-\alpha/2}^2}{2}$, they do not improve the asymptotic behaviour of the CIs, and for small samples they even have a negative effect on the coverage probabilities of the CIs.
Figure 2.1. Coverage probabilities of the Wald CIs.

\[ \kappa(0.1) = 0.2 \quad Se = 0.157 \quad Sp = 0.959 \quad p = 10\% \]

\[ \kappa(0.9) = 0.4 \quad Se = 0.686 \quad Sp = 0.657 \quad p = 25\% \]

\[ \kappa(0.9) = 0.6 \quad Se = 0.94 \quad Sp = 0.66 \quad p = 10\% \]

\[ \kappa(0.9) = 0.8 \quad Se = 0.9 \quad Sp = 0.9 \quad p = 50\% \]

2.3.4.3. Sample sizes

A very important question in the study of statistical methods for diagnosis is the determination of the sample size necessary to be able to estimate a parameter of a BDT with a determined precision to a confidence \( 100(1-\alpha)\% \). Then, based on the Wald CI
(which is an interval that in general terms performs well, both with small samples and with large ones), a method is proposed to calculate sample sizes \( n_1 \) and \( n_0 \) needed to estimate \( \kappa \) with a precision \( \delta \) and a confidence \( 100(1-\alpha)\% \). Based on the asymptotic normality of the estimator \( \hat{\kappa}(c) \), it is verified that

\[
\hat{\kappa}(c) \in \kappa(c) \pm z_{1-\alpha/2}\sqrt{\text{Var}\left[\hat{\kappa}(c)\right]},
\]

(2.106)
i.e. the probability of obtaining an estimator \( \hat{\kappa}(c) \) is in interval (2.106) with a probability \( 100(1-\alpha)\% \). Let \( \phi = n_0/n_1 \). Substituting \( n_0 \) with \( n_1\phi \) in equation (3.32) (see Appendix 2.1), and setting the value of \( \phi \) and precision \( \delta \), then it is possible to calculate sample size \( n_1 \) using the following expression

\[
\delta = z_{1-\alpha/2}\sqrt{\text{Var}\left[\hat{\kappa}(c)\right]}.
\]

(2.107)

Clearing \( n_1 \) it holds that

\[
n_1 = \left(\frac{z_{1-\alpha/2}\kappa(c)}{pqY\delta}\right)^2 \times \left[pq - \kappa(c)\left(p(q-c)\right)\right]^2 Se(1-Se) + \left[pq + \kappa(c)\left(q(c-q)\right)\right]^2 Sp(1-Sp) \phi.
\]

(2.108)

If an estimation of the disease prevalence is known and setting the value of the weighting index \( c \), this method requires knowledge of the values of \( Se, Sp \) and \( \kappa(c) \), i.e. it is necessary to know some estimators of these parameters (e.g. from pilot samples or other previous studies). The procedure would consist of the following steps:

**Step 1.** Take two pilot samples, one with \( n'_1 \) individuals who have the disease and another one with \( n'_0 \) individuals who do not have the disease, and based on these we
obtain \( \hat{Se} \), \( \hat{Sp} \), \( \hat{\kappa}(c) \) and the Wald CI. If the Wald CI calculated has a precision \( \delta \), i.e. if \( \frac{\text{Upper limit} - \text{Lower limit}}{2} \leq \delta \), then with the two pilot samples precision has been achieved and the process is finished; if precision is not achieved \( \left( \frac{\text{Upper limit} - \text{Lower limit}}{2} > \delta \right) \) then we must go to the next step.

**Step 2.** From the values estimated in Step 1, calculate the new sample size \( n_i \) with equation (2.108) (in this equation each parameter is substituted with its estimated value in Step 1) and calculate \( n_i = n_i \phi \).

**Step 3.** Take two samples, one composed of \( n_i \) individuals with the disease and another composed of \( n_{ii} \) individuals without the disease (to samples \( n_i' \) from Step 1 individuals are added until the completion of new sample sizes \( n_i \)). From these new samples, \( \hat{Se} \), \( \hat{Sp} \), \( \hat{\kappa}(c) \) and the Wald CI are calculated. If the Wald CI calculated has a precision \( \delta \) then with the two new samples precision has been achieved and the process has finished. If the Wald CI does not have the desired precision, then consider these two samples to be pilot samples and go to Step 1.

The method propose to calculate the sample size is an iterative method that depends on the pilot samples and does not guarantee that with the sizes calculated \( \kappa(c) \) is estimated with precision \( \delta \). Therefore, when applying this method it is necessary to check that with the sample sizes calculated the precision required is achieved.

This method can be applied considering \( h = 0 \) when the sizes of the pilot samples are \( \geq 100 \) or with \( h = 0.5 \) when these sizes are lower than 100.
2.3.4.4. Example

The results obtained in the previous Sections were applied to the study by Patil et al (2013) on the assessment of ischaemia modified albumin in the diagnosis of acute coronary syndrome in India. This diagnostic test was applied to a sample of 102 individuals diagnosed with coronary disease and to a sample of 110 healthy control individuals. In Table 2.5 (Frequencies), the data from the study are shown (the frequencies were calculated from the results provided by the authors). In India, the prevalence of coronary heart disease has been estimated at a range of 1.6% to 7.4% in rural populations and from 1% to 13.2% in urban populations (Gupta et al, 2008). Taking into account the conclusions of the simulation experiments obtained in Section 2.3.4.2 and considering the prevalence value to be 5%, in Table 2.5 (Estimation of the weighted kappa coefficient) we can see the estimations of the weighted kappa coefficient for different values of the weighting index $c$. As for the Bayesian CI, two sample sizes have been supposed to estimate the prevalence: $n^* = 1000$ and $n^* = 5000$.

If the clinician considers that the false positives are more important than the false negatives (as is the case when ischaemia modified albumin is going to be used as a definitive test before a risky treatment, e.g. a surgical operation), then $0 \leq c < 0.5$ and the weighted kappa coefficient has a level of clinical significance (Cicchetti, 2001) which varies mainly between ‘poor’ and ‘fair’ (in terms of the 95% CIs) depending on the $c$ index. For example, if the clinician considers that $c = 0.1$, the false positives are 9 times more important than the false negatives, then the 95% Wald CI is $(0.21; 0.55)$ and the level of clinical significance of ischaemia modified albumin is a value between ‘poor’ and ‘fair’ and, therefore, ischaemia modified albumin is not a useful test for the definitive diagnosis of coronary disease.
Table 2.5. Data from the study by Patil et al (2013) and estimations of $\kappa(c)$.

<table>
<thead>
<tr>
<th>Frequencies</th>
<th>Negative test</th>
<th>Positive test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>12</td>
<td>90</td>
<td>102</td>
</tr>
<tr>
<td>Control</td>
<td>102</td>
<td>8</td>
<td>110</td>
</tr>
</tbody>
</table>

Estimation of the weighted kappa coefficient

<table>
<thead>
<tr>
<th>$c$</th>
<th>$h = 0$</th>
<th>$h = 0.5$</th>
<th>$h = 0$</th>
<th>$h = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.38 ± 0.087</td>
<td>0.37 ± 0.084</td>
<td>0.21 - 0.55</td>
<td>0.20 - 0.53</td>
</tr>
<tr>
<td>0.4</td>
<td>0.47 ± 0.089</td>
<td>0.45 ± 0.086</td>
<td>0.29 - 0.64</td>
<td>0.28 - 0.62</td>
</tr>
<tr>
<td>0.5</td>
<td>0.51 ± 0.087</td>
<td>0.49 ± 0.085</td>
<td>0.34 - 0.68</td>
<td>0.32 - 0.66</td>
</tr>
<tr>
<td>0.6</td>
<td>0.55 ± 0.084</td>
<td>0.54 ± 0.082</td>
<td>0.39 - 0.72</td>
<td>0.38 - 0.70</td>
</tr>
<tr>
<td>0.9</td>
<td>0.76 ± 0.050</td>
<td>0.75 ± 0.050</td>
<td>0.66 - 0.86</td>
<td>0.65 - 0.85</td>
</tr>
</tbody>
</table>

95% Logit CI

<table>
<thead>
<tr>
<th>$c$</th>
<th>$h = 0$</th>
<th>$h = 0.5$</th>
<th>$h = 0$</th>
<th>$h = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.23 - 0.56</td>
<td>0.22 - 0.54</td>
<td>0.22 - 0.55</td>
<td>0.21 - 0.53</td>
</tr>
<tr>
<td>0.4</td>
<td>0.30 - 0.64</td>
<td>0.29 - 0.62</td>
<td>0.30 - 0.64</td>
<td>0.29 - 0.62</td>
</tr>
<tr>
<td>0.5</td>
<td>0.34 - 0.67</td>
<td>0.33 - 0.65</td>
<td>0.34 - 0.67</td>
<td>0.33 - 0.66</td>
</tr>
<tr>
<td>0.6</td>
<td>0.39 - 0.71</td>
<td>0.38 - 0.69</td>
<td>0.39 - 0.71</td>
<td>0.38 - 0.69</td>
</tr>
<tr>
<td>0.9</td>
<td>0.65 - 0.84</td>
<td>0.64 - 0.83</td>
<td>0.65 - 0.85</td>
<td>0.65 - 0.84</td>
</tr>
</tbody>
</table>

95% Arcsine CI

<table>
<thead>
<tr>
<th>$c$</th>
<th>$h = 0$</th>
<th>$h = 0.5$</th>
<th>$h = 0$</th>
<th>$h = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.21 - 0.55</td>
<td>0.21 - 0.55</td>
<td>0.22 - 0.55</td>
<td>0.22 - 0.55</td>
</tr>
<tr>
<td>0.4</td>
<td>0.29 - 0.63</td>
<td>0.29 - 0.63</td>
<td>0.30 - 0.63</td>
<td>0.30 - 0.63</td>
</tr>
<tr>
<td>0.5</td>
<td>0.32 - 0.66</td>
<td>0.32 - 0.66</td>
<td>0.33 - 0.66</td>
<td>0.33 - 0.66</td>
</tr>
<tr>
<td>0.6</td>
<td>0.37 - 0.69</td>
<td>0.37 - 0.69</td>
<td>0.38 - 0.69</td>
<td>0.38 - 0.69</td>
</tr>
<tr>
<td>0.9</td>
<td>0.63 - 0.83</td>
<td>0.63 - 0.83</td>
<td>0.63 - 0.83</td>
<td>0.63 - 0.83</td>
</tr>
</tbody>
</table>

95% Bayesian CI

$n^* = 1000, p \rightarrow Beta(pn^*, qn^*)

<table>
<thead>
<tr>
<th>$c$</th>
<th>$\hat{\kappa}(c)$</th>
<th>95% CI</th>
<th>$\hat{\kappa}(c)$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.37</td>
<td>0.21 - 0.55</td>
<td>0.37</td>
<td>0.22 - 0.55</td>
</tr>
<tr>
<td>0.4</td>
<td>0.45</td>
<td>0.29 - 0.63</td>
<td>0.45</td>
<td>0.30 - 0.63</td>
</tr>
<tr>
<td>0.5</td>
<td>0.49</td>
<td>0.32 - 0.66</td>
<td>0.49</td>
<td>0.33 - 0.66</td>
</tr>
<tr>
<td>0.6</td>
<td>0.53</td>
<td>0.37 - 0.69</td>
<td>0.53</td>
<td>0.38 - 0.69</td>
</tr>
<tr>
<td>0.9</td>
<td>0.74</td>
<td>0.63 - 0.83</td>
<td>0.74</td>
<td>0.63 - 0.83</td>
</tr>
</tbody>
</table>
If the clinician considers that the false negatives are more important than the false positives (as is the case when ischaemia modified albumin is going to be used as a screening test), then \(0.5 < c < 1\) and the level of clinical significance (Cicchetti, 2001) of the weighted kappa coefficient varies between ‘fair’ and ‘excellent’ (in terms of the 95% CIs) depending on the \(c\) index. For example, for \(c = 0.9\), the false negatives are 9 times more important than the false positives, the 95% Wald CI is \((0.66; 0.86)\) and the level of clinical significance of ischaemia modified albumin is a value between ‘good’ and ‘excellent’ and, therefore, this diagnostic test is very useful as a screening test.

As for the CIs, in this example as the sample sizes \(n_i\) and \(n_0\) are relatively large \((n_i > 100)\), all of them are very similar. Regarding the Bayesian CI, the sample size from which the prevalence is estimated does not have any effect upon the intervals obtained.

In order to illustrate the method described in Section 2.3.4.3 about the calculation of the sample sizes, let us consider that \(c = 0.9\). In this situation, the 95% Wald CI with \(h = 0\) is \((0.66; 0.86)\) and the precision is \(0.10 \left(\frac{0.86 - 0.66}{2} = 0.10\right)\). As an example, let us consider that the clinician wishes to estimate the weighted kappa coefficient with a precision \(\delta = 0.05\) and that the case sample and the control sample are the same size \((\phi = 1)\). As with the samples of 102 and 110 individuals, the desired precision \((\delta = 0.05 < 0.10)\) was not achieved, then using the two previous samples as pilot samples \((\hat{Se} = 0.88, \hat{Sp} = 0.93\) and \(\hat{\kappa}(0.9) = 0.76\) applying equation (2.108) it is obtained that \(n_i = 227\) and \(n_0 = \phi n_i = 227\). Therefore, to the initial case sample it is necessary to add 125 individuals with the disease and 117 individuals without the
disease must be added to the control sample. Once these new samples have been taken, it is necessary to check that the desired precision is verified.

2.3.4.5. Discussion

The weighted kappa coefficient is a valid measure to assess and compare the performance of BDTs, and it depends on the sensitivity and the specificity of the test, on the prevalence of the disease and the relative importance between the false positives and the false negatives (weighting index). The sensitivity and the specificity of the test are easily estimated, both in cross-sectional studies and in case-control studies, since they are estimators of binomial proportions. Regarding the weighting index, this is set by the clinician depending on his or her knowledge of the problem in question. The estimation of the weighted kappa coefficient in case-control studies requires knowledge of an estimation of the prevalence of the disease based on another study (or from the literature, etc.), since from this design it is impossible to estimate the prevalence. In this article, we have studied the estimation of the weighted kappa coefficient of a single BDT subject to case-control study assuming that we have an estimation of the prevalence of the disease. Fourteen approximate CIs were studied and compared for this parameter (thirteen were frequentist and one was Bayesian). The results of the simulation experiments carried out demonstrated that, in very general terms, for small samples \( n_i \leq 75 \) it is possible to use the Wald CI with \( h = 0.5 \), and that for samples with \( n_i \geq 100 \) the Wald, logit, arcsine (with \( h = 0 \)) and Bayesian (with \( \text{Beta}(p n^*, q n^*) \)) intervals can be used. The Bayesian CI performs well when for prevalence an a priori distribution \( \text{Beta}(p n^*, q n^*) \) is considered, where \( n^* \) is the sample size from which the
prevalence has been estimated. If sample size $n^\ast$ is not known, any value (e.g. 500 or 5000) can be used, since the simulation experiments have demonstrated that this value has practically no effect upon the coverage probability and length of the interval.

Another important question that has been studied is the calculation of the sample sizes $n_1$ and $n_0$ needed to estimate the weighted kappa coefficient with a determined precision. Therefore, and once we have set the values of index $c$, of the desired precision $\delta$ and of the relation between $n_1$ and $n_0$ ($\phi$), a method based on the Wald CI has been proposed that requires knowledge of the sensitivity, the specificity and the weighted kappa coefficient of the BDT (e.g. from pilot samples or other studies). The method proposed depends on the auxiliary information (pilot samples or other studies) and, therefore, it does not guarantee the estimation of the weighted kappa coefficient with the desired precision, and thus it is necessary to check this condition once the new samples have been taken.

The estimation of the weighted kappa coefficient requires knowledge of an estimation of the prevalence obtained from another study e.g. a health survey. To study the effect of a misspecification of the prevalence in the estimation of the weighted kappa coefficient, we carried out simulation experiments similar to those made in Section 2.3.4.2. For this we took as the prevalence for the inference an overestimation (and underestimation) equal to 10% and to 20% of the value of the prevalence set, and we studied the asymptotic behaviour of the intervals recommended (Wald with $h = 0.5$ for small samples and Wald with $h = 0$ for large samples). In the Tables 2.6 and 2.7 we show some of the results obtained. The results demonstrated that slight (10%) and
moderate (20%) misspecifications of the disease prevalence do not have an important effect on the probabilities of coverage of the Wald CIs.

Table 2.6. Coverage probabilities and lengths of the Wald CIs for $\kappa(c) = 0.1$.

<table>
<thead>
<tr>
<th>$n_t$</th>
<th>$n_0$</th>
<th>Cov.</th>
<th>Leng.</th>
<th>Cov.</th>
<th>Leng.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.962</td>
<td>0.443</td>
<td>0.977</td>
<td>0.399</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.957</td>
<td>0.374</td>
<td>0.974</td>
<td>0.333</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.974</td>
<td>0.402</td>
<td>0.982</td>
<td>0.361</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.955</td>
<td>0.331</td>
<td>0.973</td>
<td>0.295</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>0.965</td>
<td>0.278</td>
<td>0.979</td>
<td>0.246</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.961</td>
<td>0.249</td>
<td>0.979</td>
<td>0.220</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.965</td>
<td>0.190</td>
<td>0.977</td>
<td>0.167</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_t$</th>
<th>$n_0$</th>
<th>Cov.</th>
<th>Leng.</th>
<th>Cov.</th>
<th>Leng.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.953</td>
<td>0.460</td>
<td>0.987</td>
<td>0.371</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.943</td>
<td>0.391</td>
<td>0.979</td>
<td>0.308</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.969</td>
<td>0.417</td>
<td>0.987</td>
<td>0.336</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.947</td>
<td>0.345</td>
<td>0.984</td>
<td>0.273</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>0.958</td>
<td>0.291</td>
<td>0.987</td>
<td>0.227</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.957</td>
<td>0.260</td>
<td>0.983</td>
<td>0.202</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.956</td>
<td>0.199</td>
<td>0.963</td>
<td>0.154</td>
</tr>
</tbody>
</table>

Cov.: coverage probability. Leng.: length.
Chapter 2. Estimation of the parameters of a BDT

In this Section we have studied the estimation of the weighted kappa coefficient through CIs. Further research should be carried out when, subject to case-control study, a hypothesis test on this parameter is made i.e. $H_0 : \kappa_c = \kappa_{c_0}$ vs $H_1 : \kappa_c \neq \kappa_{c_0}$, proposing different methods to solve this hypothesis test and also studying the sample sizes necessary to solve the hypothesis test to an error $\alpha$ and to a power $\theta$.

Table 2.7. Coverage probabilities and lengths of the Wald CIs for $\kappa(c) = 0.9$.

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$n_0$</th>
<th>Cov.</th>
<th>Leng.</th>
<th>Cov.</th>
<th>Leng.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.982</td>
<td>0.287</td>
<td>0.981</td>
<td>0.280</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.991</td>
<td>0.269</td>
<td>0.991</td>
<td>0.259</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.979</td>
<td>0.205</td>
<td>0.979</td>
<td>0.203</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.986</td>
<td>0.188</td>
<td>0.986</td>
<td>0.182</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>0.941</td>
<td>0.150</td>
<td>0.947</td>
<td>0.144</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.944</td>
<td>0.125</td>
<td>0.933</td>
<td>0.120</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.941</td>
<td>0.090</td>
<td>0.939</td>
<td>0.086</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$n_0$</th>
<th>Cov.</th>
<th>Leng.</th>
<th>Cov.</th>
<th>Leng.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.982</td>
<td>0.292</td>
<td>0.981</td>
<td>0.278</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.986</td>
<td>0.275</td>
<td>0.991</td>
<td>0.255</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.962</td>
<td>0.208</td>
<td>0.979</td>
<td>0.204</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.974</td>
<td>0.192</td>
<td>0.986</td>
<td>0.179</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>0.960</td>
<td>0.154</td>
<td>0.947</td>
<td>0.142</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.944</td>
<td>0.129</td>
<td>0.934</td>
<td>0.118</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.948</td>
<td>0.092</td>
<td>0.933</td>
<td>0.084</td>
</tr>
</tbody>
</table>

Cov.: coverage probability. Leng.: length.
Chapter 3

Comparison and Combination of two Binary Diagnostic Tests

3.1 Introduction

This Chapter studies the comparison of parameters of two BDTs subject to a paired design and the combination of two BDTs. The comparison of parameters of two BDTs is an important topic in the field of Statistical Methods for Diagnosis in Medicine. In this Chapter, the hypothesis tests and the CIs to compare the parameters of two BDTs subject to a paired design are presented. In this Chapter we also study the combination of the parameters of two BDTs. In practice, it is common to combine two BDTs in order to increase the accuracy of the diagnosis of the disease. The combination of the sensitivities and of specificities, likelihoods ratios, predictive values and weighted kappa coefficient of two BDTs are presented. The combination of the weighted kappa
coefficients of two BDTs is the second contribution of this Thesis. Each one of these objectives will now be studied.

3.2. Comparison of two BDTs

Let us consider two BDTs whose performance is compared in relation to the same GS. A paired design consists of applying the two BDTs and the GS to all the individuals in a random sample sized \( n \), and the most common type of sampling when we want to compare the performance of two BDTs. Let \( T_1 \) and \( T_2 \) be the random binary variables that model the results of Test 1 and Test 2 respectively, in such a way that \( T_k = 1 \) when the result of the BDT is positive and \( T_k = 0 \) when the result is negative. Let \( D \) be the random binary variable that models the result of the GS, in such a way that \( D = 1 \) when the individual is diseased and \( D = 0 \) when the individual is non-diseased. Let \( Se_k = P(T_k = 1|D = 1) \) be the sensitivity of the \( k \)th BDT and \( Sp_k = P(T_k = 0|D = 0) \) be the specificity; \( LR_k^+ = Se_k/(1-Sp_k) \) the positive likelihood ratio and \( LR_k^- = (1-Se_k)/Sp_k \) the negative likelihood ratio; \( PPV_k = P(D = 1|T_k = 1) \) the positive predictive value and \( NPV_k = P(D = 0|T_k = 0) \) the negative predictive value; and \( \kappa_k(c) \) the weighted kappa coefficient of the \( k \)th BDT. Let us consider that the two BDTs and the GS are applied independently to all of the individuals in a sample sized \( n \), leading to Table 3.1, where \( s_{ij} \) is the number of diseased individuals in which \( T_1 = i \) and \( T_2 = j \), and \( r_{ij} \) is the number of non-diseased individuals in which \( T_1 = i \) and \( T_2 = j \), with \( i, j = 0,1 \).
The data in Table 3.1 are the product of a multinomial distribution whose probabilities are shown in Table 3.2, where

\[ p_{hk} = p\left[ Se_1^h (1 - Se_1)^{1-h} Se_2^k (1 - Se_2)^{1-k} + \delta_{hk} \varepsilon_1 \right] \]  \hspace{1cm} (3.1)

and

\[ q_{hk} = q\left[ Sp_1^{1-h} (1 - Sp_1)^h Sp_2^{1-k} (1 - Sp_2)^{1-k} + \delta_{hk} \varepsilon_0 \right] \], \hspace{1cm} (3.2)

with \( p = P(D = 1) = \sum_{h,k=0}^{1} p_{hk} \), \( q = 1 - p = \sum_{h,k=0}^{1} q_{hk} \) and \( h, k = 0,1 \). The parameter \( \varepsilon_1 \) is the covariance (Vacek, 1985) between the two \( BDTs \) when \( D = 1 \), and the parameter \( \varepsilon_0 \) is the covariance between the two \( BDTs \) when \( D = 0 \), with \( \delta_{hk} = 1 \) if \( h = k \) and \( \delta_{hk} = -1 \) if \( h \neq k \), and verifying that \( 0 \leq \varepsilon_1 \leq \min\{Se_1(1-Se_2), Se_2(1-Se_1)\} \) and \( 0 \leq \varepsilon_0 \leq \min\{Sp_1(1-Sp_2), Sp_2(1-Sp_1)\} \). If \( \varepsilon_1 = \varepsilon_0 = 0 \) then the two \( BDTs \) are conditionally independent on the disease. In practice the supposition of conditional independence is not realistic, so that usually \( \varepsilon_1 > 0 \) and/or \( \varepsilon_0 > 0 \). The Vacek model (1985) treats conditional dependence as an additive factor. Using the transformations 

\[ \varepsilon_1 = Se_1 Se_2 (\alpha_1 - 1) \] \hspace{1cm} and \hspace{1cm} \[ \varepsilon_0 = (1 - Sp_1)(1 - Sp_2)(\alpha_0 - 1) \], \hspace{1cm} (3.3)

the model treats conditional dependence as a multiplicative factor. Parameter \( \alpha_1 \) is the covariance (Berry et al, 2002) between the \( BDTs \) when \( D = 1 \), and \( \alpha_0 \) is the covariance (Berry et al, 2002) between the \( BDTs \) when \( D = 0 \), and it is verified that \( 1 \leq \alpha_i \leq 1/\max\{Se_1, Se_2\} \) and \( 1 \leq \alpha_0 \leq 1/\max\{(1 - Sp_1), (1 - Sp_2)\} \). In the case that \( \alpha_1 = \alpha_0 = 1 \) the two \( BDTs \) are conditionally independent on the disease.
Table 3.1. Frequencies when comparing two BDTs subject to a paired design.

<table>
<thead>
<tr>
<th></th>
<th>$T_1 = 0$</th>
<th>$T_1 = 1$</th>
<th></th>
<th>$T_2 = 0$</th>
<th>$T_2 = 1$</th>
<th></th>
<th>$T_2 = 0$</th>
<th>$T_2 = 1$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 1$</td>
<td>$s_{00}$</td>
<td>$s_{01}$</td>
<td>$s_{10}$</td>
<td>$s_{11}$</td>
<td>$s$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D = 0$</td>
<td>$r_{00}$</td>
<td>$r_{01}$</td>
<td>$r_{10}$</td>
<td>$r_{11}$</td>
<td>$r$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$n_{00}$</td>
<td>$n_{01}$</td>
<td>$n_{10}$</td>
<td>$n_{11}$</td>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2. Probabilities subject to a paired design.

<table>
<thead>
<tr>
<th></th>
<th>$T_1 = 1$</th>
<th>$T_1 = 0$</th>
<th></th>
<th>$T_2 = 1$</th>
<th>$T_2 = 0$</th>
<th></th>
<th>$T_2 = 1$</th>
<th>$T_2 = 0$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 1$</td>
<td>$p_{11}$</td>
<td>$p_{10}$</td>
<td>$p_{01}$</td>
<td>$p_{00}$</td>
<td>$p$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D = 0$</td>
<td>$q_{11}$</td>
<td>$q_{10}$</td>
<td>$q_{01}$</td>
<td>$q_{00}$</td>
<td>$q$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$p_{11} + q_{11}$</td>
<td>$p_{10} + q_{10}$</td>
<td>$p_{01} + q_{01}$</td>
<td>$p_{00} + q_{00}$</td>
<td>$1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.1. Sensitivities and specificities

The hypothesis test to compare two sensitivities is

$$H_0 : Se_1 = Se_2 \ vs \ H_1 : Se_1 \neq Se_2,$$

that it is equivalent to checking

$$H_0 : p_{01} = p_{10} \ vs \ H_1 : p_{01} \neq p_{10}.$$
Therefore, the comparison of two sensitivities is equivalent to the comparison of two paired binomial proportions. If \( s_{01} + s_{10} > 10 \) this hypothesis test is solved by applying the McNemar’s test (with correction continuity), i.e.

\[
z = \frac{|s_{01} - s_{10}| - 0.5}{\sqrt{s_{01} + s_{10}}} \to N(0,1).
\] (3.3)

If \( s_{01} + s_{10} \leq 10 \), the hypothesis test is solved by applying the exact test to compare two paired binomial proportions. In this situation, the exact \( p\)-value is

\[
p-value = 2 \times \sum_{i=0}^{h} \left( \frac{s_{01} + s_{10}}{2^i} \right)^{s_{01} + s_{10}}, \quad h = \min(s_{01}, s_{10}).
\] (3.4)

The comparison of the two specificities is solved in a similar way. The hypothesis test to compare two specificities is

\[H_0 : Sp_1 = Sp_2 \text{ vs } H_1 : Sp_1 \neq Sp_2,\]

that it is equivalent to checking

\[H_0 : q_{01} = q_{10} \text{ vs } H_1 : q_{01} \neq q_{10}.\]

If \( r_{01} + r_{10} > 10 \), the McNemar’s test (with correction continuity) is

\[
z = \frac{|r_{01} - r_{10}| - 0.5}{\sqrt{r_{01} + r_{10}}} \to N(0,1).
\] (3.5)

If \( r_{01} + r_{10} \leq 10 \), the exact \( p\)-value is

\[
p-value = 2 \times \sum_{i=0}^{h} \left( \frac{r_{01} + r_{10}}{2^i} \right)^{r_{01} + r_{10}}, \quad h = \min(r_{01}, r_{10}).
\] (3.6)
Chapter 3. Comparison and combination of two BDTs

The estimation by CIs of the difference between two paired proportions has been object of many studies. Agresti and Min (2005) have proposed a CI called Wald + 2 for the differences of two paired binomial proportions, which is a CI that has good asymptotic coverage. In terms of data of Table 3.1, the 100(1 − α)% Wald + 2 CI for the difference between the two sensitivities is

\[
Se_1 - Se_2 \in \frac{s_{01} - s_{01}}{s + 2} \pm z_{1-\alpha/2} \sqrt{\frac{(s_{10} + s_{01} + 1) - \frac{(s_{10} - s_{01})^2}{s + 2}}{s + 2}},
\] (3.7)

and for the difference between the two specificities the CI is

\[
Sp_1 - Sp_2 \in \frac{r_{01} - r_{01}}{r + 2} \pm z_{1-\alpha/2} \sqrt{\frac{(r_{01} + r_{10} + 1) - \frac{(r_{01} - r_{10})^2}{r + 2}}{r + 2}}.
\] (3.8)

3.2.2. Likelihood ratios

Roldán-Nofuentes and Luna del Castillo (2007) have studied the comparison of the likelihood ratios of two BDTs subject to a paired design. When the results of the two BDTs are positive, the hypothesis test to check the equality of the LRs is

\[
H_0 : \omega^* = 0 \ vs \ H_1 : \omega^* \neq 0,
\]

and when the results of both BDTs are negative, the hypothesis test is

\[
H_0 : \omega^- = 0 \ vs \ H_1 : \omega^- \neq 0,
\]

where \( \omega^* = \ln \left( \frac{LR_1^*}{LR_2^*} \right) \) and \( \omega^- = \ln \left( \frac{LR_1^-}{LR_2^-} \right) \). In terms of the probabilities given in Table 3.2, the positive LRs are written as
Chapter 3. Comparison and combination of two BDTs

\[ LR^+_1 = \frac{1 - p}{p} \frac{p_{11} + p_{10}}{q_{11} + q_{10}} \quad \text{and} \quad LR^+_2 = \frac{1 - p}{p} \frac{p_{11} + p_{01}}{q_{11} + q_{01}}, \quad (3.9) \]

and the negative LRs are written as

\[ LR^-_1 = \frac{1 - p}{p} \frac{p_{01} + p_{00}}{q_{01} + q_{00}} \quad \text{and} \quad LR^-_2 = \frac{1 - p}{p} \frac{p_{10} + p_{00}}{q_{10} + q_{00}}. \quad (3.10) \]

Therefore, it holds that

\[ \omega^+ = \ln \left\{ \frac{p_{11} + p_{10}}{q_{11} + q_{10}} \right\} \quad \text{and} \quad \omega^- = \ln \left\{ \frac{p_{01} + p_{00}}{q_{01} + q_{00}} \right\}. \quad (3.11) \]

Let \( \mathbf{\pi} = (p_{00}, p_{01}, p_{10}, p_{11}, q_{00}, q_{01}, q_{10}, q_{11})^T \) be a vector whose components are the probabilities in Table 3.2. As the probabilities \( p_{ij} \) and \( q_{ij} \) are probabilities of a multinomial distribution, their MLEs are \( \hat{p}_{ij} = \frac{s_{ij}}{n_j} \) and \( \hat{q}_{ij} = \frac{n_j}{n} \), and the estimated variance-covariance matrix of \( \hat{\mathbf{\pi}} \) is \( \hat{\Sigma}_{\mathbf{\pi}} = \left\{ \text{diag}(\hat{\mathbf{\pi}}) - \hat{\mathbf{\pi}}^T \hat{\mathbf{\pi}} \right\} / n \). The MLEs of the positive LRs are

\[ LR^+_1 = r \frac{r_{11} + r_{10}}{s (r_{11} + r_{10})} \quad \text{and} \quad LR^+_2 = r \frac{s (r_{11} + r_{01})}{s (r_{11} + r_{01})}, \quad (3.12) \]

and the MLEs of the negative LRs are

\[ LR^-_1 = r \frac{s_{01} + s_{00}}{s (r_{01} + r_{00})} \quad \text{and} \quad LR^-_2 = r \frac{s (r_{10} + r_{00})}{s (r_{10} + r_{00})}. \quad (3.13) \]

Then the MLEs of \( \omega^+ \) and \( \omega^- \) are

\[ \hat{\omega}^+ = \ln \left\{ \frac{(s_{11} + s_{10}) (r_{11} + r_{10})}{(s_{11} + s_{01}) (r_{11} + r_{10})} \right\} \quad \text{and} \quad \hat{\omega}^- = \ln \left\{ \frac{(s_{01} + s_{00}) (r_{10} + r_{00})}{(s_{01} + s_{00}) (r_{10} + r_{00})} \right\}. \quad (3.14) \]
and applying the delta method the estimated variance of $\hat{\omega}$ is

$$\hat{\text{Var}}(\hat{\omega}) = \left( \frac{\partial \omega}{\partial \pi} \right)_{\pi=\pi} \sum_{z} \left( \frac{\partial \omega}{\partial \pi} \right)_{\pi=\pi}^{T}, \tag{3.15}$$

where $\hat{\omega}$ is $\hat{\omega}^+$ or $\hat{\omega}^-$. Finally, the statistic for the hypothesis test $H_0 : \omega = 0$ vs $H_1 : \omega \neq 0$ is

$$z = \frac{\hat{\omega}}{\sqrt{\hat{\text{Var}}(\hat{\omega})}} \xrightarrow{n \to \infty} N(0,1). \tag{3.16}$$

An approximate CI for $\omega$ is obtained by inverting the contrast statistics (3.16), i.e.

$$\omega \in \hat{\omega} \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}(\hat{\omega})}, \tag{3.17}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the normal standard distribution. The 100(1-$\alpha$)% CI for the ratio of the two positive (negative) LRs is

$$\frac{LR_1}{LR_2} \in \exp \left[ \hat{\omega} \pm z_{1-\alpha/2} \sqrt{\hat{\text{Var}}(\hat{\omega})} \right], \tag{3.18}$$

where $LR_i$ is $LR_i^+$ or $LR_i^-$. Roldán-Nofuentes et al (2007) also studied the simultaneous comparison of the LRs of the two BDTs. This hypothesis test consists of solving the global hypothesis test

$$H_0 : (\omega^- = 0 \text{ and } \omega^+ = 0) \text{ vs } H_1 : (\omega^- \neq 0 \text{ and/or } \omega^+ \neq 0).$$

Applying the central theorem of the multivariate limit it holds that

$$\sqrt{n} \left( \begin{array}{c} \hat{\omega}^+ - \omega^+ \\ \hat{\omega}^- - \omega^- \end{array} \right) \xrightarrow{n \to \infty} N \left[ \begin{array}{c} 0 \\ 0 \end{array} \right], \Sigma_{\pi} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \tag{3.19}$$
where the elements of the variance-covariance matrix are estimated applying the delta method. In this way, if $\omega = (\omega^+, \omega^-)^T$ then

$$\hat{\omega} \approx \left( \frac{\partial \omega}{\partial \pi} \right)_{\pi = \hat{\pi}} \hat{\pi}^{-1} \left( \frac{\partial \omega}{\partial \pi} \right)_{\pi = \hat{\pi}}^T,$$

and the test statistic of the global hypothesis test is

$$Q^2 = \hat{\omega}^T \hat{\pi}^{-1} \hat{\omega} \xrightarrow{n \to \infty} \chi^2_2.$$

Finally, the comparison of the $LRs$ of the two $BDTs$ subject to a paired design is realised following the next steps: 1) Solve the global hypothesis test $H_0: (\omega^- = 0 \text{ and } \omega^+ = 0)$ to an error of $\alpha$ using the statistic $Q^2 = \hat{\omega}^T \hat{\pi}^{-1} \hat{\omega}$; 2) If the global hypothesis test is not significant to an error of $\alpha$, then the homogeneity of the $LRs$ of the two $BDTs$ is not rejected, but if the global hypothesis test is significant to an error of $\alpha$, then the study of the causes of the significance is performed by solving hypothesis tests $H_0: \omega^+ = 0$ and $H_0: \omega^- = 0$ along with a multiple comparison method (e.g. Bonferroni (1936), Holm (1979) or Hochberg (1988)) to an error of $\alpha$.

### 3.2.3. Predictive values

The comparison of the predictive values of two binary diagnostic tests is a topic of great interest in the study of Statistical Methods for Diagnosis, and has been the subject of many papers in the literature of Statistics. The most recent studies are: Leisenring et al (2000), Wang et al (2006), Roldán-Nofuentes et al (2012) and Kosinski (2013). From Tables 3.1 and 3.2, the $MLEs$ of $PVs$ are
Chapter 3. Comparison and combination of two BDTs

\[
P\hat{PV}_1 = \frac{s_{10} + s_{11}}{s_{10} + s_{11} + r_{10} + r_{11}} \quad \text{and} \quad P\hat{PV}_2 = \frac{r_{00} + r_{01}}{s_{00} + s_{01} + r_{00} + r_{01}} \tag{3.22}
\]

for Test 1, and

\[
N\hat{PV}_1 = \frac{s_{01} + s_{11}}{s_{01} + s_{11} + r_{01} + r_{11}} \quad \text{and} \quad N\hat{PV}_2 = \frac{r_{00} + r_{01}}{s_{00} + s_{01} + r_{00} + r_{01}} \tag{3.23}
\]

for Test 2. Applying the delta method, the estimated variances-covariances of the estimators of the predictive values are:

\[
\hat{\text{Var}}(P\hat{PV}_1) = \frac{(s_{10} + s_{11})(r_{10} + r_{11})}{n(s_{10} + s_{11} + r_{10} + r_{11})^2}, \quad \hat{\text{Var}}(P\hat{PV}_2) = \frac{(s_{01} + s_{11})(r_{01} + r_{11})}{n(s_{01} + s_{11} + r_{01} + r_{11})^2},
\]

\[
\hat{\text{Var}}(N\hat{PV}_1) = \frac{(s_{00} + s_{01})(r_{00} + r_{01})}{n(s_{00} + s_{01} + r_{00} + r_{01})^2}, \quad \hat{\text{Var}}(N\hat{PV}_2) = \frac{(s_{00} + s_{10})(r_{00} + r_{01})}{n(s_{00} + s_{10} + r_{00} + r_{01})^2},
\]

\[
\hat{\text{Cov}}(P\hat{PV}_1, P\hat{PV}_2) = \frac{s_{01}s_{10}r_{11} + s_{11}\left[r_{01}(r_{10} + r_{11}) + r_{11}(s_{01} + s_{10} + s_{11} + r_{01} + r_{11})\right]}{(s_{01} + s_{11} + r_{01} + r_{11})^2(s_{10} + s_{11} + r_{01} + r_{11})^2},
\]

\[
\hat{\text{Cov}}(P\hat{PV}_1, N\hat{PV}_1) = 0,
\]

\[
\hat{\text{Cov}}(P\hat{PV}_2, N\hat{PV}_1) = -\frac{s_{00}(s_{10} + s_{11})r_{10} + s_{10}r_{10}(s_{10} + s_{11} + r_{00} + r_{10}) + s_{10}(r_{00} + r_{10})r_{11}}{(s_{00} + s_{10} + r_{00} + r_{10})^2(s_{10} + s_{11} + r_{00} + r_{11})^2},
\]

\[
\hat{\text{Cov}}(P\hat{PV}_2, N\hat{PV}_2) = 0,
\]

\[
\hat{\text{Cov}}(N\hat{PV}_1, N\hat{PV}_2) = \frac{s_{00}(r_{00} + r_{01})r_{00} + r_{00}\left[r_{00}^2 + r_{01}s_{10} + s_{00}\left(s_{01} + s_{00} + r_{00} + r_{01}\right)\right]}{(s_{00} + s_{01} + r_{00} + r_{01})^2(s_{00} + s_{10} + r_{00} + r_{10})^2}. \tag{3.24}
\]

We will now describe each of these four methods to compare the PVs of two BDTs subject to a paired design.
3.2.3.1. Study of Leisenring et al

Leisenring et al (2000) studied the comparison of the PPVs and NPVs of two BDTs through marginal regression models. These authors deduced score statistics to compare the PPVs and NPVs of two BDTs in paired designs. The score statistic for the test

\[ H_0 : PPV_1 = PPV_2 \quad \text{vs} \quad H_1 : PPV_1 \neq PPV_2 \quad (3.25) \]

is

\[ T_{ppv} = \frac{(s_i(1-2Z_i)+s_i(1-Z_i)-s_{i-1})^2}{s_i(1-D_i)(1-2Z_i)+s_i(1-D_i)(1-Z_i)+s_i(1-D_i)(1-2Z_i)+s_i(1-D_i)(1-Z_i)+s_i(1-Z_i)+s_i(1-Z_i)} \quad (3.26) \]

and the score statistic for the test

\[ H_0 : NPV_1 = NPV_2 \quad \text{vs} \quad H_1 : NPV_1 \neq NPV_2 \quad (3.27) \]

is

\[ T_{npv} = \frac{(r_i(1-2Z_i)+r_i(1-Z_i)-r_{i-1})^2}{r_i(1-D_i)(1-2Z_i)+r_i(1-D_i)(1-Z_i)+r_i(1-D_i)(1-2Z_i)+r_i(1-D_i)(1-Z_i)+r_i(1-Z_i)+r_i(1-Z_i)} \quad (3.28) \]

The score statistics have a chi-squared distribution with 1 degree of freedom when the null hypothesis is true, and where

\[ Z_1 = \frac{s_{11}+s_{01}+r_{11}+r_{01}}{2s_{11}+s_{01}+s_{10}+2r_{11}+r_{01}+r_{01}}, \quad \bar{D}_1 = \frac{2s_{11}+s_{01}+s_{00}}{2s_{11}+s_{01}+s_{10}+2r_{11}+r_{01}+r_{01}}, \]

\[ Z_2 = \frac{s_{00}+s_{10}+r_{00}+r_{10}}{2s_{00}+s_{10}+s_{10}+2r_{00}+r_{01}+r_{10}} \quad \text{and} \quad \bar{D}_2 = \frac{2r_{00}+r_{01}+r_{10}}{2s_{00}+s_{01}+s_{10}+2r_{00}+r_{01}+r_{10}}. \]
Chapter 3. Comparison and combination of two BDTs

3.2.3.2. Study of Wang et al

Wang et al (2006) studied the comparison of the PVs of two BDTs through a weighted least square method. The statistics proposed for hypothesis tests (3.25) and (3.27) are

\[ \chi^2_{PPV} = \frac{(P\hat{PV}_1 - P\hat{PV}_2)^2}{\text{Var}(P\hat{PV}_1 - P\hat{PV}_2)} \] and \[ \chi^2_{NPV} = \frac{(N\hat{PV}_1 - N\hat{PV}_2)^2}{\text{Var}(N\hat{PV}_1 - N\hat{PV}_2)}, \] (3.29)

respectively. Both statistics have a chi-squared distribution with 1 degree of freedom, and the variances are estimated by applying the delta method (equations (3.24)).

CIs for the difference between the two PPVs and the two NPVs are obtained by inverting each contrast statistics (3.29), i.e.

\[ PPV_1 - PPV_2 \in \left( P\hat{PV}_1 - P\hat{PV}_2 \right) \pm z_{1-\alpha/2} \sqrt{\text{Var}(P\hat{PV}_1 - P\hat{PV}_2)} \] (3.30)

and

\[ NPV_1 - NPV_2 \in \left( N\hat{PV}_1 - N\hat{PV}_2 \right) \pm z_{1-\alpha/2} \sqrt{\text{Var}(N\hat{PV}_1 - N\hat{PV}_2)}. \] (3.31)

Wang et al have compared this method with the method of Leisenring et al (2000), and they recommend using the weighted least square method.

3.2.3.3. Study of Kosinski

Kosinski (2013) has proposed a weighted generalized score statistic to solve the hypothesis tests of comparison of the PVs. The weighted generalized score statistic for the hypothesis test (3.25) is
Chapter 3. Comparison and combination of two BDTs

\[ T_{WGS}^{PPV} = \frac{\left( P\hat{P}V_1 - P\hat{P}V_2 \right)^2}{\left\{ P\hat{P}V_p \left( 1 - P\hat{P}V_p \right) - 2C_{ppv}^p \right\} \left[ \frac{1}{n_{10} + n_{11}} \right. \left. + \frac{1}{n_{01} + n_{11}} \right]}, \quad (3.32) \]

and the weighted generalized score statistic for the hypothesis test (3.27) is

\[ T_{WGS}^{NPV} = \frac{\left( N\hat{P}V_1 - N\hat{P}V_2 \right)^2}{\left\{ N\hat{P}V_p \left( 1 - N\hat{P}V_p \right) - 2C_{npv}^p \right\} \left[ \frac{1}{n_{00} + n_{01}} \right. \left. + \frac{1}{n_{00} + n_{10}} \right]}, \quad (3.33) \]

Both statistics have a chi-squared distribution with 1 degree of freedom when the null hypothesis is true, and where

\[ P\hat{P}V_p = \frac{2s_{11} + s_{10} + s_{01}}{2n_{11} + n_{10} + n_{01}} \quad \text{and} \quad N\hat{P}V_p = \frac{2r_{10} + r_{01} + r_{10}}{2n_{00} + n_{01} + n_{10}} \quad (3.34) \]

are the pooled PPV and pooled NPV respectively, and

\[ C_{ppv}^p = \frac{s_{11} \left( 1 - P\hat{P}V_p \right)^2 + r_{11} P\hat{P}V_p^2}{2n_{11} + n_{10} + n_{01}} \quad \text{and} \quad C_{npv}^p = \frac{s_{00} N\hat{P}V_p^2 + r_{00} \left( 1 - N\hat{P}V_p \right)^2}{2n_{00} + n_{01} + n_{10}}. \quad (3.35) \]

CIs for the difference between the two PPVs and the two NPVs are obtained by inverting the contrast statistics (3.32) and (3.33), i.e.

\[ PPV_1 - PPV_2 \in \left( P\hat{P}V_1 - P\hat{P}V_2 \right) \pm z_{1-\alpha/2} \sqrt{P\hat{P}V_p \left( 1 - P\hat{P}V_p \right) - 2C_{ppv}^p \left[ \frac{1}{n_{10} + n_{11}} \right. \left. + \frac{1}{n_{01} + n_{11}} \right]} \quad (3.36) \]

and

\[ NPV_1 - NPV_2 \in \left( N\hat{P}V_1 - N\hat{P}V_2 \right) \pm z_{1-\alpha/2} \sqrt{N\hat{P}V_p \left( 1 - N\hat{P}V_p \right) - 2C_{npv}^p \left[ \frac{1}{n_{00} + n_{01}} \right. \left. + \frac{1}{n_{00} + n_{10}} \right]}. \quad (3.37) \]

Kosinski has demonstrated that his method performed better in terms of the type I error than methods of Leisenring et al (2000) and of Wang et al (2006).
3.2.3.4. Study of Roldán-Nofuentes et al

Roldán-Nofuentes et al (2012) have studied the simultaneous comparison of the PVs of two BDTs in paired design, and these authors have shown that the comparison of the PVs should be performed simultaneously and not independently (as are the previous methods). The simultaneous comparison of the PVs of two BDTs consists of solving the hypothesis test

\[ H_0 : (PPV_1 = PPV_2 \text{ and } NPV_1 = NPV_2) \]
\[ H_1 : (PPV_1 \neq PPV_2 \text{ and/or } NPV_1 \neq NPV_2). \]

The contrast statistics for the hypothesis test \( H_0 : (PPV_1 = PPV_2 \text{ and } NPV_1 = NPV_2) \) is

\[ Q^2 = \hat{\eta}^T \varphi^T \left( \varphi \hat{\Sigma} \varphi^T \right)^{-1} \varphi \hat{\eta} \xrightarrow{n \to \infty} \chi^2_2 \]  

(3.39)

where \( \hat{\eta} = (P\hat{PV}_1, P\hat{PV}_2, N\hat{PV}_1, N\hat{PV}_2)^T \), \( \hat{\Sigma} \) is the estimated variance-covariance matrix of \( \hat{\eta} \) (its elements are shown in equations (3.24)) and \( \varphi = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \) is the design matrix. The statistic \( Q^2 \) (3.39) is distributed asymptotically according to a central chi-square distribution with two degrees of freedom if \( H_0 \) is true. To apply this method it is necessary that all PVs can be estimated and that matrix \( \varphi \hat{\Sigma} \varphi^T \) is non-singular. Therefore, the method cannot be applied if there are many observed frequencies that are equal to zero. If this global hypothesis test is significant to an error rate of \( \alpha \), the investigation of the causes of the significance is carried out by comparing the PPVs and the NPVs independently (for example, applying the method of Kosinski (2013)) and subsequently applying a method of multiple comparisons (method of Holm (1979) or method of Hochberg (1988)) to the same error rate of \( \alpha \). Simulation
experiments performed have shown that samples of between 300 and 500 individuals are necessary in order for the power of the global hypothesis test to be high (≥ 80%).

3.2.4. Weighted kappa coefficients

Bloch (1997) has studied the comparison of the weighted kappa coefficients of two BDTs subject to a paired design. The hypothesis test is

$$H_0 : \kappa_1 (c) = \kappa_2 (c) \ vs \ H_1 : \kappa_1 (c) \neq \kappa_2 (c).$$

(3.40)

In terms of the probabilities of the Table 3.2, the weighted kappa coefficient of Test 1 is

$$\kappa_1 (c) = \frac{(p_{11} + p_{10})(q_{01} + q_{00}) - (p_{01} + p_{00})(q_{10} + q_{11})}{pc \sum_{k=0} (p_{0k} + q_{0k}) + q(1-c) \sum_{k=0} (p_{ik} + q_{ik})}.$$  

(3.41)

and that of Test 2 is

$$\kappa_2 (c) = \frac{(p_{11} + p_{01})(q_{10} + q_{00}) - (p_{10} + p_{00})(q_{11} + q_{11})}{pc \sum_{h=0} (p_{h0} + q_{h0}) + q(1-c) \sum_{h=0} (p_{h1} + q_{h1})}.$$  

(3.42)

As the probabilities $p_{hk}$ and $q_{hk}$ are the probabilities of the multinomial distribution, their estimators are $\hat{p}_{hk} = s_{hk} / n$ and $\hat{q}_{hk} = r_{hk} / n$. By substituting each parameter in the expressions of the weighted kappa coefficients with its estimator, the estimators of the weighted kappa coefficients are

$$\hat{\kappa}_1 (c) = \frac{(s_{11} + s_{10})(r_{01} + r_{00}) - (s_{01} + s_{00})(r_{10} + r_{11})}{sc \sum_{k=0} (s_{0k} + r_{0k}) + r(1-c) \sum_{k=0} (s_{ik} + r_{ik})}.$$  

(3.43)

and
\[ \hat{\kappa}_2(c) = \frac{1}{sc} \sum_{h=0}^{1} \left( s_{h0} + r_{h0} \right) + r \left( 1 - c \right) \sum_{h=0}^{1} \left( s_{h1} + r_{h1} \right). \]  

(3.44)

Finally, the statistic for the hypothesis test (3.40) is

\[ z = \frac{\hat{\kappa}_1(c) - \hat{\kappa}_2(c)}{\sqrt{\text{Var}[\hat{\kappa}_1(c)] + \text{Var}[\hat{\kappa}_2(c)] - 2\text{Cov}[\hat{\kappa}_1(c), \hat{\kappa}_2(c)]}} \xrightarrow{n \to \infty} N(0,1), \quad (3.45) \]

where the expressions of the variances and the covariance have been obtained by Bloch applying the delta method, i.e.

\[ \text{Var}[\hat{\kappa}_1(c)] = \frac{1}{n} \left[ c p \frac{n_{00} + n_{01}}{n} + \left( 1 - c \right) q \frac{n_{11} + n_{10}}{n} \right]^2 \times \]

\( \left[ \hat{q} \left( 1 - \hat{\kappa}_1(c) \right) \frac{n_{00} + n_{01}}{n} \right]^2 \frac{s_{11} + s_{10}}{n} + \left[ \hat{p} \left( 1 - \hat{\kappa}_1(c) \right) \frac{n_{00} + n_{01}}{n} + \left( 1 - c \right) \hat{\kappa}_1(c) \right]^2 \frac{r_{11} + r_{10}}{n} + \]

\[ \left[ \frac{n_{11} + n_{10}}{n} \hat{q} \left( 1 - \hat{\kappa}_1(c) \right) + c \hat{\kappa}_1(c) \right]^2 \frac{s_{01} + s_{00}}{n} + \left[ \frac{n_{11} + n_{10}}{n} \hat{p} \left( 1 - \hat{\kappa}_1(c) \right) \right]^2 \frac{r_{01} + r_{00}}{n} \]. \]

(3.46)

\[ \text{Var}[\hat{\kappa}_2(c)] = \frac{1}{n} \left[ c p \frac{n_{00} + n_{01}}{n} + \left( 1 - c \right) q \frac{n_{11} + n_{10}}{n} \right]^2 \times \]

\( \left[ \hat{q} \left( 1 - \hat{\kappa}_2(c) \right) \frac{n_{00} + n_{01}}{n} \right]^2 \frac{s_{11} + s_{10}}{n} + \left[ \hat{p} \left( 1 - \hat{\kappa}_2(c) \right) \frac{n_{00} + n_{01}}{n} + \left( 1 - c \right) \hat{\kappa}_2(c) \right]^2 \frac{r_{11} + r_{10}}{n} + \]

\[ \left[ \frac{n_{11} + n_{10}}{n} \hat{q} \left( 1 - \hat{\kappa}_2(c) \right) + c \hat{\kappa}_2(c) \right]^2 \frac{s_{01} + s_{00}}{n} + \left[ \frac{n_{11} + n_{10}}{n} \hat{p} \left( 1 - \hat{\kappa}_2(c) \right) \right]^2 \frac{r_{01} + r_{00}}{n} \]. \]

(3.47)

and
Chapter 3. Comparison and combination of two BDTs

\[ \hat{\text{Cov}}[\hat{\kappa}_1(c), \hat{\kappa}_2(c)] = \]

\[ \left( \frac{1}{n} \left[ \hat{c}_p (n_{00} + n_{01}) + (1 - \hat{c}) \hat{q} (n_{11} + n_{10}) \right] \right)^x \]

\[ \left( 1 - \hat{\kappa}_1(c) \right) \hat{\kappa}_2(c) \left[ (1 - \hat{c}) \hat{p} \frac{r_{12}}{n} - (1 - \hat{c}) \hat{p} \left( \frac{r_{11} + r_{10}}{n^2} \right) \right] \]

\[ c \hat{q} \frac{s_{00}}{n} - c \hat{q} \left( \frac{s_{10} + s_{01}}{n^2} \right) \]

\[ \hat{\kappa}_1(c) \hat{\kappa}_2(c) \left[ (1 - \hat{c}) \hat{p} \frac{r_{12}}{n} - (1 - \hat{c}) \hat{p} \left( \frac{r_{11} + r_{10}}{n^2} \right) \right] \]

\[ c \hat{q} \frac{s_{00}}{n} - c \hat{q} \left( \frac{s_{10} + s_{01}}{n^2} \right) \]

\[ \hat{\kappa}_1(c) \hat{\kappa}_2(c) \left[ (1 - \hat{c}) \hat{p} \frac{r_{12}}{n} - (1 - \hat{c}) \hat{p} \left( \frac{r_{11} + r_{10}}{n^2} \right) \right], \] (3.48)

where \( \hat{q} = 1 - \hat{p} \). An approximate CI for \( \kappa_1(c) - \kappa_2(c) \) is obtained by inverting the contrast statistics (3.45), i.e.

\[ \hat{\kappa}_1(c) - \hat{\kappa}_2(c) \pm z_{1 - \alpha/2} \sqrt{\text{Var}[\hat{\kappa}_1(c)] + \text{Var}[\hat{\kappa}_2(c)] - 2 \hat{\text{Cov}}[\hat{\kappa}_1(c), \hat{\kappa}_2(c)]}. \] (3.49)

3.3. Combination of two BDTs

In disease diagnosis it is common to combine two BDTs in order to increase the accuracy of the diagnosis. The combination of two BDTs can be carried out in parallel testing or in serial testing, and in each case, with the AND rule or the OR rule (Zhou et al, 2002). In the parallel testing the two BDTs are applied to all the individuals in a sample, and in the serial testing the application of a BDT depends on the result of the
other BDT. The AND rule implies that the diagnosis is positive if both BDTs are positive and the diagnosis is negative if one of the two BDTs is negative (or both are negative). The OR rule implies that the diagnosis is positive if one of the two BDTs is positive (or both are positive), and the diagnosis is negative if both BDTs are negative. From a clinical perspective (Zhou et al, 2002), the serial testing has the advantage of cost-effectiveness, because only a single test needs to be applied to some of the individuals; however its main disadvantage is the delay in applying the treatment for the disease, because this cannot be started until the result of the second test is obtained (where that is necessary). Marshall (1989) has studied the effect of combining two binary tests on the PVs of the combination and has proposed some criteria for testing when combining increases the value of the PVs. Lin (1999) has studied the effect of the two diagnostic tests on the Se and Sp of the combination of both tests. Macaskill et al (2002) have studied the estimation of the LRs of the combination of two BDTs in parallel testing. In this Chapter we summarize the results of these authors, and we make the second contribution of this thesis: the combination of the weighted kappa coefficients of two BDTs when these are applied in parallel testing with the AND rule or with the OR rule.

3.3.1. Study of Marshall

Marshall (1989) studied the combination of the predictive values of two BDTs in parallel testing with the AND rule and with the OR rule, proposing criteria, and their graphic representations, to determine when there is an increase in the PVs of the combination of both BDTs. The criteria proposed depend on the covariances between the two BDTs. Using the notation followed in Chapters 1 and 2, and supposing that
\( PPV_1 \geq PPV_2 \) and that the covariances \( \alpha_i \) and \( \alpha_0 \) verify \( 1 \leq \alpha_i \leq 1/\max\{Se_1, Se_2\} \) and \( 1 \leq \alpha_0 \leq 1/\max\{(1-Sp_1), (1-Sp_2)\} \), since it is assumed that the association between the two BDTs is positive \( (\alpha_i > 1) \), the regions where there is and is not an increase in the \( PPVs \) of the combination of the two BDTs are:

a) \( R\{P\hat{PV}_{\text{AND}} > P\hat{PV}_i\} = \{\hat{\alpha}_i, \hat{\alpha}_0; \hat{\alpha}_i > \hat{\phi}_1\} \), with \( \hat{\phi}_1 = (1-\hat{Sp}_2)\hat{\alpha}_0/\hat{Se}_2 \).

b) \( R\{P\hat{PV}_{\text{OR}} > P\hat{PV}_i\} = \{\hat{\alpha}_i, \hat{\alpha}_0; \hat{\alpha}_i < \hat{\phi}_2\} \), with

\[
\hat{\phi}_2 = (1-\hat{Sp}_2)\hat{\alpha}_0/\hat{Se}_2 + \left[ (1-\hat{Sp}_1)\hat{Se}_2 - \hat{Se}_1\left(1-\hat{Sp}_2\right) \right] \left[ \hat{Se}_1\left(1-\hat{Sp}_1\right)\hat{Se}_2\right].
\]

c) \( R\{P\hat{PV}_i \geq \max\{P\hat{PV}_{\text{AND}}, P\hat{PV}_{\text{OR}}\}\} = \{\hat{\alpha}_i, \hat{\alpha}_0; \hat{\phi}_1 \leq \hat{\alpha}_i \leq \hat{\phi}_2\} \).

Regarding the \( NPVs \), the regions are:

a) \( R\{N\hat{PV}_{\text{AND}} > N\hat{PV}_i\} = \{\hat{\alpha}_i, \hat{\alpha}_0; \hat{\alpha}_i > \hat{\phi}_1\} \), with

\[
\hat{\phi}_1 = \hat{e}\hat{\alpha}_0 + \left[ \hat{Se}_1 - (1-\hat{Sp}_1) \right] \left[ \hat{Se}_1\hat{Se}_2\hat{Sp}_1 \right]
\]

and

\[
\hat{e} = (1-\hat{Sp}_1)(1-\hat{Se}_1)(1-\hat{Sp}_2) / (\hat{Se}_1\hat{Se}_2\hat{Sp}_1).
\]

b) \( R\{N\hat{PV}_{\text{OR}} > N\hat{PV}_i\} = \{\hat{\alpha}_i, \hat{\alpha}_0; \hat{\alpha}_i < \hat{\phi}_2\} \), with

\[
\hat{\phi}_2 = \hat{e}\hat{\alpha}_0 + \left[ \hat{Se}_1\left(1-\hat{Sp}_2\right) - \hat{Se}_2\left(1-\hat{Sp}_1\right) + \hat{Se}_2 - (1-\hat{Sp}_2) \right] / (\hat{Se}_1\hat{Se}_2\hat{Sp}_1).
\]

c) \( R\{N\hat{PV}_i \geq \max\{N\hat{PV}_{\text{AND}}, N\hat{PV}_{\text{OR}}\}\} = \{\hat{\alpha}_i, \hat{\alpha}_0; \hat{\phi}_1 \leq \hat{\alpha}_i \leq \hat{\phi}_2\} \).
In the previous regions, the estimators of $\alpha_1$ and $\alpha_0$ are $\hat{\alpha}_i = \frac{\hat{p}_i}{\hat{p}S_{e_1}S_{e_2}}$ and

$$\hat{\alpha}_0 = \frac{\hat{q}_{11}}{\hat{q}(1-\hat{S}_{p_1})(1-\hat{S}_{p_2})}.$$ 

### 3.3.2. Study of Lin

Lin (1999) studied the sensitivity and specificity of the combination of two BDTs in parallel testing with the AND rule, discussing the dependence effect between the two BDTs. The results obtained by Lin are obvious and they are obtained from the probabilities given by equations (3.1) and (3.2). With the AND rule, the sensitivity and the specificity of the combination of the two BDTs are

$$Se_{AND} = \frac{p_{11}}{p} \quad \text{and} \quad Sp_{AND} = \frac{q_{10} + q_{01} + q_{00}}{q}. \quad (3.50)$$

Lin also discussed the dependence effect between the two BDTs on the PVs, as well as the estimation of all the parameters (sensitivity, specificity and PVs).

### 3.3.3. Study of Macaskill et al

Macaskill et al (2002) studied the combination of the likelihood ratios of two BDTs (a main test and another adjunct test) in parallel testing with the AND rule and with the OR rule, and compared the combined LRs in relation to those of the main test. In this situation, the estimators of the LRs are

$$\hat{LR}^*_i = \frac{n_{i1}^+n_D^-}{n_1^+n_D^-} \quad \text{and} \quad \hat{LR}^-_i = \frac{(n_{i1}^- + n_{i0}^-)n_D^+}{(n_{i1}^- + n_{i0}^-)n_D^+}. \quad (3.51)$$
for the main test, and

\[
\hat{L}_{R_{\text{AND}}}^{+} = \frac{n_{i1} n_{nD}}{n_{11} n_{D}} \quad \text{and} \quad \hat{L}_{R_{\text{AND}}}^{-} = \frac{\left(n_{i0} + n_{00}\right) n_{D}}{\left(n_{00} + n_{10}\right) n_{D}}
\]  

(3.52)

for the combination of the positive LRs with the AND rule, and

\[
\hat{L}_{R_{\text{OR}}}^{+} = \frac{\left(n_{i1} + n_{00}\right) n_{D}}{\left(n_{11} + n_{01}\right) n_{D}} \quad \text{and} \quad \hat{L}_{R_{\text{OR}}}^{-} = \frac{n_{00} n_{nD}}{n_{0n} n_{10}}
\]  

(3.53)

for the combination of the positive LRs with the OR rule, when \(n_{ij}^{+} \left(n_{ij}^{-}\right)\) is the number of diseased (non-diseased) individuals in which the combination of the two BDTs gives a result \(i\) and the main test gives a result \(j\), with \(i, j = 0, 1\), and \(n_{D} \left(n_{D}\right)\) is the total number of diseased (non-diseased) individuals.

In the situation studied by these authors, there structural zeros, since the combination of the two BDTs is compared in relation to a main test. With the AND rule, the number of diseased (non-diseased) individuals for whom the test is negative and the combination of the two BDTs is positive is equal to 0 \(\left(n_{01}^{-} - n_{01}^{+} = 0\right)\). With the AND rule, in terms of the frequencies in Table 3.1, it is verified that \(n_{11}^{+} = s_{11}, \ n_{10}^{+} = s_{10}, \ n_{00}^{-} = s_{00}, \ n_{11}^{-} = r_{11}, \ n_{10}^{-} = r_{10} \) and \(n_{00}^{+} = r_{01} + r_{00}\). With the OR rule, the number of diseased (non-diseased) individuals for whom the test is positive and the combination of the two BDTs is negative is equal to 0 \(\left(n_{10}^{+} - n_{10}^{-} = 0\right)\). With the OR rule, in terms of the frequencies in Table 3.1, it is verified that \(n_{11}^{+} = s_{11} + s_{10}, \ n_{01}^{+} = s_{01}, \ n_{00}^{+} = s_{00}, \ n_{11}^{-} = r_{11} + r_{10}, \ n_{01}^{-} = r_{01} \) and \(n_{00}^{-} = r_{00}\). Macaskill et al proposed CIs for \(LR_{\text{AND}}/LR_{1}\) and \(LR_{\text{OR}}/LR_{1}\), i.e.
Chapter 3. Comparison and combination of two BDTs

\[
\frac{LR_{\text{AND}}}{LR_i} \in \exp \left[ \log \left( \hat{L}_{R_{\text{AND}}} \right) - \log \left( \hat{L}_R \right) \right] \pm z_{1-\alpha/2} \sqrt{\text{Var} \left( \log \left( \hat{L}_{R_{\text{AND}}} \right) - \log \left( \hat{L}_R \right) \right)}
\]

(3.54)

and

\[
\frac{LR_{\text{OR}}}{LR_i} \in \exp \left[ \log \left( \hat{L}_{R_{\text{OR}}} \right) - \log \left( \hat{L}_R \right) \right] \pm z_{1-\alpha/2} \sqrt{\text{Var} \left( \log \left( \hat{L}_{R_{\text{OR}}} \right) - \log \left( \hat{L}_R \right) \right)}
\]

(3.55)

where \( LR \) is \( LR^+ \) or \( LR^- \), and the variances are estimated applying the delta method. If the interval is greater than 1 (the lower limit if greater than 1), then the combination of the two BDTs with the corresponding rule is better than the main test.

3.3.4. Combination of two weighted kappa coefficients

Let us consider two BDTs, Test 1 and Test 2, which are evaluated in relation of the same GS. Let \( T_h \) be the random variable which models the result of the \( h \)th BDT, in such a way that \( T_h = 1 \) when the result is positive and \( T_h = 0 \) when it is negative; and let \( D \) be the random variable which models the result of the GS, so that \( D = 1 \) when the individual does have the disease and \( D = 0 \) when the individual does not have the disease. Let \( Se_h \) and \( Sp_h \) be the sensitivity and specificity of the \( h \)th BDT. When both BDTs are applied in parallel testing the probabilities given in Table 3.2 are obtained. In Table 3.3 the associated losses when the AND rule or the OR rule are used are shown. In this situation the observed frequencies (Table 3.1) are the product of a multinomial distribution with probabilities \( \pi = \left( p_{11}, p_{10}, p_{01}, p_{00}, q_{11}, q_{10}, q_{01}, q_{00} \right)^T \), verifying that

\[
p = \sum_{h,k} p_{hk} , \quad q = \sum_{h,k} q_{hk} \quad \text{and} \quad p + q = 1 ,
\]

where \( p \) is the prevalence of the disease. In the following the weighted kappa coefficient for the combination of the two BDTs is
defined and the conditions under which the combination of both BDTs increases the combined weighted kappa coefficient are studied.

### Table 3.3. Losses in parallel testing.

<table>
<thead>
<tr>
<th></th>
<th>$T_1 = 1$</th>
<th>$T_1 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_2 = 1$</td>
<td>$T_2 = 0$</td>
<td>$T_2 = 1$</td>
</tr>
<tr>
<td>$D = 1$</td>
<td>0</td>
<td>$L$</td>
</tr>
<tr>
<td>$D = 0$</td>
<td>$L'$</td>
<td>0</td>
</tr>
</tbody>
</table>

### 3.3.4.1. Combined weighted kappa coefficient

In parallel testing with the AND rule the combined weighted kappa coefficient is

$$
\kappa_{\text{AND}}(c) = \frac{p_{11}q_{11}p}{p(1-Q_{\text{AND}})c + qQ_{\text{AND}}(1-c)},
$$

where $Q_{\text{AND}} = p_{11} + q_{11}$ is the probability that the combination of the two BDTs will be positive and $1 - Q_{\text{AND}} = 1 - p_{11} - q_{11}$ is the probability that it will be negative. With the OR rule the combined weighted kappa coefficient is
Chapter 3. Comparison and combination of two BDTs

\[ \kappa_{OR}(c) = \frac{q_{00}p - p_{00}q}{p(1-Q_{OR})c + qQ_{OR}(1-c)}, \]  

(3.57)

where \( Q_{OR} = 1 - p_{00} - q_{00} \) is the probability that the combination of the two BDTs will be positive and \( 1 - Q_{OR} = p_{00} + q_{00} \) that it will be negative. The proofs are shown in Appendix 3.1. Expressions (3.56) and (3.57) also can be written in terms of sensitivities and specificities by replacing in these equations \( p_{ij} \) and \( q_{ij} \) by the expressions (3.1) and (3.2) respectively. Then

\[ \kappa_{AND}(c) = \frac{pqY_{AND}}{pc + (q-c)[\alpha_{AND}Se_{AND}Se_{AND} - qY_{AND}]} \]

(3.58)

where \( Y_{AND} = Se_{AND} + Sp_{AND} - 1 \) is the Youden index of the combination of the two BDTs in parallel testing with the AND rule, and \( Se_{AND} = p_{11}/p \) and \( Sp_{AND} = (q_{10} + q_{01} + q_{00})/q \) are the sensitivity and the specificity of the combination of the two BDTs. For the OR rule,

\[ \kappa_{OR}(c) = \frac{pqY_{OR}}{pc + (q-c)(Se_{OR} + Se_{OR} - \alpha_{OR}Se_{OR}Se_{OR} - qY_{OR})} \]

(3.59)

where \( Y_{OR} = Se_{OR} + Sp_{OR} - 1 = Y_1 + Y_2 - Y_{AND} \) is the Youden index in parallel testing with the OR rule, and where \( Se_{OR} = (p_{11} + p_{10} + p_{01})/p \) and \( Sp_{OR} = q_{00}/q \) are the sensitivity and specificity of the combination of the two BDTs, and \( Y_h = Se_h + Sp_h - 1 \) is the Youden index of the \( h \)th BDT.

If \( \kappa_{c}(c) \) is \( \kappa_{AND}(c) \) or \( \kappa_{OR}(c) \) depending on the rule used, then the combined weighted kappa coefficient \( \kappa_{c}(c) \) verifies the following properties:
a) \(-1 \leq \kappa_i(c) \leq 1\), however as every \(BDDT\) must have a Youden index between 0 and 1 \((0 < Y_h < 1)\) then \(0 \leq \kappa_i(c) \leq 1\).

b) \(\kappa_i(c)\) is an function increasing in \(c\) if \(Q_p > p\), decreasing if \(Q_p < p\) and it is a constant function equal to the combined Youden index \(Y_i\) if \(Q_p = p\), where \(Q_p\) is equal to \(Q_{AND}\) or \(Q_{OR}\), and \(Y_i\) is equal to \(Y_{AND}\) or \(Y_{OR}\), respectively.

These properties are similar to those of the weighted kappa coefficient of a \(BDDT\). The object pursued when two \(BDDTs\) are combined is to increase the accuracy of the combination of both tests. In the following, the conditions under which the combination in parallel testing increases the value of the weighted kappa coefficient of the combination of both \(BDDTs\) are studied, both with the \(AND\) rule and with the \(OR\) rule.

Let us consider that the combination in parallel testing of the two \(BDDTs\) increases the value of the combined weighted kappa coefficient if it is verified that \(\kappa_i(c) > \max\{\kappa_1(c), \kappa_2(c)\}\) for a fixed value of the weighting index \(c\). If this is not the case, \(\kappa_i(c) \leq \max\{\kappa_1(c), \kappa_2(c)\}\) for a fixed value of \(c\), the combination of the \(BDDTs\) does not increase the value of the weighted kappa coefficient of the combination. In this situation, the combination of the two \(BDDTs\) should not be carried out for this value of \(c\), because it does not increase the value of \(\kappa_i(c)\).

In all that follows we shall consider that if the subindex \(i\) is equal to 1 (or 2) then the subindex \(j\) is equal to 2 (or 1). Moreover it is assumed that the Youden indices of the two \(BDDTs\) are between 0 and 1 \((0 < Y_h < 1, \ h=1,2)\). The situation in which a Youden index (or both) is equal to 1 is not contemplated, since the weighted kappa coefficient of the corresponding \(BDDT\) is always equal to 1, and then it is always verified that
\( \kappa_{\text{AND}}(c) \leq \kappa_i(c) \) and \( \kappa_{\text{OR}}(c) \leq \kappa_i(c) \) (an increase in the combined weighted kappa coefficient is never produced).

### 3.3.4.1.1. Increase with the AND rule

Let \( c \) be the weighting index fixed by the clinician and let

\[
c_h = \frac{q(Q_{\text{AND}}^h - Q_{\text{AND}}^i) - Q_{\text{AND}}^h(p - Q_{\text{AND}}^j)}{Y_h(p - Q_{\text{AND}}^j) - Y_{\text{AND}}^h(p - Q_{\text{AND}}^h)}
\]

with \( h = 1,2 \). Then the combination of the two BDTs with the AND rule increases the value of the combined weighted kappa coefficient for the weighting index \( c \), that is \( \kappa_{\text{AND}}(c) > \max \{ \kappa_i(c), \kappa_j(c) \} \), if one of the two following conditions is verified:

a) \( 0 < c < \min \{ c_1, c_2 \} \), with \( 0 < c_1 < 1 \) and \( 0 < c_2 < 1 \). Moreover, it is always verified that \( \kappa_{\text{AND}}(1) < \min \{ \kappa_i(1), \kappa_j(1) \} \) and \( \kappa_{\text{AND}}(0) > \max \{ \kappa_i(0), \kappa_j(0) \} \).

b) \( 0 \leq c < c_i < 1 \), \( c_j > 1 \) (or \( c_j < 0 \)) and \( Y_j < Y_{\text{AND}} \). In this situation it is verified that \( \kappa_{\text{AND}}(c) > \kappa_i(c) \) for \( 0 \leq c < c_i < 1 \), and \( \kappa_{\text{AND}}(c) > \kappa_j(c) \) for \( 0 \leq c \leq 1 \).

In any other situation that is different to the previous ones, the combination of the two BDTs in parallel testing with the AND rule never increases the value of \( \kappa_{\text{AND}}(c) \) for any value of \( c \). The proofs of these results may be seen in Appendix 3.2.

In Figure 3.1 two graphics are shown for \( \kappa_{\text{AND}}(c) \) and for different values of sensitivities, specificities, prevalence and covariances. Graphic 1 corresponds to the situation in which \( c_1 \) and \( c_2 \) are between 0 and 1, and hence \( \kappa_{\text{AND}}(c) \) is greater than \( \kappa_i(c) \) and \( \kappa_j(c) \) for \( 0 \leq c < \min \{ c_1, c_2 \} = 0.583 \); and in Graphic 2 it is verified that \( c_i \)
is between 0 and 1, $c_2$ is greater than 1 and $Y_2 < Y_{AND}$, and therefore $\kappa_{AND}(c) > \kappa_1(c)$ for $0 \leq c < 0.095$ and $\kappa_{AND}(c) > \kappa_2(c)$ for any value of $c \in [0,1]$.

### Figure 3.1. Examples of combined weighted kappa coefficients with the AND rule.

<table>
<thead>
<tr>
<th>Graphic 1</th>
<th>Graphic 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se_1 = 0.75  Sp_1 = 0.85  Se_2 = 0.85  Sp_2 = 0.90  $p = 10%$</td>
<td>Se_1 = 0.90  Sp_1 = 0.70  Se_2 = 0.60  Sp_2 = 0.80  $p = 50%$</td>
</tr>
<tr>
<td>$\alpha_1$ = 1.09  $\alpha_6$ = 3.83  $Se_{AND} = 0.695$  $Sp_{AND} = 0.943$</td>
<td>$\alpha_4$ = 1.09  $\alpha_6$ = 2.87  $Se_{AND} = 0.589$  $Sp_{AND} = 0.828$</td>
</tr>
<tr>
<td>$Y_1 = 0.60$  $Y_2 = 0.75$  $Y_{AND} = 0.637$</td>
<td>$Y_1 = 0.60$  $Y_2 = 0.40$  $Y_{AND} = 0.417$</td>
</tr>
<tr>
<td>$c_1 = 0.959$  $c_2 = 0.583$</td>
<td>$c_1 = 0.095$  $c_2 = 1.161$</td>
</tr>
</tbody>
</table>

#### 3.3.4.1.2. Increase with the OR rule

Let $c$ be the weighting index fixed by the clinician and let

$$c_h' = \frac{q(Y_{OR} - Q_{OR})}{Y_h(p - Q_{OR}) - Y_{OR}(p - Q_h)}$$

with $h = 1, 2$. Then the combination in parallel testing with the OR rule increases the value of the combined weighted kappa coefficient for the weighting index $c$, that is $\kappa_{OR}(c) > \max \{\kappa_1(c), \kappa_2(c)\}$, if one of the two following conditions is verified:
Chapter 3. Comparison and combination of two BDTs

a) \( \max \{c'_1, c'_2\} < c < 1 \), with \( 0 < c'_1 < 1 \) and \( 0 < c'_2 < 1 \). Moreover, it is always verified that \( \kappa_{OR}(0) < \min \{\kappa_1(0), \kappa_2(0)\} \) and \( \kappa_{OR}(1) > \max \{\kappa_1(1), \kappa_2(1)\} \).

b) \( 0 \leq c'_1 < c < 1 \), \( c'_1 > 1 \) (or \( c'_j < 0 \)) and \( Y_j < Y_{OR} \). In this situation it is verified that \( \kappa_{OR}(c) > \kappa_j(c) \) for \( 0 \leq c < c'_1 < 1 \), and \( \kappa_{OR}(c) > \kappa_j(c) \) for \( 0 \leq c \leq 1 \).

In any other situation that is different to the previous ones, the combination of the two BDTs in parallel with the OR rule never increases the value of \( \kappa_{OR}(c) \) for any value of \( c \). The proofs of these results may be seen in Appendix 3.3.

In Figure 3.2 two graphics for \( \kappa_{OR}(c) \) are shown, and for different values of sensitivities, specificities, prevalence and covariances. Graphic 1 corresponds to the situation in which \( c'_1 \) and \( c'_2 \) are between 0 and 1, and therefore \( \kappa_{OR}(c) \) is greater than \( \kappa_1(c) \) and \( \kappa_2(c) \) for \( \max \{c'_1, c'_2\} = 0.423 < c \leq 1 \); and in Graphic 2 it is verified that \( c'_1 \) is between 0 and 1, \( c'_2 \) is greater than 1 and \( Y_2 < Y_{OR} \), so that \( \kappa_{OR}(c) > \kappa_1(c) \) for \( 0.972 < c \leq 1 \) and \( \kappa_{OR}(c) > \kappa_2(c) \) for any value of \( c \in [0,1] \).
Figure 3.2. Examples of combined weighted kappa coefficients with the OR rule.

**Graphic 1**

- $S_{e_1} = 0.80$  $S_{p_1} = 0.95$  $S_{e_2} = 0.85$  $S_{p_2} = 0.90$  $p = 50\%$
- $\alpha_1 = 1.14$  $\alpha_2 = 8.2$  $S_{c_{or}} = 0.875$  $S_{p_{or}} = 0.891$
- $Y_1 = 0.75$  $Y_2 = 0.75$  $Y_{or} = 0.766$
- $c'_1 = 0.423$  $c'_2 = 0.198$

**Graphic 2**

- $S_{e_1} = 0.75$  $S_{p_1} = 0.95$  $S_{e_2} = 0.80$  $S_{p_2} = 0.70$  $p = 10\%$
- $\alpha_1 = 1.13$  $\alpha_2 = 2.17$  $S_{c_{or}} = 0.872$  $S_{p_{or}} = 0.683$
- $Y_1 = 0.70$  $Y_2 = 0.50$  $Y_{or} = 0.555$
- $c'_1 = 0.972$  $c'_2 = 3.147$

### 3.3.4.2. Estimation and CIs

By substituting each parameter in the expressions of the combined weighted kappa coefficients with its estimator, the estimators of the combined weighted kappa coefficients are obtained, i.e.

\[
\hat{\kappa}_{AND}(c) = \frac{s_{11}r - r_{1}s}{s(n-s_{11} - r_{11})c + r(s_{11} + r_{11})(1-c)}
\]

(3.60)

and

\[
\hat{\kappa}_{OR}(c) = \frac{r_{00}s - s_{00}r}{s(s_{00} + r_{00})c + r(n-s_{00} - r_{00})(1-c)}.
\]

(3.61)
Their asymptotic variances are obtained by applying the delta method. As \( \pi \) is the vector of probabilities of a multinomial distribution, then \( \Sigma_x = \{ \text{diag}(\pi) - \pi\pi^T \}/n \) and applying the delta method we obtain that

\[
\text{Var}[\hat{\kappa}_c(c)] = \left( \frac{\partial \kappa_c(c)}{\partial \pi} \right)^T \Sigma_x \left( \frac{\partial \kappa_c(c)}{\partial \pi} \right),
\]

(3.62)

where \( \hat{\kappa}_c(c) \) is \( \hat{\kappa}_\text{AND}(c) \) or \( \hat{\kappa}_\text{OR}(c) \). By realizing the algebraic operations and substituting each parameter by its estimator, the expressions of the estimated variance of \( \hat{\kappa}_c(c) \) is obtained. These are long and complex expressions that require the use of software for their calculation.

In Section 3.3.4.1 the situations in which the combination of the two BDTs produce an increase of the combined weighted kappa coefficient have been analysed. These conditions studied are theoretical (for the parameters), so that its practical application does not guarantee that the combination of the two BDTs produces and increase in the combined weighted kappa coefficient. Thus, in practice it is necessary to study the increase in \( \kappa_c(c) \) using CIs. Therefore, once the two BDTs have been applied to all the \( n \) individuals of a random sample and the value of the weighting index \( c \) has been fixed by the clinician, it is necessary to study if the combination of the two BDTs produce such an increase. For this we propose a CI for \( \theta_h = \kappa_c(c)/\kappa_h(c) \) with \( h = 1, 2 \). If the CI is greater than 1 (that is, the lower limit is greater than 1) then \( \kappa_c(c) \) is (to the fixed confidence) larger than \( \kappa_h(c) \). If this happens for \( \theta_1 \) and \( \theta_2 \), then \( \kappa_c(c) \) is greater that \( \kappa_1(c) \) and that \( \kappa_2(c) \), and the combination of the two BDTs increases the value of the combined weighted kappa coefficient. In the case that a CI (or both) contains the value
1 (or is lower than the said value), then the combination of the two BDTs does not increase the value of the combined weighted kappa coefficient. In the following, a CI for $\theta_{\hat{h}}$ is proposed applying the Fieller method (1940). The Fieller method is a method that has been used for calculating a CI for the ratio of two parameters. For this, we shall base it on the approximation of the estimators to the normal bivariate distribution, i.e.

$$\begin{pmatrix} \hat{\kappa}_1(c), & \hat{\kappa}_h(c) \end{pmatrix}^T \xrightarrow{n \to \infty} N(0, \Sigma_h),$$

where the matrix

$$\Sigma_h = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

is estimated by applying the delta method, i.e.

$$\Sigma_h = \left[ \frac{\partial (\kappa_1(c), \kappa_h(c))}{\partial \pi} \right] \Sigma \left[ \frac{\partial (\kappa_1(c), \kappa_h(c))}{\partial \pi} \right]^T,$$

where $\sigma_{11} = \text{Var} [\kappa_1(c)]$, $\sigma_{12} = \sigma_{21} = \text{Cov} [\kappa_1(c), \kappa_h(c)]$ and $\sigma_{22} = \text{Var} [\kappa_h(c)]$. By applying the Fieller method, it is verified that

$$\hat{\kappa}_1(c) - \theta_{\hat{h}} \hat{\kappa}_h(c) \xrightarrow{n \to \infty} N(0, \sigma_{11} - 2\theta_{\hat{h}} \sigma_{12} + \theta_{\hat{h}}^2 \sigma_{22}),$$

The Fieller CI is obtained by searching for the set of values for $\theta_{\hat{h}}$ that satisfy the inequality

$$\frac{[\hat{\kappa}_1(c) - \theta_{\hat{h}} \hat{\kappa}_h(c)]^2}{\hat{\sigma}_{11} - 2\theta_{\hat{h}} \hat{\sigma}_{12} + \theta_{\hat{h}}^2 \hat{\sigma}_{22}} < z_{1-\alpha/2}^2,$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution. By resolving the equation (3.64) one obtains that the CI for $\theta_{\hat{h}} = \kappa_1(c)/\kappa_h(c)$ is

$$\theta_{\hat{h}} \in \frac{\hat{\Delta}_3 \pm \sqrt{\hat{\Delta}_3^2 - \hat{\Delta}_2 \hat{\Delta}_4}}{\hat{\Delta}_3},$$
where $\hat{\Delta}_1 = \hat{\kappa}_i(c)\hat{\kappa}_h(c) - \hat{\sigma}_{12}z_{1-a/2}, \hat{\Delta}_2 = \hat{\kappa}_i(c) - \hat{\sigma}_{11}z_{1-a/2}$ and $\hat{\Delta}_3 = \hat{\kappa}_h(c) - \hat{\sigma}_{22}z_{1-a/2}$. This interval is valid when $\hat{\Delta}_1^2 > \hat{\Delta}_2\hat{\Delta}_3$ and $\hat{\Delta}_3 \neq 0$.

On the other hand, if the combination of the two BDTs increases the value of the combined weighted kappa coefficient then it is of interest to calculate a CI for this parameter. As the values of $\kappa_{AND}(c)$ and of $\kappa_{OR}(c)$ are between 0 and 1, the logit transformations can be applied. The logit transformation of $\hat{\kappa}_i(c)$ is closer to a normal distribution with mean $\logit[\kappa_i(c)]$. Thus, the $100(1-\alpha)%$ CI for $\logit[\kappa_i(c)]$ is $\logit[\hat{\kappa}_i(c)] \pm z_{1-a/2}\sqrt{\text{Var}[\logit(\hat{\kappa}_i(c))]}$, where $\text{Var}[\logit(\hat{\kappa}_i(c))]$ is obtained by applying the delta method. Finally the $100(1-\alpha)%$ logit CI for $\kappa_i(c)$ is

$$
\kappa_i(c) \in \frac{\exp\left\{\logit(\hat{\kappa}_i(c)) \pm z_{1-a/2}\sqrt{\text{Var}[\logit(\hat{\kappa}_i(c))]}\right\}}{1 + \exp\left\{\logit(\hat{\kappa}_i(c)) \pm z_{1-a/2}\sqrt{\text{Var}[\logit(\hat{\kappa}_i(c))]}\right\}}.
$$

### 3.3.4.3. Simulation experiments

Monte Carlo simulation experiments have been carried out for studying the coverage probabilities of the proposed CIs in the previous Section. For this, 10000 random samples of multinomial distributions have been generated, whose probabilities have been calculated from expressions (3.1) and (3.2), and with the sizes $n = \{100, 200, 300, 400, 500, 1000\}$. The random samples have been generated from values of weighted kappa coefficients, and not setting the values of Se and of Sp, from equations (1.11) and (1.12), i.e.
in the following way. As disease prevalence we have taken the values 10\%, 25\% and 50\%; as weighting index \(c\) we have taken values 0.1, 0.5 and 0.9; and for each one of the two BDTs we have taken as \(\kappa_h(0)\) and \(\kappa_h(1)\) the values \(\{0.10, 0.20, ..., 0.80, 0.90\}\), with \(h = 1, 2\). Once the values of \(p\), \(\kappa_h(0)\) and \(\kappa_h(1)\) were set, the values of \(Se\) and of \(Sp\) of each BDT were calculated solving (through the Newton-Raphson method) the system formed by equations (1.11) and (1.12). Finally, the value of the weighted kappa coefficient was calculated with equation (1.10) (or (1.13)), i.e.

\[
\kappa_h(c) = \frac{pq(Se_h + Sp_h - 1)}{p(1 - Q_h)c + qQ_h(1 - c)},
\]

where \(Q_h = pSe_h + q(1 - Sp_h)\). As values of each weighted kappa coefficient only the values \(\kappa_h(c) = \{0.2, 0.4, 0.6, 0.8\}\) have been considered. Therefore, considering the classification of the values of the weighted kappa coefficient given by Cicchetti (2001), values of \(\kappa_h(c)\) were considered with different levels of clinical significance: poor \((\kappa_h(c) < 0.40)\), fair \((0.40 \leq \kappa_h(c) \leq 0.59)\), good \((0.60 \leq \kappa_h(c) \leq 0.74)\) and excellent \((0.75 \leq \kappa_h(c) \leq 1)\). As covariances \(\alpha_i\) and \(\alpha_0\) we have taken intermediate and high values. All the samples have been generated in such a way that in all of them it has been possible to estimate all the parameters and their variances-covariances. For the whole study the confidence level has been taken as 95\%.

In Table 3.4 (Fieller CI) the coverage probabilities and the average lengths of the CI of \(\kappa_{AND}(c)/\kappa_i(c)\) are shown, for different values of \(\kappa_1(c)\) and \(\kappa_2(c)\), indicating in
Chapter 3. Comparison and combination of two BDTs

each case the values of \( p, S_{e_i}, S_{p_i}, \alpha_i \) and \( \alpha_0 \) with those with which we have calculated the value of \( \kappa_1(c), \kappa_2(c) \) and \( \kappa_{AND}(c) \). We show situations in which there is an increase of \( \kappa_{AND}(c) \) and situations in which there is no an increase of \( \kappa_{AND}(c) \) for the values \( c = \{0.1, 0.5, 0.9\} \). The results for the CI of \( \kappa_{AND}(c)/\kappa_2(c) \) are not shown because they are very similar to those obtained for \( \kappa_{AND}(c)/\kappa_1(c) \). In general terms the Fieller CI presents a coverage probability that fluctuates around 95\%, although the coverage probability may overestimate the desired coverage of 95\% when the sample size is relatively small \( (n = 100-200) \). With regard to the CI for \( \kappa_{AND}(c) \) the results are shown in Table 3.4 (Logit CI). The scenarios are the same as for the Fieller CI. In general terms, the coverage probabilities of the logit CI fluctuate around 95\%, although they may overestimate the desired coverage of 95\% when the sample size is 100-200.

In Table 3.5 the coverage probabilities and the average lengths of the CI of \( \kappa_{OR}(c)/\kappa_1(c) \) (Fieller CI) and of the CI of \( \kappa_{OR}(c) \) (Logit CI) are shown, for different values of all parameters. We show situations in which there is an increase of \( \kappa_{OR}(c) \) and situations in which there is no an increase of \( \kappa_{OR}(c) \) for \( c = \{0.1, 0.5, 0.9\} \). The results for the CI of \( \kappa_{OR}(c)/\kappa_2(c) \) are not shown because they are very similar to those obtained for \( \kappa_{OR}(c)/\kappa_1(c) \). In general terms the Fieller CI of \( \kappa_{OR}(c)/\kappa_1(c) \) and the Logit CI of \( \kappa_{OR}(c) \) have a behaviour very similar to the CIs of \( \kappa_{AND}(c)/\kappa_1(c) \) and of \( \kappa_{AND}(c) \).
Table 3.4. Coverage probabilities and average lengths of the CI for $\kappa_{\AND}(c)/\kappa_{i}(c)$ and of the CI for $\kappa_{\AND}(c)$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.941</td>
<td>1.357</td>
<td>0.987</td>
<td>0.639</td>
<td>0.978</td>
<td>1.702</td>
<td>0.980</td>
<td>0.485</td>
<td>0.985</td>
<td>0.795</td>
<td>0.989</td>
<td>0.522</td>
</tr>
<tr>
<td>200</td>
<td>0.951</td>
<td>0.877</td>
<td>0.969</td>
<td>0.528</td>
<td>0.975</td>
<td>0.843</td>
<td>0.974</td>
<td>0.370</td>
<td>0.981</td>
<td>0.440</td>
<td>0.978</td>
<td>0.403</td>
</tr>
<tr>
<td>300</td>
<td>0.959</td>
<td>0.719</td>
<td>0.959</td>
<td>0.458</td>
<td>0.953</td>
<td>0.594</td>
<td>0.956</td>
<td>0.306</td>
<td>0.963</td>
<td>0.318</td>
<td>0.961</td>
<td>0.338</td>
</tr>
<tr>
<td>400</td>
<td>0.957</td>
<td>0.619</td>
<td>0.958</td>
<td>0.407</td>
<td>0.950</td>
<td>0.491</td>
<td>0.952</td>
<td>0.268</td>
<td>0.961</td>
<td>0.260</td>
<td>0.959</td>
<td>0.296</td>
</tr>
<tr>
<td>500</td>
<td>0.947</td>
<td>0.552</td>
<td>0.958</td>
<td>0.370</td>
<td>0.947</td>
<td>0.429</td>
<td>0.956</td>
<td>0.241</td>
<td>0.956</td>
<td>0.226</td>
<td>0.957</td>
<td>0.267</td>
</tr>
<tr>
<td>1000</td>
<td>0.946</td>
<td>0.385</td>
<td>0.953</td>
<td>0.269</td>
<td>0.950</td>
<td>0.292</td>
<td>0.944</td>
<td>0.172</td>
<td>0.948</td>
<td>0.155</td>
<td>0.954</td>
<td>0.191</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.947</td>
<td>0.494</td>
<td>0.969</td>
<td>0.435</td>
<td>0.971</td>
<td>1.065</td>
<td>0.976</td>
<td>0.367</td>
<td>0.978</td>
<td>0.328</td>
<td>0.939</td>
<td>0.348</td>
</tr>
<tr>
<td>200</td>
<td>0.951</td>
<td>0.293</td>
<td>0.953</td>
<td>0.319</td>
<td>0.957</td>
<td>0.525</td>
<td>0.962</td>
<td>0.265</td>
<td>0.971</td>
<td>0.185</td>
<td>0.959</td>
<td>0.224</td>
</tr>
<tr>
<td>300</td>
<td>0.955</td>
<td>0.226</td>
<td>0.957</td>
<td>0.264</td>
<td>0.944</td>
<td>0.408</td>
<td>0.952</td>
<td>0.218</td>
<td>0.960</td>
<td>0.141</td>
<td>0.961</td>
<td>0.174</td>
</tr>
<tr>
<td>400</td>
<td>0.957</td>
<td>0.193</td>
<td>0.953</td>
<td>0.230</td>
<td>0.954</td>
<td>0.347</td>
<td>0.945</td>
<td>0.190</td>
<td>0.958</td>
<td>0.119</td>
<td>0.956</td>
<td>0.148</td>
</tr>
<tr>
<td>500</td>
<td>0.946</td>
<td>0.169</td>
<td>0.948</td>
<td>0.207</td>
<td>0.951</td>
<td>0.305</td>
<td>0.957</td>
<td>0.170</td>
<td>0.953</td>
<td>0.104</td>
<td>0.957</td>
<td>0.131</td>
</tr>
<tr>
<td>1000</td>
<td>0.949</td>
<td>0.118</td>
<td>0.955</td>
<td>0.147</td>
<td>0.958</td>
<td>0.211</td>
<td>0.947</td>
<td>0.121</td>
<td>0.935</td>
<td>0.073</td>
<td>0.941</td>
<td>0.092</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
<td>Fieller CI</td>
<td></td>
<td>logit CI</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.943</td>
<td>2.519</td>
<td>0.966</td>
<td>0.362</td>
<td>0.939</td>
<td>0.249</td>
<td>0.957</td>
<td>0.275</td>
<td>0.980</td>
<td>0.307</td>
<td>0.935</td>
<td>0.317</td>
</tr>
<tr>
<td>200</td>
<td>0.941</td>
<td>0.997</td>
<td>0.958</td>
<td>0.265</td>
<td>0.942</td>
<td>0.166</td>
<td>0.956</td>
<td>0.195</td>
<td>0.972</td>
<td>0.172</td>
<td>0.951</td>
<td>0.201</td>
</tr>
<tr>
<td>300</td>
<td>0.944</td>
<td>0.746</td>
<td>0.953</td>
<td>0.218</td>
<td>0.947</td>
<td>0.133</td>
<td>0.953</td>
<td>0.160</td>
<td>0.961</td>
<td>0.131</td>
<td>0.960</td>
<td>0.156</td>
</tr>
<tr>
<td>400</td>
<td>0.947</td>
<td>0.620</td>
<td>0.946</td>
<td>0.190</td>
<td>0.951</td>
<td>0.114</td>
<td>0.958</td>
<td>0.139</td>
<td>0.959</td>
<td>0.111</td>
<td>0.956</td>
<td>0.133</td>
</tr>
<tr>
<td>500</td>
<td>0.945</td>
<td>0.541</td>
<td>0.951</td>
<td>0.171</td>
<td>0.948</td>
<td>0.101</td>
<td>0.941</td>
<td>0.124</td>
<td>0.956</td>
<td>0.098</td>
<td>0.952</td>
<td>0.118</td>
</tr>
<tr>
<td>1000</td>
<td>0.956</td>
<td>0.372</td>
<td>0.941</td>
<td>0.121</td>
<td>0.952</td>
<td>0.071</td>
<td>0.958</td>
<td>0.088</td>
<td>0.947</td>
<td>0.069</td>
<td>0.948</td>
<td>0.083</td>
</tr>
</tbody>
</table>

Cov.: coverage probability. Leng.: average length.
Table 3.5. Coverage probabilities and average lengths of the CI for $\kappa_{OR}(c)/\kappa(c)$ and of the CI for $\kappa_{OR}(c)$.

<table>
<thead>
<tr>
<th>$\kappa_i(0.1)$</th>
<th>$\kappa_i(0.5)$</th>
<th>$\kappa_i(0.9)$</th>
<th>$\kappa_{OR}(0.1)$</th>
<th>$\kappa_{OR}(0.5)$</th>
<th>$\kappa_{OR}(0.9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.6$</td>
<td>$0.4$</td>
<td>$0.8$</td>
<td>$0.392$</td>
<td>$0.474$</td>
<td>$0.828$</td>
</tr>
<tr>
<td>$0.4$</td>
<td>$0.5$</td>
<td>$0.8$</td>
<td>$0.4$</td>
<td>$0.5$</td>
<td>$0.4$</td>
</tr>
<tr>
<td>$0.8$</td>
<td>$0.9$</td>
<td>$0.9$</td>
<td>$0.3368$</td>
<td>$0.82$</td>
<td>$0.98$</td>
</tr>
<tr>
<td>$0.28$</td>
<td>$0.92$</td>
<td>$0.96$</td>
<td>$0.953$</td>
<td>$1.538$</td>
<td>$7.171$</td>
</tr>
<tr>
<td>$1.281$</td>
<td>$6.75$</td>
<td>$13.0$</td>
<td>$1.096$</td>
<td>$0.98$</td>
<td>$0.96$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.943</td>
<td>0.975</td>
<td>0.977</td>
<td>0.442</td>
<td>0.980</td>
<td>5.117</td>
<td>0.985</td>
<td>0.490</td>
<td>0.966</td>
<td>3.121</td>
<td>0.979</td>
<td>0.425</td>
</tr>
<tr>
<td>200</td>
<td>0.946</td>
<td>0.456</td>
<td>0.967</td>
<td>0.334</td>
<td>0.971</td>
<td>3.323</td>
<td>0.983</td>
<td>0.393</td>
<td>0.938</td>
<td>1.070</td>
<td>0.974</td>
<td>0.297</td>
</tr>
<tr>
<td>300</td>
<td>0.947</td>
<td>0.335</td>
<td>0.957</td>
<td>0.278</td>
<td>0.953</td>
<td>1.331</td>
<td>0.960</td>
<td>0.333</td>
<td>0.945</td>
<td>0.794</td>
<td>0.961</td>
<td>0.239</td>
</tr>
<tr>
<td>400</td>
<td>0.953</td>
<td>0.283</td>
<td>0.951</td>
<td>0.242</td>
<td>0.943</td>
<td>0.929</td>
<td>0.957</td>
<td>0.293</td>
<td>0.938</td>
<td>0.670</td>
<td>0.942</td>
<td>0.207</td>
</tr>
<tr>
<td>500</td>
<td>0.950</td>
<td>0.251</td>
<td>0.952</td>
<td>0.218</td>
<td>0.946</td>
<td>0.662</td>
<td>0.955</td>
<td>0.264</td>
<td>0.947</td>
<td>0.577</td>
<td>0.945</td>
<td>0.185</td>
</tr>
<tr>
<td>1000</td>
<td>0.948</td>
<td>0.173</td>
<td>0.947</td>
<td>0.155</td>
<td>0.948</td>
<td>0.405</td>
<td>0.956</td>
<td>0.189</td>
<td>0.949</td>
<td>0.393</td>
<td>0.947</td>
<td>0.132</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\kappa_i(0.1)$</th>
<th>$\kappa_i(0.5)$</th>
<th>$\kappa_i(0.9)$</th>
<th>$\kappa_{OR}(0.1)$</th>
<th>$\kappa_{OR}(0.5)$</th>
<th>$\kappa_{OR}(0.9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.6$</td>
<td>$0.4$</td>
<td>$0.8$</td>
<td>$0.478$</td>
<td>$0.410$</td>
<td>$0.514$</td>
</tr>
<tr>
<td>$0.4$</td>
<td>$0.5$</td>
<td>$0.8$</td>
<td>$0.5$</td>
<td>$0.5$</td>
<td>$0.4$</td>
</tr>
<tr>
<td>$0.8$</td>
<td>$0.9$</td>
<td>$0.9$</td>
<td>$0.40$</td>
<td>$0.78$</td>
<td>$0.492$</td>
</tr>
<tr>
<td>$0.55$</td>
<td>$0.85$</td>
<td>$1.386$</td>
<td>$\alpha_0 = 0.6$</td>
<td>$1.185$</td>
<td>$2.286$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.941</td>
<td>0.420</td>
<td>0.969</td>
<td>0.368</td>
<td>0.942</td>
<td>3.218</td>
<td>0.976</td>
<td>0.328</td>
<td>0.976</td>
<td>0.483</td>
<td>0.980</td>
<td>0.393</td>
</tr>
<tr>
<td>200</td>
<td>0.940</td>
<td>0.270</td>
<td>0.956</td>
<td>0.270</td>
<td>0.944</td>
<td>1.162</td>
<td>0.959</td>
<td>0.238</td>
<td>0.966</td>
<td>0.284</td>
<td>0.979</td>
<td>0.266</td>
</tr>
<tr>
<td>300</td>
<td>0.948</td>
<td>0.216</td>
<td>0.958</td>
<td>0.223</td>
<td>0.945</td>
<td>0.782</td>
<td>0.960</td>
<td>0.195</td>
<td>0.955</td>
<td>0.225</td>
<td>0.959</td>
<td>0.213</td>
</tr>
<tr>
<td>400</td>
<td>0.953</td>
<td>0.186</td>
<td>0.954</td>
<td>0.195</td>
<td>0.946</td>
<td>0.640</td>
<td>0.952</td>
<td>0.170</td>
<td>0.953</td>
<td>0.191</td>
<td>0.957</td>
<td>0.185</td>
</tr>
<tr>
<td>500</td>
<td>0.947</td>
<td>0.166</td>
<td>0.955</td>
<td>0.175</td>
<td>0.948</td>
<td>0.558</td>
<td>0.947</td>
<td>0.152</td>
<td>0.948</td>
<td>0.169</td>
<td>0.948</td>
<td>0.165</td>
</tr>
<tr>
<td>1000</td>
<td>0.949</td>
<td>0.116</td>
<td>0.954</td>
<td>0.124</td>
<td>0.947</td>
<td>0.377</td>
<td>0.951</td>
<td>0.108</td>
<td>0.948</td>
<td>0.118</td>
<td>0.946</td>
<td>0.117</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\kappa_i(0.1)$</th>
<th>$\kappa_i(0.5)$</th>
<th>$\kappa_i(0.9)$</th>
<th>$\kappa_{OR}(0.1)$</th>
<th>$\kappa_{OR}(0.5)$</th>
<th>$\kappa_{OR}(0.9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.4$</td>
<td>$0.8$</td>
<td>$0.6$</td>
<td>$0.447$</td>
<td>$0.7$</td>
<td>$0.824$</td>
</tr>
<tr>
<td>$0.3368$</td>
<td>$0.82$</td>
<td>$0.98$</td>
<td>$0.2286$</td>
<td>$0.9429$</td>
<td>$0.5846$</td>
</tr>
<tr>
<td>$0.4$</td>
<td>$0.8$</td>
<td>$0.8$</td>
<td>$0.2286$</td>
<td>$0.9429$</td>
<td>$0.2286$</td>
</tr>
<tr>
<td>$1.288$</td>
<td>$9.25$</td>
<td>$1.1$</td>
<td>$1.016$</td>
<td>$1.148$</td>
<td>$0.981$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.961</td>
<td>3.521</td>
<td>0.978</td>
<td>0.469</td>
<td>0.940</td>
<td>0.251</td>
<td>0.954</td>
<td>0.275</td>
<td>0.979</td>
<td>0.601</td>
<td>0.981</td>
<td>0.462</td>
</tr>
<tr>
<td>200</td>
<td>0.959</td>
<td>1.491</td>
<td>0.974</td>
<td>0.347</td>
<td>0.942</td>
<td>0.166</td>
<td>0.958</td>
<td>0.195</td>
<td>0.961</td>
<td>0.389</td>
<td>0.976</td>
<td>0.328</td>
</tr>
<tr>
<td>300</td>
<td>0.957</td>
<td>0.887</td>
<td>0.958</td>
<td>0.285</td>
<td>0.946</td>
<td>0.132</td>
<td>0.952</td>
<td>0.160</td>
<td>0.958</td>
<td>0.305</td>
<td>0.960</td>
<td>0.266</td>
</tr>
<tr>
<td>400</td>
<td>0.956</td>
<td>0.569</td>
<td>0.954</td>
<td>0.249</td>
<td>0.950</td>
<td>0.114</td>
<td>0.956</td>
<td>0.139</td>
<td>0.949</td>
<td>0.260</td>
<td>0.958</td>
<td>0.229</td>
</tr>
<tr>
<td>500</td>
<td>0.949</td>
<td>0.477</td>
<td>0.957</td>
<td>0.224</td>
<td>0.943</td>
<td>0.101</td>
<td>0.949</td>
<td>0.124</td>
<td>0.950</td>
<td>0.230</td>
<td>0.952</td>
<td>0.205</td>
</tr>
<tr>
<td>1000</td>
<td>0.951</td>
<td>0.305</td>
<td>0.950</td>
<td>0.160</td>
<td>0.958</td>
<td>0.071</td>
<td>0.957</td>
<td>0.088</td>
<td>0.947</td>
<td>0.162</td>
<td>0.947</td>
<td>0.146</td>
</tr>
</tbody>
</table>

Cov.: coverage probability. Leng.: average length.
3.3.4.4. Program in R

A program has been written in R to solve this problem. This program, called “cwkc” (Combination of Weighted Kappa Coefficients), is executed with the command

\[
\text{cwkc}(s_{i1}, s_{i0}, s_{01}, s_{00}, r_{i1}, r_{i0}, r_{01}, r_{00}, \text{indexc}, \text{rule})
\]

when the CIs are calculated to a confidence of 95%, where \(s_{ij}\) and \(r_{ij}\) are the observed frequencies, \(\text{indexc}\) the value of the weighting index \(c\), and \(\text{rule}\) is equal to “and” (including commas) when the AND rule is used and is equal to “or” when the OR rule is used. The command is

\[
\text{cwkc}(s_{i1}, s_{i0}, s_{01}, s_{00}, r_{i1}, r_{i0}, r_{01}, r_{00}, \text{indexc}, \text{rule}, \text{conflevel})
\]

when the confidence level is \((100 \times \text{conflevel})\)%.

First, the program always checks that all the initial values are correct; where this is not so, an error message is given. The program produces the estimations of the weighted kappa coefficients of the BDTs and their standard errors, the estimation of the combined weighted kappa coefficient and its standard error, and the CIs proposed in Section 3.3.4.2. The program also carries out a similar graph to those shown in Figures 3.1 and 3.2, and which is kept in a file called “Graph_cwkc.jpg” in the same folder from where the program is run. In this graph the values of \(c_h\) (or \(c'_h\)) are indicated, provided that their values are between 0 and 1.

Similarly, the results obtained when running the program are kept in a file called “Results_cwkc.txt” in the same folder from where the program is run. The program is written in the Appendix 3.4.
3.3.4.5. Example

The results obtained have been applied to the study of Weiner et al (1979) on the diagnosis of coronary disease, using the clinical history (Test 1) and the exercise test (Test 2) as diagnostic tests, and the coronary angiography as GS. Weiner et al applied the two BDTs and the GS to a sample of 871 individuals. The results of this study are shown in Table 3.6, where the variable \( T_1 \) models the result of the clinical history, \( T_2 \) models the result of the exercise test, and \( D \) the result of the coronary angiography. In the following, the results of the two BDTs being combined with the AND rule and with the OR are analysed.

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_1 = 1 )</td>
</tr>
<tr>
<td>D = 1</td>
</tr>
<tr>
<td>D = 0</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

3.3.4.5.1. AND rule

By executing the program “cwkc” with the command

\[
\text{cwkc}(473,81,29,25,22,44,46,151,\text{indexc},"and")
\]

one obtains the Figure 3.3, independently of the index \( c \) chosen by the researcher, as well as the values of \( c_1 \), where \( c_1 = 0.377 \) and \( c_2 = 0.921 \). By analysing this graph one
finds that $\hat{\kappa}_{\text{AND}}(c) > \{\hat{\kappa}_1(c), \hat{\kappa}_2(c)\}$ if $c < 0.377$. This result indicates that, in terms of the estimators, the combination of the two BDTs increases the value of $\hat{\kappa}_{\text{AND}}(c)$ if the clinician is more concerned with the false positives of the combination of the two BDTs (for example, when the combination of the two BDTs is used as a previous step to a risky treatment), but always for $0 \leq c < 0.377$. If the clinician is more concerned with the false negatives of the combination of the two BDTs (for example, when the combination of the two BDTs is used as a screening test), then $\hat{\kappa}_{\text{AND}}(c) < \hat{\kappa}_1(c)$ for any value $0.5 < c \leq 1$ and the combination of the two BDTs never increases the value of $\hat{\kappa}_{\text{AND}}(c)$.

Figure 3.3. Graphic of the study of Weiner et al with the AND rule.
In order to illustrate the method, let us consider the values 0.1 and 0.3, that are lower values than 0.377, as weighting indices. In Table 3.7 the results are shown. For $c = 0.1$ the CIs for $\kappa_{\text{AND}}(0.1)/\kappa_1(0.1)$ and $\kappa_{\text{AND}}(0.1)/\kappa_2(0.1)$ are larger than 1 (the lower limit is greater than 1) and this indicates that $\kappa_{\text{AND}}(0.1) > \{\kappa_1(0.1), \kappa_2(0.1)\}$, and so, the combination of the two BDTs increases (to a 95% confidence level) the value of the weighted kappa coefficient. By interpreting the logit CI one obtains that $\kappa_{\text{AND}}(0.1)$ is, with the confidence of 95%, a value between 0.738 and 0.839, which indicates that the level of clinical significance between the combination of the two BDTs and the GS is excellent (Cicchetti, 2001). For $c = 0.3$ the CI for $\kappa_{\text{AND}}(0.3)/\kappa_1(0.3)$ contains the value 1, which indicates that $\kappa_{\text{AND}}(0.3)$ is not (to a 95% confidence level) greater than $\kappa_1(0.3)$. Therefore, the combination of the two BDTs does not increase (to a 95% confidence level) the value of the combined weighted kappa coefficient. The CI for $\kappa_{\text{AND}}(0.3)/\kappa_2(0.3)$ is greater than 1, and this indicates that $\kappa_{\text{AND}}(0.3)$ is (to a 95% confidence level) greater than $\kappa_2(0.3)$. Therefore the combination of the two BDTs has a weighted kappa coefficient greater than that of the Test 2, but is not higher than that of the Test 1.

For any value of the index $c$ greater than 0.377, the value of $\hat{\kappa}_{\text{AND}}(c)/\hat{\kappa}_1(c)$ is less than 1. In this situation, the CI contains the value $\hat{\kappa}_{\text{AND}}(c)/\hat{\kappa}_1(c)$, so that the CI is less than 1 (the upper limit of the CI is less than 1) or the CI contains the value 1, and in both cases the combination of the two BDTs with the AND rule does not produce (to a 95% confidence level) an increase of $\kappa_{\text{AND}}(c)$. 

112
Table 3.7. Results obtained when combining the two weighted kappa coefficients with the AND rule.

<table>
<thead>
<tr>
<th></th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 1</th>
<th>Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{k}_h(c) \pm SE$</td>
<td>0.652 ± 0.0322</td>
<td>0.592 ± 0.0345</td>
<td>0.660 ± 0.0291</td>
<td>0.567 ± 0.0311</td>
</tr>
<tr>
<td>$\hat{Y}_h$</td>
<td>0.660</td>
<td>0.567</td>
<td>0.660</td>
<td>0.567</td>
</tr>
<tr>
<td>$\hat{k}_{AND}(c) \pm SE$</td>
<td>0.793 ± 0.0256</td>
<td>0.695 ± 0.0249</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{AND}$</td>
<td>0.694</td>
<td>0.694</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fieller CI for $\kappa_{AND}(c)/\kappa_2(c)$

- $\{1.121, 1.328\}$ for $c = 0.1$
- $\{0.972, 1.144\}$ for $c = 0.3$

Fieller CI for $\kappa_{AND}(c)/\kappa_1(c)$

- $\{1.222, 1.483\}$ for $c = 0.1$
- $\{1.128, 1.340\}$ for $c = 0.3$

Logit CI for $\kappa_{AND}(c)$

- $\{0.738, 0.839\}$ for $c = 0.1$
- $\{0.644, 0.742\}$ for $c = 0.3$

3.3.4.5.2. OR rule

By executing the program “cwkc” with the command

```
cwkc 473,81,29,25,22,44,46,151,indexc,"or")
```

one obtains the Figure 3.4, independently of the index $c$ chosen, as well as the values of $c'_h$, with $c'_1 = 0.759$ and $c'_2 = 0.392$. By analysing this graph one finds that

$\hat{k}_{OR}(c) > \{\hat{k}_1(c), \hat{k}_2(c)\}$ if $c > 0.759$. This result indicates that, in terms of the estimators, the combination of the two BDTs increases the value of $\hat{k}_{OR}(c)$ if the clinician has more concern for the false negative of the combination of the two BDTs, but always for $0.759 < c \leq 1$. If the clinician has more concern for the false positives
Chapter 3. Comparison and combination of two BDTs

\(0 \leq c < 0.5\) of the combination of the two BDTs, then \(\hat{\kappa}_{OR}(c) < \{\hat{\kappa}_1(c), \hat{\kappa}_2(c)\}\) and the combination of the two BDTs never increases the value of \(\hat{\kappa}_{OR}(c)\).

Figure 3.4. Graphic of the study of Weiner et al with the OR rule.

In order to illustrate the method, we shall take as weighting indices the values 0.8 and 0.9, which are lower values than 0.759. In Table 3.8 the results obtained are shown. For \(c = 0.8\) the CI for \(\kappa_{OR}(0.8)/\kappa_1(0.8)\) contains the value 1, which indicates that \(\kappa_{OR}(0.8)\) is not (95% confidence level) greater than \(\kappa_1(0.8)\). Therefore, the combination of the two BDTs does not increase (to a 95% confidence level) the value of the combined weighted kappa coefficient. The CI for \(\kappa_{OR}(0.8)/\kappa_2(0.8)\) is greater than 1, which indicates that \(\kappa_{OR}(0.8)\) is (to a 95% confidence level) greater than \(\kappa_2(0.8)\). Hence, the combination of the two BDTs has a weighted kappa coefficient greater than
that of Test 2, but is not greater than that of Test 1. For $c=0.9$ the CIs for
\[ \kappa_{or}(0.9)/\kappa_1(0.9) \] and \[ \kappa_{or}(0.9)/\kappa_2(0.9) \] are greater than 1, which indicates that
\[ \kappa_{or}(0.9) > \{ \kappa_1(0.9), \kappa_2(0.9) \} \] and therefore the combination of the two BDTs increases
(to a 95% confidence level) the value of the combined weighted kappa coefficient. By
interpreting the logit CI one obtains that \[ \kappa_{or}(0.9) \] is, with a 95% confidence level, a
value between 0.676 and 0.802, which indicates that the level of clinical significance
between the combination of the two BDTs and the GS is a value between good and
excellent (Cicchetti, 2001).

For any value of the index $c$ lower than 0.759, the value of \[ \hat{\kappa}_{or}(c)/\hat{\kappa}_1(c) \] is less than
1. The CI will contains the value \[ \hat{\kappa}_{or}(c)/\hat{\kappa}_1(c) \], so that the CI is less than 1 or it
contains the value 1, and in both cases the combination of the two BDTs with the OR
rule does not produce (to a 95% confidence level) an increase of \[ \kappa_{or}(c) \].
Table 3.8. Results obtained when combining the two weighted kappa coefficients with the OR rule.

<table>
<thead>
<tr>
<th></th>
<th>$c = 0.8$</th>
<th></th>
<th>$c = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test 1</td>
<td>Test 2</td>
<td>Test 1</td>
</tr>
<tr>
<td>$\hat{\kappa}_h(c) \pm SE$</td>
<td>0.682 ± 0.0300</td>
<td>0.514 ± 0.0318</td>
<td>0.687 ± 0.0319</td>
</tr>
<tr>
<td>$\hat{Y}_h$</td>
<td>0.660</td>
<td>0.567</td>
<td>0.660</td>
</tr>
<tr>
<td>$\hat{\kappa}_{OR}(c) \pm SE$</td>
<td>0.698 ± 0.0304</td>
<td>0.744 ± 0.0322</td>
<td></td>
</tr>
<tr>
<td>$\hat{Y}_{OR}$</td>
<td>0.533</td>
<td>0.533</td>
<td>0.533</td>
</tr>
</tbody>
</table>

Fieller CI for $\kappa_{OR}(c)/\kappa_1(c)$
- (0.951, 1.099)
- (1.008, 1.164)

Fieller CI for $\kappa_{OR}(c)/\kappa_2(c)$
- (1.225, 1.515)
- (1.325, 1.654)

Logit CI for $\kappa_{OR}(c)$
- (0.635, 0.754)
- (0.676, 0.802)

3.3.4.6. Discussion

The combination of BDTs for incrementing the accuracy of the diagnosis of a disease is common in clinical practice. It consists in combining the results of the two tests using a suitable method and then estimating the parameters of the combination of both BDTs. In Section 3.3.4 we study the combination of the weighted kappa coefficients of two BDTs when the combination is in parallel testing under the rules AND and OR. The combined weighted kappa coefficient and its properties have been defined and the conditions under which the combination of the two BDTs increases the value of the combined weighted kappa coefficient have been studied. The conditions studied are theoretical
conditions (for the parameters), so that its practical application does not guarantee that
the increase in the combined weighted kappa coefficient will be produced, although
they can help to give a view of the problem. Thus, in practice, and as we carried out in
the example of Weiner et al (1979), it is necessary to study the increase in \( \kappa_c \) using
CIs.

The CIs proposed are approximate intervals, based on the asymptotic normality of
the estimators. The simulation experiments carried out have shown that the coverage
probabilities of these CIs can exceed the confidence level when the samples are not very
large (sizes 100-200), and this may be due to the fact that the convergence of the
multinomial distribution (which is the probability distribution inherent of the observed
data) with the normal distribution is slow and requiring larger sample sizes. For larger
sample sizes, the CIs proposed have a coverage probability fluctuating around the 95%
confidence level.

An alternative CI to Fieller method is obtained by applying the transformation of the
naperian logarithm. Assuming that \( \ln(\hat{\theta}_h) \xrightarrow{a.s.} N(0, \text{Var}[\ln(\theta_h)]) \), then an
approximate CI for \( \theta_h \) is

\[
\hat{\theta}_h \exp \left\{ \pm z_{1-\alpha/2} \sqrt{\text{Var}[\ln(\hat{\theta}_h)]} \right\},
\]

where \( \text{Var}[\ln(\hat{\theta}_h)] \) is obtained by applying the delta method in a similar way as is indicated in the equation
(3.62) but with \( \ln(\hat{\theta}_h) \). Simulation experiments (similar to those made in Section
3.3.4.3) have been carried out, obtaining that this CI has a very similar behaviour to the
Fieller method when the sample size is \( \geq 300 \); for samples of sizes 100 or 200, the
coverage probability of this CI exceeds 95% or does not attain that coverage (and in
some cases is below 90%).
Chapter 3. Comparison and combination of two BDTs

We have only studied the combination of the two BDTs in parallel testing. A different scenario is when the serial testing is used, in which case the application of Test 2 depends on the result of Test 1. From a statistical point of view, with the serial testing we cannot estimate [8]: all the probabilities $p_y$ and $q_y$, the sensitivity and specificity of Test 2, and the dependency factors between the two BDTs.
Conclusions

The main parameters to estimate and compare the accuracy of binary diagnostic tests are sensitivity and specificity, which only depend on the intrinsic ability of the diagnostic test to distinguish between diseased individuals and non-diseased individuals. Other parameters are likelihood ratios which only depend on the sensitivity and specificity of the binary diagnostic test. Both sensitivity and specificity, like likelihood ratios, are superior measures of the accuracy of a binary diagnostic test, as they depend on the characteristics of the test. The predictive values of a binary diagnostic test are also parameters that are widely used to assess and compare binary diagnostic tests, and they represent the clinical accuracy of the test. The predictive values depend on the sensitivity and specificity of the test and the disease prevalence.

When considering the losses or costs of an erroneous classification with the binary diagnostic test, the performance of a diagnostic test is measured through the weighted kappa coefficient. The weighted kappa coefficient depends on the sensitivity and specificity of the diagnostic test, on the disease prevalence and the weighting index $c$. The weighting index $c$ represents the relative importance between the false positives and
the false negatives. In Chapter 1, all these parameters are defined and their properties are studied.

In Chapter 2, we study the estimations and confidence intervals of the parameters defined in Chapter 1, subject to two types of study or sample: a cross-sectional study and a case-control study. The contribution made by this Chapter is the estimation of the weighted kappa coefficient subject to a case-control study. Different approximate confidence intervals are proposed for the weighted kappa coefficient: a Wald type interval, an interval with logit transformation, an interval with arcsine transformation, an interval using bootstrap and a Bayesian interval. For the first three intervals, we used four corrections which are frequent in this type of studies \( h = 0, 0.5, 2, \frac{z_{1-\alpha/2}^2}{2} \). As a bootstrap interval, we used a bias-corrected one, and for the Bayesian interval we used an interval based on the Monte Carlo method based on quartiles with distribution which a priori are non-informative and also with distribution which a priori are informative. Simulation experiments to study the asymptotic coverage of these intervals, from which some general rules of application were given: when the sample sizes are small \( n_i \leq 75 \) use the Wald CI with \( h = 0.5 \), and for other sample sizes \( n_i \geq 100 \) use the Wald CI with \( h = 0 \). A method has also been proposed to calculate the sample sizes (of cases and of controls) necessary to estimate the weighted kappa coefficient with a determined precision \( \delta \) to the confidence \( 100(1-\alpha)\% \). The results obtained have been applied to a real example of the diagnosis of coronary disease.
Finally, in Chapter 3 we studied the comparison of the parameters defined in Chapter 1 of two binary diagnostic tests subject to a paired design, presenting hypothesis tests to check the equality of the corresponding parameters and CIs for the respective differences. We also studied the combination of parameters of two binary diagnostic tests in parallel testing. The contribution of this Chapter is the combination of the weighted kappa coefficients of two binary diagnostic tests in parallel testing, both with the AND rule and with the OR rule. The combined weighted kappa coefficient was defined when using the AND rule and when using the OR rule, and we studied its properties, finding them to be similar to those of the weighted kappa coefficient of a binary diagnostic test. We studied the conditions under which the combination of two binary diagnostic tests produces an increase in the combined weighted kappa coefficient. Whether or not an increase occurs in the combined weighted kappa coefficient will depend on the value of the weighting index \( c \) set by the researcher, although in some situations such an increase never occurs. These conditions studied are theoretical, and therefore in a practical problem, whether or not there is an increase in the combined weighted kappa coefficient must be studied through a confidence interval. For this purpose, Fieller’s theorem was applied, obtaining a confidence interval for the ratio between the combined weighted kappa coefficient and each weighted kappa coefficient. Furthermore, a logit confidence interval was studied for the combined weighted kappa coefficient. Monte Carlo simulation experiments were carried out to study the asymptotic behaviour of the confidence intervals proposed, finding that, in general terms, these intervals have an asymptotic coverage that fluctuates around 95% even with relatively small samples, although in some situations the average coverage
can exceed 95%. An $R$ programme was written to solve the problem posed. The results have been applied to a real example on the diagnosis of coronary disease.
Appendices

Appendix 2.1: Estimation of the variance

The variance of the weighted kappa coefficient can be estimated applying the delta method (Agresti, 2002), i.e.

$$\text{Var} \left[ \hat{\kappa}(c) \right] \approx \left( \frac{\partial \kappa(c)}{\partial \text{Se}} \right)^2 \text{Var} \left( \hat{\text{Se}} \right) + \left( \frac{\partial \kappa(c)}{\partial \text{Sp}} \right)^2 \text{Var} \left( \hat{\text{Sp}} \right),$$

since \( \text{Cov}(\hat{\text{Se}}, \hat{\text{Sp}}) = 0 \) (as the sensitivity is estimated from the case sample and the specificity from the control sample, and both samples are independent), and where

$$\text{Var} \left( \hat{\text{Se}} \right) = \frac{\text{Se}(1-\text{Se})}{n_1}$$
and
$$\text{Var} \left( \hat{\text{Sp}} \right) = \frac{\text{Sp}(1-\text{Sp})}{n_0}.$$

Carrying out the algebraic operations, it holds that

$$\frac{\partial \kappa(c)}{\partial \text{Se}} = \frac{\kappa(c)}{pqY} \left[ pq - \kappa(c) \{ p(q-c) \} \right]$$
and
$$\frac{\partial \kappa(c)}{\partial \text{Sp}} = \frac{\kappa(c)}{pqY} \left[ pq + \kappa(c) \{ q(c-q) \} \right].$$
where \( Y = Se + Sp - 1 \) is the Youden index. Thus, it holds that

\[
Var[\hat{\kappa}(c)] = \left( \frac{\kappa(c)}{pqY} \right)^2 \times \left[ \left( pq - \kappa(c) \right) \left( p(q-c) \right) \right]^2 \frac{Se(1-Se)}{n_1} + \left[ \left( pq + \kappa(c) \right) \left( q(c-q) \right) \right]^2 \frac{Sp(1-SP)}{n_0}.
\]

Substituting in this equation each parameter with its estimator and carrying out the algebraic operations we obtain equation (2.71). The variances of the logit of \( \kappa(c) \) and of the arcsine are estimated in a similar way to the previous case, i.e.

\[
Var\{\logit[\hat{\kappa}(c)]\} \approx \left( \frac{\partial \logit[\kappa(c)]}{\partial Se} \right)^2 Var(\hat{Se}) + \left( \frac{\partial \logit[\kappa(c)]}{\partial Sp} \right)^2 Var(\hat{Sp})
\]

and

\[
Var\{\arcsin[\sqrt{\hat{\kappa}(c)}]\} \approx \left( \frac{\partial \arcsin[\sqrt{\kappa(c)}]}{\partial Se} \right)^2 Var(\hat{Se}) + \left( \frac{\partial \arcsin[\sqrt{\kappa(c)}]}{\partial Sp} \right)^2 Var(\hat{Sp}).
\]

The expression of the variance obtained here (equation (2.71)) is different to that obtained by Roldán-Nofuentes et al (2009). Equation (2.71) has a term of variability less than the variance deduced by Roldán-Nofuentes et al (2009), because in a case-control study it is necessary to know a value for the prevalence in order to estimate \( \kappa(c) \). This means that the CIs for \( \kappa(c) \) have a length which is lower in the case-control sample than in the cross-sectional sample.
Appendix 2.2: Comparison of the performance of the CIs

“Step 1” of this method to choose the optimum CI establishes that the probability of coverage should be higher than 93%, or in other words, that the CI does not fail. 

\[ \Delta \alpha = \alpha - \alpha^* = \gamma^* - \gamma \]

is defined, where \( \gamma = 1 - \alpha = 0.95 \) is the nominal confidence of the CI and \( \gamma^* \) the probability of coverage calculated. The hypothesis test (two-tailed) for the weighted kappa coefficient is \( H_0: \kappa(c) = \kappa'(c) \) vs \( H_1: \kappa(c) \neq \kappa'(c) \), where \( \kappa'(c) \) is the value of \( \kappa(c) \) subject to \( H_0 \). This test can be solved through different methods.

If the test is solved applying the Wald method, the contrast statistic is:

\[ z_w = \frac{|\hat{\kappa}(c) - \kappa(c)|}{\sqrt{\text{Var}[\hat{\kappa}(c)]}}. \]

Other alternatives are to use the logit and arcsine transformations. Applying the logit transformation, the test is \( H_0: \text{logit}[\kappa(c)] = \text{logit}[\kappa'(c)] \) vs \( H_1: \text{logit}[\kappa(c)] \neq \text{logit}[\kappa'(c)] \), and the contrast statistic is

\[ z_L = \frac{|\text{logit}[\hat{\kappa}(c)] - \text{logit}[\hat{\kappa}'(c)]|}{\sqrt{\text{Var}[\text{logit}[\hat{\kappa}(c)]]}}. \]

and applying the arcsine transformation, the test \( H_0: \text{arcsin}[\kappa(c)] = \text{arcsin}[\kappa'(c)] \) vs \( H_1: \text{arcsin}[\kappa(c)] \neq \text{arcsin}[\kappa'(c)] \) is solved with the statistic

\[ z_A = \frac{|\sin^{-1}\sqrt{\hat{\kappa}(c)} - \sin^{-1}\sqrt{\hat{\kappa}'(c)}|}{\sqrt{\text{Var}[\sin^{-1}\sqrt{\hat{\kappa}(c)}]}}. \]
In Step 1 of the method, a CI has a failure if its probability of coverage is \( \leq 93\% \), i.e. if \( \Delta \alpha \leq -2 \). In this situation, the type I error of the corresponding hypothesis test is \( \geq 7\% \), and therefore it is a very liberal hypothesis test and can give false significances. The 93\% criteria has been used by other authors (Agreti and Coull, 1998; Price and Bonnett, 2004; Martín-Andrés and Álvarez-Hernández, 2014a and 2014b). If \( \Delta \alpha > 2\% \), i.e. the probability of coverage is greater than 97\%, then the corresponding hypothesis test is very conservative (its type I error is very small, < 3\%), but at least it does not give false significances. Consequently, the choice of the optimum CI is linked to the decisions of the corresponding hypothesis test, and it is preferable to choose a conservative test rather than a very liberal one (then there will not be any false significances as its type I error is lower than the nominal one).

### Appendix 3.1: Combined weighted kappa coefficient

The weighted kappa coefficient of BDT is defined [1] as

\[
\kappa = \frac{\text{Random loss} - \text{Expected loss}}{\text{Random loss}}. \quad (3.68)
\]

The expected loss is the average loss which occurs when erroneously classifying a diseased or non-diseased patient with the BDT. The random expected loss is the expected loss which occurs when the BDT and the GS are independent, i.e. when \( P(T = i|D = j) = P(T = i) \). In parallel testing with the AND rule, the random loss is

\[
\text{Random loss} = LP(p_{10} + p_{01} + p_{00} + q_{01} + q_{00}) + LP(q_{11}(p_{10} + q_{11}) = LP(1 - p_{11} - q_{11}) + LP(q_{11} + q_{11}),
\]

\[
\text{Random loss} = LP(p_{10} + p_{01} + p_{00} + q_{01} + q_{00}) + LP(q_{11}(p_{10} + q_{11}) = LP(1 - p_{11} - q_{11}) + LP(q_{11} + q_{11}),
\]
and the expected loss is

$$\text{Expected loss} = L(p_{10} + p_{01} + p_{00}) + L'q_{11} = L(p - p_{11}) + L'q_{11}.$$ 

By substituting the two previous expressions in equation (3.68) and realizing the algebraic operations, equation (3.56) is obtained. With the OR rule, the random loss is

$$\text{Random loss} = Lp(p_{00} + q_{00}) + L'q(p_{11} + p_{10} + p_{01} + q_{11} + q_{10} + q_{01}) = Lp(p_{00} + q_{00}) + L'(1 - p_{00} - q_{00}),$$

and the expected loss is

$$\text{Expected loss} = Lp_{00} + L'(q_{11} + q_{01} + q_{00}) = Lp_{00} + L'(q - q_{00}).$$

By substituting these two previous expressions in equation (3.68) and realizing the algebraic operations, equation (3.57) is obtained.

### Appendix 3.2: The AND rule

In all of the following, it will be taken that all the Youden indices are between 0 and 1 ($0 < Y < 1$), and that the prevalence of the disease is also a value between 0 and 1 ($0 < p < 1$). Let us consider the following equalities:

$$\omega_1 = Q_1Y_{\text{AND}} - Q_{\text{AND}}Y_1 = \frac{p_{11}q_{10} - p_{10}q_{11}}{pq}, \quad \omega_2 = Q_2Y_{\text{AND}} - Q_{\text{AND}}Y_2 = \frac{p_{11}q_{01} - p_{01}q_{11}}{pq},$$

$$\xi_1 = Y_{\text{AND}} - Y_1 = \frac{q_{10}p - p_{10}q}{pq}, \quad \xi_2 = Y_{\text{AND}} - Y_2 = \frac{q_{01}p - p_{01}q}{pq},$$

$$Q_4 = p_{11} + p_{10} + q_{11} + q_{10}, \quad Q_2 = p_{11} + p_{01} + q_{11} + q_{01},$$
\[ \tau_1 = Q_1 - Q_{\text{AND}} = p_{10} + q_{10}, \quad \tau_2 = Q_2 - Q_{\text{AND}} = p_{01} + q_{01}, \]

\[ \nu_{\text{AND}} = pc(1 - Q_{\text{AND}}) + q(1 - c)Q_{\text{AND}} \] and \( \nu_h = pc(1 - Q_h) + q(1 - c)Q_h, \quad h = 1, 2, \)

where \( \nu_{\text{AND}} \) and \( \nu_h \) are the denominators of \( \kappa_{\text{AND}}(c) \) and \( \kappa_h(c) \) respectively. It is verified that \(-1 < \omega_h < 1, -1 < \xi_h < 1, 0 < Q_h < 1 \) and \( 0 < \tau_h < 1 \). By realizing algebraic operations it is verified that

\[ \kappa_{\text{AND}}(c) - \kappa_h(c) = \frac{pq\xi_h - (c - q)\tau_h\kappa_h(c)}{\nu_{\text{AND}}} = \frac{pq\xi_h\nu_h - (c - q)\tau_hY_h}{\nu_{\text{AND}}\nu_h}. \quad (3.69) \]

Then \( \kappa_{\text{AND}}(c) = \kappa_h(c) \) if

\[ c = c_h = \frac{q(Q_hY_{\text{AND}} - Q_{\text{AND}}Y_h)}{Y_h(p - Q_{\text{AND}}) - Y_{\text{AND}}(p - Q_h)} = \frac{q\omega_h}{\omega_h - p\xi_h} = \frac{q\omega_h}{\psi_h}, \quad h = 1, 2, \quad (3.70) \]

where \( \psi_h = Y_h(p - Q_{\text{AND}}) - Y_{\text{AND}}(p - Q_h) = \omega_h - p\xi_h \). The values \( c_1 \) and \( c_2 \) can take any real value. In terms of \( c_h \) the equation (3.69) is written as

\[ \kappa_{\text{AND}}(c) - \kappa_h(c) = \frac{pq^2\omega_h(c_h - c)}{c_h\nu_{\text{AND}}\nu_h}. \quad (3.71) \]

In what follows the conditions under which the combination of the two BDTs with the AND rule produces an increase of the combined weighted kappa coefficient are shown.

\( a) \) If \( 0 < c_h < 1 \) with \( h = 1, 2 \), then \( \kappa_h(c_h) \) and \( \kappa_{\text{AND}}(c_h) \), can be calculated, so obtaining \( \kappa_h(c_h) = \frac{Y_h}{(\omega_h - \xi_hQ_h)/\psi_h} \) and \( \kappa_{\text{AND}}(c_h) = \frac{Y_{\text{AND}}}{(\omega_h - \xi_hQ_{\text{AND}})/\psi_h} \). It is verified that

\[ \frac{\omega_h - \xi_hQ_h}{\psi_h} > 0 \quad \text{and} \quad \frac{\omega_h - \xi_hQ_{\text{AND}}}{\psi_h} > 0, \]
since both expressions are the denominators of weighted kappa coefficients. As it is always verified that \( Q_h > Q_{AND} \), if \( \xi_h = Y_{AND} - Y_h > 0 \) then

\[
\frac{\omega_h - \xi_h Q_{AND}}{\psi_h} - \frac{\omega_h - \xi_h Q_h}{\psi_h} > 0, 
\]

By realizing algebraic operations one obtains

\[
\frac{\omega_h - \xi_h Q_{AND}}{\psi_h} - \frac{\omega_h - \xi_h Q_h}{\psi_h} = \frac{\xi_h \tau_h}{\psi_h} > 0, 
\]

and as \( \xi_h \) and \( \tau_h \) are greater than 0, then \( \psi_h > 0 \). If \( \xi_h = Y_{AND} - Y_h < 0 \) then

\[
\frac{\omega_h - \xi_h Q_{AND}}{\psi_h} - \frac{\omega_h - \xi_h Q_h}{\psi_h} > 0, 
\]

and by realizing algebraic operations one obtains

\[
\frac{\omega_h - \xi_h Q_{AND}}{\psi_h} - \frac{\omega_h - \xi_h Q_h}{\psi_h} = -\frac{\xi_h \tau_h}{\psi_h} > 0. 
\]

As \( -\xi_h \) and \( \tau_h \) are greater than 0, then \( \psi_h > 0 \). Thus, whatever the value of \( \xi_h \), it is always verified that \( \psi_h > 0 \). As \( c_h = \frac{q \omega_h}{\psi_h} > 0 \) then \( \omega_h > 0 \). If \( 0 \leq c < c_h \) then, from equation (3.71), \( \kappa_{AND}(c) > \kappa_h(c) \) with \( h = 1, 2 \). Finally, if \( 0 \leq c < \min\{c_1, c_2\} \) then \( \kappa_{AND}(c) > \max\{\kappa_1(c), \kappa_2(c)\} \).

If \( c=1 \) then \( c_h - 1 < 0 \), and from equation (3.71) it is verified that \( \kappa_{AND}(1) < \kappa_h(1) \) with \( h = 1, 2 \). Therefore \( \kappa_{AND}(1) < \min\{\kappa_1(1), \kappa_2(c)\} \).

b) From a) it can be verified that \( \kappa_{AND}(c) > \kappa_i(c) \) if \( 0 \leq c < c_i < 1 \). Let us suppose that \( Y_j < Y_{AND} \), then \( \xi_j = Y_{AND} - Y_j > 0 \). By multiplying by \( Q_j \) one obtains
\[ \xi_j Q_j > 0 \Rightarrow Q_j Y_{AND} - Q_j Y_j = Q_j Y_{AND} - (\tau_j + Q_{AND}) Y_j = \omega_j - \tau_j Y_j > 0 \Rightarrow \omega_j > \tau_j Y_j > 0 , \]

and by multiplying by \( Q_{AND} \) one obtains

\[ \xi_j Q_{AND} > 0 \Rightarrow Q_{AND} Y_{AND} - Q_{AND} Y_j = (Q_j - \tau_j) Y_{AND} - Q_{AND} Y_j = \omega_j - \tau_j Y_{AND} > 0 \]
\[ \Rightarrow \omega_j > \tau_j Y_{AND} > 0. \]

Hence, it can be verified that \( \omega_j > \tau_j Y_{AND} > \tau_j Y_j > 0 \). As \( c_j > 1 \) and \( \omega_j > 0 \), then \( \psi_j = \omega_j - p\xi_j \) must be greater than 0, and it is verified that \( \xi_j > \omega_j \). Finally, as \( \omega_j > 0 \) then, from equation (3.71), it is verified that \( \kappa_{AND}(c) > \kappa_j(c) \) for \( 0 \leq c \leq 1 \).

In the case that \( Y_j > Y_{AND} \), then \( \xi_j < 0 \). Let us suppose that \( \omega_j > 0 \), then as \( c_j > 1 \) it must be verified that \( \omega_j - p\xi_j > 0 \), and so one obtains

\[ q\omega_j > \omega_j - p\xi_j \Rightarrow \xi_j > \omega_j > 0 , \]

which is contradictory to \( \xi_j < 0 \). Thus, if \( Y_j > Y_{AND} \) then \( \omega_j < 0 \), and from equation (3.71) it can be verified that \( \kappa_j(c) > \kappa_{AND}(c) \) for \( 0 \leq c \leq 1 \).

If \( c_j < 0 \) the reasoning is the same as for \( c_j > 1 \). In this case \( \omega_j - p\xi_j < 0 \), so that it is also verified that \( \xi_j > p\xi_j > \omega_j \).

For any situation different to those that have been dealt with in points \( a \) and \( b \), the combination of the two \( BDTs \) does not increase the value of \( \kappa_{AND}(c) \) for any value of \( c \). These situations are:

1) \( 0 \leq c < c_i < 1 \), \( c_j > 1 \) (or \( c_j < 0 \)) and \( Y_{AND} \leq Y_j \),
2) $\min\{c_1,c_2\} > 1,$

3) $\max\{c_1,c_2\} < 0,$

4) $c_i > 1$ and $c_j < 0.$

In the following each of these is analysed.

1) The situation $0 \leq c < c_i < 1,$ $c_j > 1$ (or $c_j < 0$) and $Y_{\text{AND}} < Y_j$ has been dealt with in point b) above. Let us study the same situation but supposing that $Y_{\text{AND}} = Y_j.$ If $Y_{\text{AND}} = Y_j$ then $\xi_j = 0.$ As $c_j = \frac{q\omega_j}{\omega_j - p\xi_j},$ if $\xi_j = 0$ then $c_j = q,$ but this is incompatible with $c_j > 1$ or $c_j < 0,$ because $0 < q < 1.$

2) Let us consider that $c_i > 1$ and $c_2 > 1,$ i.e. $\min\{c_1,c_2\} > 1.$ Each equation $c_h > 1,$ $h = 1,2,$ has two solutions:

$$0 < p\xi_h < \omega_h < \xi_h < 1$$

(3.72)

and

$$-1 < \xi_h < \omega_h < p\xi_h < 0.$$  

(3.73)

Let us consider

$$c_h(1 - c_h) = pq \frac{\omega_h(\omega_h - \xi_h)}{\psi_h^2} < 0,$$

then $c_1(1 - c_1)c_2(1 - c_2) > 0,$ i.e.
\[ c_1(1-c_1)c_2(1-c_2) = p^2 q^2 \frac{\omega_1 \omega_2 (\omega_1 - \xi_1)(\omega_2 - \xi_2)}{\psi_1 \psi_2} > 0. \] (3.74)

From equation (3.74) it is verified that \( c_1(1-c_1)c_2(1-c_2) > 0 \) if
\[ \omega_1 \omega_2 (\omega_1 - \xi_1)(\omega_2 - \xi_2) > 0. \] (3.75)

Moreover,
\[ c_1c_2 = q^2 \frac{\omega_1 \omega_2}{\psi_1 \psi_2} > 1. \] (3.76)

Solving the system formed by equations (3.75) and (3.76), the solutions for this system which are compatible with solutions (3.72) and (3.73) are

\[ 0 < p \xi_i < \omega_i < \xi_i < 1 \] and \( -1 < \xi_2 < \omega_2 < p \xi_2 < 0 \)

and
\[ \xi_i < \omega_i < p \xi_i < 0 \] and \( 0 < p \xi_2 < \omega_2 < \xi_2. \)

In general terms, the solution is
\[ -1 < \xi_j < \omega_j < p \xi_j < 0 < p \xi_i < \omega_i < \xi_i < 1. \]

As \( \omega_j < 0 \) y \( c_j > 1 \), then applying equation (3.34) it holds that \( \kappa_{\text{AND}}(c) < \kappa_j(c) \) for \( 0 \leq c \leq 1 \). As \( \omega_i > 0 \) and \( c_i > 1 \), then \( \kappa_{\text{AND}}(c) > \kappa_i(c) \) for \( 0 \leq c \leq 1 \). Finally,
\[ \kappa_i(c) < \kappa_{\text{AND}}(c) < \kappa_j(c), \ 0 \leq c \leq 1. \]

Consequently, if \( \min\{c_1, c_2\} > 1 \) then the combination of the two weighted kappa coefficients never increases the value of the combined weighted kappa coefficient. Furthermore, as \( \xi_j < 0 \) then \( Y_{\text{AND}} < Y_j \), and as \( \xi_i > 0 \) then \( Y_{\text{AND}} > Y_i \), i.e.
$Y_i < Y_{\text{AND}} < Y_j$.

3) Another possible situation is that $c_i < 0$ and $c_2 < 0$, i.e. $\max\{c_1, c_2\} < 0$, then the solutions for each equation $c_h < 0$, $h = 1, 2$, are $0 < \omega_h < p_1 \xi_h < \xi_h < 1$ or $-1 < \xi_h < p_2 \xi_h < \omega_h < 0$. Solving the system formed by the equations

$$c_1 (1-c_1) c_2 (1-c_2) = p^2 q^2 \frac{\omega_1 \omega_2 (\omega_1 - \xi_1) (\omega_2 - \xi_2)}{\psi_1^2 \psi_2^2} > 0$$

and

$$c_1 c_2 = q^2 \frac{\omega_1 \omega_2}{\psi_1 \psi_2} > 0,$$

the appropriate solution is $-1 < \xi_j < p_1 \xi_j < \omega_j < 0 < \omega_j < p_2 \xi_j < \xi_j < 1$. Finally, it holds that

$$\kappa_j (c) < \kappa_{\text{AND}} (c) < \kappa_j (c), \quad 0 \leq c \leq 1,$$

and

$$Y_i < Y_{\text{AND}} < Y_j.$$

4) Finally, another possible situation is that $c_i > 1$ and $c_j < 0$. Solving the system formed by the equations

$$c_1 (1-c_1) c_2 (1-c_2) = p^2 q^2 \frac{\omega_1 \omega_2 (\omega_1 - \xi_1) (\omega_2 - \xi_2)}{\psi_1^2 \psi_2^2} > 0$$

and
the appropriate solution is \(-1 < \xi_j < p\xi_j < \omega_j < 0 < p\xi_i < \omega_i < \xi_i < 1\). In this situation, it is verified that if \(\xi_j < 0\) then \(-1 < \xi_j < p\xi_j < \omega_j < q\omega_j < 0\) and \(\psi_j = \omega_j - p\xi_j > 0\), and therefore \(c_j < 0\). Consequently, if \(c_i > 1\) and \(c_j < 0\), the solution is \(-1 < \xi_j < p\xi_j < \omega_j < 0 < p\xi_i < \omega_i < \xi_i < 1\), and it is again verified that

\[ \kappa'_i(c) < \kappa_{AND}(c) < \kappa_j(c), \quad 0 \leq c \leq 1, \]

and

\[ Y_i < Y_{AND} < Y_j. \]

**Appendix 3.3: The OR rule**

In all of the following, it will be taken that all the Youden indices are between 0 and 1 (\(0 < Y < 1\)), and that the prevalence of the disease is also a value between 0 and 1 (\(0 < p < 1\)). Let us take the following equalities

\[ \omega'_1 = Q_1 Y_{OR} - Q_{OR} Y_1 = \frac{p_{01}(q_{11} + q_{10}) - q_{01}(p_{11} + p_{10})}{pq}, \]

\[ \omega'_2 = Q_2 Y_{OR} - Q_{OR} Y_2 = \frac{p_{10}(q_{11} + q_{10}) - q_{10}(p_{11} + p_{01})}{pq}, \]
\[
\xi'_1 = Y_{OR} - Y_1 = \frac{p_{01}q - q_{10}p}{pq}, \quad \xi'_2 = Y_{OR} - Y_2 = \frac{p_{10}q - q_{10}p}{pq}
\]

\[
\tau'_2 = Q_2 - Q_{OR} = -(p_{10} + q_{10}) \quad \text{and} \quad v_{OR} = pc(1 - Q_{OR}) + q(1 - c)Q_{OR}.
\]

It is verified that \(-1 < \alpha'_h < 1, \quad -1 < \xi'_h < 1\) and \(-1 < \tau'_h < 0\). By realizing algebraic operations it can be verified that

\[
\kappa_{OR}(c) - \kappa_h(c) = \frac{pq\xi'_h - (c-q)\tau'_h\kappa_h(c)}{v_{OR}} = \frac{pq\xi'_h - (c-q)\tau'_hY_{h}}{V_{OR}V'_h}, \quad (3.77)
\]

so verifying that \(\kappa_{OR}(c) = \kappa_h(c)\) if

\[
c = c'_h = \frac{q(Q_hY_{OR} - Q_{OR}Y_{h})}{Y_h(p - Q_{OR}) - Y_{OR}(p - Q_{h})} = \frac{q\alpha'_h}{\alpha'_h - p\xi'_h} = \frac{q\alpha'_h}{\psi'_h}, \quad h = 1, 2, \quad (3.78)
\]

with \(\psi'_h = Y_h(p - Q_{OR}) - Y_{OR}(p - Q_{h}) = \alpha'_h - p\xi'_h\). The values \(c'_1\) and \(c'_2\) may take any real value. In terms of \(c'_h\), equation (3.77) is written as

\[
\kappa_{OR}(c) - \kappa_h(c) = \frac{pq\xi'_h - (c - \alpha'_h)\tau'_h}{c'_hV_{OR}V'_h}, \quad (3.79)
\]

In what follows the conditions under which the combination of the two BDTs with the AND rule produces an increase of the combined weighted kappa coefficient are shown.

\text{a)} \quad \text{If} \quad 0 < c'_h < 1 \quad \text{with} \quad h = 1, 2, \quad \text{then} \quad \kappa_h(c'_h) \quad \text{and} \quad \kappa_{OR}(c'_h), \quad \text{can be calculated}, \quad \text{so obtaining} \quad \kappa_h(c'_h) = Y_h \quad \text{and} \quad \kappa_{OR}(c'_h) = Y_{OR} \quad \text{where} \quad \frac{\alpha'_h - \xi'_hQ_{h}}{\psi'_h} > 0 \quad \text{and} \quad \frac{\alpha'_h - \xi'_hQ_{OR}}{\psi'_h} > 0.
as both expressions are the denominators of weighted kappa coefficients. As it is always verified that $Q_{OR} > Q_h$, if $\xi'_h = Y_{OR} - Y_h > 0$ then

$$\frac{\omega_h' - \xi'_h Q_{OR}}{\psi_h'} - \frac{\omega_h' - \xi'_h Q_h}{\psi_h'} > 0.$$  

By realizing algebraic operations one obtains

$$\frac{\omega_h' - \xi'_h Q_{OR}}{\psi_h'} - \frac{\omega_h' - \xi'_h Q_h}{\psi_h'} = \frac{\xi'_h \tau_h}{\psi_h'} > 0,$$

and as $\xi'_h > 0$ and $\tau_h < 0$, then $\psi_h > 0$. If $\xi'_h = Y_{OR} - Y_h < 0$ then

$$\frac{\omega_h' - \xi'_h Q_{OR}}{\psi_h'} - \frac{\omega_h' - \xi'_h Q_h}{\psi_h'} < 0,$$

and by realizing algebraic operations one obtains

$$\frac{\omega_h' - \xi'_h Q_{OR}}{\psi_h'} - \frac{\omega_h' - \xi'_h Q_h}{\psi_h'} = \frac{\xi'_h \tau_h}{\psi_h'} < 0.$$  

As $\xi'_h < 0$ and $\tau'_h < 0$, then $\psi'_h < 0$. Thus, whatever the value of $\xi'_h$, it is always verified that $\psi'_h < 0$. As $c'_h = \frac{q \omega_h'}{\psi_h'} > 0$ then $\omega'_h < 0$. If $0 < c'_h < c \leq 1$ then, from equation (3.79), $\kappa_{OR}(c) > \kappa_h(c)$ with $h = 1, 2$. Finally, if $0 < \max\{c'_1, c'_2\} < c \leq 1$ then $\kappa_{OR}(c) > \max\{\kappa_1(c), \kappa_2(c)\}$.  

If $c = 0$ then $c'_h - 1 < 0$, and from equation (3.79) it is verified that $\kappa_{OR}(0) < \kappa_h(0)$ with $h = 1, 2$. Therefore $\kappa_{OR}(0) < \min\{\kappa_1(0), \kappa_2(0)\}$.  

136
b) From a) it can be verified that \( \kappa_{OR}(c) > \kappa_i(c) \) if \( 0 < c' < c \leq 1 \). Let us suppose that \( c' > 1 \) and that \( Y_j < Y_{OR} \), then \( \xi'_j = Y_{OR} - Y_j > 0 \). Let us suppose that \( \omega'_j < 0 \), as \( c' > 1 \) then \( \psi'_j = \omega'_j - p\xi'_j \) must be smaller than 0 and \( q\omega'_j < \omega'_j - p\xi'_j < 0 \), and therefore it is verified that \( p(\xi'_j - \omega'_j) < 0 \) and therefore \( \xi'_j < \omega'_j \). As \( \xi'_j > 0 \) it is not possible that \( \omega'_j < 0 \). Finally \( \omega'_j > 0 \) and from equation (3.79) it is verified that \( \kappa_{OR}(c) > \kappa_j(c) \) for \( 0 \leq c \leq 1 \).

In the case that \( Y_j > Y_{OR} \), then \( \xi'_j < 0 \). By multiplying by \( Q_j \) one obtains

\[
\xi'_j Q_j < 0 \Rightarrow Q_j Y_{OR} - Q_j Y_j = Q_j Y_{OR} - (\tau'_j + Q_{OR}) Y_j = \omega'_j - \tau'_j Y_j < 0 \Rightarrow \omega'_j < \tau'_j Y_j < 0 ,
\]

since \( \tau'_j < 0 \). Multiplying by \( Q_{OR} \) one obtains

\[
\xi'_j Q_{OR} < 0 \Rightarrow Q_{OR} Y_{OR} - Q_{OR} Y_j = (Q_j - \tau'_j) Y_{OR} - Q_{OR} Y_j = \omega'_j - \tau'_j Y_{OR} < 0 \Rightarrow \omega'_j < \tau'_j Y_{OR} < 0 .
\]

Hence, it is verified that \( \omega'_j < \tau'_j Y_j < \tau'_j Y_{OR} < 0 \). Therefore, if \( Y_j < Y_{OR} \) then \( \omega'_j < 0 \), and from equation (3.79) it can be verified that \( \kappa_j(c) > \kappa_{OR}(c) \) for \( 0 \leq c \leq 1 \).

If \( c' < 0 \) the reasoning is the same as for \( c' > 1 \). If \( \xi_j > 0 \) then \( \omega'_j > 0 \), and from equation (3.79) it is verified that \( \kappa_{OR}(c) > \kappa_j(c) \) for \( 0 \leq c \leq 1 \). Therefore, if \( \xi_j < 0 \) then \( \omega'_j < 0 \), and from equation (3.79) it can be verified that \( \kappa_j(c) > \kappa_{OR}(c) \) for \( 0 \leq c \leq 1 \).

For any situation different to those that have been dealt with in points a) and b), the combination of the two BDTs does not increase the value of \( \kappa_{OR}(c) \) for any value of \( c \). These situations are:
1) $0 < c'_i < c \leq 1$, $c'_j > 1$ (or $c'_j < 0$) and $Y_{OR} \leq Y_j$, \\

2) $\min\{c'_1, c'_2\} > 1$, \\

3) $\max\{c'_1, c'_2\} < 0$, \\

4) $c'_i > 1$ and $c'_j < 0$. \\

In the following each of these is analysed.

1) The situation $0 < c'_i < c \leq 1$, $c'_j > 1$ (or $c'_j < 0$) and $Y_{OR} < Y_j$ has been dealt with in point b) above. Let us study the same situation but supposing that $Y_{OR} = Y_j$. If $Y_{OR} = Y_j$ then $\xi' = 0$. As $c'_j = \frac{q \alpha'_j}{\alpha'_j - p \xi'_j}$, if $\xi' = 0$ then $c'_j = q$, but this is incompatible with $c'_i > 1$ or $c'_j < 0$, because $0 < q < 1$.

2) Let us consider that $c'_1 > 1$ and $c'_2 > 1$, i.e. $\min\{c'_1, c'_2\} > 1$. Each equation $c'_h > 1$, $h = 1, 2$, has two solutions:

\[
0 < p \xi'_h < \alpha'_h < \xi'_h < 1 \quad (3.80)
\]

and

\[
-1 < \xi'_h < \alpha'_h < p \xi'_h < 0 \quad (3.81)
\]

Let us consider

\[
c'_h (1-c'_h) = pq \frac{\alpha'_h (\alpha'_h - \xi'_h)}{\nu_h^2} < 0,
\]

then $c'_i (1-c'_i) c'_j (1-c'_j) > 0$, i.e.
\[ c'_1(1 - c'_1)c'_2(1 - c'_2) = p^2q^2 \frac{\omega'_1\omega'_2(\omega'_1 - \xi'_1)(\omega'_2 - \xi'_2)}{\psi'_1^2\psi'_2^2} > 0. \] (3.82)

From equation (3.74) it is verified that \( c'_1(1 - c'_1)c'_2(1 - c'_2) > 0 \) if

\[ \omega'_1\omega'_2(\omega'_1 - \xi'_1)(\omega'_2 - \xi'_2) > 0. \] (3.83)

Moreover,

\[ c'_1c'_2 = q^2 \frac{\omega'_1\omega'_2}{\psi'_1\psi'_2^2} > 1. \] (3.84)

Solving the system formed by equations (3.83) and (3.84) with \( 0 < p < 1 \), the solutions for this system which are compatible with solutions (3.80) and (3.81) are

\[ 0 < p\xi'_1 < \omega'_1 < \xi'_1 < 1 \text{ and } -1 < \xi'_2 < \omega'_2 < p\xi'_2 < 0 \]

and

\[ -1 < \xi'_1 < \omega'_1 < p\xi'_1 < 0 \text{ and } 0 < p\xi'_2 < \omega'_2 < \xi'_2 < 1. \]

In general terms, the solution is

\[ -1 < \xi'_j < \omega'_j < p\xi'_j < 0 < p\xi'_i < \omega'_i < \xi'_i < 1. \]

As \( \omega'_j < 0 \text{ y } c'_j > 1 \), then applying equation (3.79) it holds that \( \kappa_{OR}(c) < \kappa_j(c) \) for \( 0 \leq c \leq 1 \). As \( \omega'_j > 0 \text{ and } c'_j > 1 \), then \( \kappa_{OR}(c) > \kappa_j(c) \) for \( 0 \leq c \leq 1 \). Finally,

\[ \kappa_i(c) < \kappa_{OR}(c) < \kappa_j(c), \text{ } 0 \leq c \leq 1. \]

Consequently, if \( \min\{c'_1, c'_2\} > 1 \) then the combination of the two weighted kappa coefficients never increases the value of the combined weighted kappa coefficient. Furthermore, as \( \xi'_j < 0 \) then \( Y_{OR} < Y_j \), and as \( \xi'_i > 0 \) then \( Y_{OR} > Y_j \), i.e.
$Y_i < Y_{OR} < Y_j$. 

3) Another possible situation is that $c'_i < 0$ and $c'_2 < 0$, i.e. $\max\{c'_i, c'_2\} < 0$, then the solutions for each equation $c'_h < 0$, $h = 1, 2$, are $0 < \omega'_h < p\xi'_h < \xi'_h < 1$ or $-1 < \xi'_h < p\xi'_h < \omega'_h < 0$. Solving the system formed by the equations

$$c'_1 (1 - c'_1) c'_2 (1 - c'_2) = p^2 q^2 \frac{\omega'_1 \omega'_2 (\omega'_1 - \xi'_1)(\omega'_1 - \xi'_2)}{\psi'_1 \psi'_2} > 0$$

and

$$c'_1 c'_2 = q^2 \frac{\omega'_1 \omega'_2}{\psi'_1 \psi'_2} > 0,$$

the appropriate solution is $-1 < \xi'_j < p\xi'_j < \omega'_j < 0 < \omega'_i < p\xi'_i < \xi'_i < 1$. Finally, it holds that

$$\kappa_i (c) < \kappa_{OR} (c) < \kappa_j (c), \quad 0 \leq c \leq 1,$$

and

$$Y_i < Y_{OR} < Y_j.$$

4) Finally, another possible situation is that $c_i > 1$ and $c_j < 0$. Solving the system formed by the equations

$$c'_1 (1 - c'_1) c'_2 (1 - c'_2) = p^2 q^2 \frac{\omega'_1 \omega'_2 (\omega'_1 - \xi'_1)(\omega'_1 - \xi'_2)}{\psi'_1 \psi'_2} > 0$$

and
$c'_1c'_2 = q^2 \frac{\omega_l \omega_j'}{\psi'_l \psi'_j} < 0,$

the solution is $-1 < \xi'_j < p\xi'_j < \omega'_j < 0 < p\xi'_j < \omega'_j < \xi'_i < 1$. In this situation, it is verified that if $\xi'_j < 0$ then $p\xi'_j < \omega'_j < q\omega'_j < 0$ and $\psi'_j = \omega'_j - p\xi'_j > 0$, and therefore $c'_j < 0$. Consequently, if $c'_i > 1$ and $c'_j < 0$, the solution is $-1 < \xi'_j < p\xi'_j < \omega'_j < 0 < p\xi'_j < \omega'_j < \xi'_i < 1$, and it is again verified that

$$\kappa'_i(c) < \kappa_{or}(c) < \kappa'_j(c), \quad 0 \leq c \leq 1,$$

and

$$Y_i < Y_{or} < Y_j.$$ 

### Appendix 3.4: The program “cwkc”

cwkc <- function(s11, s10, s01, s00, r11, r10, r01, r00, indexc, rule, conflevel = 0.95) {

library(graphics)

if (s11 < 0 | s10 < 0 | s01 < 0 | s00 < 0 | r11 < 0 | r10 < 0 | r01 < 0 | r00 < 0) {
  cat("n")
  stop("Any observed frequency can be negative. Introduces new values \n")
  cat("n")
}

if (abs(s00 - trunc(s00)) > 0  | abs(s01 - trunc(s01)) > 0  | abs(s10 - trunc(s10)) > 0  |
abs(s11 - trunc(s11)) > 0  | abs(r00 - trunc(r00)) > 0  | abs(r01 - trunc(r01)) > 0  |
abs(r10 - trunc(r10)) > 0  | abs(r11 - trunc(r11)) > 0)
{
  cat("\n")
  stop("Observed frequencies can not have decimals. Introduces new values \n")
  cat("\n")
}

if ((s11 + s10 + s01 + s00) == 0 | (r11 + r10 + r01 + r00) == 0)
{
  cat("\n")
  stop("Accuracy of a Binary Test cannot be estimated. There are many observed 
frequencies equal to zero. Introduces new values \n")
  cat("\n")
}

if (indexc > 1 | indexc < 0)
{
  cat("\n")
  stop("Weighting index c should take a value between 0 and 1. Introduces a new 
value \n")
  cat("\n")
}

n <- s00 + s01 + s10 + s11 + r00 + r01 + r10 + r11

p <- (s00 + s01 + s10 + s11) / n

Se1 <- (s11 + s10) / (s11 + s10 + s01 + s00)

Se2 <- (s11 + s01) / (s11 + s10 + s01 + s00)
Sp1 <- (r01 + r00) / (r11 + r10 + r01 + r00)

Sp2 <- (r10 + r00) / (r11 + r10 + r01 + r00)

Y1 <- Se1 + Sp1 - 1

Y2 <- Se2 + Sp2 - 1

if (Y1 <= 0 | Y2 <= 0)
{
  cat("n")
  cat("Estimated Youden index of Binary Test 1 is ", Y1, ",n")
  cat("Estimated Youden index of Binary Test 2 is ", Y2, ",n")
  stop("Estimated Youden index of a Binary Test must be greater than zero. Introduces new values ",n")

  cat("n")
}

if (Y1 == 1 | Y2 == 1)
{
  cat("n")
  cat("Estimated Youden index of Binary Test 1 is ", Y1, ",n")
  cat("Estimated Youden index of Binary Test 2 is ", Y2, ",n")
  stop("A Binary Test is a gold standard. Introduces new values ",n")
  cat("n")
}

y <- as.character(rule)
y1 <- "and"
y2 <- "or"
if(y != y1 && y != y2)
{
  cat("n")
  stop("Write and for AND rule. Write or for OR rule \n")
  cat("n")
}

r1 <- identical(y, y1)

r2 <- identical(y, y2)

if(isTRUE(r1))
{
  kappacomb <- expression((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + q11)))

  quot1 <- expression(log(((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + q11)))) / (((p10 + p11) * (q00 + q01) - (p00 + p01) * (q10 + q11)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - p01 - q11 - q01) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + p01 + q11 + q01))))

  quot2 <- expression(log(((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + q11)))) / (((p01 + p11) * (q00 + q10) - (p00 + p10) * (q01 + q11)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - p01 - q11 - q01) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + p01 + q11 + q01))))

  arcsink <- expression(asin(sqrt((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + q11)))))
logitk <- expression(log(((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + q11))) / (1 - ((p11 * (q11 + q10 + q01 + q00) - q11 * (p11 + p10 + p01 + p00)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - q11) + (q11 + q10 + q01 + q00) * (1 - indexc) * (p11 + q11)))))

} else
{
  kappacomb <- expression((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) * (1 - indexc) * (1 - p00 - q00)))
  quot1 <- expression(log(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) * (1 - indexc) * (1 - p00 - q00))) / (((p10 + p11) * (q00 + q01) - (p00 + p10) * (q01 + q11)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - p10 - q11 - q10) + (q11 + q10 + q01 + q00) * (1 - indexc)* (p11 + p10 + q11 + q10))))
  quot2 <- expression(log(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) * (1 - indexc) * (1 - p00 - q00))) / (((p10 + p11) * (q00 + q01) - (p00 + p10) * (q01 + q11)) / ((p11 + p10 + p01 + p00) * indexc * (1 - p11 - p01 - q11 - q01) + (q11 + q10 + q01 + q00) * (1 - indexc)* (p11 + p01 + q11 + q10))))
  arcsink <- expression(asin(sqrt(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) * (1 - indexc) * (1 - p00 - q00)))))
  logitk <- expression(log(((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) * (1 - indexc) * (1 - p00 - q00))) / (1 - ((q00 * (p11 + p10 + p01 + p00) - p00 * (q11 + q10 + q01 + q00)) / ((p11 + p10 + p01 + p00) * indexc * (p00 + q00) + (q11 + q10 + q01 + q00) * (1 - indexc) * (1 - p00 - q00))))
\begin{align*}
+ q_01 + q_00)) / ((p_{11} + p_{10} + p_{01} + p_{00}) * \text{indexc} * (p_{00} + q_{00}) + (q_{11} + q_{10} + q_{01} + q_{00}) * (1 - \text{indexc}) * (1 - p_{00} - q_{00}))))
\end{align*}

\begin{align*}
Q_1 & \gets p * \text{Se1} + (1 - p) * (1 - \text{Sp1}) \\
k_1 & \gets (p * (1 - p) * Y_1) / (p * \text{indexc} * (1 - Q_1) + (1 - p) * (1 - \text{indexc}) * Q_1) \\
Q_2 & \gets p * \text{Se2} + (1 - p) * (1 - \text{Sp2}) \\
k_2 & \gets (p * (1 - p) * Y_2) / (p * \text{indexc} * (1 - Q_2) + (1 - p) * (1 - \text{indexc}) * Q_2) \\
z & = \text{qnorm}((1 + \text{conflevel})/2,0,1) \\
\alpha & \gets 1 - \text{conflevel}
\end{align*}

\begin{align*}
kappa_1 & \gets \text{expression}(((p_{10} + p_{11}) * (q_{00} + q_{01}) - (p_{00} + p_{01}) * (q_{10} + q_{11})) / ((p_{11} + p_{10} + p_{01} + p_{00}) * \text{indexc} * (1 - p_{11} - p_{10} - q_{11} - q_{10}) + (q_{11} + q_{10} + q_{01} + q_{00}) * (1 - \text{indexc}) * (p_{11} + p_{10} + q_{11} + q_{10}))) \\
kappa_2 & \gets \text{expression}(((p_{01} + p_{11}) * (q_{00} + q_{10}) - (p_{00} + p_{10}) * (q_{01} + q_{11})) / ((p_{11} + p_{10} + p_{01} + p_{00}) * \text{indexc} * (1 - p_{11} - p_{01} - q_{11} - q_{01}) + (q_{11} + q_{10} + q_{01} + q_{00}) * (1 - \text{indexc}) * (p_{11} + p_{01} + q_{11} + q_{01}))) \\
# \text{Derivatives}
\end{align*}

derivk1p00 & \gets \text{deriv}(kappa1, "p00") \\
derivk1p01 & \gets \text{deriv}(kappa1, "p01") \\
derivk1p10 & \gets \text{deriv}(kappa1, "p10") \\
derivk1p11 & \gets \text{deriv}(kappa1, "p11") \\
derivk1q00 & \gets \text{deriv}(kappa1, "q00") \\
derivk1q01 & \gets \text{deriv}(kappa1, "q01")
derivk1q10 <- deriv(kappa1, "q10")
derivk1q11 <- deriv(kappa1, "q11")

derivk2p00 <- deriv(kappa2, "p00")
derivk2p01 <- deriv(kappa2, "p01")
derivk2p10 <- deriv(kappa2, "p10")
derivk2p11 <- deriv(kappa2, "p11")
derivk2q00 <- deriv(kappa2, "q00")
derivk2q01 <- deriv(kappa2, "q01")
derivk2q10 <- deriv(kappa2, "q10")
derivk2q11 <- deriv(kappa2, "q11")

derivkappap00 <- deriv(kappacomb, "p00")
derivkappap01 <- deriv(kappacomb, "p01")
derivkappap10 <- deriv(kappacomb, "p10")
derivkappap11 <- deriv(kappacomb, "p11")
derivkappaq00 <- deriv(kappacomb, "q00")
derivkappaq01 <- deriv(kappacomb, "q01")
derivkappaq10 <- deriv(kappacomb, "q10")
derivkappaq11 <- deriv(kappacomb, "q11")

dlogitkappap00 <- deriv(logitk, "p00")
dlogitkappap01 <- deriv(logitk, "p01")
dlogitkappap10 <- deriv(logitk, "p10")
dlogitkappap11 <- deriv(logitk, "p11")
dlogitkappaq00 <- deriv(logitk, "q00")
dlogitkappaq01 <- deriv(logitk, "q01")
dlogitkappaq10 <- deriv(logitk, "q10")
dlogitkappaq11 <- deriv(logitk, "q11")

# Variances and covariances

\[ p_{00} \leftarrow \frac{s_{00}}{n} \]
\[ p_{01} \leftarrow \frac{s_{01}}{n} \]
\[ p_{10} \leftarrow \frac{s_{10}}{n} \]
\[ p_{11} \leftarrow \frac{s_{11}}{n} \]
\[ q_{00} \leftarrow \frac{r_{00}}{n} \]
\[ q_{01} \leftarrow \frac{r_{01}}{n} \]
\[ q_{10} \leftarrow \frac{r_{10}}{n} \]
\[ q_{11} \leftarrow \frac{r_{11}}{n} \]

\text{vec1} \leftarrow \text{vector("numeric", 8)}

\text{vec1}[1] \leftarrow p_{00}
\text{vec1}[2] \leftarrow p_{01}
\text{vec1}[3] \leftarrow p_{10}
\text{vec1}[4] \leftarrow p_{11}
\text{vec1}[5] \leftarrow q_{00}
\text{vec1}[6] \leftarrow q_{01}
\text{vec1}[7] \leftarrow q_{10}
\text{vec1}[8] \leftarrow q_{11}

\text{matp} \leftarrow \text{matrix(0, 8, 8)}

\text{matp}[1,1] \leftarrow p_{00}
\text{matp}[2,2] \leftarrow p_{01}
\text{matp}[3,3] \leftarrow p_{10}
\text{matp}[4,4] \leftarrow p_{11}
\text{matp}[5,5] \leftarrow q_{00}
\text{matp}[6,6] \leftarrow q_{01}
\text{matp}[7,7] \leftarrow q_{10}
matp[8,8] <- q11

sigmap <- matrix(0, 8, 8)

sigmap <- (1 / n) * (matp - vec1 %*% t(vec1))

vec2 <- vector("numeric", 8)

vec2[1] <- attr(eval(derivk1p00), "gradient")[1]
vec2[2] <- attr(eval(derivk1p01), "gradient")[1]
vec2[3] <- attr(eval(derivk1p10), "gradient")[1]
vec2[4] <- attr(eval(derivk1p11), "gradient")[1]
vec2[5] <- attr(eval(derivk1q00), "gradient")[1]
vec2[6] <- attr(eval(derivk1q01), "gradient")[1]
vec2[7] <- attr(eval(derivk1q10), "gradient")[1]
vec2[8] <- attr(eval(derivk1q11), "gradient")[1]

vec3 <- vector("numeric", 8)

vec3[1] <- attr(eval(derivk2p00), "gradient")[1]
vec3[2] <- attr(eval(derivk2p01), "gradient")[1]
vec3[3] <- attr(eval(derivk2p10), "gradient")[1]
vec3[4] <- attr(eval(derivk2p11), "gradient")[1]
vec3[5] <- attr(eval(derivk2q00), "gradient")[1]
vec3[6] <- attr(eval(derivk2q01), "gradient")[1]
vec3[7] <- attr(eval(derivk2q10), "gradient")[1]
vec3[8] <- attr(eval(derivk2q11), "gradient")[1]

mat1 <- rbind(vec2, vec3)
sigma1 <- matrix(0, 2, 2)

sigma1 <- mat1 %*% sigmap %*% t(mat1)

vec4 <- vector("numeric", 8)

vec4[1] <- attr(eval(derivkappap00), "gradient")[1]
vec4[2] <- attr(eval(derivkappap01), "gradient")[1]
vec4[3] <- attr(eval(derivkappap10), "gradient")[1]
vec4[4] <- attr(eval(derivkappap11), "gradient")[1]

vec4[5] <- attr(eval(derivkappaq00), "gradient")[1]
vec4[6] <- attr(eval(derivkappaq01), "gradient")[1]
vec4[7] <- attr(eval(derivkappaq10), "gradient")[1]
vec4[8] <- attr(eval(derivkappaq11), "gradient")[1]

mat2 <- rbind(vec4, vec2)

sigma2 <- matrix(0, 2, 2)

sigma2 <- mat2 %*% sigmap %*% t(mat2)

mat3 <- rbind(vec4, vec3)

sigma3 <- matrix(0, 2, 2)

sigma3 <- mat3 %*% sigmap %*% t(mat3)

vec5 <- vector("numeric", 8)

vec5[1] <- attr(eval(dlogitkappap00), "gradient")[1]
vec5[2] <- attr(eval(dlogitkappap01), "gradient")[1]
vec5[3] <- attr(eval(dlogitkappap10), "gradient")[1]
vec5[4] <- attr(eval(dlogitkappap11), "gradient")[1]
vec5[5] <- attr(eval(dlogitkappaq00), "gradient")[1]
vec5[6] <- attr(eval(dlogitkappaq01), "gradient")[1]
vec5[7] <- attr(eval(dlogitkappaq10), "gradient")[1]
vec5[8] <- attr(eval(dlogitkappaq11), "gradient")[1]

var1 <- t(vec5) %*% sigmap %*% vec5

if(isTRUE(r1))
{
  combSe <- p11 / p

  combSp <- (q10 + q01 + q00) / (1 - p)

  combQ <- p11 + q11
}

else
{
  combSe <- (p11 + p10 + p01) / p

  combSp <- q00 / (1 - p)

  combQ <- 1 - p00 - q00
}

combY <- combSe + combSp - 1

combwkc <- (p * (1 - p) * combY) / (p * indexc * (1 - combQ) + (1 - p) * (1 - indexc) * combQ)
logitcwkc <- log(combwkc / (1 - combwkc))

# Fieller CI

d11 <- combwkc * k1 - sigma2[1,2] * z^2

d21 <- combwkc^2 - sigma2[1,1] * z^2

d31 <- k1^2 - sigma2[2,2] * z^2

LFiellerkappak1 <- (d11 - sqrt(d11^2 - d21 * d31)) / d31

UFiellerkappak1 <- (d11 + sqrt(d11^2 - d21 * d31)) / d31


d12 <- combwkc * k2 - sigma3[1,2] * z^2

d22 <- combwkc^2 - sigma3[1,1] * z^2

d32 <- k2^2 - sigma3[2,2] * z^2

LFiellerkappak2 <- (d12 - sqrt(d12^2 - d22 * d32)) / d32

UFiellerkappak2 <- (d12 + sqrt(d12^2 - d22 * d32)) / d32

# Logit CI

Llogitkappa <- exp(logitcwkc - z * sqrt(var1[1])) / (1 + exp(logitcwkc - z * sqrt(var1[1])))

Ulogitkappa <- exp(logitcwkc + z * sqrt(var1[1])) / (1 + exp(logitcwkc + z * sqrt(var1[1])))
# Solutions for Kand = Ki and Kor = Ki

c1 <- (1 - p) * (Q1 * combY - combQ * Y1) / (Y1 * (p - combQ) - combY * (p - Q1))

c2 <- (1 - p) * (Q2 * combY - combQ * Y2) / (Y2 * (p - combQ) - combY * (p - Q2))

# Graphics

funct1 <- function (x) (p * (1 - p) * Y1) / (p * x * (1 - Q1) + (1 - p) * (1 - x) * Q1)

funct2 <- function (x) (p * (1 - p) * Y2) / (p * x * (1 - Q2) + (1 - p) * (1 - x) * Q2)

functkappa <- function (x) (p * (1 - p) * combY) / (p * x * (1 - combQ) + (1 - p) * (1 - x) * combQ)

if(isTRUE(r1)) tex <- "AND rule" else tex <- "OR rule"

jpeg("Graph_cwkc.jpg")

curve(functkappa, 0, 1, lty = 1, ylim = c(0, 1), xlim = c(0, 1), xaxp = c(0, 1, 1), main = tex, xlab = "Weighting index", ylab = "Weighted kappa coefficient")

curve(funct1, 0, 1, lty = 2, add = TRUE)

curve(funct2, 0, 1, lty = 3, add = TRUE)

axis(1, at = c(round(c1, digits = 3), round(c2, digits = 3)))

abline(v = c(round(c1, digits = 3), round(c2, digits = 3)), lty = 2, lwd = .1, col = "gray90")


legend("top", lab, lty = c(1, 2, 3), ncol = 1, merge = TRUE)

dev.off()

#Result

sink("Results_cwkc.txt", split=TRUE)

cat("\n")

cat(" R E S U L T S 
")

cat("---------------
")

cat("     WEIGHTED KAPPA COEFFICIENTS OF THE BDTs \n")
cat("\n")
cat("Weighting index c is: ", indexc, "\n")
cat("\n")
cat("Estimated weighted kappa coefficient of Test 1 is ", k1, " and its standard error is ", sqrt(sigma1[1,1]), "\n")
cat("\n")
cat("Estimated weighted kappa coefficient of Test 2 is ", k2, " and its standard error is ", sqrt(sigma1[2,2]), "\n")
cat("\n")
cat("Estimated Youden index of Test 1 is: ", Y1, "\n")
cat("\n")
cat("Estimated Youden index of Test 2 is: ", Y2, "\n")
cat("\n")
cat(" COMBINATION OF THE TWO BDTs WITH THE ", rule, " RULE \n")
cat("\n")
cat("Estimated combined weighted kappa coefficient is ", combwkc, " and its standard error is ", sqrt(sigma2[1,1]), "\n")
cat("\n")
cat("Estimated Youden index is: ", combY, "\n")
cat("\n")
cat("Estimated combined weighted kappa coefficient is equal to weighted kappa coefficient of Test 1 if c index is: ", c1, "\n")
cat("\n")
cat("Estimated combined weighted kappa coefficient is equal to weighted kappa coefficient of Test 2 if c index is: ", c2, "\n")
cat("\n")
cat(100 * conflevel, "% Fieller CI for combined weighted kappa coefficient / K1(c) is: (", LFiellerkappak1, "; ", UFiellerkappak1, ") \n")
cat("\n")
cat(100 * conflevel, "% Fieller CI for combined weighted kappa coefficient / K2(c) is: (", LFiellerkappak2, "; ", UFiellerkappak2, ") \n")
cat("\n")
cat(100 * conflevel,"% Logit CI for combined weighted kappa coefficient is:
("Llogitkappa, "; ",Ulogitkappa,\") \n")
cat("\n")
sink()
}
References


References


