ABSTRACT

Integrated-Likelihood-Ratio Confidence Intervals Obtained from Data Via a Double-Sampling Scenario

Briceön Curtis Wiley, Ph.D.

Chairperson: James D. Stamey, Ph.D.

Hypothesis testing has been a primary focus of statistical inference. Recently, confidence intervals (CIs) have been suggested as a superior inference form because of the additional information they provide to a scientist to aid decision making. For public health data, business data, and other types of data, misclassification is often present and can cause estimators to be biased, thus leading to incorrect conclusions. Tenenbein (1970) has provided a double-sampling scheme to correct for misclassification through the use of an infallible data set that is combined with a larger fallible data set subject to misclassification. Many authors have utilized the double-sampling procedure to correct for misclassification in their data. When constructing confidence intervals, for instance, Rahardja and Yang (2015) derived Wald intervals for one-sample binomial problems, and Lyles (2002) proposed a Wald interval for two-sample binomial problems. Also, Riggs et al. (2009) provided confidence intervals for one-sample Poisson rate parameters. In addition, Li (2009) built similar intervals for the difference of two Poisson rate parameters.

We derive integrated-likelihood-ratio (ILR) confidence intervals, first proposed by Severini (2010), for each of these situations to demonstrate their effectiveness in estimating parameters from data subject to misclassification. In chapter one, we

derive an ILR CI for a one-sample binomial data set and demonstrate that it has at least nominal coverage while providing narrow average interval widths when the binomial parameter is small. In chapter two, we apply a transformation related to one from Fisher and Robbins (2019) to make the ILR CI less conservative when estimating a one-sample binomial parameter, thus providing closer-to-nominal coverage while decreasing the average interval width. In chapter three, we extend the ILR CI to estimate the log odds-ratio of two binomial parameters when the binary data are subject to misclassification. Finally, in chapter four we demonstrate the ILR CI's efficacy versus the Wald and score CIs for estimating the ratio of two Poisson rate parameters using data sampled via a double-sampling scenario.

Integrated-Likelihood-Ratio Confidence Intervals Obtained from Data Via a Double-Sampling Scenario

by

Briceön Curtis Wiley, B.B.A, M.S.

A Dissertation

Approved by the Department of Statistical Science

_____

James D. Stamey, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

_____

James D. Stamey, Ph.D., Chairperson

_____

Dean M. Young, Ph.D.

_____

David J. Kahle, Ph.D.

_____

J. Allen Seward, Ph.D.

Accepted by the Graduate School
December 2020

_____

Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I want to thank my Lord and Savior Jesus Christ for allowing me to discover a morsel of the treasure mentioned in Colossians 2:2-3. I also want to thank Curtis and Tomitra for being the magnificent gifts of God that you are. Nothing that I have done has been possible without the love, guidance, patience, and prayers of you two. God has taken me to all sorts of places to do all sorts of things, and I have succeeded in each endeavor thanks to the foundation that you have built in my life. You are living proof of the importance of children having strong, God-fearing parents, and I am eternally grateful that He blessed me with both of you.

Thank you, Jordan and Ethan, for being two of the best friends I have on the planet. Without question, I would not be where I am today without you to keep me grounded and joyous. I pray that God will continue to bless me with opportunities to support and encourage you, and I cannot wait to see the amazing ways you both outshine me in the years to come.

桂さん、恵一さん、歩さん、幸一さん、桂子さん、家族に迎えてくださりありがとうございます。 Thank you, Homer, Jennie, Bernard, and Annie, for investing deeply in my life and education. Thank you, Jaylon and Arielle, Jesaiah and Abbey, Pierre and Monica, and Leon and Shawn, for sticking close and helping mold me into the man I am today. In addition, thank you, Grant, Melissa, Chris, Jinjie, Minho, Qida, Phil, and Gabriel for all of the conversations, meals, guidance, and encouragement that helped me complete the program. I also wish to thank Ryan and Christy, Wil and Jin, Kody and Laken, and Jon and Amanda for teaching me about community.

Thank you, Jane Harvill, Jack Tubbs, Amanda Hering, John Seaman, Joon Jin Song, and Dennis Johnston, for great instruction and wisdom both in and out

of the classroom. Thank you Michael Boerm, Alexandre Thiltges, Yoshiko Fujii Gaines, Yuko Prefume, and Hajime Kumahata for teaching me all about the beauty of language. Thank you to Lisa Garrett, Polly Flippin, Jennifer Crosslin, and Ronda Kruse for being my home away from home during my time in Waco. Also, thank you, Joy Young, for your magnificent help with transforming this dissertation from a scientific report to a work of art.

To David Kahle, thank you for being an awesome instructor who made me want to always give my best and never stop learning. To Allen Seward, thank you for inviting me to Business Fellows so many years ago and walking alongside me ever since. Your office and heart were always open to me, and I am eternally grateful that I had the chance to meet you back in 2012.

Thank you, Dean Young, for seeing potential in me when I never could. You pushed me further than I ever expected I could go and helped open my eyes to the wonder of statistics. I deeply appreciate you for still believing in hard work and always reminding me that "necessity breeds innovation" when difficulties arose. Your work ethic taught me how to be a strong statistician, and your character showed me what goes into being a good man.

To James Stamey, thank you for literally everything. I never thought I could succeed in graduate school, but you welcomed me in and convinced me not only that success was possible, but also that I was beyond capable. When I became too distracted, you helped reel me back in; when I was unsure how to move forward, you helped set me back on track; and when I was afraid that I would not be able to complete the program, you helped me find the finish line. This program is truly blessed to have you, and I pray you continue to be such a blessing to many more students for years to come.

DEDICATION

楓へ

まず、心からありがとうと伝えたいです。この論文は私一人のものでなく、楓のものでもあります。いつも変わらない楓の愛と根気強さがなかったら、忙しい毎日に完徹は絶対無理でした。楓は仕事でも家でも頑張っていて、夜遅くまで研究に追われる私に付き合ってくれて、面倒見よく何度もお茶を入れてくれました。この博士号は私にとって一番難しい挑戦でした。それでも、この博士号に捧げた以上に、私が楓との絆のために努力と労力を捧げていることを祈っています。

CHAPTER ONE

An Integrated-Likelihood-Ratio Confidence Interval for a Proportion Based on Misclassified Data from a Double-Sampling Scenario

This chapter is pending publication:

Wiley, B., Elrod, C., Young, P. D., and Young, D. M. (2021). An integrated-likelihood-ratio confidence interval for a proportion based on under-reported and infallible data. *Statistica Neerlandica*

1.1 Introduction

Estimating a binomial probability of success is one of the most common applications of statistical analysis. A frequently overlooked and violated assumption is that all outcomes are correctly classified. However, such an infallible classification method may not exist in practice. Several authors have addressed misclassification problems associated with inference for a single proportion parameter. For instance, Bross (1954) demonstrated that traditional estimation of a proportion parameter in the presence of misclassification produces a biased estimator. Tenenbein (1970) accounted for misclassification with a double-sampling scheme. Lie et al. (1994) and Moors et al. (2000) considered misclassification on a proportion parameter when undercounts are obtained. Boese et al. (2006) developed several interval estimators for the case where double sampling was used when false-positive counts occurred in the fallible sample. In addition, Riggs (2015) derived Wald and Score interval estimates for a proportion parameter using double sampling with inverse sampling of fallible data.

Here, suppose we have a data set that originates from two different classifiers: the larger sample is produced by a fallible classifier, while the smaller validation data set contains results from both the infallible and fallible classifiers. In this double-sampling method, similar to Raats and Moors (2003), the main study and validation

study are produced independently. Table 1.1 summarizes the structure of this double-sampling scheme. We denote m as the total number of observations for the main,

Table 1.1. One-sample misclassified binary data obtained using double sampling.

		Fallible Classifier		
Study	Infallible Classifier	0	1	Total
Sub	0	n_{00}	n_{01}	$n_{0\cdot}$
	1	n_{10}	n_{11}	$n_{1\cdot}$
	Total	$n_{\cdot 0}$	$n_{\cdot 1}$	n
Main	Not available	y	x	m

fallible data set, while n is the total number of observations for the infallible data set. Thus, the total sample size is $N = m + n$. Also, x and y are the positive and negative observation counts, respectively, from the main data set. The cell counts n_{ij} are the total number of cases where the infallible classifier (i) and fallible classifier (j) classify observations. The corresponding probabilities on each of these cell counts is used to construct the likelihood function. Similar to Rahardja and Yang (2015), we denote the infallible and fallible classifier outcomes as T and F, respectively, with the understanding that we have observed data for both classifiers in our validation study, but only results for the fallible classifier in the main data set. We let $F = 1$ and $T = 1$ denote positive outcomes for the fallible and infallible classifiers, respectively, while $F = 0$ and $T = 0$ denote negative outcomes. Furthermore, we have $p := P(T = 1)$ and $\pi := P(F = 1)$, as well as the false-positive and false-negative probabilities $\phi := P(F = 1|T = 0)$ and $\theta := P(F = 0|T = 1)$, respectively. We use the law of total probability to find that $\pi = p(1 - \theta) + (1 - p)\phi$.

Table 1.2 displays the probabilities for each of the cells for both the fallible and infallible counts. Rahardja and Yang (2015) derived an adjusted or modified Wald (mWald) CI for p that yields superior coverage properties and often better average interval widths than the usual Wald CI. Here, we derive a new integrated-likelihood-

ratio (ILR) CI for p. We demonstrate that our derived ILR CI provides superior coverage properties when contrasted with the coverage properties of the mWald CI. Furthermore, although our ILR CI is slightly conservative in terms of coverage, it yields narrower average interval widths when $p < 0.10$ or $p > 0.90$. Thus, the ILR CI is a notable improvement over the mWald CI proposed by Rahardja and Yang (2015). Using a Monte Carlo simulation and real data, we demonstrate that the ILR CI should be implemented instead of the mWald CI, especially when $n > 200$ for all values of p and when $n < 200$ and $p \notin [0.40,\ 0.60]$.

Table 1.2. Cell probabilities for the double-sample multinomial distribution.

		Fallible Classifier		
Study	Infallible Classifier	0	1	Total
Sub	0	$(1-p)(1-\phi)$	$(1-p)\phi$	$(1-p)$
	1	$p\theta$	$p(1-\theta)$	p
Main	Not available	$1-\pi$	π	1

The remainder of the paper is organized as follows. In Section 1.2, we describe the ILR, nWald, and mWald CIs that we contrast here. In Section 1.3, we describe the design and results of four Monte Carlo simulations examining average interval widths and coverage properties. In Section 1.4, we contrast the three considered CIs on real data. Finally, we briefly discuss our results in Section 1.5.

1.2 CI Methods for Estimating p

Using the information from the double-sampling scheme summarized in Tables 1.1 and 1.2, we obtain the likelihood function from which all CIs considered here are derived, namely,

$$L(p|\theta,\phi) = [(1-p)(1-\phi)]^{n_{00}}[(1-p)\phi]^{n_{01}}[p\theta]^{n_{10}}[p(1-\theta)]^{n_{11}}\pi^{x}(1-\pi)^{y}. \qquad (1.1)$$

From (1.1), we construct our ILR CI and present two competing CIs derived by Rahardja and Yang (2015). The goal for each CI is to estimate the binomial proportion parameter p using both the fallible and infallible data.

1.2.1 An Integrated-Likelihood-Ratio CI for p

Motivated by Berger et al. (1999) and Severini (2010), we construct an ILR CI for a single binomial parameter p using the double-sampling method described in Section 1.1. The integrated likelihood is

$$L_I(p) = \int_0^1 \int_0^1 L(p, \theta, \phi) g(\theta, \phi | p) d\theta d\phi, \tag{1.2}$$

where θ and ϕ are the false negative and false positive rates, respectively, and $g(\theta, \phi|p)$ is a weighting function for θ and ϕ. In Section 1.6, we have derived a closed-form expression for (1.2), which is

$$\begin{aligned} L_I(p) = \sum_{i=0}^{x} \sum_{j=0}^{y} \binom{x}{i} \binom{y}{j} (1-p)^{n_{00}+n_{01}-i-j+x+y} p^{n_{10}+n_{11}+i+j} \\ \times B(n_{00}+y-j+1,\ n_{01}+x-i+1) B(n_{10}+j+1,\ n_{11}+i+1), \end{aligned} \tag{1.3}$$

where $B(a, b)$ denotes a Beta function with parameters a and b. Without vectorization, (1.3) can prove to be computationally complex. Therefore, the use of numerical integration with (1.2) can provide results more quickly. However, with vectorization, both numerical integration of (1.2) and direct usage of (1.3) prove to be comparable in speed in R version 3.6.1. One can find the ILR CI by determining all values of p satisfying

$$\left\{ p : -2 \log \left(\frac{L_I(p)}{L_I(\hat{p}_{IL})} \right) < \chi^2_{(1,1-\alpha)} \right\}, \tag{1.4}$$

where $\hat{p}_{IL}$ is the *MLE* of p determined numerically from the likelihood function (1.2), and $\chi^2_{(1,\ 1-\alpha)}$ denotes the $(1-\alpha)$th quantile of a central chi-square distribution with one degree of freedom.

1.2.2 A Wald-type CI for p

Rahardja and Yang (2015) have introduced what they refer to as a "naïve" $100(1-\alpha)\%$ Wald CI for p, or more succinctly, an nWald interval for p. This CI is

$$\hat{p} \pm z_{\alpha/2}\hat{\sigma}, \tag{1.5}$$

where $\hat{p} = \hat{\pi}\hat{\lambda}_1 + (1-\hat{\pi})\hat{\lambda}_2$ is the *MLE* of the likelihood for p, $z_{\alpha/2}$ is the $(\alpha/2)$th quantile of the standard normal distribution, and

$$\hat{\sigma} = \sqrt{\frac{\hat{\pi}\hat{\lambda}_1(1-\hat{\lambda}_1)}{n} + \frac{(1-\hat{\pi})\hat{\lambda}_2(1-\hat{\lambda}_2)}{n} + \frac{(\hat{\lambda}_1-\hat{\lambda}_2)^2\hat{\pi}(1-\hat{\pi})}{N}}.$$

The estimators $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\pi}$ are each calculated under a re-parameterization of the likelihood function (1.1) found in Rahardja and Yang (2015). These *MLE*s are $\hat{\lambda}_1 = n_{11}/n_{\cdot 1}$, $\hat{\lambda}_2 = n_{10}/n_{\cdot 0}$, and $\hat{\pi} = (x + n_{\cdot 1})/N$. Rahardja and Yang (2015) refer to this CI for p as "naïve" because the CI (1.5) can result in bounds that are outside of the interval $[0, 1]$, thus resulting in truncated CIs when p is relatively small or large.

1.2.3 A Modified Wald CI for p

In an effort to improve upon their nWald CI for p, Rahardja and Yang (2015) have suggested the use of the logit transformation so that

$$\hat{\delta} = \text{logit}(\hat{p}) = \log\left(\frac{\hat{p}}{1-\hat{p}}\right). \tag{1.6}$$

This transformation, when applied to the *MLE* of p, yields an improved approximation of the transformed sampling distribution of $\hat{p}$ to a normal distribution. Rahardja and Yang (2015) have shown via the delta method that $\hat{\tau}^2 := \widehat{\text{Var}}(\hat{\delta}) \approx \hat{\sigma}^2/[\hat{p}(1-\hat{p})]^2$. Therefore, a $100(1-\alpha)\%$ CI for δ is

$$\hat{\delta} \pm z_{\alpha/2}\hat{\tau}. \tag{1.7}$$

Rahardja and Yang (2015) exponentiated the endpoints of the CI in (1.7) to obtain what they labeled as a modified Wald, or mWald, CI

$$[\exp(\hat{\delta} - z_{\alpha/2}\hat{\tau})/(\exp(\hat{\delta} - z_{\alpha/2}\hat{\tau}) + 1),\ \exp(\hat{\delta} + z_{\alpha/2}\hat{\tau})/(\exp(\hat{\delta} + z_{\alpha/2}\hat{\tau}) + 1)]. \tag{1.8}$$

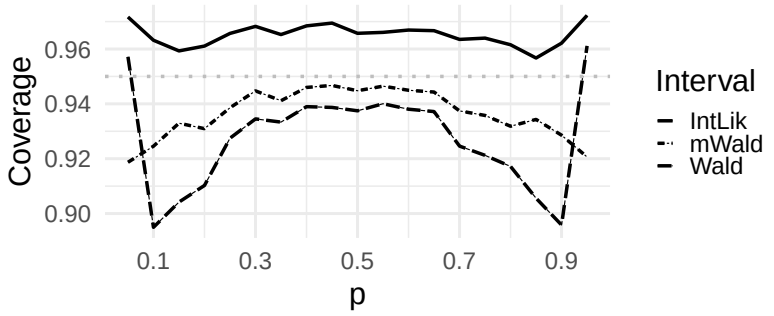
1.3 Monte Carlo Simulations

To determine the CI coverage properties and average interval widths for each of the three CIs described above, we performed four simulation studies. The goal of the first simulation study was to determine coverage properties and average widths of the three competing CIs as p varies. The purpose of the second simulation study was to demonstrate each CI's coverage and interval-width performance as the overall sample size, N, increased. Our third simulation study provided a direct coverage contrast (as derived in Graham et al. (2003)) between our ILR CI and Rahardja and Yang's mWald CI. The fourth simulation study contrasted the coverage properties of the ILR and mWald CIs over a range of both sample sizes and values of p. The simulations were performed via the computer language R 3.6.0.

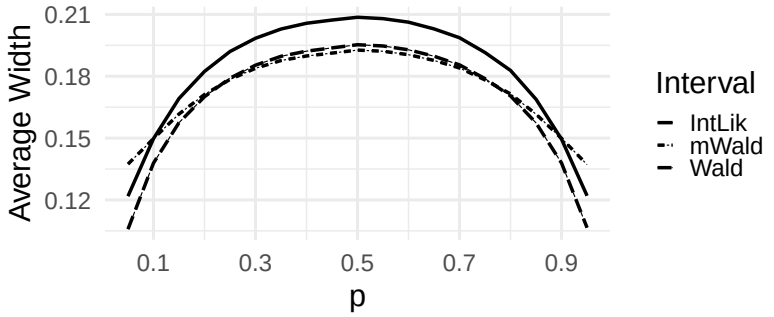
1.3.1 Simulation 1: Design and Results

We performed the first simulation study to display the coverage properties and average widths of each competing CI as the parameter of interest, p, varied. We chose our initial value of p to be 0.05 and increased it by 0.05 to 0.95. The overall sample size was fixed at $N = 400$, with $n = 40$ and $m = 360$, and the two nuisance parameters were set so that $\theta = \phi = 0.10$. We simulated the three CIs for 10,000 multinomial data sets and displayed our results in Figure 1.1.

Figure 1.1a shows that the ILR CI yielded superior coverage properties when contrasted to both the nWald and mWald CIs. In particular, we see that the mWald and nWald CIs yielded considerable under-coverage for $p < 0.20$ and $p > 0.80$. Furthermore, in Figure 1.1b, one can determine that the average width of the ILR CI was comparable to the average widths of the nWald and mWald CIs for values of p, where $0.15 < p < 0.85$. Moreover, the ILR CI had a smaller average width than the mWald CI for $p < 0.10$ and $p > 0.90$.



(a) Coverage curves as p increases from 0.05 to 0.95 when $N = 400$, $m = 360$, and $n = (1/9)m$.

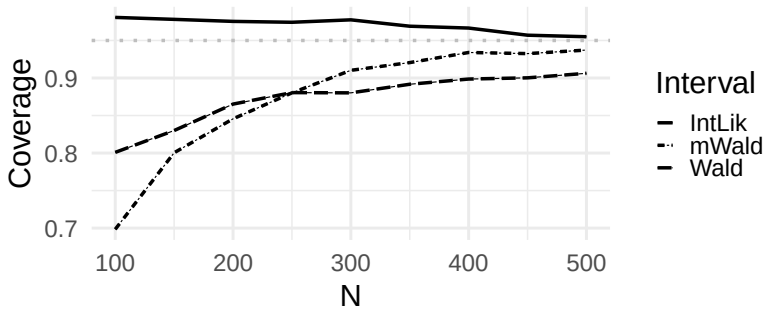


(b) Average-width curves as p increases from 0.05 to 0.95 when $N = 400$, $m = 360$, and $n = (1/9)m$.

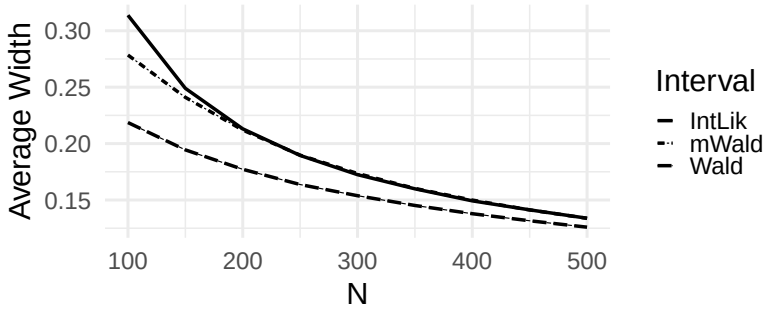
Figure 1.1. Coverage and interval-width curves for $0.05 \leq p \leq 0.95$ and $\theta = \phi = 0.10$.

1.3.2 Simulation 2: Design and Results

Our second simulation study increased N from 100 to 500 by 50 with p, θ, and ϕ fixed at 0.10. For each value of N, $n = 0.10 \times N$ and $m = 0.90 \times N$ so n increased from 10 to 50 and m increased from 90 to 450. We simulated the three CIs for 10,000 multinomial data sets. Figure 1.2 shows that as N increased, the ILR CI displayed slightly above nominal coverage. However, the rival mWald CI considerably under-covered the parameter value $p = 0.10$ when $N < 400$ and did not quite achieve nominal 0.95 coverage for $N = 500$. Also, in Figure 1.2b, note that average widths for the ILR and Wald CIs were essentially identical for $N > 150$.



(a) Coverage curves as N increases from 100 to 500 with $m = 0.90N$ and $n = (1/9)m$ with $p = \theta = \phi = 0.10$.

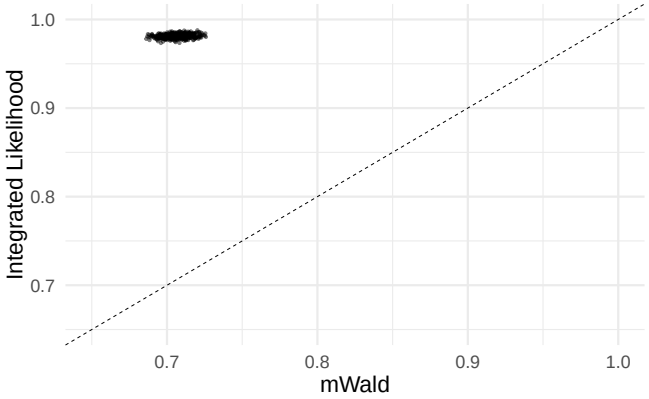


(b) Average-width curves as N increases from 100 to 500 with $m = 0.90N$ and $n = (1/9)m$ with $p = \theta = \phi = 0.10$.

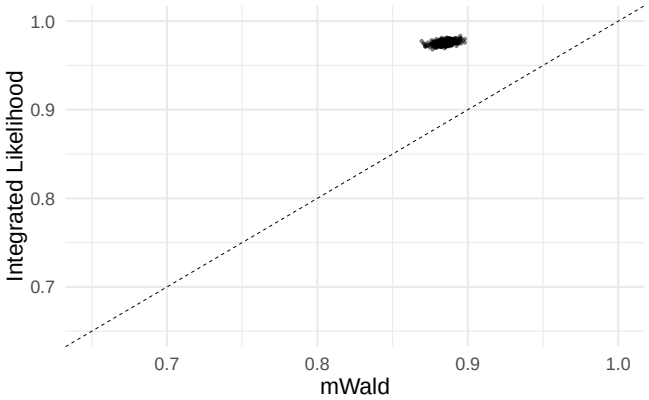
Figure 1.2. Coverage and average interval-width contrasts for $100 \leq N \leq 500$ and $p = \theta = \phi = 0.10$.

1.3.3 Simulation 3: Design and Results

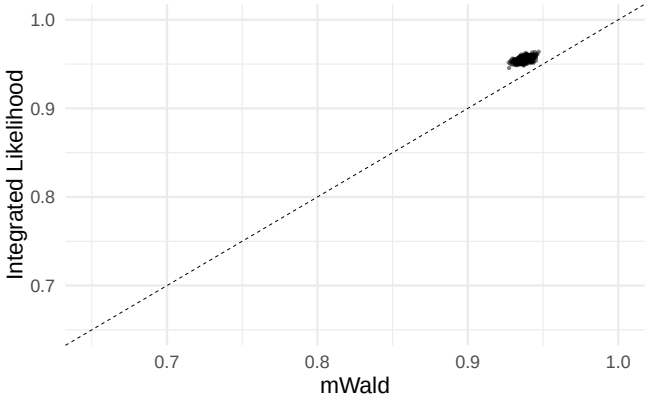
In the third simulation study, we contrasted the coverage properties of the ILR and mWald CIs for multiple simulations at three sample sizes: $N = 100$ ($n = 10,\ m = 90$), 250 ($n = 25,\ 225$), and 500 ($n = 50,\ m = 450$). More specifically, we performed Monte Carlo simulations of 1000 repetitions consisting of 5000 intervals at each value of N. In Figure 1.3, we see that the mWald CI consistently under-covers compared to the new ILR CI. Moreover, we see that for $N < 250$, the coverage of the mWald CI is well below the nominal value of 0.95. However, the ILR CI's coverage performance is slightly conservative for $N = 250$ and more conservative for $N = 100$.



(a) Coverage comparison of ILR CI and mWald CI when $N = 100$, $m = 90$, and $n = (1/9)m$.



(b) Coverage comparison of ILR CI and mWald CI when $N = 250$, $m = 225$, and $n = (1/9)m$.



(c) Coverage comparison of ILR CI and mWald CI when $N = 500$, $m = 450$, and $n = (1/9)m$.

Figure 1.3. Point cloud plots for contrast of coverage probabilities for the ILR and mWald CIs, where $N = 100,\ 250,\ 500$ and $p = \theta = \phi = 0.10$.

1.3.4 Simulation 4: Design and Results

Additionally, in Figure 1.4, we display a three-dimensional comparison of the coverage properties of the ILR and mWald CIs plotted over N and p. For every combination of N and p from $N = 100$ to $N = 500$ and from $p = 0.05$ to $p = 0.95$, respectively, the ILR CIs displayed above nominal coverage while the mWald CIs displayed below nominal coverage. Moreover, we see in Figure 1.4 that the mWald CI drastically under-covered when $p < 0.2$ and $p > 0.8$ for all sample sizes shown here, while the ILR CI is somewhat conservative for $N < 300$, especially for $p \in [0.20,\ 0.80]$.

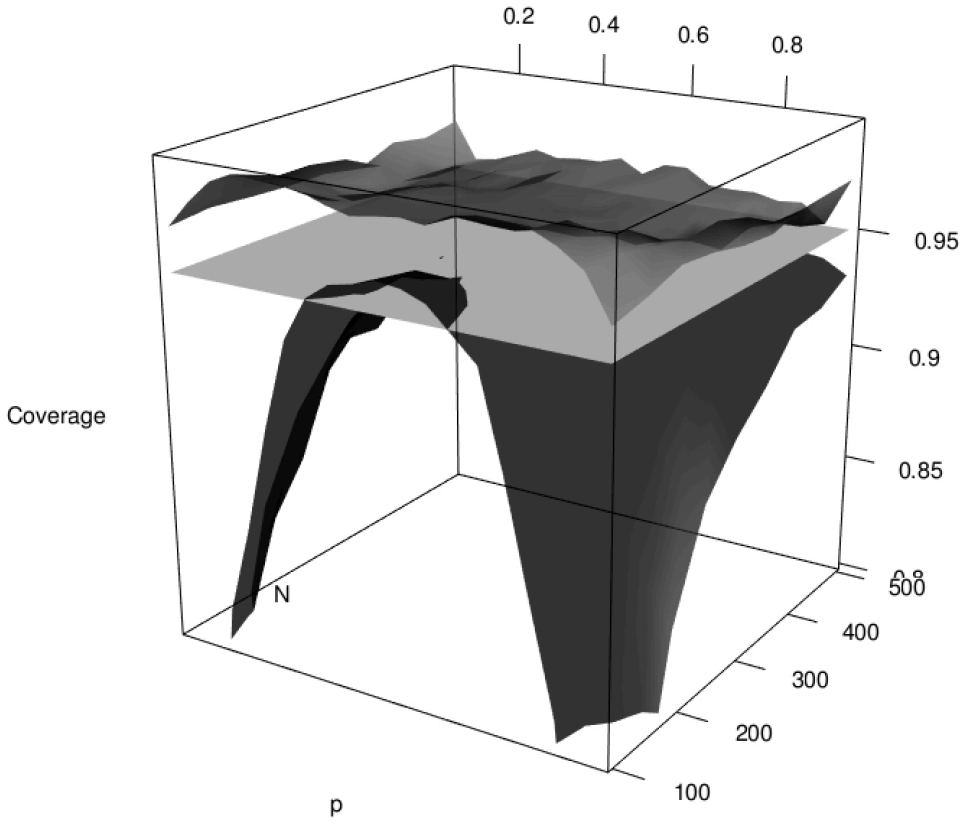


Figure 1.4. Three-dimensional plot of the ILR coverage probability (light plot) versus the mWald coverage probability (dark plot) as p varies from 0.05 to 0.95 and N varies from 100 to 500, $m = 0.90N$, $n = (1/9)m$, and $\theta = \phi = 0.10$. We include a reference nominal coverage plane of 0.95 (gray plane).

1.4 A Real-Data Example

Here, we consider the social security payment data in Raats and Moors (2003) for estimating p using double sampling with under-counted and over-counted data. Rahardja and Yang (2015) have used this data to demonstrate the efficacy of the mWald CI on real data. We computed the *MLE* $\hat{p}$ in addition to the three competing CIs: the nWald CI, mWald CI, and ILR CI. We observed that six companies were responsible for social security payments in the Netherlands. Because of the nature of this business model mistakes were possible, therefore they wanted to estimate the probability of mispayment. To estimate this parameter, an internal auditor checked a random sample of 500 social security payments, among which he found seventeen mistakes. However, because the internal auditor was also fallible, a supervising institution was hired to double check 53 of the 500 observed payments. As we can see in Table 1.3, the resulting data from the double-sampling scheme are $n_{00} = 49$, $n_{01} = 1$, $n_{10} = 1$, $n_{11} = 2$, $x = 14$, and $y = 433$. We then applied each of the three competing CI methods discussed in Section 1.3 and summarized the results, which can be found in Table 1.4.

In Table 1.4, we see that each of the CIs returned upper bounds that were close to or less than 0.10, thus suggesting that the value of p is relatively small. Based on the coverage and average interval widths examined in the above simulation studies, we have substantial evidence that the ILR CI will yield coverage slightly greater than the nominal level of 0.95. Furthermore, from the "Width" column of Table 1.4, we see that the ILR CI also yielded the smallest interval width. Hence, our proposed ILR CI yielded superior coverage and a shorter interval width than the mWald and nWald CIs derived by Rahardja and Yang (2015).

Table 1.3. Social security payment data.

		Internal Auditor		
Study	Infallible Auditor	0	1	Total
Sub	0	49	1	50
	1	1	2	3
	Total	50	3	53
Main	Not available	433	14	447

Table 1.4. CIs and Interval Widths.

Interval	Lower End Point	Upper End Point	Width
ILR CI	0.0114	0.0911	0.0797
nWald CI	0.0000	0.0870	0.0870
mWald CI	0.0141	0.1183	0.1042

1.5 Discussion

We have derived an alternative CI for the parameter p of one-sample, misclassified, binary data using a double-sampling scheme involving independent data sets produced by fallible and infallible classifiers. From our simulation results and the results from a real data set, we suggest that our ILR CI be used instead of the mWald or nWald CIs, especially when $p < 0.10$ or $p > 0.90$ for all sample sizes. Not only does the ILR CI provide greater than nominal coverage over the support of p, but it does so more efficiently for $p \notin [0.10,\ 0.90]$ by providing narrower CIs for p than both the nWald and mWald CIs.

1.6 *Appendix: A Derivation of a Closed-Form Integrated Likelihood Function for p*

Here, we derive a closed-form integrated likelihood function for a binomial parameter using double sampling. We take the weighting function to be $g(\theta, \phi) = h_1(\theta)h_2(\phi)$, where $h_i(*)$ is a Beta$(1, 1)$ density function with $i = 1, 2$.

$$
\begin{aligned}
L_I(p) &= \int_0^1 \int_0^1 L(p|\theta,\phi) g(\theta,\phi|p) d\theta d\phi \\
&= \int_0^1 \int_0^1 [(1-p)(1-\phi)]^{n_{00}} [(1-p)\phi]^{n_{01}} [p\theta]^{n_{10}} [p(1-\theta)]^{n_{11}} \pi^x (1-\pi)^y \\
&\quad \times h_1(\theta) h_2(\phi) d\theta d\phi \\
&= \int_0^1 \int_0^1 [(1-p)(1-\phi)]^{n_{00}} [(1-p)\phi]^{n_{01}} [p\theta]^{n_{10}} [p(1-\theta)]^{n_{11}} [p(1-\theta)+(1-p)\phi]^x \\
&\quad \times [1-p(1-\theta)-(1-p)\phi]^y (1)(1) d\theta d\phi \\
&= \int_0^1 \int_0^1 [(1-p)(1-\phi)]^{n_{00}} [(1-p)\phi]^{n_{01}} [p\theta]^{n_{10}} [p(1-\theta)]^{n_{11}} [p(1-\theta)+(1-p)\phi]^x \\
&\quad \times [1-p+p\theta-(1-p)\phi]^y d\theta d\phi \\
&= \int_0^1 \int_0^1 [(1-p)(1-\phi)]^{n_{00}} [(1-p)\phi]^{n_{01}} [p\theta]^{n_{10}} [p(1-\theta)]^{n_{11}} [p(1-\theta)+(1-p)\phi]^x \\
&\quad \times [(1-p)+p\theta-(1-p)\phi]^y d\theta d\phi \\
&= \int_0^1 \int_0^1 [(1-p)(1-\phi)]^{n_{00}} [(1-p)\phi]^{n_{01}} [p\theta]^{n_{10}} [p(1-\theta)]^{n_{11}} [p(1-\theta)+(1-p)\phi]^x \\
&\quad \times [p\theta+(1-p)-(1-p)\phi]^y d\theta d\phi \\
&= \int_0^1 \int_0^1 [(1-p)(1-\phi)]^{n_{00}} [(1-p)\phi]^{n_{01}} [p\theta]^{n_{10}} [p(1-\theta)]^{n_{11}} [p(1-\theta)+(1-p)\phi]^x \\
&\quad \times [p\theta+(1-p)(1-\phi)]^y d\theta d\phi \\
&= \int_0^1 \int_0^1 [(1-p)(1-\phi)]^{n_{00}} [(1-p)\phi]^{n_{01}} [p\theta]^{n_{10}} [p(1-\theta)]^{n_{11}} \\
&\quad \times \left[\sum_{i=0}^{x} \binom{x}{i} [p(1-\theta)]^i [(1-p)\phi]^{x-i} \right] \left[\sum_{j=0}^{y} \binom{y}{j} [p\theta]^j [(1-p)(1-\phi)]^{y-j} \right] d\theta d\phi \\
&= \int_0^1 \int_0^1 \sum_{i=0}^{x} \sum_{j=0}^{y} \binom{x}{i} \binom{y}{j} (1-p)^{n_{00}+n_{01}-i-j+x+y} p^{n_{10}+n_{11}+i+j} \\
&\quad \times (1-\phi)^{n_{00}+y-j} \phi^{n_{01}+x-i} \theta^{n_{10}+j} (1-\theta)^{n_{11}+i} d\theta d\phi
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=0}^{x}\sum_{j=0}^{y}\binom{x}{i}\binom{y}{j}(1-p)^{n_{00}+n_{01}-i-j+x+y}p^{n_{10}+n_{11}+i+j} \\
&\quad \times \frac{B(n_{00}+y-j+1,\ n_{01}+x-i+1)}{B(n_{00}+y-j+1,\ n_{01}+x-i+1)}\int_0^1 (1-\phi)^{n_{00}+y-j}\phi^{n_{01}+x-i}d\phi \\
&\quad \times \frac{B(n_{10}+j+1,\ n_{11}+i+1)}{B(n_{10}+j+1,\ n_{11}+i+1)}\int_0^1 \theta^{n_{10}+j}(1-\theta)^{n_{11}+i}d\theta \\
&= \sum_{i=0}^{x}\sum_{j=0}^{y}\binom{x}{i}\binom{y}{j}(1-p)^{n_{00}+n_{01}-i-j+x+y}p^{n_{10}+n_{11}+i+j} \\
&\quad \times B(n_{00}+y-j+1,\ n_{01}+x-i+1)B(n_{10}+j+1,\ n_{11}+i+1) \\
&\quad \times \int_0^1 \frac{(1-\phi)^{n_{00}+y-j}\phi^{n_{01}+x-i}}{B(n_{00}+y-j+1,\ n_{01}+x-i+1)}d\phi \\
&\quad \times \int_0^1 \frac{\theta^{n_{10}+j}(1-\theta)^{n_{11}+i}}{B(n_{10}+j+1,\ n_{11}+i+1)}d\theta \\
&= \sum_{i=0}^{x}\sum_{j=0}^{y}\binom{x}{i}\binom{y}{j}(1-p)^{n_{00}+n_{01}-i-j+x+y}p^{n_{10}+n_{11}+i+j} \\
&\quad \times B(n_{00}+y-j+1,\ n_{01}+x-i+1)B(n_{10}+j+1,\ n_{11}+i+1)(1)(1) \\
&= \sum_{i=0}^{x}\sum_{j=0}^{y}\binom{x}{i}\binom{y}{j}(1-p)^{n_{00}+n_{01}-i-j+x+y}p^{n_{10}+n_{11}+i+j} \\
&\quad \times B(n_{00}+y-j+1,\ n_{01}+x-i+1)B(n_{10}+j+1,\ n_{11}+i+1).
\end{aligned}
$$

1.7 Appendix: Attributions

Below we provide the contributions of each co-author to the published version of this chapter.

- Chris Elrod, Ph.D., contributed by aiding in creation and review of computer code needed to produce the various intervals that are used here, particularly the integrated-likelihood-ratio confidence interval.
- Phil D. Young, Ph.D., contributed by aiding in reviewing and editing.
- Dean M. Young, Ph.D., contributed by aiding in discovery and derivation of the ideas presented in this chapter as well as in reviewing and editing.

CHAPTER TWO

Coverage Correction for an Integrated-Likelihood-Ratio Confidence Interval for a Proportion from Misclassified Data using a Double-Sampling Scheme

2.1 Introduction

A major focus in statistical analysis is classification of unlabeled data. Often, researchers seek to learn enough information about their subjects to correctly classify a given individual in a category. However, rarely do observations receive perfect classification. For many studies, researchers are working with fallible classifiers that cause biased estimators because of misclassification. In fact, Bross (1954) and Hansen et al. (1960) showed that traditional estimation of the population proportion parameter in the presence of misclassification produced biased estimators that greatly affect summarizing information gathered from surveys and censuses. To account for the errors that fallible classifiers introduce, Tenenbein (1970) proposed a double-sampling scheme that employed both a fallible classifier and a gold-standard, or infallible, classifier. This sampling scheme was used by Boese et al. (2006) to develop various interval estimators for a binomial proportion parameter while Greenland (2008) produced interval estimators for the odds ratio when data are subject to misclassification. In addition, Riggs (2015) has used inverse sampling in combination with the double-sampling scheme to derive various interval estimators for a binomial parameter.

Here, as illustrated by Tenenbein (1970) and Raats and Moors (2003), we assume that we have data produced from both fallible and infallible classifiers via a double-sampling plan. The fallible data consist of a larger sample that is used for our main information source, while the infallible data set is smaller and used for validation and bias correction. These two samples are taken independently of each other.

The structure of this double-sampling scheme is displayed in Table 2.1. We allow m to be the sample size for the fallible data set, and we let n represent the sample size of the infallible data set, thus resulting in a total sample size of $N = m + n$. Our cell counts are represented as n_{ij}, where i represents the classification label of the infallible classifier and j represents the classification label of the fallible classifier. The counts x and y are the positive and negative counts, respectively, from the main study.

Table 2.1. One-sample misclassified binary data obtained using double sampling.

		Internal Auditor		
Study	Infallible Auditor	0	1	Total
Sub	0	n_{00}	n_{01}	$n_{0.}$
	1	n_{10}	n_{11}	$n_{1.}$
	Total	$n_{.0}$	$n_{.1}$	n
Main	Not available	y	x	m

Table 2.2 provides the corresponding probabilities that we use to construct the likelihood function derived from our double-sampling scheme. We label our outcomes

Table 2.2. Cell probabilities for the double-sampling multinomial distribution.

		Internal Auditor		
Study	Infallible Auditor	0	1	Total
Sub	0	$(1-p)(1-\phi)$	$(1-p)\phi$	$(1-p)$
	1	$p\theta$	$p(1-\theta)$	p
Main	Not available	$1-\pi$	π	1

for the infallible and fallible classifiers as T and F, respectively. Table 2.2 also displays data for both classifiers in the validation study, but only data for the fallible classifier in the main study. We define the parameters as follows: $p := P(T = 1)$, $\pi := P(F = 1)$, $\phi := P(F = 1|T = 0)$, and $\theta := P(F = 0|T = 1)$. The parameters ϕ and θ are the

probabilities of a false-positive and a false-negative classification, respectively. Using this information and the law of total probability, we have that $\pi = p(1-\theta)+(1-p)\phi$.

Rahardja and Yang (2015) recently derived a modified version of the classic Wald CI for a binomial parameter estimated with misclassified data to provide better coverage properties than a Wald CI while decreasing the average interval width. In addition, Wiley et al. (2021) derived an integrated-likelihood-ratio (ILR) CI that yielded better coverage properties than the mWald CI. The ILR CI also provided narrower average interval widths for extreme values of p. However, the ILR CI was consistently conservative in its coverage properties for sample sizes less than $N = 600$. Here, by following the adjustments related to that proposed by Fisher and Robbins (2019), we derive an adjusted ILR CI that provides closer-to-nominal coverage when compared to the non-adjusted ILR CI while still displaying at least nominal coverage properties. Through Monte Carlo simulations, we demonstrate the superiority of the adjusted integrated-likelihood-ratio (aILR) CI when contrasted to the ILR CI.

The remainder of this paper is organized as follows. In Section 2.2, we describe the ILR, aILR, and mWald CIs for a binomial parameter with possible misclassification in the fallible data set. In Section 2.3, we describe the design and results of our Monte Carlo simulations that examine the average interval widths and coverage properties of each CI. We contrast the ILR, aILR, and mWald CIs on real data in Section 2.4. Finally, we present a brief discussion of our results in Section 2.5.

2.2 CI Methods for Estimating p

Using the information given in the double-sampling scheme summarized in Tables 2.1 and 2.2, we obtain the likelihood function from which all CIs considered here are derived, namely

$$L(p|\theta,\phi) = [(1-p)(1-\phi)]^{n_{00}}[(1-p)\phi]^{n_{01}}[p\theta]^{n_{10}}[p(1-\theta)]^{n_{11}}\pi^{x}(1-\pi)^{y}. \qquad (2.1)$$

The goal of each CI is to estimate the binomial proportion parameter p using both fallible and infallible data sets obtained via a double-sampling scenario.

2.2.1 An Integrated-Likelihood-Ratio CI for p

Motivated by the work of Berger et al. (1999) and Severini (2010), and fully derived in Wiley et al. (2021), we present an ILR CI for a single binomial parameter p using the double-sampling scheme expressed in Tables 2.1 and 2.2. An integrated likelihood function for p is

$$L_I(p) = \int_0^1 \int_0^1 L(p, \theta, \phi) g(\theta, \phi | p) d\theta d\phi, \tag{2.2}$$

where $L(p, \theta, \phi)$ comes from (2.1) and $g(\theta, \phi|p)$ is a weighting function for θ and ϕ. Wiley et al. (2021) derived a closed-form expression for (2.2), which is

$$\begin{aligned} L_I(p) = \sum_{i=0}^{x} \sum_{j=0}^{y} \binom{x}{i} \binom{y}{j} (1-p)^{n_{00}+n_{01}-i-j+x+y} p^{n_{10}+n_{11}+i+j} \\ \times B(n_{00} + y - j + 1,\ n_{01} + x - i + 1) B(n_{10} + j + 1,\ n_{11} + i + 1), \end{aligned} \tag{2.3}$$

consisting of Beta functions, $B(a, b)$ with parameters a and b. The ILR CI is all values of p that satisfy

$$\left\{ p : -2 \log \left(\frac{L_I(p)}{L_I(\hat{p}_{IL})} \right) < \chi^2_{(1,1-\alpha)} \right\}, \tag{2.4}$$

where $\hat{p}_{IL}$ is the *MLE* of p determined numerically from the likelihood function (2.2), and $\chi^2_{1,(1-\alpha)}$ denotes the $(1-\alpha)$th quantile of the central chi-square distribution with one degree of freedom.

2.2.2 An Adjusted Integrated-Likelihood-Ratio CI for p

Wiley et al. (2021) derived an ILR CI for the binomial parameter p using data subject to misclassification and obtained from a double-sampling scenario. They demonstrated that for most values of p, their ILR CI has conservative coverage. That is, their ILR CI yields CIs whose coverage exceeds the nominal coverage probability

for almost all combinations of p and $N \leq 700$. Moreover, the ILR CI yields very conservative coverage properties when $p < 0.30$ and $N < 400$. To shrink the ILR CI interval width, we use a transformation motivated by Fisher and Robbins (2019). Let $\chi^2_{(1,1-\alpha)}$ denote the $(1-\alpha)$th quantile of the central chi-square distribution with one degree of freedom. We transform the quantile $\chi^2_{(1,1-\alpha)}$ to get

$$c_{(1-\alpha)} = n^k \left[1 - \exp(-\chi^2_{(1-\alpha)}/n^k)\right], \tag{2.5}$$

where $k \in (0, 1)$, to shorten the ILR CI interval width and, therefore, to reduce the over-coverage of the ILR CI. We use Monte Carlo simulations to choose an appropriate value of k for a fixed value of the sample size N so that the adjusted ILR CI is

$$\left\{p : -2\log\left(\frac{L_I(p)}{L_I(\hat{p}_{IL})}\right) < c_{(1-\alpha)}\right\}, \tag{2.6}$$

where $\hat{p}_{IL}$ is the maximum integrated likelihood estimator of p numerically optimized from the likelihood function (2.2), and $c_{(1-\alpha)}$ is defined in (2.5).

2.2.3 A Modified Wald CI for p

Rahardja and Yang (2015) first introduced a "naive" $100(1-\alpha)\%$ Wald CI for p by using a double sample to correct for under-reporting. Their Wald CI for p is

$$\hat{p} \pm z_{\alpha/2}\hat{\sigma}, \tag{2.7}$$

where $\hat{p} = \hat{\pi}\hat{\lambda}_1 + (1-\hat{\pi})\hat{\lambda}_2$ is the *MLE* for p of the likelihood (2.1), $z_{\alpha/2}$ is the $(\alpha/2)$th quantile of the standard normal distribution, and

$$\hat{\sigma} = \sqrt{\frac{\hat{\pi}\hat{\lambda}_1(1-\hat{\lambda}_1)}{n} + \frac{(1-\hat{\pi})\hat{\lambda}_2(1-\hat{\lambda}_2)}{n} + \frac{(\hat{\lambda}_1 - \hat{\lambda}_2)^2\hat{\pi}(1-\hat{\pi})}{N}}$$

Rahardja and Yang (2015) used a re-parameterization of (2.1) to derive $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\pi}$. Because (2.7) tends to produce CIs with bounds outside of $[0, 1]$, they proposed a modified interval to correct this issue. By using the logit transformation

$$\hat{\delta} = \text{logit}(\hat{p}) = \log\left(\frac{\hat{p}}{1-\hat{p}}\right), \tag{2.8}$$

Rahardja and Yang (2015) were able to better approximate the transformed sampling distribution of $\hat{p}$ to an approximate normal distribution. Also, using the delta method, they obtained the estimator $\hat{\tau}^2 := \widehat{\text{Var}}(\hat{\delta}) \approx \hat{\sigma}^2/[\hat{p}(1-\hat{p})]^2$, and therefore constructed an approximate $100(1-\alpha)\%$ CI for δ, which is

$$\hat{\delta} \pm z_{\alpha/2}\hat{\tau}. \tag{2.9}$$

Finally, by exponentiating the boundaries of (2.9), they obtained the CI

$$[\exp(\hat{\delta} - z_{\alpha/2}\hat{\tau})/(\exp(\hat{\delta} - z_{\alpha/2}\hat{\tau}) + 1),\ \exp(\hat{\delta} + z_{\alpha/2}\hat{\tau})/(\exp(\hat{\delta} + z_{\alpha/2}\hat{\tau}) + 1)], \tag{2.10}$$

which is the Rahardja and Yang (2015) mWald CI for p.

2.3 Monte Carlo Simulations

We performed two simulation studies to determine the CI coverage properties and average interval-width characteristics for each of the three previously described CIs: the mWald, the ILR, and the aILR. For each of the studies, we have chosen k for the aILR to be

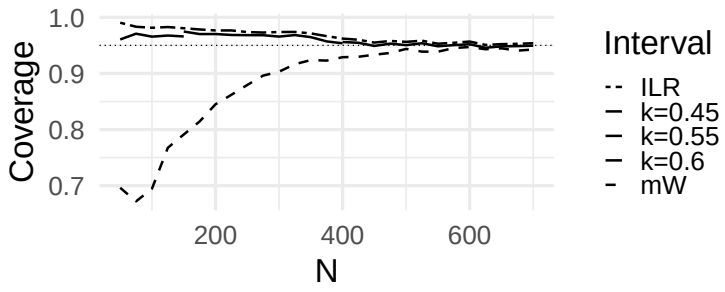
$$k = \begin{cases} 0.45 & N \leq 150, \\ 0.55 & 150 < N \leq 400, \\ 0.60 & 400 > N. \end{cases}$$

The first simulation study served to examine the coverage properties and average CI widths of the three CIs as N varied from 50 to 700 and p was fixed at $p = 0.10$, 0.30, and 0.50. The second simulation study provided the coverage properties for the three CI methods as both N and p were varied. Each simulation was performed with the computer language R.

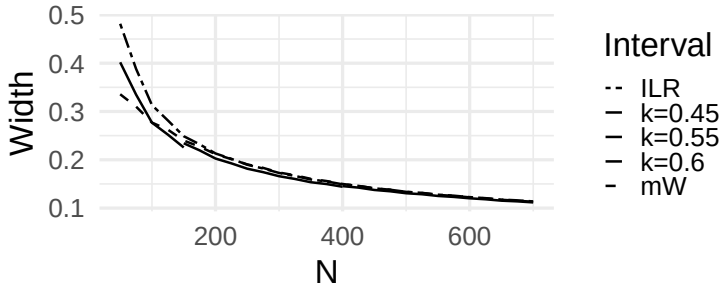
2.3.1 Simulation 1: Design and Results

The first simulation study demonstrated the coverage properties and average interval widths of the three competing CIs as the sample size, N, varied for three different fixed values of p. We began with $N = 50$ and increased N by 25 to $N = 700$

while we let $m = 0.9 \times N$ and $n = 0.1 \times N$. In addition, we set $\theta = \phi = 0.10$ and simulated the Wald, ILR, and aILR CIs for 10,000 multinomial data sets. The simulation results for $p = 0.10$, $p = 0.30$, and $p = 0.50$ are displayed in Figures 2.1, 2.2, and 2.3, respectively.



(a) Coverage curves as N increases from 50 to 700 with $m = 0.9N$ and $n = (1/9)m$ for $p = 0.10$ and $\theta = \phi = 0.10$.

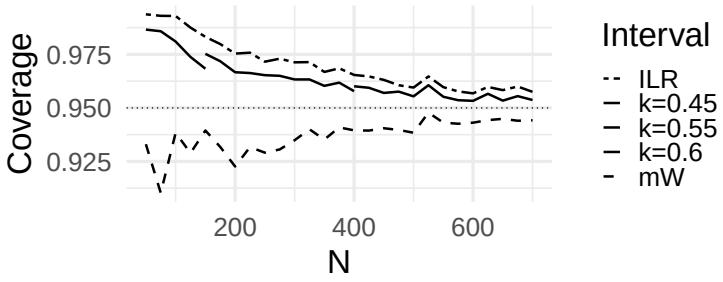


(b) Average-width curves as N increases from 50 to 700 with $m = 0.9N$ and $n = (1/9)m$ for $p = 0.10$ and $\theta = \phi = 0.10$.

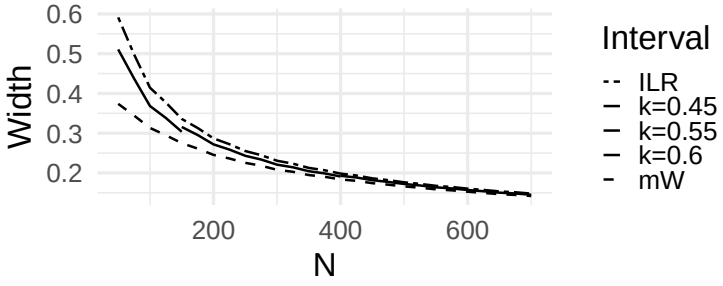
Figure 2.1. Coverage and average interval-width plots for $50 \leq N \leq 700$ when $m = 0.9N$, $n = (1/9)m$, $p = 0.10$, and $\theta = \phi = 0.10$.

Figure 2.1a provides the simulation coverage properties for the three competing CIs when $p = 0.10$ and $50 \leq N \leq 700$. We observed that all three CIs covered at least nominally or near-nominally when $N \geq 500$. However, the aILR CI was less conservative in terms of coverage than the ILR CIs for all sample sizes. Figure 2.1b also provides plots for the average interval widths and shows that the aILR CI has narrower average interval widths when $N < 400$.

Figures 2.2 and 2.3 provide very similar coverage and average interval-width results for the cases when $p = 0.30$ and $p = 0.50$, respectively. As discussed previously, in Figure 2.2a we see that the aILR CI outperformed the mWald and ILR CIs by yielding closer to nominal coverage without having less than nominal coverage for all values of N. The average width of the mWald CI is too narrow when $N \leq 700$. The aILR CI's average interval width was also moderate compared to that of the mWald and the ILR CIs, as seen in Figure 2.2b. When $p = 0.50$, Figure 2.3a shows



(a) Coverage curves as N increases from 50 to 700 with $m = 0.9N$ and $n = (1/9)m$ for $p = 0.30$ and $\theta = \phi = 0.10$.

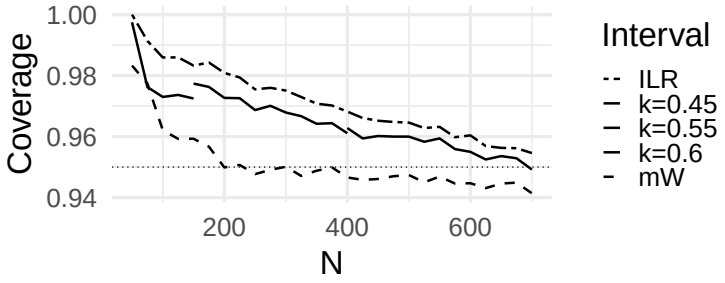


(b) Average-width curves as N increases from 50 to 700 with $m = 0.9N$ and $n = (1/9)m$ for $p = 0.30$ and $\theta = \phi = 0.10$.

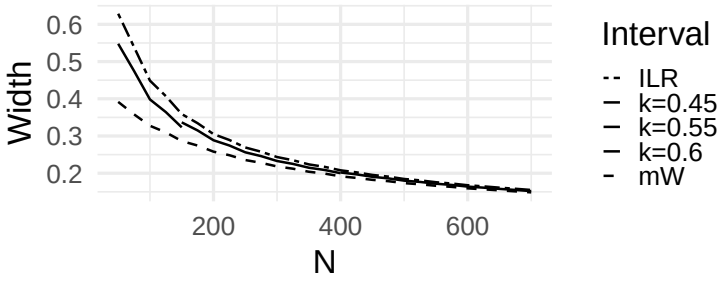
Figure 2.2. Coverage and average interval-width contrasts for $50 \leq N \leq 700$ when $p = 0.30$ and $\theta = \phi = 0.10$.

that the mWald CI demonstrated better coverage properties than the mWald CI did when $p = 0.10$. However, for all N considered here, the coverage for the mWald CI was less than nominal while the aILR CI steadily approached nominal coverage as N increased. As stated in Wiley et al. (2021), when p is very small or very large, as

is the case in Figure 2.1, the mWald CI performed very poorly in terms of coverage. However, the aILR-based CI is somewhat conservative in coverage while the ILR CI is very conservative, especially when $N < 600$. Thus, we have shown here that the conservativeness of the regular ILR CI can be reduced by the aILR CI to achieve closer-to-nominal coverage for sample sizes where $N < 700$.



(a) Coverage curves as N increases from 50 to 700 with $m = 0.9N$ and $n = (1/9)m$ for $p = 0.50$ and $\theta = \phi = 0.10$.

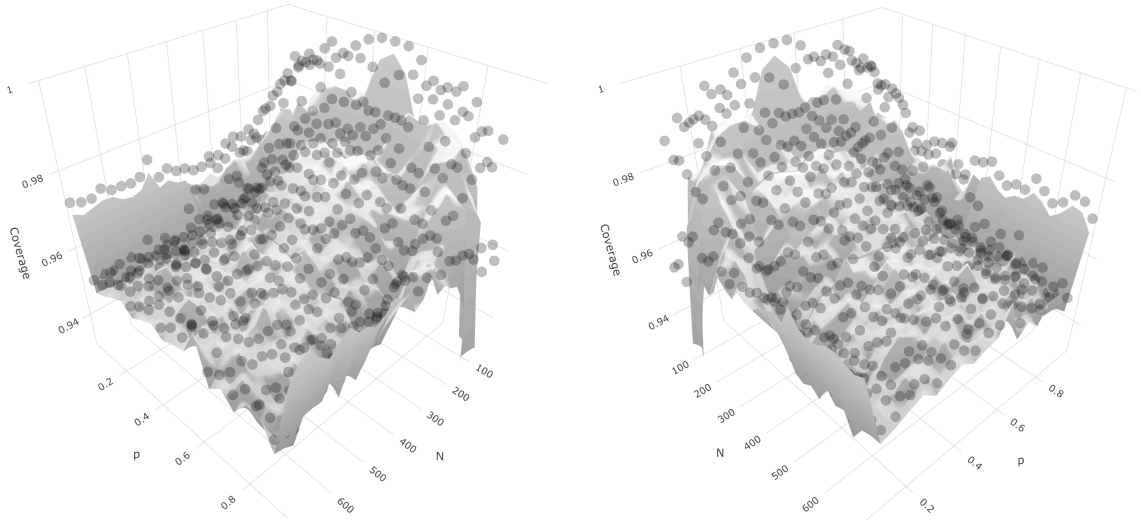


(b) Average-width curves as N increases from 50 to 700 with $m = 0.9N$ and $n = (1/9)m$ for $p = 0.50$ and $\theta = \phi = 0.10$.

Figure 2.3. Coverage and average interval-width contrasts for $50 \leq N \leq 700$ when $p = 0.50$ and $\theta = \phi = 0.10$.

2.3.2 Simulation 2: Design and Results

The second simulation study focused on the two ILR-based CIs. We varied N from 50 to 700 and p from 0.05 to 0.95 to provide a three-dimensional coverage plot for the aILR and ILR CIs and then contrast their respective coverage properties. These results appear in Figure 2.4.



(a) Perspective one.

(b) Perspective two.

Figure 2.4. Three-dimensional plot of the ILR (black points) and aILR (gray plane) CIs coverage curves for $k \in \{0.45,\ 0.55,\ 0.60\}$ as N increases from 50 to 700 with $m = 0.9N$ and p increases from 0 to 1 with $n = (1/9)m$ and $\theta = \phi = 0.10$.

First, we notice that the ILR CI coverage, represented by the points in Figures 2.4a and 2.4b, is all above the gray surface that represents the aILR CI coverage. As seen in our previous simulations, the aILR CI is consistently less conservative than the ILR CI while providing at least nominal coverage for all values of N and p, except when p is close to one. Both CIs approached nominal coverage as N increased; however, if one desires shorter CI interval widths with at least nominal coverage, the aILR CI is superior when compared to the ILR CI as an omnibus CI, regardless of the values of p and N.

2.4 A Real-Data Example

Next, we revisit the social security payment data that was introduced by Raats and Moors (2003) for estimating p using double sampling with fallible data consisting of both under-counted and over-counted data. Rahardja and Yang (2015) initially used this data to demonstrate the efficacy of the mWald CI on real data. Wiley et al. (2021) also utilized this data to demonstrate the superiority of the ILR CI coverage

and width to the mWald CI coverage and width when $p < 0.05$. Here, we computed the *MLE* $\hat{p}$ along with five competing CIs: the mWald CI, the ILR CI, and three aILR CIs with $k \in \{0.45,\ 0.55,\ 0.60\}$. The data come from six companies that were responsible for social security payments in the Netherlands. Because of the complications that come with social security payments, mis-payments are often made. Therefore, we wish to estimate the probability of issuing an incorrect payment. To estimate this proportion of mis-payments, an internal auditor checked a random sample of 500 social security payments and found 17 mistakes. Because of the fallibility of the internal auditor, a supervising institution was hired to review a subset of 53 observed payments. Table 2.3 displays the resulting counts from the double-sampling scheme, which are $n_{00} = 49,\ n_{01} = 1,\ n_{10} = 1,\ n_{11} = 2,\ x = 14,\ y = 433$. The results of applying each of the five CIs to this data are provided in Table 2.4.

Table 2.3. Social security payment data

		Internal Auditor		
Study	Infallible Auditor	0	1	Total
Sub	0	49	1	50
	1	1	2	3
	Total	50	3	53
Main	Not available	433	14	447

We see that Table 2.4 suggests that the actual value for p must be small because the upper bound of each of our CIs is at or below 0.10. From our simulation studies in Section 2.3, we know that we can achieve near-nominal coverage based on an appropriate value k. We also see in Table 2.4 that the aILR CI provides a narrower interval than the mWald and ILR CIs. Because our real data example has $N = 500$, we suggest using an aILR CI with k between 0.45 and 0.60 to determine an appropriate CI for p. This choice of k provides a narrower CI than the ILR CI, which has at least nominal coverage.

Table 2.4. CIs and Interval Widths

Interval	Lower End Point	Upper End Point	Width
aILR CI $k = 0.45$	0.0124	0.0868	0.0744
aILR CI $k = 0.55$	0.0120	0.0887	0.0768
aILR CI $k = 0.60$	0.0117	0.0898	0.0781
ILR CI	0.0114	0.0911	0.0797
mWald CI	0.0141	0.1183	0.1042

2.5 Discussion

The ILR CI, introduced by Wiley et al. (2021), for estimating the parameter p using misclassified, binary data from a double-sampling scheme can be quite conservative. Therefore, we propose an aILR CI that yields much narrower CIs but still produces close-to-nominal coverage. Through our simulation results and an application to a real data set, we determined that an aILR CI with our adjustment should be used. We also remark that the value of k that shortens the ILR CI depends on the values of N, p, ϕ and θ. Using Monte Carlo simulation, one can determine aILR CIs that have near nominal coverage while they possess narrower interval widths, thus allowing for much more efficient CIs than both the mWald and ILR CIs.

CHAPTER THREE

Integrated-Likelihood-Ratio Confidence Interval for the Log Odds-Ratio and Odds-Ratio Derived using Misclassified Data from a Double-Sampling Scenario

3.1 Introduction

In case-control studies, a common measure of interest is the odds-ratio. Szumilas (2010) explained that researchers are primarily interested in contrasting how the presence or absence of an exposure, typically a disease, affects the odds of a particular outcome. In many situations, especially in public health applications, outcomes are mislabeled. Spencer et al. (2018) showed that when error is present in the measurement of outcomes and not accounted for, then misclassification bias is introduced into the analyses' conclusions.

Currently, many statistical methods exist that aid one in dealing with misclassification bias. Greenland (1988) considered methods for the calculation of variance estimators for epidemiologic-effect estimators in epidemiological studies where data are subject to misclassification. Carroll et al. (1993) explored the estimation of parameters for logistic models when the data for a rare disease are derived from a case-control study subject to misclassification. Also, Morrissey and Spiegelman (1999) considered corrections for the odds-ratio that adjust for bias by using both the matrix and inverse matrix methods. Prescott and Garthwaite (2002) then provided a Bayesian approach to label misclassification when considering case-control studies.

In addition, Tenenbein (1970) introduced a double-sampling approach to correct for misclassification. Table 3.1 shows counts for a double-sampling plan for case-control data. Here, m_C is the total number of observations present in our main, fallible data set for sample C, where $C = 1$ represents the cases and $C = 0$ represents the controls, while n_C is the total number of observations in the infallible data set.

Table 3.1. Single-sample misclassified binary data obtained using double sampling where C represents the cases ($C = 1$) or controls ($C = 0$).

		Fallible Classifier		
Study	Infallible Classifier	$F = 0$	$F = 1$	Total
Sub	$T = 0$	n_{C00}	n_{C01}	$n_{C0\cdot}$
	$T = 1$	n_{C10}	n_{C11}	$n_{C1\cdot}$
	Total	$n_{C\cdot 0}$	$n_{C\cdot 1}$	n_C
Main	Not available	y_C	x_C	m_C

Our total sample size, therefore, is $N_C = m_C + n_C,\ C = 0, 1$. We let x_C and y_C denote the total positive and negative counts that we observe in the fallible data set, respectively. Also, we let each of the cell counts be denoted as n_{CTF}, where C represents the group the count is sampled from, T represents the label of the infallible classifier, and F represents the label of the fallible classifier. The corresponding proportions for Table 3.1 are given in Table 3.2. We define $p_C =: P(T = 1)$ as the

Table 3.2. Cell probabilities for the double-sample multinomial distribution.

		Fallible Classifier		
Study	Infallible Classifier	$F = 0$	$F = 1$	Total
Sub	$T = 0$	$(1 - p_C)(1 - \phi_C)$	$(1 - p_C)\phi_C$	$(1 - p_C)$
	$T = 1$	$p_C\theta_C$	$p_C(1 - \theta_C)$	p_C
Main	Not available	$1 - \pi_C$	π_C	1

prevalence and $\pi_C := P(F = 1)$ as the probability of the fallible classifier denoting a positive occurrence. Our false-positive and false-negative probabilities are $\phi_C := P(F = 1|T = 0)$ and $\theta_C := P(F = 0|T = 1)$, respectively.

Several confidence intervals (CIs) for the odds-ratio have been proposed. Woolf (1955) presented what is considered to be the first confidence interval for the odds-ratio, while Mantel and Haenszel (1959) provided an improvement. Haldane (1956) and Gart and Zweifel (1967) discussed methods for improving the performance of a typical Wald CI for the odds-ratio. Some of these improvements were further

explored by Agresti and Coull (1998). Cornfield (1956) first introduced a score CI for the odds-ratio, and Agresti (2011) provided various extensions. Agresti (2003) also discussed the calculation of exact CIs for the odds-ratio through application of the hypergeometric function.

Although one commonly estimates the odds-ratio in case-control studies by using CIs, few CIs exist that consider or correct for the impact of misclassification. Lyles (2002) provided a Wald CI that accounts for misclassification through "crude" estimation of the variance; however, apparently no other CIs exist for the odds-ratio that account for misclassification. Therefore, here we introduce integrated-likelihood-ratio (ILR) CIs for the log odds-ratio and the odds-ratio. We determine that the ILR CI yields better coverage properties than Lyles' Wald CI for the log odds-ratio and is comparable in average width to the Wald CI for sufficiently large samples sizes, N_0 and N_1.

The remainder of the paper is organized as follows. In Section 3.2, we derive an ILR CI and describe a Wald CI for the log odds-ratio and odds-ratio. In Section 3.3, we describe the design and results of our Monte Carlo simulations examining average interval widths and interval coverage properties of the two CIs for the log odds-ratio. We contrast the two competing CIs on real data in Section 3.4, and finally, we briefly discuss our results in Section 3.5.

3.2 Two Confidence Intervals for Estimating the Odds-Ratio

If we assume the double-sampling scheme summarized in Tables 3.1 and 3.2, we can produce the relevant likelihood function for $p_C,\ C = 0, 1$, the probability of the prevalence of the disease of interest, which is

$$\begin{aligned} L(p_C|\theta_C, \phi_C, \mathbf{d}_C) = [(1-p_C)(1-\phi_C)]^{n_{C00}}[(1-p_C)\phi_C]^{n_{C01}}[p_C\theta_C]^{n_{C10}} \\ \times [p_C(1-\theta_C)]^{n_{C11}}\pi_C^{x_C}(1-\pi_C)^{y_C}, \end{aligned} \tag{3.1}$$

where $\mathbf{d}_C := (n_{C00}, n_{C01}, n_{C10}, n_{C11}, x_C, y_C)'$ and $\pi_C = p_C(1-\theta_C) + (1-p_C)\phi_C$, $C = 0, 1$. Agresti (1999) explained that large-sample CIs based on the log odds-ratio often outperform CIs for the odds-ratio. We let

$$\gamma = \log\left[\frac{p_1/(1-p_1)}{p_0/(1-p_0)}\right] \tag{3.2}$$

be the log odds-ratio. Then,

$$\gamma = \log\left[\frac{p_1/(1-p_1)}{p_0/(1-p_0)}\right] \implies p_0 = \frac{p_1}{(1-p_1)e^\gamma + p_1},$$

and the likelihood function for the control contribution, $C = 0$, in terms of γ is

$$\begin{aligned} L(\gamma|\mathbf{\Theta}_0, \mathbf{d}_0) = &\left[\left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)(1-\phi_0)\right]^{n_{000}} \left[\left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)\phi_0\right]^{n_{001}} \\ &\times \left[\frac{p_1}{(1-p_1)e^\gamma + p_1}\theta_0\right]^{n_{010}} \left[\frac{p_1}{(1-p_1)e^\gamma + p_1}(1-\theta_0)\right]^{n_{011}} \\ &\times \left[\frac{p_1}{(1-p_1)e^\gamma + p_1}(1-\theta_0) + \left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)\phi_0\right]^{x_0} \\ &\times \left[1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}(1-\theta_0) - \left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)\phi_0\right]^{y_0}, \end{aligned} \tag{3.3}$$

where $\mathbf{\Theta}_0 := (p_1, \theta_0, \phi_0)'$ represents the parameter vector for $C = 0$. Therefore, the full likelihood function for γ for our two-sample problem is

$$\begin{aligned} L(\gamma|\mathbf{\Theta}, \mathbf{d}) &= L(p_1|\theta_1, \phi_1) \times L(\gamma|p_1, \theta_0, \phi_0) \\ &= [(1-p_1)(1-\phi_1)]^{n_{100}}[(1-p_1)\phi_1]^{n_{101}}[p_1\theta_1]^{n_{110}}[p_1(1-\theta_1)]^{n_{111}} \\ &\quad \times [p_1(1-\theta_1) + (1-p_1)\phi_1]^{x_1}[1 - p_1(1-\theta_1) - (1-p_1)\phi_1]^{y_1} \\ &\quad \times \left[\left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)(1-\phi_0)\right]^{n_{000}} \left[\left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)\phi_0\right]^{n_{001}} \\ &\quad \times \left[\frac{p_1}{(1-p_1)e^\gamma + p_1}\theta_0\right]^{n_{010}} \left[\frac{p_1}{(1-p_1)e^\gamma + p_1}(1-\theta_0)\right]^{n_{011}} \\ &\quad \times \left[\frac{p_1}{(1-p_1)e^\gamma + p_1}(1-\theta_0) + \left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)\phi_0\right]^{x_0} \\ &\quad \times \left[1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}(1-\theta_0) - \left(1 - \frac{p_1}{(1-p_1)e^\gamma + p_1}\right)\phi_0\right]^{y_0}, \end{aligned} \tag{3.4}$$

where $\boldsymbol{\Theta} := (p_1, \theta_0, \phi_0, \theta_1, \phi_1)'$ represents our full parameter vector for the transformed likelihood function and $\mathbf{d} := (\mathbf{d}_0, \mathbf{d}_1)'$. Using the likelihood function (3.4), we can construct an ILR CI for γ, given in (3.2), and $\psi = \frac{p_1/(1-p_1)}{p_0/(1-p_0)}$. Here, we are interested in deriving an ILR CI for γ and contrasting the coverage and average interval widths to those of the Wald CI for γ.

3.2.1 An Integrated-Likelihood-Ratio CI for γ

Motivated by Berger et al. (1999) and Severini (2010), we construct an ILR CI to estimate the log odds-ratio using both fallible and infallible data from a double-sampling method like the one described by Tenenbein (1970). An integrated likelihood function for γ is

$$\begin{aligned} L_I(\gamma) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 \int_0^1 & L(\gamma | p_1, \theta_1, \theta_0, \phi_1, \phi_0) \\ & \times g_1(\theta_1, \phi_1 | p_1) g_0(\theta_0, \phi_0 | p_0) dp_1 d\theta_1 d\theta_0 d\phi_1 d\phi_0, \end{aligned} \tag{3.5}$$

where θ_C and ϕ_C, $C = 0, 1$, are the false negative and false positive proportions, respectively, for each of our two populations. In addition, $g_C(\theta_C, \phi_C | p_C)$, $C = 0, 1$ are weighting functions for the two sets of misclassification parameters. In Section 3.6, we have derived a closed-form expression for (3.5), which is

$$\begin{aligned} L_I(\gamma) = & \sum_{i=0}^{x_1} \sum_{j=0}^{y_1} \sum_{k=0}^{x_0} \sum_{l=0}^{y_0} \binom{x_1}{i} \binom{y_1}{j} \binom{x_0}{k} \binom{y_0}{l} [e^\gamma]^{-n_{010} - n_{011} - k - l} \\ & \times B(y_1 + n_{100} - j + 1,\ x_1 + n_{101} - i + 1) \\ & \times B(y_0 + n_{000} - l + 1,\ x_0 + n_{001} - k + 1) \\ & \times B(n_{110} + j + 1,\ n_{111} + i + 1) B(n_{010} + l + 1,\ n_{011} + k + 1) \\ & \times \Gamma(n_{100} + n_{101} + n_{000} + n_{001} + x_1 + y_1 + x_0 + y_0 - i - j - k - l + 1) \\ & \times \Gamma(n_{110} + n_{111} + n_{010} + n_{011} + i + j + k + l + 1) \\ & \times {}_2F_1(n_{110} + n_{111} + n_{010} + n_{011} + i + j + k + l + 1,\ N_0,\ 2 + N_1 + N_0; \\ & \qquad 1 - \cosh(\gamma) + \sinh(\gamma)), \end{aligned} \tag{3.6}$$

where $B(a,b)$ denotes a Beta function with parameters a and b, and $\Gamma(a)$ denotes a Gamma function with parameter a. Also, ${}_2F_1(a,b,c;z)$ denotes the regularized hypergeometric function defined in Clyde et al. (2011) and Hankin (2015), as

$${}_2F_1(a,\ b,\ c;\ z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}dt.$$

The calculation of (3.6) is computationally demanding. Therefore, we recommend numerical integration of (3.5) for calculating the ILR CI for γ, which is

$$\left\{\gamma : -2\log\left(\frac{L_I(\gamma)}{L_I(\hat{\gamma}_{IL})}\right) < \chi^2_{(1,1-\alpha)}\right\}, \tag{3.7}$$

where $\hat{\gamma}_{IL}$, is the *MLE* of γ derived from the integrated likelihood function $L_I(\gamma)$ found by numerical optimization, and $\chi^2_{(1,1-\alpha)}$ denotes the $(1-\alpha)$th quantile of a central chi-square distribution with one degree of freedom. This ILR CI for γ can be transformed into a CI for the odds-ratio $\psi = \frac{p_1/(1-p_1)}{p_0/(1-p_0)}$ through exponentiation of the lower and upper bounds.

3.2.2 A Wald-type CI for γ

Lyles (2002) introduced a Wald CI for estimating crude odds-ratios in case-control studies. Lyles' Wald CI for γ is

$$\hat{\gamma} \pm z_{\alpha/2}\hat{\sigma}^2\left\{\hat{\gamma}\right\}, \tag{3.8}$$

where

$$\begin{aligned}\hat{\sigma}^2\left\{\hat{\gamma}\right\} = \sum_{C=0}^{1}\left\{\hat{p}_C(1-\hat{p}_C)\right\}^{-2}\Bigg[&\left(\widehat{PPV}_C + \widehat{NPV}_C - 1\right)^2\hat{\sigma}^2\left\{\hat{\pi}_C\right\} \\ &+ \left(\hat{\pi}_C\right)^2\hat{\sigma}^2\left\{\widehat{PPV}_C\right\} + \left(1-\hat{\pi}_C\right)^2\hat{\sigma}^2\left\{\widehat{NPV}_C\right\}\Bigg],\end{aligned} \tag{3.9}$$

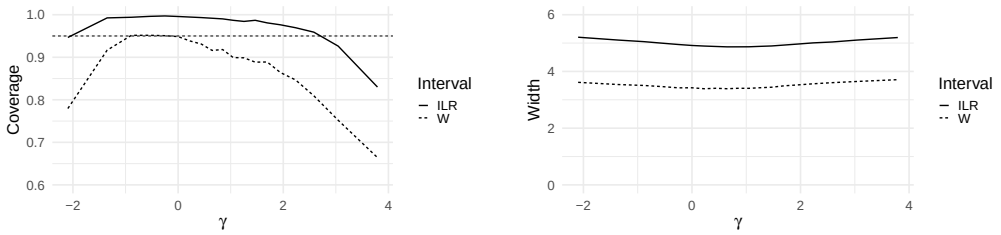
with

$$\begin{aligned}
\hat{p}_C &= \hat{\pi}_C\widehat{PPV}_C + (1-\hat{\pi}_C)(1-\widehat{NPV}_C), \\
\hat{\pi}_C &= (x_C + n_{C.1})/N_C, \qquad \hat{\sigma}^2\left\{\hat{\pi}_C\right\} = \hat{\pi}_C(1-\hat{\pi}_C)/N_C, \\
\widehat{PPV}_C &= n_{C11}/n_{C.1}, \qquad \hat{\sigma}^2\left\{\widehat{PPV}_C\right\} = \widehat{PPV}_C(1-\widehat{PPV}_C)/n_{C.1}, \\
\widehat{NPV}_C &= n_{C00}/n_{C.0}, \qquad \hat{\sigma}^2\left\{\widehat{NPV}_C\right\} = \widehat{NPV}_C(1-\widehat{NPV}_C)/n_{C.0}.
\end{aligned}$$

3.3 Monte Carlo Simulations

We performed a simulation study to compare and contrast the coverage properties and average interval widths of the ILR CI and the Wald CI where γ was allowed to vary. For the simulation, we chose to vary p_1 from 0.05 to 0.95 while holding p_0 constant at 0.30. We simulated data sets for five different sample sizes $N_i \in \{50,\ 100,\ 150,\ 200,\ 250\},\ i = 0, 1$ with $N_0 = N_1$, $m_i = 0.9 \times N_i$, and $n_i = 0.1 \times N_i$ for $i \in \{0,\ 1\}$. Finally, our misclassification parameters were set as $\phi_0 = 0.20$, $\theta_0 = 0.25$, $\phi_1 = 0.10$, and $\theta_1 = 0.15$. We simulated 10,000 multinomial data sets for each unique sample size, $N_i,\ i = 0, 1$. All calculations were performed via the computer language R version 3.6.1.

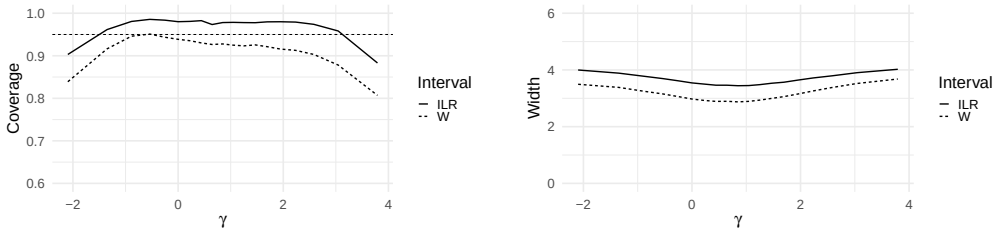
Figure 3.1 provides us with the simulation results when $N_0 = N_1 = 50$. We see a large difference between the coverage properties for the two CIs. Figure 3.1b shows that the Wald CI was considerably narrower. However, we see the ramifications of this fact in Figure 3.1a in that the narrower Wald CI has less than nominal coverage for most values of γ considered here. We can see that the Wald CI attained nominal coverage or slightly above nominal coverage when $p_0 \approx p_1$, but otherwise, the Wald CI coverage was mostly less than nominal and was very poor for $\gamma > 1$. However, the ILR CI attained above nominal coverage for most values of γ but dipped below nominal coverage when p_1 was considerably greater than p_0.



(a) Coverage curves (b) Average-width curves

Figure 3.1. Coverage and average-interval-width curves for $0.05 \leq p_1 \leq 0.95$ when $N_0 = N_1 = 50$, $m_1 = 45$, $n_1 = 5$, $\phi_1 = 0.10$, $\theta_1 = 0.15$, $p_0 = 0.30$, $\phi_0 = 0.20$, and $\theta_0 = 0.25$.

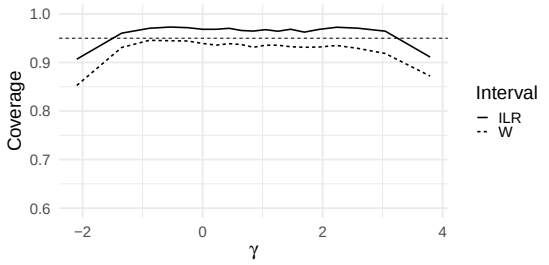
In Figure 3.2 we see that the increase in sample sizes from $N_0 = N_1 = 50$ to $N_0 = N_1 = 100$ caused an increase in the coverage of the Wald CI, shown in Figure 3.2a, and a decrease in the average-width of the ILR CI, shown in Figure 3.2b. Though the Wald CI coverage increased, it still provided mostly less than nominal coverage, while the ILR CI gave conservative coverage and produced closer to nominal coverage. We also see that the average CI widths for the two CIs began to converge, so the issue of producing a relatively wider ILR CI quickly dissipated as the sample size was increased.



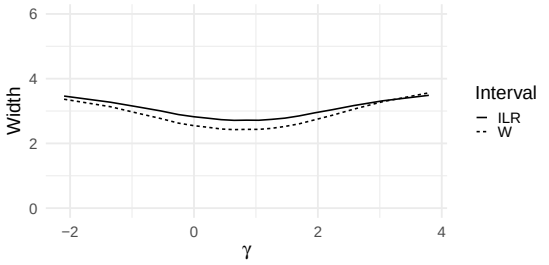
(a) Coverage curves (b) Average-width curves

Figure 3.2. Coverage and average-interval-width curves for $0.05 \leq p_1 \leq 0.95$ when $N_0 = N_1 = 100$, $m_1 = 90$, $n_1 = 10$, $\phi_1 = 0.10$, $\theta_1 = 0.15$, $p_0 = 0.30$, $\phi_0 = 0.20$, and $\theta_0 = 0.25$.

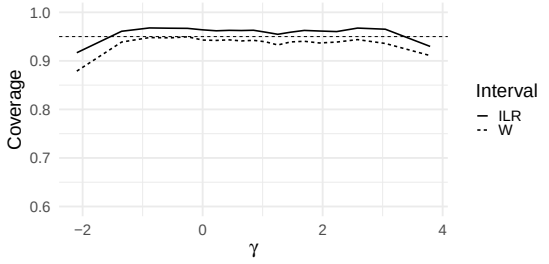
Figure 3.3 displays the simulation results for $N_i \in \{150,\ 200,\ 250\}$, $i = 0, 1$. When the sample sizes were increased to $N_0 = N_1 = 150$, Figures 3.3a and 3.3b show that the coverages and average interval widths of the Wald and ILR CIs become more similar. This fact is further supported through the sample size increases shown in Figures 3.3c, 3.3b, 3.3e and 3.3f. Also, we see in each of these instances that the ILR CI provided slightly above nominal coverage, falling below 95% only when p_1 deviated largely from p_0. The Wald CI more often provided slightly below nominal coverage. More interestingly, we see that as the sample sizes increased, the average interval widths of both the Wald and ILR CIs decreased, but the ILR CI provided narrower CIs for extreme γ values and the Wald CI was slightly narrower than the



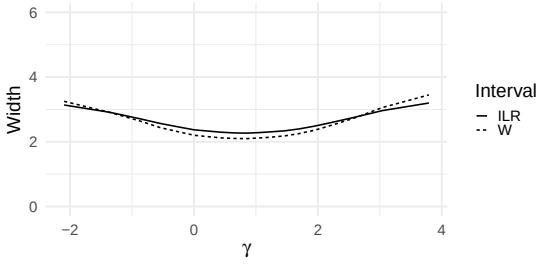
(a) Coverage curves for $N_0 = N_1 = 150$



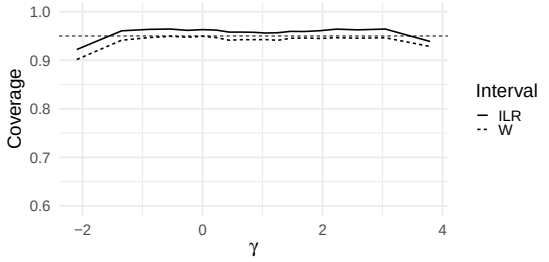
(b) Average-width curves for $N_0 = N_1 = 150$



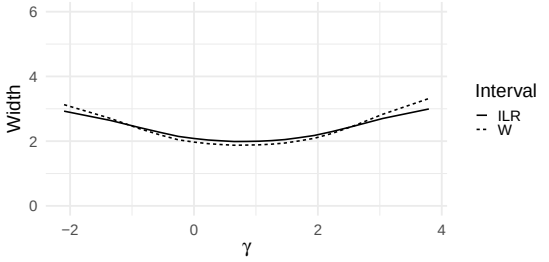
(c) Coverage curves for $N_0 = N_1 = 200$



(d) Average-width curves for $N_0 = N_1 = 200$



(e) Coverage curves for $N_0 = N_1 = 250$



(f) Average-width curves for $N_0 = N_1 = 250$

Figure 3.3. Coverage and average-interval-width curves for $0.05 \leq p_1 \leq 0.95$ when $\phi_1 = 0.10,\ \theta_1 = 0.15,\ p_0 = 0.30,\ \phi_0 = 0.20,$ and $\theta_0 = 0.25$, with $m_i = 0.9N_i$ and $n_i = (1/9)m,\ i = 0, 1$.

ILR CI for moderate values of γ. This result is especially evident in Figures 3.3d and 3.3f.

3.4 A Real-Data Example

Here, we compare and contrast the performance of the ILR and Wald CIs on a simulated sample version of a real-data example. Our motivation comes from the SIDS data set provided by Drews et al. (1990). As stated in Greenland (2008), we use the false-positive rate $\hat{\phi}_1 = 0.13$, sensitivity $\widehat{SE}_1 = 0.63$, and positive predictive-

value $\widehat{PPV}_1 = P(T = 1|F = 1) = 0.57$ to estimate the probability of death by SIDS, which was $\hat{p} = P(T = 1) = 0.2184$. We followed a similar procedure to estimate these probabilities for the control group. The resulting estimated probabilities are provided in Table 3.3.

Table 3.3. Probabilities for example where $N_C = 500$.

Case		Control	
$\hat{p}_1$	0.2148	$\hat{p}_0$	0.1792
$\hat{\phi}_1$	0.1300	$\hat{\phi}_0$	0.0700
$\hat{\theta}_1$	0.3700	$\hat{\theta}_0$	0.4300

We then used the information provided in Table 3.3 to simulate a data set to display the effectiveness of the two CIs for estimating ψ when our data were grouped by these probabilities and when $N_C = 500,\ C = 0, 1$. These data are provided in Table 3.4. Next, we then used the data in Table 3.4 to construct Wald and ILR CIs

Table 3.4. Cell counts for example where $N_C = 500$.

Study	$C = 1$	$F = 0$	$F = 1$	$C = 0$	$F = 0$	$F = 1$
Sub	$T = 0$	67	8	$T = 0$	78	6
	$T = 1$	14	11	$T = 1$	5	11
Main		311	89		329	71

for the odds of death by SIDS in this mock case-control study. The MLE for ψ was $\hat{\psi} = \frac{\hat{p}_1(1-\hat{p}_0)}{\hat{p}_0(1-\hat{p}_1)} = 1.253$. We then calculated the ILR and Wald CIs given in Table 3.5. We see that the two CIs are comparable in both location and interval width. The ILR CI is slightly more conservative and, thus, is slightly wider in width than the Wald CI for ψ. We conclude that, based on our results in Section 3.3, one should adopt the ILR CI in this situation and similar ones where $N \geq 200$ and $p_0 \not\approx p_1$ because it should more consistently capture ψ while remaining comparable in interval width.

Table 3.5. CIs and Interval Widths

Interval	Lower End Point	Upper End Point	Width
Wald	0.972	3.342	2.370
ILR	1.002	3.466	2.464

3.5 Discussion

We have derived a new CI for the log odds-ratio and odds-ratio parameters γ and ψ, respectively, when dealing with data consisting of two sample misclassified binary observations that are obtained via a double-sampling scheme involving independent data sets produced by fallible and infallible classifiers. After reviewing our results from the simulations in Section 3.3 and the real data set in Section 3.4, we recommend the use of our new ILR CI rather than the Wald CI proposed by Lyles (2002) because of its coverage properties. We have shown that the ILR CI for γ not only has conservative coverage properties, but also has very comparable average interval widths when compared to the Wald CI for γ when $p_1 \in [0.10,\ 0.90]$ and $N_0 = N_1 \geq 250$.

3.6 Appendix: A Derivation of a Closed-Form Integrated Likelihood Function for γ

Here, we derive a closed-form integrated likelihood function for the log odds-ratio of two binomial parameters using double-sampling paradigm. We use the weighting functions $g_i(\theta_i, \phi_i) = h_{i1}(\theta_i)h_{i2}(\theta_i)$, where $h_{ij}(*)$ is a Beta$(1,1)$ density function with $i = 0, 1$ and $j = 1, 2$.

$$
\begin{aligned}
L_I(\gamma) &= \int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 L(\gamma|p_1,\theta_1,\theta_0,\phi_1,\phi_0)g_1(\theta_1,\phi_1|p_1)g_0(\theta_0,\phi_0|p_0)dp_1d\theta_1d\theta_0d\phi_1d\phi_0 \\
&= \int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 [(1-p_1)(1-\phi_1)]^{n_{100}}[(1-p_1)\phi_1]^{n_{101}}[p_1\theta_1]^{n_{110}}[p_1(1-\theta_1)]^{n_{111}} \\
&\quad \times [p_1(1-\theta_1)+(1-p_1)\phi_1]^{x_1}[1-p_1(1-\theta_1)-(1-p_1)\phi_1]^{y_1} \\
&\quad \times \left[\left(1-\frac{p_1}{(1-p_1)e^\gamma+p_1}\right)(1-\phi_0)\right]^{n_{000}} \left[\left(1-\frac{p_1}{(1-p_1)e^\gamma+p_1}\right)\phi_0\right]^{n_{001}} \\
&\quad \times \left[\frac{p_1}{(1-p_1)e^\gamma+p_1}\theta_0\right]^{n_{010}} \left[\frac{p_1}{(1-p_1)e^\gamma+p_1}(1-\theta_0)\right]^{n_{011}} \\
&\quad \times \left[\frac{p_1}{(1-p_1)e^\gamma+p_1}(1-\theta_0)+\left(1-\frac{p_1}{(1-p_1)e^\gamma+p_1}\right)\phi_0\right]^{x_0} \\
&\quad \times \left[1-\frac{p_1}{(1-p_1)e^\gamma+p_1}(1-\theta_0)-\left(1-\frac{p_1}{(1-p_1)e^\gamma+p_1}\right)\phi_0\right]^{y_0} \\
&\quad dp_1d\theta_1d\theta_0d\phi_1\phi_0 \\
&= \int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 (1-p_1)^{n_{100}+n_{101}}p_1^{n_{110}+n_{111}} \\
&\quad \times \left(1-\frac{p_1}{(1-p_1)e^\gamma+p_1}\right)^{n_{000}+n_{001}} \frac{p_1}{(1-p_1)e^\gamma+p_1}^{n_{010}+n_{011}} \\
&\quad \times (1-\phi_1)^{n_{100}}\phi_1^{n_{101}}(1-\phi_0)^{n_{000}}\phi_0^{n_{001}}\theta_1^{n_{110}}(1-\theta_1)^{n_{111}}\theta_0^{n_{010}}(1-\theta_0)^{n_{011}} \\
&\quad \times [p_1(1-\theta_1)+(1-p_1)\phi_1]^{x_1}[1-p_1+p_1\theta_1-\phi_1+p_1\phi_1]^{y_1} \\
&\quad \times \left[\frac{p_1}{(1-p_1)e^\gamma+p_1}(1-\theta_0)+\left(1-\frac{p_1}{(1-p_1)e^\gamma+p_1}\right)\phi_0\right]^{x_0} \\
&\quad \times \left[1-\frac{p_1}{(1-p_1)e^\gamma+p_1}+\frac{p_1\theta_0}{(1-p_1)e^\gamma+p_1}-\phi_0+\frac{p_1\phi_0}{(1-p_1)e^\gamma+p_1}\right]^{y_0} \\
&\quad dp_1d\theta_1d\theta_0d\phi_1\phi_0 \\
&= \int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 (1-p_1)^{n_{100}+n_{101}}p_1^{n_{110}+n_{111}}p_1^{n_{010}+n_{011}} \\
&\quad \times (1-\phi_1)^{n_{100}}\phi_1^{n_{101}}(1-\phi_0)^{n_{000}}\phi_0^{n_{001}}\theta_1^{n_{110}}(1-\theta_1)^{n_{111}}\theta_0^{n_{010}}(1-\theta_0)^{n_{011}}
\end{aligned}
$$

$$
\begin{aligned}
&\times ((1-p_1)e^{\gamma}+p_1-p_1)^{n_{000}+n_{001}}((1-p_1)e^{\gamma}+p_1)^{-(n_{000}+n_{001}+n_{010}+n_{011}+x_0+y_0)} \\
&\times [p_1(1-\theta_1)+(1-p_1)\phi_1]^{x_1}[(1-p_1)(1-\phi_1)+p_1\theta_1]^{y_1} \\
&\times [p_1(1-\theta_0)+((1-p_1)e^{\gamma}+p_1-p_1)\phi_0]^{x_0} \\
&\times [(1-p_1)e^{\gamma}+p_1-p_1+p_1\theta_0-((1-p_1)e^{\gamma}+p_1)\phi_0+p_1\phi_0]^{y_0} \\
&\quad dp_1 d\theta_1 d\theta_0 d\phi_1 \phi_0 \\
=&\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 (1-p_1)^{n_{100}+n_{101}} p_1^{n_{110}+n_{111}}((1-p_1)e^{\gamma})^{n_{000}+n_{001}} p_1^{n_{010}+n_{011}} \\
&\times (1-\phi_1)^{n_{100}}\phi_1^{n_{101}}(1-\phi_0)^{n_{000}}\phi_0^{n_{001}}\theta_1^{n_{110}}(1-\theta_1)^{n_{111}}\theta_0^{n_{010}}(1-\theta_0)^{n_{011}} \\
&\times ((1-p_1)e^{\gamma}+p_1)^{-N_0} \\
&\times [p_1(1-\theta_1)+(1-p_1)\phi_1]^{x_1}[(1-p_1)(1-\phi_1)+p_1\theta_1]^{y_1} \\
&\times [p_1(1-\theta_0)+(1-p_1)e^{\gamma}\phi_0]^{x_0} \\
&\times [(1-p_1)e^{\gamma}+p_1\theta_0-(1-p_1)e^{\gamma}\phi_0]^{y_0} dp_1 d\theta_1 d\theta_0 d\phi_1 \phi_0 \\
=&\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 (1-p_1)^{n_{100}+n_{101}+n_{000}+n_{001}} p_1^{n_{110}+n_{111}+n_{010}+n_{011}} \\
&\times (1-\phi_1)^{n_{100}}\phi_1^{n_{101}}(1-\phi_0)^{n_{000}}\phi_0^{n_{001}}\theta_1^{n_{110}}(1-\theta_1)^{n_{111}}\theta_0^{n_{010}}(1-\theta_0)^{n_{011}} \\
&\times [p_1(1-\theta_1)+(1-p_1)\phi_1]^{x_1}[(1-p_1)(1-\phi_1)+p_1\theta_1]^{y_1} \\
&\times [p_1(1-\theta_0)+(1-p_1)e^{\gamma}\phi_0]^{x_0}[(1-p_1)e^{\gamma}(1-\phi_0)+p_1\theta_0]^{y_0} \\
&\times ((1-p_1)e^{\gamma}+p_1)^{-N_0}\,[e^{\gamma}]^{n_{000}+n_{001}}\, dp_1 d\theta_1 d\theta_0 d\phi_1 \phi_0 \\
=&\int_0^1\int_0^1\int_0^1\int_0^1\int_0^1 (1-p_1)^{n_{100}+n_{101}+n_{000}+n_{001}} p_1^{n_{110}+n_{111}+n_{010}+n_{011}} \\
&\times (1-\phi_1)^{n_{100}}\phi_1^{n_{101}}(1-\phi_0)^{n_{000}}\phi_0^{n_{001}}\theta_1^{n_{110}}(1-\theta_1)^{n_{111}}\theta_0^{n_{010}}(1-\theta_0)^{n_{011}} \\
&\times \sum_{i=0}^{x_1}\binom{x_1}{i} p_1^i(1-\theta_1)^i(1-p_1)^{x_1-i}\phi_1^{x_1-i} \\
&\times \sum_{j=0}^{y_1}\binom{y_1}{j}(1-p_1)^{y_1-j}(1-\phi_1)^{y_1-j}p_1^j\theta_1^j \\
&\times \sum_{k=0}^{x_0}\binom{x_0}{k} p_1^k(1-\theta_0)^k(1-p)^{x_0-k}\phi_0^{x_0-k}\,[e^{\gamma}]^{x_0-k}
\end{aligned}
$$

$$
\begin{aligned}
&\times \sum_{l=0}^{y_0} \binom{y_0}{l} (1-p_1)^{y_0-l} (1-\phi_0)^{y_0-l} \left[e^{\gamma}\right]^{y_0-l} p_1^l \theta_0^l \\
&\times \left((1-p_1)e^{\gamma} + p_1\right)^{-N_0} \left[e^{\gamma}\right]^{n_{000}+n_{001}} dp_1 d\theta_1 d\theta_0 d\phi_1 \phi_0 \\
= &\sum_{i=0}^{x_1} \sum_{j=0}^{y_1} \sum_{k=0}^{x_0} \sum_{l=0}^{y_0} \int_0^1 \binom{x_1}{i} \binom{y_1}{j} \binom{x_0}{k} \binom{y_0}{l} \left[e^{\gamma}\right]^{n_{000}+n_{001}+x_0+y_0-k-l} \\
&\times \int_0^1 (1-\phi_1)^{n_{100}+y_1-j} \phi_1^{n_{101}+x_1-i} d\phi_1 \int_0^1 (1-\phi_0)^{n_{000}+y_0-l} \phi_0^{n_{001}+x_0-k} d\phi_0 \\
&\times \int_0^1 \theta_1^{n_{110}+j} (1-\theta_1)^{n_{111}+i} d\theta_1 \int_0^1 \theta_0^{n_{010}+l} (1-\theta_0)^{n_{011}+k} d\theta_0 \\
&\times (1-p_1)^{n_{100}+n_{101}+n_{000}+n_{001}+x_1+y_1+x_0+y_0-i-j-k-l} \\
&\times p_1^{n_{110}+n_{111}+n_{010}+n_{011}+i+j+k+l} \left((1-p_1)e^{\gamma} + p_1\right)^{-N_0} dp_1 \\
= &\sum_{i=0}^{x_1} \sum_{j=0}^{y_1} \sum_{k=0}^{x_0} \sum_{l=0}^{y_0} \binom{x_1}{i} \binom{y_1}{j} \binom{x_0}{k} \binom{y_0}{l} \left[e^{\gamma}\right]^{n_{000}+n_{001}+x_0+y_0-k-l} \\
&\times B(y_1 + n_{100} - j + 1,\ x_1 + n_{101} - i + 1) \\
&\times B(y_0 + n_{000} - l + 1,\ x_0 + n_{001} - k + 1) \\
&\times B(n_{110} + j + 1,\ n_{111} + i + 1) B(n_{010} + l + 1,\ n_{011} + k + 1) \\
&\times \int_0^1 (1-p_1)^{n_{100}+n_{101}+n_{000}+n_{001}+x_1+y_1+x_0+y_0-i-j-k-l} \\
&\times p_1^{n_{110}+n_{111}+n_{010}+n_{011}+i+j+k+l} \left((1-p_1)e^{\gamma} + p_1\right)^{-N_0} dp_1 \\
= &\sum_{i=0}^{x_1} \sum_{j=0}^{y_1} \sum_{k=0}^{x_0} \sum_{l=0}^{y_0} \binom{x_1}{i} \binom{y_1}{j} \binom{x_0}{k} \binom{y_0}{l} \left[e^{\gamma}\right]^{n_{000}+n_{001}+x_0+y_0-k-l} \\
&\times B(y_1 + n_{100} - j + 1,\ x_1 + n_{101} - i + 1) \\
&\times B(y_0 + n_{000} - l + 1,\ x_0 + n_{001} - k + 1) \\
&\times B(n_{110} + j + 1,\ n_{111} + i + 1) B(n_{010} + l + 1,\ n_{011} + k + 1) \\
&\times {}_2F_1(n_{110} + n_{111} + n_{010} + n_{011} + i + j + k + l + 1,\ N_0,\ 2 + N_1 + N_0; \\
&\qquad 1 - \cosh(\gamma) + \sinh(\gamma)) \left[e^{\gamma}\right]^{-N_0} \\
&\times \Gamma(n_{100} + n_{101} + n_{000} + n_{001} + x_1 + y_1 + x_0 + y_0 - i - j - k - l + 1) \\
&\times \Gamma(n_{110} + n_{111} + n_{010} + n_{011} + i + j + k + l + 1)
\end{aligned}
$$

$$= \sum_{i=0}^{x_1}\sum_{j=0}^{y_1}\sum_{k=0}^{x_0}\sum_{l=0}^{y_0}\binom{x_1}{i}\binom{y_1}{j}\binom{x_0}{k}\binom{y_0}{l}\,[e^{\gamma}]^{-n_{010}-n_{011}-k-l}$$

$$\times B(y_1 + n_{100} - j + 1,\ x_1 + n_{101} - i + 1)$$

$$\times B(y_0 + n_{000} - l + 1,\ x_0 + n_{001} - k + 1)$$

$$\times B(n_{110} + j + 1,\ n_{111} + i + 1)B(n_{010} + l + 1,\ n_{011} + k + 1)$$

$$\times \Gamma(n_{100} + n_{101} + n_{000} + n_{001} + x_1 + y_1 + x_0 + y_0 - i - j - k - l + 1)$$

$$\times \Gamma(n_{110} + n_{111} + n_{010} + n_{011} + i + j + k + l + 1)$$

$$\times {}_2F_1(n_{110} + n_{111} + n_{010} + n_{011} + i + j + k + l + 1,\ N_0,\ 2 + N_1 + N_0;$$

$$1 - \cosh(\gamma) + \sinh(\gamma)).$$

CHAPTER FOUR

Confidence Intervals for the Ratio of Two Poisson Rate Parameters with Under-Reported Data Using a Double-Sampling Scenario

4.1 Introduction

Misclassified data is relatively common in research problems in many disciplines. For instance, Fujisawa and Izumi (2000) discussed the importance of managing misclassification in quality control for repeated measures in a laboratory and manufacturing setting. Also, Ji et al. (2006) discussed the effects of misclassification among phenotypes and genotypes on the power of tests that detect genetic association. In addition, DiBartolomeo and Witkowski (1997) examined the costs of misclassification due to ignored external factors involved in building mutual funds. Often, the researcher is focused on correcting misclassification bias in point estimators. One can find examples of this type in Tenenbein (1970), Whittemore and Gong (1991), Viana (1994), and Joseph et al. (1995). Though point estimators are useful in answering proposed research questions, interval estimators can provide more information for guiding the actions and conclusions of decision makers, as explained by Altman (2005) and Wasserstein and Lazar (2016).

Less research has been published concerning interval estimators for data subject to misclassificaiton. A contrast of various confidence intervals (CIs) for a one-population binomial parameter with under-reported successes was performed by Boese et al. (2006), and a two-population binomial model with under-reporting was proposed by Lyles (2002). Here, we consider a Poisson model with under-reported counts first introduced by Sposto et al. (1992) that focuses on two Poisson rate parameters. Closed-form maximum likelihood estimators for this Poisson model were developed by Stamey et al. (2005b), where the data were observed via a double-sampling scheme.

This scheme, introduced by Tenenbein (1970), is a standard technique for correcting misclassification where information observed by both fallible and infallible samples is combined to improve the quality and efficiency of the estimators of interest.

Using this Poisson model and a double-sampling scheme, Riggs et al. (2009) derived maximum-likelihood-based CIs for an individual Poisson rate parameter while accounting for the misclassification in the individual counts. Li (2009) then derived CIs for the difference between the two Poisson rate parameters of interest. In both papers, the authors derived CIs by inverting the Wald, score, and profile log-likelihood statistics. Here, we wish to construct a new CI for the ratio of the two Poisson rate parameters with misclassified data based on an integrated-likelihood-ratio (ILR) statistic. This ILR CI is motivated by the work in Berger et al. (1999) and Severini (2010).

We have organized the remainder of the paper as follows. In Section 4.2, we describe the Wald, score, and ILR CIs that we examine and contrast here. In Section 4.3, we describe the design and results of Monte Carlo simulations examining average interval widths and coverage properties. We then compare and contrast the three CIs on real data in Section 4.4. Finally, we briefly discuss the utility of the proposed ILR CI in Section 4.5.

4.2 Confidence Intervals for Estimating the Ratio of Poisson Rate Parameters

We begin by discussing the double-sampling based model that was first employed by Sposto et al. (1992) and further explored by Stamey et al. (2005a), Stamey et al. (2005b), and Riggs et al. (2009). For this model, we let z_i denote the observed counts from our fallible classifier for population i, we let m_i denote the unobserved correctly classified counts in population i from our fallible classifier, and we let y_i denote the unobserved mislabeled counts in population i that are said to belong to population j from our fallible classifier, where $i,\ j = 1,\ 2, i \neq j$ and $z_i = m_i - y_i + y_j$.

These counts are modeled as

$$z_i \sim Poisson(N[\lambda_i(1-\theta_i) + \lambda_j\theta_j]),$$

$$m_i \sim Poisson(N\lambda_i),$$

and

$$y_i|m_i \sim binomial(m_i, \theta_i),$$

where N is the sample size of our error-prone sample. Though N is referred to as the sample size, it is often given as person-years, machine-time, etc.

Because of the misclassified counts affecting z_i, the model is over parameterized, and the observed data allows us to estimate only the parameter $\mu = N[\lambda_i(1-\theta_i) + \lambda_j\theta_j]$. Therefore, we use a training sample to estimate the parameters λ_i and θ_i. This additional sample yields m_{0i}, the count of correctly classified individuals in our sample obtained from an infallible classifier, and y_{0i}, the count of incorrectly classified individuals obtained from the infallible classifier. The infallibly labeled sample size is denoted as N_0. The observed data are modeled as

$$z_i \sim Poisson(N[\lambda_i(1-\theta_i) + \lambda_j\theta_j]),$$

$$m_{0i} \sim Poisson(N_0\lambda_i),$$

and

$$y_{0i}|m_{0i} \sim binomial(m_{0i}, \theta_i),$$

where $i,\ j = 1,\ 2, i \neq j$. The corresponding likelihood function is

$$\begin{aligned} L(\mathbf{\Theta}|\mathbf{d}) = \lambda_1^{m_{01}}\lambda_2^{m_{02}}e^{-N_0(\lambda_1+\lambda_2)}\theta_1^{y_{01}}(1-\theta_1)^{m_{01}-y_{01}}\theta_2^{y_{02}}(1-\theta_2)^{m_{02}-y_{02}} \\ \times [\lambda_1(1-\theta_1) + \lambda_2\theta_2]^{z_1}[\lambda_2(1-\theta_2) + \lambda_1\theta_1]^{z_2}e^{-N(\lambda_1+\lambda_2)}, \end{aligned} \tag{4.1}$$

where $\mathbf{\Theta} := (\lambda_1, \lambda_2, \theta_1, \theta_2)'$ and $\mathbf{d} := (z_1, z_2, m_{01}, m_{02}, y_{01}, y_{02})'$. However, because we are interested in the ratio $\phi := \lambda_1/\lambda_2$ we perform the reparameterization

$$\phi = \frac{\lambda_1}{\lambda_2} \implies \lambda_1 = \phi\lambda_2$$

so that λ_1 is expressed in terms of ϕ. This transformation changes the vector of parameters to $\mathbf{\Theta}^* := (\phi, \lambda_2, \theta_1, \theta_2)'$, and we obtain the transformed likelihood function

$$\begin{aligned} L(\mathbf{\Theta}^*|\mathbf{d}) = \phi^{m_{01}} \lambda_2^{m_{01}+m_{02}+z_1+z_2} e^{-\lambda_2(\phi+1)(N_0+N)} [\phi(1-\theta_1)+\theta_2]^{z_1} [(1-\theta_2)+\phi\theta_1]^{z_2} \\ \times \theta_1^{y_{01}} (1-\theta_1)^{m_{01}-y_{01}} \theta_2^{y_{02}} (1-\theta_2)^{m_{02}-y_{02}}. \end{aligned} \tag{4.2}$$

Thus, the log-likelihood function is

$$\begin{aligned} \ell(\mathbf{\Theta}^*|\mathbf{d}) &= \log(L(\mathbf{\Theta}^*|\mathbf{d})) \\ &= m_{01}\log(\phi) + (m_{01}+m_{02}+z_1+z_2)\log(\lambda_2) - \lambda_2(\phi+1)(N_0+N) \\ &\quad + z_1\log[\phi(1-\theta_1)+\theta_2] + z_2\log[(1-\theta_2)+\phi\theta_1] + y_{01}\log(\theta_1) \\ &\quad + (m_{01}-y_{01})\log(1-\theta_1) + y_{02}\log(\theta_2) + (m_{02}-y_{02})\log(1-\theta_2). \end{aligned} \tag{4.3}$$

From (4.3), we derive the Hessian and the Fisher's information matrices,

$$\mathbf{H} = \begin{bmatrix} h_\phi & h_{\phi\lambda_2} & h_{\phi\theta_1} & h_{\phi\theta_2} \\ h_{\lambda_2\phi} & h_{\lambda_2} & h_{\lambda_2\theta_1} & h_{\lambda_2\theta_2} \\ h_{\theta_1\phi} & h_{\theta_1\lambda_2} & h_{\theta_1} & h_{\theta_1\theta_2} \\ h_{\theta_2\phi} & h_{\theta_2\lambda_2} & h_{\theta_2\theta_1} & h_{\theta_2} \end{bmatrix} \quad \text{and} \quad \mathbf{I}(\mathbf{\Theta}^*) = \begin{bmatrix} i_\phi & i_{\phi\lambda_2} & i_{\phi\theta_1} & i_{\phi\theta_2} \\ i_{\lambda_2\phi} & i_{\lambda_2} & i_{\lambda_2\theta_1} & i_{\lambda_2\theta_2} \\ i_{\theta_1\phi} & i_{\theta_1\lambda_2} & i_{\theta_1} & i_{\theta_1\theta_2} \\ i_{\theta_2\phi} & i_{\theta_2\lambda_2} & i_{\theta_2\theta_1} & i_{\theta_2} \end{bmatrix},$$

for the parameter vector $\mathbf{\Theta}^*$, which one needs to obtain the Wald and score CIs for ϕ. One can find these derivations for $\mathbf{H}$ and $\mathbf{I}(\mathbf{\Theta}^*)$ in Section 4.6 and Section 4.7, respectively.

4.2.1 A Wald CI for ϕ

A Wald CI for each of our Poisson rate parameters $\lambda_i,\ i = 1, 2$, was derived by Riggs et al. (2009). To calculate a Wald CI for ϕ, one needs the unrestricted maximum likelihood estimates for each of our parameters in $\mathbf{\Theta}^* = (\phi, \lambda_2, \theta_1, \theta_2)'$, which are represented as $\hat{\mathbf{\Theta}}^* = (\hat{\phi}, \hat{\lambda}_2, \hat{\theta}_1, \hat{\theta}_2)'$. Stamey et al. (2005b) have shown that

$$\hat{\lambda}_i = \frac{z_i + m_{0i} + \frac{y_{0i}z_j}{z_{0j}} - \frac{y_{0j}z_i}{z_{0i}}}{N + N_0} \tag{4.4}$$

and

$$\hat{\theta}_i = \frac{y_{0i} z_{0i}(z_j + z_{0j})}{y_{0i} z_{0i}(z_j + z_{0j}) + (m_{0i} - y_{0i}) z_{0j}(z_i + z_{0i})}, \tag{4.5}$$

where $z_{0i} = m_{0i} - y_{0i} + y_{0j}$ and $i,\ j = 1,\ 2, i \neq j$. Because of the invariance property of *MLE's*, we have $\hat{\phi} = \hat{\lambda}_1 / \hat{\lambda}_2$.

The Wald CI is based on the fact that when n is sufficiently large, $\hat{\phi} \dot{\sim} N(\phi, I^{11}(\mathbf{\Theta}^*))$, where $I^{11}(\mathbf{\Theta}^*)$ is the $(1,1)$ element of $[\mathbf{I}(\mathbf{\Theta}^*)]^{-1}$. Thus, Riggs et al. (2009) inverted the statistic $Z = \frac{\hat{\phi} - \phi}{\sqrt{I^{11}(\mathbf{\Theta}^*)}} \sim N(0,1)$ to determine the Wald CI for ϕ, which is

$$\hat{\phi} \pm z_{\alpha/2} \sqrt{I^{11}(\mathbf{\Theta}^*)}, \tag{4.6}$$

where $z_{\alpha/2}$ is the $(1 - \frac{\alpha}{2})$th quantile of the standard normal distribution. We remark that this Wald CI is somewhat crude at times, providing bounds that are outside of the support of ϕ; that is, for small values of ϕ, the Wald CI can have a lower bound that is less than zero.

4.2.2 A Score CI for ϕ

Riggs et al. (2009) and Li (2009) derived a score CI using the double-sampling procedure described in Section 4.2. An important difference when one is finding the score CI compared to the Wald CI is that one needs $\hat{\mathbf{\Theta}}^*_\phi = (\hat{\lambda}_{2\phi}, \hat{\theta}_{1\phi}, \hat{\theta}_{2\phi})'$, the restricted *MLE's*, for a given ϕ. Riggs et al. (2009) and Li (2009) both utilized the EM algorithm to calculate these restricted *MLE's*. Here, for a specific data set, we use optimization and root-finding methods available in the computer software R to calculate these values numerically. To construct the score CI, one determines the score function, $u_\phi\left(\hat{\mathbf{\Theta}}^*_\phi\right)$, which is the first derivative of the log-likelihood in (4.3) taken with respect to ϕ. Then one evaluates this function with the restricted *MLE's*.

Using the fact that the score statistic has the approximate distribution

$$[u_\phi(\hat{\mathbf{\Theta}}^*_\phi)]^2 I^{11}(\hat{\mathbf{\Theta}}^*_\phi) \dot{\sim} \chi^2_1,$$

where χ^2_1 is the central chi-square distribution with one degree of freedom, we obtain the score CI

$$\left\{\phi : [u_\phi(\hat{\boldsymbol{\Theta}}^*_\phi)]^2 I^{11}(\hat{\boldsymbol{\Theta}}^*_\phi) \leq \chi^2_{(1,1-\alpha)}\right\},$$

where $\chi^2_{(1,1-\alpha)}$ is the $(1-\alpha)$th quantile of the central chi-square distribution with one degree of freedom.

4.2.3 An Integrated-Likelihood-Ratio CI for ϕ

We next consider the integrated-likelihood-ratio (ILR) CI for ϕ, the ratio of two Poisson rate parameters. The motivation for this CI comes from Berger et al. (1999) and Severini (2010) and we use the likelihood function (4.2) to determine the integrated-likelihood function

$$L_I(\phi) = \int_0^1 \int_0^1 \int_0^\infty L(\boldsymbol{\Theta}^*|\mathbf{d}) f(\lambda_2, \theta_1, \theta_2|\phi) d\lambda_2 d\theta_1 d\theta_2. \tag{4.7}$$

Here, we use $f(\lambda_2, \theta_1, \theta_2|\phi) = g(\lambda_2)h_1(\theta_1)h_2(\theta_2)$ as a weighting function for λ_2, θ_1, and θ_2, respectively. We let $g(\lambda_2)$ be a Gamma(0.001, 0.001) density function and let $h_i(*)$ be a Beta(1, 1) density function for $i = 1,\ 2$. In Section 4.8, we have derived a closed-form for (4.7), which is

$$\begin{aligned} L_I(\phi) = {} & \frac{1}{[(\phi+1)(N_0+N)+0.001]^{m_{01}+m_{02}+z_1+z_2+0.001}} \sum_{i=0}^{z_1}\sum_{j=0}^{z_2} \binom{z_1}{i}\binom{z_2}{j} \phi^{m_{01}+i+j} \\ & \times B\left(y_{01}+j+1,\ m_{01}-y_{01}+i+1\right) \\ & \times B\left(y_{02}+z_1-i+1,\ m_{02}-y_{02}+z_2-j+1\right), \end{aligned} \tag{4.8}$$

where $B(a, b)$ denotes a Beta function with parameters a and b. We recommend the use of (4.8) because of its decreased computational complexity contrasted to (4.7).The ILR CI is

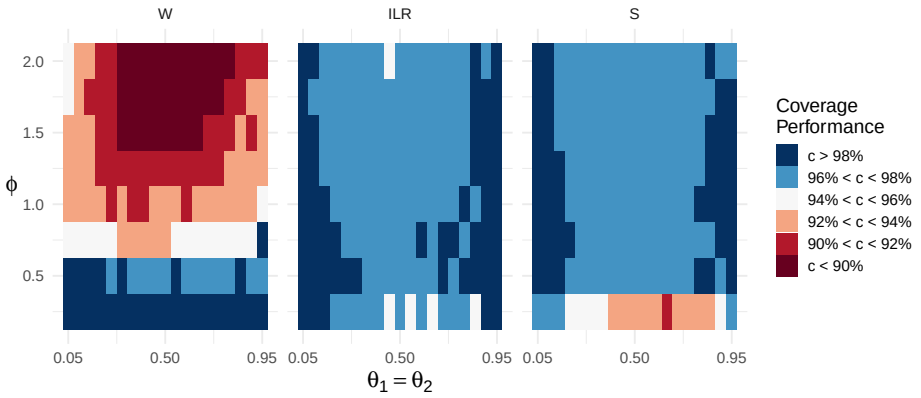
$$\left\{\phi : -2\log\left(\frac{L_I(\phi)}{L_I(\hat{\phi}_{IL})}\right) < \chi^2_{(1,1-\alpha)}\right\}, \tag{4.9}$$

where we find $\hat{\phi}_{IL}$ by maximizing $L_I(\phi)$ over the parameter space of ϕ, and $\chi^2_{(1,1-\alpha)}$ denotes the $(1-\alpha)$th quantile of a central chi-square distribution with one degree of freedom.

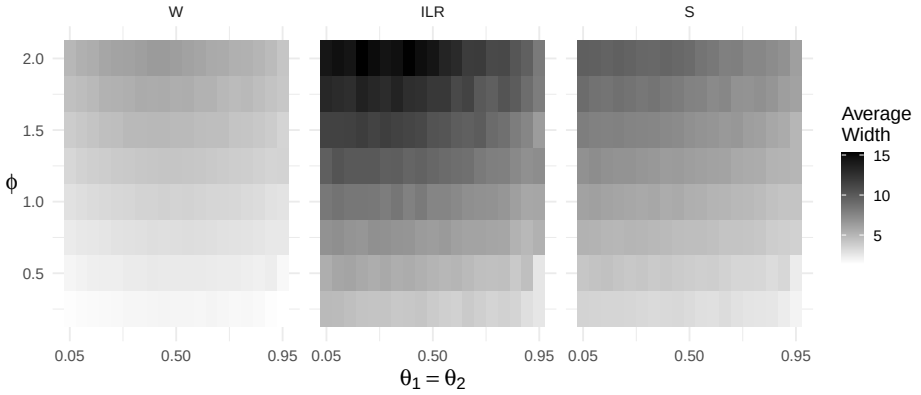
4.3 Monte Carlo Simulations

To compare and contrast the coverage properties and interval widths of the Wald, score, and ILR CIs, we performed a simulation study for various sample sizes, N_0 and N. We performed seven separate simulations with the (N_0, N) pairs $(1,2)$, $(1,3)$, $(2,4)$, $(2,6)$, $(3,6)$, $(3,9)$, and $(4,8)$. For each simulation, we let λ_1 vary from 1 to 8 by 1 while λ_2 was fixed at 4; therefore, we let ϕ vary from 0.25 to 2.00 by 0.25. Also, we assumed $\theta_1 = \theta_2$, which were allowed to vary from 0.05 to 0.95 by 0.05. Thus, we examined 152 unique combinations of $\phi, \theta_1,$ and θ_2 at each separate sample-size pair (N_0, N). We generated 10,000 simulated data sets for each parameter-sample-size combination using the computer language R version 3.6.2.

In this first simulation, because of missing data issues, particularly with the Wald and score intervals, we added 0.50 to the counts m_{01}, m_{02}, y_{01}, or y_{02} when those counts were zero. Figure 4.1 contains the coverage and interval-width results for the simulation where $N_0 = 1$ and $N = 2$. In particular, one can view the coverage results for this simulation in Figure 4.1a and observe that the three CIs rarely obtained approximately nominal coverage. The Wald CI consistently under-covered for larger values of ϕ but over-covered when $\phi < 0.75$. We see that the score and ILR CIs had similar coverage properties. In particular, the score CI under-covered whenever $\phi < 0.25$ and $\theta_1 = \theta_2$ were marginal. In Figure 4.1b, we display the simulation results for the interval widths. The Wald CI, on average, produced the narrowest CIs, as one might expect, and the score CI had moderate interval widths while the ILR CI produced the widest intervals. However, in the situations where the ILR CI gave the widest interval width, we see in Figure 4.1a that the ILR CI over-covered



(a) Heat map of coverage properties for Wald, ILR, and score CIs



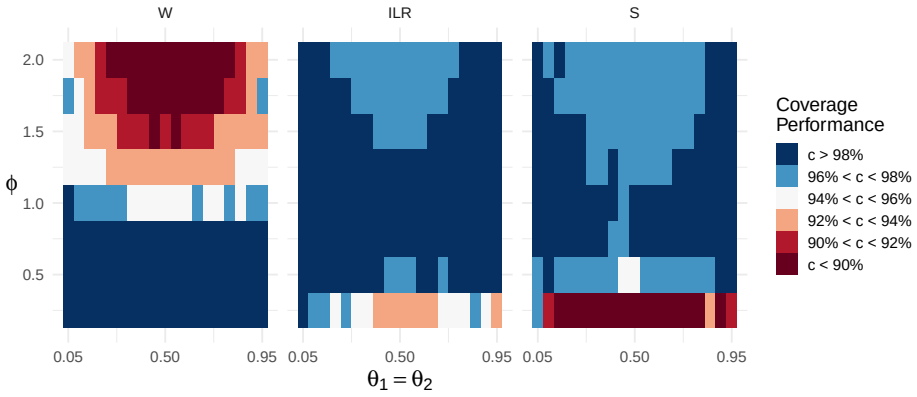
(b) Heat map of average interval widths for Wald, ILR, and score CIs

Figure 4.1. Coverage and average interval-width heat maps for the ratio of two Poisson rate parameters for $N_0 = 1$, $N = 2$, $0.25 \leq \phi \leq 2.00$, and $0.05 \leq \theta_1, \theta_2 \leq 0.95$.

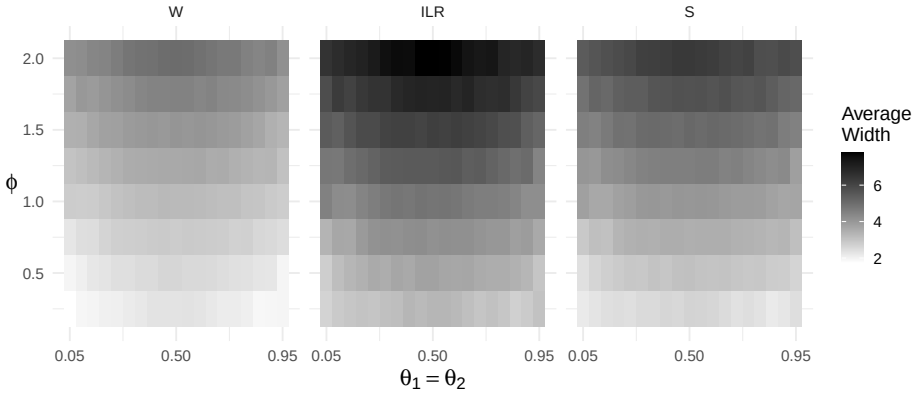
only slightly more than the score CI and had better coverage properties for several parameter configurations, especially for the case where $\phi = 0.25$.

Figure 4.2 shows the coverage and average interval-width simulation results for $N_0 = 1$ and $N = 3$. Here, we see much of the same behavior for the Wald CI as in Figure 4.1, that is, large over-coverage when ϕ was small for considerable under-coverage as ϕ approached 2 with $0.20 \leq \theta_i \leq 0.80$, $\theta_i = 1, 2$. The ILR CI's behavior showed mostly slight over-coverage for values of ϕ, where $0.50 \leq \phi \leq 1.25$. However, among the three CIs, the ILR demonstrated the best coverage for $\phi = 0.25$. The score CI also demonstrated slight over-coverage but slightly outperformed the ILR

for moderate values of $\theta_1 = \theta_2$ and ϕ when $\phi \geq 1.25$. The average interval widths shown in Figure 4.2b are shorter than those shown in Figure 4.1b. However, similar patterns were observed; that is, the Wald CI provided the narrowest intervals while the ILR and score CIs were comparable in average interval width with the ILR CI typically being slightly wider.



(a) Heat map of coverage properties for Wald, ILR, and score CIs



(b) Heat map of average interval widths for Wald, ILR, and score CIs

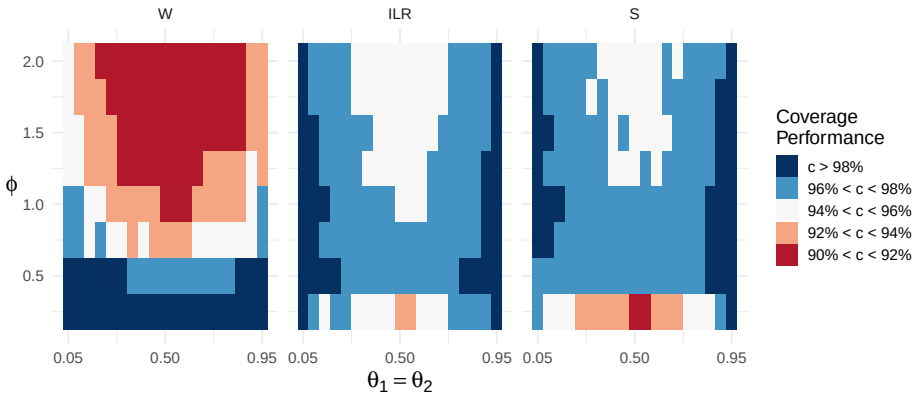
Figure 4.2. Coverage and average interval-width heat maps for the ratio of two Poisson rate parameters for $N_0 = 1$, $N = 3$, $0.25 \leq \phi \leq 2.00$, and $0.05 \leq \theta_1, \theta_2 \leq 0.95$.

In Figure 4.3, we find the coverage and average interval-width results for $N_0 = 2$ and $N = 4$. Here, all three CIs began to show evidence of convergence to the nominal coverage. The Wald CI still over-covered when $\phi \leq 0.50$ and under-covered

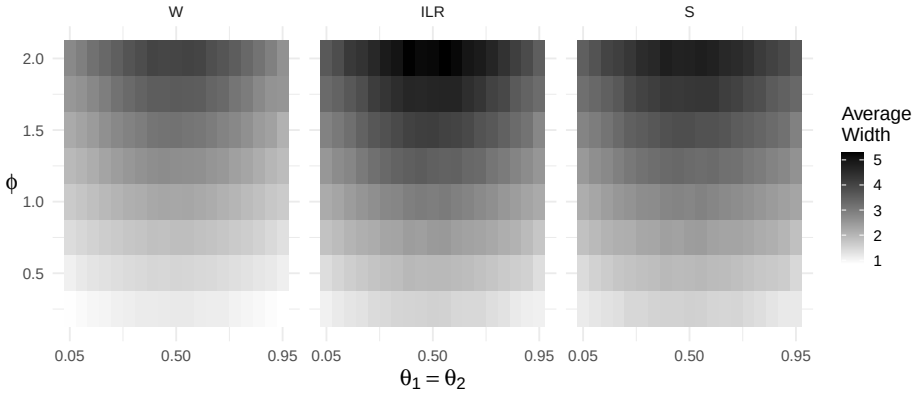
for $\phi \geq 1.00$, but we see in Figure 4.2a that this coverage result was to a lesser extent than in Figure 4.1a. Both the ILR and score CIs showed much more evidence of approximately nominal coverage, with the ILR CI providing nearer-to-nominal coverage than the score CI. We remark that fewer instances occurred of the score CI achieving nominal coverage than for the ILR CI. For instance, when the ILR CI under-covered, specifically when $\theta_1 = \theta_2 = 0.50$ and $\phi = 0.25$, the score CI under-covered for more values of $\theta_1 = \theta_2$ when $\phi = 0.25$. Figure 4.2b also shows evidence that the average interval widths of the three CIs became more similar to each other than those observed in Figure 4.1, with the Wald CI still yielding the narrowest intervals while often having less-than-nominal coverage.

Figure 4.4 displays the coverage properties for the remaining (N_0, N) combinations: $\{(2,6),\ (3,6),\ (3,9),\ \text{and}\ (4,8)\}$. When $(N_0, N) = (2,6)$, we see that the Wald CI continued the under-coverage that we observed in previous simulations. However, when $(N_0, N) = (4,8)$, the Wald CI's began to yield closer-to-nominal coverage, albeit still under-covering for most of the parameter configurations, and over-covering for $\phi = 0.25$ and $0.05 \leq \theta_1, \theta_2 \leq 0.95$. We also see evidence that the score CI was somewhat conservative for the four sample-size pairs (N_0, N), becoming less conservative as the sample sizes increased. However, the score CI still over-covered for most parameter configurations. Meanwhile, the ILR CI demonstrated the best overall coverage properties. Similar to its competitors, the ILR CI over-covered for extreme values of $\theta_1 = \theta_2$, i.e., when $\theta_i < 0.25$ and $\theta_i > 0.75$, $i = 1, 2$, with $\phi < 0.10$. However, the ILR CI appears to consistently provide near-nominal coverage for most combinations of ϕ and $\theta_1 = \theta_2$.

Figure 4.5 displays the average interval-width behavior of the three competing CIs for $(N_0, N) \in \{(2,6),\ (3,6),\ (3,9),\ \text{and}\ (4,8)\}$. When $(N_0, N) = (2,6)$, we observe that the Wald CI produced the narrowest CIs while the score CI and ILR CI were comparable in terms of average widths, with the score CI often being slightly



(a) Heat map of coverage properties for Wald, ILR, and score CIs



(b) Heat map of average interval widths for Wald, ILR, and score CIs

Figure 4.3. Coverage and average interval-width heat maps for the ratio of two Poisson rate parameters for $N_0 = 2$, $N = 4$, $0.25 \leq \phi \leq 2.00$, and $0.05 \leq \theta_1, \theta_2 \leq 0.95$.

narrower than the ILR CI. However, as the sample sizes increased, the three CIs began to converge to the same interval widths. When $(N_0, N) = (4, 8)$, we see what appears to be very similar behavior for the CI widths among the three competing CIs. Taking what we see here in consideration with the coverage properties visible in Figure 4.4, we concluded that the ILR CI was comparable in width to the score CI while also being, on average, less conservative. Moreover, the ILR CI demonstrated closer-to-nominal coverage compared to the score CI coverage for many combinations of ϕ and θ_i, $i = 1, 2$.

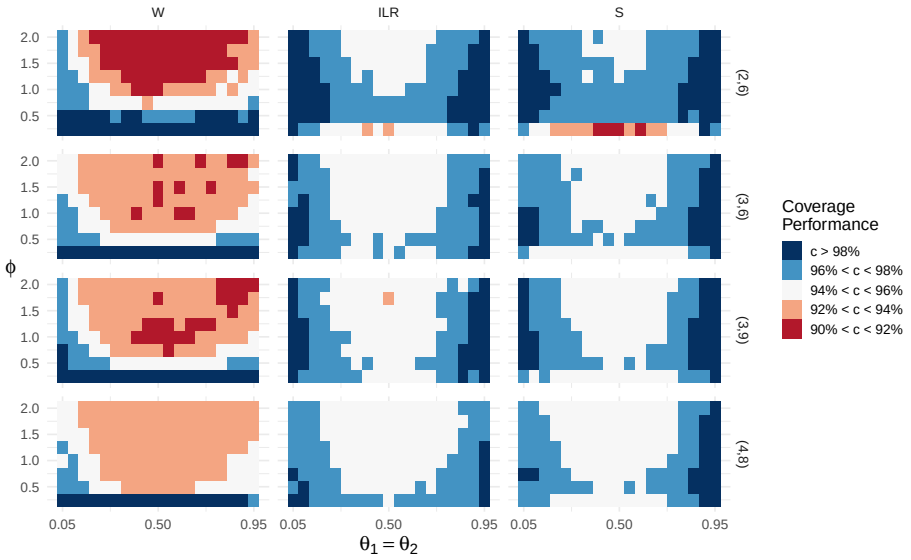


Figure 4.4. Coverage heat maps for Wald, ILR, and score CIs of the ratio of two Poisson rate parameters for $(N_0, N) \in \{(2,6),\ (3,6),\ (3,9),\ \text{and}\ (4,8)\}$ when $0.25 \leq \phi \leq 2.00$ and $0.05 \leq \theta_1, \theta_2 \leq 0.95$.

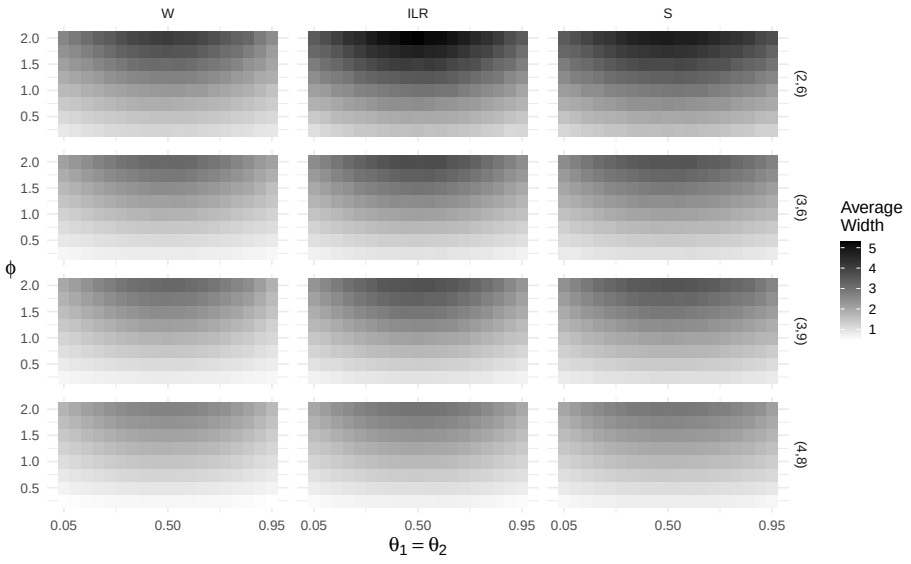


Figure 4.5. Average interval-width heat maps for Wald, ILR, and score CIs of the ratio of two Poisson rate parameters for $(N_0, N) \in \{(2,6),\ (3,6),\ (3,9),\ \text{and}\ (4,8)\}$ when $0.25 \leq \phi \leq 2.00$ and $0.05 \leq \theta_1, \theta_2 \leq 0.95$.

4.4 A Real-Data Example

Next, we analyze data derived from Whittemore and Gong (1991) in order to demonstrate the efficacy of the Wald, score, and ILR CIs for the ratio of two Poisson rate parameters that follow a double-sampling scheme when counts are subject to misclassification. The data are based on reports of death by cervical cancer in European countries from 1969 to 1973. Though Whittemore and Gong observed data from England and Wales, Belgium, France, and Italy, we will focus only on France and Italy. We consider the information reported at the country level as the fallible sample and the further investigation done by country-specific physicians as the infallible sample.

We desire to estimate $\phi = \lambda_1/\lambda_2$, where λ_1 represents the rate of deaths by cervical cancer in France and λ_2 represents the same for Italy. For our situation, the value of ϕ allows us to determine whether or not a statistically significant difference exists between the rate of deaths in France and Italy. Table 4.1 shows the counts for the fallible and infallible samples. The notation z_i represents the count of deaths from population i, $i = 1, 2$, from the fallible classifier. The notation m_{0i} represents the count of deaths from population i, $i = 1, 2$, that were correctly labeled by the fallible classifier and confirmed by the infallible classifier. Finally, y_{0i} represents the count of deaths from population i, $i = 1, 2$, that were incorrectly labeled by the fallible classifier and then correctly identified by the infallible classifier. Here, $i = 1$ represents deaths from France and $i = 2$ represents deaths from Italy. To emphasize the performance of the Wald, score, and ILR CIs on small samples, we take the data provided in Whittemore and Gong (1991) and, using a portion of the fallible sample results, calculate the *MLE's* for the original data set. We then use these *MLE's* to simulate the data provided in Table 4.2.

The three CIs based on the data given in Table 4.2 are displayed in Table 4.3. We observe that each CI contains the value $\phi = 1$, which corresponds to $\lambda_1 = \lambda_2$.

Table 4.1. Original cervical cancer data with training sample size $N_0 = 2301$ and validation sample size $N = 15,246$

		Investigated Counts	
Group	Country Data	Correct	Misclassified
France	$z_1 = 1117$	$m_{01} = 38$	$y_{01} = 15$
Italy	$z_2 = 839$	$m_{02} = 41$	$y_{02} = 9$

Table 4.2. Simulated cervical cancer data with training sample size $N_0 = 2.301$ and validation sample size $N = 15.246$

		Investigated Counts	
Group	Country Data	Correct	Misclassified
Injured	$z_1 = 137$	$m_{01} = 17$	$y_{01} = 5$
Uninjured	$z_2 = 105$	$m_{02} = 26$	$y_{02} = 8$

Because our CIs contain one, we found no evidence of a statistical difference between the rate parameters of death by cervical cancer for French and Italian women at the 5% significance level. We also observe that all of the CIs were comparable in size. However, the ILR CI was the narrowest.

Table 4.3. CIs and Interval Widths

Interval	Lower End Point	Upper End Point	Width
Wald CI	0.304	1.193	0.889
score CI	0.416	1.274	0.858
ILR CI	0.568	1.335	0.767

4.5 Discussion

Here, we have derived an ILR CI for the ratio of two complementary Poisson rate parameters and contrasted its efficacy to two other CIs for $\phi = \lambda_1/\lambda_2$ using data containing counts that are subjected to misclassification. To account for the misclassified observations, we used a double-sampling scheme that provides us with

both fallible and infallible samples with which we construct the CIs. We used Monte Carlo simulations to examine the coverage and average interval-width properties of each CI and to compare and contrast their individual performances for various sample size and parameter configurations. We observed that the Wald CI typically provided narrower intervals that tended to under-cover while the score and ILR CIs provided similar coverage results. However, the ILR CI most consistently provided nominal coverage results, and the score CI had slightly narrower average interval-widths. Finally, we observed that for a real-data example involving rates of death by cervical cancer in France and Italy, the ILR CI gave the narrowest CI from among all of the CIs considered here.

4.6 Appendix: A Derivation of the Hessian Matrix from the Likelihood Function of ϕ

Here, we derive the Hessian matrix, or the observed information matrix, for the log-likelihood function (4.2) described in Section 4.2. We have

$$\mathbf{H} := \begin{bmatrix} \frac{\partial^2 l}{\partial \phi^2} & \frac{\partial^2 l}{\partial \phi \partial \lambda_2} & \frac{\partial^2 l}{\partial \phi \partial \theta_1} & \frac{\partial^2 l}{\partial \phi \partial \theta_2} \\ \frac{\partial^2 l}{\partial \lambda_2 \partial \phi} & \frac{\partial^2 l}{\partial \lambda_2^2} & \frac{\partial^2 l}{\partial \lambda_2 \partial \theta_1} & \frac{\partial^2 l}{\partial \lambda_2 \partial \theta_2} \\ \frac{\partial^2 l}{\partial \theta_1 \partial \phi} & \frac{\partial^2 l}{\partial \theta_1 \partial \lambda_2} & \frac{\partial^2 l}{\partial \theta_1^2} & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 l}{\partial \theta_2 \partial \phi} & \frac{\partial^2 l}{\partial \theta_2 \partial \lambda_2} & \frac{\partial^2 l}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_2^2} \end{bmatrix} = \begin{bmatrix} h_\phi & h_{\phi\lambda_2} & h_{\phi\theta_1} & h_{\phi\theta_2} \\ h_{\lambda_2\phi} & h_{\lambda_2} & h_{\lambda_2\theta_1} & h_{\lambda_2\theta_2} \\ h_{\theta_1\phi} & h_{\theta_1\lambda_2} & h_{\theta_1} & h_{\theta_1\theta_2} \\ h_{\theta_2\phi} & h_{\theta_2\lambda_2} & h_{\theta_2\theta_1} & h_{\theta_2} \end{bmatrix},$$

where

$$h_\phi = -\frac{m_{01}}{\phi^2} - \frac{z_1(1-\theta_1)^2}{(\theta_2 + (1-\theta_1)\phi)^2} - \frac{z_2\theta_1^2}{(1-\theta_2+\theta_1\phi)^2},$$

$$h_{\phi\lambda_2} = -N - N_0,$$

$$h_{\phi\theta_1} = \frac{z_1(1-\theta_1)\phi}{(\theta_2 + (1-\theta_1)\phi)^2} - \frac{z_1}{\theta_2 + (1-\theta_1)\phi} - \frac{z_2\theta_1\phi}{(1-\theta_2+\theta_1\phi)^2} + \frac{z_2}{1-\theta_2+\theta_1\phi},$$

$$h_{\phi\theta_2} = -\frac{z_1(1-\theta_1)}{(\theta_2 + (1-\theta_1)\phi)^2} + \frac{z_2\theta_1}{(1-\theta_2+\theta_1\phi)^2},$$

$$h_{\lambda_2} = -\frac{m_{01} + m_{02} + z_1 + z_2}{\lambda_2^2},$$

$$h_{\lambda_2\theta_1} = 0,$$

$$h_{\lambda_2\theta_2} = 0,$$

$$h_{\theta_1} = -\frac{m_{01} - y_{01}}{(1-\theta_1)^2} - \frac{y_{01}}{\theta_1^2} - \frac{z_1\phi^2}{(\theta_2 + (1-\theta_1)\phi)^2} - \frac{z_2\phi^2}{(1-\theta_2+\theta_1\phi)^2},$$

$$h_{\theta_1\theta_2} = \frac{z_1\phi}{(\theta_2 + (1-\theta_1)\phi)^2} + \frac{z_2\phi}{(1-\theta_2+\theta_1\phi)^2},$$

and

$$h_{\theta_2} = -\frac{m_{02} - y_{02}}{(1-\theta_2)^2} - \frac{y_{02}}{\theta_2^2} - \frac{z_1}{(\theta_2 + (1-\theta_1)\phi)^2} - \frac{z_2}{(1-\theta_2+\theta_1\phi)^2}$$

with $h_{\phi\lambda_2} = h_{\lambda_2\phi}$, $h_{\phi\theta_1} = h_{\theta_1\phi}$, $h_{\phi\theta_2} = h_{\theta_2\phi}$, $h_{\lambda_2\theta_1} = h_{\theta_1\lambda_2}$, $h_{\lambda_2\theta_2} = h_{\theta_2\lambda_2}$, and $h_{\theta_1\theta_2} = h_{\theta_2\theta_1}$.

4.7 Appendix: A Derivation of the Information Matrix from the Likelihood Function of ϕ

Here, we derive Fisher's Information Matrix associated with the log-likelihood function given in (4.2). We have

$$\mathbf{I}(\phi) := \begin{bmatrix} -E\left(\frac{\partial^2 l}{\partial \phi^2}\right) & -E\left(\frac{\partial^2 l}{\partial \phi \partial \lambda_2}\right) & -E\left(\frac{\partial^2 l}{\partial \phi \partial \theta_1}\right) & -E\left(\frac{\partial^2 l}{\partial \phi \partial \theta_2}\right) \\ -E\left(\frac{\partial^2 l}{\partial \lambda_2 \partial \phi}\right) & -E\left(\frac{\partial^2 l}{\partial \lambda_2^2}\right) & -E\left(\frac{\partial^2 l}{\partial \lambda_2 \partial \theta_1}\right) & -E\left(\frac{\partial^2 l}{\partial \lambda_2 \partial \theta_2}\right) \\ -E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \phi}\right) & -E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \lambda_2}\right) & -E\left(\frac{\partial^2 l}{\partial \theta_1^2}\right) & -E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2}\right) \\ -E\left(\frac{\partial^2 l}{\partial \theta_2 \partial \phi}\right) & -E\left(\frac{\partial^2 l}{\partial \theta_2 \partial \lambda_2}\right) & -E\left(\frac{\partial^2 l}{\partial \theta_2 \partial \theta_1}\right) & -E\left(\frac{\partial^2 l}{\partial \theta_2^2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} i_\phi & i_{\phi\lambda_2} & i_{\phi\theta_1} & i_{\phi\theta_2} \\ i_{\lambda_2\phi} & i_{\lambda_2} & i_{\lambda_2\theta_1} & i_{\lambda_2\theta_2} \\ i_{\theta_1\phi} & i_{\theta_1\lambda_2} & i_{\theta_1} & i_{\theta_1\theta_2} \\ i_{\theta_2\phi} & i_{\theta_2\lambda_2} & i_{\theta_2\theta_1} & i_{\theta_2} \end{bmatrix},$$

where

$$\begin{aligned} i_\phi &= -E\left(-\frac{m_{01}}{\phi^2} - \frac{z_1(1-\theta_1)^2}{(\theta_2+(1-\theta_1)\phi)^2} - \frac{z_2\theta_1^2}{(1-\theta_2+\theta_1\phi)^2}\right) \\ &= \frac{N_0\phi\lambda_2}{\phi^2} + \frac{(1-\theta_1)^2 N[\phi\lambda_2(1-\theta_1)+\lambda_2\theta_2]}{(\theta_2+(1-\theta_1)\phi)^2} + \frac{\theta_1^2 N[\lambda_2(1-\theta_2)+\phi\lambda_2\theta_1]}{(1-\theta_2+\theta_1\phi)^2} \\ &= \frac{N_0\lambda_2}{\phi} + \frac{(1-\theta_1)^2 N\lambda_2}{\theta_2+(1-\theta_1)\phi} + \frac{\theta_1^2 N\lambda_2}{1-\theta_2+\theta_1\phi}, \\ i_{\phi\lambda_2} &= -E\left(-N-N_0\right) \\ &= N+N_0, \\ i_{\phi\theta_1} &= -E\left(\frac{z_1(1-\theta_1)\phi}{(\theta_2+(1-\theta_1)\phi)^2} - \frac{z_1}{\theta_2+(1-\theta_1)\phi} - \frac{z_2\theta_1\phi}{(1-\theta_2+\theta_1\phi)^2} + \frac{z_2}{1-\theta_2+\theta_1\phi}\right) \\ &\quad + \frac{\theta_1\phi N[\lambda_2(1-\theta_2)+\phi\lambda_2\theta_1]}{(1-\theta_2+\theta_1\phi)^2} - \frac{N[\lambda_2(1-\theta_2)+\phi\lambda_2\theta_1]}{1-\theta_2+\theta_1\phi} \\ &= -\frac{(1-\theta_1)\phi N\lambda_2}{\theta_2+(1-\theta_1)\phi} + N\lambda_2 + \frac{\theta_1\phi N\lambda_2}{1-\theta_2+\theta_1\phi} - N\lambda_2 \\ &= -\frac{(1-\theta_1)\phi N\lambda_2}{\theta_2+(1-\theta_1)\phi} + \frac{\theta_1\phi N\lambda_2}{1-\theta_2+\theta_1\phi}, \\ i_{\phi\theta_2} &= -E\left(-\frac{z_1(1-\theta_1)}{(\theta_2+(1-\theta_1)\phi)^2} + \frac{z_2\theta_1}{(1-\theta_2+\theta_1\phi)^2}\right) \end{aligned}$$

$$= \frac{(1-\theta_1)N\lambda_2}{\theta_2+(1-\theta_1)\phi} - \frac{\theta_1 N\lambda_2}{1-\theta_2+\theta_1\phi},$$

$$i_{\lambda_2} = -E\left(-\frac{m_{01}+m_{02}+z_1+z_2}{\lambda_2^2}\right)$$
$$= \frac{1}{\lambda_2}\left(N_0\phi + N_0 + N[\phi(1-\theta_1)+\theta_2+N[(1-\theta_2)+\phi\theta_1]\right)$$
$$= \frac{(N_0+N)(\phi+1)}{\lambda_2},$$

$$i_{\lambda_2\theta_1} = -E\left(0\right) = 0,$$

$$i_{\lambda_2\theta_2} = -E\left(0\right) = 0,$$

$$i_{\theta_1} = -E\left(-\frac{m_{01}-y_{01}}{(1-\theta_1)^2} - \frac{y_{01}}{\theta_1^2} - \frac{z_1\phi^2}{(\theta_2+(1-\theta_1)\phi)^2} - \frac{z_2\phi^2}{(1-\theta_2+\theta_1\phi)^2}\right)$$
$$= \frac{N_0\phi\lambda_2 - m_{01}\theta_1}{(1-\theta_1)^2} + \frac{m_{01}}{\theta_1} + \frac{\phi^2 N\lambda_2}{\theta_2+(1-\theta_1)\phi} + \frac{\phi^2 N\lambda_2}{1-\theta_2+\theta_1\phi},$$

$$i_{\theta_1\theta_2} = -E\left(\frac{z_1\phi}{(\theta_2+(1-\theta_1)\phi)^2} + \frac{z_2\phi}{(1-\theta_2+\theta_1\phi)^2}\right)$$
$$= -\frac{\phi N\lambda_2}{\theta_2+(1-\theta_1)\phi} - \frac{\phi N\lambda_2}{1-\theta_2+\theta_1\phi},$$

and

$$i_{\theta_2} = -E\left(-\frac{m_{02}-y_{02}}{(1-\theta_2)^2} - \frac{y_{02}}{\theta_2^2} - \frac{z_1}{(\theta_2+(1-\theta_1)\phi)^2} - \frac{z_2}{(1-\theta_2+\theta_1\phi)^2}\right)$$
$$= \frac{N_0\lambda_2 - m_{02}\theta_2}{(1-\theta_2)^2} + \frac{m_{02}}{\theta_2} + \frac{N\lambda_2}{\theta_2+(1-\theta_1)\phi} + \frac{N\lambda_2}{1-\theta_2+\theta_1\phi},$$

with $i_{\phi\lambda_2} = i_{\lambda_2\phi}$, $i_{\phi\theta_1} = i_{\theta_1\phi}$, $i_{\phi\theta_2} = i_{\theta_2\phi}$, $i_{\lambda_2\theta_1} = i_{\theta_1\lambda_2}$, $i_{\lambda_2\theta_2} = i_{\theta_2\lambda_2}$, and $i_{\theta_1\theta_2} = i_{\theta_2\theta_1}$.

4.8 Appendix: A Derivation of a Closed-Form Integrated-Likelihood-Function for ϕ

Here, we derive a closed-form integrated-likelihood-function kernel for the ratio of two Poisson rate parameters under a double-sampling paradigm. We use the weighting function $f(\lambda_2, \theta_1, \theta_2) = g(\lambda_2)h_1(\theta_1)h_2(\theta_2)$, where $g(\lambda_2)$ is a Gamma$(0.001, 0.001)$ density function and $h_i(*)$ is a Beta$(1, 1)$ density function with $i = 1, 2$. Let $\mathbf{d} = (z_1, z_2, m_{01}, m_{02}, y_{01}, y_{02})'$ and $\mathbf{\Theta}^* = (\phi, \lambda_2, \theta_1, \theta_2)'$. Then,

$$
\begin{aligned}
L_I(\phi) &= \int_0^1 \int_0^1 \int_0^\infty L(\mathbf{\Theta}^*|\mathbf{d}) f(\lambda_2, \theta_1, \theta_2|\phi) d\lambda_2 d\theta_1 d\theta_2 \\
&= \int_0^1 \int_0^1 \int_0^\infty L(\mathbf{\Theta}^*|\mathbf{d}) g(\lambda_2) h_1(\theta_1) h_2(\theta_2) d\lambda_2 d\theta_1 d\theta_2 \\
&\propto \int_0^1 \int_0^1 \int_0^\infty \phi^{m_{01}} \lambda_2^{m_{01}+m_{02}+z_1+z_2} e^{-\lambda_2(\phi+1)(N_0+N)} [\phi(1-\theta_1)+\theta_2]^{z_1} \\
&\quad \times [(1-\theta_2)+\phi\theta_1]^{z_2} \theta_1^{y_{01}} (1-\theta_1)^{m_{01}-y_{01}} \theta_2^{y_{02}} (1-\theta_2)^{m_{02}-y_{02}} \\
&\quad \times \lambda_2^{0.001-1} e^{-0.001\lambda_2} d\lambda_2 d\theta_1 d\theta_2 \\
&= \int_0^1 \int_0^1 \int_0^\infty \phi^{m_{01}} \lambda_2^{m_{01}+m_{02}+z_1+z_2+0.001-1} e^{-\lambda_2[(\phi+1)(N_0+N)+0.001]} \\
&\quad \times [\phi(1-\theta_1)+\theta_2]^{z_1} [(1-\theta_2)+\phi\theta_1]^{z_2} \theta_1^{y_{01}} (1-\theta_1)^{m_{01}-y_{01}} \\
&\quad \times \theta_2^{y_{02}} (1-\theta_2)^{m_{02}-y_{02}} d\lambda_2 d\theta_1 d\theta_2 \\
&= \frac{\Gamma(m_{01}+m_{02}+z_1+z_2+0.001)\phi^{m_{01}}}{[(\phi+1)(N_0+N)+0.001]^{m_{01}+m_{02}+z_1+z_2+0.001}} \\
&\quad \times \int_0^1 \int_0^1 [\phi(1-\theta_1)+\theta_2]^{z_1} [(1-\theta_2)+\phi\theta_1]^{z_2} \\
&\quad \times \theta_1^{y_{01}} (1-\theta_1)^{m_{01}-y_{01}} \theta_2^{y_{02}} (1-\theta_2)^{m_{02}-y_{02}} d\theta_1 d\theta_2 \\
&= \frac{\Gamma(m_{01}+m_{02}+z_1+z_2+0.001)\phi^{m_{01}}}{[(\phi+1)(N_0+N)+0.001]^{m_{01}+m_{02}+z_1+z_2+0.001}} \\
&\quad \times \int_0^1 \int_0^1 \theta_1^{y_{01}} (1-\theta_1)^{m_{01}-y_{01}} \theta_2^{y_{02}} (1-\theta_2)^{m_{02}-y_{02}} \\
&\quad \times \sum_{i=0}^{z_1} \binom{z_1}{i} \phi^i (1-\theta_1)^i \theta_2^{z_1-i} \sum_{j=0}^{z_2} \binom{z_2}{j} (1-\theta_2)^{z_2-j} \phi^j \theta_1^j d\theta_1 d\theta_2 \\
&= \frac{\Gamma(m_{01}+m_{02}+z_1+z_2+0.001)\phi^{m_{01}}}{[(\phi+1)(N_0+N)+0.001]^{m_{01}+m_{02}+z_1+z_2+0.001}} \sum_{i=0}^{z_1} \sum_{j=0}^{z_2} \binom{z_1}{i} \binom{z_2}{j}
\end{aligned}
$$

$$
\begin{aligned}
&\times \phi^{i+j} \int_0^1 \int_0^1 \theta_1^{y_{01}+j}(1-\theta_1)^{m_{01}-y_{01}+i}\theta_2^{y_{02}+z_1-i}(1-\theta_2)^{m_{02}-y_{02}+z_2-j} d\theta_1 d\theta_2 \\
&= \frac{\Gamma(m_{01}+m_{02}+z_1+z_2+0.001)\phi^{m_{01}}}{[(\phi+1)(N_0+N)+0.001]^{m_{01}+m_{02}+z_1+z_2+0.001}} \sum_{i=0}^{z_1}\sum_{j=0}^{z_2} \binom{z_1}{i}\binom{z_2}{j} \\
&\quad \times \phi^{i+j} B\left(y_{01}+j+1,\ m_{01}-y_{01}+i+1\right) \\
&\quad \times B\left(y_{02}+z_1-i+1,\ m_{02}-y_{02}+z_2-j+1\right) \\
&= \frac{1}{[(\phi+1)(N_0+N)+0.001]^{m_{01}+m_{02}+z_1+z_2+0.001}} \sum_{i=0}^{z_1}\sum_{j=0}^{z_2} \binom{z_1}{i}\binom{z_2}{j} \phi^{m_{01}+i+j} \\
&\quad \times B\left(y_{01}+j+1,\ m_{01}-y_{01}+i+1\right) \\
&\quad \times B\left(y_{02}+z_1-i+1,\ m_{02}-y_{02}+z_2-j+1\right).
\end{aligned}
$$

BIBLIOGRAPHY

Agresti, A. (1999). On logit confidence intervals for the odds ratio with small samples. *Biometrics*, 55(2):597–602.

Agresti, A. (2003). Dealing with discreteness: making exact confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research*, 12(1):3–21.

Agresti, A. (2011). Score and pseudo-score confidence intervals for categorical data analysis. *Statistics in Biopharmaceutical Research*, 3(2):163–172.

Agresti, A. and Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126.

Altman, D. G. (2005). Why we need confidence intervals. *World Journal of Surgery*, 29(5):554–556.

Berger, J. O., Liseo, B., Wolpert, R. L., et al. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1):1–28.

Boese, D. H., Young, D. M., and Stamey, J. D. (2006). Confidence intervals for a binomial parameter based on binary data subject to false-positive misclassification. *Computational Statistics & Data Analysis*, 50(12):3369–3385.

Bross, I. (1954). Misclassification in 2 x 2 tables. *Biometrics*, 10(4):478–486.

Carroll, R. J., Gail, M. H., and Lubin, J. H. (1993). Case-control studies with errors in covariates. *Journal of the American Statistical Association*, 88(421):185–199.

Clyde, M. A., Ghosh, J., and Littman, M. L. (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, 20(1):80–101.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pages 135–148. University of California Press, Berkeley, CA.

DiBartolomeo, D. and Witkowski, E. (1997). Mutual fund misclassification: Evidence based on style analysis. *Financial Analysts Journal*, 53(5):32–43.

Drews, C. D., Kraus, J. F., and Greenland, S. (1990). Recall bias in a case-control study of sudden infant death syndrome. *International Journal of Epidemiology*, 19(2):405–411.

Fisher, T. J. and Robbins, M. W. (2019). A cheap trick to improve the power of a conservative hypothesis test. *The American Statistician*, 73(3):232–242.

Fujisawa, H. and Izumi, S. (2000). Inference about misclassification probabilities from repeated binary responses. *Biometrics*, 56(3):706–711.

Gart, J. J. and Zweifel, J. R. (1967). On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika*, pages 181–187.

Graham, P., Mengersen, K., and Morton, A. (2003). Confidence limits for the ratio of two rates based on likelihood scores: non-iterative method. *Statistics in Medicine*, 22(12):2071–2083.

Greenland, S. (1988). Variance estimation for epidemiologic effect estimates under misclassification. *Statistics in Medicine*, 7(7):745–757.

Greenland, S. (2008). Maximum-likelihood and closed-form estimators of epidemiologic measures under misclassification. *Journal of Statistical Planning and Inference*, 138(2):528–538.

Haldane, B. J. (1956). The estimation and significance of the logarithm of a ratio of frequencies. *Annals of Human Genetics*, 20(4):309–311.

Hankin, R. K. (2015). Numerical evaluation of the gauss hypergeometric function with the hypergeo package. *The R Journal*, 7:81–88.

Hansen, M. H., Hurwitz, W. N., and Bershad, M. A. (1960). Measurement errors in censuses and surveys. *Bulletin de l'Institut International de Statistique*, 38(2):359–374.

Ji, F., Yang, Y., Haynes, C., Finch, S. J., and Gordon, D. (2006). Computing asymptotic power and sample size for case-control genetic association studies in the presence of phenotype and/or genotype misclassification errors. *Statistical Applications in Genetics and Molecular Biology*, 4(1).

Joseph, L., Gyorkos, T. W., and Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, 141(3):263–272.

Li, M. (2009). *Maximum-Likelihood-Based Confidence Intervals for the Difference of Poisson Rate Parameters Where the Data Is Subject to Misclassification*. Stephen F. Austin State University.

Lie, R. T., Heuch, I., and Irgens, L. M. (1994). Maximum likelihood estimation of the proportion of congenital malformations using double registration systems. *Biometrics*, pages 433–444.

Lyles, R. H. (2002). A note on estimating crude odds ratios in case-control studies with differentially misclassified exposure. *Biometrics*, 58(4):1034–1036.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4):719–748.

Moors, J., Van Der Genugten, B., and Strijbosch, L. (2000). Repeated audit controls. *Statistica Neerlandica*, 54(1):3–13.

Morrissey, M. J. and Spiegelman, D. (1999). Matrix methods for estimating odds ratios with misclassified exposure data: extensions and comparisons. *Biometrics*, 55(2):338–344.

Prescott, G. J. and Garthwaite, P. H. (2002). A simple Bayesian analysis of misclassified binary data with a validation substudy. *Biometrics*, 58(2):454–458.

Raats, V. and Moors, J. (2003). Double-checking auditors: a Bayesian approach. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):351–365.

Rahardja, D. and Yang, Y. (2015). Maximum likelihood estimation of a binomial proportion using one-sample misclassified binary data. *Statistica Neerlandica*, 69(3):272–280.

Riggs, K. (2015). Confidence intervals for a proportion using inverse sampling when the data is subject to false-positive misclassification. *Journal of Data Science*, 13(2):623–636.

Riggs, K., Young, D., and Stamey, J. (2009). Likelihood-based confidence intervals for complementary Poisson rate parameters with misclassified data. *Communications in Statistics—Theory and Methods*, 38(2):159–172.

Severini, T. A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika*, 97(2):481–496.

Spencer, E. A., Mahtani, K. R., Brassey, J., and Heneghan, C. (2018). Misclassification bias. `https://catalogofbias.org/biases/misclassification-bias/`.

Sposto, R., Preston, D. L., Shimizu, Y., and Mabuchi, K. (1992). The effect of diagnostic misclassification on non-cancer and cancer mortality dose response in A-bomb survivors. *Biometrics*, pages 605–617.

Stamey, J., Seaman, J., and Young, D. (2005a). Bayesian analysis of complementary Poisson rate parameters with data subject to misclassification. *Journal of Statistical Planning and Inference*, 134(1):36–48.

Stamey, J. D., Young, D. M., and Stephens, D. (2005b). Maximum likelihood estimation of two inversely related Poisson rate parameters with misclassified data. *American Journal of Mathematical and Management Sciences*, 25(1-2):65–81.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3):227.

Tenenbein, A. (1970). A double sampling scheme for estimating from binomial data with misclassifications. *Journal of the American Statistical Association*, 65(331):1350–1361.

Viana, M. (1994). Bayesian small-sample estimation of misclassified multinomial data. *Biometrics*, pages 237–243.

Wasserstein, R. L. and Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133.

Whittemore, A. S. and Gong, G. (1991). Poisson regression with misclassified counts: Application to cervical cancer mortality rates. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 40(1):81–93.

Wiley, B., Elrod, C., Young, P. D., and Young, D. M. (2021). An integrated-likelihood-ratio confidence interval for a proportion based on under-reported and infallible data. *Statistica Neerlandica*.

Woolf, B. (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics*, 19(4):251–253.