8-2013

Objective Bayesian Hypothesis Testing and Estimation for the Risk Ratio in a Correlated 2x2 Table with Structural Zero

Xiaohua Bai
Clemson University, xbai@clemson.edu

Objective Bayesian hypothesis testing and estimation for the risk ratio in a correlated 2×2 table with structural zero

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mathematical Science

by
Xiaohua Bai
August 2013

Accepted by:
Dr. Xiaoqian Sun, Committee Chair
Dr. Colin Gallagher
Dr. Chanseok Park

Abstract

In this project, we illustrate the construction of an objective Bayesian hypothesis testing and point estimation for the risk ratio in a correlated 2×2 table with structural zero. We solve the problem using Bayesian method through the reference prior distribution, and the corresponding posterior distribution of the risk ratio can be derived. Then combined the intrinsic discrepancy, an invariant information-based loss function, provides an integrated objective Bayesian solution to both hypothesis testing and point estimation problems.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Background

In some clinical trials, one may encounter a problem that correlated 2×2 table with structural zero. It means one of the four cells of the table is a logical impossibility. One famous example was given in Agresti(1990)[1], where the sample of 156 dairy calves were first classified according to whether they had a primary pneumonia infection. After the first infection was cleared up, they reclassified the calves according to whether they developed a secondary infection within 2 weeks. If the calve did not accept the first infection, then it could not get the second infection. That means it is impossible that a calve gets a second infection without accepting the first infection, which combines the correlated 2×2 table with structural zero in one of the off-diagonal cells. Table 1.1 is the data structure of this problem.

For Table 1.1, we use "yes" or "no" to represent whether the calve accepts the infection or not. Without loss of generality, let X_{11} be the amount of trials that accept the two pneumonia infections, X_{12} be the amount of trials that only infect the first one, and X_{22} be the amount of trials

	Second Infection(Yes)	**Second Infection(No)**	**Total**
First Infection(Yes)	$X_{11}(P_{11})$	$X_{12}(P_{12})$	$X_{1+}(P_{1+})$
First Infection(No)		$X_{22}(P_{22})$	$X_{22}(P_{22})$
Total	$X_{11}(P_{11})$	$X_{12}+X_{22}(P_{12}+P_{22})$	n(1)

Table 1.1: Data and probability of the calves for the First and Second Infections

that do not infect any infection. Since all trials must be one of the three cases that we mentioned above, then this statistical event can be described using the multinomial distribution, which P_{11}, P_{12}, P_{22} are the cell probabilities that correspond to the X_{11}, X_{12}, X_{22}.

The probability mass function for the multinomial distribution of this statistical event is

$$P(X_{11} = x_{11}, X_{12} = x_{12}, X_{22} = x_{22}) = \frac{n!}{x_{11}x_{12}x_{22}} p_{11}^{x_{11}} p_{12}^{x_{12}} p_{22}^{x_{22}}, \tag{1.1}$$

where x_{11}, x_{12}, x_{22} are nonnegative integers, $x_{11} + x_{12} + x_{22} = n$, $p_{11} + p_{12} + p_{22} = 1$, n is the total number of the trials for all cases, and p_{11}, p_{12}, p_{22} are the probabilities with respect to the x_{11}, x_{12}, x_{22}[4].

The risk ratio (RR) between the secondary infection given the primary infection, and the primary infection is defined by $\Phi = P_{11}/(P_{1+})^2$, where $P_{1+} = P_{11} + P_{12}$. Some statisticians proposed the frequentist statistical inference for this problem. Liu (1998) studied the RR's interval estimation based on Wald's test[6]. Tang and Tang (2002) established the interval estimation by score's test based on Liu[9]. Some statisticians also provided statistical inference for this problem using Bayesian analysis. Stamey (2006) studied Bayesian inference for the RR under the beta prior distribution[8]. Shi (2009) obtained a Bayesian credible interval for the RR under the Dirichlet prior distribution[7].

Often, we have two methods to choose the prior distribution for the Bayesian inference. One is a subjective prior distribution, the other is an objective prior distribution. Since the beta prior distribution and the Dirichlet prior distribution also belong to the class of subjective prior distributions, then I choose an objective prior distribution to perform the Bayesian hypothesis testing and point estimation for the risk ratio in the correlated 2×2 table with structural zero. In the multinomial distribution with the parameters (P_{11}, P_{12}), we transform the corresponding parameters to (Φ, Λ) with Φ as the parameter of interest. That means, we need add one nuisance parameter Λ. Here, we choose $\Lambda = P_{1+}$. Then we derive the reference prior, the posterior distribution and the intrinsic discrepancy loss function. In addition, in order to derive the reference prior for (Φ, Λ), we use the method of transformation from the reference prior with parameters (P_{11}, P_{12}).

Chapter 2

Bayesian Inference

2.1 Bayesian inference summaries

Bayesian inference is based on the posterior distribution. Suppose the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is generated from model

$$\mathcal{M} = \{p(\mathbf{x} \mid \boldsymbol{\theta}); \mathbf{x} \in \mathbf{X}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}, \tag{2.1}$$

where $\mathbf{X}$ is the sample space and $\boldsymbol{\theta}$ is an unknown parameter. Then we can obtain the posterior distribution of $\boldsymbol{\theta}$ given data $\mathbf{x}$ by Bayes formula

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{\pi(\boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\mu}) p(\mathbf{x} \mid \boldsymbol{\mu}) d\boldsymbol{\mu}}, \tag{2.2}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution[5]. Since the denominator $\int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\mu}) p(\mathbf{x} \mid \boldsymbol{\mu}) d\boldsymbol{\mu}$ is the marginal density of $\mathbf{x}$, so the integration of the denominator is a constant, then we have

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}) \propto \pi(\boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}). \tag{2.3}$$

Now, refer to the model (2.1) and it generates the data $\mathbf{x}$. However, for the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ in the Bayesian problems of hypothesis testing and point estimation, we often may be interested in $\boldsymbol{\omega}(\boldsymbol{\theta})$, a function of θ. In this way, we could rewrite the model as

$$\mathcal{M} = \{p(\mathbf{x} \mid \boldsymbol{\omega}, \lambda); \boldsymbol{\omega} \in \boldsymbol{\Omega}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}, \tag{2.4}$$

where $\mathbf{\Omega} \times \mathbf{\Lambda} = \mathbf{\Theta}$. In order to perform the Bayesian inference, it is very important to choose an appropriate loss function for the problem.

Let $\delta\{\boldsymbol{\omega}_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\}$ be the loss function, where the value ω_0 acts as a proxy for the parameter $\boldsymbol{\omega}$ which we are interested in. Then, we can use both the posterior distribution and the loss function to solve the problems for point estimation, credible interval and hypothesis testing through the corresponding posterior expected loss,

$$d(\boldsymbol{\omega}_0 \mid \mathbf{x}) = \int_{\mathbf{\Omega}} \int_{\mathbf{\Lambda}} \delta\{\boldsymbol{\omega}_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\} \pi(\boldsymbol{\omega}, \boldsymbol{\lambda} \mid \mathbf{x}) d\boldsymbol{\lambda} d\boldsymbol{\omega}, \tag{2.5}$$

where $\pi(\boldsymbol{\omega}, \boldsymbol{\lambda} \mid \mathbf{x})$ is the posterior distribution of $(\boldsymbol{\omega}, \boldsymbol{\lambda})$.

We can easily see $d(\boldsymbol{\omega}_0 \mid \mathbf{x})$ is a function of $\boldsymbol{\omega}_0$, and the generated data $\mathbf{x}$ is the only information that we need to provide for the formula (2.5). Also, we know the value of expected loss depends on both the loss function and the prior distribution.

2.1.1 Hypothesis testing

Consider the hypothesis testing for $\boldsymbol{\omega}$ as following:

$$\mathbf{H}_0 : \boldsymbol{\omega} = \boldsymbol{\omega}_0 \qquad \text{vs} \qquad \mathbf{H}_1 : \boldsymbol{\omega} \neq \boldsymbol{\omega}_0, \tag{2.6}$$

where the true model is

$$\mathcal{M} = \{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}), \mathbf{x} \in \mathbf{X}, \boldsymbol{\omega} \in \mathbf{\Omega}, \boldsymbol{\lambda} \in \mathbf{\Lambda}\}, \tag{2.7}$$

where $\boldsymbol{\omega}$ is an unknown parameter of interest.

We define the decision problem for testing the null hypothesis, and two elements are included in the action space: to accept the null hypothesis or to reject the null hypothesis. Without loss of generality, we use the a_0 to represent the acceptance, else use the a_1. Then, it is necessary to specify the appropriate loss functions $\{\delta\{a_i, (\boldsymbol{\omega}, \boldsymbol{\lambda})\}, i = 0, 1\}$ to measure the values of parameters $(\boldsymbol{\omega}, \boldsymbol{\lambda})$. Therefore, if the data $\mathbf{x}$ is given, the decision rule is to reject the null hypothesis if and only if the value of expected posterior loss for accepting the null hypothesis $\delta\{a_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\}$ is larger than that for rejecting $\delta\{a_1, (\boldsymbol{\omega}, \boldsymbol{\lambda})\}$. The formula is represented as following:

$$\int_{\mathbf{\Omega}} \int_{\mathbf{\Lambda}} \mathbf{\Delta}\delta\{\mathbf{H}_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\} \pi(\boldsymbol{\omega}, \boldsymbol{\lambda} \mid \mathbf{x}) d\boldsymbol{\lambda} d\boldsymbol{\omega} > 0. \tag{2.8}$$

where $\boldsymbol{\Delta}\delta\{\mathbf{H}_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\}$, the difference for the loss, is defined as

$$\boldsymbol{\Delta}\delta\{\mathbf{H}_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\} = \delta\{a_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\} - \delta\{a_1, (\boldsymbol{\omega}, \boldsymbol{\lambda})\}. \tag{2.9}$$

Then the decision criterion can be written as

$$\int_{\boldsymbol{\Omega}} \int_{\boldsymbol{\Lambda}} \{\delta\{a_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\} - \delta\{a_1, (\boldsymbol{\omega}, \boldsymbol{\lambda})\}\}\pi(\boldsymbol{\omega}, \boldsymbol{\lambda} \mid \mathbf{x}) d\boldsymbol{\lambda} d\boldsymbol{\omega} > 0. \tag{2.10}$$

2.1.2 Point estimation

Bayesian estimation includes two aspects, which are the point estimation and the credible interval (region estimation). We focus on the problem of point estimation because the interval estimation is essentially equivalent the hypothesis testing[4]. According to the model from (2.7), the best estimator for $\boldsymbol{\omega}$ is the value that minimizes the expected loss of the decision problem. Let $\boldsymbol{\omega}^*$ be the best estimator for $\boldsymbol{\omega}$, then $\boldsymbol{\omega}^*$ is determined by

$$\boldsymbol{\omega}^*(\mathbf{x}) = \text{arginf}_{\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}} \underset{\cdot}{(}\boldsymbol{\omega}_0 \mid \mathbf{x}). \tag{2.11}$$

In addition, the value of $\boldsymbol{\omega}^*(\mathbf{x})$ is a constant when the data $\mathbf{x}$ is given[2].

2.2 Objective Bayesian priors

The prior distribution plays on important role in Bayesian inference. In general, there are two kinds of prior distributions, one is the subjective priors, and the other is the objective priors. A subjective prior means choosing the prior distribution base on the prior knowledge. However, it is often difficult to have these prior knowledge in practice and thus an objective prior is always welcome. An objective prior is also called the noninformative prior, since it only needs little prior information. Thus, we adopt the reference prior to solve the above decision problems of the Bayesian inference.

2.2.1 Reference prior

Refer to the model (2.1), the algorithm of deriving the reference prior is given as follows[3]:

(1): Define the vector parameter $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ of interest with order from the above model, where θ_1 is parameter of first interest, θ_2 is the parameter of second interest, etc.

(2): Calculate the Fisher information matrix of parameter $\boldsymbol{\theta}$, denoted by $\mathbf{I}(\boldsymbol{\theta})$. Define $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{I}^{-1}(\boldsymbol{\theta})$. Here, we use $\mathbf{I}$ and $\mathbf{S}$ to represent them.

(3): Suppose we have known the form of $\mathbf{I}$ and $\mathbf{S}$. Let $\mathbf{S}_j$ be the $j \times j$ upper left submatrix of $\mathbf{S}$. Then, let $\mathbf{H}_j = {\mathbf{S}_j}^{-1}$ and h_j be the (j,j)th element of the matrix $\mathbf{H}_j$. In addition, given $(\theta_1, \theta_2, \ldots, \theta_j)$, we define the conditional prior for $(\theta_{j+1}, \theta_{j+2}, \ldots, \theta_k)$ by $p(\theta_{j+1}, \theta_{j+2}, \ldots, \theta_k \mid \theta_1, \theta_2, \ldots, \theta_j)$.

(4): The function $\phi_j(\theta_1, \theta_2, \ldots, \theta_j)$ is defined as below,

$$\phi_j(\theta_1, \theta_2, \ldots, \theta_j) = \exp\{ \int \frac{1}{2} \log h_j(\theta) p(\theta_{j+1}, \theta_{j+2}, \ldots, \theta_k \mid \theta_1, \theta_2, \ldots, \theta_j) d\theta_{j+1} d\theta_{j+2} \ldots d\theta_k \}. \tag{2.12}$$

Here, the formula of the conditional prior for θ_j is given as

$$p(\theta_j \mid \theta_1, \theta_2, \ldots, \theta_{j-1}) = c(\theta_1, \theta_2, \ldots, \theta_{j-1}) \phi_j(\theta_1, \theta_2, \ldots, \theta_j), \tag{2.13}$$

where $c(\theta_1, \theta_2, \ldots, \theta_{j-1})$ is fixed, such that

$$\int_{\boldsymbol{\Theta}'_j} p(\theta_j \mid \theta_1, \theta_2, \ldots, \theta_{j-1}) d\theta_j = 1. \tag{2.14}$$

For example, suppose $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Then we can derive the conditional reference prior for θ_2 by using the algorithm above, thus

$$p(\theta_2 \mid \theta_1) = c(\theta_1) \phi_2(\theta_1, \theta_2), \tag{2.15}$$

and

$$p(\theta_1) = c \phi_1(\theta_1, \theta_2), \tag{2.16}$$

where $c(\theta_1)$ and c are both fixed, which we can obtain the values using the formula (2.14). Hence, we can derive the reference prior for (θ_1, θ_2) by (2.15) and (2.16), which is given by

$$\pi(\theta_1, \theta_2) = p(\theta_2 \mid \theta_1) p(\theta_1). \tag{2.17}$$

If $h_j(\theta)$ only depends on $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_j)$, then we can simplify the formula (2.13) as

$$\phi_j(\theta_1, \theta_2, \ldots, \theta_j) = h_j^{\frac{1}{2}}(\theta). \tag{2.18}$$

2.3 The objective Bayesian criterion

2.3.1 The intrinsic discrepancy loss

Suppose $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ are two probability densities. Define $k(p_2 \mid p_1)$ as the Kullback-Leibler (KL) directed logarithmic divergence of $p_2(\mathbf{x})$ from $p_1(\mathbf{x})$, that is,

$$k(p_2 \mid p_1) = \int_{\mathbf{X}} p_1(\mathbf{x}) \log \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} d\mathbf{x}. \tag{2.19}$$

Similarly, $k(p_1 \mid p_2)$ is the KL directed logarithmic divergence of $p_1(\mathbf{x})$ from $p_2(\mathbf{x})$. Therefore, we can define the intrinsic discrepancy loss function by the definition below.

Definition. The intrinsic discrepancy loss $\delta(p_1, p_2)$ between two probability densities $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, $\mathbf{x} \in \mathbf{X}$ is

$$\begin{aligned} \delta\{p_1, p_2\} &= \min\big\{k(p_1 \mid p_2), k(p_2 \mid p_1)\big\} \\ &= \min\Big\{ \int_{\mathbf{X}} p_1(\mathbf{x}) \log \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} d\mathbf{x}, \int_{\mathbf{X}} p_2(\mathbf{x}) \log \frac{p_2(\mathbf{x})}{p_1(\mathbf{x})} d\mathbf{x} \Big\}. \end{aligned}$$

In addition, we can use $\delta\{\mathcal{M}_1, \mathcal{M}_2\}$ to describe the intrinsic discrepancy loss between the two families, which is

$$\delta(\mathcal{M}_1, \mathcal{M}_2) = \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\phi} \in \boldsymbol{\Phi}} \delta\{p_1(\mathbf{x} \mid \boldsymbol{\theta}), p_2(\mathbf{x} \mid \boldsymbol{\phi})\}, \tag{2.20}$$

where the two families are $\mathcal{M}_1 = \{p_1(\mathbf{x} \mid \boldsymbol{\theta}), \mathbf{x} \in \mathbf{X}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ and $\mathcal{M}_2 = \{p_2(\mathbf{x} \mid \boldsymbol{\phi}), \mathbf{x} \in \mathbf{X}, \boldsymbol{\phi} \in \boldsymbol{\Phi}\}$.

It is easy to observe some properties from the definition above. The intrinsic discrepancy for $\delta\{p_1, p_2\}$ is non-negative, its value could be zero if and only if $p_1 = p_2$ almost everywhere. Also, it has the symmetric property that $\delta\{p_1, p_2\} = \delta\{p_2, p_1\}$.

In fact, we may not be interested in all the parameters from a model. Let the value $\boldsymbol{\omega}_0$ be a proxy for the vector parameter $\boldsymbol{\omega}$ from the model

$$\mathcal{M}_0 = \{p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}), \mathbf{x} \in \mathbf{X}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}.$$

Refer to the model (2.7), we can define the intrinsic discrepancy loss $\delta(\boldsymbol{\omega}_0, \boldsymbol{\omega}, \boldsymbol{\lambda})$, which is used to measure the distance between the probability density $p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})$ and the family of probability densities $\{p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}), \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}$,

$$\begin{aligned}
& \delta(\boldsymbol{\omega}_0, \boldsymbol{\omega}, \boldsymbol{\lambda}) \\
= \; & \min\nolimits_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}} \delta\{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}), p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)\} \\
= \; & \min\nolimits_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}} \big\{\min\{k\big(p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}) \mid p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)\big), k\big(p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \mid p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})\big)\}\big\} \\
= \; & \min\nolimits_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}} \Big\{\min\{ \int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})}{p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0)} d\mathbf{x}, \int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0)}{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})} d\mathbf{x}\}\Big\}.
\end{aligned}$$

However, the value obtained from $k\big(p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}) | p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)\big)$ may not always bigger than the value obtained from $k\big(p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) | p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})\big)$, vice versa. So, in order to find the value for

$$\min\nolimits_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}} \Big\{\min\{ \int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})}{p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0)} d\mathbf{x}, \int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0)}{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})} d\mathbf{x}\}\Big\},$$

we need first find the relative minimal values for $\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}$ by

$$\int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})}{p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0)} d\mathbf{x},$$

and

$$\int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}_\mathbf{0}, \boldsymbol{\lambda}_0)}{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})} d\mathbf{x}[3].$$

Therefore, we have the intrinsic discrepancy $\delta(\boldsymbol{\omega}_0, \boldsymbol{\omega}, \boldsymbol{\lambda})$ shown as below

$$\begin{aligned}
\delta(\boldsymbol{\omega}_0, \boldsymbol{\omega}, \boldsymbol{\lambda}) \quad = \quad & \min\Big\{\min\nolimits_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}}\{ \int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})}{p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)} d\mathbf{x}\}, \\
& \min\nolimits_{\boldsymbol{\lambda}_0 \in \boldsymbol{\Lambda}}\{ \int_{\mathbf{X}} p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0) \log \frac{p(\mathbf{x} \mid \boldsymbol{\omega}_0, \boldsymbol{\lambda}_0)}{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})} d\mathbf{x}\}\Big\}. \qquad (2.21)
\end{aligned}$$

2.3.2 The objective Bayesian criterion

Refer to the model (2.7), we can obtain the objective Bayesian criterion to the decision problem for the hypothesis testing $\{H_0 : \boldsymbol{\omega} = \boldsymbol{\omega}_0\}$ vs $\{H_1 : \boldsymbol{\omega} \neq \boldsymbol{\omega}_0\}$. That is to reject H_0 if and only if,

$$d(\boldsymbol{\omega}_0, \mathbf{x}) = \int_{\boldsymbol{\Omega}} \int_{\boldsymbol{\Lambda}} \delta\{\boldsymbol{\omega}_0, (\boldsymbol{\omega}, \boldsymbol{\lambda})\} \pi(\boldsymbol{\omega}, \boldsymbol{\lambda} \mid \mathbf{x}) d\boldsymbol{\lambda} d\boldsymbol{\omega} > d_0, \tag{2.22}$$

where d_0 is a given positive constant, $\delta(\boldsymbol{\omega}_0, \boldsymbol{\omega}, \boldsymbol{\lambda})$ is the intrinsic discrepancy loss, and $\pi(\boldsymbol{\omega}, \boldsymbol{\lambda} \mid \mathbf{x})$ is the posterior distribution. In addition, the reference prior of the formula (2.3) can be derived by the algorithm in Section (2.2.1), since it can determine the posterior distribution $\pi(\boldsymbol{\omega}, \boldsymbol{\lambda} \mid \mathbf{x})$ with the density $p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda})$.

Following[3], we choose d_0 as 2.5, 5, 10, where $d_0 = 2.5$ is regarded as mild evidence against the null hypothesis H_0, $d_0 = 5$ is provided strong evidence against the null hypothesis H_0, and $d_0 = 10$ is secured to reject the null hypothesis H_0.

Now, let's focus on the problem of point estimation. By applying the same model $\mathcal{M} = \{p(\mathbf{x} \mid \boldsymbol{\omega}, \boldsymbol{\lambda}), \mathbf{x} \in \mathbf{X}, \boldsymbol{\omega} \in \boldsymbol{\Omega}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}\}$, the Bayes estimator $\boldsymbol{\omega}^*$ for $\boldsymbol{\omega}$ is

$$\boldsymbol{\omega}^*(\mathbf{x}) = \text{arginf}_{\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}} d(\boldsymbol{\omega}_0, \mathbf{x}). \tag{2.23}$$

Chapter 3

Method

3.1 Multinomial distribution

If the joint distribution of random variables $X_1, X_2, \ldots, X_m$ is a multinomial distribution, then we write as $(X_1, X_2, \ldots, X_m) \sim Mul(n, \theta_1, \theta_2, \ldots, \theta_m)$ and the probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_m) = \frac{n!}{\prod_{j=1}^{m} x_j!} \prod_{j=1}^{m} \theta_j^{x_j}, \tag{3.1}$$

where each x_i is a nonnegative integer, $\sum_{j=1}^{m} x_j = n$, $\sum_{j=1}^{m} \theta_j = 1$, $\theta_j > 0$, n is the total number of the trials for all cases, $x_1, x_2, \ldots, x_m$ are the numbers for each case, and θ_j is the probability with respect to the x_j[4]. Now, let's discuss an important property for the marginal distribution of the multinomial distribution.

Multinomial Theorem. Let m and n be positive integers. Let $\mathbb{A}$ be the set of vectors $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ such that each x_i is a nonnegative integer and $\sum_{j=1}^{m} x_j = n$. Then, for any real numbers $\theta_1, \theta_2, \ldots, \theta_m$, we have

$$(\theta_1 + \theta_2 + \ldots, +\theta_m)^n = \sum_{\mathbf{x} \in \mathbb{A}} \frac{n!}{\prod_{j=1}^{m} x_j!} \prod_{j=1}^{m} \theta_j^{x_j}. \tag{3.2}$$

Now, by using the Multinomial Theorem, we get an important property[4]: If $(X_1, X_2, \ldots, X_m) \sim$

$Mul(n, \theta_1, \theta_2, \ldots, \theta_m)$, then $X_j \sim \text{Bin}(n, x_j)$, where

$$P(X_j = x_j) = \frac{n!}{x_j!(n-x_j)!}\theta_j^{x_j}(1-\theta_j)^{n-x_j}. \tag{3.3}$$

Therefore, we can easily get $\text{E}(X_j) = n\theta_j$, since $X_j \sim \text{Bin}(n, x_j)$.

3.2 Reference prior

In this section, we try to derive the reference prior for the multinomial distribution in (1.1), where the algorithm was discussed in Section 2.2.1.

Let $\mathbf{X} = (X_{11}, X_{12})$, $\mathbf{P} = (P_{11}, P_{12})$ and the total number of the trials is n. Then the conditional probability density function of $\mathbf{X}$ given $\mathbf{P} = \mathbf{p} = (p_{11}, p_{12})$ is a multinomial distribution as we mentioned in the beginning,

$$f(\mathbf{x} \mid \mathbf{p}) = \begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} p_{11}^{x_{11}} p_{12}^{x_{12}} (1 - p_{11} - p_{12})^{n - x_{11} - x_{12}},$$

where n is a positive integer, $x = (x_{11}, x_{12})$, $x_{11} + x_{12} \leq n$ and $p_{11} + p_{12} < 1$.

Now, we want to derive the Fisher Information matrix for (p_{11}, p_{12}). From the multinomial distribution, we obtain

$$\log f(\mathbf{x} \mid \mathbf{p}) = \log \begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} + x_{11} \log p_{11} + x_{12} \log p_{12} + (n - x_{11} - x_{12}) \log(1 - p_{11} - p_{12}).$$

Then,

$$\begin{aligned}
\frac{\partial \log f}{\partial p_{11}} &= \frac{x_{11}}{p_{11}} - \frac{n - x_{11} - x_{12}}{1 - p_{11} - p_{12}}, \\
\frac{\partial \log f}{\partial p_{12}} &= \frac{x_{12}}{p_{12}} - \frac{n - x_{11} - x_{12}}{1 - p_{11} - p_{12}}, \\
\frac{\partial^2 \log f}{\partial p_{11}^2} &= -\frac{x_{11}}{p_{11}^2} - \frac{n - x_{11} - x_{12}}{(1 - p_{11} - p_{12})^2}, \\
\frac{\partial^2 \log f}{\partial p_{12}^2} &= -\frac{x_{12}}{p_{12}^2} - \frac{n - x_{11} - x_{12}}{(1 - p_{11} - p_{12})^2}, \\
\frac{\partial^2 \log f}{\partial p_{11} \partial p_{12}} &= \frac{\partial^2 \log f}{\partial p_{12} \partial p_{11}} = -\frac{n - x_{11} - x_{12}}{(1 - p_{11} - p_{12})^2}.
\end{aligned}$$

Since $\mathrm{E}(X_j) = n\theta_j$ by the property in section 3.1, we have

$$\begin{aligned}
\mathrm{E}\Big[-\frac{\partial^2 \log f}{\partial p_{11}^2}\Big] &= \frac{\mathrm{E}(X_{11})}{p_{11}^2} + \frac{\mathrm{E}(n - X_{11} - X_{12})}{(1-p_{11}-p_{12})^2} \\
&= \frac{n}{p_{11}} + \frac{n}{1-p_{11}-p_{12}} \\
&= \frac{n(1-p_{12})}{p_{11}(1-p_{11}-p_{12})}, \\
\mathrm{E}\Big[-\frac{\partial^2 \log f}{\partial p_{12}^2}\Big] &= \frac{\mathrm{E}(X_{12})}{p_{12}^2} + \frac{\mathrm{E}(n - X_{11} - X_{12})}{(1-p_{11}-p_{12})^2} \\
&= \frac{n}{p_{12}} + \frac{n}{1-p_{11}-p_{12}} \\
&= \frac{n(1-p_{11})}{p_{12}(1-p_{11}-p_{12})}, \\
\mathrm{E}\Big[-\frac{\partial^2 \log f}{\partial p_{11}\partial p_{12}}\Big] &= \frac{\mathrm{E}(n - X_{11} - X_{12})}{(1-p_{11}-p_{12})^2} \\
&= \frac{n}{1-p_{11}-p_{12}}, \\
\mathrm{E}\Big[-\frac{\partial^2 \log f}{\partial p_{11}\partial p_{12}}\Big] &= \frac{n}{1-p_{11}-p_{12}}.
\end{aligned}$$

Therefore, the Fisher information matrix for (p_{11}, p_{12}) is given by

$$I(p_{11}, p_{12}) = \begin{pmatrix} \frac{n(1-p_{12})}{p_{11}(1-p_{11}-p_{12})} & \frac{n}{1-p_{11}-p_{12}} \\ \frac{n}{1-p_{11}-p_{12}} & \frac{n(1-p_{11})}{p_{12}(1-p_{11}-p_{12})} \end{pmatrix}.$$

Furthermore,

$$S(p_{11}, p_{12}) = I(p_{11}, p_{12})^{-1} = \begin{pmatrix} \frac{p_{11}(1-p_{11})}{n} & -\frac{p_{11}p_{12}}{n} \\ -\frac{p_{11}p_{12}}{n} & \frac{p_{12}(1-p_{12})}{n} \end{pmatrix}.$$

Hence,

$$\begin{aligned}
H_1 &= S_1^{-1} = \frac{n}{p_{11}(1-p_{11})}, \\
H_2 &= S_2^{-1} = I(p_{11}, p_{12}).
\end{aligned}$$

Thus,

$$h_1 = \frac{n}{p_{11}(1-p_{11})} \quad \textit{and} \quad h_2 = \frac{n(1-p_{11})}{p_{12}(1-p_{11}-p_{12})}.$$

Since, h_1 and h_2 only depends on $\mathbf{p} = (p_{11}, p_{12})$, then

$$\begin{aligned}
\pi(p_{12} \mid p_{11}) &= \frac{|h_2|^{\frac{1}{2}}}{\int_0^1 |h_2|^{\frac{1}{2}} dp_{12}} = \frac{\{\frac{n(1-p_{11})}{p_{12}(1-p_{11}-p_{12})}\}^{\frac{1}{2}}}{\int_0^1 \{\frac{n(1-p_{11})}{p_{12}(1-p_{11}-p_{12})}\}^{\frac{1}{2}} dp_{12}} \\
&\propto {p_{12}}^{-\frac{1}{2}}(1-p_{11}-p_{12})^{-\frac{1}{2}}, \\
\pi(p_{11}) &= \frac{|h_1|^{\frac{1}{2}}}{\int_0^1 |h_1|^{\frac{1}{2}} dp_{11}} \propto {p_{11}}^{-\frac{1}{2}}(1-p_{11})^{-\frac{1}{2}}.
\end{aligned}$$

So, the reference prior for the $\mathbf{p} = (p_{11}, p_{12})$ is

$$\pi(p_{11}, p_{12}) \propto {p_{11}}^{-\frac{1}{2}}(1-p_{11})^{-\frac{1}{2}}{p_{12}}^{-\frac{1}{2}}(1-p_{11}-p_{12})^{-\frac{1}{2}}. \tag{3.4}$$

3.3 Posterior distribution

Using the method of transformation to obtain the reference prior for (ϕ, λ) by applying formula (3.4), then we can derive the posterior distribution for (ϕ, λ) by applying formula (2.3). Since $\phi = p_{11}/(p_{11}+p_{12})^2$ and $\lambda = p_{11}+p_{12}$,

$$\begin{aligned}
p_{11} &= \phi(p_{11}+p_{12})^2 = \phi\lambda^2, \\
p_{12} &= (p_{11}+p_{12})[1 - \frac{p_{11}}{(p_{11}+p_{12})^2}(p_{11}+p_{12})] = \phi(1-\phi\lambda).
\end{aligned}$$

So the Jacobian matrix is

$$\mathbf{J} = \begin{pmatrix} \frac{\partial p_{11}}{\partial \phi} & \frac{\partial p_{11}}{\partial \lambda} \\ \frac{\partial p_{12}}{\partial \phi} & \frac{\partial p_{12}}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} \lambda^2 & 2\phi\lambda \\ \lambda^2 & 1-2\phi\lambda \end{pmatrix},$$

and thus

$$|\mathbf{J}| = \lambda^2.$$

Therefore, the reference prior for (ϕ, λ) is

$$\begin{aligned}
\pi(\phi, \lambda) &\propto (\phi\lambda^2)^{-\frac{1}{2}}(1-\phi\lambda^2)^{-\frac{1}{2}}[\phi(1-\phi\lambda)]^{-\frac{1}{2}}[1-\phi\lambda^2-\phi(1-\phi\lambda)]^{-\frac{1}{2}}\lambda^2 \\
&\propto \phi^{-\frac{1}{2}}\lambda^{\frac{1}{2}}(1-\phi\lambda^2)^{-\frac{1}{2}}(1-\phi\lambda)^{-\frac{1}{2}}(1-\lambda)^{-\frac{1}{2}}.
\end{aligned}$$

Since,

$$f(\mathbf{x} \mid \mathbf{p}) = \begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} p_{11}^{x_{11}} p_{12}^{x_{12}} (1 - p_{11} - p_{12})^{n - x_{11} - x_{12}},$$

the conditional probability density function of $\mathbf{x}$ given (ϕ, λ) becomes

$$f(\mathbf{x} \mid \phi, \lambda) = \begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} \phi^{x_{11}} \lambda^{2x_{11} + x_{12}} (1 - \phi\lambda)^{x_{12}} (1 - \lambda)^{n - x_{11} - x_{12}}.$$

Now, combining the conditional probability density function $f(\mathbf{x} \mid \phi, \lambda)$ and the reference prior $\pi(\phi, \lambda)$, the posterior distribution of (ϕ, λ) is

$$\pi(\phi, \lambda \mid \mathbf{x}) \propto \phi^{x_{11} - \frac{1}{2}} \lambda^{2x_{11} + x_{12} + \frac{1}{2}} (1 - \lambda)^{n - x_{11} - x_{12} - \frac{1}{2}} (1 - \phi\lambda)^{x_{12} - \frac{1}{2}} (1 - \phi\lambda^2)^{-\frac{1}{2}}. \tag{3.5}$$

3.4 Intrinsic discrepancy loss

By Section 2.3.2, the model is $\mathcal{M} = \{p(\mathbf{x}|\omega, \lambda), \mathbf{x} \in \mathbf{X}, \omega \in \mathbf{\Omega}, \lambda \in \mathbf{\Lambda}\}$, then the objective Bayesian criterion to the decision problem for the hypothesis testing $\{H_0 : \omega = \omega_0\}$ vs $\{H_1 : \omega \neq \omega_0\}$ is given by formula (2.22) and the point estimation of ω is given by formula (2.23). Since we want to test $\{H_0 : \phi = \phi_0\}$ vs $\{H_1 : \phi \neq \phi_0\}$ and estimate the parameter ϕ, then the intrinsic discrepancy loss $\delta(\phi_0, \phi, \lambda)$ can be derived from formula (2.21).

First of all, we want to obtain the minimal from the following two equations $\min_{\lambda_0 \in \mathbf{\Lambda}} k\big(f(\mathbf{x} \mid \phi_0, \lambda_0) \mid f(\mathbf{x} \mid \phi, \lambda)\big)$ and $\min_{\lambda_0 \in \mathbf{\Lambda}} k\big(f(\mathbf{x} \mid \phi, \lambda) \mid f(\mathbf{x} \mid \phi_0, \lambda_0)\big)$.

Note that,

$$\begin{aligned} & k\big(f(\mathbf{x} \mid \phi_0, \lambda_0) \mid f(\mathbf{x} \mid \phi, \lambda)\big) \\ = & \sum_{x_{11}} \sum_{x_{12}} f(\mathbf{x} \mid \phi, \lambda) \log \frac{f(x_{11}, x_{12} \mid \phi, \lambda)}{f(x_{11}, x_{12} \mid \phi_0, \lambda_0)} \\ = & \sum_{x_{11}} \sum_{x_{12}} f(\mathbf{x} \mid \phi, \lambda) \log \frac{\begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} \phi^{x_{11}} \lambda^{2x_{11} + x_{12}} (1 - \phi\lambda)^{x_{12}} (1 - \lambda)^{n - x_{11} - x_{12}}}{\begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} \phi_0^{x_{11}} \lambda_0^{2x_{11} + x_{12}} (1 - \phi_0\lambda_0)^{x_{12}} (1 - \lambda_0)^{n - x_{11} - x_{12}}}, \end{aligned}$$

where $x_{11} + x_{12} \leq n$, $\phi > 0$, $0 < \lambda < 1$, $0 < \phi\lambda < 1$, $\phi_0 > 0$, $0 < \lambda_0 < 1$ and $0 < \phi_0\lambda_0 < 1$. Since $\mathrm{E}(X_j) = n\theta_j$, it follows

$$\mathrm{E}(X_{11}) = np_{11} = n\phi\lambda^2,$$
$$\mathrm{E}(X_{12}) = np_{12} = n\lambda(1-\phi\lambda),$$

and thus $k\big(f(\mathbf{x} \mid \phi_0, \lambda_0) \mid f(\mathbf{x} \mid \phi, \lambda)\big)$ can be simplified as

$$n\phi\lambda^2 \log \frac{\phi}{\phi_0} + n\lambda(1-\phi\lambda) \log \frac{1-\phi\lambda}{1-\phi_0\lambda_0} + (n\phi\lambda^2 + n\lambda) \log \frac{\lambda}{\lambda_0} + (n - n\lambda) \log \frac{1-\lambda}{1-\lambda_0}. \tag{3.6}$$

Moreover, we want to obtain the value for $\min_{\lambda_0 \in \mathbf{\Lambda}} k\big(f(\mathbf{x} \mid \phi_0, \lambda_0) \mid f(\mathbf{x} \mid \phi, \lambda)\big)$. Let

$$f(\lambda_0) = n\phi\lambda^2 \log \frac{\phi}{\phi_0} + n\lambda(1-\phi\lambda) \log \frac{1-\phi\lambda}{1-\phi_0\lambda_0} + (n\phi\lambda^2 + n\lambda) \log \frac{\lambda}{\lambda_0} + (n - n\lambda) \log \frac{1-\lambda}{1-\lambda_0}. \tag{3.7}$$

Therefore,

$$\begin{aligned} f'(\lambda_0) &= n\lambda(1-\phi\lambda)\frac{\phi_0}{1-\phi_0\lambda_0} - (n\phi\lambda^2 + n\lambda)\frac{1}{\lambda_0} + (n-n\lambda)\frac{1}{1-\lambda_0} \\ &= -\frac{n}{(1-\phi_0\lambda_0)\lambda_0(1-\lambda_0)}[\phi_0(\lambda+1)\lambda_0^2 - (2\phi_0\lambda + \phi\lambda^2 + 1)\lambda_0 + \phi\lambda^2 + \lambda]. \end{aligned}$$

It is easy to see that

$$\begin{aligned} \Delta &= (2\phi_0\lambda + \phi\lambda^2 + 1)^2 - 4\phi_0(\lambda+1)(\phi\lambda^2 + \lambda) \\ &= (1-\lambda^2)(1-\lambda^2\phi^2) + 4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2 > 0, \end{aligned}$$

so the two roots of $f'(\lambda_0) = 0$ are

$$\begin{aligned} \lambda_0^{(1)} &= \frac{2\phi_0\lambda + \phi\lambda^2 + 1 - \sqrt{(2\phi_0\lambda + \phi\lambda^2 + 1)^2 - 4\phi_0(\lambda+1)(\phi\lambda^2 + \lambda)}}{2\phi_0(\lambda+1)} \\ &= \frac{2\phi_0\lambda + \phi\lambda^2 + 1 - \sqrt{(1-\lambda^2)(1-\lambda^2\phi^2) + 4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2}}{2\phi_0(\lambda+1)}, \\ \lambda_0^{(2)} &= \frac{2\phi_0\lambda + \phi\lambda^2 + 1 + \sqrt{(2\phi_0\lambda + \phi\lambda^2 + 1)^2 + 4\phi_0(\lambda+1)(\phi\lambda^2 + \lambda)}}{2\phi_0(\lambda+1)} \\ &= \frac{2\phi_0\lambda + \phi\lambda^2 + 1 + \sqrt{(1-\lambda^2)(1-\lambda^2\phi^2) + 4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2}}{2\phi_0(\lambda+1)}. \end{aligned}$$

Now, we will show $\lambda_0^{(1)} < 1$ and $\lambda_0^{(1)} < \frac{1}{\phi_0}$.

Firstly, let

$$Q = \Big(\sqrt{(1-\lambda^2)(1-\lambda^2\phi^2) + 4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2}\Big)^2 - (2\phi_0\lambda + \phi\lambda^2 + 1 - 2\phi_0(\lambda+1))^2.$$

Since

$$\begin{aligned} Q &= (1-\lambda^2)(1-\lambda^2\phi^2) + 4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2 - (\phi\lambda^2 + 1 - 2\phi_0)^2 \\ &= 4\phi_0(\lambda+1)(1-\phi_0)(1-\lambda) > 0, \end{aligned}$$

then

$$\Big(\sqrt{(1-\lambda^2)(1-\lambda^2\phi^2) + 4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2}\Big)^2 > (2\phi_0\lambda + \phi\lambda^2 + 1 - 2\phi_0(\lambda+1))^2,$$

$$2\phi_0\lambda + \phi\lambda^2 + 1 - \sqrt{(1-\lambda^2)(1-\lambda^2\phi^2) + 4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2} < 2\phi_0(\lambda+1)),$$

and therefore $\lambda_0^{(1)} < 1$.

Secondly, if $\phi_0 > \frac{\phi+1}{2}$, then

$$\begin{aligned} \lambda_0^{(1)} &< \frac{2\phi_0\lambda + \phi\lambda^2 + 1 - \sqrt{4\lambda^2(\phi_0 - \frac{\phi+1}{2})^2}}{2\phi_0(\lambda+1)} \\ &= \frac{2\phi_0\lambda + \phi\lambda^2 + 1 - 2\lambda(\phi_0 - \frac{\phi+1}{2})}{2\phi_0(\lambda+1)} \\ &= \frac{(1+\phi\lambda)(1+\lambda)}{2\phi_0(\lambda+1)} \\ &< \frac{1}{\phi_0}. \end{aligned}$$

If $\phi_0 \leq \frac{\phi+1}{2}$, then

$$\begin{aligned} \lambda_0^{(1)} &< \frac{(\phi+1)\lambda + \phi\lambda^2 + 1}{2\phi_0(\lambda+1)} \\ &< \frac{(1+\phi\lambda)(1+\lambda)}{2\phi_0(\lambda+1)} \\ &< \frac{1}{\phi_0}. \end{aligned}$$

Therefore,

$$\lambda_0^{(1)} < \min(\lambda_0^{(2)}, \frac{1}{\phi_0}, 1).$$

Thus, $\lambda_0^{(1)}$ is a limit point for $f(\lambda_0)$ and

$$\min_{\lambda_0 \in \mathbf{\Lambda}} k\big(f(\mathbf{x} \mid \phi_0, \lambda_0) \mid f(\mathbf{x} \mid \phi, \lambda)\big) = \min\{f(\lambda_0^{(1)}), f\big(\min(1, \frac{1}{\phi_0})\big)\}. \tag{3.8}$$

In addition,

$$\begin{aligned}
& k\big(f(\mathbf{x} \mid \phi, \lambda) \mid f(\mathbf{x} \mid \phi_0, \lambda_0)\big) \\
= & \sum_{x_{11}} \sum_{x_{12}} f(\mathbf{x} \mid \phi_0, \lambda_0) \log \frac{f(x_{11}, x_{12} \mid \phi_0, \lambda_0)}{f(x_{11}, x_{12} \mid \phi, \lambda)} \\
= & \sum_{x_{11}} \sum_{x_{12}} f(\mathbf{x} \mid \phi_0, \lambda_0) \log \frac{\begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} \phi_0^{x_{11}} \lambda_0^{2x_{11}+x_{12}} (1-\phi_0\lambda_0)^{x_{12}} (1-\lambda_0)^{n-x_{11}-x_{12}}}{\begin{pmatrix} n \\ x_{11}, x_{12} \end{pmatrix} \phi^{x_{11}} \lambda^{2x_{11}+x_{12}} (1-\phi\lambda)^{x_{12}} (1-\lambda)^{n-x_{11}-x_{12}}} \\
= & n\phi_0\lambda_0^2 \log \frac{\phi_0}{\phi} + n\lambda_0(1-\phi_0\lambda_0) \log \frac{1-\phi_0\lambda_0}{1-\phi\lambda} + (n\phi_0\lambda_0^2 + n\lambda_0) \log \frac{\lambda_0}{\lambda} + \\
& (n - n\lambda_0) \log \frac{1-\lambda_0}{1-\lambda},
\end{aligned}$$

where $x_{11} + x_{12} \leq n$, $\phi > 0$, $0 < \lambda < 1$, $0 < \phi\lambda < 1$, $\phi_0 > 0$, $0 < \lambda_0 < 1$, $0 < \phi_0\lambda_0 < 1$ and

$$\mathrm{E}(X_{11}) = np_{11} = n\phi_0\lambda_0^2,$$
$$\mathrm{E}(X_{12}) = np_{12} = n\lambda_0(1-\phi_0\lambda_0).$$

Similarly, we want to obtain the value for $\min_{\lambda_0 \in \mathbf{\Lambda}} k\big(f(\mathbf{x} \mid \phi, \lambda) \mid f(\mathbf{x} \mid \phi_0, \lambda_0)\big)$. Let

$$\begin{aligned}
g(\lambda_0) \quad = \quad & n\phi_0\lambda_0^2 \log \frac{\phi_0}{\phi} + n\lambda_0(1-\phi_0\lambda_0) \log \frac{1-\phi_0\lambda_0}{1-\phi\lambda} + (n\phi_0\lambda_0^2 + n\lambda_0) \log \frac{\lambda_0}{\lambda} + \\
& (n - n\lambda_0) \log \frac{1-\lambda_0}{1-\lambda}.
\end{aligned} \tag{3.9}$$

Then

$$\begin{aligned} g'(\lambda_0) &= 2n\phi_0\lambda_0 \log\frac{\phi_0}{\phi} + (n - 2n\phi_0\lambda_0)\log\frac{1-\phi_0\lambda_0}{1-\phi\lambda} - n\phi_0\lambda_0 + (n + 2n\phi_0\lambda_0)\log\frac{\lambda_0}{\lambda} + \\ &\quad n + n\phi_0\lambda_0 - n\log\frac{1-\lambda_0}{1-\lambda} - n \\ &= 2n\phi_0\lambda_0 \log\frac{\phi_0}{\phi} + (n - 2n\phi_0\lambda_0)\log\frac{1-\phi_0\lambda_0}{1-\phi\lambda} + (n + 2n\phi_0\lambda_0)\log\frac{\lambda_0}{\lambda} - n\log\frac{1-\lambda_0}{1-\lambda} \\ &= n[2\phi_0\lambda_0 \log\frac{\phi_0\lambda_0(1-\phi\lambda)}{\phi\lambda(1-\phi_0\lambda_0)} + \log\frac{\lambda_0(1-\phi_0\lambda_0)(1-\lambda)}{\lambda(1-\phi\lambda)(1-\lambda_0)}]. \end{aligned}$$

Now, we observe the formula of $g'(\lambda_0)$ is too complex for us to find a theoretical solution of the limit point for $g(\lambda_0)$. However, if ϕ, ϕ_0 and λ are given, then we can use R to calculate the roots of $g'(\lambda_0) = 0$, where they are all the limit points for $g(\lambda_0)$ and the domain for λ_0 is $\big(0, \min(1, \frac{1}{\phi_0})\big)$. Hence, we can find $\min_{\lambda_0 \in \boldsymbol{\Lambda}} k\big(f(\mathbf{x} \mid \phi, \lambda) \mid f(\mathbf{x} \mid \phi_0, \lambda_0)\big)$ through these points.

Therefore, we derive the intrinsic discrepancy loss function $\delta(\phi_0, \phi, \lambda)$ by comparing the values of $\min_{\lambda_0 \in \boldsymbol{\Lambda}} k\big(f(\mathbf{x} \mid \phi_0, \lambda_0) \mid f(\mathbf{x} \mid \phi, \lambda)\big)$ and $\min_{\lambda_0 \in \boldsymbol{\Lambda}} k\big(f(\mathbf{x} \mid \phi, \lambda) \mid f(\mathbf{x} \mid \phi_0, \lambda_0)\big)$. In addition, from the posterior distribution $\pi(\phi, \lambda)$ and the intrinsic discrepancy $\delta(\phi_0, \phi, \lambda)$, we can obtain the objective Bayesian criterion to the decision problem for the hypothesis testing $\{H_0 : \phi = \phi_0\}$: We will reject H_0 if and only if

$$d(\phi_0, \mathbf{x}) = \int_{\boldsymbol{\Phi}} \int_{\boldsymbol{\Lambda}} \delta\{\phi_0, (\phi, \lambda)\} \pi(\phi, \lambda \mid \mathbf{x}) d\lambda d\phi > d_0, \tag{3.10}$$

and the formula of the point estimation problem is

$$\phi^*(\mathbf{x}) = \text{arginf}_{\phi_0 \in \boldsymbol{\Phi}} d(\phi_0, \mathbf{x}), \tag{3.11}$$

where ϕ_0 is the parameter that is used to act as a proxy to the parameter ϕ.

3.5 Modify objective Bayesian criterion

Since the intrinsic discrepancy loss $\delta(\phi_0, \phi, \lambda)$ is too complex to find a theoretical value, then by "*Law of Large Numbers for Markov Chain*", we could obtain an approximate value for

$\delta(\phi_0, \phi, \lambda)$ as $m \to \infty$, which the formula is

$$\frac{\sum_{j=1}^{m} \delta_j(\phi_0, \phi^j, \lambda^j)}{m} \to d(\phi_0, \mathbf{x}), \tag{3.12}$$

where ϕ^j, λ^j are generated by the given data $\mathbf{x}$.

Therefore, using the approximate value for $\delta(\phi_0, \phi, \lambda)$, we modify the objective Bayesian criterion as the decision problem, which we reject H_0 if and only if

$$d(\phi_0, \mathbf{x}) \approx \frac{\sum_{j=1}^{m} \delta_j(\phi_0, \phi^j, \lambda^j)}{m} > d_0. \tag{3.13}$$

Also, the formula of the point estimation problem is modified as

$$\phi^*(\mathbf{x}) \approx \text{arginf}_{\phi_0 \in \mathbf{\Phi}} d(\phi_0, \mathbf{x}). \tag{3.14}$$

Chapter 4

Simulation Study

4.1 Hypothesis testing

In this section, we first present a simulation for the hypothesis testing $\{H_0 : \phi = \phi_0\}$ vs $\{H_1 : \phi \neq \phi_0\}$. We illustrate this simulation on the model

$$\mathcal{M} = \{p(\mathbf{x}|\phi, \lambda), \mathbf{x} \in \mathbf{X}, \phi \in \mathbf{\Phi}.\lambda \in \mathbf{\Lambda}\}. \tag{4.1}$$

The propose of this simulation is to obtain the percentage of acceptance for the null hypothesis $\{H_0 : \phi = \phi_0\}$. This simulation is shown as following four steps.

Step 1: Fix p_{11}, p_{12}, we generate a data $\mathbf{x}$ from the probability density function $f(\mathbf{x} \mid p_{11}, p_{12})$, where $\mathbf{x} \mid p_{11}, p_{12} \sim \text{Mul}(n, p_{11}, p_{12})$ and $\mathbf{x} = (x_{11}, x_{12})$. Here, we use $f(\mathbf{x} \mid p_{11}, p_{12})$ to generate the data $\mathbf{x}$, because $f(\mathbf{x} \mid p_{11}, p_{12})$ is a one to one function to $f(\mathbf{x} \mid \phi, \lambda)$.

Step 2: For each $j \in \{1, 2, \ldots, m\}$, we generate the parameters $(p_{11}^{(j)}, p_{12}^{(j)})$ from the posterior distributions $\pi(p_{11}^{(j)} \mid \mathbf{x})$ and $\pi(p_{12}^{(j)} \mid p_{11}, \mathbf{x})$ for the data $\mathbf{x}$. Then, we can obtain the parameters $(\phi^{(j)}, \lambda^{(j)})$ from the generated parameters $(p_{11}^{(j)}, p_{12}^{(j)})$ by the method of transformation. After that, let's substitute the parameters $(\phi^{(j)}, \lambda^{(j)})$ into the formulas (3.8) and (3.9) to get the value of $\lambda_0^{(j)}$. Also, the value of the intrinsic discrepancy loss $\delta_j(\phi_0, \phi, \lambda)$ can be obtained by substituting the $\lambda_0^{(j)}$

into the formulas (3.7) and (3.9). Here, we claim that

$$\begin{aligned} p_{11} \mid \mathbf{x} &\sim \text{Beta}\Big(x_{11}+\frac{1}{2}, n-x_{11}+\frac{1}{2}\Big), &(4.2)\\ p_{12} \mid p_{11}, \mathbf{x} &\sim (1-p_{11})\text{Beta}\Big(x_{12}+\frac{1}{2}, n-x_{11}-x_{12}+\frac{1}{2}\Big). &(4.3)\end{aligned}$$

Now, we prove the above formulas (4.2) and (4.3). From the reference prior function of (p_{11}, p_{12}) by formula (3.4), we can obtain the posterior distribution for (p_{11}, p_{12}), which is

$$\pi(p_{11}, p_{12} \mid \mathbf{x}) \propto p_{11}^{x_{11}-\frac{1}{2}}(1-p_{11})^{-\frac{1}{2}} p_{12}^{x_{12}-\frac{1}{2}}(1-p_{11}-p_{12})^{n-x_{11}-x_{12}-\frac{1}{2}}. \tag{4.4}$$

Since

$$\begin{aligned} \pi(p_{11}, p_{12} \mid \mathbf{x}) &= \pi(p_{11} \mid \mathbf{x})\pi(p_{12} \mid \mathbf{x}, p_{11}),\\ \pi\Big(\frac{p_{12}}{1-p_{11}} \mid \mathbf{x}, p_{11}\Big) &\propto \Big(\frac{p_{12}}{1-p_{11}}\Big)^{x_{12}-\frac{1}{2}}\Big(1-\frac{p_{12}}{1-p_{11}}\Big)^{n-x_{11}-x_{12}-\frac{1}{2}},\\ \pi(p_{11} \mid \mathbf{x}) &= \int_0^{1-p_{11}} \pi(p_{11}, p_{12} \mid \mathbf{x}) dp_{12}\\ &\propto p_{11}^{x_{11}-\frac{1}{2}}(1-p_{11})^{-\frac{1}{2}} \int_0^{1-p_{11}} p_{12}^{x_{12}-\frac{1}{2}}(1-p_{11}-p_{12})^{n-x_{11}-x_{12}-\frac{1}{2}} dp_{12}\\ &\propto p_{11}^{x_{11}-\frac{1}{2}}(1-p_{11})^{n-x_{11}-\frac{1}{2}}, \end{aligned}$$

we have

$$\begin{aligned} \pi(p_{11} \mid \mathbf{x}) &\sim \text{Beta}\Big(x_{11}+\frac{1}{2}, n-x_{11}+\frac{1}{2}\Big),\\ \pi\Big(\frac{p_{12}}{1-p_{11}} \mid p_{11}, \mathbf{x}\Big) &\sim \text{Beta}\Big(x_{12}+\frac{1}{2}, n-x_{11}-x_{12}+\frac{1}{2}\Big). \end{aligned}$$

Step 3: We substitute the values of m groups for the intrinsic discrepancy loss $\delta_j(\phi_0, \phi, \lambda)$ into the formula (3.13). Then we can check whether d_0 has a bigger value by comparing the value of $d(\phi_0, \mathbf{x})$ with d_0. In addition, let the constant $d_0 = 2.5$ and r be the counter of index. If $d(\phi_0, \mathbf{x}) < 2.5$, $r = 1$. Otherwise, $r = 0$.

Step 4: By repeating the above process for w times, we can find the percentage of acceptance for the null hypothesis $\{H_0 : \phi = \phi_0\}$ by z/w, where z is the total for the counter of index r and $w = 1000$.

However, we can only obtain one value for the percentage of acceptance for the null hypothesis $\{H_0 : \phi = \phi_0\}$ from above four steps. If we fix p_{11}, then we can get a group of values from different p_{12} values by repeating this process, since the value of ϕ also varies.

4.2 Point estimation

Now, we present the simulation example for the point estimation by applying the same model

$$\mathcal{M} = \{p(\mathbf{x}|\phi, \lambda), \mathbf{x} \in \mathbf{X}, \phi \in \mathbf{\Phi}.\lambda \in \mathbf{\Lambda}\}.$$

The aim of this simulation is to find the best estimator for the parameter ϕ. Similarly, we start this simulation by following the same procedures as we did for step 1 and step 2 of hypothesis testing in Section 4.1.

Step 3: We fix an interval of the value for ϕ_0. Let the both end points from this interval be the minimal and maximal from the parameters ϕ^j we generated, where $j \in \{1, 2, \ldots, m\}$. After that, we divide this interval into s subintervals. Therefore, we scale the interval by using s+1 numbers of values for ϕ_0 in order.

Step 4: We can find the best estimator for ϕ by using the formula (3.14).

Step 5: By repeating the above process for w times, we use the mean of the w groups to obtain the best estimator for ϕ, where $w = 1000$.

Furthermore, by fixing the value of p_{11} and let p_{12} be difference, we can find a group best estimators for the different values of parameter ϕ.

4.3 Numerical tests

Firstly, we proceed on simulation study to examine our methods. Following the above procedures, we generate data when p_{11} and ϕ_0 are set to be different values. The following plots in Figure 4.1 and Figure 4.2 each displays how the percentage of acceptance changes along with the underlying ϕ value or p_{12} value. Each point in the plots represents the result from $w = 1000$.

The plots are displayed in pairs. Graphs in a pair represent the results from the same set of data, but is plotted against different arguments, for example ϕ and p_{12}, so as to give a clearer view of the changes. Note that the graphs in a pair are hence equivalent. This can be explained by the one-to-one relationship between ϕ and p_{12} (see Figure 4.3). Among different pairs for the values of p_{11} and(or) ϕ_0 are different. In Figure 4.1, $\phi_0 = 0.8$, where p_{11} are 0.2, 0.3 and 0.5 in the 3 pairs. In Figure 4.2, $\phi_0 = 1.5$, where p_{11} are 0.2, 0.3 and 0.5 in the 3 pairs.

We can see that, for the set ϕ_0, the graph becomes "fatter" as p_{11} increase. This implies the percentage of acceptance decreases slower as p_{11} increase. Besides, for the set ϕ_0 and p_{11}, the percentage of acceptance becomes higher as the value of ϕ getting closer to the value of ϕ_0.

Also, the results of point estimation of ϕ are shown below in Figure 4.4. Three graphs are plotted corresponding to p_{11} are 0.2, 0.3 and 0.5. Each of these graphs displays the comparison between the estimated value and the real value of ϕ against varying p_{12}. Here, we use dot '.' to present the real value and '$\times$' to present the estimated value. We can see that the estimated curve fits the real one pretty well.

Secondly, refer the example from Agresti (1990), we take a real data objective Bayesian hypothesis testing $\{H_0 : \phi = \phi_0\}$ vs $\{H_1 : \phi \neq \phi_0\}$ and point estimation for the parameter ϕ. For that problem, the total number of trials is 156, $x_{11} = 30$, $x_{12} = 63$, $X_{22} = 63$. Table 4.1 is the result of the hypothesis testing and Table 4.2 is the result of the point estimation.

ϕ_0	ϕ	**Accept**$H_0(\%)$
0.5411	0.5411	0.953
0.5411	0.2138	0.126
0.5411	0.1015	0.062
0.5411	1.1365	0.006

Table 4.1: Hypothesis testing for RR of real data

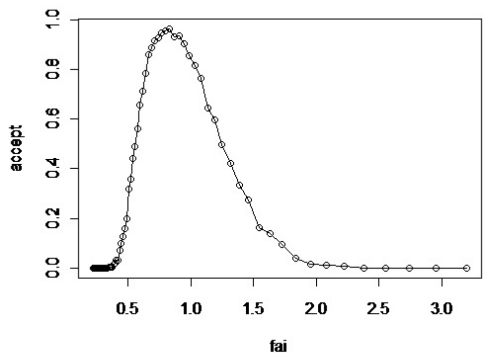
Method	**Estimator of RR**
Tail	0.5418
Score	0.5411
Objective prior	0.5388

Table 4.2: Point estimation for RR of real data

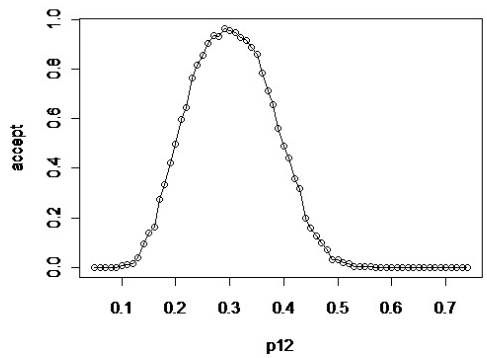
4.4 Summary

From the Figure 4.1, Figure 4.2, let p_{11} be fixed, we can see that the percentage of acceptance for the hypothesis testing $\{H_0 : \phi = \phi_0\}$ vs $\{H_1 : \phi \neq \phi_0\}$ exceeds 0.95 either when the value of ϕ is close to the value of ϕ_0, or when the value of p_{12} is close to the value of $\left(\frac{p_{11}^2}{\phi_0} - p_{11}\right)$. In addition, from the Figure 4.3, we can see the ϕ and p_{12} are one-to-one. Moreover, from the Figure 4.4, we can see that the estimated curve fits the real one pretty well. For the result in Table 4.1, it states the same situation which we mentioned above. Table 4.2 shown the result of real data comparison from three methods, and it seems that all the estimations for RR are similar.

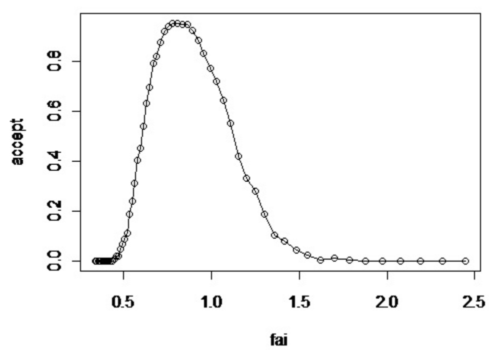
Furthermore, as to the future work, I will try to use the objective Bayesian prior function to take a credible interval estimation of the risk ratio in a correlated 2×2 table with structural zero, which can compare the other results for the different methods of the interval estimation's problem.



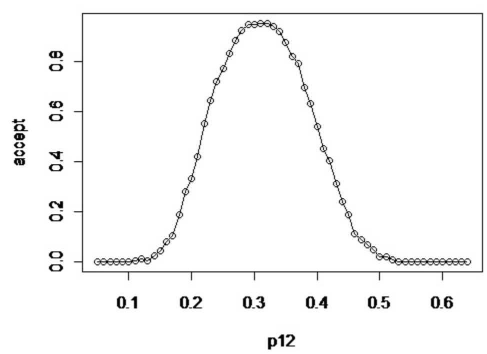
(a) $p_{11} = 0.2, \phi_0 = 0.8$



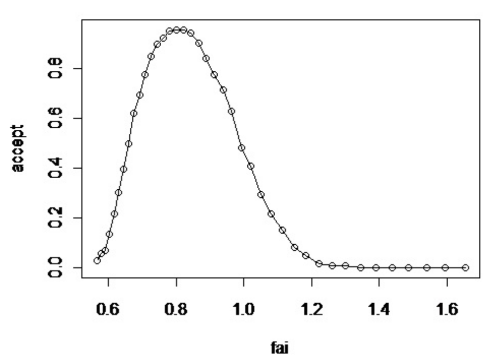
(b) $p_{11} = 0.2, \phi_0 = 0.8$



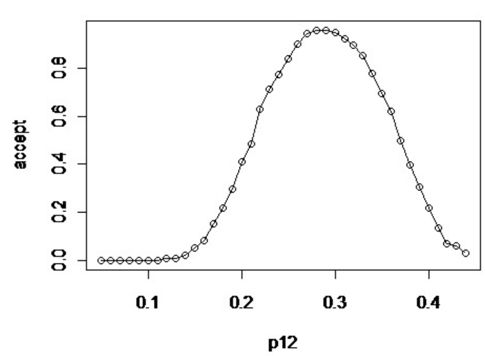
(c) $p_{11} = 0.3, \phi_0 = 0.8$



(d) $p_{11} = 0.3, \phi_0 = 0.8$

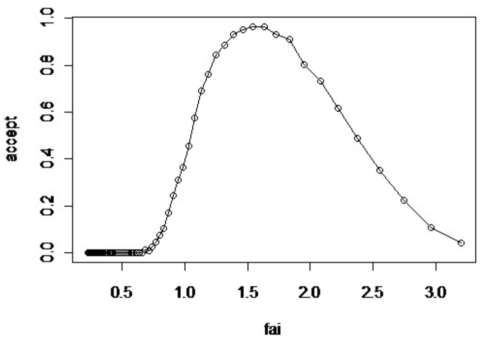


(e) $p_{11} = 0.5, \phi_0 = 0.8$

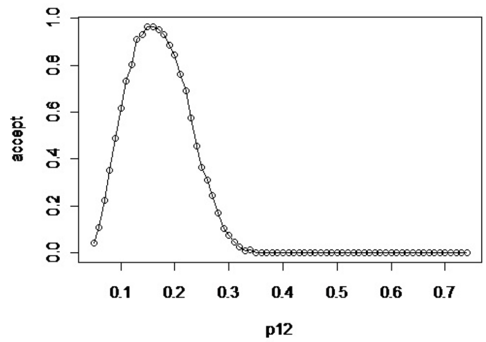


(f) $p_{11} = 0.5, \phi_0 = 0.8$

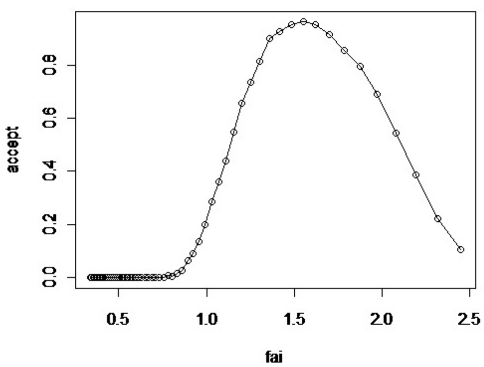
Figure 4.1



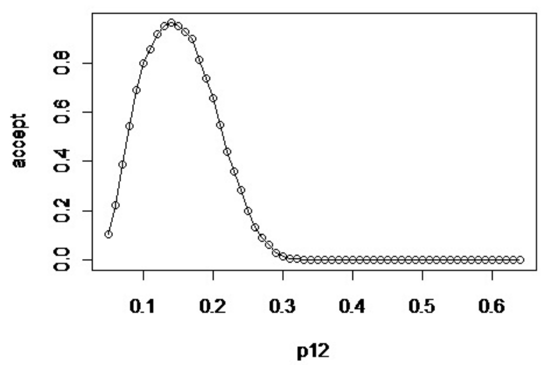
(a) $p_{11} = 0.2, \phi_0 = 1.5$



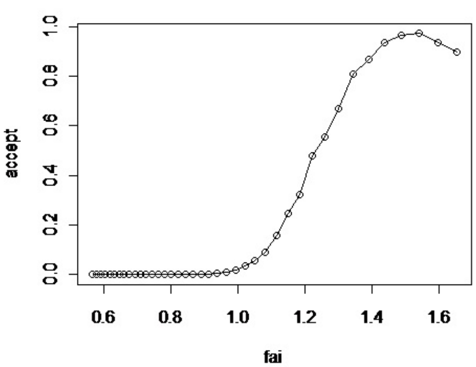
(b) $p_{11} = 0.2, \phi_0 = 1.5$



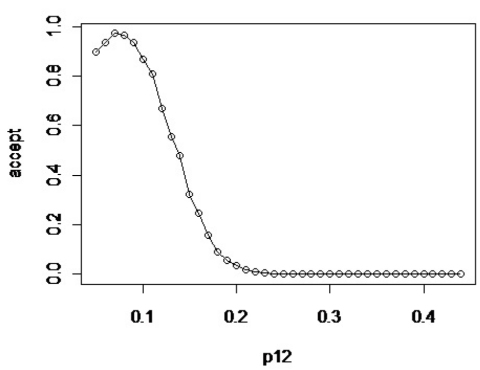
(c) $p_{11} = 0.3, \phi_0 = 1.5$



(d) $p_{11} = 0.3, \phi_0 = 1.5$

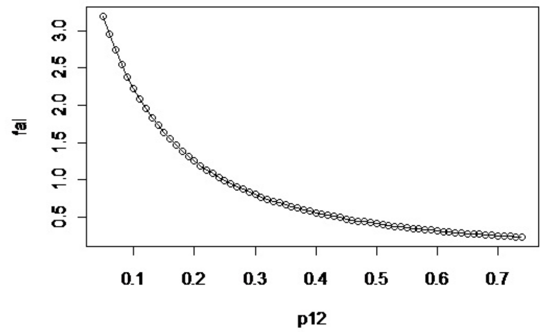


(e) $p_{11} = 0.5, \phi_0 = 1.5$

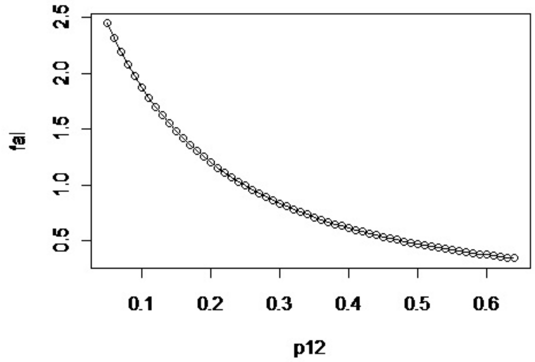


(f) $p_{11} = 0.5, \phi_0 = 1.5$

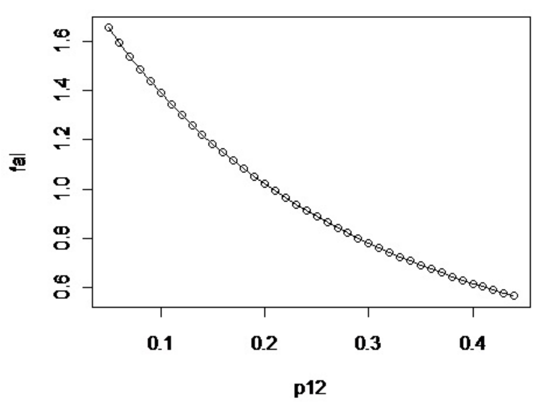
Figure 4.2



(a) $p_{11} = 0.2$

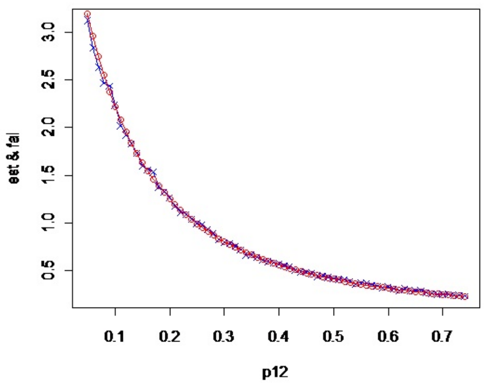


(b) $p_{11} = 0.3$

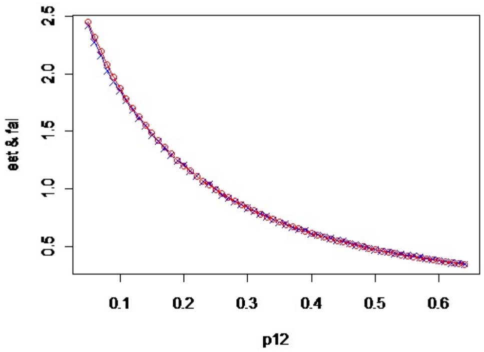


(c) $p_{11} = 0.5$

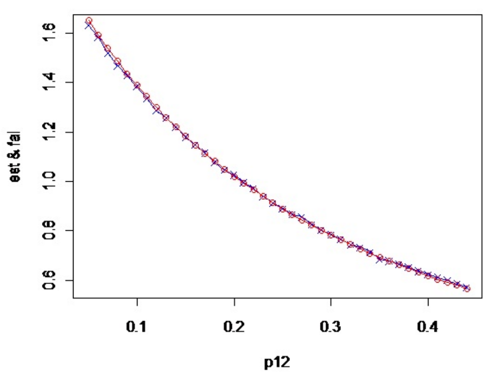
Figure 4.3



(a) $p_{11} = 0.2$



(b) $p_{11} = 0.3$



(c) $p_{11} = 0.5$

Figure 4.4

Bibliography

[1] A. Agresti. *An introduction to categorical data analysis*, volume 423. Wiley-Interscience, 2007.

[2] J.M. Bernardo. Integrated objective bayesian estimation and hypothesis testing. *Bayesian Statistics*, 9:1–68, 2011.

[3] J.M. Bernardo and R. Rueda. Bayesian hypothesis testing: A reference approach. *International Statistical Review*, 70(3):351–372, 2002.

[4] G. Casella and R.L. Berger. Statistical inference. 2001.

[5] J.K. Ghosh, Mohan. Delampady, and Tapas. Samanta. *An introduction to Bayesian analysis*. Springer New York, 2006.

[6] K.J. Lui. Interval estimation of the risk ratio between a secondary infection, given a primary infection, and the primary infection. *Biometrics*, pages 706–711, 1998.

[7] L. Shi, H. Sun, and P. Bai. Bayesian confidence interval for difference of the proportions in a 2×2 table with structural zero. *Journal of Applied Statistics*, 36(5):483–494, 2009.

[8] J.D. Stamey, J.W. Seaman, and D.M. Young. Bayesian inference for a correlated 2×2 table with a structural zero. *Biometrical journal*, 48(2):233–244, 2006.

[9] N.S. Tang and M.L. Tang. Exact unconditional inference for risk ratio in a correlated 2×2 table with structural zero. *Biometrics*, 58(4):972–980, 2002.