2013-01-28

Binocular Sensitivity and Specificity of Screening Tests in Prospective Studies of Paired Organs

Perera, Singappuli

UNIVERSITY OF CALGARY

Binocular Sensitivity and Specificity of Screening Tests in Prospective Studies of Paired Organs

by

Singappuli Perera

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS & STATISTICS

CALGARY, ALBERTA

JANUARY, 2013

Abstract

Diagnostic and screening studies in ophthalmology frequently involve binocular data where pairs of eyes are evaluated, through some diagnostic procedure, for the presence of certain diseases or pathologies. It is usually sufficient in practice that at least one eye is positively diagnosed for the patient to be sent for further and more extensive eye examination. More relevant diagnostic accuracy measures in these cases are therefore the probability of at least one correct positive diagnosis in patients with one or both eyes truly diseased and the probability of two correct negative diagnoses for patients with both eyes truly un-diseased. The former is analogous to sensitivity and the latter to specificity. Predictive values may be similarly re-defined.

The thesis proposes these new sensitivity and specificity measures as alternatives to conventional ones for paired binocular binary diagnostic data arising from screening studies with cross-sectional sampling. The measures are defined for flexible models based on copulas and extensions of existing models for correlated binary data. The proposed methodology is illustrated with data from a study on diabetic retinopathy.

Acknowledgements

I would like to express my deepest gratitude to many people who offered their generous guidance and help during my study and research at University of Calgary. First of all, my sincerest and deepest thanks go to my supervisor Dr. Alexander R. de Leon for his invaluable assistance, encouragement, excellent advice, guidance, and support during my graduate studies. I really enjoyed working with him and learned quite a lot from him as well. His carefulness, diligence, and critical thinking had a significant impact on my research work. It is my great honour to work with him.

Secondly, I would like to express my deepest gratitude to Drs. Jingjing Wu, Hyang Kim, and Jian Kang in my thesis committee, for their time and comments.

I am thankful to the Faculty of Graduate Studies for awarding me a Queen Elizabeth II Graduate Scholarship from the Alberta Government, and the Department of Mathematics and Statistics for their generous financial support during my graduate studies.

I would also like to thank everyone who supported me throughout the course of the thesis, especially, Niroshan Withanage, Yanmei Fei, and all statistics graduate students who gave me inspiration and helpful suggestions.

Last but not least, my deepest love and gratitude go to my husband, my parents, my lovely son, my sister, and my brother for their incredible support, encouragement, and understanding during my studies.

Singappuli Perera
January 2013
Calgary AB, Canada

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Thesis background

Patients frequently undergo several diagnostic tests or are subjected to the same test on repeated occasions. This is especially true in situations where test outcomes are reader-based assessments, in which case two or more readers are necessary to minimize, if not eliminate, so-called reader bias. Simply assuming that the test results are independent is unwise, since there are correlations among the resulting evaluations. For instance, test results from the same patient are frequently correlated while test results from different patients are still independent, as in the diabetic retinopathy study described below. The proper accounting of associations which intuitively exist between measurements taken from fellow eyes of a patient is an interesting statistical question. Failure to account for such correlations by treating eyes as independent may consequently yield incorrect inferences. Since valid information contained in correlated observations is less than expected from independent observations, underestimation of standard errors (SEs) could happen, yielding inflated Type I errors of significance tests. Thus, methods that account for this correlation are needed.

The development of this thesis is motivated by a study on diabetic patients in Alberta, Canada, who suffer from treatable diabetic retinopathy. The study entailed screening for the presence or absence of certain ophthalmologic pathologies such as clinically significant macular edema (CSME), microaneurysms, intra-retinal haemorrhage, and others, that are indicative of retinal thickening [1]. Due to advances in digital imaging in recent years, this diagnosis can be performed at a distance using

digital images, a cost-effective screening approach for health-care burdens compared to the accepted 'gold standard' of contact lens biomicroscopy to identify patients who need further assessment. In such a tele-ophthalmologic screening system, digital images of patients' eyes are read remotely by specialists, and patients are diagnosed as either positive or negative for diseases. As a result, only patients who need treatment would have to travel to specialists in urban centres like Edmonton and Calgary so that transportation cost is also reduced. For example, in Canada, where a disproportionate number of diabetic patients are aboriginal Canadians living in reserves in far-flung remote areas, sending retinal specialists on remote clinics can be costly and inefficient. Using distance evaluation of retinopathy-related pathologies based on digital images of diabetic patients eyes is a potentially cost-effective alternative to clinical examination. However, before wide implementation of any potential new diagnostic methodology, its accuracy must first be examined. The purpose of the study is thus to determine whether diabetic retinopathy can be identified with high-resolution stereoscopic digital photography, and whether this identification correlates well with the accepted gold standard of clinical examination.

The data set-up for the case of one pathology, one reader, and N patients is shown in Table 1, where Y_{i1L} and Y_{i1R} indicate the binary test results (i.e., 1 for positive, 0 for negative) for the left (L) and right (R) eyes, respectively, of patient $i = 1, \cdots, N$. Furthermore, let Y_{i2L} and Y_{i2R} denote the true disease status (i.e., 1 for positive, 0 for negative) for the left and right eyes, respectively, of patient i, as determined by the gold standard. Note that there are correlations between the reader diagnoses and between the disease status of fellow eyes.

The accuracy of a medical test for diagnosing the presence (positive) or absence (negative) of a disease can be described by several measures, the most common of which are given by the test's sensitivity and specificity with respect to a traditionally

Table 1.1: Data set-up for a pathology with one reader for the left and right eyes of N patients.

Patient	Diagnosis		Disease status	
	Left eye	Right eye	Left eye	Right eye
1	Y_{11L}	Y_{11R}	Y_{12L}	Y_{12R}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	Y_{i1L}	Y_{i1R}	Y_{i2L}	Y_{i2R}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	Y_{N1L}	Y_{N1R}	Y_{N2L}	Y_{N2R}

adopted test regarded as the 'gold standard.' Sensitivity is the probability that the new test indicates presence of the disease when the gold standard indicates that it is present, while specificity is the probability that the new test shows absence of disease when the gold standard also gives a negative result. Given binary variables Y_1 and Y_2 denoting the respective results from the new test and the disease status as determined by the gold standard, with 0 and 1 indicating negative and positive results, respectively, the new test's sensitivity and specificity are then given by $\text{Sen} = P(Y_1 = 1|Y_2 = 1)$ and $\text{Sp} = P(Y_1 = 0|Y_2 = 0)$, respectively. The sensitivity Sen and specificity Sp are also known by several names. For example, Sen is also referred to as the true positive fraction (TPF) or the true alarm rate (TAR), while $1 - \text{Sp}$ is called the false positive fraction (FPF) or the false alarm rate (FAR); Sen is also known as the hit rate in engineering applications. In statistical hypothesis testing, Sen is known as the statistical power while Sp corresponds to the statistical significance.

From a statistical point of view, there is no difference between screening and diagnostic tests; they are used interchangeably in the context of statistics, with both treated as medical tests. While this may be appropriate in many applications, screening for diseases based on paired organs (e.g., fellow ears, fellow eyes) calls for a modification of the above conventional measures of diagnostic accuracy. This is because, in practice,

a positive screening test can be allowed to be somewhat less than perfect, and in fact, is traditionally followed not by treatment but by a confirmatory diagnosis via a better, possibly more expensive and more invasive, test. Hence, a diabetic patient in the above study, for example, needs only to be correctly positively screened for retinopathy in one eye, instead of both eyes, before he or she can be sent for further examination. That is, a screening test in this case requires only that its 'partial' TPF be 'acceptably high'. Conversely, a diabetic patient should be considered 'negative' for a disease only when both his eyes are negatively screened, which suggests that the screening test's 'full' FPF (i.e., both negative eyes are screened as negative) be 'acceptably low'.

The above modifications are necessary even in the case of exchangeability, where one organ is indistinguishable from the other. While diagnostic accuracy measures such as Sen and Sp do not depend on the particular organ in this case, a joint measure akin to the partial TPF and full FPF above is useful in capturing the diagnostic power of screening tests. The main objective of this thesis is thus to propose new diagnostic accuracy measures for screening tests based on paired organs. The new 'binocular' measures are defined for two particular models for correlated binocular binary outcomes. The first model is an extension of the so-called common correlation model (CCM) and the second is constructed from the Gaussian copula. We discuss these models in Chapters 2 and 3 after we present a brief review of the literature in what follows.

1.2 Literature review

Traditional estimation methods that ignore correlations in clustered diagnostic data have been shown to be inadequate as they underestimate standard errors and lead to incorrect estimates [2]; methods that account for these correlations are thus needed.

There has been previous work that address this problem on the estimation of sensitivity and specificity and the calculation of their estimates' standard errors in the context of clustered binary diagnostic data. These include simple adjustments to standard errors; likelihood-based approaches via beta-binomial models [3, 4], various correlated binomial models, and the common correlation model (CCM) for binary data; and regression-based methods such as the generalized estimating equations (GEE) approach [5], variations of logistic regression, and generalized linear models (GLMs) with specified joint distribution. Before we review statistical approaches for incorporating correlation in binocular (or paired) data settings, further details on diagnostic and screen tests for binocular settings, especially in ophthalmological studies, are discussed below.

1.2.1 Diagnostic and screening tests for binocular data settings

While multivariate methods such as multiple regression for normally distributed data or multiple logistic regression for binary data are commonly used in epidemiologic research, they are often not directly suitable for ophthalmological (or otolaryngological) studies because these methods work under the assumption of independence between individual sample points. Yet eyes are the fundamental units of analysis in these applications, and frequently exhibit strong correlations between outcomes on the right and left eyes (or ears) [4]. A number of strategies for dealing with this problem in binocular diagnostic data in ophthalmology include the following [4]:

1. analyze data on an eye-specific basis by ignoring the inter-eye correlation;

2. analyze data on a patient-specific basis by using the result only from the better (or worse) eye, or the left (or right) eye;

3. analyze data on a patient-specific basis by using the average or difference of results

for fellow eyes; or

4. analyze data on an eye-specific basis without ignoring the correlation between fellow eyes.

The first strategy considers that a sample of N patients contribute data from $2N$ independent eyes, if the data are available from both eyes for each patient. When correlation exists between fellow eyes from the same patient, the analysis usually yield invalid inferences; for instance, p-values of significance tests are too low, thus misstating the true significance. In addition, the standard errors of estimates are also too low, which yield confidence intervals that are too narrow.

The second strategy is applicable but inefficiently uses data, since 50% of the observations are discarded. Additionally, it may lead to definitional problems when several eye-specific correlated variables are under study. For instance, in the diabetic retinopathy study, it is possible for one patient to have the left eye showing absence of CSME and presence of microaneurysms, and the right eye showing presence of CSME and absence of microaneurysms. Finally, if the right and left eyes are treated separately, differing results may be obtained. Therefore, the standard for 'better' becomes problematic and should be carefully considered.

The third strategy also represents a valid but inefficient method of analysis because averaging precludes some eye-specific covariates, which is often of primary interest, and differencing only allows eye-specific covariates. In addition, like the second strategy, this strategy becomes inapplicable when there are missing data.

The fourth strategy is potentially the best but most difficult to use in practice. Since it utilizes data on both eyes and accounts for the correlation between fellow eyes in performing significance tests and constructing confidence intervals, it yields the most complete analysis of the data. The difficulty here lies in specifying a flexible model

for the data on both eyes that incorporates the association between fellow eyes. Most recent research have focused on this strategy with particular efforts to analyze data on fellow eyes marginally as well as to account for their correlations. We adopt this strategy in this thesis and consider two models for the correlated binocular binary diagnostic data from the diabetic retinopathy study described above.

1.2.2 Statistical analysis of binocular data

1.2.2.1 Adjustments to standard errors

Models based on independence assumptions between fellow eyes usually overstate the informative sample size. Asymptotic theory on maximum likelihood estimates (MLEs), for example, states that asymptotic standard errors are inversely related to $\sqrt{N}$, where N is the number of patients (i.e., clusters). An overstated sample size thus leads to underestimated standard errors, thus yielding inflated Type I errors and p-values for significance tests. Adjustments to standard errors as a remedy for this problem can be interpreted as adjustments to the effective sample size. These include simple adjustments to standard errors introduced in [6, 7] to account for the intra-cluster correlation, and a weighted estimator proposed in [8], for handling unbalanced cluster sizes. All these methods utilize weighted sample sizes calculated based on data, and standard errors are obtained according to weighted coefficients. While they provide simple calculations by weighted estimation, which work with any kind of correlation structure, these calculations fail to provide correlation estimates and are incapable of dealing with data with varying cluster sizes. They have been shown to be inferior to GEE in moderate to large samples and moderate intra-cluster correlations [9].

1.2.2.2 Regression-based approaches

Regression models are commonly adopted in diagnostic studies, since analyses based on them allow direct incorporation of covariate effects, making comparison among dif-

ferent tests straightforward. Such models often consist of two parts, the marginal models and the joint distribution for likelihood representation. For example, polychotomous logistic regression was adapted to correlated ophthalmologic data in [4]. A similar approach based on weighting was recently discussed in [10] in the context of logistic regression analysis of binocular ophthalmologic data. A regression method based on the GEE approach [11] to deal with clustered binary diagnostic data is described in [5, 12, 13]. A hybrid marginal GLM based on a correlated binary distribution to estimate sensitivity and specificity in binocular settings is studied recently in [14]. Although GEE methods yield consistent and asymptotically normally distributed estimates, one drawback is the dearth of model selection and assessment criteria.

1.2.2.3 Model-based approaches

Several families of distributions for correlated binomial data can be adopted for the study of diagnostic sensitivity and specificity. An extension to the common correlation model (CCM) was recently introduced in [15], where simple and easy-to-calculate estimates of measures of diagnostic accuracy for binocular binary data were studied. The method is based on the CCM for correlated binary data [16], which was previously studied in [17]. The methodology is flexible enough to delineate the different associations in the binocular data; in addition, it yields convenient marginalization properties, and thus can be viewed as generalizing other simpler commonly used binary data models [15]. One disadvantage of the model is that it works under the assumption of exchangeability of the binocular data.

1.2.2.4 Copula-based approaches

An alternative strategy involves the use of copulas, as recently discussed by [18], among others. While not new, applications of copulas to discrete data [18–23] have only recently been elucidated and clarified in [24]. Problems arising from the use of

copulas to construct discrete distributions were discussed in [25], where practioners are cautioned that "everything that can go wrong, will go wrong." As showed in [24], a number of complications arise from the direct application of copula models to discrete data. A number of recent work on correlated data analysis with biostatistical applications [26, 27] have adopted copula functions to indirectly specify associated joint distributions. This is only a recent phenomenon in modeling correlated outcomes in health and medicine, and unresolved issues, both methodological and practical, abound, especially as they apply to discrete data.

These models all rely on likelihood-based estimation that incorporate intra-cluster correlations; they are flexible and lend themselves easily to straightforward adaptations of conventional asymptotic tests. However, restrictions on parameter spaces such as those for correlations might be a drawback of these models, as when only positive correlations are permitted by a model. Moreover, the asymptotic theory of MLEs may break down when some boundary conditions are not met, a distinct possibility in models with restricted parameter spaces.

1.3 Thesis overview

This thesis is concerned with the development of models and associated methodologies for the analysis of correlated binocular binary diagnostic data and is motivated by the diabetic retinopathy study described in Section 1.1. It entails the construction of joint models that flexibly represent the complex dependence structures that characterize the nature of relationships between eye measurements on the same patient.

The approach adopted in the first part of the thesis makes use of the extended CCM (ECCM) to model the data from the diabetic retinopathy study. The general formulation of the model includes the well-known beta-binomial model as well as the correlated

binomial models in [28]. The second part of the thesis is based on a model constructed from the Gaussian copula, which is used to analyze the same data. Modified sensitivity and specificity measures for binocular data from the two models are introduced and their estimation is illustrated via the diabetic retinopathy data.

The thesis is organized as follows. Chapter 2 develops the ECCM and includes a discussion of model properties and of the binocular measures of diagnostic accuracy. Chapter 3 begins with a brief review of copula models, followed by a discussion of the copula model for binary binocular diagnostic data; an adaptation of the binocular measures of diagnostic accuracy is studied via a simulation study. In chapter 4, the methodologies are applied to data from the diabetic retinopathy study. Finally, Chapter 5 concludes the thesis with brief discussion of the merits and demerits of the proposed methodologies, and identifies some promising future work.

Chapter 2

Extended common correlation model

2.1 Introduction

The extended common correlation model (ECCM) is a simple model for correlated binocular binary data that yields easy-to-calculate estimates of measures of diagnosis accuracy along with their corresponding standard errors [15]. It is a generalization of the CCM for correlated binary data introduced in [16] and was first introduced in [17]. This general formulation of the CCM includes the well-known beta-binomial model as well as the correlated binomial models in [28]. The model is flexible enough to delineate the different correlations in the binocular data; in addition, it yields convenient marginalization properties which include other simpler commonly used binary data models as special cases [15].

The chapter studies the ECCM for correlated binocular binary diagnostic data and uses it to obtain new binocular measures of diagnostic accuracy as alternatives to conventional measures. The chapter is organized as follows. Section 2.1 discusses the development of the ECCM for a cross-sectional diagnostic study involving paired organs; specifically, the setting of the retinopathy study in Chapter 1 is adopted to fix ideas. Section 2.2 explores the properties of ECCM, including its marginal distributions. Section 2.3 adopts the method of moments estimation for the model. These moments estimates are then used in Section 2.4 to develop new binocular versions of sensitivity and specificity, two measures of diagnostic accuracy most commonly used in applications. Finally, a small-scale empirical comparison of the new and conventional measures is reported in Section 2.6.

2.2 Model development

Define Y_{i1L} and Y_{i1R} as 1 (i.e., positive), if the reader indicates the presence of the disease in the left and right eyes, respectively, of patient $i = 1, \cdots, N$, and as 0 (i.e., negative), otherwise. Furthermore, let Y_{i2L} and Y_{i2R} denote the true disease status (equal to 1, if positive, and 0, if negative) of the left and right eyes, respectively, of patient i, as determined by the gold standard. We assume a cross-sectional design, where both diagnosis and status are determined for a patient after enrolment. Random effects are used to flexibly model the joint distribution of Y_{i1L}, Y_{i1R}, Y_{i2L}, and Y_{i2R}, and to capture the key features of the correlations between them. For $j = 1,2$, let the random effect $P_j = P(Y_{ijL} = 1|P_j) = P(Y_{ijR} = 1|P_j)$ be the common conditional probability of a positive result, and let the conditional distribution of $(Y_{ijL}, Y_{ijR})^\top$, given the random effects P_j, be a CCM with intra-pair correlation κ_j. Assuming $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$ are conditionally independent, given the random effects, the unconditional joint probability $P_{\ell_1 r_1 \ell_2 r_2} = P(Y_{i1L} = \ell_1, Y_{i1R} = r_1, Y_{i2L} = \ell_2, Y_{i2R} = r_2)$ is then given by

$$\begin{aligned} P_{\ell_1 r_1 \ell_2 r_2} &= \int_0^1 \int_0^1 P(Y_{i1L} = \ell_1, Y_{i1R} = r_1|p_1) P(Y_{i2L} = \ell_2, Y_{i2R} = r_2|p_2) \\ &\quad \times f_{P_1,P_2}(p_1,p_2) dp_1 dp_2, \end{aligned} \tag{2.1}$$

where $f_{P_1,P_2}(p_1,p_2) = f_{P_1}(p_1) f_{P_2}(p_2)\{1 + (p_1 - \pi_1)(p_2 - \pi_2)/\sqrt{var(P_1)var(P_2)}\}$ is the joint density of P_1 and P_2 in canonical form [29], $f_{P_j}(\cdot)$ is the density of P_j, $\pi_1 = P(Y_{i1L} = 1) = P(Y_{i1R} = 1)$ is the probability of a positive diagnosis by the reader, and $\pi_2 = P(Y_{i2L} = 1) = P(Y_{i2R} = 1)$ is the prevalence of the disease. Taking $P_j \sim \text{beta}(\alpha_j = (1-\rho)\pi_j/\rho, \beta_j = (1-\rho)(1-\pi_j)/\rho)$, with ρ as a "correlation" parameter, it is shown in [17] that (2.1) reduces to 9 distinct probabilities, which can be written in terms of κ_1, κ_2, and the non-central product moments

$$\psi(m_1,m_2) = E(P_1^{m_1} P_2^{m_2}) = E(P_1^{m_1}) E(P_2^{m_2}) \left\{ 1 + \frac{(\gamma_1^{(m_1)} - \pi_1)(\gamma_2^{(m_2)} - \pi_2)}{\rho\tau} \right\}, \tag{2.2}$$

where $\tau^2 = \pi_1\pi_2(1-\pi_1)(1-\pi_2)$,

$$E(P_j^{m_j}) = \frac{\Gamma(\alpha_j+m_j)\Gamma(\alpha_j+\beta_j)}{\Gamma(\alpha_j)\Gamma(\alpha_j+\beta_j+m_j)} \quad \text{and} \quad \gamma_j^{(m_j)} = \frac{(1-\rho)\pi_j+\rho m_j}{(1-\rho)\pi_j+(1-\rho)(1-\pi_j)+\rho m_j},$$

with $\Gamma(\cdot)$ as the gamma function. These 9 distinct probabilities are given as follows:

$$\begin{aligned}
P_{1111} &= (1-\kappa_1)(1-\kappa_2)\psi(2,2)+(1-\kappa_1)\kappa_2\psi(2,1)+\kappa_1(1-\kappa_2)\psi(1,2)+\kappa_1\kappa_2\psi(1,1)\\
P_{1110} &= P_{1101} = \kappa_1(1-\kappa_2)\{\psi(1,1)-\psi(1,2)\}+(1-\kappa_1)(1-\kappa_2)\psi(2,1)\\
&\quad -(1-\kappa_1)(1-\kappa_2)\psi(2,2)\\
P_{1100} &= (1-\kappa_1)\psi(2,0)+\kappa_1\psi(1,0)-(1-\kappa_1)(2-\kappa_2)\psi(2,1)\\
&\quad +(1-\kappa_1)(1-\kappa_2)\psi(2,2)-\kappa_1(2-\kappa_2)\psi(1,1)+\kappa_1(1-\kappa_2)\psi(1,2)\\
P_{1011} &= P_{0111} = (1-\kappa_1)(1-\kappa_2)\{\psi(1,2)-\psi(2,2)\}+(1-\kappa_1)\kappa_2\{\psi(1,1)-\psi(2,1)\}\\
P_{1010} &= P_{1001} = P_{0110} = P_{0101} = (1-\kappa_1)(1-\kappa_2)\{\psi(1,1)-\psi(1,2)\\
&\quad -\psi(2,1)+\psi(2,2)\}\\
P_{1000} &= P_{0100} = (1-\kappa_1)\psi(1,0)-(1-\kappa_1)(2-\kappa_2)\psi(1,1)+(1-\kappa_1)(1-\kappa_2)\psi(1,2)\\
&\quad -(1-\kappa_1)\psi(2,0)+(1-\kappa_1)(2-\kappa_2)\psi(2,1)-(1-\kappa_1)(1-\kappa_2)\psi(2,2)\\
P_{0011} &= (1-\kappa_2)\psi(0,2)-(2-\kappa_1)(1-\kappa_2)\psi(1,2)+(1-\kappa_1)(1-\kappa_2)\psi(2,2)+\kappa_2\psi(0,1)\\
&\quad -(2-\kappa_1)\kappa_2\psi(1,1)+(1-\kappa_1)\kappa_2\psi(2,1)\\
P_{0010} &= P_{0001} = (1-\kappa_2)\{\psi(0,1)-\psi(0,2)\}-(2-\kappa_1)(1-\kappa_2)\psi(1,1)\\
&\quad +(2-\kappa_1)(1-\kappa_2)\psi(1,2)+(1-\kappa_1)(1-\kappa_2)\psi(2,1)-(1-\kappa_1)(1-\kappa_2)\psi(2,2)\\
P_{0000} &= 1-\sum_{(\ell_1,r_1,\ell_2,r_2)\neq(0,0,0,0)} P_{\ell_1 r_1 \ell_2 r_2}.
\end{aligned}$$

A potential drawback of the ECCM is the exchangeability assumption $P_{1010} = P_{0101} = P_{1001} = P_{0110}$, which suggests that perfect agreement between diagnosis and disease status for exactly one eye is equally likely as perfect disagreement. It is quite possible that this assumption may not hold in cases concerning chronic diseases. Note that this equivalence arose from the assumption of exchangeability of fellow eyes, and as

such, can possibly be remedied by adopting a richer family of densities for P_1 and P_2 in model (2.1); generalized versions of the beta distribution [30] involving additional parameters for added flexibility are possible choices.

2.3 Model properties

We list the properties of the ECCM below. Note that the ECCM possesses certain desirable and convenient marginalization properties, which yield the CCM as a special case.

Property 2.1. *The unconditional marginal distribution of* $(Y_{ijL}, Y_{ijR})^\top$ *is a CCM with intra-pair correlation* $\rho_j = \rho + \kappa_j(1-\rho)$ *and common probability* π_j *of a positive result.*

It is easy to show by direct marginalization [17] (or by integrating out P_j in the conditional CCM for $(Y_{ijL}, Y_{ijR})^\top$ given P_j) that the unconditional marginal probabilities $P(Y_{ijL} = \ell_j, Y_{ijR} = r_j) = \sum_{\ell_{j'}, r_{j'}} P_{\ell_1 r_1 \ell_2 r_2}$ are given by

$$\begin{aligned}
P(Y_{ijL}=1, Y_{ijR}=1) &= \pi_j^2 + \rho_j \pi_j (1-\pi_j) \\
P(Y_{ijL}=1, Y_{ijR}=0) &= P(Y_{ijL}=0, Y_{ijR}=1) = (1-\rho_j)\pi_j(1-\pi_j) \\
P(Y_{ijL}=0, Y_{ijR}=0) &= (1-\pi_j)^2 + \rho_j \pi_j (1-\pi_j).
\end{aligned}$$

Note that the above is a CCM with intra-pair correlation ρ_j and common success probability π_j. In this sense, model (2.1) can be viewed as an extension of CCM to paired binocular binary data. Note as well that the ECCM implicitly assumes exchangeability of Y_{ijL} and Y_{ijR}.

Property 2.2. *The conditional intra-pair correlation* κ_j *between* Y_{ijL} *and* Y_{ijR}*, given the random effect* P_j*, can be expressed as* $\kappa_j = (\rho_j - \rho)/(1-\rho)$*, so that* κ_1 *can be interpreted as a Cohen's kappa-like [31, 32] measure of agreement between the reader's left- and*

right-eye diagnoses, and κ_2 can be represented as the corresponding agreement between the disease status of fellow eyes.

Property 2.2 follows directly from Property 2.1, since $\rho_j = \rho + \kappa_j(1-\rho)$. Note that $\rho_1 = \rho$ if and only if $\kappa_1 = 0$, which indicates the conditional independence between the reader's diagnoses, given P_1. The unconditional correlation ρ_1 between the diagnoses can then be mainly attributed to the parameter ρ. Similarly, $\rho_2 = \rho$ if and only if $\kappa_2 = 0$, which implies the conditional independence of disease status of fellow eyes, given P_2, and suggests that the main source of association is ρ [15]. Indeed, as we show in Property 2.3 below, the parameter ρ is in fact a correlation.

Property 2.3. *The parameter ρ is the correlation between the number $Y_{i1\cdot} = Y_{i1L} + Y_{i1R}$ of diagnosis-positive eyes and the number $Y_{i2\cdot} = Y_{i2L} + Y_{i2R}$ of status-positive eyes.*

Property 2.3 easily follows from model (2.1) and Property 2.1. The correlation ρ thus provides a global measure of association between the diagnoses and the disease status of fellow eyes. To ensure that the joint probabilities $P_{\ell_1 r_1 \ell_2 r_2}$ determined by model (2.1) define a proper probability distribution (i.e., the probabilities are non-negative and sum to unity), the correlation ρ must satisfy the following inequality

$$\frac{1}{\kappa_1\kappa_2} max\left\{-\frac{1}{\pi_1\pi_2}, -\frac{1}{(1-\pi_1)(1-\pi_2)}\right\} \leq \rho \leq \frac{1}{\kappa_1\kappa_2} min\left\{-\frac{1}{\pi_1(1-\pi_2)}, -\frac{1}{(1-\pi_1)\pi_2}\right\};$$

see [17] for a proof.

An advantage of the ECCM in (2.1) is that the joint distribution of the binocular data $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$ is completely determined by the marginal densities of P_1 and P_2. Beta distributions are very versatile so that by choosing beta marginal densities for P_1 and P_2, a variety of uncertainties regarding their distributions (hence, in Y_{ijL} and Y_{ijR}, $j = 1,2$) can be usefully modelled [33].

Table 2.1: Joint probability distribution of $Y_{i1\cdot} = Y_{i1L} + Y_{i1R}$ and $Y_{i2\cdot} = Y_{i2L} + Y_{i2R}$.

y_1	y_2			$P(Y_{i1\cdot} = y_1)$
	0	1	2	
0	p_{00}	p_{01}	p_{02}	$p_{0\cdot}$
1	p_{10}	p_{11}	p_{12}	$p_{1\cdot}$
2	p_{20}	p_{21}	p_{22}	$p_{2\cdot}$
$P(Y_{i2\cdot} = y_2)$	$p_{\cdot 0}$	$p_{\cdot 1}$	$p_{\cdot 2}$	1

2.4 Moments estimation

Suppose $n_{y_1y_2}$ is the number of patients having $y_1 = 0,1,2$ eyes diagnosed as positive by the reader and $y_2 = 0,1,2$ eyes with positive status, so that $\sum_{y_1=0}^{2}\sum_{y_2=0}^{2} n_{y_1y_2} = N$. Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_5)^\top = (\pi_1, \pi_2, \rho_1, \rho_2, \rho)^\top$, and putting $p_{y_1y_2} = P(Y_{i1\cdot} = y_1, Y_{i2\cdot} = y_2)$ with $p_{y_1\cdot} = \sum_{y_2=0}^{2} p_{y_1y_2}$ and $p_{\cdot y_2} = \sum_{y_1=0}^{2} p_{y_1y_2}$, then $p_{y_1y_2}$ in Table 2.1 can be represented by the following: $p_{00} = P_{0000}$, $p_{10} = P_{1000} + P_{0100}$, $p_{20} = P_{1100}$, $p_{01} = P_{0001} + P_{0010}$, $p_{11} = P_{1010} + P_{1001} + P_{0101} + P_{0110}$, $p_{21} = P_{1110} + P_{1101}$, $p_{02} = P_{0011}$, $p_{12} = P_{0111} + P_{1011}$, and $p_{22} = P_{1111}$. Then, their moments estimates are given by $\widehat{p}_{00} = n_{00}/N, \cdots, \widehat{p}_{22} = n_{22}/N$.

To obtain the moments estimate of $\boldsymbol{\theta}$, we need to express the parameters in $\boldsymbol{\theta}$ in terms of the probabilities $p_{y_1y_2}$ so that we can apply the plug-in principle. Under the assumption of exchangeability of fellow eyes, the probability $\theta_1 = \pi_1 = P(Y_{i1L} = 1) = P(Y_{i1R} = 1) = P_{1\cdots} = P_{\cdot 1\cdot\cdot}$ of positive diagnosis can be calculated as follows:

$$\theta_1 \;=\; \frac{1}{2}\left\{p_{10} + p_{11} + p_{12} + 2(p_{20} + p_{21} + p_{22})\right\}.$$

Similarly, we have $\theta_2 = \pi_2 = \{p_{01} + p_{11} + p_{21} + 2(p_{02} + p_{12} + p_{22})\}/2$, where $\pi_2 = P(Y_{i2L} = 1) = P(Y_{i2R} = 1) = P_{\cdot\cdot 1\cdot} = P_{\cdots 1}$ is the prevalence of disease.

In addition, the correlations $\theta_3 = \rho_1 = corr(Y_{i1L}, Y_{i1R})$ and $\theta_4 = \rho_2 = corr(Y_{i2L}, Y_{i2R})$ between left- and right-eye diagnoses and between left- and right-eye status, respec-

tively, can be calculated as follows:

$$\begin{aligned}\theta_3 &= \frac{p_{20}+p_{21}+p_{22}-\theta_1^2}{\theta_1(1-\theta_1)} = 1-\frac{p_{10}+p_{11}+p_{12}}{2\theta_1(1-\theta_1)}\\ \theta_4 &= 1-\frac{p_{01}+p_{11}+p_{21}}{2\theta_2(1-\theta_2)}.\end{aligned}$$

Finally, noting that $E(Y_{ij.}) = 2\pi_j$ and $var(Y_{ij.}) = 2\pi_j(1-\pi_j)(1+\rho_j)$, $j = 1,2$, the correlation $\theta_5 = \rho = corr(Y_{i1.}, Y_{i2.})$ can be expressed as

$$\theta_5 = \frac{E(Y_{i1.}Y_{i2.}) - \theta_1\theta_2}{2\sqrt{\theta_1\theta_2(1-\theta_1)(1-\theta_2)(1+\theta_3)(1+\theta_4)}},$$

where $E(Y_{i1.}Y_{i2.}) = p_{11}+2p_{12}+2p_{21}+4p_{22}$. From the above derivations, it is clear that $\boldsymbol{\theta}$ is a function of $\mathbf{p} = (p_{00},\cdots,p_{22})^\top$, and the moments estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is easily obtained by plug-in method using $\widehat{\mathbf{p}} = (\widehat{p}_{00},\cdots,\widehat{p}_{22})^\top$. With $n_{y_1.} = n_{y_10}+n_{y_11}+n_{y_12}$ and $n_{.y_2} = n_{0y_2}+n_{1y_2}+n_{2y_2}$, the moment estimates of $\theta_1,\cdots,\theta_5$ are

$$\begin{aligned}\widehat{\theta}_1 &= \widehat{\pi}_1 = \frac{1}{2N}(n_{1.}+2n_{2.})\\ \widehat{\theta}_2 &= \widehat{\pi}_2 = \frac{1}{2N}(n_{.1}+2n_{.2})\\ \widehat{\theta}_3 &= \widehat{\rho}_1 = 1-\frac{n_{1.}}{2N\widehat{\theta}_1(1-\widehat{\theta}_1)}\\ \widehat{\theta}_4 &= \widehat{\rho}_2 = 1-\frac{n_{.1}}{2N\widehat{\theta}_2(1-\widehat{\theta}_2)}\\ \widehat{\theta}_5 &= \widehat{\rho} = \frac{n_{11}+2n_{12}+2n_{21}+4n_{22}-4N\widehat{\theta}_1\widehat{\theta}_2}{2N\widehat{\tau}\sqrt{(1+\widehat{\theta}_3)(1+\widehat{\theta}_4)}},\end{aligned}$$

where $\widehat{\tau}^2 = \widehat{\theta}_1\widehat{\theta}_2(1-\widehat{\theta}_1)(1-\widehat{\theta}_2)$. Estimates of κ_1 and κ_2 can likewise be written as $\widehat{\kappa}_1 = (\widehat{\theta}_3-\widehat{\theta}_5)/(1-\widehat{\theta}_5)$ and $\widehat{\kappa}_2 = (\widehat{\theta}_4-\widehat{\theta}_5)/(1-\widehat{\theta}_5)$.

From standard asymptotic theory, it follows that $\widehat{\boldsymbol{\theta}}$ has an asymptotic multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix

$$\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}} = \frac{1}{N}\left(\frac{\partial\boldsymbol{\theta}}{\partial\mathbf{p}}\right)(\mathbf{D}-\mathbf{p}\mathbf{p}^\top)\left(\frac{\partial\boldsymbol{\theta}}{\partial\mathbf{p}}\right)^\top,$$

where $\mathbf{D} = diag(\mathbf{p})$. Alternatively, the bootstrap method can be used to obtain standard errors for the estimates [15]. Corresponding estimates $\widehat{P}_{\ell_1 r_1 \ell_2 r_2}$ of the joint probabilities $P_{\ell_1 r_1 \ell_2 r_2}$ are similarly obtained by plug-in method [15].

2.5 Measures of screening accuracy

The accuracy of a medical test for diagnosing the presence or absence of a disease can be described by several measures, the most common of which are given by the test's sensitivity and specificity with respect to the true disease status as determined by a traditionally used and accepted test regarded as a 'gold standard'. In practice, ophthalmologists are generally interested in coming up with measures of accuracy independent of the particular eye under consideration. This implies that $\text{Sen}_L = \text{Sen}_R$ and $\text{Sp}_L = \text{Sp}_R$, i.e., the sensitivity and specificity of the diagnostic test is independent of the left eye or the right eye. This is true under the the so-called exchangeability assumption for fellow eyes, which holds for the ECCM. Given that $\text{Sen} = P(Y_{i1L} = 1|Y_{i2L} = 1) = P(Y_{i1R} = 1|Y_{i2R} = 1)$ and $\text{Sp} = P(Y_{i1L} = 0|Y_{i2L} = 0) = P(Y_{i1R} = 0|Y_{i2R} = 0)$, we get the following:

$$\begin{aligned}
\text{Sen} &= \frac{\sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1} P_{\ell_1 1 \ell_2 1}}{\sum_{\ell_1=0}^{1}\sum_{r_1=0}^{1}\sum_{\ell_2=0}^{1} P_{\ell_1 r_1 \ell_2 1}} = \frac{\sum_{r_1=0}^{1}\sum_{r_2=0}^{1} P_{1 r_1 1 r_2}}{\sum_{\ell_1=0}^{1}\sum_{r_1=0}^{1}\sum_{r_2=0}^{1} P_{\ell_1 r_1 1 r_2}} \\
&= \frac{P_{\cdot 1 \cdot 1}}{P_{\cdots 1}} = \frac{P_{1 \cdot 1 \cdot}}{P_{\cdot\cdot 1 \cdot}} \qquad (2.3) \\
\text{Sp} &= \frac{\sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1} P_{\ell_1 0 \ell_2 0}}{\sum_{\ell_1=0}^{1}\sum_{r_1=0}^{1}\sum_{\ell_2=0}^{1} P_{\ell_1 r_1 \ell_2 0}} = \frac{\sum_{r_1=0}^{1}\sum_{r_2=0}^{1} P_{0 r_1 0 r_2}}{\sum_{\ell_1=0}^{1}\sum_{r_1=0}^{1}\sum_{r_2=0}^{1} P_{\ell_1 r_1 0 r_2}} \\
&= \frac{P_{\cdot 0 \cdot 0}}{P_{\cdots 0}} = \frac{P_{0 \cdot 0 \cdot}}{P_{\cdot\cdot 0 \cdot}}. \qquad (2.4)
\end{aligned}$$

A test's sensitivity Sen and specificity Sp are the two most common measures of its diagnostic accuracy. For diagnostic studies with exchangeable paired organs, it may be sufficient to consider only these two measures to gauge a diagnostic test's accuracy. However, when screening for diseases in paired organs, a positive result of a screening test in practice is usually followed, not directly by treatment, but with further, more definitive diagnostic procedures. Hence, in screening a particular condition in a binocular setting as in the diabetic retinopathy study, where fellow eyes are screened for retinopathy-related pathologies, it is usually enough that at least one eye is diagnosed

as positive in order that a patient is sent for further and more extensive eye examination. Conversely, it is necessary that both eyes be diagnosed as negative before the patient can be declared as negative for the disease, and thus, in no need of further examination. As such, conventional sensitivity and specificity as defined in (2.3) and (2.4) fail to capture the true screening accuracy of binocular screening tests; hence, more relevant diagnostic measures are required in these cases. We introduce new binocular measures of screening accuracy in the next section.

2.5.1 Binocular accuracy measures

For screening tests involving paired organs resulting in binocular binary diagnostic data, we define the binocular sensitivity bSen as the probability of at least one correct positive diagnosis in patients with one or both eyes truly diseased; we likewise define the binocular specificity bSp as the probability of two correct negative diagnoses for patients with both eyes truly un-diseased.

For the binocular sensitivity bSen, we can consider three distinct scenarios as follows:

$$\begin{aligned}
\mathrm{b_1Sen} &= P(Y_{i1L}=Y_{i1R}=1|Y_{i2L}=Y_{i2R}=1)+P(Y_{i1L}=0,Y_{i1R}=1|Y_{i2L}=Y_{i2R}=1) \\
&\quad +P(Y_{i1L}=1,Y_{i1R}=0|Y_{i2L}=Y_{i2R}=1) = \frac{P_{1111}+P_{0111}+P_{1011}}{P_{\cdot\cdot 11}} \\
\mathrm{b_2Sen} &= P(Y_{i1L}=1,Y_{i1R}=0|Y_{i2L}=1,Y_{i2R}=0) = \frac{P_{1010}}{P_{\cdot\cdot 10}} \\
\mathrm{b_3Sen} &= P(Y_{i1L}=0,Y_{i1R}=1|Y_{i2L}=0,Y_{i2R}=1) = \frac{P_{0101}}{P_{\cdot\cdot 01}}.
\end{aligned}$$

We can then define bSen as a weighted average of $\mathrm{b_1Sen}$, $\mathrm{b_2Sen}$, and $\mathrm{b_3Sen}$:

$$\mathrm{bSen} = w_1\mathrm{b_1Sen}+w_2\mathrm{b_2Sen}+w_3\mathrm{b_3Sen},$$

where $0 \le w_1, w_2, w_3 \le 1$ and $w_1+w_2+w_3=1$. That is, bSen is a convex linear combi-

nation of b_1Sen, b_2Sen, and b_3Sen. For the rest of the chapter, we choose to use

$$w_1 = \frac{P_{\cdot\cdot 11}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}} \quad w_2 = \frac{P_{\cdot\cdot 10}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}} \quad w_3 = \frac{P_{\cdot\cdot 01}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}},$$

that is, the weights are the marginal probabilities $P_{\cdot\cdot 11}$, $P_{\cdot\cdot 10}$, and $P_{\cdot\cdot 01}$, normalized so that they sum to 1. In this case, the binocular sensitivity bSen is given by

$$\text{bSen} = \frac{P_{1111}+P_{0111}+P_{1011}+P_{1010}+P_{0101}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}} \tag{2.5}$$

The binocular specificity is defined in a straightforward way as follows:

$$\text{bSp} = P(Y_{i1L}=Y_{i1R}=0|Y_{i2L}=Y_{i2R}=0) = \frac{P_{0000}}{P_{\cdot\cdot 00}}. \tag{2.6}$$

Corresponding estimates of the binocular accuracy measures bSen and bSp are obtained directly by plugging estimates in Section 2.4 into (2.5) and (2.6). Using standard asymptotic theory as in Section 2.4, the estimate $(\widehat{\text{bSen}},\widehat{\text{bSp}})^\top$ of $(\text{bSen},\text{bSp})^\top$ follows, for large N, an approximate bivariate normal distribution with mean $(\text{bSen},\text{bSp})^\top$ and covariance matrix obtained by the delta method from $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$.

2.5.2 Conventional vs. binocular accuracy measures

In this section, we investigate the relationships, if any, between the conventional sensitivity and specificity Sen and Sp, respectively, and their binocular counterparts bSen and bSp. As discussed in Section 2.5.1, the conventional accuracy measures may not capture the true screening accuracy of a screening test, since the diagnostic outcomes arising from the paired organs (i.e., fellow eyes) necessitate measures that incorporate the binocular nature of the diagnoses, even in the case of exchangeability. Intuitively, we expect the conventional measure to be at least as high as the binocular measure, since the former is based on the marginal distribution while the latter arises from the joint distribution. This observation is generally confirmed in what follows.

Table 2.2: Results for Sen, Sp, bSen, and bSp, for the ECCM with $\pi_1 = \pi_2 = 0.6, 0.5, 0.4$, $\rho_1 = \rho_2 = 0.8, 0.6, 0.5$, and $\rho = 0.5, 0.3$.

$\pi_1 = \pi_2$		$\rho_1 = \rho_2 = 0.8$		$\rho_1 = \rho_2 = 0.6$		$\rho_1 = \rho_2 = 0.5$	
		$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.3$
0.6	Sen	0.800	0.720	0.800	0.720	0.800	0.720
	Sp	0.700	0.580	0.700	0.580	0.700	0.580
	bSen	0.744	0.670	0.703	0.634	0.687	0.621
	bSp	0.675	0.544	0.658	0.513	0.653	0.499
0.5	Sen	0.750	0.650	0.750	0.650	0.750	0.650
	Sp	0.750	0.650	0.750	0.650	0.750	0.650
	bSen	0.686	0.595	0.642	0.558	0.625	0.545
	bSp	0.728	0.617	0.712	0.588	0.708	0.575
0.4	Sen	0.700	0.580	0.700	0.580	0.700	0.580
	Sp	0.800	0.720	0.800	0.720	0.800	0.720
	bSen	0.630	0.523	0.584	0.487	0.567	0.475
	bSp	0.781	0.691	0.768	0.665	0.764	0.654

Noting that

$$\begin{aligned} \text{Sen} &= \left(\frac{P_{1010}+P_{1011}+P_{0111}+P_{1111}+P_{0101}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}}\right)\left(\frac{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}}\right) \\ &\quad + \frac{P_{1110}-P_{0111}-P_{0101}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}}, \end{aligned}$$

so that we get the following relationship between Sen and bSen:

$$\text{bSen} = \left(\frac{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}}\right)\text{Sen} - \frac{P_{1110}-P_{0111}-P_{0101}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}}.$$

Since $\text{Sen} \geq (P_{\cdot\cdot 11}+P_{\cdot\cdot 10})\text{Sen}/(P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01})$, it follows that

$$\text{bSen} < \text{Sen} - \frac{P_{1110}-(P_{0111}+P_{0101})}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}+P_{\cdot\cdot 01}}. \tag{2.7}$$

From (2.7), it is clear that we have bSen = Sen if and only if $P_{\cdot\cdot 01} = 0$ and

$$\begin{aligned} P_{1110} &= P(Y_{i1L} = Y_{i1R} = Y_{i2L} = 1, Y_{i2R} = 0) \\ &= P_{0111} + P_{0101} \\ &= P(Y_{i1L} = 0, Y_{i1R} = Y_{i2L} = Y_{i2R} = 1) \\ &\quad + P(Y_{i1L} = 0, Y_{i1R} = 1, Y_{i2L} = 0, Y_{i2R} = 1). \end{aligned}$$

Using the expressions for $P_{\ell_1 r_1 \ell_2 r_2}$ given in Section 2.2, the condition $P_{1110} = P_{0111} + P_{0101}$ is equivalent to the following for the ECCM:

$$\begin{aligned}(1-\kappa_1)(1-\kappa_2)\{\psi(1,1)-\psi(2,1)\} &= (1-\kappa_2)\{\psi(2,1)-\psi(2,2)\} \\ &\quad -\kappa_1\{\psi(1,2)-\psi(2,2)\}.\end{aligned}$$

Therefore, if $\kappa_2 = 1$ and $\kappa_1 = 0$, then $P_{1110} = P_{0111} + P_{0101}$. In addition, by Property 2.1, $P_{\cdot\cdot 01} = P(Y_{i2L} = 0, Y_{i2R} = 1) = (1-\rho_2)\pi_2(1-\pi_2)$ so that $P_{\cdot\cdot 01} = 0$ if and only if $\rho_2 = 1$ if and only if $\kappa_2 = 1$. Hence, a sufficient (but not necessary) condition for bSen = Sen is that $\kappa_2 = 1$ and $\kappa_1 = 0$.

To derive the relationship between Sp and bSp, we proceed as follows. Using the patient's left eye, (2.4) gives

$$\text{Sp} - \frac{P_{0001} + P_{0100} + P_{0101}}{P_{\cdot\cdot 00} + P_{\cdot\cdot 01}} = \left(\frac{P_{\cdot\cdot 00}}{P_{\cdot\cdot 00} + P_{\cdot\cdot 01}}\right)\text{bSp}.$$

Since the joint probabilities $P_{\ell_1 r_1 \ell_2 r_2}$ determined by (2.1) define a proper probability distribution, we get the following inequality:

$$\text{Sp} > \left(\frac{P_{\cdot\cdot 00}}{P_{\cdot\cdot 00} + P_{\cdot\cdot 01}}\right)\text{bSp}. \tag{2.8}$$

Note that this holds as well for the patient's right eye, since $P_{\cdot\cdot 01} = P_{\cdot\cdot 10}$, by the assumption of exchangeability of fellow eyes under the ECCM. By Property 2.1 of the ECCM, we get

$$\text{Sp} > \{1 - \pi_2(1-\rho_2)\}\text{bSp}, \tag{2.9}$$

where $\rho_2 = corr(Y_{i2L}, Y_{i2R})$. It then follows that we have bSp = Sp if and only if $\rho_2 = 1$ (i.e., $\kappa_2 = 1$) and $P_{0001} = P_{0100} = P_{0101} = 0$. Therefore, a sufficient (but not necessary) condition for bSp = Sp is that $\kappa_1 = \kappa_2 = 1$. This makes sense because if the left- and right-eye diagnoses as well as the left- and right-eye disease status are perfectly associated, one needs to consider only one of the fellow eyes, in which case the binocular nature of the data may be ignored.

2.6 Empirical comparisons of Sen, Sp, bSen, and bSp

In this section, the conventional measures Sen and Sp and the binocular measures bSen and bSp are calculated for various ECCMs. The following parameter configurations for the ECCM are considered: $\pi_1 = \pi_2 = 0.6, 0.5, 0.4$, $\rho_1 = \rho_2 = 0.8, 0.6, 0.5$, and $\rho = 0.5, 0.3$, with number $N = 1000$ of patients. These parameter configurations are relatively common in diagnostic studies in ophthalmology, such as diabetic retinopathy study described earlier. They also define proper joint probability distributions for the binocular data.

As shown in Table 2.2, Sen and Sp, on the one hand, do not depend on ρ_1 and ρ_2, since they are marginal measures and are based only on one eye. This is possible because of exchangeability, which holds for the ECCM. However, for screening purposes, it is not enough to consider only one eye; as pointed out earlier, it is sufficient for at least one diseased eye to be diagnosed correctly to have a patient declared 'positive' for further testing, and for both negative eyes to be correctly diagnosed before a patient can be declared 'negative', and therefore in no need of further examination. By ignoring the correlation between fellow eyes and using only their marginal distributions to define the accuracy measures, it is not possible to jointly evaluate the diagnoses for the fellow eyes.

The binocular measures bSen and bSp, on the other hand, depend on all the parameters. As expected from the conditions obtained in Section 2.5.2, we have Sen $>$ bSen and Sp $>$ bSp. In addition, it appears that the difference between the conventional and binocular measures becomes large as $\rho_1 = \rho_2$ get small; the same observation holds as ρ decreases. Note that these results are implied by the relationships between the two sets of measures derived in Section 2.5.2.

Lest we think that Sen $>$ bSen and Sp $>$ bSp always hold, note that this is not

always the case. In fact, it is quite easy to construct an ECCM for which Sen $<$ bSen, say. This is true, for example, for the ECCM with $\pi_{i1} = 0.7$, $\pi_{i2} = 0.6$, $\rho_1 = \rho = 0.6$, and $\rho_2 = 0.9$.

Chapter 3

A Gaussian copula model for paired binocular binary data

3.1 Introduction

Copulas have attracted a lot of interest among researchers recently [18,34], due mainly to the flexibility they provide in analysing a wide array of correlated data settings. In practice, statisticians often know very little about the joint behaviour of outcomes but can specify their marginal behaviours reasonably well, as is the case with correlated binary data. Copulas provide a means of assembling a joint distribution when only its margins are known. The approach embeds univariate marginal cumulative distribution functions (CDFs) $F_{Y_1}(\cdot), \cdots, F_{Y_p}(\cdot)$ of random variables (RVs) $Y_1, \cdots, Y_p$, into their corresponding p-dimensional CDF $F_{Y_1,\cdots,Y_p}(\cdot)$ via a copula $C(\cdot)$ as follows:

$$F_{Y_1,\cdots,Y_p}(y_1, \cdots, y_p) \quad = \quad C(u_1, \cdots, u_p),$$

where $u_k = F_{Y_k}(y_k)$ is the observed value of the so-called probability integral transform (PIT) $U_k = F_{Y_k}(Y_k)$, $k = 1, \cdots, p$. The meta distribution $F_{Y_1,\cdots,Y_p}(\cdot)$ is thus specified via its margins and a copula that "glues" them together; in parametric contexts, the margins $F_{Y_k}(\cdot)$ need not come from the same parametric families, allowing researchers great flexibility in modelling different non-Gaussian data. However, such distribution is unique only if the RVs involved are continuous. The same is not true for discrete (e.g., binary) RVs, in which case the copula is only uniquely identified on the Cartesian product of the ranges of the margins [24]; moreover, the uniformity of the PITs, a condition for Sklar's Theorem [18], does not hold in the discrete case. In fact, the class of possible copulas can be quite large, especially in the binary case. Note, however, that from a modeling perspective, the non-uniqueness of the copula is not an issue, as

the parameters of the model are still identifiable and the copula still corresponds to a proper multivariate distribution. However, common rank-based association measures like Kendall's tau, while margin-free in the purely continuous case, may now depend on the margins, and the range of their possible values may be restricted — severely in some cases — rendering interpretations of such measures problematic.

The chapter proposes a Gaussian copula approach to model the paired binocular binary diagnostic data, like those from the diabetic retinopathy study. For models based on the Gaussian copula, correlations between outcomes are not modeled directly; instead, correlations are incorporated for the normal scores $\Phi^{-1}\{F_{Y_k}(Y_k)\} = \Phi^{-1}(U_k)$ and $\Phi^{-1}\{F_{Y_{k'}}(Y_{k'})\} = \Phi^{-1}(U_{k'})$, which are transformations of the original variables, where $\Phi^{-1}(\cdot)$ is the standard normal quantile function (i.e., the inverse of $\Phi(\cdot)$, the standard normal CDF). Recent work in [35, 36] enable practitioners to recover an estimate of $corr(Y_k, Y_{k'})$ given an estimate of the normal correlation $corr\{\Phi^{-1}(U_k), \Phi^{-1}(U_{k'})\}$.

The chapter is organized as follows. Section 3.2 briefly reviews copula models and describes the development specifically of the Gaussian copula model. Section 3.3 adopts the Gaussian copula to directly construct a joint distribution for the paired binocular binary data $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$ discussed in Chapter 2. Properties of and estimation via maximum likelihood for the model are discussed. Adaptations of the binocular measures of screening accuracy introduced in Chapter 2 are also presented. Finally, Section 3.6 discusses the results of simulations on the relative bias, relative efficiency, and asymptotic normality of the MLEs for the model parameters.

3.2 Brief review of copula

A number of recent work on correlated data analysis in biostatistical applications have adopted copula functions to indirectly specify the associated joint distributions. This

is only a recent phenomenon in modelling correlated outcomes in health and medicine, and unresolved issues, both methodological and practical, abound, especially as they apply to discrete data.

The basic idea in copula modelling is to assemble a joint distribution from uniform margins via a copula function. Given a continuous RV Y, and its CDF $F_Y(\cdot)$, one can easily see that the PIT $U = F_Y(Y)$ follows a uniform distribution on $[0,1]$. Thus, modelling the joint distribution of RVs Y_i, $i = 1, \cdots, N$, by using their marginal CDFs $F_{Y_i}(\cdot)$, is equivalent to modelling the joint distribution using the marginal uniform RVs $U_i = F_{Y_i}(Y_i)$ [37].

Consider the 2-dimensional case of (possibly dependent) continuous RVs Y_1 and Y_2, with corresponding CDFs $F_{Y_1}(\cdot)$ and $F_{Y_2}(\cdot)$. Let $F_{Y_1,Y_2}(\cdot)$ be the joint CDF of Y_1 and Y_2, so that

$$F_{Y_1,Y_2}(y_1,y_2) \quad = \quad P(Y_1 \leq y_1, Y_2 \leq y_2).$$

Let $U_1 = F_{Y_1}(Y_1)$ and $U_2 = F_{Y_2}(Y_2)$, where U_1 and U_2 are uniform RVs on $[0,1]$. Note that the association between U_1 and U_2 is described by their joint CDF $C(u_1,u_2) = P(U_1 \leq u_1, U_2 \leq u_2)$. It follows that the joint CDF is given by the following:

$$F_{Y_1,Y_2}(y_1,y_2) \quad = \quad P\{F_{Y_1}(Y_1) \leq F_{Y_1}(y_1), F_{Y_2}(Y_2) \leq F_{Y_2}(y_2)\} \quad = \quad C(u_1,u_2), \qquad (3.1)$$

where $u_1 = F_{Y_1}(y_1)$ and $u_2 = F_{Y_2}(y_2)$. Thus, the joint CDF $F_{Y_1,Y_2}(\cdot)$ of Y_1 and Y_2 is equivalent to the joint CDF $C(\cdot)$ of marginally uniform RVs $U_1 = F_{Y_1}(Y_1)$ and $U_2 = F_{Y_2}(Y_2)$; thus a 2-dimensional copula $C(\cdot) : [0,1]^2 \to [0,1]$ can be defined with the following properties:

(i) for every $u_1, u_2 \in [0,1]$, $C(u_1,0) = C(0,u_2) = 0$, $C(u_1,1) = u_1$ and $C(1,u_2) = u_2$;

(ii) for every $u_1, u_1', u_2, u_2' \in [0,1]$ such that $u_1 \leq u_1'$ and $u_2 \leq u_2'$, $C(u_1',u_2') - C(u_1',u_2) - C(u_1,u_2') + C(u_1,u_2) \geq 0$.

By uniformity of PITs, $Y_1 = F_{Y_1}^{-1}(U_1)$ and $Y_2 = F_{Y_2}^{-1}(U_2)$, so that $C(u_1,u_2) = P\{F_{Y_1}^{-1}(U_1) \leq F_{Y_1}^{-1}(u_1), F_{Y_2}^{-1}(U_2) \leq F_{Y_2}^{-1}(u_2)\} = F_{Y_1,Y_2}(y_1,y_2)$. Hence, to a specified joint distribution with continuous margins, a unique function $C(\cdot)$ is associated. Copulas thus provide a strategy of specifying a joint distribution when only margins are known or easily constructed.

3.2.1 Gaussian copula

The choice of copula reflects the margins' dependence structure, which is captured by a dependence parameter embedded in the copula. A number of copula families have been studied in the literature, each imposing a variety of ways of modelling the dependence between margins [38]. One such family is the Gaussian family [39]. It has been widely used in applications because of its convenient marginal and conditional properties.

In the bivariate case, with $\Phi_{\widetilde{\rho}}(\cdot)$ as the standard bivariate normal CDF, with zero means and unit variances, and correlation coefficient $\widetilde{\rho}$, the Gaussian copula function is given by

$$\begin{aligned} \Phi_{\widetilde{\rho}}\{\Phi^{-1}(u_1),\Phi^{-1}(u_2)\} &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\widetilde{\rho}^2}} \\ &\quad \times \exp\left\{-\frac{s^2+t^2-2\widetilde{\rho}st}{2(1-\widetilde{\rho}^2)}\right\} dsdt. \end{aligned} \tag{3.2}$$

The parameter $\widetilde{\rho}$ is not the correlation between Y_1 and Y_2; instead, it is the normal correlation given by $\widetilde{\rho} = corr\{\Phi^{-1}(U_1),\Phi^{-1}(U_2)\}$, i.e., the correlation between the normal scores $\Phi^{-1}(U_1)$ and $\Phi^{-1}(U_2)$, and it determines the degree of dependence between Y_1 and Y_2. With continuous margins $Y_1 \sim F_{Y_1}(\cdot)$ and $Y_2 \sim F_{Y_2}(\cdot)$, and with PITs $U_1 = F_{Y_1}(Y_1)$ and $U_2 = F_{Y_2}(Y_2)$ whose joint CDF is given by (3.2), the joint CDF $F_{Y_1,Y_2}(\cdot)$ of Y_1 and Y_2 is

$$F_{Y_1,Y_2}(y_1,y_2) = \Phi_{\widetilde{\rho}}(\Phi^{-1}\{F_{Y_1}(y_1)\},\Phi^{-1}\{F_{Y_2}(y_2)\}). \tag{3.3}$$

If Y_1 and Y_2 are both normal RVs, then the joint distribution of Y_1 and Y_2 is exactly standard bivariate normal with correlation $\widetilde{\rho}$. One of the advantages of the Gaussian copula is that it can be easily extended to higher dimensions to construct the multivariate Gaussian copula

$$F_{Y_1,\cdots,Y_p}(y_1,\cdots,y_p) \quad = \quad \Phi_{\widetilde{\mathbf{R}}}(\Phi^{-1}\{F_{Y_1}(y_1)\},\cdots,\Phi^{-1}\{F_{Y_p}(y_p)\}),$$

where $\Phi_{\widetilde{\mathbf{R}}}(\cdot)$ is the standard multivariate normal CDF with correlation matrix $\widetilde{\mathbf{R}}$, which may be flexibly parametrically specified according to the data's dependence structure [39].

In the case of multivariate binary-$\{0,1\}$ RVs $Y_k \sim F_{Y_k}(\cdot)$, $i = k,\cdots,p$, the multivariate Gaussian copula $\Phi_{\widetilde{\mathbf{R}}}(\cdot)$ is uniquely determined only in the product range $\mathbb{R}_{F_{Y_1}} \times \cdots \times \mathbb{R}_{F_{Y_p}}$, where $\mathbb{R}_{F_{Y_k}}$ is the range of $F_{Y_k}(\cdot)$. The joint probability distribution of $Y_1,\cdots,Y_k$ is then

$$P(Y_1 = y_1,\cdots,Y_p = y_p) = \sum_{j_1=0}^{1} \cdots \sum_{j_p=0}^{1} (-1)^{p+j_1+\cdots+j_p} \Phi_{\widetilde{\mathbf{R}}}\{\Phi^{-1}(u_{1j_1}),\cdots\Phi^{-1}(u_{pj_p})\}, \quad (3.4)$$

where $u_{k0} = F_{Y_k}(y_k)$ and $u_{k1} = F_{Y_k}(y_k^-) = \lim_{y \uparrow y_k} F_{Y_k}(y)$ [40]. Note that even though (3.4) is unique only in $\mathbb{R}_{F_{Y_1}} \times \cdots \times \mathbb{R}_{F_{Y_p}}$, it is still a proper multivariate distribution in that $P(Y_1 = y_1,\cdots,Y_p = y_p) \geq 0$ and $\sum_{y_1,\cdots,y_p} P(Y_1 = y_1,\cdots,Y_p = y_p) = 1$. However, as in many multivariate discrete distributions, the condition $\sum_{y_1,\cdots,y_p} P(Y_1 = y_1,\cdots,Y_p = y_p) = 1$ restricts the correlations $\widetilde{\rho}_{kk'}$ in $\widetilde{\mathbf{R}}$ to lie in possibly narrow ranges $[h^u_{kk'}, h^\ell_{kk'}]$, with $h^u_{kk'} \gg -1$ and $h^\ell_{kk'} \ll 1$, in addition to the conventional assumption that they be such that $\widetilde{\mathbf{R}}$ is positive definite.

3.3 Model for paired binocular binary outcomes

Let $\pi_{1j} = P(Y_{i1j} = 1)$ be the probability of a positive diagnosis for eye $j = L,R$, of patient i, and let $\pi_{2j} = P(Y_{i2j} = 1)$ be the corresponding probability that the patient's eye j is

positive. Note that unlike in the ECCM, we do not assume exchangeability of fellow eyes, unless $\pi_{1L} = \pi_{1R}$ and $\pi_{2L} = \pi_{2R}$. Adopting the joint probability (3.4) given by the Gaussian copula, the joint probability $P_{\ell_1 r_1 \ell_2 r_2} = P(Y_{i1L} = \ell_1, Y_{i1R} = r_1, Y_{i2L} = \ell_2, Y_{i2R} = r_2)$ is then given by

$$\begin{aligned} P_{\ell_1 r_1 \ell_2 r_2} &= \sum_{j_1=0}^{1} \cdots \sum_{j_4=0}^{1} (-1)^{4+j_1+\cdots+j_4} \Phi_{\widetilde{\mathbf{R}}} \left\{ \Phi^{-1}(u_{1Lj_1}), \Phi^{-1}(u_{1Rj_1}), \right. \\ &\quad \left. \Phi^{-1}(u_{2Lj_1}), \Phi^{-1}(u_{2Rj_4}) \right\}, \end{aligned} \tag{3.5}$$

where, for example, $u_{1L0} = P(Y_{i1L} \leq \ell_1)$ and $u_{1R1} = \lim_{\ell \uparrow \ell_1} P(Y_{i1L} \leq \ell) = P(Y_{i1L} \leq \ell_1 - 1)$, for $\ell_1 = 0, 1$; note that for $\ell_1 = 0$, $u_{1L0} = P(Y_{i1L} \leq 0) = 1 - \pi_{1L}$ and $u_{iL1} = 0$, and for $\ell_1 = 1$, we have $u_{1L0} = P(Y_{i1L} \leq 1) = 1$ and $u_{iL1} = 1 - \pi_{1L}$. In addition, we assume the following form for $\widetilde{\mathbf{R}}$:

$$\widetilde{\mathbf{R}} = \begin{pmatrix} 1 & \widetilde{\rho}_1 & \widetilde{\rho}_3 & 0 \\ & 1 & 0 & \widetilde{\rho}_4 \\ & & 1 & \widetilde{\rho}_2 \\ & & & 1 \end{pmatrix}, \tag{3.6}$$

where $\widetilde{\rho}_1 = corr(\Phi^{-1}\{F_{Y_{i1L}}(Y_{i1L})\}, \Phi^{-1}\{F_{Y_{i1R}}(Y_{i1R})\})$ is the correlation between the normal scores of reader diagnoses of fellow eyes, $\widetilde{\rho}_2 = corr(\Phi^{-1}\{F_{Y_{i2L}}(Y_{i2L})\}, \Phi^{-1}\{F_{Y_{i2R}}(Y_{i2R})\})$ is the correlation between the normal scores of the status of fellow eyes, and $\widetilde{\rho}_3 = corr(\Phi^{-1}\{F_{Y_{i1L}}(Y_{i1L})\}, \Phi^{-1}\{F_{Y_{i2L}}(Y_{i2L})\})$ and $\widetilde{\rho}_4 = corr(\Phi^{-1}\{F_{Y_{i1R}}(Y_{i1R})\}, \Phi^{-1}\{F_{Y_{i2R}}(Y_{i2R})\})$ are the correlations between the normal scores of reader diagnoses and disease status of the left and right eyes, respectively. Note that $Y_{i1j} \sim$ bernoulli(π_{1j}) and $Y_{i2j} \sim$ bernoulli(π_{2j}), $j = L, R$. Note as well that the normal correlations $\widetilde{\rho}_1$, $\widetilde{\rho}_2$, $\widetilde{\rho}_3$, and $\widetilde{\rho}_4$ are proxies for the respective Pearson's correlations $\rho_1 = corr(Y_{i1L}, Y_{i1R})$, $\rho_2 = corr(Y_{i2L}, Y_{i2R})$, $\rho_3 = corr(Y_{i1L}, Y_{i2L})$, and $\rho_4 = corr(Y_{i1R}, Y_{i2R})$, where for example, we have

$$corr(Y_{i2L}, Y_{i2R}) = \frac{cov(Y_{i1L}, Y_{i1R})}{\sqrt{var(Y_{i1L}) var(Y_{i1R})}} = \frac{P(Y_{i1L} = Y_{i1R} = 1) - \pi_{1L}\pi_{1R}}{\sqrt{\pi_{1L}\pi_{1R}(1-\pi_{1L})(1-\pi_{1R})}}. \tag{3.7}$$

A final remark concerns the marginal distributions of $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$. Since the Gaussian copula is closed under marginalization, it follows that the (marginal) distributions of $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$ are given by the bivariate Gaussian copula and are of the same form as (3.5).

Table 3.1: Comparison of true joint probabilities $P_{\ell_1 r_1 \ell_2 r_2}$ and corresponding relative frequencies for the Gaussian copula model with $\pi_{1L} = 0.53$, $\pi_{1R} = \pi_{2L} = \pi_{2R} = 0.5$, and $\widetilde{\rho}_1 = \widetilde{\rho}_2 = 0.82$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0$, for $N = 1000$.

Joint probability	True value	Relative frequency	Frequency
P_{0000}	0.1561	0.1390	139
P_{0001}	0.0376	0.0420	42
P_{0010}	0.0376	0.0410	41
P_{0100}	0.0333	0.0400	40
P_{1000}	0.0454	0.0400	40
P_{0101}	0.0080	0.0070	7
P_{0110}	0.0080	0.0060	6
P_{1001}	0.0109	0.0110	11
P_{1010}	0.0109	0.0160	16
P_{0011}	0.1561	0.1790	179
P_{1100}	0.1682	0.1510	151
P_{1110}	0.0405	0.0330	33
P_{1101}	0.0405	0.0540	54
P_{1011}	0.0454	0.0440	44
P_{0111}	0.0333	0.0390	39
P_{1111}	0.1682	0.1580	158

3.3.1 Maximum likelihood estimation

Consider N patients providing paired binocular binary diagnostic data $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$, $i = 1, \cdots, N$. With $\boldsymbol{\theta} = (\pi_{1L}, \pi_{1R}, \pi_{2L}, \pi_{2R}, \widetilde{\rho}_1, \widetilde{\rho}_2, \widetilde{\rho}_3, \widetilde{\rho}_4)^\top$, the likelihood function $L(\boldsymbol{\theta})$ is represented as

$$L(\boldsymbol{\theta}) \;=\; \prod_{\ell_1=0}^{1} \prod_{r_1=0}^{1} \prod_{\ell_2=0}^{1} \prod_{r_2=0}^{1} P_{\ell_1 r_1 \ell_2 r_2}^{n_{\ell_1 r_1 \ell_2 r_2}}, \tag{3.8}$$

where $n_{\ell_1 r_1 \ell_2 r_2}$ is the number of patients with $Y_{i1L} = \ell_1, Y_{i1R} = r_1, Y_{i2L} = \ell_2, Y_{i2R} = r_2$. The corresponding log-likelihood function $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$ is then

$$\ell(\boldsymbol{\theta}) = \sum_{\ell_1=0}^{1}\sum_{r_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{r_2=0}^{1} n_{\ell_1 r_1 \ell_2 r_2} \log P_{\ell_1 r_1 \ell_2 r_2}. \tag{3.9}$$

Putting $s(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ as the score function, the MLE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by solving $s(\boldsymbol{\theta}) = \mathbf{0}$ iteratively via a Newton-Raphson updating scheme. Using (3.5), the elements of $s(\boldsymbol{\theta})$ of the Gaussian copula model are as follows:

$$\frac{\partial}{\partial \pi_{kj}}\ell(\boldsymbol{\theta}) = \sum_{\ell_1=0}^{1}\sum_{r_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{r_2=0}^{1} \left(\frac{n_{\ell_1 r_1 \ell_2 r_2}}{P_{\ell_1 r_1 \ell_2 r_2}}\right) \frac{\partial}{\partial \pi_{kj}} P_{\ell_1 r_1 \ell_2 r_2}, \tag{3.10}$$

for $k = 1, 2$, and $j = L, R$; and

$$\frac{\partial}{\partial \widetilde{\rho}_m}\ell(\boldsymbol{\theta}) = \sum_{\ell_1=0}^{1}\sum_{r_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{r_2=0}^{1} \left(\frac{n_{\ell_1 r_1 \ell_2 r_2}}{P_{\ell_1 r_1 \ell_2 r_2}}\right) \frac{\partial}{\partial \widetilde{\rho}_m} P_{\ell_1 r_1 \ell_2 r_2}, \tag{3.11}$$

for where $m = 1, \cdots, 4$.

With the help of symbolic computing software, the Hessian $H(\boldsymbol{\theta}) = \partial^2 \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ of (3.9) can be obtained; however, such usually tedious work can be undertaken automatically by numerical algorithms embedded in many statistical software. Usual properties of MLEs still apply in this case: under standard regularity conditions, $\widehat{\boldsymbol{\theta}}$ is a consistent estimate of $\boldsymbol{\theta}$ and asymptotically follows a multivariate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of the Fisher information matrix $I_N(\boldsymbol{\theta}) = E\{-H(\boldsymbol{\theta})\} = E\{s(\boldsymbol{\theta}) s^\top(\boldsymbol{\theta})\}$. Frequently in practice, the Fisher information matrix is estimated by the observed Fisher information $s(\widehat{\boldsymbol{\theta}}) s^\top(\widehat{\boldsymbol{\theta}})$, or by $-H(\widehat{\boldsymbol{\theta}})$, in which case standard errors of estimates are obtained from the diagonals of $\{s(\widehat{\boldsymbol{\theta}}) s^\top(\widehat{\boldsymbol{\theta}})\}^{-1}$ or $-H^{-1}(\widehat{\boldsymbol{\theta}})$.

Corresponding estimates of $\text{Sen}_L = P(Y_{i1L} = 1 | Y_{i2L} = 1)$ and $\text{Sen}_R = P(Y_{i1R} = 1 | Y_{i2R} = 1)$, the respective left- and right-eye sensitivities, and of $\text{Sp}_L = P(Y_{i1L} = 0 | Y_{i2L} = 0)$ and $\text{Sp}_R = P(Y_{i1R} = 0 | Y_{i2R} = 0)$, the respective left- and right-eye specificities, are obtained

Table 3.2: Relative bias (in %), standard deviation (SD), large-sample standard error (SE), and relative efficiency of MLEs for $\boldsymbol{\theta}$ and the accuracy measures for the Gaussian copula model with $\pi_{1L} = 0.53$, $\pi_{1R} = \pi_{2L} = \pi_{2R} = 0.5$, $\widetilde{\rho}_1 = \widetilde{\rho}_2 = 0.82$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0$, for $N = 1000$.

Parameter	True value	Mean estimate	Relative bias	SD	SE	Relative efficiency
π_{1L}	0.5300	0.5303	0.0549	0.0049	0.0047	0.9708
π_{1R}	0.5000	0.5478	3.5594	0.0047	0.0048	1.0160
π_{2L}	0.5000	0.5178	3.5671	0.0056	0.0056	0.9903
π_{2R}	0.5000	0.5079	1.5834	0.0057	0.0056	0.9903
$\widetilde{\rho}_1$	0.8200	0.8389	2.2989	0.0050	0.0050	0.9984
ρ_1	0.6105	0.6331	3.7074	0.0058	0.0059	1.0251
$\widetilde{\rho}_2$	0.8200	0.8389	2.2989	0.0050	0.0051	1.0257
ρ_2	0.6120	0.6104	0.2625	0.0059	0.0060	1.0099
Sen_L	0.5300	0.5303	0.0542	0.0049	0.0048	0.9850
Sen_R	0.5000	0.5178	3.5609	0.0047	0.0046	0.9774
Sp_L	0.4670	0.4697	0.0617	0.0049	0.0048	0.9866
Sp_R	0.5000	0.4822	3.5554	0.0047	0.0047	1.0035
bSen	0.4453	0.4485	0.7157	0.0040	0.0040	0.9784
bSp	0.3874	0.3844	0.7658	0.0043	0.0044	1.0132

directly by plug-in method using the MLEs $\widehat{\pi}_{1L}$, $\widehat{\pi}_{1R}$, $\widehat{\pi}_{2L}$, $\widehat{\pi}_{2R}$, $\widehat{\widetilde{\rho}}_1$, $\widehat{\widetilde{\rho}}_2$, $\widehat{\widetilde{\rho}}_3$, and $\widehat{\widetilde{\rho}}_4$. Standard errors are then calculated from the large-sample covariance matrix obtained by the delta method as

$$cov(\boldsymbol{\eta}) \quad = \quad \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}}\right) I_N^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}}\right)^{\top}, \tag{3.12}$$

where $\boldsymbol{\eta} = (\text{Sen}_L, \text{Sen}_R, \text{Sp}_L, \text{Sp}_R)^{\top}$.

3.4 Binocular accuracy measures

Without assuming exchangeability of fellow eyes, $\text{Sen}_L \neq \text{Sen}_R$ and $\text{Sp}_L \neq \text{Sp}_R$ because the joint probabilities are such that $P_{0001} \neq P_{0010}$, $P_{0100} \neq P_{1000}$, $P_{1110} \neq P_{1101}$, $P_{1011} \neq P_{0111}$, $P_{0010} \neq P_{0001}$, and $P_{1010} \neq P_{1001} \neq P_{0101} \neq P_{0110}$. In such a case, we obtain Sen_L,

Sen_R, Sp_L, and Sp_R as follows:

$$\text{Sen}_R = \frac{P_{\cdot 1 \cdot 1}}{P_{\cdots 1}} = \frac{P_{0101}+P_{0111}+P_{1101}+P_{1111}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 01}} \tag{3.13}$$

$$\text{Sen}_L = \frac{P_{1 \cdot 1 \cdot}}{P_{\cdot 1 \cdot \cdot}} = \frac{P_{1010}+P_{1011}+P_{1110}+P_{1111}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 10}} \tag{3.14}$$

$$\text{Sp}_R = \frac{P_{\cdot 0 \cdot 0}}{P_{\cdots 0}} = \frac{P_{0000}+P_{0010}+P_{1000}+P_{1010}}{P_{\cdots 0}} \tag{3.15}$$

$$\text{Sp}_L = \frac{P_{0 \cdot 0 \cdot}}{P_{\cdot 0 \cdot \cdot}} = \frac{P_{0000}+P_{0001}+P_{0100}+P_{0101}}{P_{\cdot 0 \cdot \cdot}}. \tag{3.16}$$

Note that in the absence of exchangeability, sensitivities and specificities need to be calculated separately for each of the fellow eyes. This presents an inconvenience to practitioners, since a single measure is preferred in clinical practice. The binocular sensitivity and specificity bSen and bSp defined in Chapter 2 become even more relevant in this case, as they yield eye-independent accuracy measures that, at the same time, incorporate the association between fellow eyes.

It is easy to show the following relationships between the two sets of accuracy measures:

$$\text{bSen} = (w_1+w_3)\text{Sen}_R + (w_1+w_2)\text{Sen}_L - \frac{P_{1111}+P_{1110}+P_{1101}}{P_{\cdot\cdot 11}+P_{\cdot\cdot 01}+P_{\cdot\cdot 10}} \tag{3.17}$$

$$\begin{aligned}\text{bSp} = {} & \frac{1}{2}\left(1+\frac{P_{\cdot\cdot 10}}{P_{\cdot\cdot 00}}\right)\left(\text{Sp}_R - \frac{P_{0010}+P_{1000}+P_{1010}}{P_{\cdot\cdot 00}+P_{\cdot\cdot 10}}\right) \\ & + \frac{1}{2}\left(1+\frac{P_{\cdot\cdot 01}}{P_{\cdot\cdot 00}}\right)\left(\text{Sp}_L - \frac{P_{0001}+P_{0100}+P_{0101}}{P_{\cdot\cdot 00}+P_{\cdot\cdot 01}}\right),\end{aligned} \tag{3.18}$$

where w_1, w_2, and w_3 are the weights defined in Chapter 2, and (3.18) follows by adding the following:

$$\begin{aligned}\text{Sp}_R &= \left(\frac{P_{\cdot\cdot 00}}{P_{\cdot\cdot 00}+P_{\cdot\cdot 10}}\right)\text{bSp} + \frac{P_{0010}+P_{1000}+P_{1010}}{P_{\cdot\cdot 00}+P_{\cdot\cdot 10}} \\ \text{Sp}_L &= \left(\frac{P_{\cdot\cdot 00}}{P_{\cdot\cdot 00}+P_{\cdot\cdot 01}}\right)\text{bSp} + \frac{P_{0001}+P_{0100}+P_{0101}}{P_{\cdot\cdot 00}+P_{\cdot\cdot 01}}.\end{aligned}$$

MLEs of bSen and bSp are obtained by plug-in method due to the invariance property of MLEs. The corresponding tandard errors are calculated by delta method or by the bootstrap approach.

Table 3.3: Comparison of true joint probabilities $P_{\ell_1 r_1 \ell_2 r_2}$ and corresponding relative frequencies for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.43$, $\pi_{2R} = 0.52$, and $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.34$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0.21$, for $N = 1000$.

Joint probability	True value	Relative frequency	Frequency
P_{0000}	0.1502	0.1620	162
P_{0001}	0.0913	0.0910	91
P_{0010}	0.0595	0.0620	62
P_{0100}	0.0451	0.0430	43
P_{1000}	0.0332	0.0270	27
P_{0101}	0.0600	0.0600	60
P_{0110}	0.0092	0.0100	10
P_{1001}	0.0102	0.0120	12
P_{1010}	0.0307	0.0290	29
P_{0011}	0.0962	0.1020	102
P_{1100}	0.0993	0.0960	96
P_{1110}	0.0738	0.0720	72
P_{1101}	0.0807	0.0740	74
P_{1011}	0.0286	0.0300	30
P_{0111}	0.0386	0.0330	33
P_{1111}	0.0931	0.0970	97

3.5 Simulation study

In this section, we report the results of simulations to explore the small sample properties of MLEs for the Gaussian copula model, including the accuracy measures. The simulations are based on paired binocular binary diagnostic data $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$, $i = 1, \cdots, N$, with joint probabilities $P_{\ell_1 r_1 \ell_2 r_2}$ from (3.5). Given N, data on disease status Y_{i2L} and Y_{i2R} and on diagnoses Y_{i1L} and Y_{i1R}, for fixed $\boldsymbol{\theta} = (\pi_{1L}, \pi_{1R}, \pi_{2L}, \pi_{2R}, \widetilde{\rho}_1, \widetilde{\rho}_2, \widetilde{\rho}_3, \widetilde{\rho}_4)^\top$, are generated using the following algorithm:

1. generate uniform random number U_i and assign people to the respective group by using the probabilities calculated in Step 1;

2. obtain the MLE of $\boldsymbol{\theta} = (\pi_{1L}, \pi_{1R}, \pi_{2L}, \pi_{2R}, \widetilde{\rho}_1, \widetilde{\rho}_2, \widetilde{\rho}_3, \widetilde{\rho}_4)^\top$ using the optimization function `optim` in R with "`method=Nelder-Mead`", where the function `pmvnorm` in R

Table 3.4: Relative bias (in %), standard deviation (SD), large-sample standard error (SE), and relative efficiency of MLEs for $\boldsymbol{\theta}$ and the accuracy measures for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.43$, $\pi_{2R} = 0.52$, $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.34$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0.21$, for $N = 2000$.

Parameter	True value	Mean estimate	Relative bias	SD	SE	Relative efficiency
π_{1L}	0.4500	0.4691	4.2422	0.0037	0.0036	0.9811
π_{1R}	0.5000	0.5105	2.1001	0.0035	0.0035	0.9992
π_{2L}	0.4300	0.4276	0.5481	0.0035	0.0035	0.9971
π_{2R}	0.5200	0.5324	2.3867	0.0039	0.0039	1.0063
$\widetilde{\rho}_1$	0.7000	0.7217	3.1004	0.0059	0.0060	1.0275
ρ_1	0.4915	0.5116	4.1626	0.0054	0.0055	1.0068
$\widetilde{\rho}_2$	0.3400	0.3349	1.4919	0.0089	0.0090	1.0063
ρ_2	0.1709	0.1672	1.9094	0.0040	0.0038	0.9636
$\widetilde{\rho}_3$	0.2100	0.2078	1.0592	0.0061	0.0059	0.9712
ρ_3	0.1325	0.1323	1.0830	0.0040	0.0039	0.9807
$\widetilde{\rho}_4$	0.2100	0.2078	1.0592	0.0061	0.0059	0.9712
ρ_4	0.0928	0.0927	0.2459	0.0055	0.0056	1.0337
Sen_L	0.5265	0.5456	3.6373	0.0042	0.0040	0.9742
Sen_R	0.5462	0.5558	1.7507	0.0044	0.0043	0.9904
Sp_L	0.6080	0.5880	3.2871	0.0042	0.0044	1.0485
Sp_R	0.5464	0.5369	1.6848	0.0045	0.0044	0.9861
bSen	0.3737	0.3740	0.0963	0.0042	0.0043	1.0141
bSp	0.4585	0.4483	2.1739	0.0040	0.0041	1.0211

is used to evaluate the 4-dimensional multivariate normal CDF to evaluate $P_{\ell_1 r_1 \ell_2 r_2}$ in (3.5);

3. go to 1 and repeat $R = 1000$ times.

For computational convenience, we re-parametrize $\widetilde{\rho}_m$ as Fisher's z-transformation $\widetilde{\eta}_m = \log\{(1+\widetilde{\rho}_m)/(1-\widetilde{\rho}_m)\}$, to remove the constraints on $\widetilde{\rho}_m$; this implies that $\widetilde{\rho}_m = (e^{2\widetilde{\eta}_m} - 1)/(e^{2\widetilde{\eta}_m} + 1)$. For each of the parameter configurations, $R = 1000$ simulation repeats are used to examine the performance in finite samples of MLEs and their large-sample standard errors described in Section 3.3.1. Given a parameter θ_h and an

estimate $\widehat{\theta}_h$, the relative bias (in %) and relative efficiency of $\widehat{\theta}_h$ are

$$\begin{aligned} \text{Relative bias of } \widehat{\theta}_h &= \left(\frac{\text{Mean of } \widehat{\theta}_h - \theta_h}{\theta_h} \right) \times 100\% \\ \text{Relative efficiency of } \widehat{\theta}_h &= \frac{\text{Mean of SE}(\widehat{\theta}_h)}{\text{Empirical SD of } \widehat{\theta}_h}, \end{aligned}$$

where the mean of $\widehat{\theta}_h$ is obtained as the average of the $R = 1000$ realizations of $\widehat{\theta}_h$. All the Gaussian copula models considered in the simulations yield valid joint distributions for $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$, in that the joint probabilities $P_{\ell_1 r_1 \ell_2 r_2}$ are all positive and sum to one. In fact, this is the main criteria for selecting the parameter configurations in the simulations, since it is imperative that the resulting joint distributions be proper distributions.

Table 3.1 displays a comparison of the joint probabilities and their corresponding empirical relative frequencies for the Gaussian copula model with $\pi_{1L} = 0.53$, $\pi_{1R} = \pi_{2L} = \pi_{2R} = 0.5$, and $\widetilde{\rho}_1 = \widetilde{\rho}_2 = 0.82$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0$, for one repeat with $N = 1000$ patients; note that $P_{\ell_1 r_1 \ell_2 r_2} > 0$ and $\sum_{\ell_1, r_1, \ell_2, r_2} P_{\ell_1 r_1 \ell_2 r_2} = 1$, so that the joint probability distribution of $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$ is a valid distribution. Table 3.2 shows the empirical performance of the MLEs in terms of relative bias and relative efficiency. Inspection of the parameter-specific biases of $\widehat{\theta}_h$ clearly indicates that maximum likelihood estimation yields reasonably unbiased estimates. This result is to be expected based on theoretical arguments, considering that $N = 1000$, a fairly large sample size. The parameter-specific relative efficiencies in Table 3.2 are likewise very close to one, indicating that the MLEs for the Gaussian copula model have large-sample standard errors which reflect the true sampling variability.

Table 3.3 displays a comparison of the joint probabilities and their corresponding empirical relative frequencies for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.43$, $\pi_{2R} = 0.52$, and $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.34$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0.21$, for one repeat with $N = 1000$ patients; note that $P_{\ell_1 r_1 \ell_2 r_2} > 0$ and $\sum_{\ell_1, r_1, \ell_2, r_2} P_{\ell_1 r_1 \ell_2 r_2} = 1$, so that the model

Table 3.5: Relative bias (in %), standard deviation (SD), large-sample standard error (SE), and relative efficiency of MLEs for $\boldsymbol{\theta}$ and the accuracy measures for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.42$, $\pi_{2R} = 0.52$, $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.33$, $\widetilde{\rho}_3 = 0.13$, $\widetilde{\rho}_4 = 0.22$, for $N = 2000$.

Parameter	True value	Mean estimate	Relative bias	SD	SE	Relative efficiency
π_{1L}	0.4500	0.4689	4.1914	0.0028	0.0028	0.9956
π_{1R}	0.5000	0.5147	–4.6892	0.0028	0.0027	0.9830
π_{2L}	0.4200	0.4274	1.7643	0.0026	0.0025	0.9990
π_{2R}	0.5200	0.5375	3.3664	0.0030	0.0031	1.0128
$\widetilde{\rho}_1$	0.7000	0.7278	3.9715	0.0046	0.0045	0.9950
ρ_1	0.4865	0.5068	4.1737	0.0043	0.0044	1.0322
$\widetilde{\rho}_2$	0.3300	0.3221	–2.4023	0.0066	0.0067	1.0049
ρ_2	0.1701	0.1630	–4.1582	0.0024	0.0025	1.0129
$\widetilde{\rho}_3$	0.1300	0.1358	4.4378	0.0074	0.0077	1.0447
ρ_3	0.0827	0.0862	4.1474	0.0047	0.0044	0.9326
$\widetilde{\rho}_4$	0.2200	0.2096	–4.7119	0.0037	0.0042	1.1401
ρ_4	0.1075	0.1029	–4.3370	0.0039	0.0040	1.0185
Sen_L	0.4983	0.5186	4.0696	0.0039	0.0038	0.9756
Sen_R	0.5934	0.5709	–3.7984	0.0033	0.0032	0.9709
Sp_L	0.5851	0.5683	–2.8651	0.0035	0.0032	0.9293
Sp_R	0.5138	0.5352	4.1621	0.0035	0.0035	0.9936
bSen	0.3760	0.3611	–3.9786	0.0025	0.0025	0.9971
bSp	0.4267	0.4385	2.7584	0.0030	0.0029	0.9782

is a proper one. Table 3.4 shows the empirical performance of the MLEs in terms of relative bias and relative efficiency. Inspection of the parameter-specific biases of $\widehat{\theta}_h$ clearly indicates that the MLEs are asymptotically unbiased. Again, this result is to be expected based on theoretical arguments, since $N = 1000$ is fairly large. The parameter-specific relative efficiencies in Table 3.4 are all uniformly close to unity, again indicating that the standard errors of the MLEs for the Gaussian copula model can be captured quite well.

Figures 3.1 and 3.3 depict the histograms of MLEs $\widehat{\pi}_{1L}$, $\widehat{\pi}_{1R}$, $\widehat{\pi}_{2L}$, and $\widehat{\pi}_{2R}$, along with MLEs $\widehat{\rho}_1$, $\widehat{\rho}_2$, $\widehat{\rho}_3$, and $\widehat{\rho}_4$ for the Pearson's correlations (those for $\widetilde{\rho}_1, \cdots, \widetilde{\rho}_4$ are not

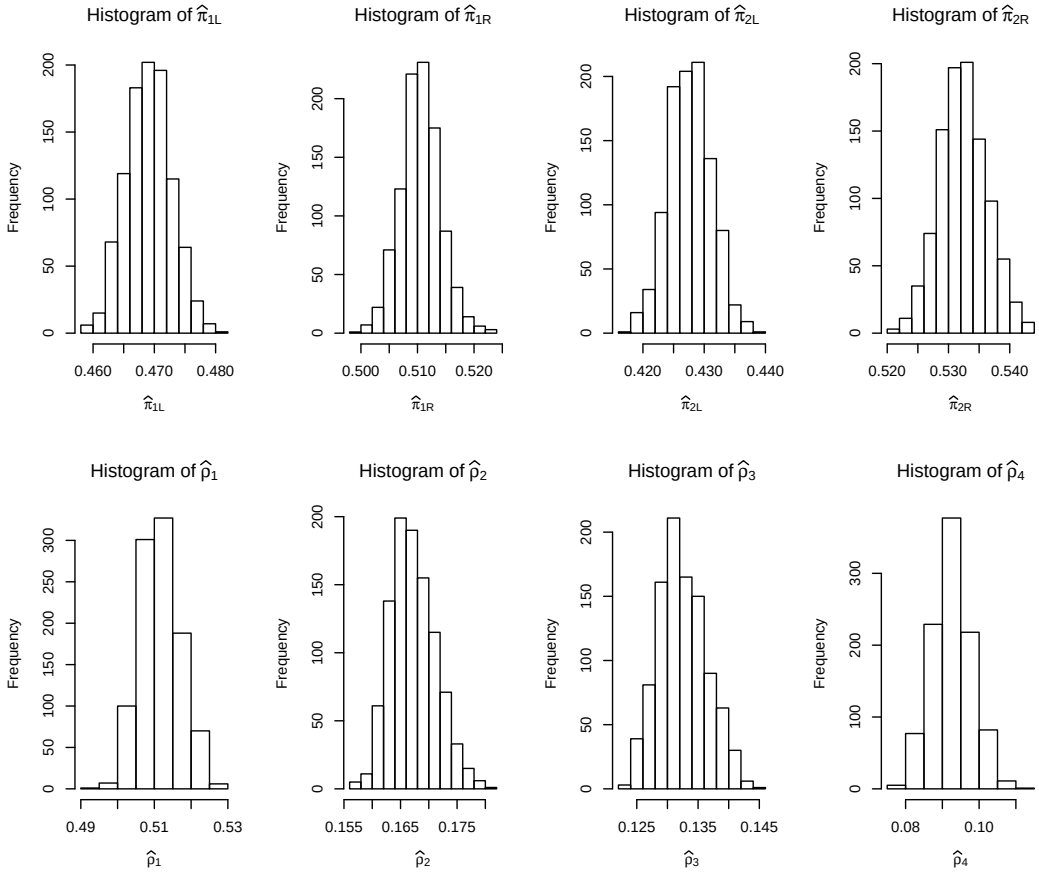

Figure 3.1: Histograms of MLEs for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.43$, $\pi_{2R} = 0.52$, $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.34$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0.21$, for $N = 1000$

reported anymore), and the resulting MLEs of conventional and binocular accuracy measures, for the Gaussian copula model considered in Table 3.4. QQ plots for the MLEs are also shown in Figures 3.2 and 3.4. Both histograms and QQ plots indicate support for normality of MLEs.

Finally, Table 3.5 reports the empirical bias and efficiency of the MLEs for the case where $\widetilde{\rho}_3 \neq \widetilde{\rho}_4$ such that $\widetilde{\rho}_3 \neq 0$ and $\widetilde{\rho}_4 \neq 0$; the same conclusions apply here as in Tables 3.2 and 3.4.

From Tables 3.2, 3.4, and 3.5, we can see that both binocular measures bSen and bSp are less than either of their eye-specific counterparts. In fact, they are smaller than the average of the left- and right-eye measures. The same observations apply

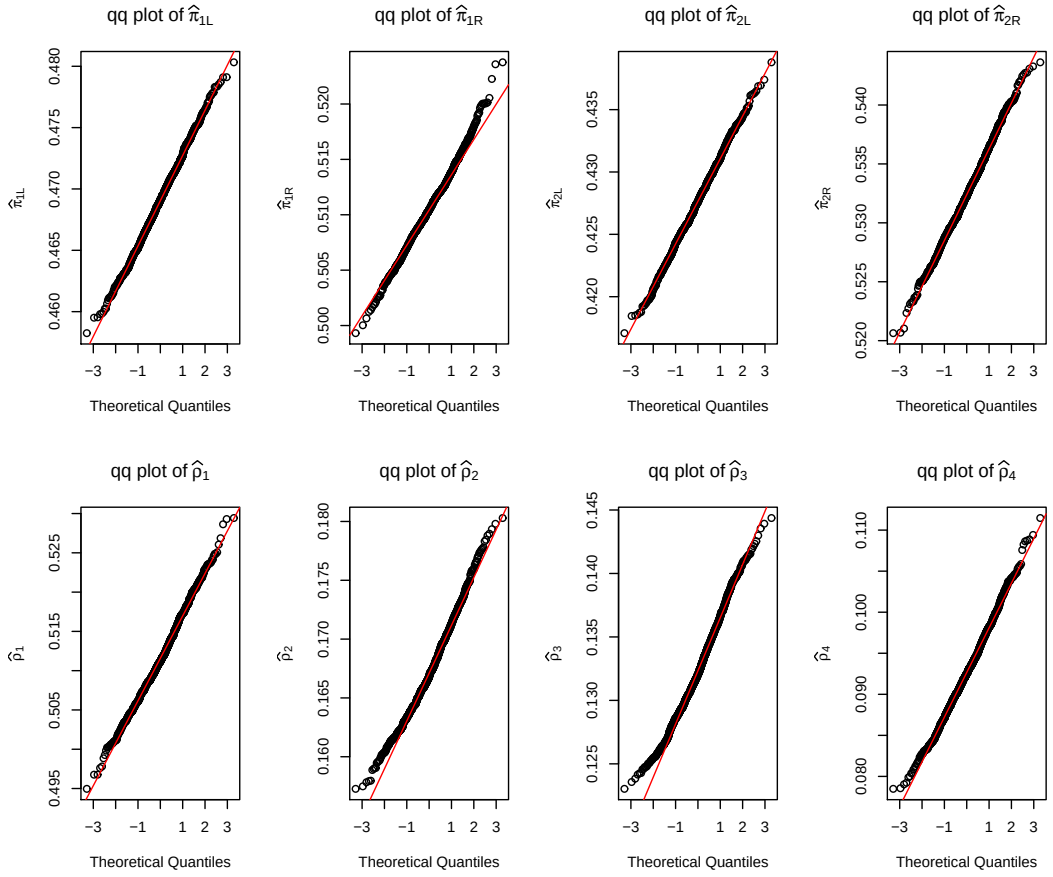

Figure 3.2: QQ plots of MLEs for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.43$, $\pi_{2R} = 0.52$, $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.34$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0.21$, for $N = 1000$.

to the results displayed in Table 3.4. While it is clear that a single measure based on exchangeability will not suffice in this case (as the left- and right-eye measures are clearly different), using single eye-specific measures Sen_j and Sp_j is not what clinicians prefer in practice. Binocular measures are especially advantageous in this regard.

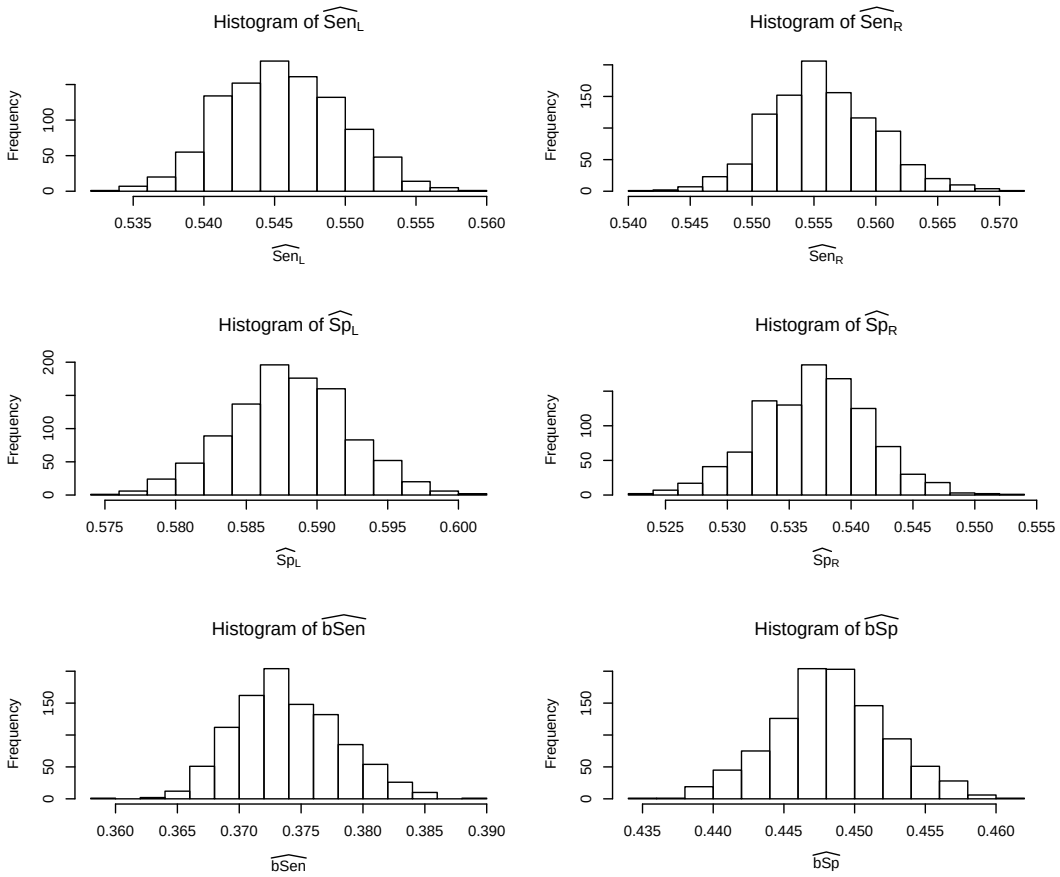

Figure 3.3: Histograms of accuracy measures Sen_L, Sen_R, Sp_L, Sp_R, bSen, and bSp for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.43$, $\pi_{2R} = 0.52$, $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.34$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0.21$, for $N = 1000$.

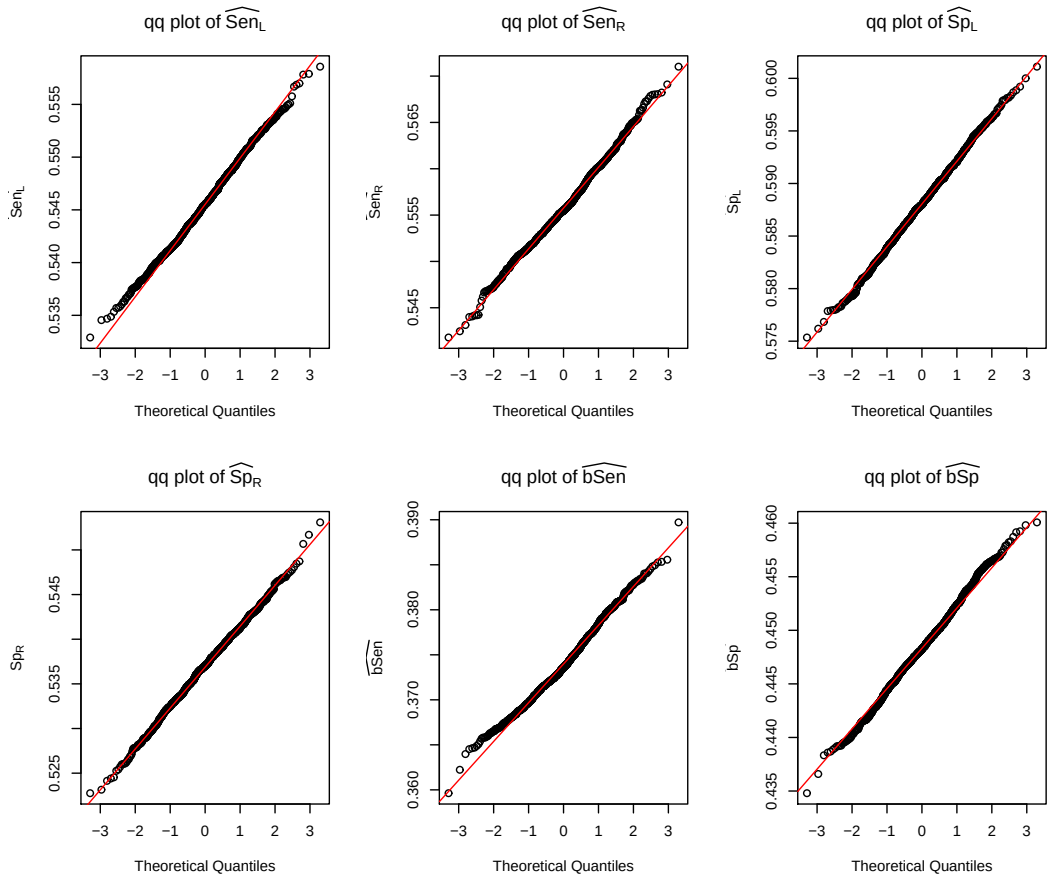

Figure 3.4: QQ plots of accuracy measures Sen_L, Sen_R, Sp_L, Sp_R, bSen, and bSp for the Gaussian copula model with $\pi_{1L} = 0.45$, $\pi_{1R} = 0.5$, $\pi_{2L} = 0.43$, $\pi_{2R} = 0.52$, $\widetilde{\rho}_1 = 0.7$, $\widetilde{\rho}_2 = 0.34$, $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0.21$, for $N = 1000$.

Chapter 4

Application to diabetic retinopathy data

4.1 Introduction

In this chapter, data from the diabetic retinopathy study in [1] are used to illustrate the calculation of the new binocular measures of screening accuracy for the two models studied in Chapters 2 and 3, namely, the ECCM and the Gaussian copula model.

The study involved $N = 94$ diabetic patients in Alberta, Canada. Its purpose was to compare high-resolution digital photography with contact lens biomicroscopy (CLBM) for screening retinal thickening and other retinopathy-related pathologies. On the same day after the patients underwent clinical retinal examination with CLBM, which is considered to be the 'gold standard', by a retinal specialist, they received high-resolution stereoscopic digital imaging of the macula. The digital images were stored uncompressed and then graded by an experienced reader at least two months after they were taken. They were assessed in random order, with a minimum of two months in between review of the left eye images and those of the right eyes to minimize reader recall. Retinal thickening as well as several pathologies that are indicative of retinal thickening were marked as either present (positive) or absent (negative). These pathologies included CSME, microaneurysms, intra-retinal haemorrhage, hard exudates, and others. The results were then compared with those from CLBM. Digital photography provides a cost-effective tele-ophthalmology system, given that adequate identification of potentially treatable retinopathy be ensured before wide implementation.

In what follows, the pathologies macular edema and hard exudate are considered. Macular edema pertains to the thickening and swelling of the eye's macula due to fluid

Table 4.1: Number $n_{y_1y_2}$ of patients with y_1 diagnosis-positive eyes and y_2 status-positive eyes for the pathologies macular edema and hard exudate.

Macular edema					Hard exudate				
y_1	y_2			$n_{y_1\cdot}$	y_1	y_2			$n_{y_1\cdot}$
	0	1	2			0	1	2	
0	42	2	3	47	0	42	5	1	48
1	6	10	2	18	1	4	8	5	17
2	3	2	24	29	2	1	3	25	29
$n_{\cdot y_2}$	51	14	29	94	$n_{\cdot y_2}$	47	16	31	94

and protein deposits, while hard exudate involves the leakage of fluid and lipoprotein into the retina of the eye. We first adopt the ECCM to model the data, thus assuming exchangeability of left- and right-eye diagnoses. The Gaussian copula model is then applied to the data without the exchangeability assumption, with eye-specific sensitivity and specificity along with their binocular analogues. Table 4.1 displays the data concerning the numbers $n_{y_1y_2}$ of patients with $y_1 = 0,1,2$ eyes diagnosed as positive by the reader, and $y_2 = 0,1,2$ eyes with positive status for a pathology from a total of $N = 94$ diabetic patients.

4.2 Results for ECCM

The ECCM in Chapter 2 is adopted in what follows to model the paired binocular binary diagnostic and status data in Table 4.1. Estimates shown in Table 4.2 are moments estimates of the parameters of the ECCM for the pathologies macular edema and hard exudate; the large-sample standard errors of the estimates for the two pathologies are also reported. The estimates yield a valid joint distribution for the paired binocular binary data, as the estimated joint probabilities $\widehat{P}_{\ell_1 r_1 \ell_2 r_2}$ sum to one and are all positive.

The fellow-eye status correlation estimates of $\widehat{\rho}_2 = 0.6855$ and $\widehat{\rho}_2 = 0.6498$ for macular edema and hard exudate, respectively, suggest a moderately strong correlation

between the disease status of the left and right eyes [15]. Also, the reader diagnoses of the two pathologies for the left and right eyes appear to be reasonably moderately correlated, as indicated by the correlation estimates $\widehat{\rho}_1 = 0.6030$ and $\widehat{\rho}_1 = 0.6235$ for macular edema and hard exudate, respectively. The aggregate correlations $\widehat{\rho} = 0.7467$ for macular edema and $\widehat{\rho} = 0.8231$ for hard exudate indicate that the associations can be mainly attributed to the correlation between the aggregate counts of diagnosis-positive and status-positive eyes of patients [15].

Estimates of measures of diagnostic accuracy are also shown in Table 4.2. The estimated sensitivity is 0.8693 for macular edema and 0.8783 for hard exudate. In addition, the estimated specificity is in the range between 88% and 94% for both pathologies. In contrast, the estimates of binocular sensitivity and specificity range from 76% to 94% for both pathologies, a slightly wider range compared to their conventional versions. Note that, on the one hand, the conventional sensitivity and specificity, while based on the joint distribution of the diagnosis and status data $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$, are effectively only marginal measures. The binocular measures, on the other hand, are joint measures that consider both left- and right-eye diagnoses simultaneously along with the corresponding left- and right-eye status. As such, they provide more complete evaluations of the screening accuracy of the test.

4.3 Results for Gaussian copula model

In this section, the Gaussian copula model in Chapter 3 is adopted to analyze the examination results in Table 4.1. We first consider the full model with both $\widetilde{\rho}_3 \neq 0$ and $\widetilde{\rho}_4 \neq 0$. We then consider the reduced models with $\widetilde{\rho}_4 = 0$, and with $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0$. Estimates shown in Tables 4.3, 4.4, and 4.5 are the MLEs obtained by maximizing the joint likelihood of model in (3.8) and (3.9); also shown are their large-sample standard

Table 4.2: Moments estimates and their large-sample standard errors (SE) for the ECCM for pathologies macular edema and hard exudate.

Parameter	Macular edema		Hard exudate	
	Estimate	SE	Estimate	SE
π_1	0.4062	0.0384	0.4010	0.0324
π_2	0.3853	0.0423	0.4166	0.0219
ρ_1	0.6030	0.0805	0.6235	0.0667
ρ_2	0.6856	0.0767	0.6498	0.1155
ρ	0.7467	0.0435	0.8231	0.0372
Sen	0.8693	0.0443	0.8784	0.0447
Sp	0.8842	0.0282	0.9399	0.0194
bSen	0.7597	0.0373	0.7558	0.0513
bSp	0.8568	0.0360	0.9374	0.0246

errors calculated using the bootstrap method with $B = 100$ bootstrap samples. The estimates yield valid joint distributions for the paired binocular binary data, since the estimated joint probabilities $\widehat{P}_{\ell_1 r_1 \ell_2 r_2}$ sum to one and are all positive.

Noting that the reduced copula models are nested within the full model, it is straightforward to determine which of the models provide the best fit via standard likelihood ratio tests. For example, we can test $H_0 : \widetilde{\rho}_4 = 0$ against $H_1 : \widetilde{\rho}_4 \neq 0$ to determine if the improvement in the model fit provided by the full model over the reduced model with $\widetilde{\rho}_4 = 0$ is statistically significant. The same can be done to compare the two reduced models. With p-value $= 0.0002$ (log-likelihood $= -189.92$ for full model vs. log-likelihood $= -196.05$ for reduced model with $\widetilde{\rho}_4 = 0$) and p-value $= 0.026$ (log-likelihood $= -195.79$ for full model vs. log-likelihood $= -197.82$ for reduced model with $\widetilde{\rho}_4 = 0$), respectively, for macular edema and hard exudate for the test $H_0 : \widetilde{\rho}_4 = 0$ against $H_1 : \widetilde{\rho}_4 \neq 0$, we determined that the full model is statistically significantly better than the reduced model with $\widetilde{\rho}_4 = 0$, which in turn is statistically significantly better than the reduced model with $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0$ for both pathologies, with p-values ≈ 0; in this latter test, log-likelihood $= -196.05$ (-197.82) for reduced model $\widetilde{\rho}_4 = 0$ vs.

Table 4.3: MLEs and their large-sample standard errors (SE) for the full copula model for pathologies macular edema and hard exudate.

Parameter	Macular edema		Hard exudate	
	Estimate	SE	Estimate	SE
π_{1L}	0.3432	0.0026	0.4213	0.0024
π_{1R}	0.4182	0.0027	0.4797	0.0027
π_{2L}	0.4283	0.0025	0.5024	0.0022
π_{2R}	0.2304	0.0029	0.4332	0.0029
$\widetilde{\rho}_1$	0.4551	0.0051	0.5709	0.0044
ρ_1	0.2937	0.0045	0.3840	0.0040
$\widetilde{\rho}_2$	–0.1501	0.0074	0.2052	0.0072
ρ_2	0.1375	0.0028	0.2293	0.0024
$\widetilde{\rho}_3$	0.7074	0.0070	0.5167	0.0066
ρ_3	0.4886	0.0043	0.3430	0.0042
$\widetilde{\rho}_4$	0.5396	0.0009	0.5006	0.0043
ρ_4	0.5169	0.0043	0.4337	0.0040
Sen_L	0.6116	0.0041	0.5897	0.0043
Sen_R	0.7932	0.0029	0.7040	0.0032
Sp_L	0.8576	0.0028	0.7488	0.0032
Sp_R	0.7553	0.0041	0.7293	0.0037
bSen	0.5905	0.0024	0.5851	0.0021
bSp	0.6448	0.0028	0.5758	0.0027

log-likelihood $= -210.53$ (-213.2) for reduced model with $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0$, for macular edema (hard exudate). The discussion below is thus confined to results for the full model.

From Table 4.3, the fellow-eye status correlation estimates of $\widehat{\rho}_2 = 0.1375$ and $\widehat{\rho}_2 = 0.2293$ for macular edema and hard exudate, respectively, suggest a relatively weak positive correlation between the left- and right-eye disease status. Also, the reader diagnoses for left- and right-eyes for the two pathologies are similarly associated, as indicated by the correlation estimates $\widehat{\rho}_1 = 0.2937$ and $\widehat{\rho}_1 = 0.384$ for macular edema and hard exudate, respectively. The estimated correlations between diagnosis and disease status for the same eyes given by $\widehat{\rho}_3$ and $\widehat{\rho}_4$, for the left and right eyes, respectively, range between 0.48 and 0.51 for macular edema and between 0.34 and

0.43 from hard exudate.

Also shown in Table 4.3 are the MLEs of sensitivity of left and right eyes $\widehat{\text{Sen}}_L = 0.6116$ and $\widehat{\text{Sen}}_R = 0.7932$, respectively, for macular edema, which are comparable to $\widehat{\text{Sen}} = 0.8693$ obtained from the the ECCM; as well, note that $\widehat{\text{Sen}}_L \neq \widehat{\text{Sen}}_R$, i.e., the sensitivities are not the same for the left and right eye, thus indicating that exchangeability does not hold. Similarly, the MLEs of specificity of left and right eyes are $\widehat{\text{Sp}}_L = 0.8576$ and $\widehat{\text{Sp}}_R = 0.7553$, respectively, for the pathology macular edema. These again compare favourably with $\widehat{\text{Sp}} = 0.8841$ in Table 4.2 for the ECCM. The same conclusion can be made for the pathology hard exudate. These reinforce the earlier conclusion based on the sensitivities that the exchangeability of fellow eyes does not hold.

MLEs of the binocular diagnostic measures are also displayed in Table 4.3. Observe that while $\widehat{\text{bSen}} < min\{\widehat{\text{Sen}}_L, \widehat{\text{Sen}}_R\}$ and $\widehat{\text{bSp}} < min\{\widehat{\text{Sp}}_L, \widehat{\text{Sp}}_R\}$ for both macular edema and hard exudate. Note that in the absence of exchangeability, the binocular measures provide only a pair of measures, like in the usual diagnostic scenario, in contrast to the conventional measures that rely on four eye-specific measures that may not be appealing to clinical practitioners.

4.4 Discussion

In this chapter, data from the diabetic retinopathy study were used to illustrate the calculation of the proposed binocular measures of screening accuracy using the ECCM and the Gaussian copula model discussed in Chapters 2 and 3. Results of the analyses provide helpful illustrations of the application of the proposed methodologies. While the results are generally the same for the two sets of accuracy measures in the case of the ECCM, results for the Gaussian copula indicate that the assumption of exchangeability of fellow eyes under the ECCM does not hold. In addition, not only

are the left- and right-eye sensitivities and specificities quite different, the two sets of measures (i.e., $\{\widehat{\text{Sen}}_L, \widehat{\text{Sen}}_R, \widehat{\text{Sp}}_L, \widehat{\text{Sp}}_R\}$ and $\{\widehat{\text{bSen}}, \widehat{\text{bSp}}\}$) yielded estimates that differ substantially in magnitudes. This suggests that the binocular measures should be preferred in practice when screening paired organs, since they simultaneously consider the two organs and incorporate the organs' correlation, thus providing a better, more complete and more unified assessment of screening tests for paired organs.

Table 4.4: MLEs and their large-sample standard errors (SE) for the reduced copula model for pathologies macular edema and hard exudate with $\widetilde{\rho}_4 = 0$.

Parameter	Macular edema		Hard exudate	
	Estimate	SE	Estimate	SE
π_{1L}	0.3534	0.0025	0.3432	0.0028
π_{1R}	0.3872	0.0027	0.4182	0.0027
π_{2L}	0.3761	0.0022	0.4283	0.0024
π_{2R}	0.2198	0.0026	0.2304	0.0027
$\widetilde{\rho}_1$	0.6614	0.0048	0.4551	0.0046
ρ_1	0.4540	0.0025	0.2933	0.0025
$\widetilde{\rho}_2$	–0.4099	0.0075	–0.1502	0.0072
ρ_2	0.0639	0.0029	0.1373	0.0031
$\widetilde{\rho}_3$	0.5064	0.0059	0.7074	0.0067
ρ_3	0.3316	0.0045	0.4885	0.0044
Sen_L	0.5568	0.0009	0.6111	0.0009
Sen_R	0.7261	0.0011	0.7932	0.0010
Sp_L	0.7694	0.0036	0.8574	0.0037
Sp_R	0.7676	0.0032	0.7556	0.0041
bSen	0.4721	0.0036	0.5907	0.0037
bSp	0.6423	0.0036	0.6451	0.0036

Table 4.5: MLEs and their large-sample standard errors (SE) for the reduced copula model for pathologies macular edema and hard exudate with $\widetilde{\rho}_3 = \widetilde{\rho}_4 = 0$.

Parameter	Macular edema		Hard exudate	
	Estimate	SE	Estimate	SE
π_{1L}	0.3854	0.0125	0.3832	0.0117
π_{1R}	0.4145	0.0128	0.4173	0.0117
π_{2L}	0.3578	0.0124	0.4278	0.0101
π_{2R}	0.3818	0.0136	0.4056	0.0094
$\widetilde{\rho}_1$	0.7924	0.0193	0.8348	0.0115
ρ_1	0.5780	0.0200	0.6237	0.0133
$\widetilde{\rho}_2$	0.8501	0.0185	0.8518	0.0118
ρ_2	0.6408	0.0193	0.6461	0.0118
Sen_L	0.3854	0.0129	0.3832	0.0120
Sen_R	0.4146	0.0134	0.4172	0.0117
Sp_L	0.6146	0.0125	0.6167	0.0112
Sp_R	0.5854	0.0130	0.5827	0.0115
bSen	0.3544	0.0127	0.3530	0.0110
bSp	0.4984	0.0193	0.5089	0.0126

Chapter 5

Conclusion

5.1 Summary

The primary purpose of this thesis was to introduce new binocular accuracy measures as alternatives to conventional measures that can be used to evaluate screening tests in studies involving paired organs (e.g., eyes, ears). Particular attention is given to cross-sectional designs where both diagnoses and disease status are determined after study enrolment or sampling. Specifically, we considered screening/diagnostic studies yielding paired binocular binary data $(Y_{i1L}, Y_{i1R})^\top$ and $(Y_{i2L}, Y_{i2R})^\top$, where the former pertains to the left- and right-organ diagnoses and the latter to the left- and right-organ disease status.

To define the new binocular measures, we considered two models, namely, the extended common correlation model (ECCM) and one generated directly from the Gaussian copula. The first relies on the assumption of exchangeability of fellow organs while the second is more flexible. However, while the ECCM lends itself easily to moments estimation, thus obviating the need for iterative calculations of estimates, numerically implemented maximum likelihood estimation is readily adapted to the Gaussian copula model. Standard asymptotic results for MLEs and moments estimates apply, so that traditional large-sample standard errors can be used.

Binocular versions of sensitivity and specificity are defined and obtained for the ECCM and the Gaussian copula model. Under exchangeability, the binocular measures represent a simultaneous assessment of accuracy of binocular screening tests in contrast to common sensitivity and specificity that, although independent of eyes,

are based on marginal evaluations. As such, they fail to consider the joint assessments of the paired organs. This becomes a problem in practice, where it is enough for screening tests to have at least one organ to be positive before a patient is sent for further confirmatory testing.

Without the assumption of exchangeability of fellow organs, the binocular measures are still to be preferred to eye-specific sensitivities and specificities for the following reasons. One, they provide the usual pair of measures 'binocular sensitivity' and 'binocular specificity', akin to the traditional sensitivity-specificity pair, while conventional eye-specific measures yield four values, a sensitivity-specificity pair for each organ. This is not attractive to practitioners and clinicians, since they usually prefer as few measures as possible.

Two, the binocular measures provide a more complete and unified means of evaluating the screening accuracy of tests by accounting for the correlation between fellow organs. They are also based on a joint assessment of the paired organs, and the result on one of the organs is therefore considered in relation to that of the other.

Comparisons between the conventional and binocular sensitivity and specificity were also carried out for both models. While values for the measures were close in many cases, it is entirely possible that they will be quite different from the conventional measures. Such cases are potentially tricky since the two sets of accuracy measures may lead to divergent results. In such an event, we believe that the binocular measures should take precedence.

Simulation studies conducted to investigate the relative bias, precision of estimates (i.e., moments estimates for the ECCM and MLEs for the Gaussian copula model), and asymptotic normality show good performance of the methodology. We also illustrated our proposed methodology via data from a retinopathy diabetic study.

The proposed methodology studied in this thesis provides a viable alternative ap-

proach to conventional ways of assessing diagnostic accuracy of screening tests for paired organs. The binocular versions of sensitivity and specificity reflect the way screening tests are conducted in practice. They overcome the shortcomings of conventional measures.

5.2 Future research

A number of issues related to the manner the data are modelled should be further explored in a future work. In practice, multiple readers are used to screen or diagnose not one but several diseases, which are likely correlated. In the diabetic retinopathy study, for example, at least two readers evaluate the presence or absence of several retinopathy-related eye pathologies. This gives rise to issues concerning associations between different pathologies, repeated evaluation by readers, and possible correlations within family members. While this setting can be conceptually accommodated within the Gaussian copula framework studied in the thesis, a careful study of the correlation structure is essential; in addition, computational difficulties need to be addressed. Efficient algorithms for estimation will require study in depth. The interpretation of correlation parameters is an important issue in copula modelling. Finally, diagnostic studies often rely on certain generally accepted 'gold standards,' diagnostic tests used to establish the true disease status of patients. However, such tests may not always exist, or even if they do, they may be costly to apply to all patients. Random effects models, for instance one that involves latent classes, might be one potential solution. This may be explored in a future work.

Bibliography

[1] C. J. Rudnisky, B. J. Hinz, M. T. S. Tennant, A. R. de Leon, and M. D. J. Greve, "High-resolution stereoscopic digital fundus photography versus contact lens biomicroscopy for the detection of clinically significant macular edema," *Ophthalmology*, vol. 109, pp. 267–274, 2002.

[2] R. J. Glynn and B. Rosner, "Accounting for the correlation between fellow eyes in regression analysis," *Archives of Ophthalmology*, vol. 110, pp. 381–387, 1992.

[3] P. P. Hujoel, L. H. Moulton, and W. J. Loesche, "Estimation of sensitivity and specificity of site-specific diagnostic tests," *Journal of Periodontal Research*, vol. 25, pp. 193–196, 1990.

[4] B. Rosner, "Multivariate methods for clustered binary data with more than one level of nesting," *Journal of the American Statistical Association*, vol. 84, pp. 373–380, 1989.

[5] P. J. Smith and A. Hadgu, "Sensitivity and specificity for correlated observations," *Statistics in Medicine*, vol. 11, pp. 1503–1509, 1992.

[6] J. N. K. Rao and A. J. Scott, "A simple method for the analysis of clustered binary data," *Biometrics*, pp. 577–585, 1992.

[7] A. Donner and N. Klar, "Confidence interval construction for effect measures arising from cluster randomization trials," *Journal of Clinical Epidemiology*, vol. 46, pp. 123–131, 1993.

[8] E. W. Lee and N. Dubin, "Estimation and sample size considerations for clustered binary responses," *Statistics in Medicine*, vol. 13, pp. 1241–1252, 1994.

[9] C. Ahn, "An evaluation of methods for the estimation of sensitivity and specificity of site-specific diagnostic tests," *Biometrical Journal*, vol. 39, pp. 793–807, 1997.

[10] M. L. C. Leite and A. Nicolosi, "Statistical analysis of correlated binary data in ophthalmology: A weighted logistic regression approach," *Ophthalmic Epidemiology*, vol. 5, pp. 117–131, 1998.

[11] S. L. Zeger and K. Y. Liang, "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics*, vol. 42, pp. 121–130, 1986.

[12] M. R. Sternberg and A. Hadgu, "A GEE approach to estimating sensitivity and specificity and coverage properties of the confidence intervals," *Statistics in Medicine*, vol. 20, pp. 1529–1539, 2001.

[13] W. Leisenring, M. S. Pepe, and G. Longton, "A marginal regression modelling framework for evaluating medical diagnostic tests," *Statistics in Medicine*, vol. 16, pp. 1263–1281, 1997.

[14] A. R. de Leon, M. Guo, C. J. Rudnisky, and G. Singh, "A likelihood approach to estimating sensitivity and specificity for binocular data: Application in ophthalmology," *Statistics in Medicine*, vol. 26, pp. 3300–3314, 2007.

[15] A. R. de Leon, A. Soo, D. C. Bonzo, and C. J. Rudnisky, "Joint estimation of diagnostic accuracy measures for paired organs–application in ophthalmology," *Biometrical Journal*, vol. 51, pp. 837–850, 2009.

[16] T. K. Mak, "Analysing intraclass correlation for dichotomous variables," *Applied Statistics*, vol. 37, pp. 344–352, 1988.

[17] M. M. Shoukri and A. Donner, "Bivariate models for co-aggregation of dichotomous traits in twins," *Statistics in Medicine*, vol. 26, pp. 336–351, 2007.

[18] P. X. K. Song, *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer, 2007.

[19] A. K. Nikoloulopoulos and D. Karlis, "Finite normal mixture copulas for multivariate discrete data modeling," *Journal of Statistical Planning and Inference*, vol. 139, pp. 3878–3890, 2009.

[20] A. K. Nikoloulopoulos and D. Karlis, "Multivariate logit copula model with an application to dental data," *Statistics in Medicine*, vol. 27, pp. 6393–6406, 2008.

[21] D. M. Zimmer and P. K. Trivedi, "Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand," *Journal of Business and Economic Statistics*, vol. 24, pp. 63–76, 2006.

[22] D. Trégouët, P. Ducimetière, V. Bocquet, S. Visvikis, F. Soubrier, and L. Tiret, "A parametric copula model for analysis of familial binary data," *American Journal of Human Genetics*, vol. 64, pp. 886–893, 1999.

[23] S. G. Meester and J. Mackay, "A parametric model for cluster correlated categorical data," *Biometrics*, vol. 50, pp. 954–963, 1994.

[24] C. N. L. Genest and J. Nešlehová, "A primer on copulas for count data," *ASTIN Bulletin*, vol. 37, pp. 475–515, 2007.

[25] P. Embrechts, "Copulas: A personal view," *Journal of Risk and Insurance*, vol. 76, pp. 639–650, 2009.

[26] G. Yin and Y. Yuan, "Bayesian dose finding in oncology for drug combinations by copula regression," *Applied Statistics*, vol. 58, pp. 211–224, 2009.

[27] N. Kolev and D. Paiva, "Copula-based regression models: A survey," *Journal of Statistical Planning and Inference*, vol. 139, pp. 3847–3856, 2009.

[28] J. K. Haseman and L. L. Kupper, "Analysis of dichotomous response data from certain toxicological experiments," *Biometrics*, vol. 35, pp. 281–293, 1979.

[29] K. V. Mardia, *Families of Bivariate Distributions*. Griffin, 1970.

[30] S. Nadarajah and S. Kotz, "Multitude of beta distributions with applications," *Statistics*, vol. 41, pp. 153–179, 2007.

[31] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*. Wiley, 3rd ed., 2003.

[32] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, pp. 378–382, 1971.

[33] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*, vol. 2. Houghton Mifflin, 1970.

[34] D. Oakes and J. Ritz, "Regression in a bivariate copula model," *Biometrika*, vol. 87, pp. 345–352, 2000.

[35] D. Kugiumtzis and E. Bora-Senta, "Normal correlation coefficient of non-normal variables using piece-wise linear approximation," *Computational Statistics*, vol. 25, pp. 645–662, 2010.

[36] T. Vangeneugden, G. Molenberghs, A. Laenen, H. Geys, C. Beunckens, and C. Sotto, "Marginal correlation in longitudinal binary data based on generalized linear mixed models," *Communications in Statistics–Theory and Methods*, vol. 39, pp. 3540–3557, 2010.

[37] R. B. Nelsen, *An Introduction to Copulas*. Springer, 2006.

[38] P. K. Trivedi and D. M. Zimmer, "Copula modeling: An introduction for practitioners," *Foundations and Trends in Econometrics*, vol. 1, pp. 1–111, 2005.

[39] R. T. Clemen and T. Reilly, "Correlations and copulas for decision and risk analysis," *Management Science*, vol. 45, pp. 208–224, 1999.

[40] L. Madsen and Y. Fang, "Joint regression analysis for discrete longitudinal data," *Biometrics*, vol. 67, pp. 1171–1175, 2011.