Learning for Contingency Tables and Survival Data Using Imprecise Probabilities

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

in Partial Fulfillment of the Requirements

for the degree of Doctor of Philosophy

in the Department of Mathematics and Statistics

University of Saskatchewan

Saskatoon

By

Naeima Abdallah Ashleik

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mathematics and Statistics
University of Saskatchewan
142 McLean Hall, 106 Wiggins Road
Saskatoon, Saskatchewan S7N 5E6 Canada
OR
Dean College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

Bayesian inference is a method of statistical inference in which all forms of uncertainty are expressed in terms of probability. Classical Bayesian inference has some limitations. One of these situations is when we have little to no information about the experiment; another situation is when we have computational or time limitations. Also problematic is a situation where there are conflicts in choosing a prior distribution where we have experts giving different prior information, which results in less precise posterior probabilities. Because of these limitations, imprecise Bayesian approach takes place in Bayesian inference.

Upper and lower posterior expectations are computed in order to calculate the degree of imprecision of the log-odds ratio. This is implemented in two-way contingency tables and then generalized to three-way tables by using different families of prior distributions, is which the core of this work. Survival data including right-censored observations are generated and converted to a sequence of 2×2 tables, three-way contingency tables, each 2×2 is built at each observed death time. Here, we assume only one death happens at each time and no ties. To implement imprecise Bayesian inference, two choices of imprecise priors are chosen. A set of four Normal priors and a set of four Beta priors are used with a non-central hypergeometric likelihood to update the posterior families and then the degree of imprecision is calculated for both cases. An example of real data is applied on Ovarian Cancer Survival data where upper and lower posterior expectations are estimated in order to calculate the degree of imprecision.

We conduct simulation studies to sample from posterior distribution and estimate the log-odds ratio by using upper and lower posterior expectations. In the situation of three-way contingency tables, updating a set of priors to a set of posterior is done sequentially at each table by running MCMC method through using JAGS from `R` via `rjags` and `runjags` packages. Also, four factors (sample size, censoring rate, true parameter, and balancing rate) are studied to see how these four factors affect the degree of imprecision with the two choices of imprecise priors. A fractional factorial design of 27 runs is constructed to see which one

of these four factors is more significant. For each one of these 27 combination, upper and lower posterior expectations and the degree of imprecision of the log-odds ratio are calculated.

The findings show that the smallest value of the degree of imprecision appears at the combination where the sample size is large ($n = 200$) and small number of censored times. In contrast, the largest value of the degree of imprecision is observed at the combination where the sample size is small ($n = 40$) and large number of censored times. These conclusions are supported by the findings of ANOVA that show that main effects of the four factors are significant. The conclusion that can be summarized from the results of this work is having more information (more data) leads to less uncertainty about the parameter of interest.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Dr. M. G. Bickis for the continuous advice, patient guidance and support. His guidance helped me in all the time of research and writing of this thesis. This work could not be possible without his support and encouragement.

Besides my advisor, I would like to thank Dr. Juxin Liu for her co-supervision for a period of time during my Ph.D work. Her assistance and encouragement motivated and helped me during my comprehensive exam and computational part of my dissertation.

My sincere thanks also goes to all the committee members: Dr. Shahedul Khan, Dr. Longhai Li , Dr. Chris Soteros, and Dr. Eric Neufeld for sharing their invaluable comments and suggestions, which made the presentation of the thesis more clear.

Most importantly, I am deeply indebted to my family, who missed me back home and counted down every single day throughout the period of study. Special thanks to my husband, Rafaa, for his continuing love, support, and encouragement throughout this period. I also thank my son, Issa, and my daughters, Fatima and Joud for all the happy moments I spent with them in my spare time.

I would like to thank the Ministry of Higher Education and Scientific Research and University of Benghazi, Libya, for providing me a scholarship for my Ph.D study.

Financial supports for attending conferences and SIPTA summer school from NSERC Discovery Grant and Dr. M. G. Bickis are appreciated.

I gratefully acknowledge the College of Graduate and Postdoctoral Studies, Department of Mathematics and Statistics, University of Saskatchewan for providing me funding for the last period of this work.

To my father, the memory of my mother,

my husband,

and those who educated me.

Contents

List of Tables

List of Figures

1. Introduction

Survival data analysis is a collection of statistical procedures where the outcome variable of interest is time until an event occurs, such as death in biological organisms or failure in mechanical systems. Usually, we refer to the time variable as survival time and the event as failure because the event of interest usually is death, disease incidence, or some other negative individual experience (Cox and Oakes, 1984, chapter 1). The survival time of a subject may not be observed for the full time to failure; in this case, the survival time is said to be censored. The comparison between survival times including censored observations of two treatments, has been done by summarizing the evidence about which treatment has longer survival time. One of the frequentist methods for solving this is the *log-rank test*. In this thesis we are assuming that we have two treatments (control and test) with two outcomes (alive and dead). This kind of data is survival data where the time of death can be observed or censored and this can be conveniently displayed in three-way *contingency tables*, a sequence of 2×2 tables, one at each time of observed death, (i.e., we have set of stratified 2×2 tables). This method of combining information over a number of 2×2 tables was proposed by Mantel and Haenszel (1959). The comparison between the two treatments in these contingency tables can be done by estimating the odds ratio of the two treatments (Mantel-Haenszel statistic). Therefore, the log odds ratio is our parameter of interest. In the Bayesian approach, our uncertainty about the parameter of interest can be modelled by a probability (prior) distribution.

Bayesian inference is a method of statistical inference in which all forms of uncertainty are expressed in terms of probability. Bayesians treat parameters as random variables and

define the probability as "degree of belief," which means that the probability of an event is the degree to which you believe that the event is true. One of most important objectives in Bayesian statistical inference is making inferences about parameters of interest, where prior information about these parameters is represented in probability distributions known as *prior distributions*. Combining prior distribution and data by multiplying the prior density by the likelihood (Bayes' theorem) is called updating the prior. By updating the prior, one can get the *posterior distribution*, which contains all the information about the parameters of interest after the data are examined. However, classical (precise) Bayesian inference has some limitations. According to Walley (1991, chapter 1), there are situations where it is difficult to assign a single probability distribution for the prior. One is when we have little to no information about the experiment; another situation is when we have computational or time limitations. Also problematic is a situation where there are conflicts in choosing prior distribution where we have experts giving different prior information, which results in less precise posterior probabilities. Because of these limitations with a precise Bayesian approach, this work proposes using *imprecise probabilities*.

1.1 Contributions

Choosing a prior distribution to represent the prior information about the parameters of interest is the feature of any precise Bayesian analysis. However, choosing prior distribution encounters some difficulties as we mentioned early. Because of these limitations of the precise Bayesian approach, using *imprecise probabilities* as an alternative approach is the purpose of this work.

The implementation of an imprecise Bayesian approach is the aim of this dissertation. The contribution will be introduced in the following steps:

- Determination of upper and lower posterior expectations is investigated for the log odds ratio in two-way contingency tables to calculate the degree of imprecision in two scenarios. First, imprecise Dirichlet model where a set of product of beta priors

is proposed and degree of imprecision of log-odds ratio is computed. Second, a re-parametrization of the multinomial distribution and logit model is defined in terms of the canonical parameter since the multinomial is a member of the exponential family, and redefine log-odds ratio in terms of this canonical parameter. In this approach, precise and imprecise Bayesian inference are applied to the log-odds ratio in two-way table and empirical CDFs of posterior samples of two cases are compared.

- To address the generalization from two-way to three-way contingency tables, survival data with right-censored observations is considered in the case that we have two groups (test and control) with two outcomes (alive and dead) represented in a sequence of 2×2 tables, one at each death. The main purpose of this aim is getting the upper and lower posterior expectations of the log-odds ratio in order to compute the degree of imprecision for each table and how it gets decreased as number of tables (deaths) increased (more information). To implement this, the log-rank test is constructed; In fact, under the null hypothesis of independence, the non-central hypergeometric distribution is the base of constructing the log-rank test. Also, re-parametrization of the odds ratio is assumed based on the feature that non-central hypergeometric distribution is a member of the exponential family.

A simulation study for each aim of this work is proposed and results are summarized in Sections 3.3.1 and 4.4 and Appendix A.

1.2 Thesis Organization

In Chapter 2, background information on probability models and parametrization for two-way and three way contingency tables is provided in Section 2.1. Section 2.1.2 represents fundamental knowledge about modelling cell counts in contingency tables and Section 2.1.3 reviews the existing literature on Bayesian inference for log-linear parameters in two-way tables and the choices of the prior distributions that have been used and how each choice affects the results of those studies. Section 2.2 provides a reasonably comprehensive overview of survival data modelling and analysis and this section ends with Bayesian inference in sur-

vival data. In Section 2.3, an overview of the issue of choices of prior distributions that have place in the literature on Bayesian inference and also this section provides a theoretical introduction to imprecise probability theory supported by some definitions and examples where it ends by a section on nonparametric predictive inference as it is considered as an imprecise approach.

In Chapter 3, in Section 3.2, an imprecise Dirichlet model where a set of product of beta priors is proposed and degree of imprecision of log-odds ratio is computed. Re-parametrization and Normal priors for log-odds ratio in two-way contingency tables with a simulation study are presented in Section 3.3.

In Chapter 4, a generalization from two-way to three-way contingency tables is addressed by considering survival data with right-censored observations. This kind of data is represented in a sequence of 2×2 tables, one at each death. In Section 4.2, non-central hypergeometric model is imposed log-odds ratio as the parameter of the model. Then, in Section 4.2.1, two choices of imprecise priors are chosen in this chapter, each as a set of four priors is given to the parameter of interest. In Section 4.3, an example of real data is applied on Ovarian Cancer Survival data where upper and lower posterior expectations are estimated in order to calculate the degree of imprecision. A simulation study using MCMC methods is done, and discussion of the results is stated in Section 4.4.

Conclusion and future work are included in Chapter 5, and complete results and graphs of the simulation in Chapter 4 are included in Appendix A, and `R` codes in Appendix B.

2. Background

2.1 Contingency Tables

In statistics, a contingency (cross-classification) table is a type of table in a matrix format that displays the frequency distribution of two (or more) discrete variables. For example, comparing two medical treatments with two outcomes (success, failure) in 2×2 tables, is very popular used in biomedical and social science applications. Two-way contingency tables are used where we are interested in quantifying the strength of the association between two variables.

In two-way tables, particularly a 2×2 table, let y_{ij} denotes the observed count at cell (i, j), $i, j = 0, 1$, as follows:

	Event		
Group	0	1	Total
0	y_{00}	y_{01}	$n_{0.}$
1	y_{10}	y_{11}	$n_{1.}$
Total	$n_{.0}$	$n_{.1}$	n

where $n_{i.} = y_{i0} + y_{i1}$ is the marginal row total of the ith row, while $n_{.j} = y_{0j} + y_{1j}$ is the marginal total of the jth column, and $n = n_{..} = n_{0.} + n_{1.} = n_{.0} + n_{.1}$ is the total number of the observations.

The odds ratio (cross product ratio) is one of the common measures of association between

row variable and column variable. The ratio of success probability p vs. failure probability $1-p$, known as the odds of success

$$odds = \frac{p}{(1-p)}.$$

Odds are nonnegative, with $odds > 1$ means that success is more likely than a failure. If p_1 and p_2 are the success probabilities of two populations, then the ratio of two odds

$$\Psi = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

is called the *odds ratio* and is more useful for the comparison of p_1 and p_2 than their difference (Kateri, 2014).

The odds ratio is also defined in terms of the joint distribution of two binary random variables (which we here call "Group" and "Event") given by the four cell probabilities $p_{00}, p_{01}, p_{10}, p_{11}$ as presented in a 2×2 table:

	Event	
Group	0	1
0	p_{00}	p_{01}
1	p_{10}	p_{11}

the p_{ij} represents the probability that a subject in Group $i = 0, 1$ has the Event $j = 0, 1$. The cell probabilities p_{00}, p_{01}, p_{01}, and p_{11} are nonnegative and sum to one. The sampling models in two-way tables will be discussed in the next section. The odds can be defined in terms of the conditional probabilities,

	Event	
Group	0	1
0	$\frac{p_{00}}{p_{00}+p_{01}}$	$\frac{p_{01}}{p_{00}+p_{01}}$
1	$\frac{p_{10}}{p_{11}+p_{10}}$	$\frac{p_{11}}{p_{11}+p_{10}}$

Thus the odds ratio is

$$\begin{aligned}\Psi &= \frac{\left(\frac{p_{00}}{p_{00}+p_{01}}\right)}{\left(\frac{p_{01}}{p_{00}+p_{01}}\right)} \Bigg/ \frac{\left(\frac{p_{10}}{p_{11}+p_{10}}\right)}{\left(\frac{p_{11}}{p_{11}+p_{10}}\right)} \\ &= \frac{p_{00}/p_{01}}{p_{01}/p_{11}} = \frac{p_{00}p_{11}}{p_{01}p_{10}},\end{aligned} \tag{2.1}$$

and the sample odds ratio is

$$\widehat{\Psi} = \frac{\hat{p}_{00}\hat{p}_{11}}{\hat{p}_{01}\hat{p}_{10}} = \frac{y_{00}y_{11}}{y_{01}y_{10}}. \tag{2.2}$$

The distribution of the sample odds ratio is highly skewed; therefore, it is preferred to use the log of odds ratio particularly in two-way tables. The log of the odds ratio in (2.1) is

$$\log \Psi = \log p_{11} - \log p_{10} - \log p_{01} + \log p_{00}, \tag{2.3}$$

and for a random sample, $\log \widehat{\Psi}$ is approximately normally distributed with mean $\log \Psi$ and variance $\sum_{i,j} y_{ij}^{-1}$ (Kateri, 2014).

In the presence of more than two variables in contingency tables, multi-way tables are very common in practice. For a three-way contingency table, let y_{ijk} denote the observed cell count in the cell (i, j, k), where $i = 1, ..., I$ rows, $j = 1, ..., J$ columns, and $k = 1, ..., K$ layer levels. Consider the case when we have $2 \times 2 \times 2$ table as follows:

Layer 1:

	Event	
Group	0	1
0	y_{001}	y_{011}
1	y_{101}	y_{111}

Layer 2:

	Event	
Group	0	1
0	y_{002}	y_{012}
1	y_{102}	y_{112}

In the case when survival data with right-censored observations are presented in contingency tables, at each observed death, 2×2 table is constructed. These tables are the layers as illustrated above, and called strata. The marginal and conditional associations in three-way contingency tables can be described by the odds ratios. For the case of $2 \times 2 \times K$ table and conditioning on the third variable, the conditional odds ratio

$$\Psi_k = \frac{p_{00k} p_{11k}}{p_{01k} p_{10k}},$$

describes the conditional association in partial table k. For example, when we have $2 \times 2 \times 2$ table, the difference between two log odds ratios of two 2×2 tables is defined as follows:

$$\begin{aligned} \log \Psi_1 - \log \Psi_2 &= \{\log p_{001} - \log p_{011} - \log p_{101} + \log p_{111}\} \\ &- \{\log p_{002} - \log p_{012} - \log p_{102} + \log p_{112}\}. \end{aligned} \tag{2.4}$$

In the case of the marginal odds ratio

$$\Psi_{ij.} = \frac{p_{00.} p_{11.}}{p_{01.} p_{10.}},$$

where probabilities in the above equation $p_{ij.} = \frac{y_{ij.}}{n}$ and $y_{ij.} = \sum_k y_{ijk}$. The interest of this thesis is concerned with the case of *stratified* 2×2 tables, more details will be given later in Chapter 4.

2.1.1 Probability Models and Parametrization for Two-way Contingency Tables

Dobson and Barnett (2008) discuss probability models for contingency tables. There are four sampling cases: (1) No totals are fixed, the four cell counts are assumed to have independent Poisson distributions, $Y_{ij} \sim Poisson(\mu_{ij})$, (2) The row or column totals are fixed, then the joint probability distribution for each row or column is product of binomials; for example the row totals are fixed, so $y_{00} \sim B(n_{0.}, p_{00})$ and $y_{10} \sim B(n_{1.}, p_{10})$, (3) The grand total n is fixed, the conditional distribution of the cell counts *given* their total is multinomial with probabilities p_{ij}, (4) The row, column, and thus the grand totals are fixed, then the conditional distribution of y_{00} conditionally on $y_{00} + y_{10} = n_{.1}$ is non-central hypergeometric.

Beside that, the interesting point of view in this thesis is the *odds ratio*, therefore, we are going to consider these four situations in more details regarding the parametrization and the odds ratio in (2.1), whether we think in terms of probabilities that add to one across the table or conditional probabilities for rows, or for columns. It is interested to note that Slavkovic and Fienberg (2010) extend the idea of odds ratio and define other two odds ratios as follows: Conditioning on columns,

$$\Psi^* = \frac{p_{00}p_{01}}{p_{10}p_{11}}. \tag{2.5}$$

and conditioning on rows,

$$\Psi^{**} = \frac{p_{00}p_{10}}{p_{01}p_{11}}. \tag{2.6}$$

First, when there are no constraints on y_{ij}'s, they can be modelled under the assumption that the observations are independent. The joint distribution of the four cell counts is the product of Poisson distribution. Then the log-likelihood can be written as:

$$\begin{aligned} \ell &= \sum_i \sum_j (y_{ij} \log \mu_{ij} - \mu_{ij} - \log y_{ij}!) \\ &\propto \sum_i \sum_j y_{ij} \log \mu_{ij} - \mu_{..}, \end{aligned} \tag{2.7}$$

where $\mu_{..} = \sum_{i,j} \mu_{ij}$ and $\mu_{ij} = \mu_{..} p_{ij}$, then $p_{00} = \mu_{00}/\mu_{..}$, $p_{01} = \mu_{01}/\mu_{..}$, $p_{10} = \mu_{10}/\mu_{..}$, and $p_{11} = 1 - p_{00} - p_{01} - p_{10}$. Thus, the odds ratio in (2.1) can be written in terms of μ_{ij} as follows:

$$\Psi = \frac{p_{00}p_{11}}{p_{01}p_{10}} = \frac{\mu_{00}\mu_{11}}{\mu_{01}\mu_{10}} \tag{2.8}$$

Second, when the row or column totals are fixed, the joint distribution is a product of Binomials. For instance, when the row totals are fixed with the constraints that $\sum_{i,j} p_{ij} = 1$ and $n_{i.} = \sum_j y_{ij}$, the log-likelihood will be will be proportional to

$$\ell \propto \sum_i \sum_j y_{ij} \log p_{ij}. \tag{2.9}$$

Now, by considering what is known as conditional logits for *columns*, suppose that

$$\eta_1 = \log \frac{\frac{p_{00}}{p_{00}+p_{10}}}{\frac{p_{10}}{p_{00}+p_{10}}} = \log \frac{p_{00}}{p_{10}}, \qquad \eta_2 = \log \frac{\frac{p_{01}}{p_{01}+p_{11}}}{\frac{p_{11}}{p_{01}+p_{11}}} = \log \frac{p_{01}}{p_{11}}. \tag{2.10}$$

Then

$$\eta_1 - \eta_2 = \log \frac{p_{00}p_{11}}{p_{01}p_{10}} = \log \Psi, \tag{2.11}$$

and

$$\eta_1 + \eta_2 = \log \frac{p_{00}p_{01}}{p_{10}p_{11}} = \log \Psi^*, \tag{2.12}$$

where Ψ and Ψ^* are the odds ratios in (2.1) and (2.5) respectively.

Similarly, by considering the conditional logits for *rows*, suppose that

$$\zeta_1 = \log \frac{\frac{p_{00}}{p_{00}+p_{01}}}{\frac{p_{01}}{p_{01}+p_{00}}} = \log \frac{p_{00}}{p_{01}}, \qquad \zeta_2 = \log \frac{\frac{p_{10}}{p_{10}+p_{11}}}{\frac{p_{11}}{p_{10}+p_{11}}} = \log \frac{p_{10}}{p_{11}}. \tag{2.13}$$

Then

$$\zeta_1 - \zeta_2 = \log \frac{p_{00}p_{11}}{p_{01}p_{10}} = \log \Psi, \tag{2.14}$$

and

$$\zeta_1 + \zeta_2 = \log \frac{p_{00}p_{10}}{p_{01}p_{11}} = \log \Psi^{**}, \tag{2.15}$$

where Ψ^{**} is the odds ratio in (2.6).

Third, conditional on the grand total n, the joint distribution of the cell counts y_{ij} is multinomial with constraint $\sum_{i,j} p_{ij} = 1$ and the log-likelihood will be proportional to

$$\ell \propto \sum_i \sum_j y_{ij} \log p_{ij}, \tag{2.16}$$

and one can parametrize the multinomial distribution for a 2×2 table with cell probabilities $p_{00}, p_{01}, p_{10}, p_{11}$ with the parameters

$$\theta_1 = \frac{1}{2}(\eta_1 + \eta_2) = \log \sqrt{\Psi^*}, \tag{2.17}$$

$$\theta_2 = \frac{1}{2}(\zeta_1 + \zeta_2) = \log \sqrt{\Psi^{**}}, \tag{2.18}$$

and

$$\theta_3 = \frac{1}{2}(\eta_1 - \eta_2) = \frac{1}{2}(\zeta_1 - \zeta_2) = \log\sqrt{\Psi}, \tag{2.19}$$

then the odds ratio can be written in terms of θ_3 as follows:

$$\Psi = e^{2\theta_3}. \tag{2.20}$$

Fourth, by looking to the case when row, column, and grand totals are fixed the distribution will be the non-central hypergeometric (which is an exponentially weighted version of the central hypergeometric distribution (McCullagh and Nelder, 1989)) and parametrized by the odds ratio.

Now, it might be possible to start with the case of fixing row totals, for example, to get the fourth case of non-central hypergeometric. Suppose that we have the row totals are fixed, then the sampling model is product binomial

$$y_{00} \sim B(n_{0.}, p_{00}) \qquad y_{10} \sim B(n_{1.}, p_{10}).$$

Therefore, the conditional distribution of y_{00} conditionally on $y_{00} + y_{10} = n_{.0}$ is non-central hypergeometric with parameter Ψ as follows:

$$f(y;\Psi) = \frac{\binom{n_{0.}}{y_{00}}\binom{n_{1.}}{y_{10}}\Psi^{y_{00}}}{P_0(\Psi)}, \tag{2.21}$$

where $\max(0, n_{.0} - n_{1.}) \leq u \leq \min(n_{0.}, n_{.0})$ and $P_0(\Psi) = \sum_u \binom{n_{0.}}{u}\binom{n_{1.}}{n_{.0}-u}\Psi^u$.

In the non-central hypergeometric distribution case, the log-likelihood function will take the form

$$\ell \propto y_{00}\log\Psi - \log P_0(\Psi). \tag{2.22}$$

The non-central hypergeometric distribution in (2.21) and the log likelihood in (2.22) above can be expressed in terms of (2.20) respectively as follows:

$$f(y;\Psi) = \frac{\binom{n_{0.}}{y_{00}}\binom{n_{1.}}{y_{10}}e^{2\theta_3 y_{00}}}{P_0(e^{2\theta_3})}, \tag{2.23}$$

and

$$\ell \propto 2y_{00}\theta_3 - \log P_0(e^{2\theta_3}), \tag{2.24}$$

where $P_0(e^{2\theta_3}) = \sum_u \binom{n_{0\cdot}}{u}\binom{n_{1\cdot}}{n_{\cdot 0}-u}e^{2u\theta_3}$.

In fact, the fourth sampling case where row and column and then grand totals are fixed and under the null hypothesis of independence, the non-central hypergeometric distribution is the basis of constructing the log-rank test as discussed in Section 2.2.1 in comparing two survival functions in survival data analysis.

2.1.2 Log-linear Models for Contingency Tables

Modelling cell counts in contingency tables is based on work that deals fundamentally with the knowledge of *log-linear models*. The development of log-linear models grew primarily through the work of Birch (1963), Goodman (1963), and Bishop (1967). The log linear model is one of the specialized cases of *generalized linear models* for Poisson or multinomial-distributed data (Dobson and Barnett, 2008). The log-linear model is used to analyze the relationship between two categorical variables (two-way contingency tables) or more than two categorical variables (multi-way contingency tables). The log-linear model in two-way $I \times J$ contingency tables can take the following form:

$$E(Y_{ij}) = \mu_{ij} = np_{ij}$$

$$\log E(Y_{ij}) = \log \mu_{ij} = \log n + \log p_{ij},$$

and under independence, $p_{ij} = p_{i\cdot}p_{\cdot j}$

$$\log E(Y_{ij}) = \log \mu_{ij} = \log n + \log p_{i\cdot} + \log p_{\cdot j}.$$

Therefore, the formula for expressing independence is multiplicative, so $\log \mu_{ij}$ has the additive form and can be written as follows:

$$\log E(Y_{ij}) = \lambda + \lambda_i^1 + \lambda_j^2,$$

for a row effect λ_i^1, a column effect λ_j^2, and λ is overall mean (the 1 and 2 superscripts are labels, not "power" exponents), and in the case of dependence, the model is called a saturated (full) model and has the form:

$$\log E(Y_{ij}) = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_{ij}^{12}, \tag{2.25}$$

where λ_{ij}^{12} is the interaction effect and under the constraints

$$\sum_{i=1}^{I} \lambda_i^1 = 0, \quad \sum_{j=1}^{J} \lambda_i^2 = 0, \quad \sum_{i=1}^{I} \lambda_{ij}^{12} = \sum_{j=1}^{J} \lambda_{ij}^{12} = 0.$$

Hence the minimal model will take the form:

$$\log E(Y_{ij}) = \lambda.$$

The log of the odds ratio in (2.1) can be written in terms of the parametrization that discussed in the previous section and using (2.20) as follows:

$$2\theta_3 = \log \frac{p_{11}/p_{10}}{p_{01}/p_{00}} \tag{2.26}$$

$$\theta_3 = 1/2(\log p_{11} - \log p_{10}) - 1/2(\log p_{01} - \log p_{00}), \tag{2.27}$$

where $p_{ij} = \mu_{ij}/n_{i.}$. Then the log-odds ratio can be also expressed in terms of the expected frequencies as follows:

$$\begin{aligned} \theta_3 &= 1/2(\log \mu_{11} - \log \mu_{10}) - 1/2(\log \mu_{01} - \log \mu_{00}) \\ &= 1/2\lambda_{11}^{12} + 1/2\lambda_{00}^{12} - 1/2\lambda_{10}^{12} - 1/2\lambda_{01}^{12}, \end{aligned} \tag{2.28}$$

which means the λ_{ij}'s determine the association (Agresti, 2002). Also, one can show that the parameter λ_{ij} is equal to 1/4 log-odds ratio based on the constraints on log-linear's parameters for the 2×2 tables, $\sum_i \lambda_{ij} = 0$ and $\sum_i \lambda_{ij} = 0$, so:

$$\begin{aligned} \lambda_{00}^{12} + \lambda_{01}^{1} &= 0, \\ \lambda_{00}^{12} &= -\lambda_{01}^{1}, \end{aligned}$$

and

$$\lambda_{11}^{12} + \lambda_{10}^{2} = 0,$$

$$\lambda_{11}^{12} = -\lambda_{10}^{2},$$

where $\lambda_{01}^{1} = \lambda_{10}^{2}$, so we can write $\lambda_{00}^{12} = \lambda_{11}^{12} = -\lambda_{10}^{2}$ and substitute this in (2.28), we will have

$$\begin{aligned}
\theta_3 &= 1/2\lambda_{11}^{12} + 1/2\lambda_{00}^{12} - 1/2\lambda_{10}^{2} - 1/2\lambda_{01}^{1} \\
&= 1/2\lambda_{00}^{12} + 1/2\lambda_{00}^{12} + 1/2\lambda_{00}^{12} + 1/2\lambda_{00}^{12} \\
&= 2\lambda_{00}^{12}.
\end{aligned}$$

Now, for a three-way $I \times J \times K$ contingency table, the saturated log-linear model is

$$\log E(Y_{ijk}) = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{ijk}^{123}, \tag{2.29}$$

satisfying the constraints that

$$\sum_{i=1}^{I} \lambda_i^1 = \sum_{j=1}^{J} \lambda_i^2 = \sum_{k=1}^{K} \lambda_k^3 = 0,$$

$$\sum_{i=1}^{I} \lambda_{ij}^{12} = \sum_{j=1}^{J} \lambda_{ij}^{12} = ... = \sum_{k=1}^{K} \lambda_{jk}^{23} = 0,$$

$$\sum_{i=1}^{I} \lambda_{ijk}^{123} = \sum_{j=1}^{J} \lambda_{ijk}^{123} = \sum_{k=1}^{K} \lambda_{ijk}^{123} = 0.$$

2.1.3 Bayesian Inference for Log-linear Model

Bayesian inference is one of two dominant approaches to statistical inference. The word "Bayesian" refers to the influence of Reverend Thomas Bayes, who introduced what is now known as Bayes' theorem where model parameters, $\boldsymbol{\theta}$, are treated as being random variables and a prior distribution is assigned to these parameters. Bayesian paradigm has three components: A prior distribution $\pi(\boldsymbol{\theta})$; the given data $\mathbf{Y}$; the function $p(\mathbf{Y}|\boldsymbol{\theta})$ as the likelihood function; and posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{Y})$, which is a result of updating your prior

by multiplying the prior by the likelihood. Combining these components leads us to *Bayes' theorem*:

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{Y})} = \frac{p(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\ d\boldsymbol{\theta}} \tag{2.30}$$

In general,

$$Posterior \propto Likelihood \times Prior.$$

Bayesian inference for log-linear parameters in two-way tables has been done using a prior distribution on the parameters and expressing the results in the form of a posterior distribution.

For a multinomial random variable y_{ij} with cell probabilities $\boldsymbol{p} = (p_{00}, p_{01}, p_{10}, p_{11})'$, the Dirichlet distribution, denoted $Dir(\boldsymbol{\alpha})$, is a conjugate prior (the prior and the posterior distributions are called conjugate if they are in the same family distribution). It is a continuous multivariate distribution parametrized by a vector $\boldsymbol{\alpha}$ of positive reals, and has probability density function

$$\pi(\boldsymbol{p}) \propto \prod_{i,j} p_{ij}^{\alpha_{ij}-1}. \tag{2.31}$$

Lindley (1964) used Dirichlet (α_{ij}) prior distributions for the cell probabilities p_{ij}. He showed that the contrasts of log cell probabilities, $\sum\sum a_{ij} \log p_{ij}$ where $\sum\sum a_{ij} = 0$, such as log-odds ratio in 2×2 table with cell probabilities p_{ij}'s, have an approximate (large sample) joint normal posterior distribution with mean and variance given respectively by

$$\mu = \textstyle\sum_i \sum_j a_{ij} \log y_{ij}, \quad v = \sum_i \sum_j a_{ij}^2 y_{ij}^{-1}.$$

Lindley used this approximation to obtain the posterior density of the log-odds ratio and to develop a Bayesian statistic for testing independence in 2×2 table. Also, extensions to three-way tables; especially $2 \times 2 \times 2$ tables, were introduced. Lindley (1964) derives the posterior of the difference between the two log odds ratios of two 2×2 tables in (2.4), which is approximately normally distributed with mean

$$\begin{aligned} \mu_{\Psi_1 - \Psi_2} &= \{\log y_{111} - \log y_{211} - \log y_{121} + \log y_{221}\} \\ &- \{\log y_{112} - \log y_{212} - \log y_{122} + \log y_{222}\}, \end{aligned}$$

and variance

$$\sigma_{\Psi_1-\Psi_2} = \sum_{i,j,k} y_{ijk}^{-1}.$$

In a hierarchical Bayesian approach, for two-way contingency table with cell count y_{ij} and cell probabilities p_{ij}, Leonard (1975) considers the multivariate logits

$$\log p_{ij} = \gamma_{ij} - D(\gamma), \tag{2.32}$$

where $D(\gamma) = \log(\sum_{ij} \exp(\gamma_{ij}))$ is chosen to ensure that p_{ij} always sum to one. Then he introduces row effects λ_i^1, column effects λ_j^2, and interaction effects λ_{ij}^{12} satisfying

$$\gamma_{ij} = \lambda_i^1 + \lambda_j^2 + \lambda_{ij}^{12}.$$

To assign a prior distribution, the row effects λ_i^1 are assumed to be a priori independent of the column effects λ_j^2 and also of the interaction effects λ_{ij}^{12}. To model the belief that the set of row effects λ_i^1 is exchangeable, Leonard uses two-stage prior:
Stage I: $\lambda_1^1, ..., \lambda_I^1$ are independently normally distributed with mean μ_{λ_i} and variance $\sigma_{\lambda_i}^2$.
Stage II: The prior parameters μ_{λ_i} and $\sigma_{\lambda_i}^2$ are independent where μ_{λ_i} have an improper uniform over the whole of the real line. Given parameters τ_{λ_i}, ν_{λ_i} both positive, the $\tau_{\lambda_i}\nu_{\lambda_i}\sigma_{\lambda_i}^{-2}$ is assumed to have inverse chi-squared distribution with ν_{λ_i} degrees of freedom and τ_{λ_i} provides a prior estimate of $\sigma_{\lambda_i}^2$. Similar exchangeable prior distributions are assigned to the sets of column effects λ_j^2 and interaction effects λ_{ij}^{12}.

For computational convenience, Leonard estimated the log-linear parameters by their posterior modes, and those posterior modes were plugged into the log-linear model to get cell probability estimates.
In three-way contingency tables, Nazaret (1987) extends the work of Leonard (1975) for two-way tables by using the same Bayesian approach with a multivariate logit transformation for obtaining the Bayes estimates for the main and interaction effects. Besides that Nazaret (1987) approximates the posterior means by the posterior modes for moderate sample size. Also, Nazaret shows that the choice of the value of ν's (the degree of freedom above) and the sample size affect the speed of the convergence. For example, Leonard (1975)'s advice is to choose the values of ν's to be close to zero where these values affect the convergence to be

slow and lead to a negative-definite covariance matrix of model parameter estimates. However, according to Nazaret (1987), choosing values of ν's to be close to one and large sample size speeds up the convergence which makes the algorithm rapidly climb to the maximum of the posterior distribution.

Assuming a multinomial sampling model, Albert and Gupta (1982) assign a Dirichlet prior for cell probabilities p_{ij} with parameters $\alpha_{ij} = st_{ij}$, where s is the precision parameter and t_{ij}'s are the prior means (this is a reparameterization of Dirichlet distribution in terms of a precision parameter and means, and it is convenient because s will be fixed). The hyperparameters t_{ij} reflect a prior belief that the cell probabilities may be either symmetric or independent in two-way contingency tables. In the symmetry case, they gave a Dirichlet distribution for the parameters of the multinomial model in the first stage, and in the second, uniform distribution was given to the prior means. In the independence case, they assumed Dirichlet prior distributions in both stages. Albert and Gupta showed that large value of s, the precision parameter, indicates strong belief in symmetry or independence.

Incorporating the Dirichlet distribution as a prior for the cell probabilities in contingency tables with multinomial sampling is a tractable choice because of its computational convenience. However, according to Agresti and Hitchcock (2005) and Knuiman and Speed (1988), a one-stage Dirichlet prior does not always provide a sufficient structure to be given for cell probabilities, such as corresponding to a log-linear model. As Leonard (1975) discussed that the exchangeability within each set of log-linear parameters is more reasonable than the exchangeability of multinomial probabilities that one gets with a Dirichlet prior. Therefore, the choice of normal prior for the log probabilities is an alternative choice to Dirichlet prior for cell probabilities. In the same context, Albert and Gupta (1983) argued that the Dirichlet distribution is appropriate for representing the prior information about cell probabilities, however, it does not have enough number of parameters to combine separate prior knowledge about the marginal probabilities and an interaction parameter.
Albert and Gupta (1983) considered 2×2 tables where the cell counts are assumed to have a multinomial distribution and in which the prior information was stated in terms of two

common measures of association. To measure the association in the table, Albert and Gupta consider the the correlation coefficient, $\rho = \frac{(p_{11}-p_{1.}p_{.1})}{R}$, where $R = (p_{1.}p_{2.}p_{.1}p_{.2})^{1/2}$, and the odds ratio, $\Psi = \frac{(p_{11}p_{22})}{(p_{12}p_{21})}$. Instead of using a Dirichlet prior on the cell probabilities, they define the prior density of the form

$$\psi(p_{11}, p_{12}, p_{21}, p_{22}) = \psi_1(p_{1.})\psi_2(p_{.1})\psi_3(p_{11}, p_{12}, p_{21}, p_{22}|p_{1.}, p_{.1}),$$

where ψ_3 is the prior density of one independent parameter such as ρ or Ψ, which describes conditionally on the marginal probabilities, the interaction between the two variables.

Knuiman and Speed (1988), in the essence of Albert and Gupta (1983), used a structured multivariate normal prior for the parameters in log-linear model collectively instead of giving a univariate normal prior for each parameter individually as in Leonard (1975). They assumed the Poission log-linear model

$$\log(m) = \boldsymbol{X}\boldsymbol{\beta},$$

where $m = E(y)$, $\boldsymbol{X}$ is the model matrix, and $\boldsymbol{\beta}$ is the vector of unknown regression coefficients or effects (model parameters). The $\boldsymbol{\beta}$ was given a multivariate normal distribution, $\boldsymbol{\beta} \sim N(\boldsymbol{\beta_0}, \boldsymbol{S})$, where $\boldsymbol{\beta_0}$ and $\boldsymbol{S}$ are assumed known. Knuiman and Speed considered the posterior summary statistics that are the posterior mode which is the solution of

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log[\pi(\boldsymbol{\beta}|\boldsymbol{y})] = 0,$$

where $\pi(\boldsymbol{\beta}|\boldsymbol{y})$ is the posterior density for $\boldsymbol{\beta}$. And the dispersion matrix

$$D(\boldsymbol{\beta}) = -[\frac{\partial^2 \log[\pi(\boldsymbol{\beta}|\boldsymbol{y})]}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}].$$

For more illustration, Knuiman and Speed (1988) provided two examples of data in two-way and three way contingency table, and compared the posterior mode and measure of dispersion estimates with the likelihood estimates where they concluded that the results are very similar.

In the context of using a normal prior with multinomial data, Lindley (1964) remarked that if cell counts y_i are independent Poisson variables with means μ_i, the conditional distribution of them, given $n = \sum y_i$, would be multinomial with cell probabilities $p_i = \frac{\mu_i}{\sum \mu_i}$, and

also since the $n = \sum y_i$ has a Poisson distribution with mean $\sum \mu_i$, one can define an alternative parameterization for Poisson model like $\mu^+ = \sum \mu_i$. Therefore, if the prior distribution of μ_i can be factored into one part that depends only on μ^+ and another part depends only on p_i, then the posterior would be the same. As a result, the posterior distribution of p_i will only depend on the multinomial part of the likelihood. Thus the posterior may be obtained by the Poisson device. Based on this, Forster (2010) develops the results of Lindley (1964) to provide a general framework for the analysis of multinomial data using Poisson log-linear model. Forster's focus is particularly on multivariate normal prior distributions for the log-linear parameters.

2.2 Survival Data Modelling and Analysis

Survival analysis is the term used to describe the analysis of survival time or lifetime data. In health applications, the survival time could be the time from diagnosis of a disease till death, or the length of a disease's remission time. In engineering, survival time could be the time to failure of a part (in which case survival data may be referred to as reliability data). The usual questions of interest involve the quantiles (e.g., median) of the survival time, or the effect of covariates on survival time. We may be interested in characterizing the distribution of "time to event" for a given population, as well as comparing this "time to event" among different groups (e.g., treatment vs. control in a clinical trial). Two features of survival time data are:

- Times are non-negative.
- The survival time might be *censored*, the survival time of a subject is said to be censored when the end-point of interest has not been observed for that subject. There are some general reasons why censoring may occur:
  - A subject does not experience the event before the study ends.
  - A subject can not be followed-up on during the study period.

– A subject withdraws from the study because of death from an unrelated cause.

In any survival analysis, survival and hazard functions are two important terms. They are, in essence, opposing concepts, in which the survival function focuses on surviving, whereas the hazard focuses on failing given survival up to a certain point in time.

The *survival function*, $S(t)$, is defined as the probability of a subject surviving longer than a specific time t; that is, $S(t) = P(T > t) = 1 - F(t)$. The survival time here is assumed to be continuous. The *hazard function*, $h(t)$, is used to express the risk or hazard of death at a specific time t, and is obtained from the probability that a subject dies at time t, conditional on that subject having survived to that time,

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

Estimates of survival function and hazard function can be obtained by using methods for estimating such as an empirical estimate, life-table (LT) estimate, Kaplan-Meier estimate, and Nelson-Aalen estimate (Collett, 2003). These methods are non-parametric or distribution-free. Once the estimated survivor function has been found, the median and some other percentiles of the distribution of survival times can be estimated. For more reading about the methods of estimating the survival and hazard functions, see Chapter 2 of Collett (2003) and Chapter 2 of Kleinbaum and Klein (2005).

In clinical studies one is concerned not only with estimating survival or hazard functions, but, more often, with the comparison of the life experience of two or more groups of patients who receive different treatments (test and control treatments). In these kinds of studies, it is difficult to have a priori knowledge to make trustworthy hypotheses on the underlying survival functions; thus, the non-parametric approach is often adopted to compare survival curves. There are a number of frequentist methods that can be used for hypothesis testing. Two of the various most common non-parametric tests are the *Generalized Wilcoxon* test (Gehan, 1965) and the *log-rank* test (Mantel, 1966).

2.2.1 Frequentist Analysis of Survival Data

In the case of three-way contingency tables, Mantel-Haenzel (MH) statistic (Mantel and Haenszel, 1959) is used. MH statistic is obtained by examining the odds ratio at each table, and then combining this information across tables (the common odds ratio). Fixing the third index of three-way table gives a two-way table. These tables are called strata (as discussed early in Section 2.1) in the Mantel-Haenzel case. The log-rank test is essentially the Mantel-Haenzel statistics in which strata are replaced by slices in time and they are computationally equal. In hypothesis testing, the log-rank test is a procedure for comparing the survival functions of two groups.

Consider the following situation when we have two treatments (control and test) with two outcomes (dead and alive), and we are interested in comparing the treatment effects to know which treatment has a longer survival time. This kind of data can be conveniently displayed in three-way contingency tables, stratified 2×2 tables, one table at each time of observed death t_k, where $k = 1, ..., K$, $I = 2$ (i.e. $i = 0$ for control, and $i = 1$ for test) and $J = 2$ (i.e. $j = 0$ for alive (success) and $j = 1$ for dead (failure). In this data format, we have K 2×2 tables. The data at time t_k can be represented in a two-way contingency table as follows:

	Event		
Group	0	1	Totals
0	d_{00k}	$n_{0k} - d_{00k}$	n_{0k}
1	d_{10k}	$n_{1k} - d_{01k}$	n_{1k}
Totals	d_k	$n_k - d_k$	n_k

where n_{0k} and n_{1k} are the number of individuals at risk of death before time t_k in the first and second group respectively, and $n_k = n_{0k} + n_{1k}$ is the total number of individuals at risk of death. Also, d_k is deaths in total out of n_k. Now, under the null hypothesis that there is no difference in survival time in two groups, assessing the validity of this hypothesis is done by the log-rank test. Conditional on the 4 marginal totals in the above table, and under the null hypothesis, the four entries of this table are determined by the value of d_{00k}, the number of deaths at t_k in group 1. Therefore, d_{00k} is a random variable that takes values from 0

to the minimum of d_k and n_{0k}. Thus, d_{00k} has *hypergeometric* distribution with mean and variance

$$\mu_{00k} = \frac{n_{0k}d_k}{n_j},$$

so

$$\sigma^2_{00k} = \frac{n_{0k}n_{1k}d_k(n_k - d_k)}{n_k^2(n_k - 1)},$$

where μ_{00k} under the null hypothesis is the expected number of individuals who die at t_k in group 1. Next, by combining the information from the individual 2×2 tables for each death time, a measure of the deviation of the observed from d_{00k} from their expected values is defined as follows:

$$U_L = \sum_{k=1}^{K}(d_{00k} - \mu_{00k}), \tag{2.33}$$

and the mean value of U_L under the null hypothesis is:

$$E(U_L) = 0,$$

and

$$V_L = Var(U_L) = \textstyle\sum_{k=1}^{K} \sigma^2_{00k}.$$

So, the log-rank test statistic has the form:

$$W_L = \frac{U_L^2}{V_L}, \tag{2.34}$$

The statistic W_L summarizes the extent to which the observed deaths in two groups of data depart from those expected under the null hypothesis of no differences. Under this null hypothesis, the distribution of the statistic W_L is approximately chi-squared with one degree of freedom. The log-rank test is preferred when the assumption of proportional hazard is held. Otherwise, the Wilcoxon test is suitable one for testing the hypothesis that there is no difference between two groups of survival functions. The assumption of proportional hazards is that the ratio of hazards for two groups does not depends on time which means that the ratio of hazards for two group remains constant over time. One way of checking the proportionality is simply by plotting the log-log of the two survival functions against time. If the curves are parallel, we could assume proportional hazards.

2.2.2 Modelling Survival Data

Cox Proportional Hazard Model

The Cox proportional hazard model was introduced by Cox (1972) and it explores the relationship between the survival of a patient and one or more explanatory variables. The model is based on the assumption of proportional hazards.

In general, let us consider a situation where the hazard of death at a particular time depends on the values $x_1, ..., x_p$ of the variables $X_1, ..., X_p$ of p explanatory variables. The set of the values of the explanatory variables can be represented by the vector $\boldsymbol{x} = (x_1, x_2, ..., x_p)'$ with $h_0(t)$ as the baseline hazard function that corresponds to the probability of dying when all the explanatory variables are zero. The baseline hazard function has the same role as the intercept in ordinary regression. The general proportional hazard model for ith individual can be written as:

$$h_i(t) = h_0(t) \exp(x_i \boldsymbol{\beta}), \tag{2.35}$$

where $i = 1, .., n$ and $j = 1, ..., p$ and $\boldsymbol{\beta}$ is a vector of the coefficients in the proportional hazard model. The equation (2.35) can be re-expressed in the form:

$$\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi}, \tag{2.36}$$

The parameters of the Cox model can be estimated by the methods of partial maximum likelihood and measures of discrepancy or goodness of fit may be formed in the logarithm of a ratio of likelihoods to be called the *deviance*, which referes to the quantity $-2 \log \hat{L}$, where $\hat{L}$ is the maximized value of the likelihood function, and this is known as the *log-likelihood ratio* statistic (Collett, 2003).

In the case of comparing two survival functions of two groups, let x_i be a *binary variable*, where

$$x_i = \begin{cases} 0 & \text{if the individual is in group 1 (control);} \\ 1 & \text{if the individual is in group 2 (test),} \end{cases}$$

so, the hazard model in this case will take the form:

$$h_i(t) = h_0(t) e^{\beta x_i}. \tag{2.37}$$

The ratio of the hazards of death at time t for an individual on test treatment relative to an individual on the control treatment is:

$$\psi = \exp\{\beta\}.$$

The log odds ratio in 2×2 tables and log hazard ratio in Cox model have the same interpretation and correspond to each other, that is, the log odds ratio is equal to the log hazard ratio which equal to the regression coefficient, β. Therefore, any inference conclusions about log odds ratio is equivalent to conclusions about Cox model's parameter β.

2.2.3 Bayesian Inference for Survival Data

In Bayesian inference, priors play the important role of representing the uncertainty of the parameter of interest before the current data are observed. In the context of Bayesian survival analysis, according to Ibrahim et al. (1999) and Ibrahim et al. (2001), the very common choice of informative prior for $\boldsymbol{\beta}$ is the normal prior, and the popular noninformative one for $\boldsymbol{\beta}$ is the uniform prior. Considering Cox proportional hazard model of the form

$$h(t, \boldsymbol{x}) = h_0(t) \exp(\boldsymbol{x}'\beta), \tag{2.38}$$

where h_0 is the baseline hazard function at time t, $\boldsymbol{x}$ is a vector of covariates, and $\boldsymbol{\beta}$ denotes a vector of regression coefficients. Ibrahim et al. (1999) suggest a discrete gamma process as a prior for baseline hazard function, $h_0(t)$. To define a discrete gamma process, Ibrahim et al. build a finite partition of the time axis by letting $0 \leq t_0 < t_1 < ... < t_M$ be this finite partition and define the increment in the baseline hazard in the interval $(t_{i-1}, t_i]$, $i = 1, ..., M$, where M is the total number of the intervals, as follows

$$\delta_i = h_0(t_i) - h_0(t_{i-1}).$$

The δ_i's are a priori independent random variables with gamma distributions, $\delta_i \sim G(\alpha(t_i) - \alpha(t_{i-1}), \lambda)$. Now, let $\Delta = (\delta_1, ..., \delta_M)$. The prior density of Δ will be

$$\pi(\Delta) = \prod_{i=1}^{M} f(\delta_i),$$

where $f(\delta_i)$ is a $G(\alpha(t_i) - \alpha(t_{i-1}), \lambda)$ density, where $\alpha(t_i) - \alpha(t_{i-1})$ are the shape parameters and λ is the scale parameter, where $\alpha, \lambda > 0$. One of the prior choices that Ibrahim et al. suggest in the case of having little information about Δ is $\alpha(t) = t_i - t_{i-1}$ for $t_{i-1} \leq t \leq t_i$ and choose large λ. For a prior distribution for the regression coefficients, Ibrahim et al. assume that Δ and $\boldsymbol{\beta}$ are a priori independent having the joint prior density and then consider a multivariate normal prior for the regression coefficient, $\boldsymbol{\beta}$. To compute the posterior probabilities, they used the computational method Gibbs sampling (Lynch, 2007).

Omurlu et al. (2009) compared Bayesian survival analysis and Cox regression analysis by using simulated and breast cancer data sets where the comparison was done by comparing the parameter estimates of both methods. Omurlu considers two situations in choosing a prior distribution for the parameters, $\boldsymbol{\beta}$: informative and noninformative priors. When the sample size increased, the posterior summaries that have been obtained from the Bayesian survival analysis with proper informative prior were more unbiased with smaller standard error than Cox regression analysis. Moreover, Bayesian survival analysis had a better predictive performance than Cox analysis when the variance of informative prior was decreased, which led them to conclude that Bayesian survival analysis had better performance than Cox regression analysis in the case of informative priors.

2.3 Prior distributions and Imprecise Probabilities

The prior distribution is the key to Bayesian inference, and its determination is the most important step in drawing this inference. The choice of prior distribution is the issue that is still challenging statisticians and researchers. In practice, rarely is information precise enough to lead to the exact determination of the prior distribution. At this point, we will briefly discuss the choice of prior distributions: Subjective, objective, and conjugate priors. Because of limitations of a precise Bayesian approach as mentioned in the introduction of this thesis, the *imprecise probability* approach as a generalization is the main focus and it is introduced later in this section.

2.3.1 Choice of Prior Distributions

Subjective Priors

Subjective (informative) priors were developed by De Finetti (1937), Savage (1972), and Lindley (1956), which is more commonly called an elicited prior, and refers to the elicitation of knowledge (Kass and Waserman, 1996). Elicitation is defined as a technique of gathering expert opinion, each expert will give us his probability (belief) about an unknown quantity, which is known as subjective probability. Since subjective probabilities represent degree of beliefs, the elicitation is the technique to extract and quantify the individual belief about uncertain quantities. Savage (1972) presents three methods for elicitation techniques. He proposes that a direct question about the feeling is one way to elicitation. The second method infers probabilities from the individual's actions in an uncertain situation. A third way is asking the individual what his actions would be in the situation. One general way to do the elicitation is through specially designed methods of verbal or written communication, which can be done through individual interviews or interactive groups.

The choice of prior distribution depends on the availability of prior information. Therefore, there are various reasons that make it difficult to get precise information. First, there are a number of biases in people's probability assessments. Second, elicitation itself can be biased; for example, clinicians may be overly optimistic, or trial investigators may be more optimistic than clinicians in general. Last, the decision-maker, the client, or the statistician may not have the time or resources to determine the proper prior based on the information that they have.

Objective Priors

Objective prior is an alternative to elicitation to find a structural rules that define a prior. Jeffreys' prior (Jeffreys, 1946) is the most famous objective prior distribution in one-dimensional problems, and is considered a weakly informative prior because it, in some ways, has minimum information (Berger, 2006). Jeffreys formulated his rule by considering a number of situations. In one of these cases, when the parameter space is finite, he considered the

principle of insufficient reason and took the prior density to be constant. Also, he showed that Jeffreys prior is invariant to one-to-one transformation of the parameter. Jeffreys (1946) suggests his general rule, which is based on Fisher's information:

$$I(\theta) = -E\left[\frac{\partial^2 \log l(x|\theta)}{\partial\theta^2}\right],$$

where θ is the parameter of interest, and $l(x|\theta)$ is the log-likelihood. Then, the Jeffreys' prior density in a one-dimensional case is:

$$\pi(\theta) \propto I(\theta)^{\frac{1}{2}}.$$

Example 2.3.1 *Suppose* $x_i \sim B(n,\theta)$, $i = 1, ..., n$,

$$p(\boldsymbol{x}|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x},$$

where $0 \leq \theta \leq 1$ *and then the log likelihood function would be:*

$$l(\boldsymbol{x}|\theta) = \log\binom{n}{x} + x\log(\theta + (n-x)\log(1-\theta)$$

$$\frac{\partial}{\partial\theta}l(\boldsymbol{x}|\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

$$\frac{\partial^2}{\partial\theta^2}l(\boldsymbol{x}|\theta) = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2},$$

and

$$I(\theta) = -E\left[\frac{\partial^2 l(\boldsymbol{x}|\theta)}{\partial\theta^2}\right] = \frac{n}{\theta(1-\theta)},$$

Then the Jeffreys' prior on θ *is:*

$$\begin{aligned}\pi(\theta) &\propto |I(\theta)|^{\frac{1}{2}} \\ &\propto \left[\frac{n}{\theta(1-\theta)}\right]^{\frac{1}{2}} \\ &\propto \theta^{-1/2}(1-\theta)^{-1/2}.\end{aligned}$$

which is $Beta(1/2, 1/2)$.

Jeffreys' prior satisfies the property of being invariant to one-to-one transformation. If $\phi = h(\theta)$, then

$$\pi(\phi) = \pi(\theta)\left|\frac{d\theta}{d\phi}\right| = \pi(\theta)\left|\frac{dh(\theta)}{d\theta}\right|^{-1}.$$

In fact, Jeffreys' prior has been criticized by some Bayesians. One of these disadvantages is that in a multidimensional case, Jeffreys prior may lead to incoherences. According to Robert (2007), Jeffreys himself was mainly emphasizing the use of these kind of prior distributions in a one-dimensional case, Jeffreys' prior turns out to be the same as the so-called reference prior in the one-dimensional case.

Reference priors were proposed by Jose Bernardo in a 1979 paper (Bernardo, 1979), and further developed by Jim Berger and others from the 1980's through to the present. The idea behind reference priors is to formalize what exactly we mean by an uninformative prior: it is a function that maximizes some measure of distance or divergence between the posterior and prior as observations are made. The commonly used definition of a reference prior is a prior that maximizes the missing information in the experiment. Any of several possible divergence measures can be chosen; for example, the Kullback-Leibler divergence or the Hellinger distance. One might ask how can we choose a prior to maximize the divergence between the posterior and prior, without having seen the data first? Reference priors handle this by taking the expectation of the divergence given a model distribution for the data.

Following Bernardo (1979), consider we have a model $\mathcal{M} = \{p(x|\theta), x \in X, \theta \in \Theta\}$ parametrized by Θ, and we want a strictly positive prior function $\pi(\theta)$ that maximizes its K-L divergence; from the posterior this K-L divergence is;

$$\int \pi(\theta|x) \log \frac{\pi(\theta|x)}{\pi(\theta)} d\theta.$$

Its *expected information* about θ to be delivered by the model $\mathcal{M}$ can be written as;

$$I\{\mathcal{M}, \pi(\theta)\} = \int p(x) \int \pi(\theta|x) \log \frac{\pi(\theta|x)}{\pi(\theta)} d\theta dx,$$

where $p(x)$ is the marginal distribution. The expected information, $I\{\mathcal{M}, \pi(\theta)\}$, measures the amount of missing information about θ when the prior is $\pi(\theta)$. Therefore, choosing a reference prior involves finding $\pi^*(\theta)$ that maximizes the expected information:

$$\pi^*(\theta) = \arg\max_{\pi(\theta)} I\{\mathcal{M}, \pi(\theta)\}. \tag{2.39}$$

We note that defining reference priors in terms of mutual information implies that they are invariant under reparameterization, since the mutual information itself is invariant. However, reference prior has some drawbacks as discussed in Berger and Bernardo (1992). To discuss these drawbacks, the following definitions are considered,

Definition 2.3.1 *A strictly positive continuous prior $\pi(\theta)$ is a permissible prior for model $\mathcal{M} = \{p(x|\theta), x \in X, \theta \in \Theta\}$ if:*

1. *for all $x \in X$, $\pi(\theta|x)$ is proper, such that $\int p(x|\theta)\pi(\theta)d\theta < \infty$;*
2. *for some approximating compact sequence, the corresponding posterior sequence is expected logarithmically convergent to $\pi(\theta|x) \propto p(x|\theta)\pi(\theta)$.*

Definition 2.3.2 *(Maximizing Missing Information (MMI) Property)*
Let $\mathcal{M} \equiv p(x|\theta), x \in X, \theta \in \Theta \in R\}$, be a model with one continuous parameter, and let $\mathcal{P}$ be a class of prior function for θ for which $\int p(x|\theta)\pi(\theta)d\theta < \infty$. The function $\pi(\theta)$ is said to have the MMI property for model $\mathcal{M}$ given $\mathcal{P}$ if, for compact set $\Theta_0 \in \Theta$ and any $p \in \mathcal{P}$,

$$\lim_{k\to\infty} \{I\{\pi_0|\mathcal{M}^k\} - I\{p_0|\mathcal{M}^k\}\} \geq 0, \tag{2.40}$$

where π_0 and p_0 are the renormalized restrictions of $\pi(\theta)$ and $p(\theta)$ to Θ_0, respectively.

The first weakness that they have discussed is with the continuous parameter space where the problem of maximizing the mutual information may not be analytically tractable, and the second is when there is a case of infinite amount of information, the expected information is typically not defined on an unbounded set. To overcome these two difficulties, Berger et al. (2009) suggested and defined that a reference prior must be permissible and have a Maximizing Missing Information (MMI) property, where the latter is considered to be more essential.

Conjugate Priors

Conjugate priors are commonly used in Bayesian inference for computational convenience. In a Bayesian framework, if the posterior distribution is in the same family as the prior

distribution, then the prior and posterior are called *conjugate* distributions, and the prior is called a conjugate for the likelihood function. Where the likelihood functions happen to be an exponential family, there exists a conjugate prior, and this is one of the exponential family's properties. A conjugate family for a natural exponential family

$$f(y|\theta) = h(y)\exp(\theta y - \phi(\theta)), \tag{2.41}$$

where $\phi(\theta) = \log \int h(y)\exp(\theta y)dy$ is the cumulant function, is given by:

$$\pi(\theta|\mu,\lambda) = K(\mu,\lambda)\exp(\theta\mu - \lambda\phi(\theta)). \tag{2.42}$$

Here, the underlying measure defined by (2.42) is the Lebesgue measure, μ and λ are hyperparameters, and $K(\mu,\lambda)$ is the normalizing constant of the density (Robert, 2007). According to Diaconis and Ylvisaker (1979), it is possible to show that this distribution is normalizable if $\lambda > 0$ and μ/λ lies in the interior of the convex hull of the support of the parameter θ, $\theta \in \Theta$, where Θ is called the natural parameter space and it is assumed to be a nonempty open set on R^d. The conjugate prior exponential family has further attraction since the posterior expectation of the parameter θ,

$$E(\theta|y_1,...,y_n) = \frac{y_0 + n\bar{y}}{\lambda + n} \tag{2.43}$$

is a linear function in y, as shown by Diaconis and Ylvisaker, where y_0 is the prior information. They also have shown, additionally, under certain regularity conditions, if the dominating measure of f is continuous with respect to the Lebesgue measure, the linearity of the posterior expectation allows for the prior distribution is of the form of the natural exponential family with such standard examples as normal prior for normal location, the gamma prior for the Poisson, and beta prior for the binomial.

Each prior discussed in the previous two subsections might be a conjugate prior. As shown in Example 2.3.1, Jeffreys' prior form a beta family, which is a conjugate prior to the binomial likelihood function.

Choosing prior distribution to represent the uncertainty about the parameter of interest is a feature of Bayesian analysis. Because of limitations of a precise Bayesian approach as men-

tioned in the introduction of this thesis, the *imprecise probability* approach as a generalization is the main focus of my thesis and is introduced in the next subsection.

2.3.2 Basic concepts of Imprecise Probability Theory

The prior distribution represents the uncertainty about the parameter of interest before data is observed. The traditional (precise) probability theory has limitations, the most crucial one is when we have little or no information for assessing a single probability of an event; say A, $Pr(A) = p$. Instead of a precise (single) value of the probability of an event, a pair of lower and upper probabilities $Pr(A), [p_1, p_2]$ are used to include a set of probabilities, and this leads to the concept of *imprecise probability*. Imprecise probability theory is a generalization of classical probability theory in terms of lower and upper probabilities and lower and upper expectations.

The idea of using imprecise probability has a long history. The first work to build the theory of imprecise probability was made by Keynes (1921) when he discussed the mathematical models of upper and lower probabilities. This was followed by a large amount of literature where the imprecision of personal probabilities and utilities was stressed, in particularly by Good (Good, 1952), who also proposed axioms for upper and lower probabilities. The upper and lower probabilities were inferred as personal betting rates by Smith (1961) when he suggested some essential basics of avoiding sure loss and coherence concepts (definitions of these two concepts are provided in Subsection 2.3.2). Later on, Williams (1975, 2007) generalized Smith's results that coherent lower probabilities are lower envelopes of precise probability measures. Both Williams and Smith's work were inspired from de Finetti 's exposition (De Finetti, 1937) where de Finetti's approach is based on the idea that the price $P(X)$, where X is a random quantity, should be fair. For example, de Finetti assumes that the individual is willing to take either side of the bet, so that the bet is "fair" from the individual's point of view. After Smith and Williams' work, a large contribution was made by Shafer (1976) with his historical statement 'Dempster-Shafer theory' of belief function. Stemming from these efforts, Fine (1988) explored and developed the theory of undominated

lower probabilities to be applied to model for understanding nondeterminstic phenomena. A comprehensive collection of the foundations of all previous work in imprecise probabilities theory is provided in Walley's book (Walley, 1991) where the name of "imprecise probability" was proposed. In the preface to this book, Walley said:

> *My view of probabilistic reasoning has been especially influenced by the writings of Terrence Fine, Bruno de Finetti, Jack Good, J.M. Keynes, Glenn Shafer, Cedric Smith, and Peter Williams.*

For tracking the development of imprecise probabilities work, the "Society for Imprecise Probability: Theory and Applications" (SIPTA) (http://www.sipta.org) aims at promoting research on imprecise probabilities through a series of activities, including ISIPTA conferences every odd year since 1999 and SIPTA schools every even year since 2004.

Coherent Lower and Upper Previsions

Imprecise probability can be seen as a generalization of the traditional (precise) probability theory. Imprecise probabiliy is applicable when information is scarce, vague, or conflicting, in which case a unique probability distribution may be hard to identify. Imprecise probability theory is based on lower and upper previsions (expectations), denoted by $\underline{P}(X)$ and $\overline{P}(X)$, where the lower prevision, $\underline{P}(X)$, can be regarded as *supremum buying price*, and the upper prevision, $\overline{P}(X)$, as *infimum selling price*, where X is a *gamble*, what is the gamble? A gamble X is a bounded real-valued function (a random variable) on Ω

$$X : \Omega \to \mathbb{R} : A \mapsto X(A),$$

where Ω is the set of possible outcomes A. Walley defined the *coherent* lower and upper previsions as follows:

Definition 2.3.3 *(Coherent Lower Prevision)*
Suppose that $\mathcal{X}$ is a linear space of bounded random variables (gambles) on the sample space Ω, and the lower prevision, $\underline{P}$, is a real-valued function that maps to real numbers, $\underline{P} : \mathcal{X} \to \mathbb{R}$. Then $\underline{P}$ is said to be coherent when it satisfies the following three axioms, for all $X, Y \in \mathcal{X}$, and the positive scalar λ:

A-1 $\underline{P}(X) \geq \inf X$

A-2 $\underline{P}(\lambda X) = \lambda \underline{P}(X)$

A-3 $\underline{P}(X+Y) \geq \underline{P}(X) + \underline{P}(Y)$.

The coherent lower prevision $\underline{P}$ is a concave function by A-2 $-$ A-3, i.e., $\underline{P}(\lambda X + (1-\lambda)Y) \geq \lambda \underline{P}(X) + (1-\lambda)\underline{P}(Y)$ when $0 \leq \lambda \leq 1$.

Definition 2.3.4 *(Coherent Upper Prevision)*
An upper prevision $\overline{P}$ is said to be coherent when its conjugate lower prevision, defined by $\underline{P}(X) = -\overline{P}(-X)$, is coherent. Coherence of upper previsions is distinguished by conjugate versions of axioms A-1$-$A-3:

B-1 $\overline{P}(X) \leq \sup X$

B-2 $\overline{P}(\lambda X) = \lambda \overline{P}(X)$

B-3 $\overline{P}(X+Y) \leq \overline{P}(X) + \overline{P}(Y)$.

Axioms B-2 and B-3 imply that the coherent upper prevision, $\overline{P}$, is a convex function on $\mathcal{X}$. Upper and lower previsions do seem to be sufficiently general to model all common types of uncertainty. Upper and lower probabilities are special cases of upper and lower previsions, defined only for indicator functions of events.

Let $\mathcal{A}$ denote an arbitrary class of events, which is considered a class of 0-1 random variables. If the lower prevision is defined on a such class $\mathcal{A}$, $\underline{P}$ is called a ***lower probability*** on $\mathcal{A}$, and $\underline{P}(A)$is called the lower probability of event $\mathcal{A}$. Similarly, the conjugate upper prevision $\overline{P}$ is now called the ***upper probability***, so $\overline{P}$ is defined on $\mathcal{A}^c = \{A^c : A \in \mathcal{A}\} = \{1 - A : A \in \mathcal{A}\}$ rather than on $-\mathcal{A} = \{-A : A \in \mathcal{A}\}$. $\overline{P}$ is defined on $\mathcal{A}^c$ by $\overline{P}(A) = 1 - \underline{P}(A^c)$. As before, $\underline{P}(A)$ and $\overline{P}(A)$ are still interpreted as supremum buying price and infimum selling price, respectively.

Linear Previsions

Any coherent lower prevision defined in a linear space, which is a self-conjugate, is called a linear prevision. Self-conjugate means that $\underline{P}(X) = -\overline{P}(-X)$ for all $X \in \mathcal{X}$. When the lower and upper previsions coincide and are coherent, they will be called linear previsions and denoted by $P(X)$. The prevision $P(X)$ is called your fair price because you are both willing to buy X for any price less than $P(X)$ and to sell X for any price greater than $P(X)$. Before giving the formal definition of linear previsions, definitions of coherence and avoiding sure loss should be stated. According to Walley (1991, Chapter 2), coherence can be defined in general as following:

Definition 2.3.5 *(Coherence)*
Let $\mathcal{X}$ be a linear space of gambles on the sample space Ω, and Let $G(X)$ denote the marginal gamble $X - \underline{P}(X)$. The lower prevision $\underline{P}$ is coherent if $\sup\left[\sum_{j=1}^{n} G(X_j) - mG(X_0)\right] \geq 0$ *whenever m and n are non-negative integers and $X_0, X_1, ..., X_n$ are in $\mathcal{X}$.*

Definition 2.3.6 *(Avoiding sure loss)*
The lower prevision $\underline{P}$ avoids sure loss if $\sup \sum_{j=1}^{n} G(X_j) = \sup \sum_{j=1}^{n} [X_j(a) - \underline{P}(X_j)] \geq 0$ *whenever $n \geq 1$ and $X_1, X_2, ..., X_n$ are in $\mathcal{X}$.*

The following is a Toy example to illustrate the coherence and avoiding sure loss properties of the lower prevision.

Example 2.3.2 *Assume that you want to buy a house sometime soon, and you have looked to some online websites. Assume that there are three houses for sale, and the only information that you know about these three houses are provided through these websites. You know the asking price of the house, but because you have not seen the houses, you are uncertain about their true value. Let us say there are three possible (unknown) situations, $\Omega = \{h1, h2, h3\}$, that might affect the value of the houses as follows:*

- *$h1$: House 2 and house 3 do need repairs, but house 1 does not need repairs.*
- *$h2$: House 1 and house 3 do need repairs, but house 2 does not need repairs.*

- *$h3$: Only house 2 does need repairs, but house 1 and 3 do not need repairs.*

The uncertainty here is that you do not know if the house needs repairs or not, you have to take a decision or an action in the face of uncertainty (accepting a gamble). Now, consider the following three gambles,

- *X_1 means you are willing to accept buying house 1.*
- *X_2 means you are willing to accept buying house 2.*
- *X_3 means you are willing to accept buying house 3.*

Now, let us consider that the rewards are an extra value of the house (in thousands of dollars). For any gamble, you will receive a reward that depends on which of the situations $h1, h2, h3$, actually obtains. You know what the possible rewards are, but you do not know the actual situations. The rewards according to situations $h1$, $h2$, and $h3$ are

$$X_1(h1) = 100, \quad X_1(h2) = 70, \quad X_1(h3) = 100,$$

$$X_2(h1) = 70, \quad X_2(h2) = 100, \quad X_2(h3) = 70,$$

and

$$X_3(h1) = 70, \quad X_3(h2) = 70, \quad X_3(h3) = 100.$$

Assuming that you accept X_1, for example, accepting buying house 1 under $h1$ will give you a reward of \$100, accepting buying house 1 under $h2$ will give you a reward of \$70 (the house does need repairs), and accepting buying house 1 under $h3$ will give you a reward of \$100 (the house does not need repairs). Similarly for accepting gambles X_2 and X_3.

Assume that you are willing to pay up to \$70 for the gamble X_1, up to \$70 to get X_2, and up to \$100 to get X_3. Then $\underline{P}(X_1) = 70$, $\underline{P}(X_2) = 70$, and $\underline{P}(X_3) = 100$ are your lower previsions for the gamble X_1, X_2 and X_3 respectively.

Now, to demonstrate the avoiding sure loss property of the lower prevision, we need to look to Definition 2.3.6. In this definition, the sum, that has to hold for any $n \geq 1$, can

alternatively be written as a linear combination with non-negative integer coefficients a_j*, such that* $\sum_{j=1}^{3} a_j G(X_j) = \sum_{j=1}^{3} a_j[X_j(h) - \underline{P}(X_j)]$ *and regarding the scenario in this example, the supremum of this linear combination can be written as*

$$\sup\{a_1[X_1 - \underline{P}(X_1)] + a_2[X_2 - \underline{P}(X_2)] + a_3[X_3 - \underline{P}(X_3)]\} \geq 0, \tag{2.44}$$

and under $h1$*,* $h2$*, and* $h3$*, the inequality (2.44) can be expressed in* $30a_1 - 30a_3$*,* $30a_2 - 30a_3$*, and* $30a_1$*. To discuss the supremum over* $h1$*,* $h2$*, and* $h3$ *, we need to look at which values of* a_j*'s, as non-negative integers, this supremum will be achieved. Since* a_j *are non-negative, then* $a_1 - a_3 \leq a_1$ *which means that the supremum is not achieved at* $h1$ *(it could occur at* $h1$ *and* $h3$ *if* a_2 *and* a_3 *are zero). Thus the supremum will be achieved either at* $h2$ *or* $h3$*. In the case where* $a_2 - a_3 > a_1$*, the supremum is* $30a_2 - 30a_3$*, which is non-negative, and will be achieved at* $h2$*. If* $a_2 - a_3 > a_1$ *does not hold then the supremum is* $30a_1$*, which is non-negative, and will be achieved at* $h3$*. Therefore, the assessments of lower previsions on three gambles* X_1*,* X_2*, and* X_3 *avoid a sure loss.*

To illustrate the coherence property, we need to show that if lower previsions satisfy the three axioms A-1, A-2, and A-3 in the Definition 2.3.3, or not. The axiom A-1 is satisfied for the three gambles since $\underline{P}(X) \geq \inf X$ *(accepting sure gain). To satisfy A-2 and A-3 and ensure the coherence, we need to assign the lower prevision to multiples of* X_1*,* X_2*, and* X_3*, assuming* $\lambda = 2$*, for example,* $\underline{P}(\lambda X_1) = 140$*. Also, we need to assign the lower prevision on the sum of the three gambles such as* $\underline{P}(X_1 + X_2 + X_3) = 250$*.*

Now, $\underline{P}(2X_1) = 2\underline{P}(X_1) = 140$*, and* $\underline{P}(X_1 + X_2 + X_3) > \underline{P}(X_1) + \underline{P}(X_2) + \underline{P}(X_3)$*. Thus, axioms A-2 and A-3 are satisfied and the lower previsions are coherent according to Definition 2.3.3.*

Now, the definitions of coherence and avoiding sure loss allow to define the linear prevision as follows:

Definition 2.3.7 *(Linear Prevision)*

Suppose that P is a real-valued function defined on $\mathcal{X}$ (a class of gambles). Let $G(X)$ denote the marginal gamble $X - P(X)$. P is then called a linear prevision on $\mathcal{X}$ if

$$\sup\left[\sum_{j=1}^{n} G(X_j) - \sum_{j=1}^{m} G(Y_j)\right] \geq 0,$$

whenever m and n are non-negative integers and $X_1, .., X_n, Y_1, ..., Y_m$ are in $\mathcal{X}$.

Walley (1991) discussed de Finetti's terminology for defining the linear prevision, which is equivalent to the two axioms of de Finetti:

1. $P(X + Y) = P(X) + P(Y)$ when $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ (additivity).

2. $\inf X \leq P(X) \leq \sup X$ when $X \in \mathcal{X}$ (convexity).

Lower Envelopes of Linear Previsions

Every coherent lower prevision is a lower envelope of some class of linear previsions (Walley, 1991, Chapter 2). The next theorem shows that if there is a linear prevision, P dominates $\underline{P}$ on $\mathcal{X}$; that is, $P(X) \geq \underline{P}(X)$ for all $X \in \mathcal{X}$, then the lower prevision $\underline{P}$ avoids sure loss. Moreover, for $\underline{P}$ to be coherent it is sufficient that $\underline{P}$ is a **lower envelope** of a class $\mathcal{M}$ of linear previsions such that $\underline{P}(X) = \inf\{P(X) : P \in \mathcal{M}\}$ for all $X \in \mathcal{X}$.

Theorem 2.3.1 *(Lower envelope theorem)*

Suppose $\underline{P}$ is a lower prevision on domain $\mathcal{X}$, where $\mathcal{X}$ is an arbitrary subset of $\mathcal{L}$, where $\mathcal{L}$ is the set of all gambles on Ω.

1. *$\underline{P}$ avoids sure loss if and only if $\mathcal{M}(\underline{P})$ is non-empty (i.e., if and only if $\underline{P}$ is dominated by some linear previsions), where $\mathcal{M}(\underline{P})$ is the class of all linear previsions that dominate $\underline{P}$ on $\mathcal{X}$.*

2. *$\underline{P}$ is coherent if and only if it is the lower envelope of $\mathcal{M}(\underline{P})$ (i.e., if and only if it is the lower envelope of some class of linear previsions),*

where $\mathcal{M}(\underline{P})$ is a set of lower previsions on domain $\mathcal{X}$.

The basic idea that can be taken from this theorem is if we can define a class of linear previsions, and considering the infimum of the expectations over this class, then these will correspond to coherent lower previsions. For a clear illustration about concepts of avoiding sure loss and coherence, Augustin et al. (2014, Chapter 2) have given some examples in their recent book "Introduction to Imprecise Probabilities."

Degrees of Imprecision

The degree of imprecision is a measure of the imprecision regarding a gamble X, and it is defined as the difference between the upper and lower previsions,

$$\Delta(X) = \overline{P}(X) - \underline{P}(X), \tag{2.45}$$

The degree of imprecision is decreased as the number of observation is increased (amount of information is increased). The lack of information is the main source of imprecision (Walley, 1991, Chapter 5). In this work, the censored observations in survival data is the source of imprecision.

Credal Sets

In the theory of imprecise probabilities, upper and lower previsions (expectations) are playing the main role. A set, $\mathcal{M}$, of linear previsions is called *credal set* if it is closed and convex, then this set will be completely specified by its upper and lower previsions (Walley, 1991). To model the uncertainty about the parameter of interest, a set $\mathcal{M}$ of prior distributions is used in imprecise Bayesian approach. According to Benavoli and Zaffalon (2012), the set $\mathcal{M}$ of prior distributions should have a minimal property when there is no prior information about the parameter of interest. Such this property is that the set $\mathcal{M}$ should be large enough to model the uncertainty, and to avoid getting incoherent posterior in case of improper prior, but not too much to not allow making inference from the data. Bickis (2017) shows a number of examples of sets of priors that have Benavoli-Zaffalon (BZ) property.

2.3.3 Imprecise Bayesian Inference

In drawing imprecise Bayesian inferences from multinomial data, Walley (1996) introduced the *imprecise Dirichlet model*. The Dirichlet prior distribution in Section 2.1.3 is parameterized by a vector α with probability density function

$$\pi(\mathbf{p}) = \frac{1}{B(\alpha)} \prod_{i=1}^{k} p_i^{\alpha_i - 1},$$

where $\alpha_i > 0$ are parameters of the Dirichlet distribution, and $\sum p_i = 1$, $p_i \geq 0$. The parameters $\alpha_i = st_i$, where s is the concentration parameter and it is a positive constant, $s > 0$, and t_i is the mean of p_i such that $0 < t_i < 1$ and $\sum t_i = 1$. Walley (1996) suggests to choose s sufficiently large, $s = 1$ or 2. Walley (1996) defined the imprecise Dirichlet model as the set of all Dirichlet distributions. This set is used to model the prior ignorance about the parameter of interest.

Suppose that $\mathbf{y} \sim Multinomial(\mathbf{p}, n)$, $\mathbf{y} = (y_1, ..., y_k)$, $\mathbf{p} = (p_1, ..., p_k)$, and the probability mass function of the multinomial is;

$$f(\mathbf{y}|\mathbf{p}) = \frac{n!}{\prod_i y_i!} \prod_i p_i^{y_i},$$

where $i = 1, 2, ..., k$, $\sum y_i = n$ and $\sum p_i = 1$, $p_i \geq 0$. Then, the prior model is the set of all Dirichlet distributions that are parametrized with $(s, \mathbf{t})$, where s and $\mathbf{t} = (t_1, ..., t_k)$ are hyperparameters. This parameterization is convenient because s is fixed and $t's$ are the expectations of the parameters $p's$. The Dirichlet conjugate prior can be written as follows:

$$\pi(\mathbf{p}) \propto \prod_i p_i^{st_i - 1},$$

Then the posterior Dirichlet distribution is derived as:

$$\pi(\mathbf{p}|\mathbf{y}) \propto \prod_i p_i^{y_i + st_i - 1}.$$

This form is obtained by multiplying the prior function by the multinomial likelihood function. Therefore, the Dirichlet *posterior expectation* is:

$$t_i^* = E(\mathbf{p}|\mathbf{y}) = \frac{y_i + st_i}{n + s}.$$

By maximizing and minimizing t_i^* as $t_i \to 1$ and $t_i \to 0$, we will get the posterior upper and lower expectations:

$$\overline{P}(\mathbf{p}|\mathbf{y}) = \frac{y_i + s}{n + s}$$

and

$$\underline{P}(\mathbf{p}|\mathbf{y}) = \frac{y_i}{n + s}.$$

Walley (1996) and Walley et al. (1996) used the imprecise beta(s) model (a special case of the imprecise Dirichlet(s) model with $k = 2$ categories) to analyze data in the form of a contingency table. He illustrated this approach with an example of data from medical trials in which a comparison is made of two treatments for resistant pulmonary hypertension in newborn babies. The treatments are conventional therapy (CT) using ventilation with oxygen at high pressure, and extracorporeal membrane oxygenation (ECMO), a technique which uses heart-lung bypass technology to oxygenate blood outside the body. The babies were assigned to treatments randomly and the stage of the trial was terminated as soon as four deaths had occurred in one of the treatment groups. The outcome was that 6 of the 10 babies who received CT survived, and all 9 babies received ECMO survived. This data are displayed in a 2×2 contingency table as shown in Table 2.1.

Table 2.1: A 2×2 contingency table for CT and ECMO data.

	Event		
Group	Death	Survivor	Total
CT	4	6	10
ECMO	0	9	9
Total	4	15	19

Considering this data, two statistical problems are discussed. The first one is making inference about which treatment is more effective, and the second is taking a decision about which treatment should be preferred to another for the next patient or whether it is ethical to select the treatment by randomization.

Walley et al. assumed that babies have a constant chance of survival under each treatment, denoted by θ_c for CT and θ_e for ECMO and defined the difference between the probability of survival under each treatment, $\psi = \theta_e - \theta_c$. In the inference part for imprecise beta model, to know which treatment is more effective, a test of null hypothesis $H_0 : \theta_e \leq \theta_c$ can be formulated against $H_a : \theta_e > \theta_c$. The conclusion for this test can be derived from calculating the posterior upper and lower probabilities, for example, when $s = 2$, $\overline{P}(H_0|n) = 0.185$, equivalently, $\underline{P}(H_a|n) = 0.815$. This indicates evidence to support that ECMO is more effective than CT. Also the data support this conclusion since all babies treated with ECMO survived.

Considering the decision problem to figure out the preferred treatment. Walley et al. used the upper and lower posterior expectations of $\psi = \theta_e - \theta_c$ as follows:

$$\underline{E}(\psi, y) = \underline{E}(\theta_e|y) - \overline{E}(\theta_c|y) = 0.152$$

$$\overline{E}(\psi, y) = \overline{E}(\theta_e|y) - \underline{E}(\theta_c|y) = 0.5$$

Thus, the ECMO should be preferred over CT if $\underline{E}(\psi, y)$ is greater than 0. Similarly, CT should be preferred if $\overline{E}(\psi, y)$ is less than 0. Walley's conclusion was that ECMO is more effective than CT and is the preferred treatment.

Bayesian inference for survival data including right-censored observations has been done imprecisely. Early in this subsection, we discuss the work by Walley (1996) about using the imprecise Dirichlet model related to multinomial data in a Bayesian workframe; however, he does not consider survival data with right-censored observations. Coolen (1997) presents an update of Walley's imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations to introduce nonparametric estimates of survivor functions. Coolen considers multinomial distribution with Dirichlet priors, making the approach basically nonparametric, and the model uses a finite partition of the time-axis such that makes it becomes related to life-tables. In contrast, Coolen's work does not show a description of the upper and lower hazard functions. Bickis (2009) introduces the imprecise logit-normal model as a family of prior distributions for a binomial success probability. The

model is constructed by giving the logit of the probability a normal distribution. This leads to a three-dimensional exponential family. Bickis generalizes this model to the multivariate case where some restrictions has been made on the hyperparameters to be a suitable chosen subset. Then, the extremes of the posterior expectations are computed to give imprecise predictive probabilities. Bickis and Bickis (2007) have sought to understand phenomena such as Influenza pandemic, in their study of predicting the next pandemic: An exercise in imprecise hazards, they found that there is an increasing hazard after 25 years. In that work, Dirichlet and product Beta models are imposed and both models shows the same results. Upper and lower survivor functions are plotted comparing with Kaplam-Meier estimate where the last lies between the upper and lower survivor curves, as indicated by Coolen (1997) and Coolen and Yan (2003) and has shown later in Figure 2.1.

Also, Coolen's approach is closely related to the method of Berliner and Hill (1988), where their work concentrates more on predictive inference for a future observation, which is the focus in the next section.

2.3.4 Nonparametric Predictive Inference

Predictive inference is an approach of statistical inference that stresses the prediction of a future observation based on past observations. A simple and famous example in predictive inference is the sunrise problem. Given information about the observed weather for n days ago, the question of interest is "what is the probability that the sun will rise tomorrow ($n+1$ day)?" This emphasis has changed due to the idea of exchangeability by De Finetti (1974) that future observations should behave like past observations, and since being presented by Geisser (1993, chapter 2,3), this type of predictive inference has been called low structure inference. Considering a parametric framework, the prediction of future values of a random variable based on past observed data is obtained by calibrated prediction intervals and frequentist predictive distributions (Lawless and Fredette, 2005).

Nonparametric predictive inference (NPI) is an imprecise approach based on few assumptions and quantifies uncertainty in terms of lower and upper probabilities. This kind of inference is based on Hill's assumption $A_{(n)}$ that was introduced to make predictions about

the occurrence of a future observation given a number of observed quantities. Hill (1968) defines $A_{(n)}$ as asserting that, conditional upon the observations $X_1, ..., X_n$, the next observation X_{n+1} is equally likely to lie in the open intervals between successive order statistics of a given sample. Moreover, the definition can be given in three points:(1) exchangeability; (2) ties that have a probability of 0; (3) given data $x_i, i = 1, ..., n$, the probability that the next observation lies in the open interval $I_j = (x_j, x_{j+1})$ is $1/(n+1)$, for all $j = 0, ..., n$, where $x_0 = -\infty$ and $x_{n+1} = \infty$.

Coolen (1996) compares two populations based on related low structure assumption where such comparison is expressed in terms of comparison of future observations from two different groups using both imprecise probabilities and imprecise previsions. For example, Table 2.2 shows birth weights (in grams) for $n = 12$ male and $m = 12$ female babies.

Table 2.2: Ordered birthweights

Male (X)	2625 2628 2795 2847 2925 2968 2975 3163 3176 3292 3421 3473
Female (Y)	2412 2539 2729 2754 2817 2875 2935 2991 3126 3210 3231 3317

In classical statistics, the comparison of this kind of data has been done by testing a hypothesis; for example, that both data sets are randomly drawn from the same population. In NPI, the predictive comparison has been done by comparing the random birthweight of a future male (X_{n+1}) and a future female (Y_{m+1}), with the assumption of exchangeability with the 12 observed birth weights in two groups, and assuming $A_{(12)}$ for each group.
The NPI lower and upper probabilities are

$$\begin{aligned} \underline{P}(X_{n+1} > Y_{m+1}) &= \frac{1}{(n+1)(m+1)} \sum_{j=0}^{n-1} (n-j) s_j \\ &= \underline{P}(X_{13} > Y_{13}) = \frac{86}{169} = 0.509 \end{aligned}$$

and

$$\overline{P}(X_{n+1} > Y_{m+1}) = \frac{1}{(n+1)(m+1)} \{n + m + 1 + \sum_{j=0}^{n-1} (n-j) s_j\}$$

$$= \quad \overline{P}(X_{13} > Y_{13}) = \frac{111}{169} = 0.657,$$

where s_i is the number of observed y values per intervals bounded by sequential x values, then

$$s_i = \#\{y_j | x_i < y_j < x_{i+1}, j = 1, ..., m\}, \quad i = 0, ..., n-1,$$

such that $s_0 = \#\{y_j | -\infty < y_j < x_1\}$ and $s_n = \#\{y_j | x_n < y_j < \infty\}$.

In the conclusion of this example, Coolen indicates some sign that $X_{13} > Y_{13}$, but it is not very strong evidence.

Nonparametric Predictive Inference with Right-censored Data

The NPI approach is introduced for a prediction about future observations in the form of lower and upper probabilities and has been used in data including right-censored observations (Berliner and Hill, 1988); (Coolen and Yan, 2003), and some applications in reliability and operational research as summarized by Coolen (2010).

The assumption $A_{(n)}$ proposed by Hill in 1968, as discussed in Subsection 2.3.4 provides a partially specified predictive distribution for a future observation, X_{n+1}, given past observations. However, it does not allow right-censoring of data among the observations. Berliner and Hill (1988) and Coolen and Yan (2003) presented related nonparametric methods for dealing with survival data, including right-censoring. Berliner and Hill replaced the exact observed right-censoring times by "partial censoring information," shifting each exact right censoring time back to the nearest smaller observed time, allowing for inference on the basis of $A_{(n)}$ alone. Coolen and Yan developed this work by using exact censoring information and adding more assumption, which is called a right-censoring $A_{(n)}$ (rc-$A_{(n)}$) assumptions. The assumption rc-$A_{(n)}$ gives a partially specified predictive probability distribution for future observation when data have right-censoring observations, and the assumption is expressed via a so-called the M-function value. M-function is an approach introduced by Coolen and Yan to give probabilities for the future observation on intervals, including right-censoring times. Based on these probabilities, Coolen and Yan introduce *upper* and *lower* survival functions.

Definition (M-function): A partial specification of a probability distribution for a real valued random quantity T can be provided via probability masses assigned to intervals without any further restriction on the spread of the probability mass within each interval. A probability mass assigned in such a way to an interval (a, b) is denoted by $M_T(a, b)$ and referred to as M-function value for T on (a, b). The interval in the definition of the M-function is an open interval because it assumes no ties in the data, and the M-function values on nonspecified intervals are assumed to be zero. All M-function values for T on all intervals should sum to one, so each M-function value should be in [0,1].

The concept of M-function takes our minds to talk a little bit about the *belief function*. Dempster-Shafer theory (Shafer, 1976) is a generalization of the Bayesian theory of subjective probability; belief functions base degrees of belief (or confidence, or trust). The degree of belief is represented by the belief function which can be defined as following:
A function $Bel : 2^{\Omega} \to [0, 1]$ is a *belief function* on Ω if and only if it satisfies the following conditions:

1. $Bel(\emptyset) = 0$
2. $Bel(\Omega) = 1$
3. For every integer n and every collection $A_1, ..., A_n$ of subset of Ω,

$$Bel(A_1 \cup ... \cup A_n) \geq \sum_{\substack{I \subset 1,...,n \\ I \neq \emptyset}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i) \tag{2.46}$$

Where Ω is a finite set of all possible answers about any uncertainty of events of interest, and 2^{Ω} is the power set.

The definition of M-function and belief function are same where the values of both functions should be in the interval $[0, 1]$. Also, later in this section, we will see that upper and lower survival functions are introduced in terms of imprecise probabilities where these probabilities are defined in terms of M-function.

Before we start talk about Coolen and Yan's rc-$A_{(n)}$ assumption, we should look at the definition of M-function on an open interval without any restrictions (Coolen and Yan,

2003). According to $A_{(n)}$ assumption, predictive probabilities are represented by:

$$M_{X_{n+1}}(t_{(j)}, t_{(j+1)}) = \frac{1}{n+1}, \tag{2.47}$$

and all $j = 0, ..., n$, which is the probability distribution of a future observation X_{n+1} lies on the open interval $(t_{(j)}, t_{(j+1)})$. Since a partial specification of a probability distribution for a random quantity is available in terms of M-function values, then minimum upper and maximum lower bounds can be specified by that probability (M-function) of the bounds of this interval.

Let u be the number of the observed death, $0 \leq u \leq n$, and these times are observed at ordered times, $0 < t_{(1)} < ... < t_{(u)}$, and $v = u - n$ be the number of non-observed times (the ordered right-censored times). Let $I_{(i)} = (t_{(i)}, t_{(i+1)})$, for $i = 0, ..., u$, where $t_{(0)} = 0$ and $t_{(u+1)} = \infty$, and let $\tilde{n}_t = n_t + 1$, n_t be the number of individuals in the risk set . The ordered right-censored times within $I_{(i)}$ is denoted by $c^i_1 < c^i_2 < ... < c^i_{l_i}$, where l_i is the number of right-censored times in $I_{(i)}$. While the M-function is defined based on the assumption $A_{(n)}$, Coolen and Yan introduce their new assumption "right-censoring $A_{(n)}$", (rc-$A_{(n)}$), to provide a partially specified probability distribution for the observable random quantity $X_{(n+1)}$, via the M-function values on the intervals $(t_{(i)}, t_{(i+1)})$ and $(c^i_k, t_{(i+1)})$. These M-function values can be used to derive lower and upper probabilities for events of interest in terms of X_{n+1}.

Definition (rc-$A_{(n)}$): The assumption "right-censoring $A_{(n)}$," (rc-$A_{(n)}$), is a probability distribution for a nonnegative random quantity X_{n+1}, on the basis of data including u event times and v right-censoring times, and is partially specified by the following M-function values:

$$M_{X_{n+1}}(t_{(i)}, t_{(i+1)}) = \frac{1}{n+1} \prod_{\{r: c_{(r)} < t_{(i)}\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}}, \tag{2.48}$$

$$M_{X_{n+1}}(c^i_k, t_{(i+1)}) = \frac{1}{(n+1)\tilde{n}_{c^i_k}} \prod_{\{r: c_{(r)} < c^i_k\}} \frac{\tilde{n}_{c_{(r)}} + 1}{\tilde{n}_{c_{(r)}}}, \tag{2.49}$$

the product over an empty set is defined as 1. Coolen and Yan introduced the probabilities of $X_{n+1} \in (t_{(i)}, t_{(i+1)})$ in terms of the M-function values and considering the case of existing

of right-censoring observation, so these probabilities lead to the ability to write the following relation:

$$P(X_{n+1} \in (t_{(i)}, t_{(i+1)})) = \frac{1}{n+1} \prod_{\{r:c_{(r)}<t_{(i+1)}\}} \frac{\tilde{n}_{c_{(r)}}+1}{\tilde{n}_{c_{(r)}}} = M_{X_{(n+1)}}(t_{(i+1)}, t_{(i+2)}), \tag{2.50}$$

for $i = 0, ..., u-1$. Based on these probabilities, Coolen and Yan mentioned that lower and upper survival functions are equal at observed event time $t_{(i)}$, and their value can be derived as follows:

$$\underline{S}_{X_{n+1}}(t_{(i)}) = \overline{S}_{X_{n+1}}(t_{(i)}) = \sum_{j=i}^{u} P(X_{n+1} \in (t_{(j)}, t_{(j+1)})). \tag{2.51}$$

Coolen and Yan introduced the *upper survival function* based on the fact that the M-function values in $rc\text{-}A_{(n)}$ are all defined on intervals with an observed event time (or infinity),(t, ∞), and so the upper survival function is defined as the sum of all the $rc\text{-}A_{(n)}$-based M-function values defined on intervals starting at $t_{(i)}$ or greater values. For the intervals starting at right-censoring times $c_k^i \in (t_{(i)}, t_{(i+1)}))$, $k = 0, ..., L_i$, they are all defined on intervals $(c_k^i, t_{(i+1)}))$, which can be represented by a subinterval $(t, t_{(i+1)})$ of $(t_{(i)}, t_{(i+1)}))$ where $t > 0$. Thus that leads to:

$$\bar{S}_{X_{n+1}}(t) = \bar{S}_{X_{n+1}}(t_{(i)}), \tag{2.52}$$

for $i = 0, ..., u$ and all $t \in [t_{(i)}, t_{(i+1)})$.

The *lower survival function* for X_{n+1} at $t > 0$ is derived by summing the $rc\text{-}A_{(n)}$-based M-function values for intervals that completely lie in (t, ∞), which leads to, for $i = 0, ..., u$

$$\underline{\mathrm{S}}_{X_{n+1}}(t) = \sum_{j=i+1}^{u} P(X_{n+1} \in (t_{(j)}, t_{(j+1)})) + \sum_{\{k:c_k^i \geq t\}} M_{X_{n+1}}(c_k^i, t_{(i+1)}) \tag{2.53}$$

for $t \in [t_{(i)}, t_{(i+1)})$. The lower and upper survival function are defined in terms of the M-function which is a belief function. According to Walley (1991, chapter 5), a belief function is a special type of coherent lower probability that satisfies (2.46), the extra property of complete monotonicity. A lower probability $\underline{P}$, defined on all subsets of Ω, is a belief function if and only if it can be written in the form

$$\underline{P}(A) = \sum_{B \subset A} m(B)$$

for all sets A, where m is a probability mass function defined on all subsets of Ω such that

- $m(\emptyset) = 0$,
- $m(B) \geq 0$ for all subsets B, and
- $\sum_{B \subset \Omega} m(B) = 1$.

The function m is called probability assignment (Shafer, 1976). Then, one can say that the function m is just the lower survival function.

Coolen and Yan (2003) make a comparison between upper and lower survival functions and a Kaplan-Meier estimate. Figure 2.1 is a graphed example to show that the Kaplan-Meier estimate lies between the upper and lower survival functions. Coolen and Yan's lower survival function for the next observation X_{n+1} is going to the zero after the largest observation, which is also the case for the Kaplan-Meier estimator if this observation is an event time. The data that used to produce the following plot are from Collett (2003), prognosis for women with breast cancer. The survival times are as follows:

23 47 69 70* 71* 100* 101* 148 181 198* 208* 212* 224*.

The $(*)$ indicates censored survival times.

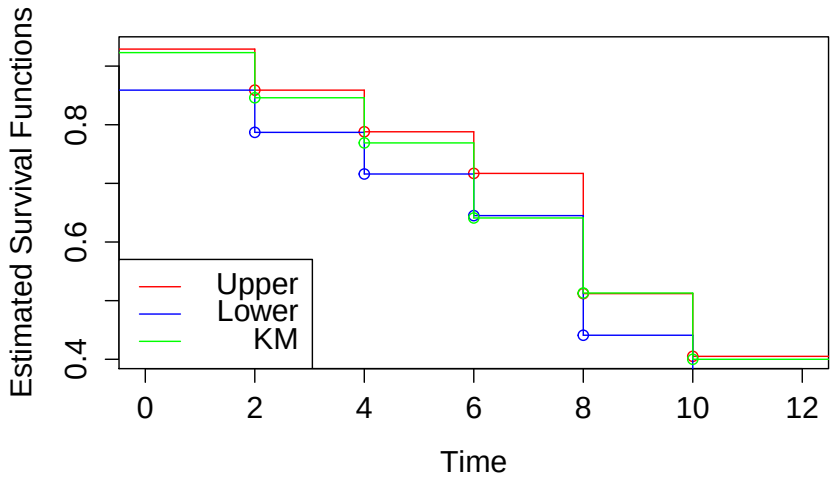


Figure 2.1: Comparison between upper and lower survival functions and Kaplan-Meier estimate

2.4 Numerical Techniques

In Bayesian inference, computing the posterior, $\pi(\theta|Y)$, requires evaluating the integral in the dominator of Eq.(2.30) analytically. Computing this integral might be intractable. For this reason, the literature on the computational methods has grown recently. One way has become a popular way of sampling from posterior distributions is the Markov Chain Monte Carlo (MCMC) methods. Markov Chain is the process of sampling a new value from the posterior distribution, given the previous value, this iterative process produce a Markov Chain of values that establish a sample of draws from the posterior. Markov Chain can be described as follows: Suppose we have a set of states, $S = \{s_1, ..., s_r\}$. The process start in one state and moves to another, and each move is called a step. The chain moves from state to the next one with a probability called transition probability and distribution called a stationary distribution. MCMC works by constructing a chain whose stationary distribution is the desired posterior distribution. MCMC methods produce an approximation of the posterior distribution, $\pi(\theta|Y)$, by sampling a large number of θ values from that distribution after running the algorithm long enough, and then these θ values can be used to estimate the central tendency of the posterior, its highest density interval (HDI), etc. The two common MCMC methods are Metropolis-Hastings (M-H) algorithm and the Gibbs sampling (Lynch, 2007).

In this work, in Section 3.2, an optimization technique is used to calculate the upper and lower posterior expectations (Eq.(3.3)) of log-odds ratio in 2×2 table. This optimization is done by considering the constrain on the values of the prior means t_{ij}, where $0 < t_{ij} < 1$ and using the function `optim` in `R`.

The implementation of MCMC method is done through using the programming language JAGS (Just Another Gibbs Sampler) from `R` via `rjags` and `runjags` packages. In general, the main goals in generating an MCMC sample from the posterior distribution is the convergence of the chain and how much information about the posterior does the chain contain?

Visual examinations of the chain's convergence are a *trace* plot, which is a graph of the sampled parameter values as a function of step in the chain. Trace, density, and empirical CDF plots are visualized in this thesis. A good background about MCMC methods and JAGS can be found in Kruschke (2015). In Sections 3.3 and 4.4, the posterior sample is sampled by running MCMC by considering the extreme points of a set of prior distribution.

In this work, normal and Beta priors are considered. The upper and lower posterior expectations are estimated as maximum and minimum of the posterior expectation and then the degree of imprecision, as defined in Section 2.3.2, is calculated. To get a 95 % imprecise Bayesian credible intervals, the .025 and .975 quantiles of the posterior sample are calculated by using the `quantile` function. Then the set of priors will give a set of intervals, the supremum of the upper limits and the infimum of the lower limits will give an imprecise credible interval. The `R` codes for this work are provided in Appendix B.

3. Bayesian Imprecise Inference for Log-odds Ratio in 2×2 Tables

3.1 Motivation

The prior distribution represents the uncertainty about the parameter of interest before data are observed. As discussed in the previous subsection, the traditional (precise) probability theory has limitations, the most crucial one is when we have little or no information for assessing a single probability of an event; say A. In Section 2.3.3 a discussion about drawing imprecise Bayesian inferences from multinomial data and imprecise Dirichlet model have taken a place. Whereas Walley (1996) and Walley et al. (1996) has focused on the parameter $\psi = \theta_e - \theta_c$, the difference between the probability of survival under each treatment in 2×2 table (Table 2.1), this thesis is mostly concerned with the odds ratio (Section 2.1) since the odds ratio is more informative for the comparison of two probabilities than their difference (Kateri, 2014), in two-way and three way contingency tables.

The contribution in this chapter will be introduced in two approaches. The first approach is in the essence of Walley (1996) and in the case we have two treatments with two outcomes and the data are displayed in a 2×2 table. The second approach is a re-parametrization and alternative priors for 2×2 table considering the four sampling schemes in two-way contingency tables. Lower and upper posterior expectations of log-odds ratio will be derived in both approaches.

3.2 Imprecise Dirichlet Approach

In Section 2.3.3, imprecise Dirichlet model by Walley (1996) is discussed with an example about comparing two treatments in the case of 2×2 table (Table 2.1). Now, consider the situation where the row totals in the table are assumed to be fixed and the joint probability distribution of the cell counts y_{ij}'s will be product of two binomials with parameters p_{ij}'s. Therefore, the appropriate prior is a product of *Beta* distributions as conjugate priors for the cell probabilities, $\boldsymbol{p}$, where $\boldsymbol{p} = (p_{00}, p_{01}, p_{10}, p_{11})'$. The beta distribution here is reparameterized in terms of the concentration parameter s, and the means of cell probabilities t_i, where $s > 0$, $0 < t_i < 1$, and $\sum_i t_i = 1$,

$$\pi(\boldsymbol{p}) \propto \prod_i p_i^{st-1}(1-p_i)^{s(1-t_i)-1}.$$

Thus, the imprecise prior model can defined as a set of conjugate prior distributions that takes the following form,

$$\mathcal{M}_0 = \{Beta(st_i, (1-t_i)) : t_i \in \Omega\},$$

where $\Omega = \{(0,1) \times (0,1)\}$ is the parameter space. The prior knowledge will then be updated via Bayes rule, which means updating each element in the set $\mathcal{M}_0$ in light of the observed sample. Thus, the posterior product of beta distribution is defined as follows:

$$\pi(\boldsymbol{p}|y, st_i) \propto \prod_i p_i^{y_{ij}+st_i-1}(1-p_i)^{n-y_{ij}+s(1-t_i)-1}.$$

The set $\mathcal{M}_|y$ of posterior distributions can take the following form:

$$\mathcal{M}_|y = \{Beta(y_{ij} + st_i, n - y_{ij} + s(1-t_i)) : t_i \in \Omega\}.$$

From here, one can start with defining the expectation of the log-odds ratio in (2.3) as follows:

$$\begin{aligned} E(\log \Psi) &= E(\log p_{11} - \log p_{10}) - E(\log p_{01} - \log p_{00}) \\ &= E(\log p_{11}) - E(\log p_{10}) - E(\log p_{01}) + E(\log p_{00}). \end{aligned} \tag{3.1}$$

Then the lower and upper expectations can be written in the forms:

$$\underline{P}(\log \Psi) \geq \underline{P}(\log p_{11} - \log p_{10}) - \overline{P}(\log p_{01} - \log p_{00}),$$

$$\geq \underline{P}(\log p_{11}) - \overline{P}(\log p_{10}) - \overline{P}(\log p_{01}) + \underline{P}(\log p_{00}),$$

and

$$\begin{aligned}\overline{P}(\log \Psi) &\leq \overline{P}(\log p_{11} - \log p_{10}) - \underline{P}(\log p_{01} - \log p_{00}),\\ &\leq \overline{P}(\log p_{11}) - \underline{P}(\log p_{10}) - \underline{P}(\log p_{01}) + \overline{P}(\log p_{00}).\end{aligned}$$

The prior of each p_{ij} can be modelled by an imprecise beta model, $p_{ij} \sim Beta(st_{ij}, s(1-t_{ij}))$. In order to get the lower and upper posterior expectations of the log-odds ratio, we need first to find the posterior expectations for each $\log p_{ij}$. For example:

$$\begin{aligned}E(\log p_{11}) &= \int_0^1 \log p_{11} \frac{\Gamma(n + st_{11} + s(1-t_{11}))}{\Gamma(y_{11} + st_{11})\Gamma(n - y_{11} + s(1-t_{11}))}\\ &\quad p_{11}^{y_{11}+st_{11}-1}(1-p_{11})^{n-y_{11}+s(1-t_{11})-1}\, dp_{11}\\ &= \psi(y_{11} + st_{11}) - \psi(n+s), \qquad (3.2)\end{aligned}$$

where $\psi(.)$ is called the digamma function, the logarithmic derivative of the gamma function. Similarly, we can get $E(\log p_{00})$, $E(\log p_{01})$, and $E(\log p_{10})$ with the same way in equation 3.2. Then we can write the $E(\log \Psi)$ by substituting each $E(\log p_{ij})$ in equation 3.1, and therefore the posterior expectation of the log-odds ratio can take the following form:

$$E(\log \Psi) = \psi(y_{11} + st_{11}) - \psi(y_{10} + st_{10}) - \psi(y_{01} + st_{01}) + \psi(y_{00} + st_{00}). \qquad (3.3)$$

The lower and upper bounds of the expectation in (3.3) over a posterior set can be obtained numerically by doing a constrained optimization problem. This optimization is done by considering the constraint on the values of the prior means t_{ij}, where $0 < t_{ij} < 1$ and $s = 1$. Consider the following 2×2 Table:

	Event		
Group	0	1	Total
0	1	2	3
1	3	1	4
Total	4	3	7

Optimization for the posterior expectation of log- odds ratio in eq.(3.3) is done and upper and lower posterior expectations and the degree of imprecision are displayed in Table 3.1,

where $\Delta_{\log \Psi}$ is 0.018. Figures 3.1 and 3.2 are contour and perspective plots of the posterior expectation of log-odds ratio in eq.(3.3). Equation (3.3) is in terms of digamma function where y's are given in Table ??, $s = 1$, and t_{ij}'s. Here, t_{00} is fixed, $t_{11} = 1 - (t_{00} + t_{01} + t_{10})$, where$0 < t_{01}, t_{10} < 1$. The interpretation of Figures 3.1 and 3.2 is that the posterior expectation function of log-odds ratio in eq.(3.3) is (almost) linear.

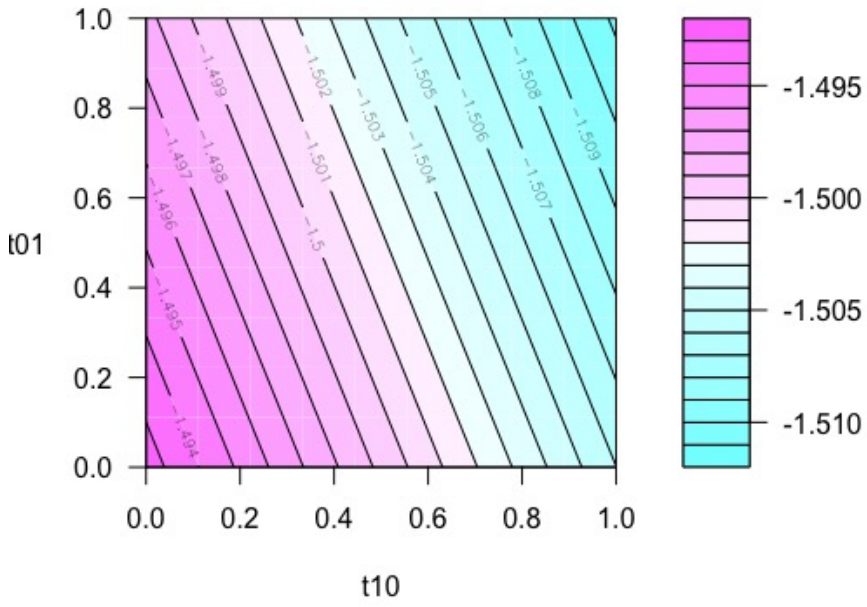


Figure 3.1: The contour plot of the posterior expectation of log odd-ratio

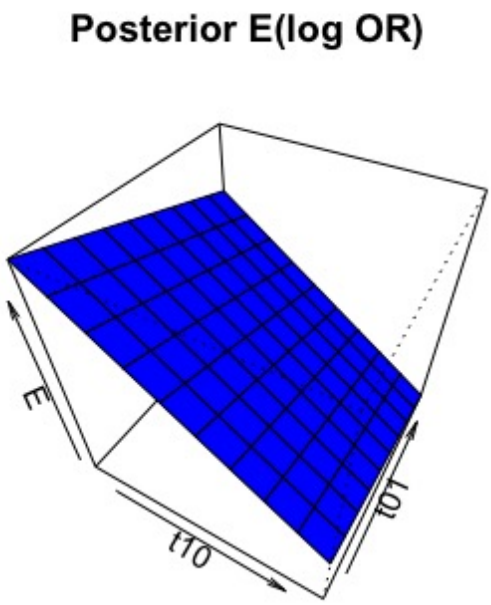


Figure 3.2: The perspective plot of the posterior expectation of log odd-ratio

The

Table 3.1: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio.

Prior	sample size	$\overline{E}(\log \Psi)$	$\underline{E}(\log \Psi)$	$\Delta_{\log \Psi}$
product of betas	7	-1.492	-1.511	0.018

3.3 Re-parametrization and Alternative Priors for 2×2 Table

The tractable choice of the prior distribution in Bayesian statistical inference for contingency tables is the Dirichlet prior for its computational convenience, as it is discussed in the

previous Section (3.2). However, according to Agresti and Hitchcock (2005) and Knuiman and Speed (1988), a one-stage Dirichlet prior does not always provide a sufficient structure to be given for cell probabilities. The choice of normal prior for the log probabilities as we see in Leonard (1975) is an alternative choice to Dirichlet prior for cell probabilities. Knuiman and Speed (1988) used a structured multivariate normal prior for the parameters in log-linear model collectively instead of giving a univariate normal prior for each parameter individually as in Leonard (1975). More recently, Forster (2010) develops the results of Lindley (1964) to provide a general framework for the analysis of multinomial data using Poisson log-linear model. Forster's focus is particularly on multivariate normal prior distributions for the log-linear parameters.

In the Subsection 2.1.1, four sampling schemes for contingency table are discussed. The nice property that all four sampling models have is that their log likelihood functions can be written in the *exponential family* form as follows:

$$l(\boldsymbol{\theta}; y) = \boldsymbol{\theta} y - \phi(\boldsymbol{\theta}), \tag{3.4}$$

where $\boldsymbol{\theta}$ is a vector of canonical parameters and $\phi(\boldsymbol{\theta})$ is the cumulant function. Considering the multinomial distribution with being a member of the exponential family will take the following form:

$$f(y|\boldsymbol{p}) = \frac{n!}{\prod_k y_k!} \exp\{\sum_{k=1}^{K} y_k \log p_k\},$$

However, in multinomial distribution we need to consider the restriction $\sum_k p_k = 1$, then we can express the likelihood as following:

$$\begin{aligned} f(y|\boldsymbol{p}) &\propto \exp\left\{ y_1 \log p_1 + y_2 \log p_2 + y_3 \log p_3 + (1 - \sum_{k=1}^{K-1} y_k) \log(1 - \sum_{k=1}^{K-1} p_k) \right\} \\ &\propto \exp\left\{ \sum_{k}^{K-1} \log(\frac{p_k}{1 - \sum_{k=1}^{K-1} p_k}) y_k + \log(1 - \sum_{k=1}^{K-1} p_k) \right\} \end{aligned}$$

From this representation we can define the canonical parameter

$$\theta_k = \log \frac{p_k}{1 - \sum_{k=1}^{K-1} p_k} = \log \frac{p_k}{p_K},$$

where $k = 1, ..., K$, and p_k can be expressed in terms of θ_k by taking the exponential of the Eq. above

$$p_k = \frac{e^{\theta_k}}{\sum_{k=1}^{K-1} e^{\theta_k}}.$$

Now, by taking the logarithm of both sides, we can get what is known as the multinomial logit as following:

$$\begin{aligned} \log p_k &= \theta_k - \log(\sum_{k=1}^{K} e^{\theta_k}) \\ &= \theta_k - \phi(\boldsymbol{\theta}), \end{aligned} \tag{3.5}$$

where $\phi(\boldsymbol{\theta}) = \log(\sum_{k=1}^{K} e^{\theta_k})$ plays the rule of a normalizing constant which guarantees that $\sum_k p_k = 1$.

Now, one can parametrize the multinomial distribution of a single observation, let's say z_{ij} that indicates which of the four cells is observed. The likelihood in this case for n independent observations would be just the product of the likelihoods of the observations, which would be of the same form. The observations of the table can be denoted as:

z_{00}	z_{01}
z_{10}	z_{11}

with a single observation which means only one cell entries is one and others are zeros. Let us consider we have n observations with new variables

$$l_1 = z_{10} + z_{11} - \frac{1}{2},$$

$$l_2 = z_{01} + z_{11} - \frac{1}{2},$$

and

$$l_3 = z_{00} + z_{11} - \frac{1}{2},$$

thus the l_k variables quantify the deviation of the observation from the uniform expected value of $\frac{1}{4}$ in the cells. Now we can rewrite the multinomial logit function in (3.5) in terms of l's as:

$$\log p_{ij} = l_1(i,j)\theta_1 + l_2(i,j)\theta_2 + l_3(i,j)\theta_3 - \phi(\boldsymbol{\theta}), \quad i,j = 0,1. \tag{3.6}$$

where

$$\begin{aligned}\phi(\boldsymbol{\theta}) &= -\frac{1}{4}\log\prod_{i,j} p_{ij} \\ &= \log(1+e^{(\theta_1-\theta_3)}+e^{(\theta_2-\theta_3)}+e^{(\theta_1+\theta_2)}) - 1/2(\theta_1+\theta_2-\theta_3). \end{aligned} \tag{3.7}$$

Where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. Therefore, we one can define a matrix L as follows:

$$L = \begin{bmatrix} -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 \end{bmatrix}$$

where the matrix L displays the relationship between z_{ij}'s and l_k's. The rows represent cells in the 2×2 table where each row's element represent a cell in a 2×2 table. The columns in L represent the l_k's above. Thus, the equation (3.6) can rewritten in matrix form as

$$\log \boldsymbol{p} = \boldsymbol{L\theta} - \phi(\boldsymbol{\theta}). \tag{3.8}$$

Now, assume we have n i.i.d. observations of z's, then $y_{ij} = \sum_{ijk} z_{ijk}$, where $k = 1, ..., n$, and $n = \sum_{ij} y_{ij}$. Therefore, the right hand side of the log-likelihood in equation (2.16) will be also written in a matrix format as following:

$$\sum_i \sum_j y_{ij} \log p_{ij} = \boldsymbol{y'L\theta} - n\phi(\boldsymbol{\theta}). \tag{3.9}$$

It can be noticed that the distributions of the 2×2 tables under multinomial sampling represent an exponential family, with θ's being canonical parameters and l's being minimal sufficient statistics. The next focus in this work is choosing a prior distributions for the canonical parameters θ's and then concentrate in the posterior distribution of the parameter of interest, log odds ratio, which is in this parametrization equals to $2\theta_3$ as defined in Eq. (2.20).

In the essence of Forster (2010) and Knuiman and Speed (1988), suppose that we put a multivariate normal prior on parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ with a prior mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$ and variance-covariance matrix $\sigma^2 I$, taking the following form

$$\pi(\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\}, \tag{3.10}$$

Therefore, the posterior density of θ's will have the expression:

$$\begin{aligned}\pi(\boldsymbol{\theta}|l,\mu,\sigma^2) &= \frac{1}{\mathcal{C}}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\}\exp\left\{\boldsymbol{y'L\theta} - n\phi(\boldsymbol{\theta})\right\} \\ &= \frac{1}{\mathcal{C}}\exp\left\{\boldsymbol{y'L\theta} - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\}\exp\left\{-n\phi(\boldsymbol{\theta})\right\} \quad , \end{aligned} \tag{3.11}$$

where $\mathcal{C}$ is the normalizing constant,

$$\mathcal{C} = \int_\theta \exp\left\{\boldsymbol{y'L\theta} - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right\}\exp\left\{-n\phi(\boldsymbol{\theta})\right\} d\theta. \tag{3.12}$$

In this work, estimating the parameter of interest has been done using MCMC methods through applying Gibbs sampling algorithm using JAGS and `R` programs.

3.3.1 An Example

Consider the log-likelihood in Eq. (3.9) with data from the following table,

Table 3.2: A 2×2 contingency table.

	Event		
Group	0 (Alive)	1 (Dead)	Total
0 (control)	3	5	8
1 (test)	7	2	9
Total	10	7	17

A normal prior with parameters μ and σ^2 for each one of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ is given and updated to the posterior using the log-likelihood in Eq. (3.9). The parameters $\boldsymbol{\theta}$ are estimated by running MCMC using `runjags` in JAGS from within `R`.

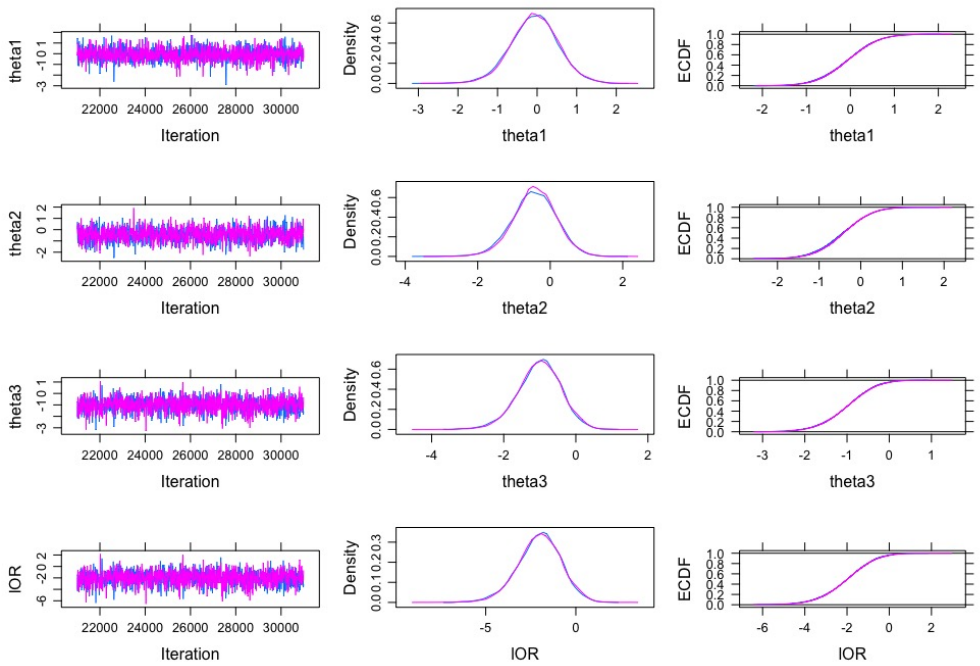


Figure 3.3: Trace plots (on the left), density plots (on the middle), and ECDF plots (on the right) of the posterior samples of the parameters $\theta_1, \theta_2, \theta_3$ and the log-odds ratio in the fourth row, using a single normal prior (precise).

Figure 3.3 shows diagnostic plots of simulated posterior samples of parameters $\boldsymbol{\theta}$ using single normal prior where $\boldsymbol{\theta} \sim N(0, 400)$. Then imprecise case where a set, $\mathcal{M}_0$, of four normal priors are given to the parameters is considered as follows:

$$\mathcal{M}_0 = \{Normal(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) : \boldsymbol{\mu} \in (-200, -2, 2, 200), \boldsymbol{\sigma^2} \in (400, 4, 4, 400)\}.$$

The four normal priors are $N(-200, 400)$, $N(-2, 4)$, $N(2, 4)$, and $N(200, 400)$. Table 3.3 presents the upper and lower posterior expectations of log-odds ratio and the degree of imprecision.

Table 3.3: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio.

Prior	sample size	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\log \Psi}$
Normal	17	-0.573	-1.312	0.739

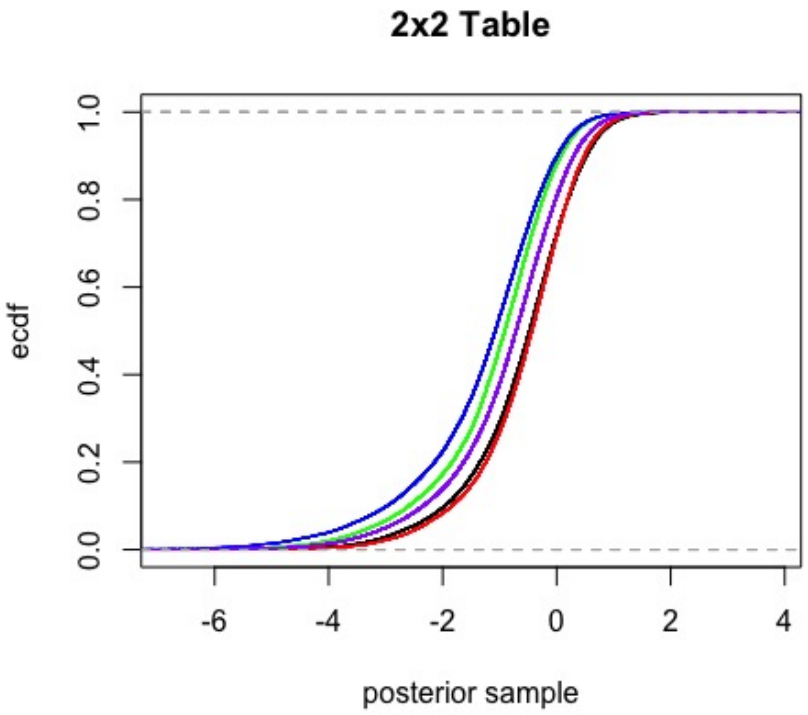


Figure 3.4: Plots of ECDFs of posterior sample of log-odds ratio in the two cases of using precise (purple ecdf) and imprecise (set of four normal priors: blue, green, red, and black).

Figure 3.4 shows the empirical CDF plots of the posterior sample for the two case, precise and imprecise, when empirical CDF curve of a single normal prior lies between the empirical CDF's of the set of normal priors with different means and variances.

4. Bayesian Imprecise Inference for Log-rank Test in Stratified 2×2 Tables

4.1 Motivation

In clinical studies one is concerned not only with estimating survival or hazard functions, but, more often, with the comparison of the life experience of two or more groups of patients who receive different treatments (test and control treatments). In these kinds of studies, it is difficult to have a priori knowledge to make trustworthy hypotheses on the underlying survival functions; thus, the non-parametric approach is usually adopted to compare survival curves. There are a number of frequentist methods that can be used for hypothesis testing. Two of the various most common non-parametric tests are the *Generalized Wilcoxon* test (Gehan, 1965) and the *log-rank* test (Mantel, 1966).

In this chapter survival data with right-censored observations is considered in the case that we have two groups (test and control) with two outcomes (alive and dead) which are represented in a sequence of 2×2 tables, one at each death. The main purpose in this chapter is getting the upper and lower posterior expectations of the log-odds ratio in order to compute the degree of imprecision for each set of tables and how it gets decreased as number of tables (deaths) increased. To implement this, the log-rank test is constructed; In fact, under the null hypothesis of independence, the central hypergeometric distribution is the base of constructing the log-rank test. Also, re-parametrization of the odds ratio is assumed based

on the feature that non-central hypergeometric distribution is a member of the exponential family.

4.2 Non-central Hypergeometric Model

In the Subsection 2.1.1, four sampling schemes for contingency tables are discussed. Consider Table 2.1 at a time of death with group 0 being the group that receives the control treatment and group 1 that receives the test treatment:

Table 4.1: A 2×2 contingency table.

	Event		
Group	0 (Alive)	1 (Dead)	Total
0 (control)	y_{00}	y_{01}	$n_{0.}$
1 (test)	y_{10}	y_{11}	$n_{1.}$
Total	$n_{.0}$	$n_{.1}$	n

Conditioning on the column totals, the sampling model is product binomial as follows:

$$y_{11} \sim B(n_{1.}, p_{11}) \qquad y_{01} \sim B(n_{.1}, p_{01}).$$

Then conditioning on having column, row (and hence grand) totals are fixed, the conditional distribution of y_{11} conditionally on $y_{01} + y_{11} = n_{.1}$ is non-central hypergeometric distribution with parameter Ψ. Consider that Ψ_0 and Ψ_1 are the odds in group 0 and 1 respectively,

$$\begin{aligned} f(y_{11}|\Psi) &= Pr(y_{11}|y_{01} + y_{11} = n_{.1}, n_{0.}, n_{1.}, \Psi_0, \Psi_1) \\ &= \frac{Pr(y_{11}|y_{01} + y_{11} = n_{.1}|n_{0.}, n_{1.}, \Psi_0, \Psi_1)}{Pr(y_{01} + y_{11} = n_{.1}|n_{0.}, n_{1.}, \Psi_0, \Psi_1)} \\ &= \frac{\binom{n_{1.}}{y_{11}} p_{11}^{y_{11}} (1-p_{11})^{(n_{1.}-y_{11})} \binom{n_{0.}}{n_{1.}-y_{11}} p_{01}^{n_{1.}-y_{11}} (1-p_{01})^{n_{0.}-(n_{1.}-y_{11})}}{Pr(y_{01} + y_{11} = n_{.1}|n_{0.}, n_{1.}, \Psi_0, \Psi_1)} \end{aligned}$$

$$= \frac{\binom{n_{1.}}{y_{11}}(\frac{p_{11}}{1-p_{11}})^{y_{11}}(1-p_{11})^{n_{1.}}\binom{n_{0.}}{n_{1.}-y_{11}}(\frac{p_{01}}{1-p_{01}})^{y_{11}}(1-p_{01})^{n_{0.}}}{Pr(y_{01}+y_{11}=n_{.1}|n_{0.},n_{1.},\Psi_0,\Psi_1)}$$

$$= \frac{\binom{n_{1.}}{y_{11}}\binom{n_{0.}}{n_{1.}-y_{11}}\Psi_1^{y_{11}}\Psi_0^{(n_{1.}-y_{11})}(1-p_{11})^{n_{1.}}(1-p_{01})^{n_{0.}}}{Pr(y_{01}+y_{11}=n_{.1}|n_{0.},n_{1.},\Psi_0,\Psi_1)}$$

$$= \frac{\binom{n_{1.}}{y_{11}}\binom{n_{0.}}{n_{1.}-y_{11}}\frac{\Psi_1}{\Psi_0}^{y_{11}}\Psi_0^{n_{1.}}(1-p_{11})^{n_{1.}}(1-p_{01})^{n_{0.}}}{Pr(y_{01}+y_{11}=n_{.1}|n_{0.},n_{1.},\Psi_0,\Psi_1)}$$

$$= \frac{\binom{n_{1.}}{y_{11}}\binom{n_{0.}}{n_{1.}-y_{11}}\Psi^{y_{11}}\Psi_0^{n_{1.}}(1-p_{11})^{n_{1.}}(1-p_{01})^{n_{0.}}}{\Psi_0^{n_{1.}}(1-p_{11})^{n_{1.}}(1-p_{01})^{n_{0.}}\sum_{j=max(0,n_{.1}-n_{0.})}^{min(n_{1.},n_{.1})}\binom{n_{1.}}{j}\binom{n_{0.}}{n_{1.}-j}\Psi^j}$$

$$= \frac{\binom{n_{1.}}{y_{11}}\binom{n_{0.}}{n_{1.}-y_{11}}\Psi^{y_{11}}}{\sum_{j=max(0,n_{.1}-n_{0.})}^{min(n_{1.},n_{.1})}\binom{n_{1.}}{j}\binom{n_{0.}}{n_{1.}-j}\Psi^j} \tag{4.1}$$

Now, let consider the case where we have only one death at each table and we are assuming no ties, that is $n_{.1}=1$, and let $y_{11}=y$ be the indicator of the event that the death is in the test group. Then Eq. (4.1) can be written as

$$f(y|\Psi) = \frac{\binom{n_{1.}}{y_{11}}\binom{n_{0.}}{n_{1.}-y_{11}}\Psi^{y}}{\sum_{j=max(0,n_{.1}-n_{0.})}^{min(n_{1.},n_{.1})}\binom{n_{1.}}{j}\binom{n_{0.}}{n_{1.}-j}\Psi^j} \tag{4.2}$$

$$= \frac{\binom{n_{1.}}{y}\binom{n_{0.}}{n_{1.}-y}\Psi^{y}}{\binom{n_{1.}}{0}\binom{n_{0.}}{n_{.1}}\Psi^0+\binom{n_{1.}}{n_{.1}}\binom{n_{0.}}{n_{.1}-n_{.1}}\Psi^{n_{1.}}}$$

$$= \frac{n_{1.}\Psi^y}{n_{0.} + \binom{n_{1.}}{1}\binom{n_{0.}}{0}\Psi}$$

$$= \frac{n_{1.}\Psi^y}{n_{0.} + n_{1.}\Psi} = \frac{(r\Psi)^y}{1 + r\Psi}, \tag{4.3}$$

where $r = n_{1.}/n_{0.}$ is the balancing rate for the table. Therefore, Eq. (4.2) is the (partial) likelihood based on the non-central hypergeometric distribution obtained by conditioning on both margins. Let $\theta = \log \Psi$ (the log-odds ratio), then Eq. (4.2) can be written as:

$$f(y|\theta) = \frac{(re^{\theta})^y}{1 + re^{\theta}}, \tag{4.4}$$

where

$$f(y|\theta) = \begin{cases} \frac{re^{\theta}}{1+re^{\theta}} & \text{if } y = 1 \text{ (the death happens in group 1 (test));} \\ \frac{1}{1+re^{\theta}} & \text{if } y = 0 \text{ (the death happens in group 0 (control)).} \end{cases}$$

The likelihood function in Eq. (4.4) can be written in terms of a transformed parameter $p = (1 + e^{-\theta})^{-1}$ as follows:

$$\frac{(re^{\theta})^y}{1 + re^{\theta}} = \frac{(rp)^y(1-p)^{1-y}}{1 + (r-1)p}. \tag{4.5}$$

Equation (4.5) is a binomial likelihood if $r = 1$ ($n_{0.} = n_{1.}$).

Now, consider we have a sequence of 2×2 tables at each time of death. To implement the Bayesian imprecise approach, a set of priors on the log-odds ratio, θ, is considered. By the proportional hazards assumption, the parameter θ is the same for all tables. At each death time (at each table), the (partial) likelihoods in Eq. (4.4) and (4.5) that follow whether the death happened in group 0 or group 1 are used.

4.2.1 Choices of Imprecise Priors

A discussion about the choice of prior distributions has taken place in Chapter 2 of this dissertation. Recalling the literature in Section 2.3.3, Walley (1996) defined the imprecise Dirichlet model as the set of all Dirichlet (s, t) distributions and Walley (1996) and Walley

et al. (1996) used the imprecise beta(s) model (a special case of the imprecise Dirichlet(s) model with $k = 2$ categories), to analyze data in the form of a contingency table. Bickis (2009) introduces the imprecise logit-normal model as a family of prior distributions for a binomial success probability. Different models and different choice of priors have been proposed by PhD theses work by Bataineh (2012) and Lee (2014).

In this dissertation, two choice of priors are used. The set of priors will be updated to a set of posteriors by using likelihoods in Eq. (4.4) and (4.5), that is, a sequence of updates to the posterior for each observed death. Therefore, these likelihoods can be rewritten as,

$$L(\theta|y) = \prod_{i=1}^{K} \frac{(r_i e^{\theta})^{y_i}}{1 + r_i e^{\theta}}, \tag{4.6}$$

and

$$L(p|y) = \prod_{i=1}^{K} \frac{(r_i p)^{y_i}(1-p)^{1-y_i}}{1 + (r_i - 1)p}. \tag{4.7}$$

First, a family of normal priors with different means and variances has been assigned directly to the log-odds ratio θ. Second, a family of beta priors with different shape parameters is given to the parameter p and then the parameter $\theta = \log(\frac{p}{1-p})$ is estimated.

Imprecise Normal Prior

A family of four normal priors has been assigned to the log-odds ratio, θ, and this set can be defined as follows:

$$\mathcal{M}_0 = \{Normal(\boldsymbol{\mu}, \boldsymbol{\sigma^2}) : \boldsymbol{\mu} \in (-200, -2, 2, 200), \boldsymbol{\sigma^2} \in (400, 4, 4, 400)\}. \tag{4.8}$$

The four normal priors are $N(-200, 400)$, $N(-2, 4)$, $N(2, 4)$, and $N(200, 400)$. At each table (observed death time), each prior in the set will be updated to a posterior by using the likelihood in Eq. (4.6), that is, a sequence of updates to the posterior for each observed death.

Imprecise Beta Prior

The second choice of an imprecise prior in this simulation study is a family of four beta priors is given to the parameter p, and this set can be defined as follows:

$$\mathcal{M}_0 = \{Beta(\boldsymbol{a}, \boldsymbol{b}) : \boldsymbol{a} \in (0.1, 0.3, 1.2, 1.6), \boldsymbol{b} \in (1.9, 1.7, 0.8, 0.4)\}. \tag{4.9}$$

The four beta priors are $Beta(0.1, 1.9)$, $Beta(0.3, 0.7)$, $Beta(1.2, 0.8)$, and $Beta(1.6, 0.4)$. At each table (observed death time), each prior in the set will be updated to a posterior by using the likelihood in Eq. (4.7), that is, a sequence of updates to the posterior for each observed death and the log-odds ratio, θ, is estimated at each update and upper and lower posterior expectations of θ are calculated.

4.3 An Example with Real Data

Considering the model presented in Section 4.2 and the two choices of priors in Subsection 4.2.1, an application of real data for this work is presented. The data set here is Ovarian Cancer Survival Data that is included with `R` in package `survival` as a data frame. These data present survival times in days for two groups of patients. The two treatments are cyclophosphamids alone (CTX) and cyclophosphamids plus adriamycin (CTX+AD). More information about this data set can be found in Edmunson et al. (1979).

An imprecise Bayesian approach is applied on these data by updating a set of priors to a set of posteriors of the parameter θ, log-odds ratio, for each observed death using the likelihood function in Eq. (4.4). This is done by running MCMC using `runjags` in JAGS from within `R`. Upper posterior expectation $\overline{E}(\theta|y)$ and lower posterior expectation $\underline{E}(\theta|y)$ are estimated in order to calculate the degree of imprecision $\Delta_{\theta|y}$ of the log-odds ratio. This is done numerically by finding the maximum and minimum over the set of simulated posterior means.

In this data set, there are a total of 26 observations, 13 in each group. The number of observed death times is 12 (12 table, one at each observed death time), and number of censored times is 14. Figure 4.1 shows the estimated Kaplan-Meier survival functions of two groups (CTX and CTX+AD).

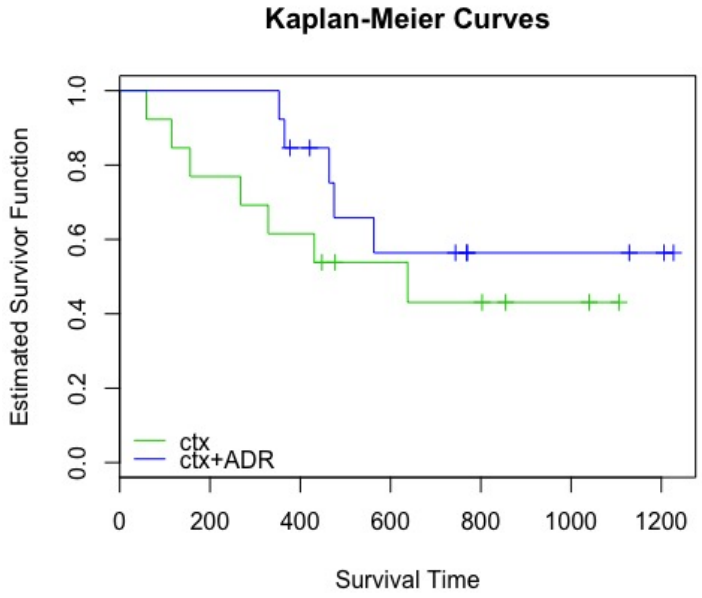


Figure 4.1: The estimated Kaplan-Meier survival functions of two groups (CTX and CTX+AD). "+" represents censored times.

Table 4.2 presents the estimates of upper and lower posterior expectations and the degree of imprecision of the log-odds ratio. The degree of imprecision $\Delta_{\log \Psi}$ is 0.411 in the case of choosing normal prior and 0.475 in beta case, which there is no big difference between the two cases. The small difference between the estimates is 0.064 which can be explained by looking to Figures 4.2 and 4.3 where the posterior ecdfs differ for Table 1, but they are quite similar for Table 12, and how the degree of imprecision becomes less as we have more tables. Figures 4.2 and 4.3 show the plots of empirical CDFs of the posterior samples for number of tables. For example, the top-left graph is ecdfs of the posterior samples of the first table while the top-right and bottom-left are tables 3 and 6, and the bottom-right one represents ecdfs of posterior sample of the last table which is table 12 (i.e. $K = 12$, 12 observed death times). The word "Table" on the top of each graph means the 2×2 table that is built at that observed death time; for example, "Table 12" means the table at the observed death time number 12. In Figures 4.2, blue, green, red, and black curves represent the posterior samples when priors are $N(-200, 400)$, $N(-2, 4)$, $N(2, 4)$, and $N(200, 400)$ respectively. In Figures 4.3, blue, green, red, and black curves represent the posterior samples when priors

are $Beta(0.1, 1.9)$, $Beta(0.3, 0.7)$, $Beta(1.2, 0.8)$, and $Beta(1.6, 0.4)$ respectively. The same convention follows for all figure of the ecdfs in rest of this dissertation. The next section will talk about simulations with different factors are considered. One of these factors is the sample size where one can see how the sample size affects the degree of imprecision.

Table 4.2: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio.

Prior	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$	Imprecise credible interval
Normal	-0.401	-0.811	0.411	(-2.071, 0.720)
Beta	-0.348	-0.823	0.475	(-2.004, 0.735)

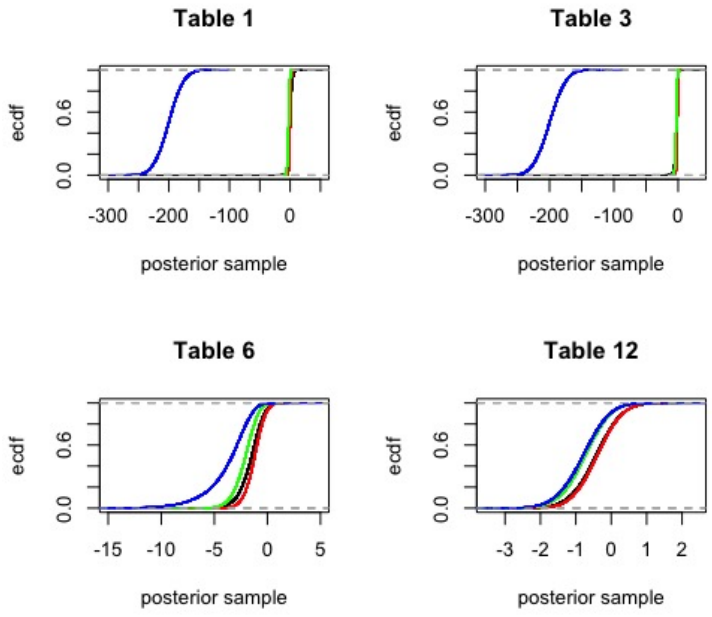


Figure 4.2: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors using Ovarian Cancer survival data. The X-axis represents the log-odds ratio. Blue, green, red, and black curves represent the posterior samples when priors are $N(-200, 400)$, $N(-2, 4)$, $N(2, 4)$, and $N(200, 400)$ respectively.

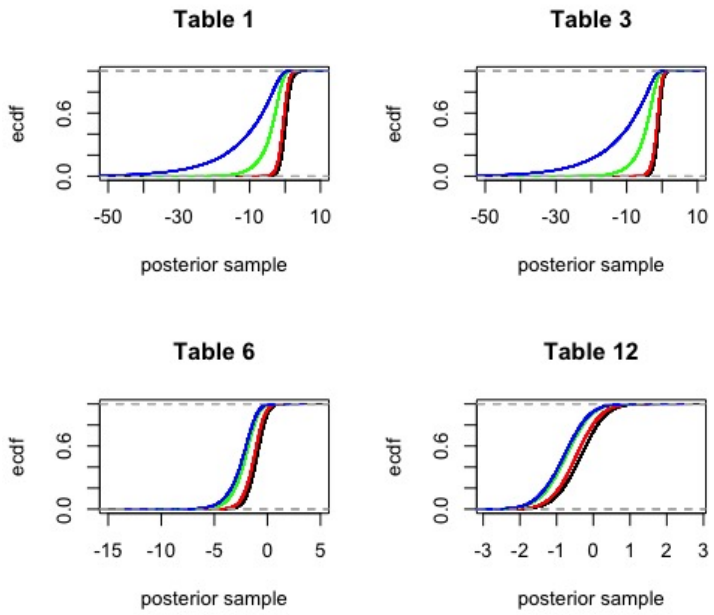


Figure 4.3: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors using Ovarian Cancer survival data. Blue, green, red, and black curves represent the posterior samples when priors are $Beta(0.1, 1.9)$, $Beta(0.3, 0.7)$, $Beta(1.2, 0.8)$, and $Beta(1.6, 0.4)$ respectively.

4.4 Simulation Study and Results

In this section, several sets of survival data with right-censored observations is generated for two groups. The survival times of Cox proportional hazard model that discussed in Section 2.2.2, in general, can be generated using some distributions like exponential, Weibull, or Gompertz. The translation of the regression effects from hazard to survival time is easy if the baseline hazard function is constant, i.e. the survival times are exponentially distributed. For this reason, the survival times in this work are generated using the exponential distribution, $Exponential(0.5)$. Similarly, censored times are also generated by using the exponential distribution with parameter λ_c. The binary covariate, X_i, is generated using the Bernoulli distribution.

The generated censored survival data are converted to K 2×2 contingency tables, where

K is the total number of observed deaths (tables). Actually, K here is a random variable, so it could be different for each simulation. An imprecise Bayesian approach is applied on simulated data by updating a set of priors to a set of posteriors of the parameter θ, log-odds ratio, for each observed death using the likelihood function in Eq. (4.4). This is done by running MCMC using `runjags` in JAGS from within `R`.

The main goal in this simulation is computing the upper posterior expectation $\overline{E}(\theta|y)$ and lower posterior expectation $\underline{E}(\theta|y)$ in order to calculate the degree of imprecision $\Delta_{\theta|y}$ of the log-odds ratio and compare the cases where $\Delta_{\theta|y}$ is reduced. Four factors in this simulation study are considered. Table 4.3 describe the factors where n is the sample size with 3 different number of subjects, 40, 100, 200. Three different values of the parameter λ_c are 0.001, 0.1, and 0.5, that is, 0.1 %, 10 %, and 50 % of the total number of subjects is censored. True values of the model parameter θ are considered as 0, -0.6, -1.2 , where the balancing rate $r = n_{1.}/n_{0.}$, the allocation ratio, is imposed for the purpose of looking at different scenarios of allocating the number of subjects in each group. The three levels of r are $r = 1$ when $n_{0.} = n_{1.}$, $r < 1$ when $n_{1.} < n_{0.}$, and $r > 1$ when $n_{1.} > n_{0.}$. To get these three cases of the balancing rate, r, the binary covariate, X_i that represents the two group who received two treatment are generated by using Bernoulli distribution as mentioned above, that is, when $X_{Control} \sim Bernoulli(0.5)$ and $X_{Test} \sim Bernoulli(0.5)$, $r = 1$, and when $X_{Control} \sim Bernoulli(0.3)$ and $X_{Test} \sim Bernoulli(0.1)$, $r < 1$, and when $X_{Control} \sim Bernoulli(0.1)$ and $X_{Test} \sim Bernoulli(0.3)$, $r > 1$.

Table 4.3: A description of the four factors, each factor with 3 levels.

Factors	Description of Factors	Levels
n	Sample size	40, 100, 200
λ_c	Censoring rate	0.001, 0.1, 0.5
θ	True value	0, -0.6, -1.2
r	Balancing rate	$r = 1$, $r < 1$, $r > 1$

The purpose of considering the four factors is to examine how the degree of imprecision

will be affected at different levels of these factors and to investigate which of these are more significant. To do that, a fractional factorial design on the the degree of imprecision for each level of the four factors is performed as in Table 4.3. In a factorial design, as the number of factors increases, the number of runs needed for a complete replicate of the design speedily enlarges the resources of most experimenters. For example, in this work, a complete replicate of the design 3^4 requires 81 runs. For this reason, a fractional factorial design $3^{4-1} = 27$ runs (combinations) are constructed in Table 4.4 and this design is a resolution IV design, 3^{4-1}_{IV}, which means no main effects are aliased with any other main effects or with any two-factor interaction, but two-factor interactions are aliased with each other (Montgomery, 2009, Chapter 8).

For each one of these 27 combination, there are K tables and MCMC has been run and upper and lower posterior expectations and the degree of imprecision of the log-odds ratio are calculated. This process has been done for two choices of imprecise priors, 27 runs for each case which means the total number of runs is 54 runs. A discussion of complete results of the simulation study beside ANOVA tables and Boxplots are provided in the following sections and Appendix A.

4.4.1 The Results in Imprecise Normal Case

Table 4.5 presents the results of running the MCMC for the 27 combinations by considering the set of normal prior defined above. The general picture emerging from results in the table is that the lowest value of $\Delta_{\theta|y}$, which is 0.026, is observed at combination number 20 when the sample size is 200, the censoring rate is 0.1 %, the true value of θ is -0.6, and $r = 1$. In contrast, The largest value of $\Delta_{\theta|y}$, which is 0.562 appears at combination number 9 where $n = 40$, $\lambda_c = 0.5$, $\theta = -1.2$, and $r < 1$. The empirical CDF plots of the posterior samples of four chosen tables each with four priors are shown in Figures 4.4 and 4.5; for instance, in Figures 4.4, the top-left graph is ecdfs of the posterior samples of the first table while the top-right and bottom-left are tables 10 and 14, and the bottom-right one represents ecdfs of posterior sample of the last table which is table 17 (i.e. $K = 17$ at combination number

9). The same way follows for all figure of the ecdfs in rest of this dissertation. The degree of imprecision in these figures is represented by the gap between the ecdfs curves and it can be seen that the ecdf plots confirm the values of $\Delta_{\theta|y}$'s. Imprecise highest posterior density credible intervals of θ for the 27 combinations considering the set of normal priors are shown in Table 4.6.

Table 4.4: The 3_{IV}^{4-1} Design.

	Factors							
Run	A	B	C	D	Sample size	Censoring rate	True parameter	Balancing rate
1	0	0	0	0	40	0.001	0	$r = 1$
2	0	0	1	1	40	0.001	-0.6	$r < 1$
3	0	0	2	2	40	0.001	-1.2	$r > 1$
4	0	1	0	1	40	0.1	0	$r < 1$
5	0	1	1	2	40	0.1	-0.6	$r > 1$
6	0	1	2	0	40	0.1	-1.2	$r = 1$
7	0	2	0	2	40	0.5	0	$r > 1$
8	0	2	1	0	40	0.5	-0.6	$r = 1$
9	0	2	2	1	40	0.5	-1.2	$r < 1$
10	1	0	0	1	100	0.001	0	$r < 1$
11	1	0	1	2	100	0.001	-0.6	$r > 1$
12	1	0	2	0	100	0.001	-1.2	$r = 1$
13	1	1	0	2	100	0.1	0	$r > 1$
14	1	1	1	0	100	0.1	-0.6	$r = 1$
15	1	1	2	1	100	0.1	-1.2	$r < 1$
16	1	2	0	0	100	0.5	0	$r = 1$
17	1	2	1	1	100	0.5	-0.6	$r < 1$
18	1	2	2	2	100	0.5	-1.2	$r > 1$
19	2	0	0	2	200	0.001	0	$r < 1$
20	2	0	1	0	200	0.001	-0.6	$r = 1$
21	2	0	2	1	200	0.001	-1.2	$r < 1$
22	2	1	0	0	200	0.1	0	$r = 1$
23	2	1	1	1	200	0.1	-0.6	$r < 1$
24	2	1	2	2	200	0.1	-1.2	$r > 1$
25	2	2	0	1	200	0.5	0	$r < 1$
26	2	2	1	2	200	0.5	-0.6	$r > 1$
27	2	2	2	0	200	0.5	-1.2	$r = 1$

Table 4.5: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio in the case of imprecise normal priors.

Combination	Sample size	Censoring rate	True parameter	Balancing rate (r)	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
1	40	0.001	0	$r = 1$	-0.012	-0.124	0.111
2			-0.6	$r < 1$	-0.306	-0.434	0.128
3			-1.2	$r > 1$	-0.650	-0.831	0.229
4		0.1	0	$r < 1$	0.072	-0.079	0.219
5			-0.6	$r > 1$	-0.562	-0.763	0.200
6			-1.2	$r = 1$	-1.202	-1.455	0.253
7		0.5	0	$r > 1$	0.172	-0.185	0.357
8			-0.6	$r = 1$	-0.551	-0.827	0.323
9			-1.2	$r < 1$	-1.158	-1.754	0.562
10	100	0.001	0	$r < 1$	0.184	0.121	0.062
11			-0.6	$r > 1$	-0.698	-0.769	0.070
12			-1.2	$r = 1$	-1.336	-1.419	0.082
13		0.1	0	$r > 1$	-0.284	-0.369	0.085
14			-0.6	$r = 1$	-0.608	-0.670	0.062
15			-1.2	$r < 1$	-1.203	-1.364	0.161
16		0.5	0	$r = 1$	0.018	-0.058	0.076
17			-0.6	$r < 1$	-0.841	-1.099	0.257
18			-1.2	$r > 1$	-1.123	-1.328	0.204
19	200	0.001	0	$r < 1$	-0.030	-0.057	0.027
20			-0.6	$r = 1$	-0.665	-0.692	0.026
21			-1.2	$r < 1$	-1.212	-1.260	0.047
22		0.1	0	$r = 1$	-0.003	-0.031	0.028
23			-0.6	$r < 1$	-0.645	-0.693	0.047
24			-1.2	$r > 1$	-1.074	-1.128	0.053
25		0.5	0	$r < 1$	-0.160	-0.217	0.057
26			-0.6	$r > 1$	-0.354	-0.436	0.081
27			-1.2	$r = 1$	-1.074	-1.162	0.088

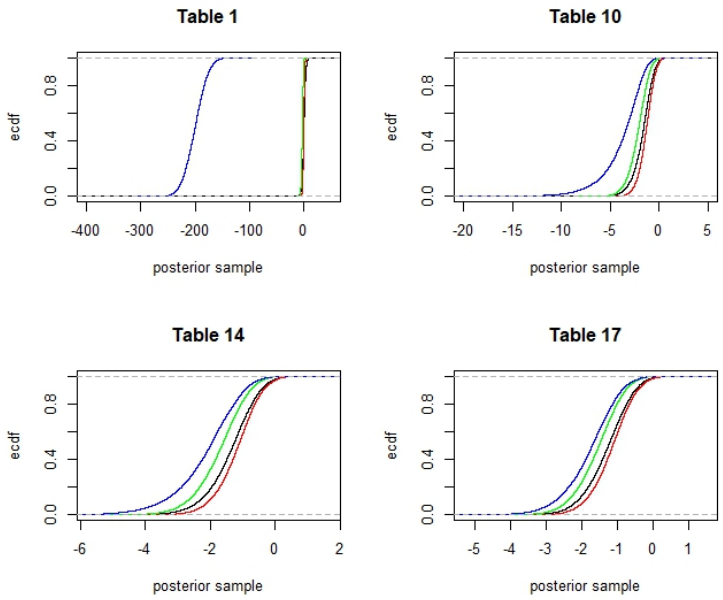


Figure 4.4: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 9.

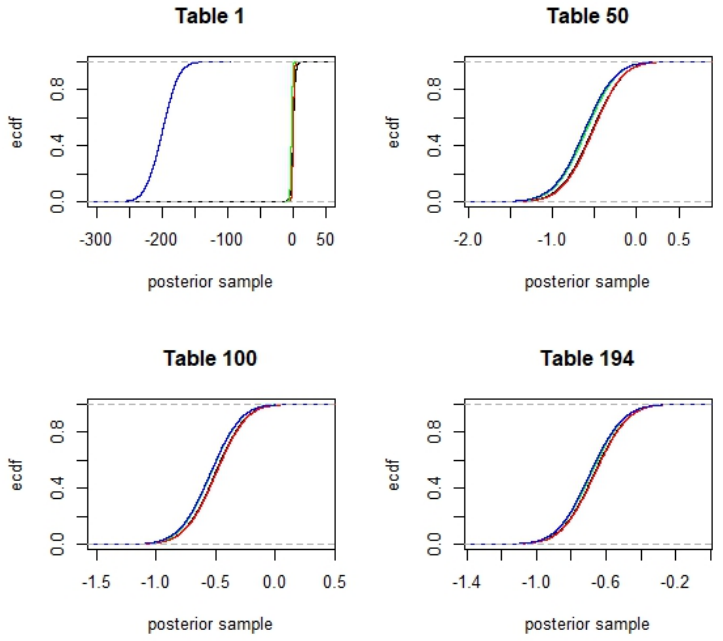


Figure 4.5: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 20.

Table 4.6: Imprecise credible intervals in the case of using imprecise normal priors.

Combination	Sample size	Censoring rate	True parameter	Balancing rate (r)	Imprecise credible interval
1	40	0.001	0	$r = 1$	(-0.348 , 0.208)
2			-0.6	$r < 1$	(-0.669 , -0.072)
3			-1.2	$r > 1$	(-2.106 , -1.129)
4		0.1	0	$r < 1$	(-0.339 , 0.332)
5			-0.6	$r > 1$	(-1.054 , -0.280)
6			-1.2	$r = 1$	(-1.752 , -0.916)
7		0.5	0	$r > 1$	(-0.581 , 0.578)
8			-0.6	$r = 1$	(-1.158 , -0.240)
9			-1.2	$r < 1$	(-2.200 , -0.756)
10	100	0.001	0	$r < 1$	(-0.039 , 0.344)
11			-0.6	$r > 1$	(-0.942 , -0.532)
12			-1.2	$r = 1$	(-1.580 , -1.173)
13		0.1	0	$r > 1$	(-0.556 , -0.098)
14			-0.6	$r = 1$	(-0.826 , -0.456)
15			-1.2	$r < 1$	(-1.601 , -0.976)
16		0.5	0	$r = 1$	(-0.241 , 0.205)
17			-0.6	$r < 1$	(-1.410 , -0.556)
18			-1.2	$r > 1$	(-1.603 , -0.857)
19	200	0.001	0	$r < 1$	(-0.167 , 0.080)
20			-0.6	$r = 1$	(-0.793 , -0.564)
21			-1.2	$r < 1$	(-1.391 , -1.084)
22		0.1	0	$r = 1$	(-0.138 , 0.104)
23			-0.6	$r < 1$	(-0.826 , -0.510)
24			-1.2	$r > 1$	(-1.266 , -0.938)
25		0.5	0	$r < 1$	(-0.377 , 0.003)
26			-0.6	$r > 1$	(-0.614 , -0.179)
27			-1.2	$r = 1$	(-0.895 , -1.337)

It can be inferred from results in Table 4.5 that the less censoring and large sample size is the more information and less imprecision we get and vice versa. To confirm that, ANOVA for different model of the factorial deign in Table 4.3 on the degree of imprecision $\Delta_{\theta|y}$ are considered to see how much these factors affect the degree of imprecision. Starting with fitting the linear regression model (say Model 1), ANOVA table is displayed in Table 4.7.

Next, ANOVA of a model includes only the main effects (say Model 2) is summerized in Table 4.8 which is followed by table of the results of the model (say Model 3) that includes main effects plus some of two-factor interactions (Table 4.9). Model assumptions are checked for the three models by looking to the graphical analysis of residuals (residuals plots) where a heteroscedascity is indicated as in Figure 4.6.

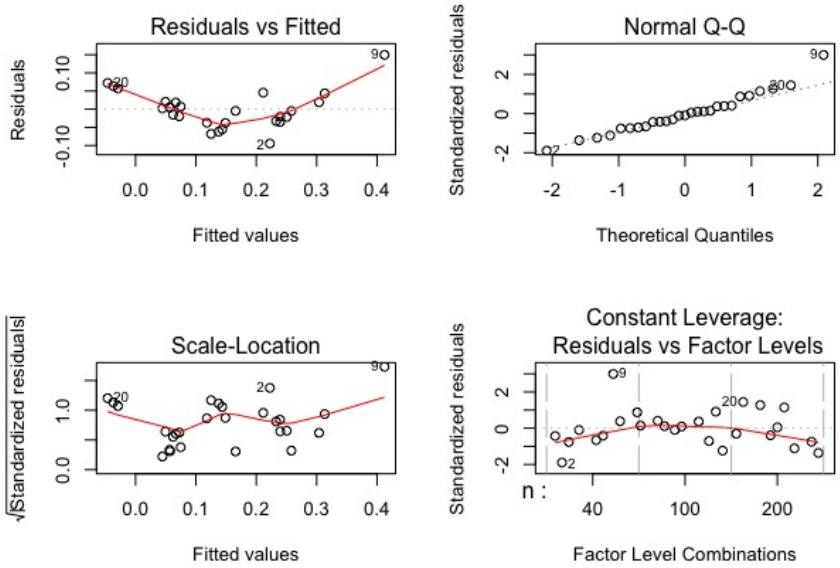


Figure 4.6: The residuals plot of Model 2.

Therefore, log transformation is applied for the three models on the response variable (degree of imprecision) and the residuals plots of Model 2 are displayed in the following figure which indicates that the spread of residuals around zero is fairly homogenous (Figure 4.7).

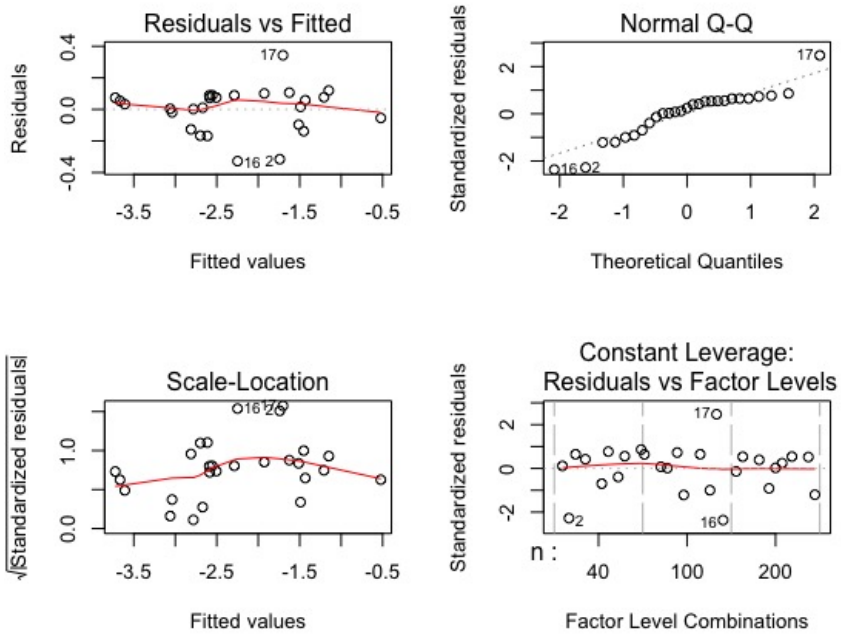


Figure 4.7: The residuals plot of Model 2 after applying log transformation.

Table 4.7: ANOVA table for factorial design in Table 4.3 in the case of using imprecise normal prior.

	Df	Sum of Square	Mean Square	F value	P-value
n	1	11.755	11.755	166.704	7.07e-10
λ_c	1	3.468	3.468	49.185	2.93e-06
θ	1	1.309	1.309	18.568	0.0005
r	1	0.301	0.301	4.265	0.055
$n \times \lambda_c$	1	0.008	0.008	0.110	0.744
$n \times \theta$	1	0.004	0.004	0.063	0.805
$n \times r$	1	0.001	0.001	0.009	0.925
$\lambda_c \times \theta$	1	0.014	0.014	0.203	0.658
$\lambda_c \times r$	1	0.014	0.014	0.195	0.664
$\theta \times r$	1	0.072	0.072	1.027	0.325
Residuals	16	1.128	0.071		

AICc = 37.179

The three models that considered above are compared by looking to corrected version of their AIC (AICc) values (Burnham and Anderson, 2002). These values are provided under each table where Model 2 has the lowest AICc (AICc= 3.639) compared to Model 1 and 3.

The three models that considered above show that all the main effects are significant

Table 4.8: ANOVA table for the main effects of factorial design in Table4.2 in the case of using imprecise normal prior.

	Df	Sum of Square	Mean Square	F value	P-value
n	2	12.059	6.030	209.41	3.43e-13
λ_c	2	3.575	1.788	62.08	8.36e-09
θ	2	1.338	0.669	23.23	1.03e-05
r	2	0.585	0.292	10.15	0.00112
Residuals	18	0.518	0.029		

AICc = 3.639

Table 4.9: ANOVA table for factorial design in Table 4.3 in the case of using imprecise normal prior.

	Df	Sum of Square	Mean Square	F value	P-value
n	2	12.059	6.030	198.358	3.31e-06
λ_c	2	3.575	1.788	58.807	0.000114
θ	2	1.338	0.669	22.005	0.001727
r	2	0.585	0.292	9.615	0.013448
$n \times \lambda_c$	4	0.038	0.010	0.315	0.858051
$n \times \theta$	4	0.113	0.028	0.929	0.505625
$n \times r$	4	0.185	0.046	1.518	0.308097
Residuals	6	0.182	0.030		

AICc =238.692

at 5% and 10% confidence levels and none of the two-factor interactions. Graphically, by looking to Boxplots, Figure 4.8 shows the four boxplots of the four factors where it can be noticed that the degree of imprecision is decreased when we have more data ($n = 200$), less censoring ($\lambda_c = 0.001$), true value is -1.2, and when $r = 1$, the number of subjects in two groups are equal.

The boxplots of two-factor interactions give more support to the results in Table 4.5. For example, Figure 4.9 clearly indicates that the degree of imprecision in the case of considering a set of normal priors happens when $n = 200$ and $\lambda_c = 0.001$, more data and less censored observations. However, The degree of imprecision is increased when $n = 40$ and $\lambda_c = 0.5$, less data and more censored observations.

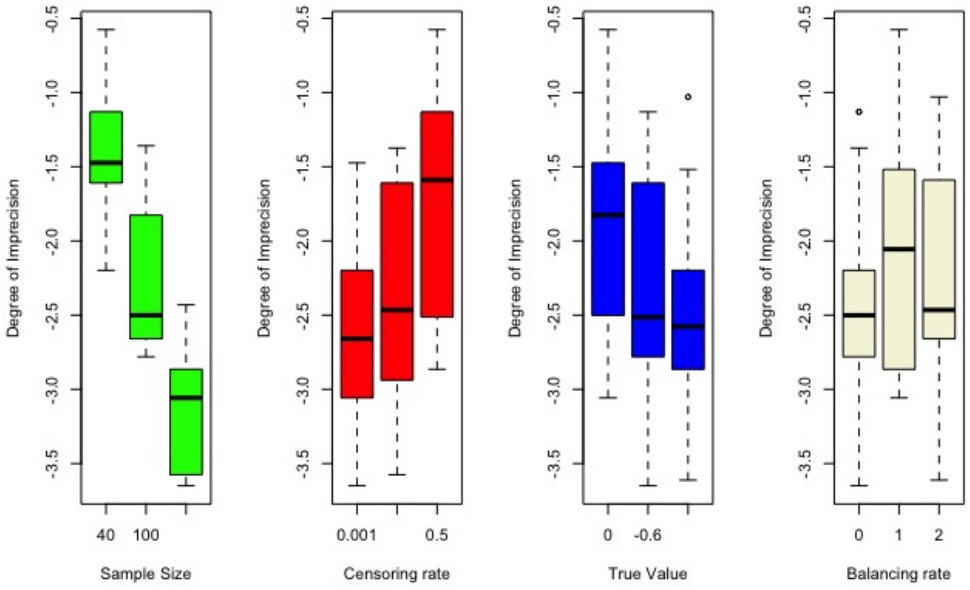


Figure 4.8: Boxplots of significant factors in Table 4.8.

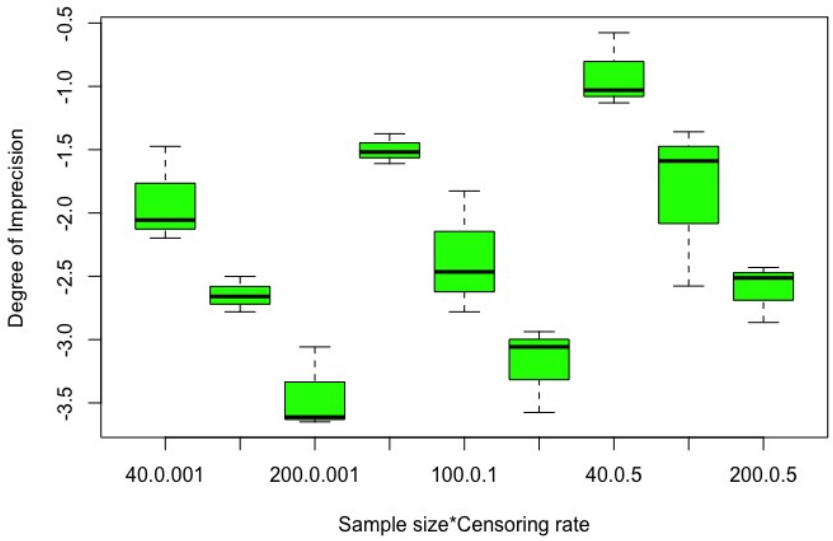


Figure 4.9: Boxplots of the interaction between the sample size factor and the censoring rate factor. The number, for example, 40.0.001 on horizontal axis means the interaction when $n = 40$ and $\lambda_c = 0.001$.

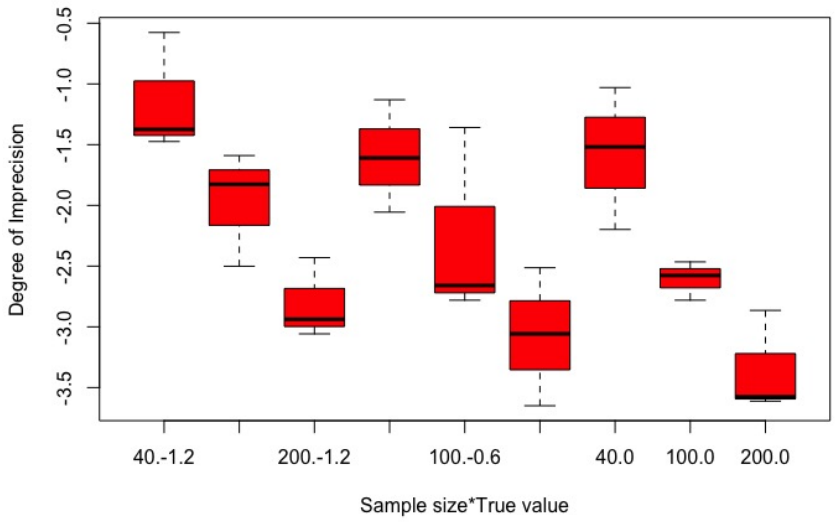


Figure 4.10: Boxplots of of the interaction between the sample size factor and the true value factor. The number, for example, 40.-1.2 on horizontal axis means the interaction when $n = 40$ and $\theta = -1.2$.

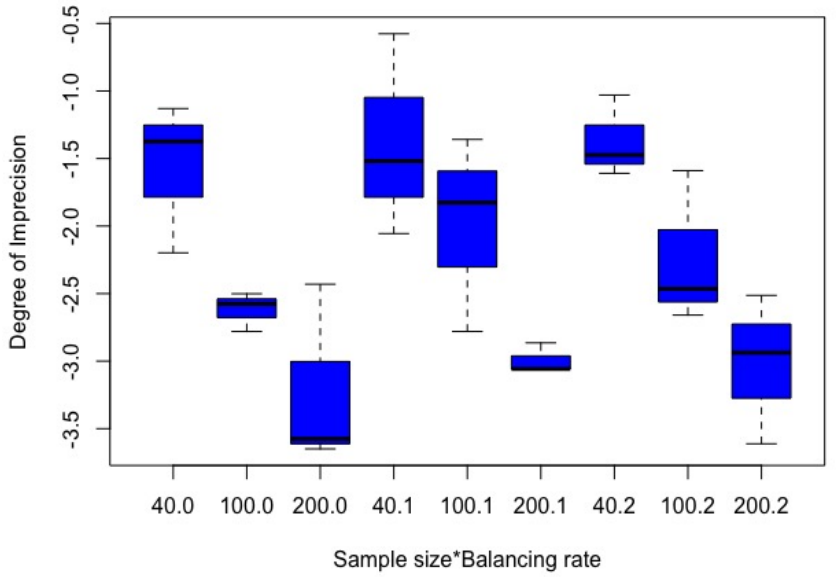


Figure 4.11: Boxplots of of the interaction between the sample size factor and the balancing rate factor. The number, for example, 40.0 on horizontal axis means the interaction when $n = 40$ and $r = 0$. The levels of balancing rate factors are $r = 0$ when $r = 1$, $r = 1$ when $r < 1$, and $r = 2$ when $r > 1$.

4.4.2 The Results in Imprecise Beta Case

The findings of the case of using beta priors support conclusions in the case of using normal priors. The results of running the MCMC for the 27 combinations by considering the set of beta prior defined above are reported in Table 4.10. The results in the table yielded some interesting finding that the reduced value of $\Delta_{\theta|y}$ is 0.031 and observed at combination number 20. However, The largest value of $\Delta_{\theta|y}$ is 0.562 appears at combination number 9. The empirical CDF plots of the posterior samples of four chosen tables each with four priors are shown in Figures 4.12 and 4.13. To clarify, for example, in Figures 4.12, the top-left graph is ecdfs of the posterior samples of the first table while the top-right and bottom-left are tables 50 and 100, and the bottom-right one represents ecdfs of posterior samples of the last table which is table 192 (i.e. $K = 192$ at combination number 20). The degree of imprecision in these figures is representing by the gap between the ecdfs curves and it can be seen that the ecdf plots confirm the values of $\Delta_{\theta|y}$'s. Imprecise highest posterior density credible intervals of θ for the 27 combinations considering the set of normal priors are shown in Table 4.11.

Table 4.10: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio in the case of imprecise Beta priors

Combination	Sample size	Censoring rate	True parameter	Balancing rate (r)	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
1	40	0.001	0	$r = 1$	-0.007	-0.162	0.155
2			-0.6	$r < 1$	-0.285	-0.460	0.174
3			-1.2	$r > 1$	-0.613	-0.843	0.229
4		0.1	0	$r < 1$	0.082	-0.136	0.219
5			-0.6	$r > 1$	-0.517	-0.782	0.265
6			-1.2	$r = 1$	-1.150	-1.416	0.266
7		0.5	0	$r > 1$	-0.605	-0.939	0.334
8			-0.6	$r = 1$	-0.501	-0.824	0.323
9			-1.2	$r < 1$	-1.057	-1.619	0.562
10	100	0.001	0	$r < 1$	0.215	0.141	0.074
11			-0.6	$r > 1$	-0.494	-0.575	0.080
12			-1.2	$r = 1$	-1.308	-1.404	0.096
13		0.1	0	$r > 1$	-0.026	-0.135	0.108
14			-0.6	$r = 1$	-0.616	-0.696	0.080
15			-1.2	$r < 1$	-1.100	-1.249	0.149
16		0.5	0	$r = 1$	0.346	0.204	0.142
17			-0.6	$r < 1$	-0.655	-0.876	0.220
18			-1.2	$r > 1$	-1.418	-1.66	0.245
19	200	0.001	0	$r < 1$	-0.028	-0.067	0.038
20			-0.6	$r = 1$	-0.595	-0.627	0.031
21			-1.2	$r < 1$	-1.250	-1.309	0.058
22		0.1	0	$r = 1$	-0.097	-0.136	0.038
23			-0.6	$r < 1$	-0.443	-0.496	0.053
24			-1.2	$r > 1$	-1.253	-1.309	0.055
25		0.5	0	$r < 1$	0.110	0.032	0.077
26			-0.6	$r > 1$	-0.603	-0.685	0.081
27			-1.2	$r = 1$	-1.120	-1.210	0.089

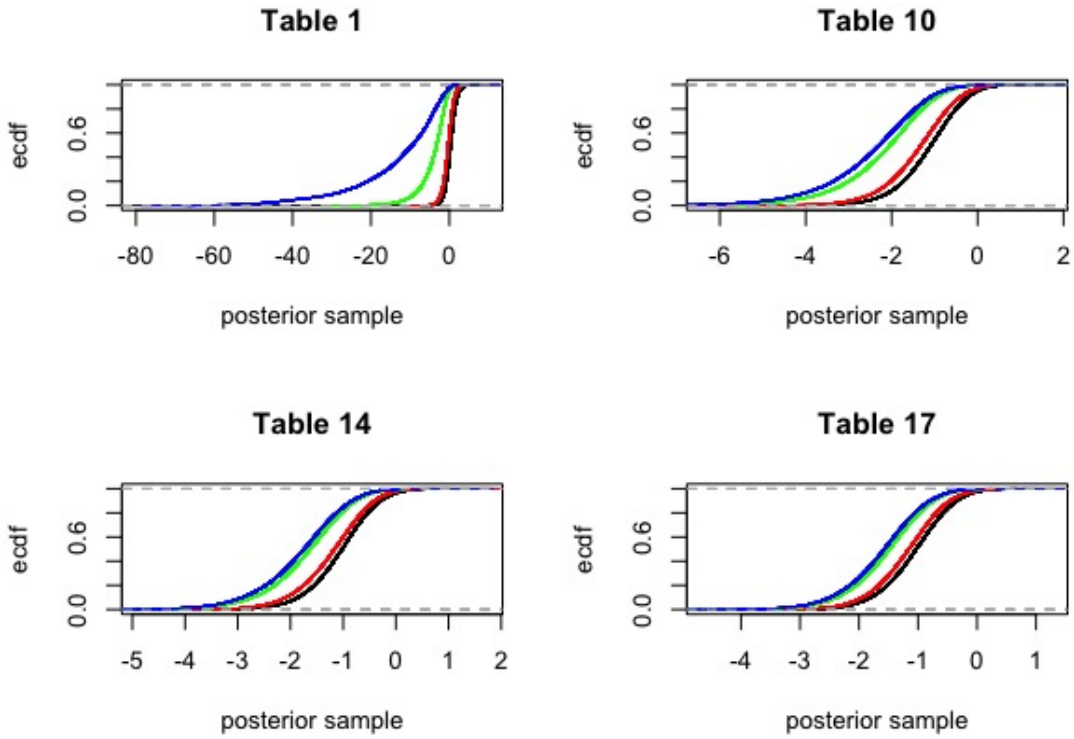


Figure 4.12: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 9.

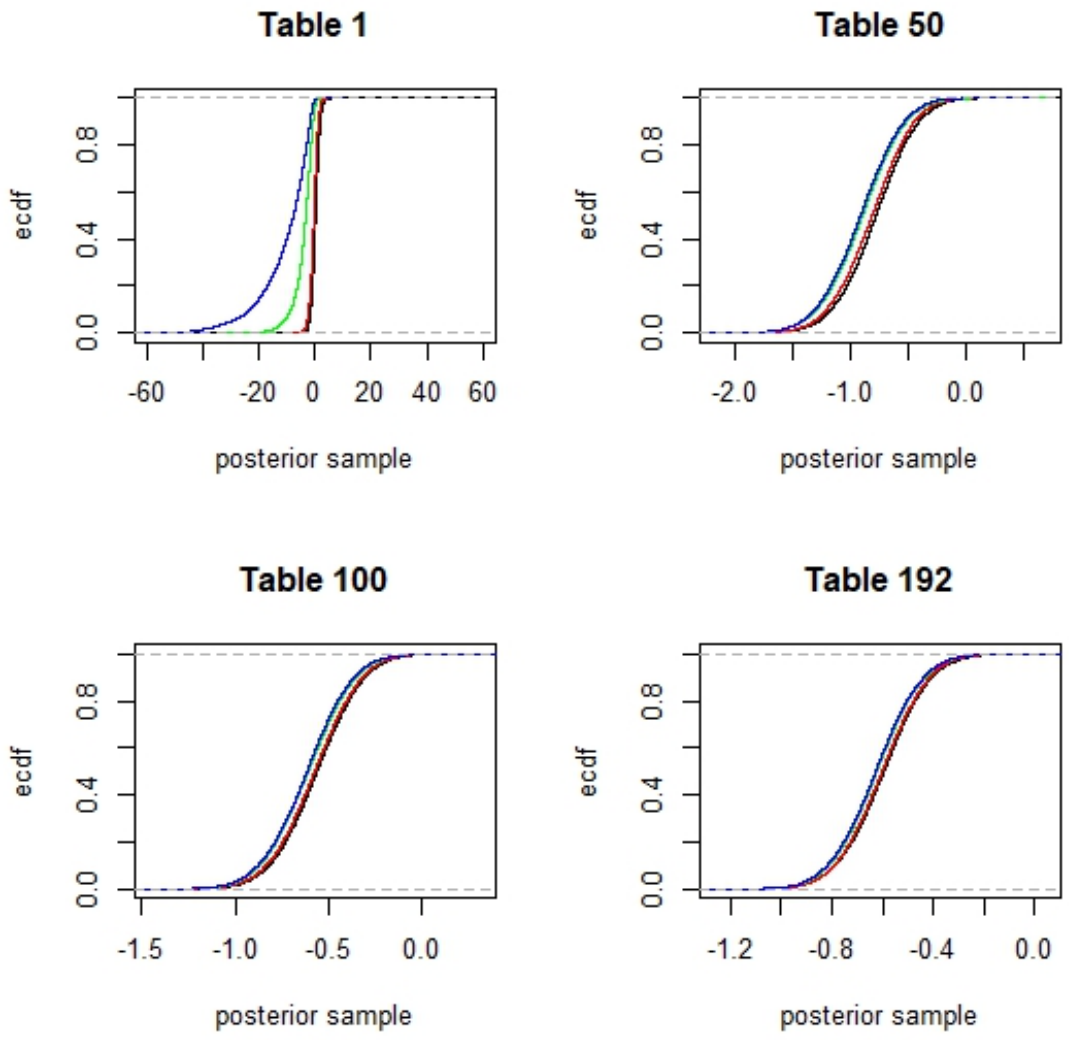


Figure 4.13: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 20.

Table 4.11: Imprecise credible intervals in the case of using imprecise beta priors.

Combination	Sample size	Censoring rate	True parameter	Balancing rate (r)	Imprecise credible interval
1	40	0.001	0	$r = 1$	(-0.377 , 0.207)
2			-0.6	$r < 1$	(-0.679 , -0.054)
3			-1.2	$r > 1$	(-1.570 , -0.749)
4		0.1	0	$r < 1$	(-0.384 , 0.333)
5			-0.6	$r > 1$	(-1.057 , -0.238)
6			-1.2	$r = 1$	(-1.696 , -0.866)
7		0.5	0	$r > 1$	(-0.471 , 0.560)
8			-0.6	$r = 1$	(-1.133 , -0.192)
9			-1.2	$r < 1$	(-2.041 , -0.666)
10	100	0.001	0	$r < 1$	(-0.008 , 0.368)
11			-0.6	$r > 1$	(-0.734 , -0.334)
12			-1.2	$r = 1$	(-1.573 , -1.144)
13		0.1	0	$r > 1$	(-0.313 , 0.155)
14			-0.6	$r = 1$	(-0.852 , -0.465)
15			-1.2	$r < 1$	(-1.456 , -0.893)
16		0.5	0	$r = 1$	(-0.004, 0.552)
17			-0.6	$r < 1$	(-1.131 , -0.394)
18			-1.2	$r > 1$	(-1.941 , -1.147)
19	200	0.001	0	$r < 1$	(-0.177, 0.081)
20			-0.6	$r = 1$	(-0.726 , -0.495)
21			-1.2	$r < 1$	(-1.441 , -1.117)
22		0.1	0	$r = 1$	(-0.245 , 0.013)
23			-0.6	$r < 1$	(-0.620 , -0.316)
24			-1.2	$r > 1$	(-1.438 , -1.127)
25		0.5	0	$r < 1$	(-0.120 , 0.264)
26			-0.6	$r > 1$	(-0.837 , -0.454)
27			-1.2	$r = 1$	(-1.373 , -0.953)

Similarly, ANOVA of the factorial design in Table 4.3 on the degree of imprecision $\Delta_{\theta|y}$ in the case of using imprecise beta priors is evaluated. The three models considered in the case of normal prior are also considered here. Also, log transformation on these model is applied to avoid heteroscedascity. The results of analysis of variance of the degree of imprecision of Model 1, Model 2, and Model 3 are demonstrated in Tables 4.12, 4.13, and 4.14, respectively.

The main effects of the factors are effective, and none of the two-factor interactions are. Besides that, three models also are compared with each other by using their AICc where Model 2 (with only main effects) is still preferred with AICc= - 30.793. Figure 4.14 is a graphic summary of the Boxplots of the significant factors and their levels. However, Figures 4.15, 4.16, and 4.17 show that the combinations of different levels of the four factors affect the value of the degree of imprecision. For instance, In Figure 4.15 clearly shows that the smallest value of $\Delta_{\theta|y}$ appears at the interaction of $n = 200$ and $\lambda_c = 0.001$. This conclusion gives a clear message that survival data with less number of censored observations is more informative with less imprecision.

Table 4.12: ANOVA table for factorial design in Table 4.3 in the case of using imprecise beta prior.

	Df	Sum of Square	Mean Square	F value	P-value
n	1	10.906	10.906	309.809	6.79e-12
λ_c	1	2.780	2.780	78.983	1.38e-07
θ	1	0.558	0.558	15.860	0.00107
r	1	0.136	0.136	3.853	0.06730
$n \times \lambda_c$	1	0.001	0.001	0.034	0.85682
$n \times \theta$	1	0.003	0.003	0.080	0.78135
$n \times r$	1	0.000	0.000	0.001	0.97185
$\lambda_c \times \theta$	1	0.004	0.004	0.121	0.73221
$\lambda_c \times r$	1	0.000	0.000	0.000	0.98847
$\theta \times r$	1	0.003	0.003	0.082	0.77894
Resdiuals	16	0.563	0.035		

AICc = 18.403

Table 4.13: ANOVA table for the main effects of factorial design in Table4.2 in the case of using imprecise beta prior.

	Df	Sum of Square	Mean Square	F value	P-value
n	2	11.101	5.550	690.04	$<$ 2e-16
λ_c	2	2.826	1.413	175.68	1.55e-12
θ	2	0.621	0.310	38.58	3.10e-07
r	2	0.262	0.131	16.30	9.11e-05
Residuals	18	0.145	0.008		

AICc = - 30.793

Table 4.14: ANOVA table for factorial design in Table 4.3 in the case of using imprecise beta prior.

	Df	Sum of Square	Mean Square	F value	P-value
n	2	11.101	5.550	375.726	4.97e-07
λ_c	2	2.826	1.413	95.655	2.81e-05
θ	2	10.621	0.310	21.004	0.00195
r	2	0.262	0.131	8.877	0.01611
$n \times \lambda_c$	4	0.036	0.009	0.602	0.67553
$n \times \theta$	4	0.004	0.001	0.071	0.98840
$n \times r$	4	0.016	0.004	0.277	0.88282
Residuals	6	0.089	0.015		

AICc = 218.207

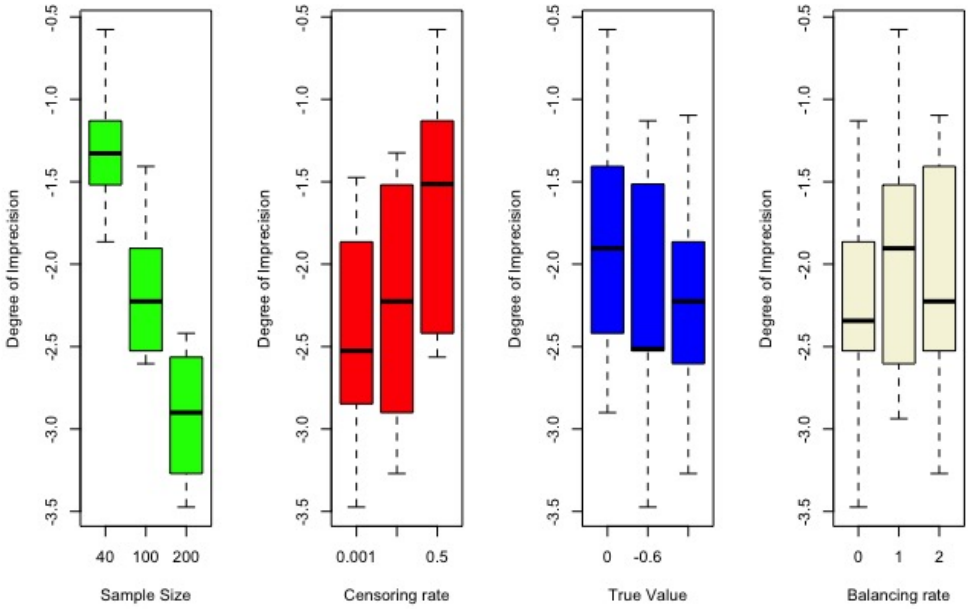


Figure 4.14: Boxplots of significant main effects of the four factors

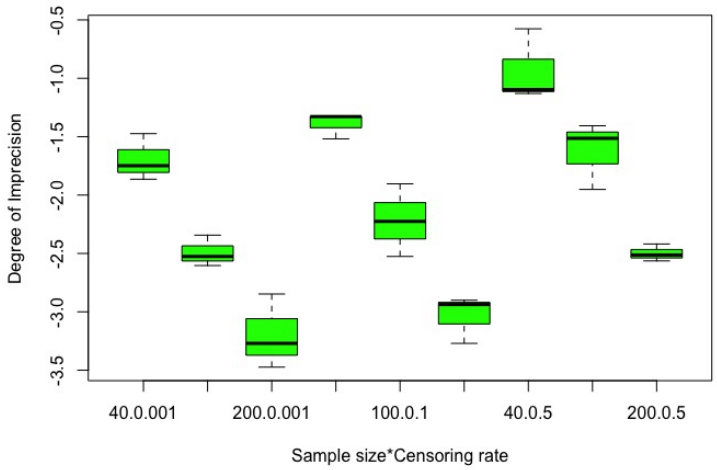


Figure 4.15: Boxplots of the interaction between the sample size factor and the censoring rate factor. The number, for example, 40.0.001 on horizontal axis means the interaction when $n = 40$ and $\lambda_c = 0.001$.

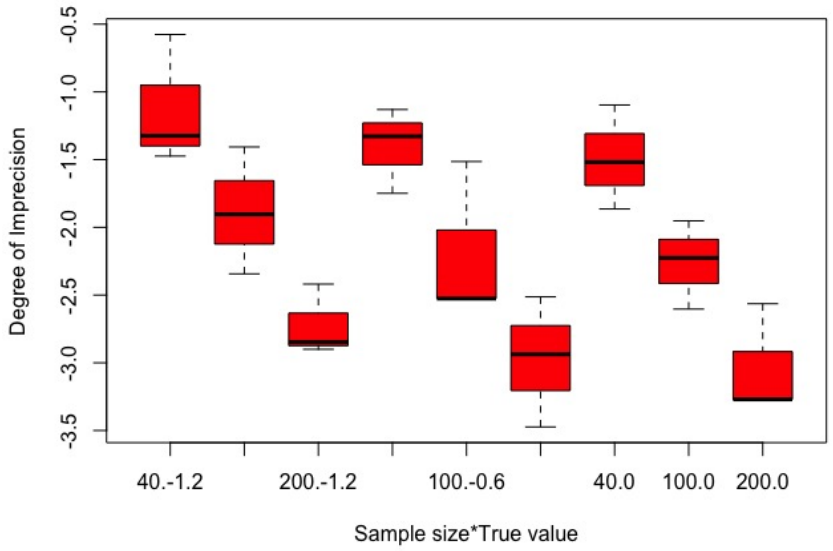


Figure 4.16: Boxplots of of the interaction between the sample size factor and the true value factor. The number, for example, 40.-1.2 on horizontal axis means the interaction when $n = 40$ and $\theta = -1.2$.

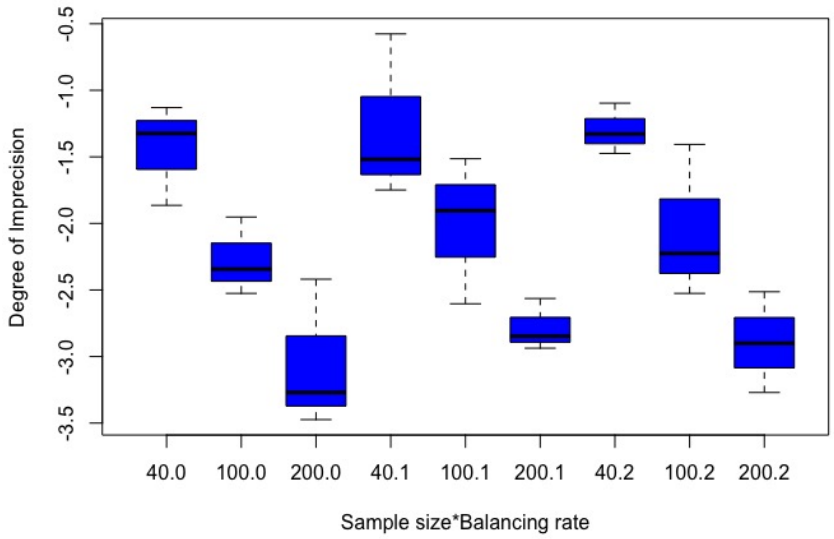


Figure 4.17: Boxplots of of the interaction between the sample size factor and the balancing rate factor. The number, for example, 40.0 on horizontal axis means the interaction when $n = 40$ and $r = 0$. The levels of balancing rate factors are $r = 0$ when $r = 1$, $r = 1$ when $r < 1$, and $r = 2$ when $r > 1$.

4.4.3 ANOVA of combining the degree of imprecision of Normal and Beta

Since two choices of imprecise priors are considered in this work, the ANOVA results in the two case of imprecise normal and beta have the same conclusion. To support this, ANOVA is applied by combining the degree of imprecision of the two cases (Normal and Beta) and add a factor that represents the prior type ("0" means normal, "1" means beta). The results of ANOVA are summarized in Table 4.15 and supported by Figures 4.18 and 4.19. The conclusion here is still the same as in the two cases. Additional conclusion is inferred by the factor of prior type which is significant as shown in Table 4.15.

Table 4.15: ANOVA table for the factorial design in Table 4.3.

	Df	Sum of Square	Mean Square	F value	P-value
n	2	23.146	11.573	628.338	$<$ 2e-16
λ_c	2	6.376	3.188	173.086	$<$ 2e-16
θ	2	1.876	0.938	50.934	3.19e-11
r	2	0.815	0.407	22.124	5.42e-07
$prior$	1	0.288	0.288	15.657	0.000342
$n \times prior$	2	0.014	0.007	0.380	0.686313
$\lambda_c \times prior$	2	0.025	0.013	0.689	0.508518
$\theta \times prior$	2	0.082	0.041	2.229	0.122315
$r \times prior$	2	0.032	0.016	0.865	0.429582
Resdiuals	36	0.663	0.018		

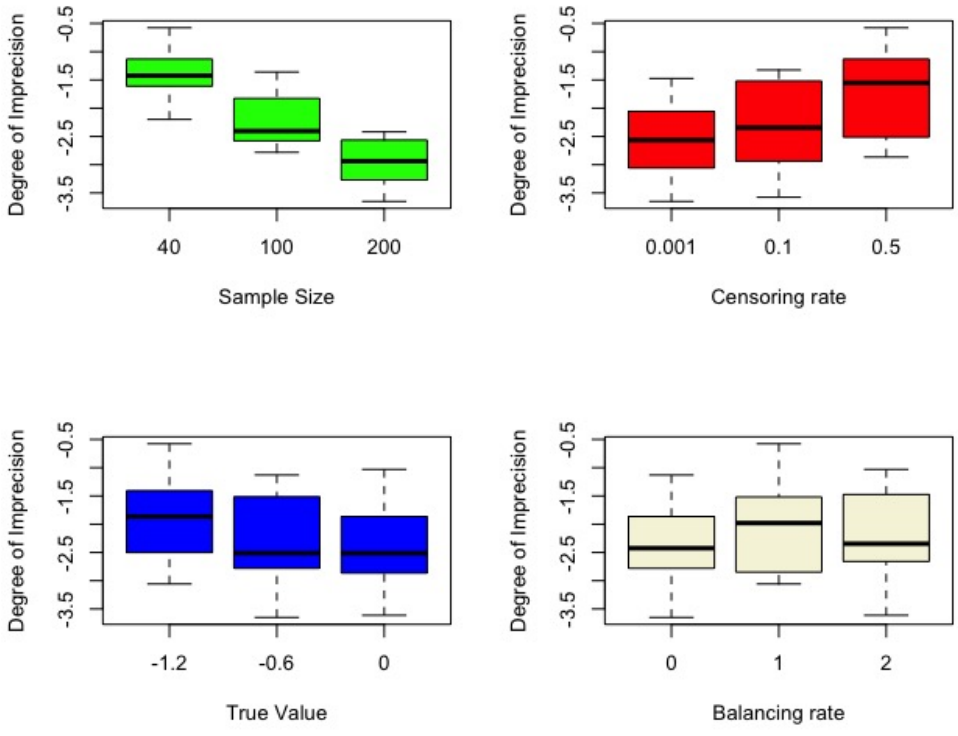


Figure 4.18: Boxplots of main effects of the sample size, censoring rate, true value of the parameter, and balancing rate.

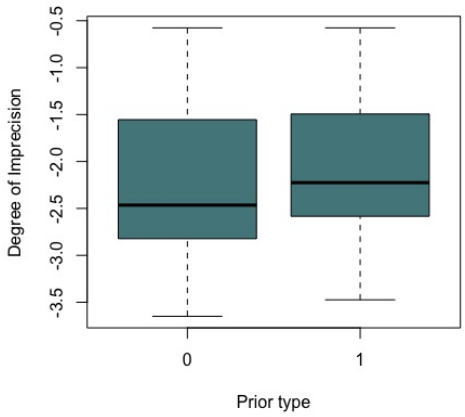


Figure 4.19: Boxplots of prior type factor where "0" means the case of imprecise normal prior and "1" means the case of imprecise beta prior

The boxplots above provide a clear picture of how the four factors plus prior type factor affect the degree of imprecision. Figure 4.19 highlights that the value of degree of imprecision in the normal case is less than its value in beta case. This provides strong support to results in Tables 4.5 and 4.10.

5. Conclusion and Future Work

This chapter represents a conclusion of the implementation of imprecise Bayesian inference on the log-odds ratio in two-way and three-way contingency tables using survival data with right-censored observations. The conclusion is presented in Section 5.1 and future plan is proposed in Section 5.2.

5.1 Conclusion

Two approaches have been applied where in both cases the degree of imprecision is calculated in Chapter 3. In Section 3.2, a family of product of Beta is given to the cell probabilities in the table as an imprecise prior and then posterior expectation of the log-odds ratio is computed and upper and lower posterior expectation are obtained to calculate the degree of imprecision. In Section 3.3, the approach of a re-parametrization of multinomial distribution and logit model is defined in terms of the canonical parameter $\boldsymbol{\theta}$ and $2\theta_3$ being the log odds ratio. Under this approach, the log-odds ratio of 2×2 table is estimated by considering using both a single and a set of prior distributions. In the situation of using an imprecise prior, the degree of imprecision is 0.739 as shown in Table 3.3. By comparing the degree of imprecision in Table 3.1 and 3.3, it is clear that the value of $\Delta_{\log \Psi}$ which is 0.018 in imprecise Dirichlet approach is less than its value in the case of re-parametrization of multinomial distribution and using normal priors.

A generalization to three-ways contingency tables is the main focus in Chapter 4. An

example of real data is applied on Ovarian Cancer Survival data where there are 12 observed death times ($2 \times 2 \times 12$ tables). The estimates of upper and lower posterior expectations and the degrees of imprecision of log-odds ratio are presented in Table 4.2 where the degree of Imprecision $\Delta_{\log \Psi}$ is 0.411 in the case of choosing normal prior and 0.475 in beta prior case, which there are no big difference between the two cases. In the simulation study, survival time data with right-censored observations are generated and displayed in stratified 2×2 tables, one table at each death with no ties. Imprecise Bayesian approach is applied by updating a set of priors to a set of posteriors by using non-central hypergeometric model with the parameter log-odds ratio as our parameter of interest. Two prior families, normal and beta, are given. Four factors with 3 levels in the simulation study are considered and represented in Table 4.3. Comparing the degree of imprecision in the situations of imposing these two kinds of prior is done. The results show that in the case of using a set of normal priors with different means and variances, the degree of imprecision is reduced in the case that the sample size is 200 and censoring rate is 0.1 %. However, when we have a small sample size and high censoring rate, the degree of imprecision is increased. The same findings are observed when a set of beta priors is used. ANOVA for three different model in the case of imprecise normal and beta priors provide the same conclusion that main effects of the four factors are significant. In the case of combining the degree of imprecision of normal and beta cases, ANOVA gives support to the conclusion when the two cases are considered separately. A prior type factor is added and provides a support to results that the degree of imprecision value is reduced slightly in imprecise normal case than in beta case. In short, the uncertainty about the parameter of interest is reduced by having more information, more data, and less censored observations as the results of this work displayed, which is intuitively what one would expect.

5.2 Future Work

While this thesis has demonstrated an implementation of drawing an imprecise Bayesian inference for log-odds ratio in contingency tables using survival data, many opportunities for extending the scope of this thesis remain. This section presents some of these directions.

- The considered survival data in this work is assumed to have no ties and only one observed death at each time, future work will have to address the case of having ties, more one death at each survival time.

- Cox regression model with several covariates would be a future investigation to see how these covariates will affect the degree of imprecision.

- In Section 2.3.4, Nonparametric predictive inference as an approach of imprecise probability theory is presented. A comparison between the methodology of this thesis and NPI would be another future work to be assessed. Also, NPI for next table conditioning on the previous tables is another point to be investigated.

- In the work of Wang's Ph.d dissertation (Wang, 1995) which followed by (Wang and Bickis, 2003) and (Yanqing and Yuan, 2013), the problem of allocating of the treatment to present patient in adaptive design clinical trials study is considered. Wang proposed a classical Bayesian approach in which the prior information is the prior knowledge on the effectiveness of the treatment. Future work could be added to this thesis is implementing imprecise probabilities to decide which treatment is more effective, that is, if the lower posterior expectation of new treatment is greater than the upper posterior expectation of the other treatment.

References

Agresti, A. (2002). *Categorical data analysis*. Wiley seriese in probability and statistics. Wiley-interscience, second edition. 13

Agresti, A. and Hitchcock, D. B. (2005). Bayesian inference for categorical data analysis. *Statistical Methods and Applications*, 14(3):297–330. 17, 56

Albert, J. H. and Gupta, A. K. (1982). Mixtures of dirichlet distributions and estimation in contingency tables. *The Annals of Statistics*, 10(4):1261–1268. 17

Albert, J. H. and Gupta, A. K. (1983). Estimation in contingency tables using prior information. *Journal of the Royal Statistical Society*, 45(1):60–69. 17, 18

Augustin, T., Coolen, F. P. A., Cooman, G., and Troffaes, M. C. M. (2014). *Introduction to imprecise probabilities*. John Wiley and Sons, West Sussex, United Kingdom. 38

Bataineh, O. (2012). *Imprecise Probability Models for Logistic Regression*. PhD thesis, University of Saskatchewan. 66

Benavoli, A. and Zaffalon, M. (2012). A model of prior ignorance for inferences in the one-parameter exponential family. *Journal of statistical planning and inference*, 142:1960–1979. 38

Berger, J. (2006). The case for objective bayesian analysis. *Bayesian Analysis*, 1(3):385–402. 26

Berger, J. O. and Bernardo, J. M. (1992). On the development of refrence priors. *Oxford University Press*, 4:35–60. 29

Berger, J. O., Bernardo, J. M., and Sun, D. (2009). The formal definition of reference priors. *The Annals of Statistics*, 37(2):905–938. 29

Berliner, L. and Hill, B. (1988). Bayesian nonparametric survival analysis. *Journal of American Statistical Association*, 83(403):772–779. 42, 44

Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society*, 41(2):113–147. 28

Bickis, M. (2009). The impresice logit-normal model and its application to estimating hazard functions. *Journal of Statistical Theory and Practice*, 3(1). 41, 66

Bickis, M. (2017). Towards a geometry of imprecise inference. *International Journal of approximate reasoning*, 38:281–297. 38

Bickis, M. and Bickis, U. (2007). Predicting the next pandemic: An exercise in imprecise hazards. In de Cooman G., Vejnarova J., Z. M., editor, *5th International Symposium on Imprecise Probability: Theories and Application*, pages 41–46, Prague, Czech Republic. ISIPTA'07. 42

Birch, Y. W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society*, 125:220–233. 12

Bishop, M. M. (1967). Maximum likelihood in three-way contingency tables. *Ph. D. thesis, De- partment of Statistics, Harvard University.* 12

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Sptinger, New York, second edition. 79

Collett, D. (2003). *Modelling survival data in medical research.* Chapman and Hall, New York, NY., second edition. 20, 23, 48

Coolen, F. (1996). Comparing two populations based on low stochastic structure assumptions. *Statistics and Probability Letters*, 29:297–305. 43

Coolen, F. (1997). An imprecise dirichlet model for bayesian analysis of failure data including rightcensored observations. *Reliability Engineering and System Safty*, 56:61–68. 41, 42

Coolen, F. (2010). *International Encyclopedia of Statistical Science*, chapter Nonparametric predictive inference. Springer. 44

Coolen, F. P. A. and Yan, K. J. (2003). Nonparametric predictive inference with right-censoring data. *Elsevier B. V.*, 126(1):25–54. 42, 44, 45, 48

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, 34(2):187–220. 23

Cox, D. R. and Oakes, D. (1984). *Analysis survival data.* Chapman and Hall, New York. 1

De Finetti, B. (1937). Foresight:its logical laws, its subjective sources. *John Wiley*, pages 93–158. 26, 31

De Finetti, B. (1974). *Theory of Probability*, volume 1. John Wiley, New York. 42

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of Statistics*, 7(2):269–281. 30

Dobson, A. J. and Barnett, A. G. (2008). *An Introduction to generalized linear models*. CRC Press, third edition. 8, 12

Edmunson, J. H., Fleming, T. R., Decker, D. G., Malkasian, G. D., Jefferies, J. A., Webb, M. J., and Kvols, L. K. (1979). Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma vs. minimal residual disease. *Cancer Treatment Reports*, 63:241–247. 67

Fine, T. L. (1988). Lower probability models for uncertainty and nondeterministic processes. *Journal of Statistical Planning and Inference*, 20:389–411. 31

Forster, J. J. (2010). Bayesian inference for poisson and multinomial log-linear model. *Statistical Methodology*, 14:210–224. 19, 56, 59

Gehan, E. A. (1965). A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1):203–223. 20, 62

Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman and Hall, Inc., New York. 42

Good, I. J. (1952). Rational decisions. *Royal Statistical Society*, 14(1):107–114. 31

Goodman, L. A. (1963). On methods for comparing contingency tables. *Journal of the Royal Statistical Society*, 126:94–108. 12

Hill, B. M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 36:677–691. 43

Ibrahim, J., Chen, M., and MacEachern, S. N. (1999). Bayesian variable selection for proportioanl hazards models. *Canadian Journal of Statistics*, 27(4):701–717. 24

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian survival analysis*. Springer, New York. 24

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London*, 186:453–461. 26, 27

Kass, R. E. and Waserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370. 26

Kateri, M. (2014). *Contingency Table Analysis: Methods and Implementation Using R*. Springer, New York. 6, 7, 51

Keynes, J. M. (1921). *A Treatise on Probability*. Machmillan and CO. 31

Kleinbaum, D. G. and Klein, M. (2005). *Survival analysis: A self-learning text*. Springer, New York, NY., third edition. 20

Knuiman, M. W. and Speed, T. (1988). Incorporating prior information into the analysis of contingency tables. *International Biometric Society*, 44(4):1061–1071. 17, 18, 56, 59

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis*. Elsevier Inc., 2nd edition. 50

Lawless, J. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika*, 92:529–542. 42

Lee, C. H. (2014). *Imprecise Prior for Imprecise Inference on Poisson Sampling Model*. PhD thesis, University of Saskatchewan. 66

Leonard, T. (1975). Bayesian estimation methods for two-contingency tables. *Journal of the Royal Statistical Society*, 37(1):23–37. 16, 17, 18, 56

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annual of Mathematical Statistics*, 27:986–1005. 26

Lindley, D. V. (1964). The bayesian analysis of contingency tables. *The Annals of Mathematical Statistics*, 35(4):1622–1643. 15, 18, 19, 56

Lynch, S. M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Sptinger, New York. 25, 49

Mantel, N. (1966). Evaluation of survival data and two new rank-order statistics arising in its consideration. *Cancel Chemotherapy Reports*, 50:163–170. 20, 62

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute*, 22(4):719–729. 1, 21

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. 519.5. Chapman and Hall, London, second edition. 11

Montgomery, D. C. (2009). *Design and Analysis of Experiments*. John Wiley and Sons,Inc., 7th edition. 72

Nazaret, W. A. (1987). Bayesian log linear estimates for three-way contingency tables. *Biometrika*, 74(2):401–410. 16, 17

Omurlu, K. I., Ozdamar, K., and Ture, M. (2009). Comparison of bayesian survival analysis and cox regression analysis in simulated and breast cancer data sets. *Expert Systems With Applications*, 36(8):11341–11346. 25

Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Springer, New York, NY., second edition. 28, 30

Savage, L. J. (1972). *The foundations of statistics*. The Foundations of Statistical Inference, London. 26

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton and London. 31, 45, 48

Slavkovic, A. B. and Fienberg, S. E. (2010). *Algebraic and Geometric Methods in Statistics*, chapter Algebraic geometry of 2x2 contingency tables. Cambridge University Press, first edition. 9

Smith, C. A. B. (1961). Consistency in statistical inference and decision. *Royal Statistical Society*, 23(1):1–37. 31

Walley, P. (1991). *Statistical reasoning with imprecise probablilities*. Chapman and Hall, New York, NY. 2, 32, 34, 37, 38, 47

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society*, 58(1):3–57. 39, 40, 41, 51, 52, 65

Walley, P., Gurrin, L., and Burton, P. (1996). Analysis of clinical data using imprecise prior probabilities. *Journal of the Royal Statistical Society*, 45(4):457–485. 40, 51, 65

Wang, X. (1995). *Sequential Selections of Treatments with Delayed Responses*. PhD thesis, University of Saskatchewan. 96

Wang, X. and Bickis, M. G. (2003). One-armed bandit models with continuous and delayed responses. *Mathematical Methods of Operations Research*, 58:209–219. 96

Williams, P. M. (1975). Coherence, strict coherence and zero probabilities. In *Proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science*, volume VI, pages 29–33. 31

Williams, P. M. (2007). Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383. 31

Yanqing, Y. and Yuan, Y. (2013). An optimal allocation for response-adaptive designs. *Journal of Applied Statistics*, 40(9):1996–2008. 96

Appendix A

Complete Results of Chapter 4

The following are the 27 combinations (runs) in two cases of normal and beta priors. The results are presented in Tables and ECDF plots.

Combination 1:

$n = 40$, $\lambda_c = 0.001$, $\theta = 0$, and $r = 1$.

Table A.1: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 1.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	199.966	0.024	199.941
	10	1.402	0.767	0.635
	20	-0.240	-0.478	0.238
	39	-0.012	-0.124	0.111
Beta	1	3.358	-0.759	4.118
	10	1.163	0.553	0.609
	20	-0.215	-0.510	0.295
	39	-0.007	-0.162	0.155

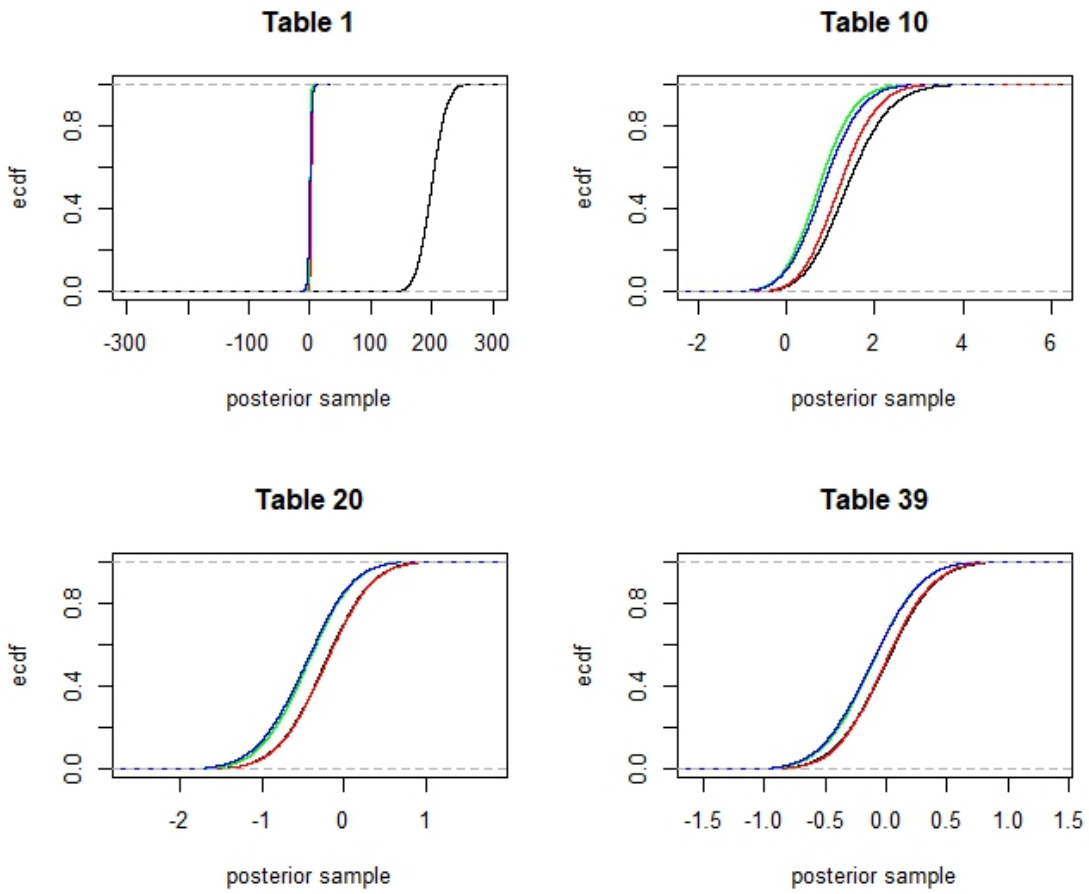


Figure A.1: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 1.

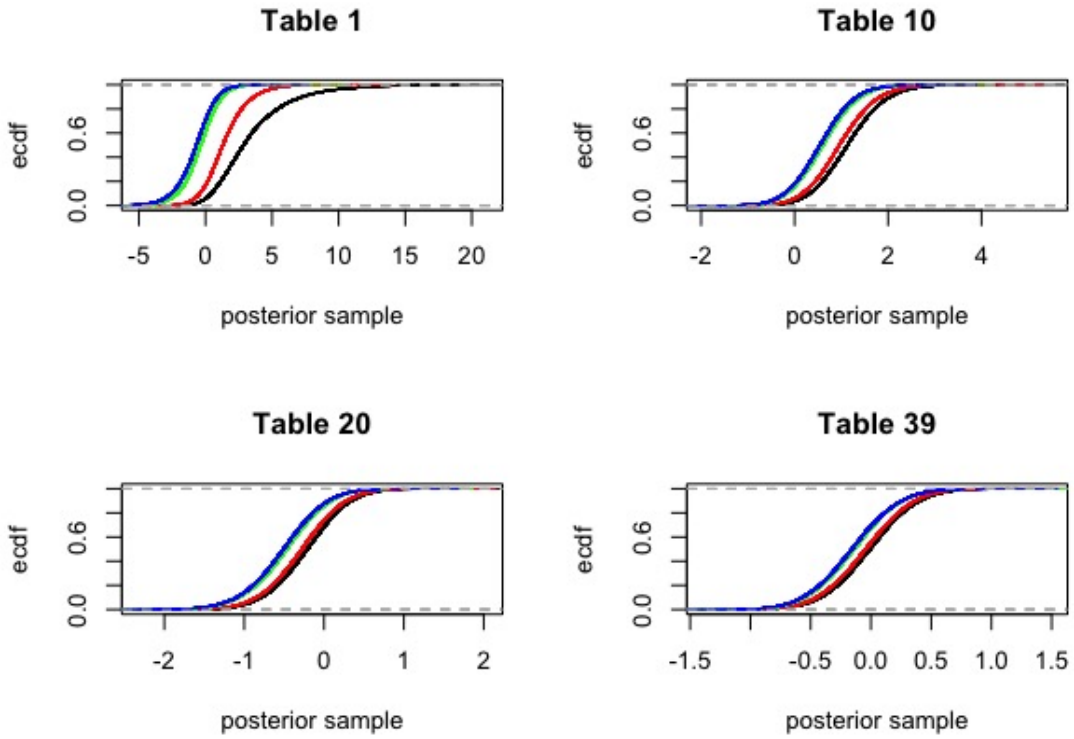


Figure A.2: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 1.

Combination 2:

$n = 40$, $\lambda_c = 0.001$, $\theta = -0.6$, and $r < 1$.

Table A.2: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 2.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	0.672	-200.090	200.762
	10	-1.124	-3.506	2.381
	20	-0.507	-0.838	0.330
	39	-0.306	-0.434	0.128
Beta	1	0.407	-11.408	11.815
	10	-0.945	-2.263	1.318
	20	-0.454	-0.827	0.372
	39	-0.285	-0.460	0.174

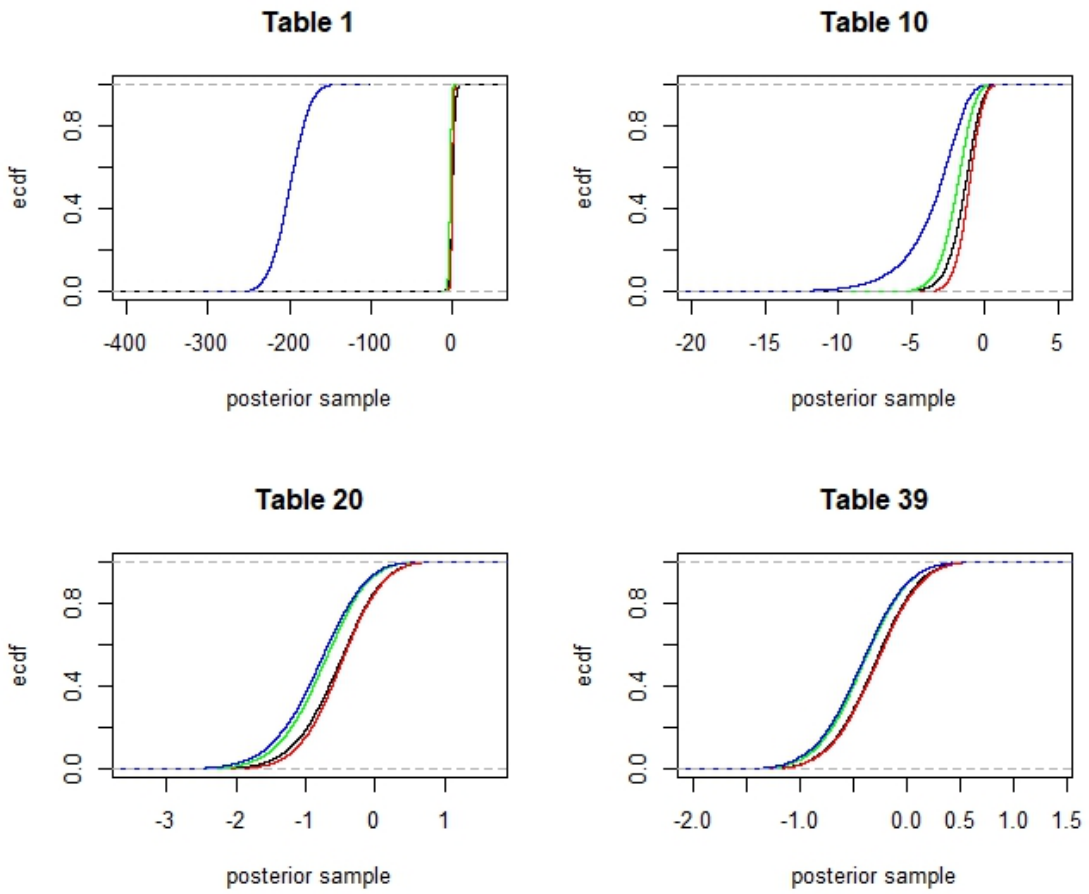


Figure A.3: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 2.

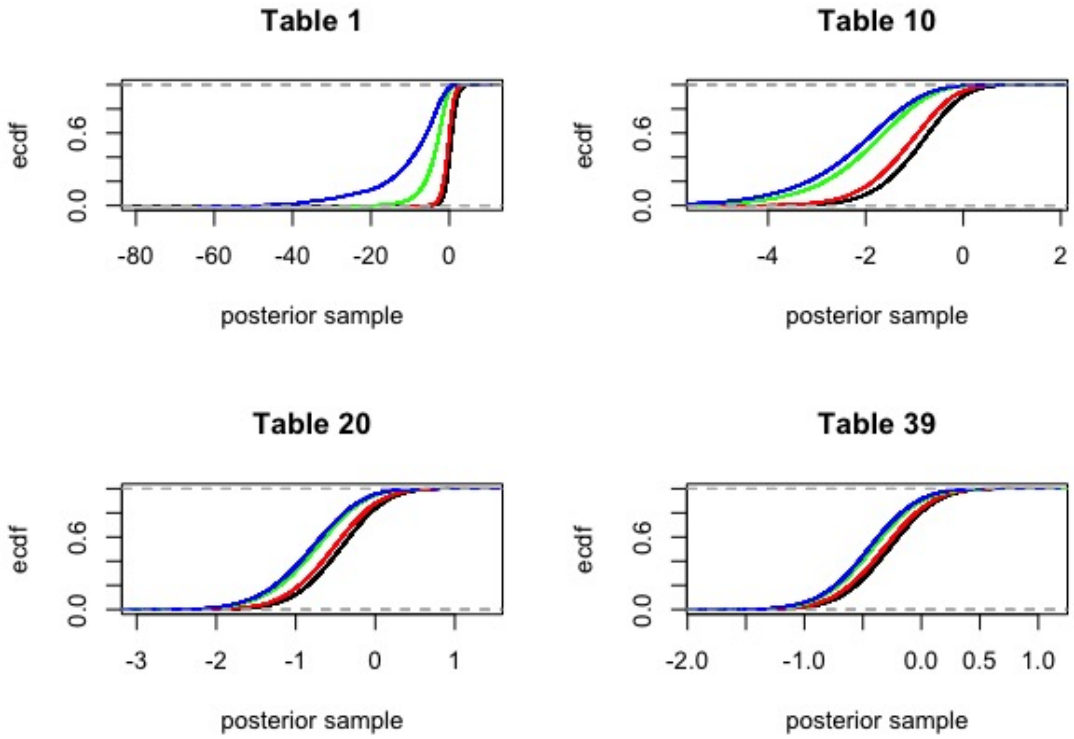


Figure A.4: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 2.

Combination 3:

$n = 40$, $\lambda_c = 0.001$, $\theta = -1.2$, and $r > 1$.

Table A.3: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 3.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	-0.488	-199.948	199.459
	10	-0.760	-1.317	0.557
	15	-1.414	-1.817	0.402
	20	-0.650	-0.831	0.180
Beta	1	2.947	-1.213	4.161
	10	-1.382	-1.989	0.607
	20	-1.227	-1.557	0.330
	38	-0.613	-0.843	0.229

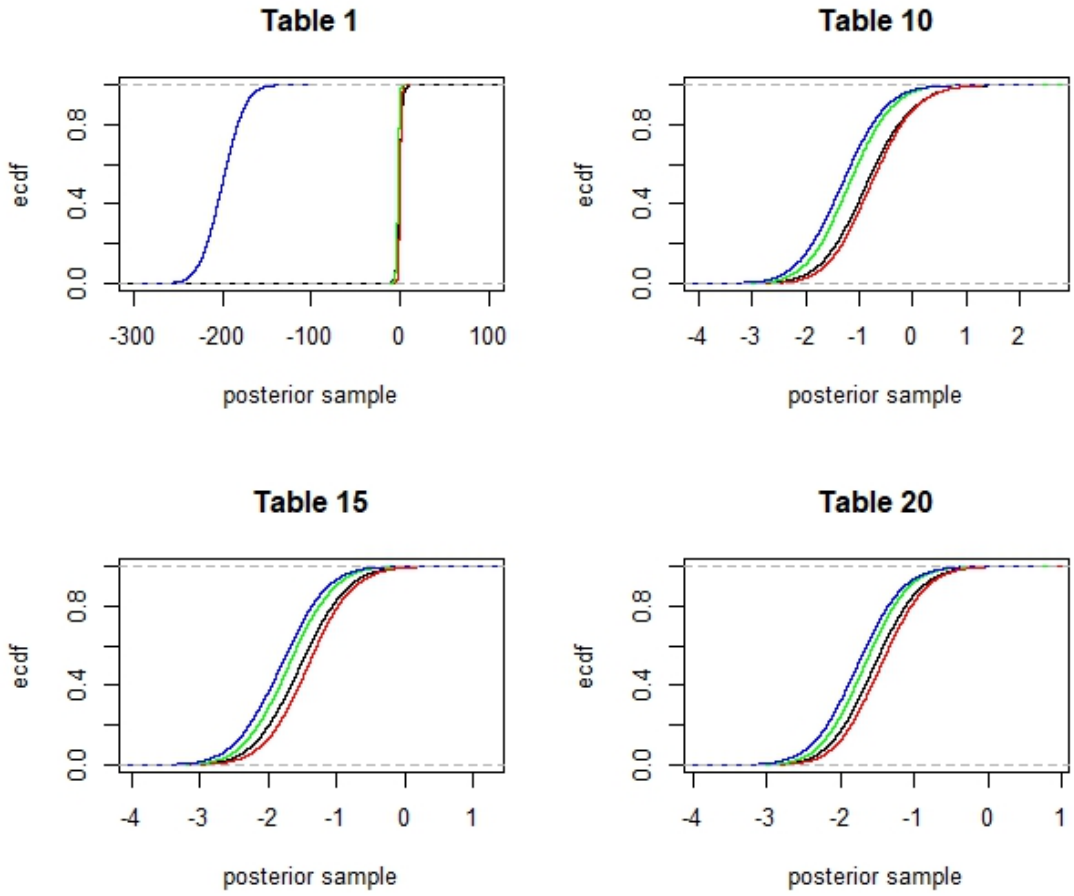


Figure A.5: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 3.

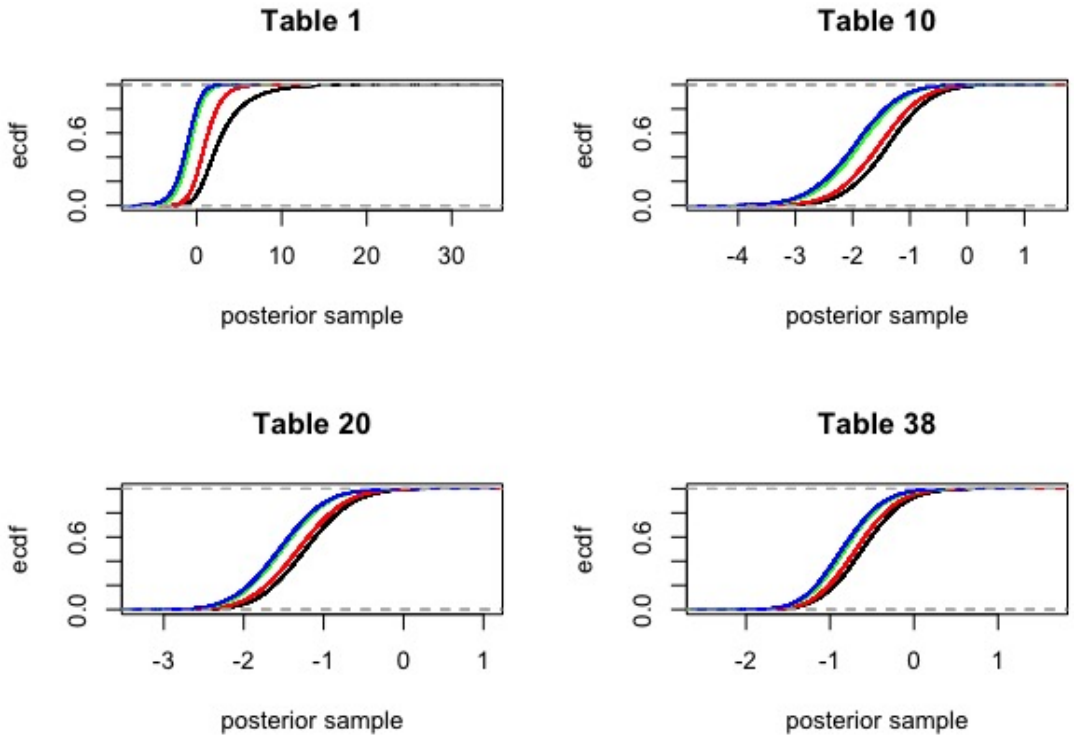


Figure A.6: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 3.

Combination 4:

$n = 40$, $\lambda_c = 0.1$, $\theta = 0$, and $r < 1$.

Table A.4: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 4.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	0.724	-199.907	200.631
	10	-0.515	-1.397	0.881
	20	-0.020	-0.263	0.243
	30	0.072	-0.079	0.152
Beta	1	0.413	-12.366	12.779
	10	-0.407	-1.246	0.839
	20	0.003	-0.316	0.319
	30	0.082	-0.136	0.219

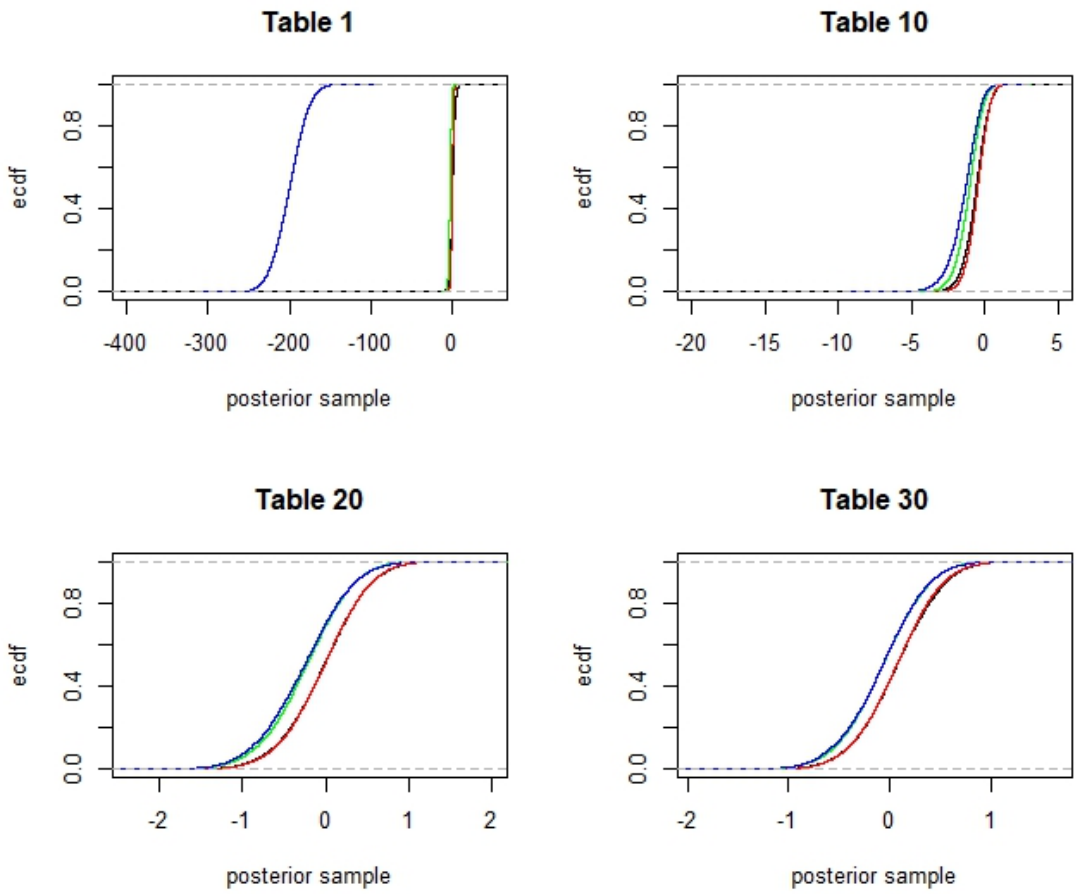


Figure A.7: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 4.

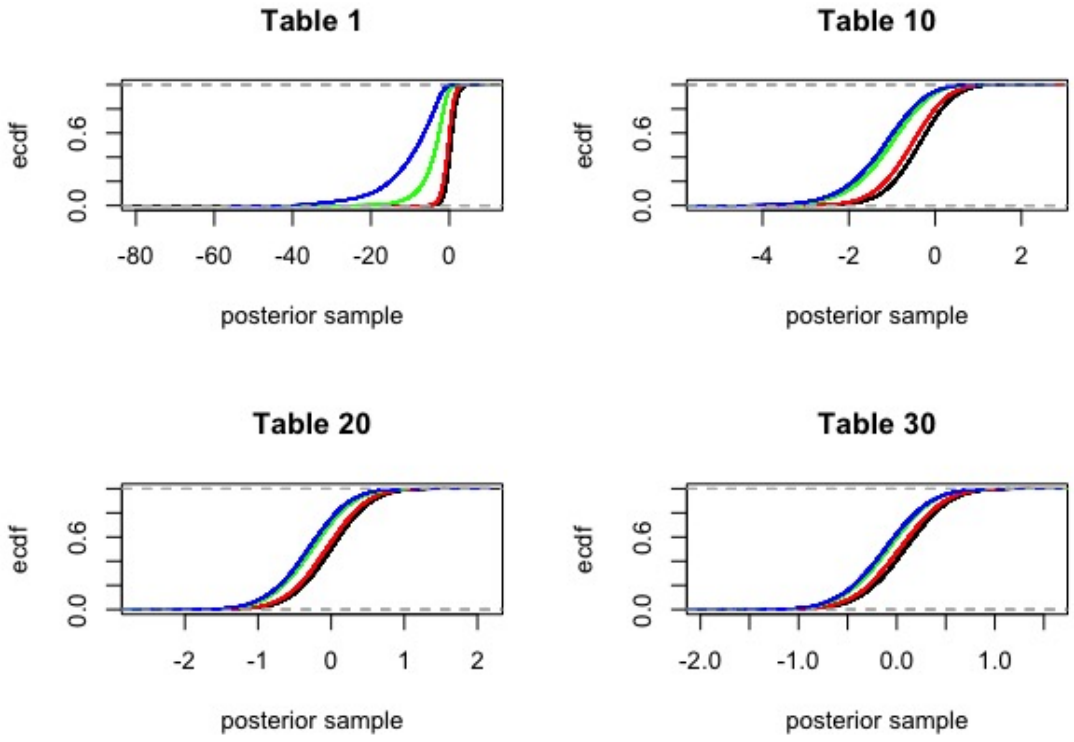


Figure A.8: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 4.

Combination 5:

$n = 40$, $\lambda_c = 0.1$, $\theta = -0.6$, and $r > 1$.

Table A.5: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 5.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	200.008	-1.221	201.229
	10	-0.564	-1.107	0.543
	20	-1.128	-1.420	0.291
	33	-0.562	-0.763	0.200
Beta	1	2.984	-1.256	4.240
	10	-0.471	-1.085	0.613
	20	-1.053	-1.381	0.328
	33	-0.517	-0.782	0.265

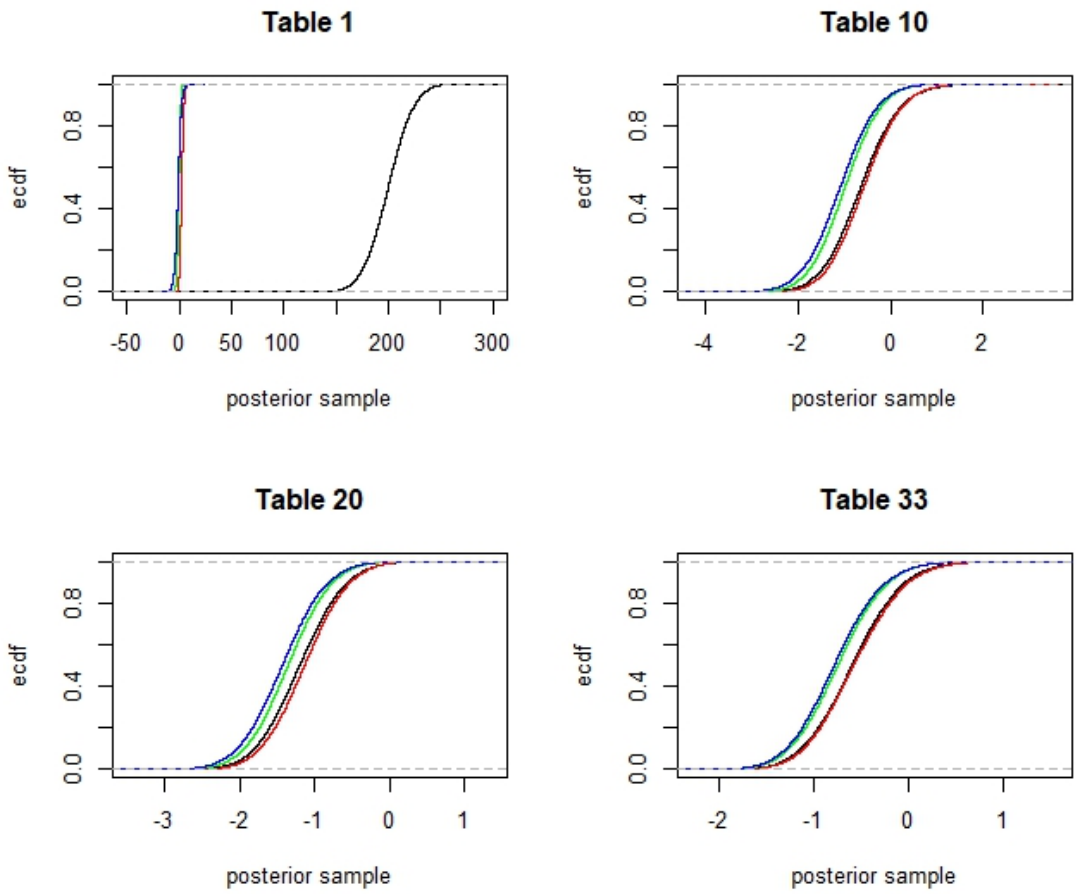


Figure A.9: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 5.

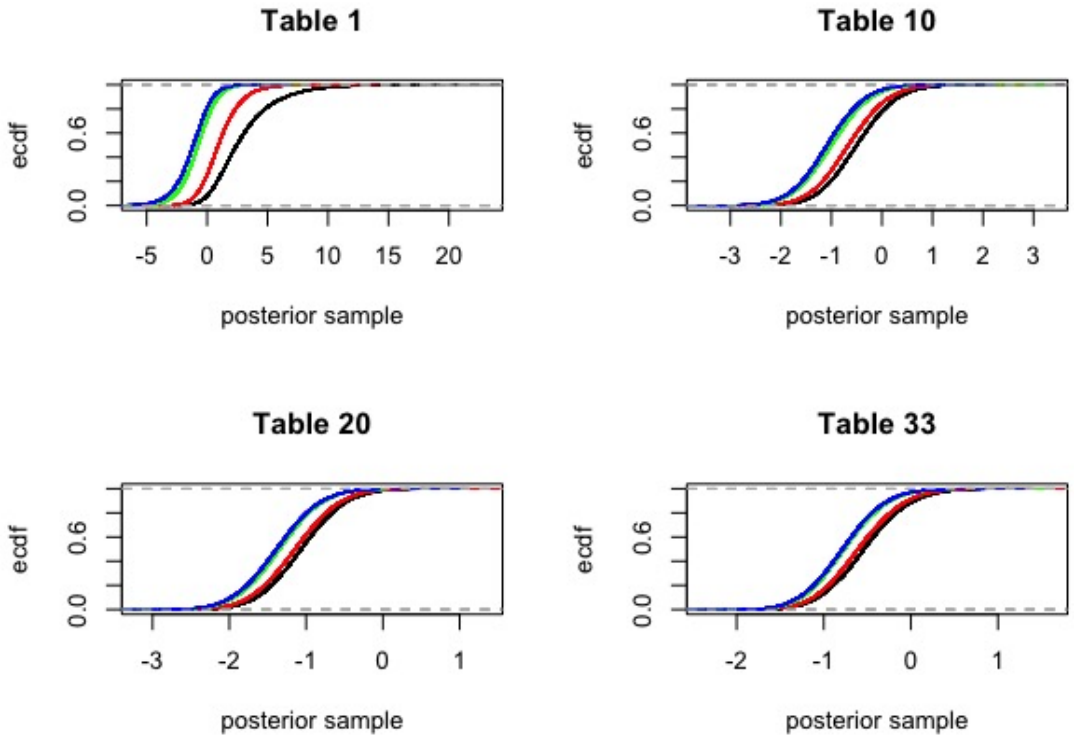


Figure A.10: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 5.

Combination 6:

$n = 40$, $\lambda_c = 0.1$, $\theta = -1.2$, and $r = 1$.

Table A.6: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 6.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	200.163	-0.041	200.205
	10	-0.666	-1.298	0.632
	20	-1.118	-1.441	0.323
	26	-1.202	-1.455	0.253
Beta	1	3.304	-0.816	4.121
	10	-0.563	-1.207	0.643
	20	-1.048	-1.396	0.347
	26	-1.150	-1.416	0.266

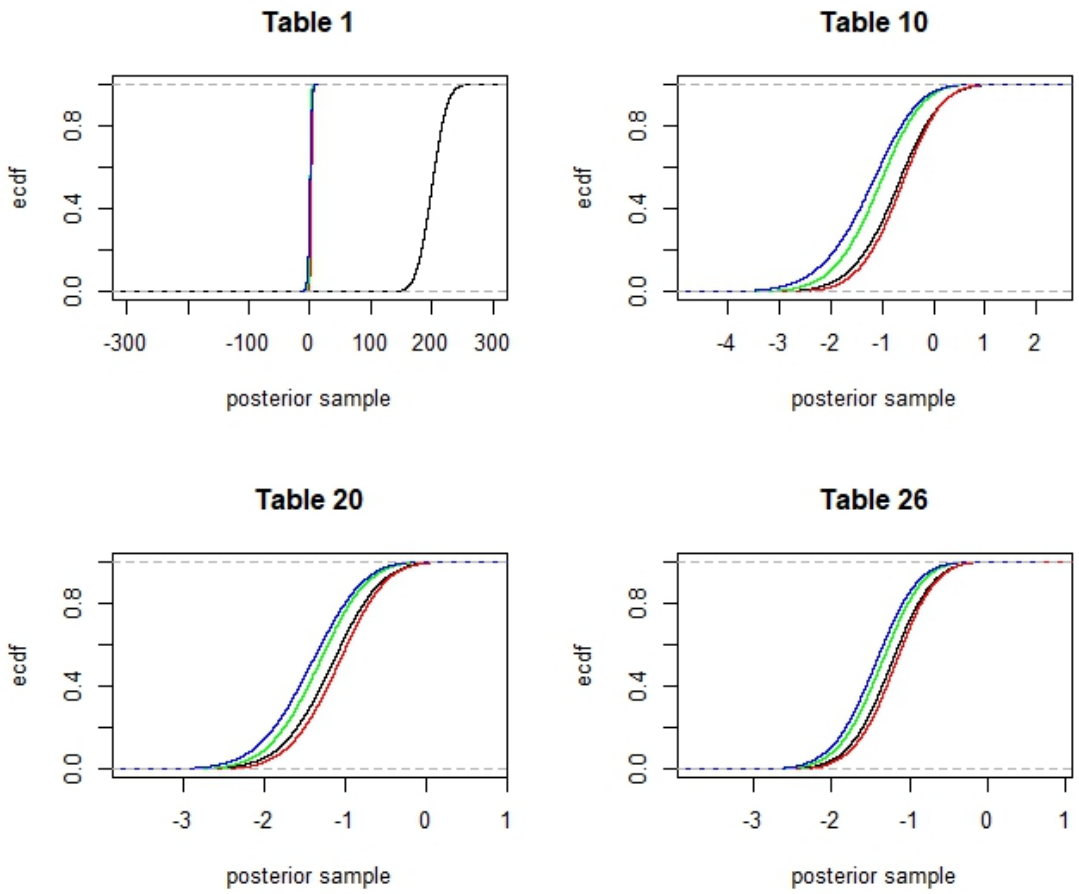


Figure A.11: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 6.

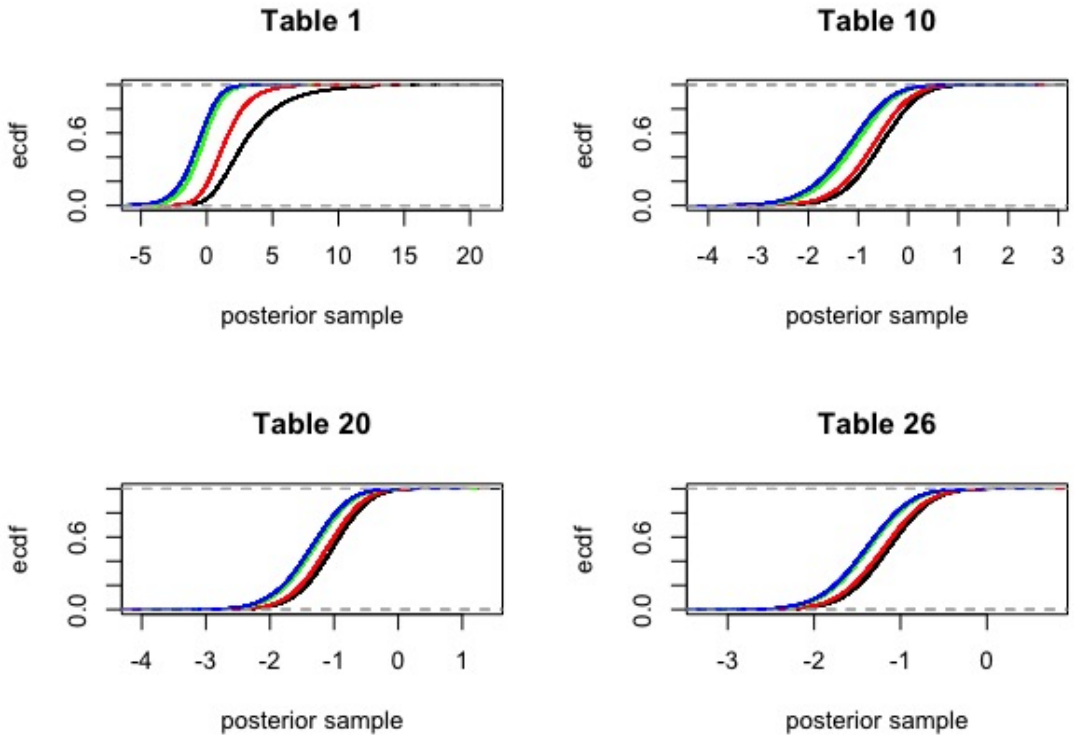


Figure A.12: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 6.

Combination 7:

$n = 40$, $\lambda_c = 0.5$, $\theta = 0$, and $r > 1$.

Table A.7: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 7.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	199.833	-1.394	201.227
	10	2.605	0.382	2.223
	15	0.431	-0.048	0.480
	17	0.172	-0.185	0.357
Beta	1	2.979	-1.233	4.212
	10	0.038	-0.598	0.637
	15	-0.322	-0.751	0.429
	19	-0.605	-0.939	0.334

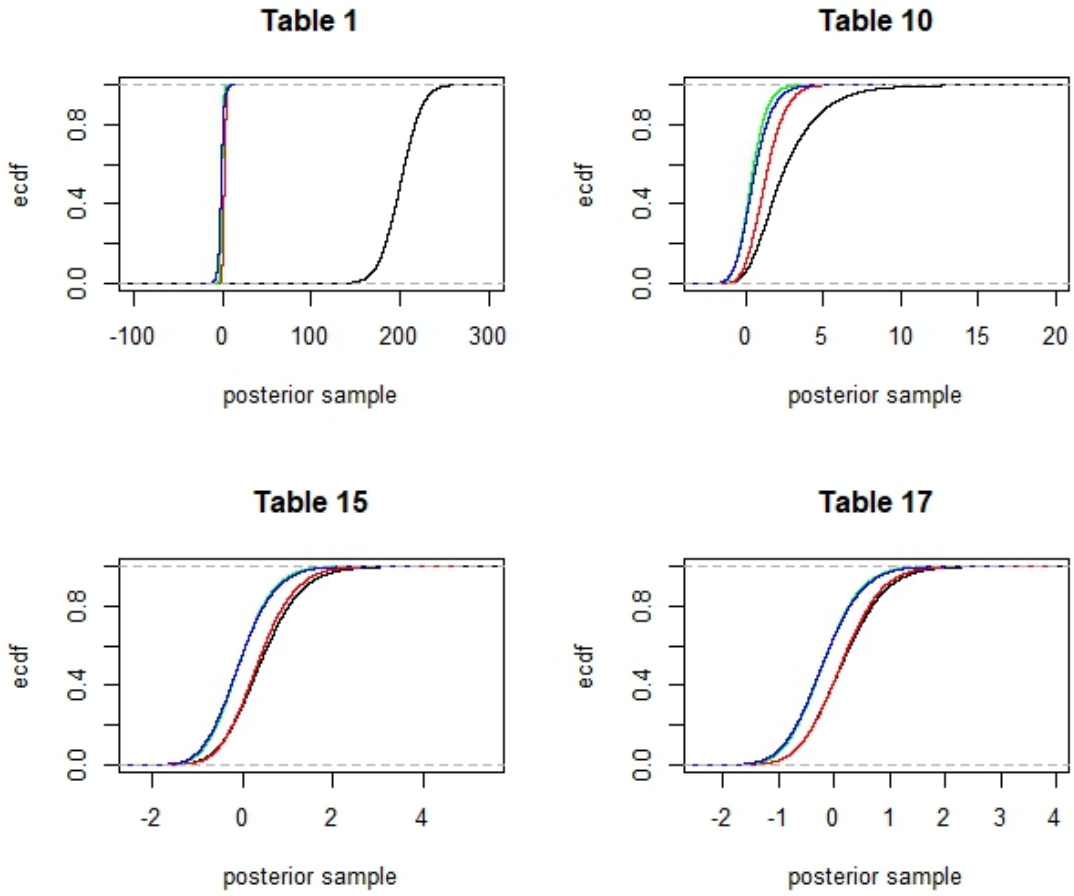


Figure A.13: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 7.

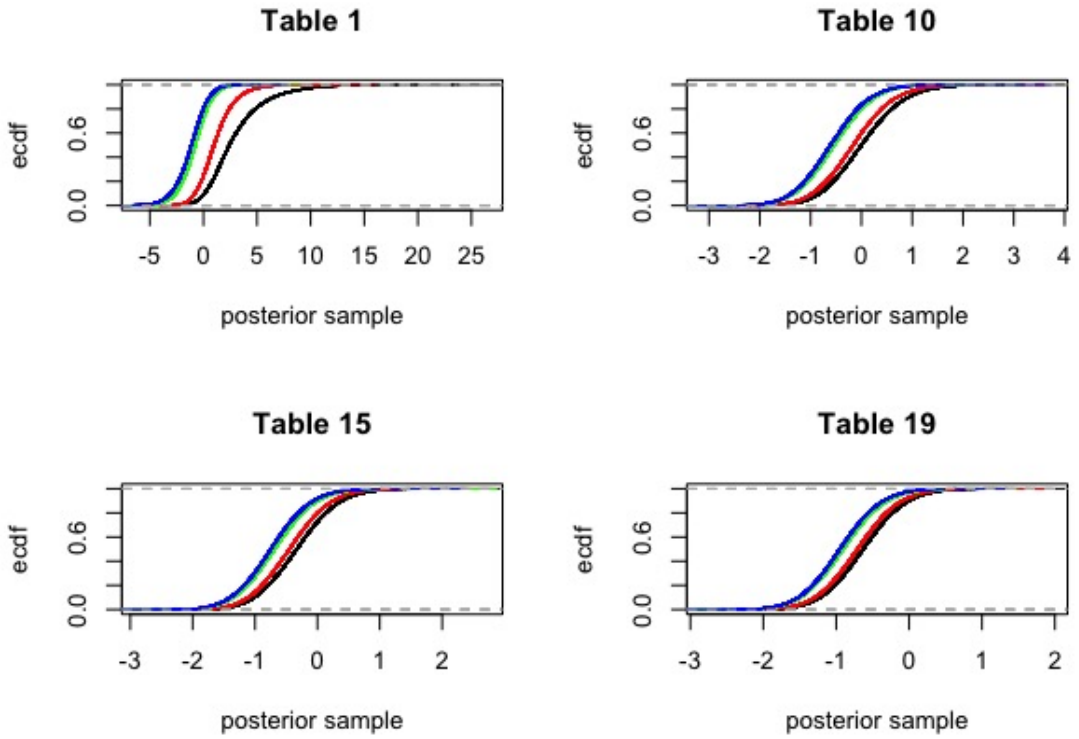


Figure A.14: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 7.

Combination 8:

$n = 40$, $\lambda_c = 0.5$, $\theta = -0.6$, and $r = 1$.

Table A.8: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 8.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	200.129	-0.012	200.142
	10	-0.184	-0.671	0.486
	15	-0.566	-0.927	0.361
	19	-0.551	-0.827	0.276
Beta	1	3.336	-0.782	4.118
	10	-0.122	-0.695	0.572
	15	-0.503	-0.929	0.425
	19	-0.501	-0.824	0.323

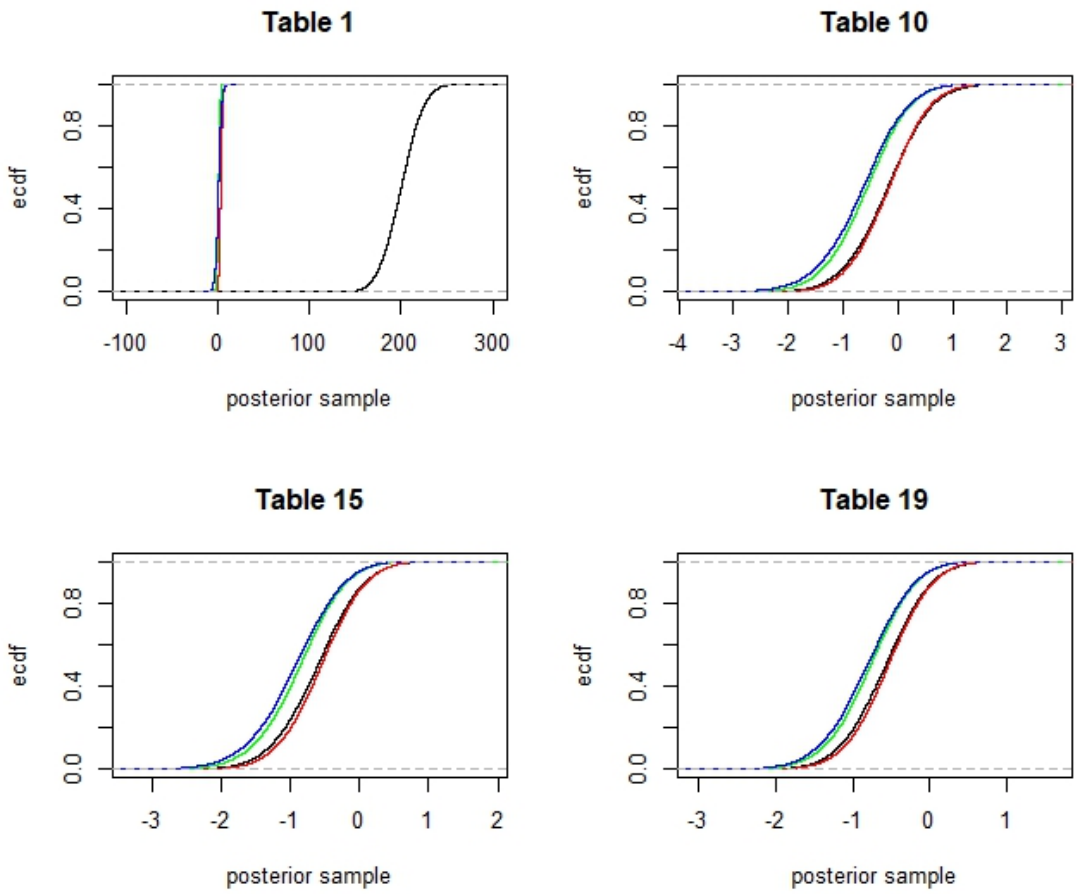


Figure A.15: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 8.

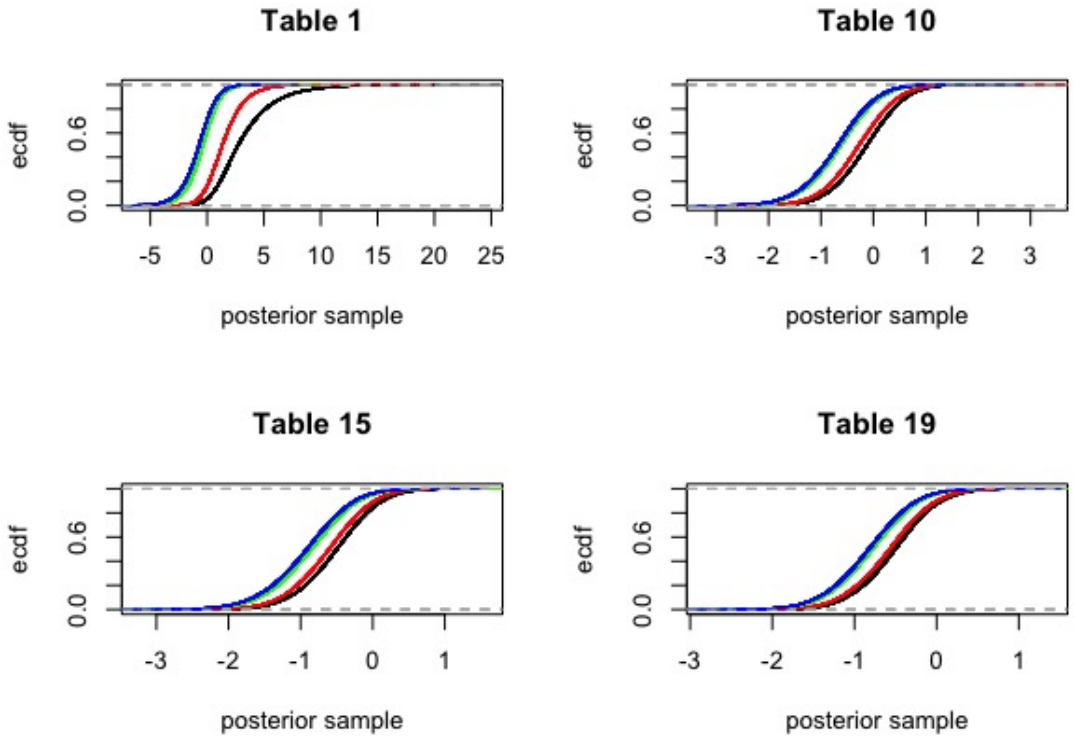


Figure A.16: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 8.

Combination 9:

$n = 40$, $\lambda_c = 0.5$, $\theta = -1.2$, and $r < 1$.

Table A.9: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 9.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	0.686	-199.649	200.335
	10	-1.300	-3.753	2.453
	14	-1.165	-2.104	0.939
	17	-1.158	-1.754	0.595
Beta	1	0.403	-10.299	10.702
	10	-1.130	-2.470	1.339
	14	-1.045	-1.820	0.774
	17	-1.057	-1.619	0.562

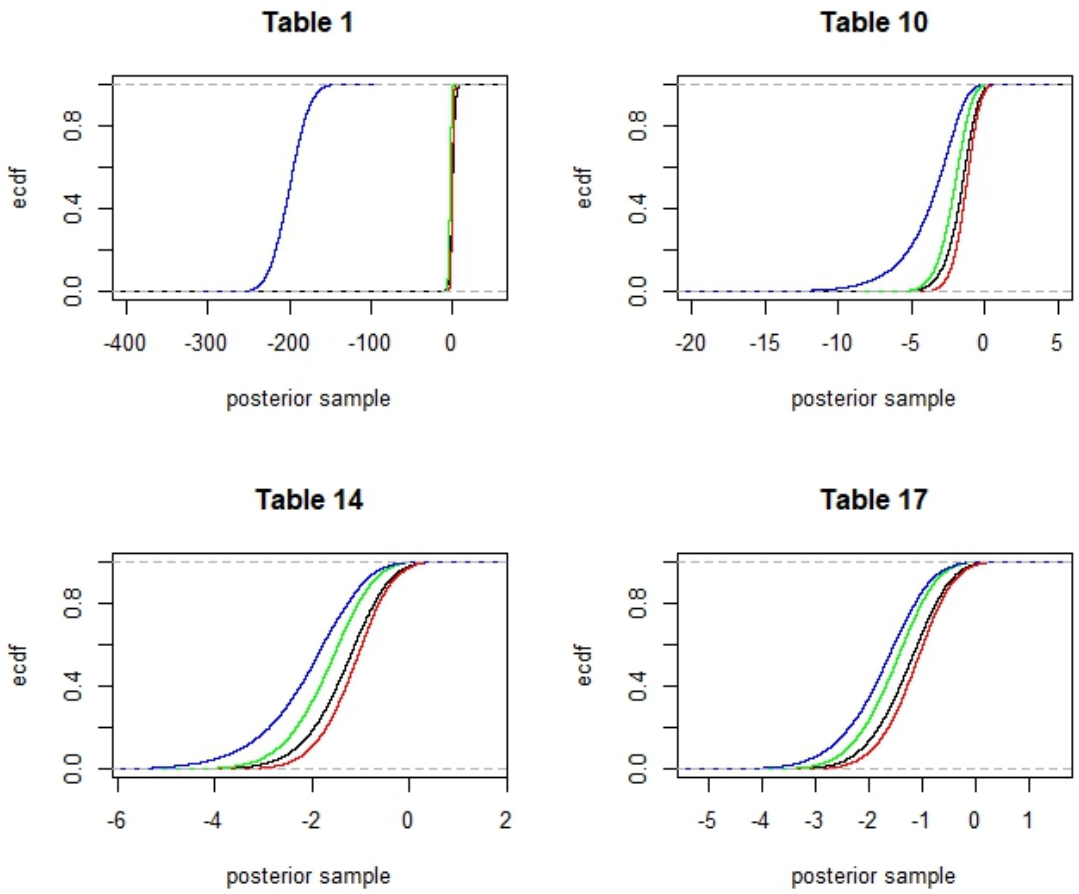


Figure A.17: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 9.

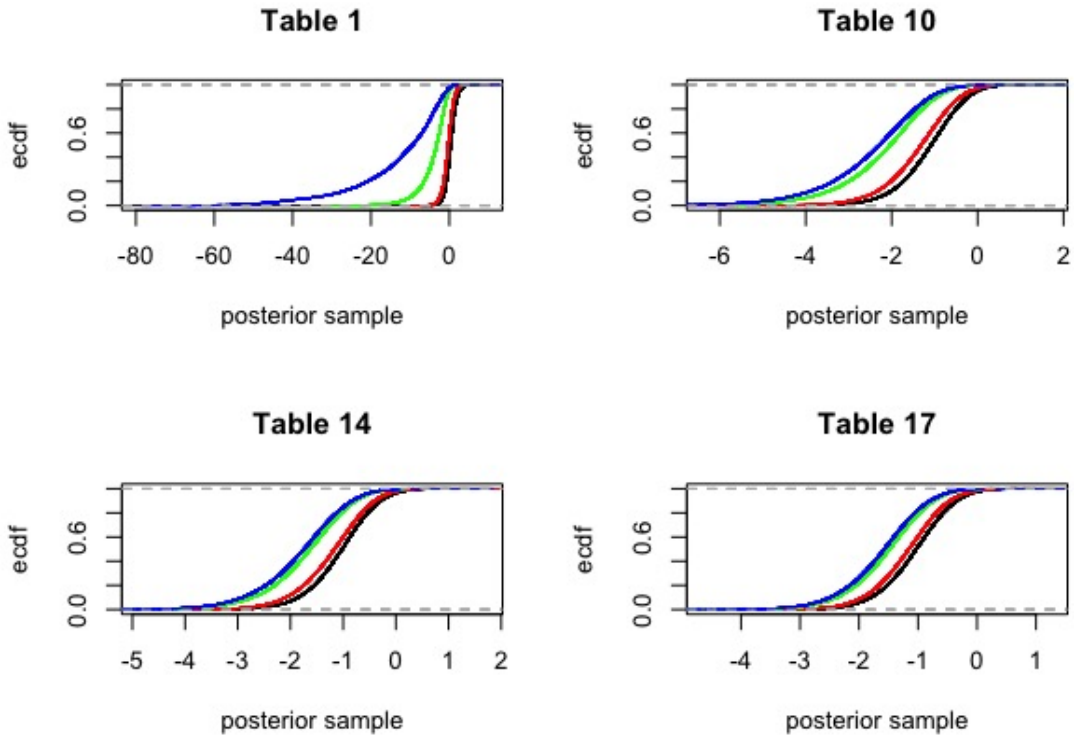


Figure A.18: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 9.

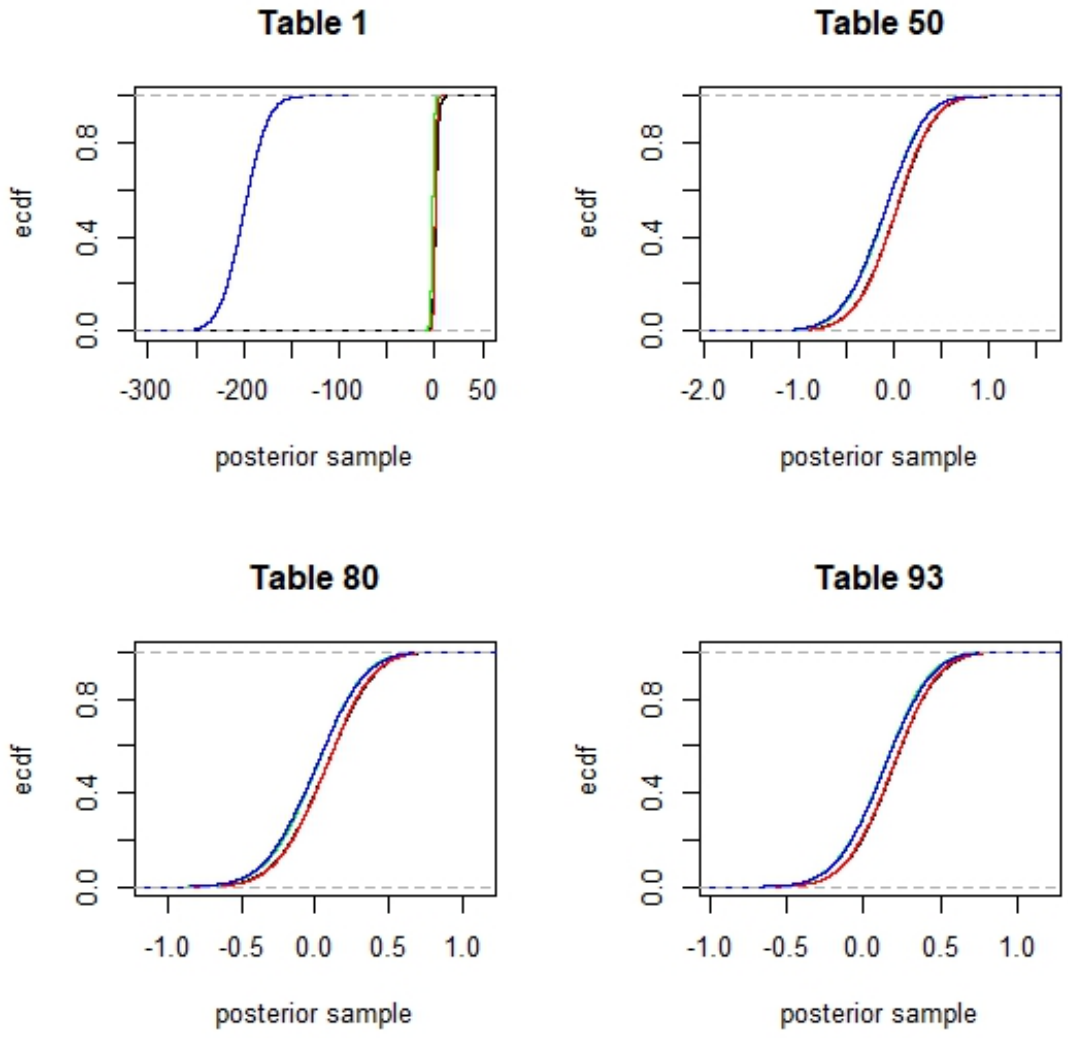
Combination 10:

$n = 100$, $\lambda_c = 0.001$, $\theta = 0$, and $r < 1$.

Table A.10: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 10.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	1.146	-199.891	201.037
	50	0.001	-0.115	0.117
	80	0.065	-0.003	0.069
	93	0.184	0.121	0.062
Beta	1	0.492	-12.106	12.598
	30	0.101	-0.136	0.237
	60	0.250	0.138	0.112
	97	0.215	0.141	0.074

Table A.11: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 10.



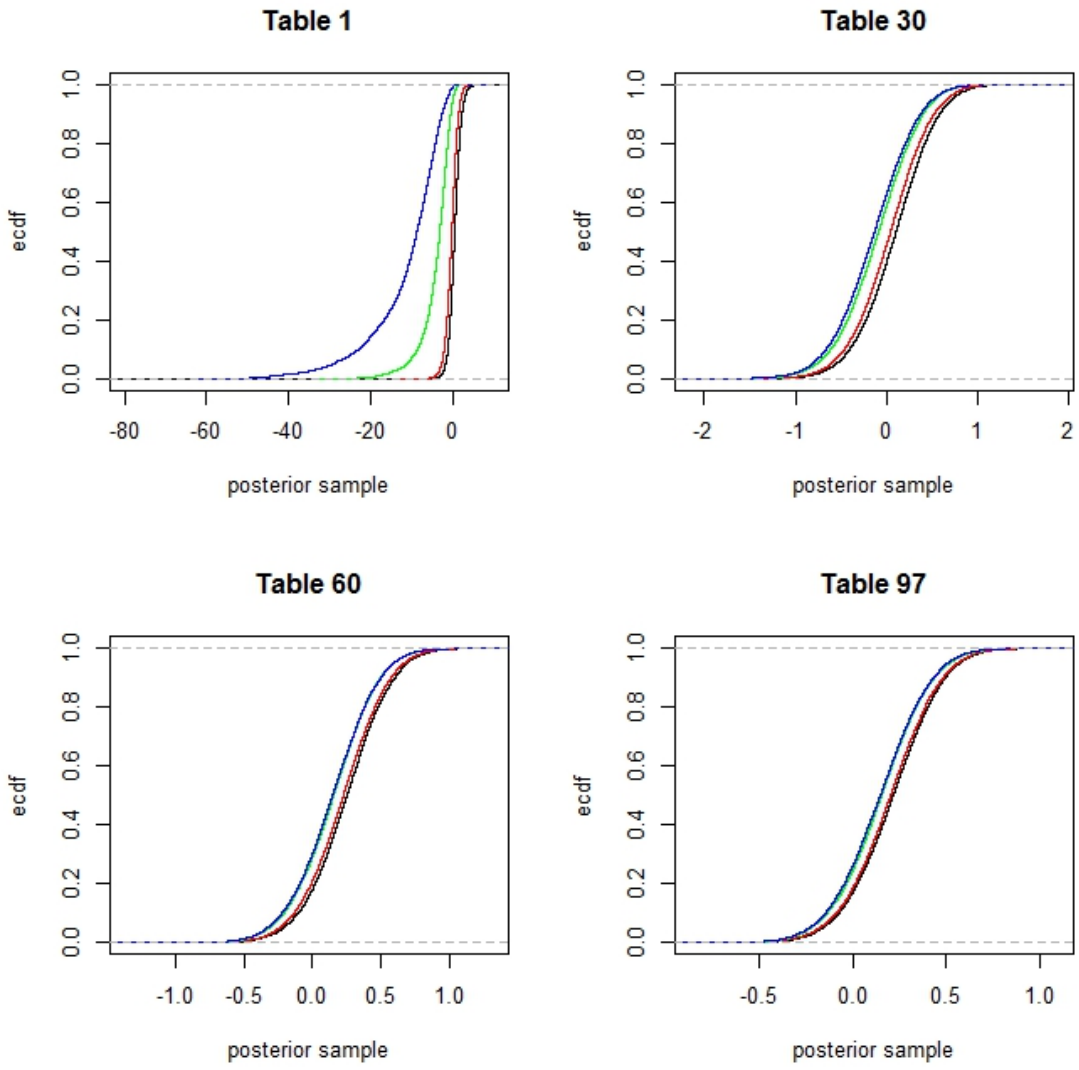


Figure A.19: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 10.

Combination 11:

$n = 100$, $\lambda_c = 0.001$, $\theta = -0.6$, and $r > 1$.

Table A.12: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 11.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	199.840	-1.112	200.953
	30	-0.347	-0.518	0.171
	60	-0.957	-0.768	0.093
	81	-0.698	-0.769	0.070
Beta Beta	1	2.974	-1.215	4.190
	30	-0.711	-0.907	0.196
	60	-0.510	-0.623	0.112
	95	-0.494	-0.575	0.080

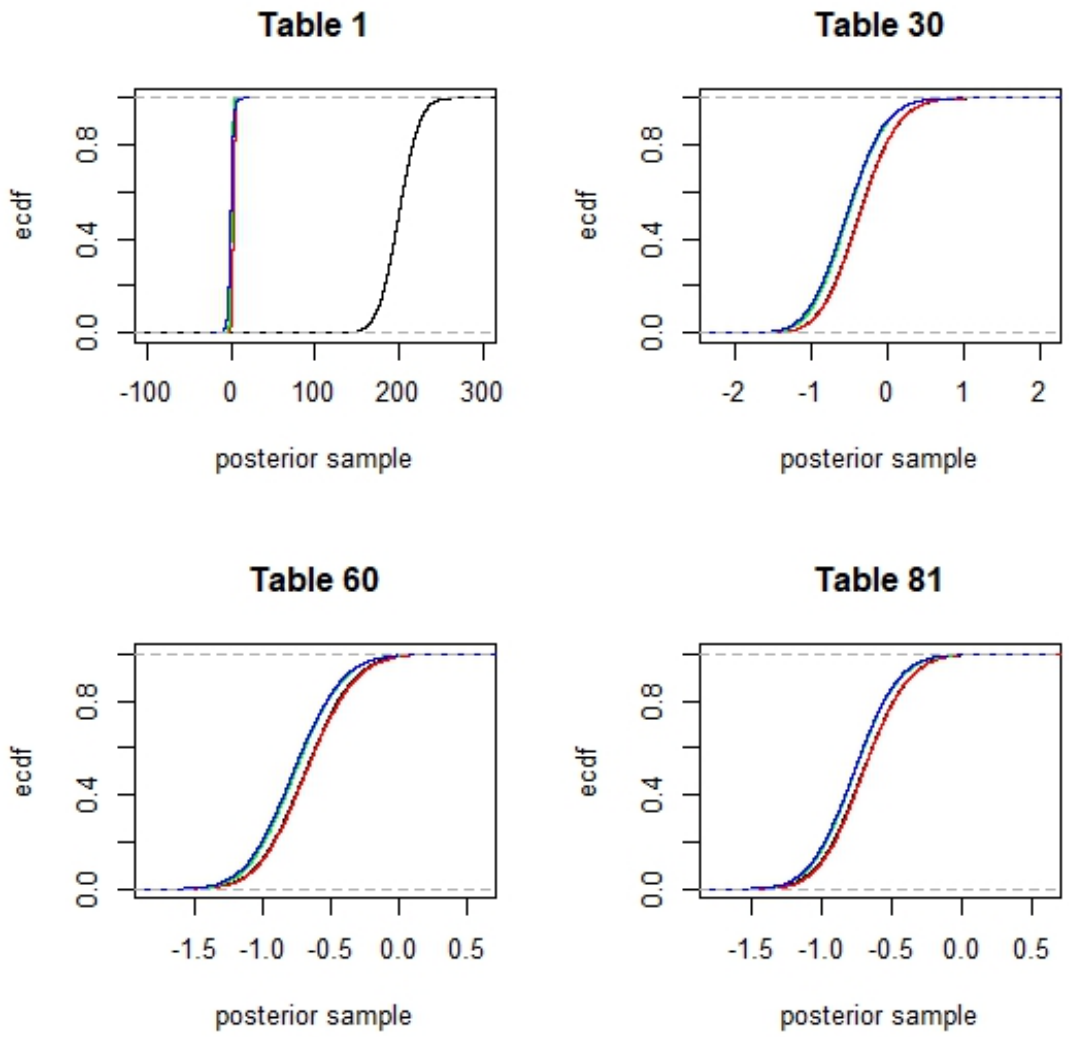


Figure A.20: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 11.

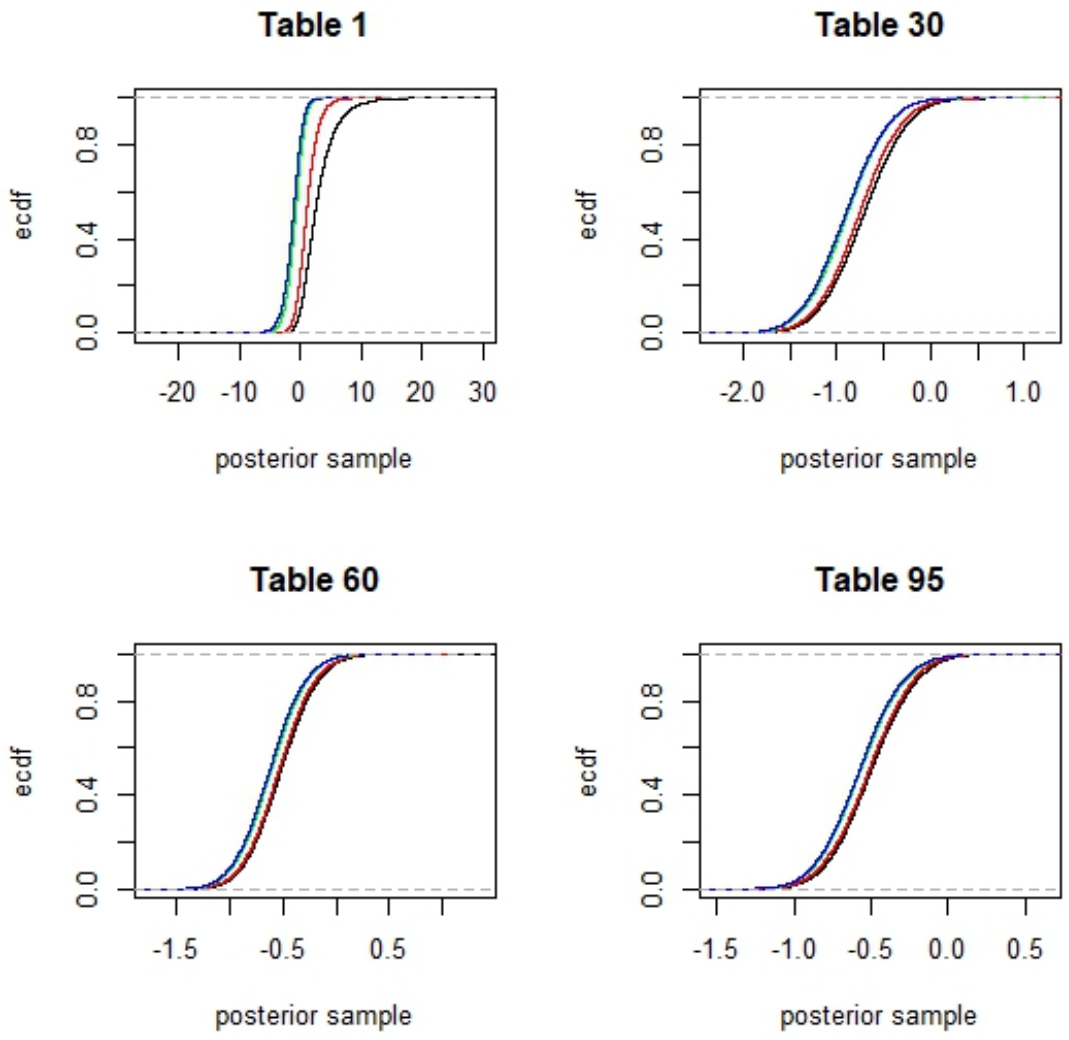


Figure A.21: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 11.

Combination 12:

$n = 100$, $\lambda_c = 0.001$, $\theta = -1.2$, and $r = 1$.

Table A.13: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 12.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	200.001	-0.021	200.023
	30	-1.332	-1.604	0.271
	50	-1.299	-1.434	0.135
	81	-1.336	-1.419	0.082
Beta	1	3.285	-0.776	4.061
	30	-0.900	-1.131	0.230
	60	-1.122	-1.240	0.118
	84	-1.308	-1.404	0.096

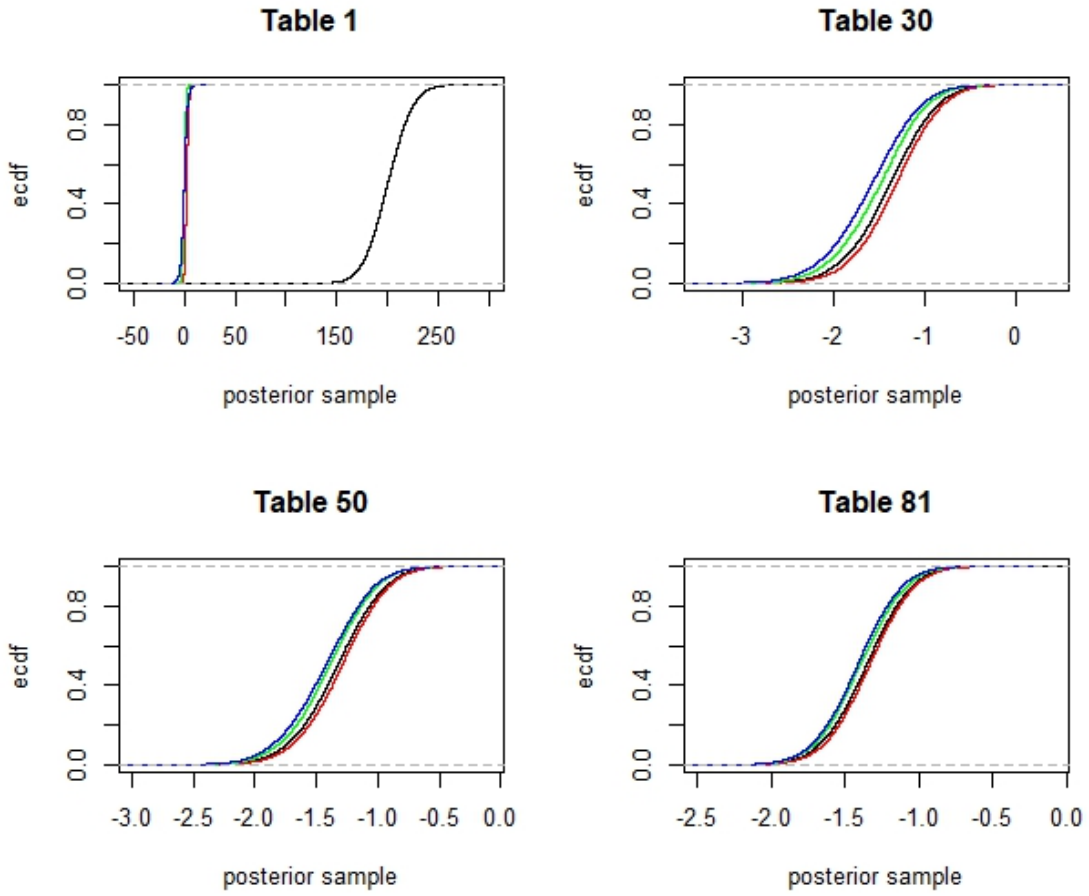


Figure A.22: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 12.

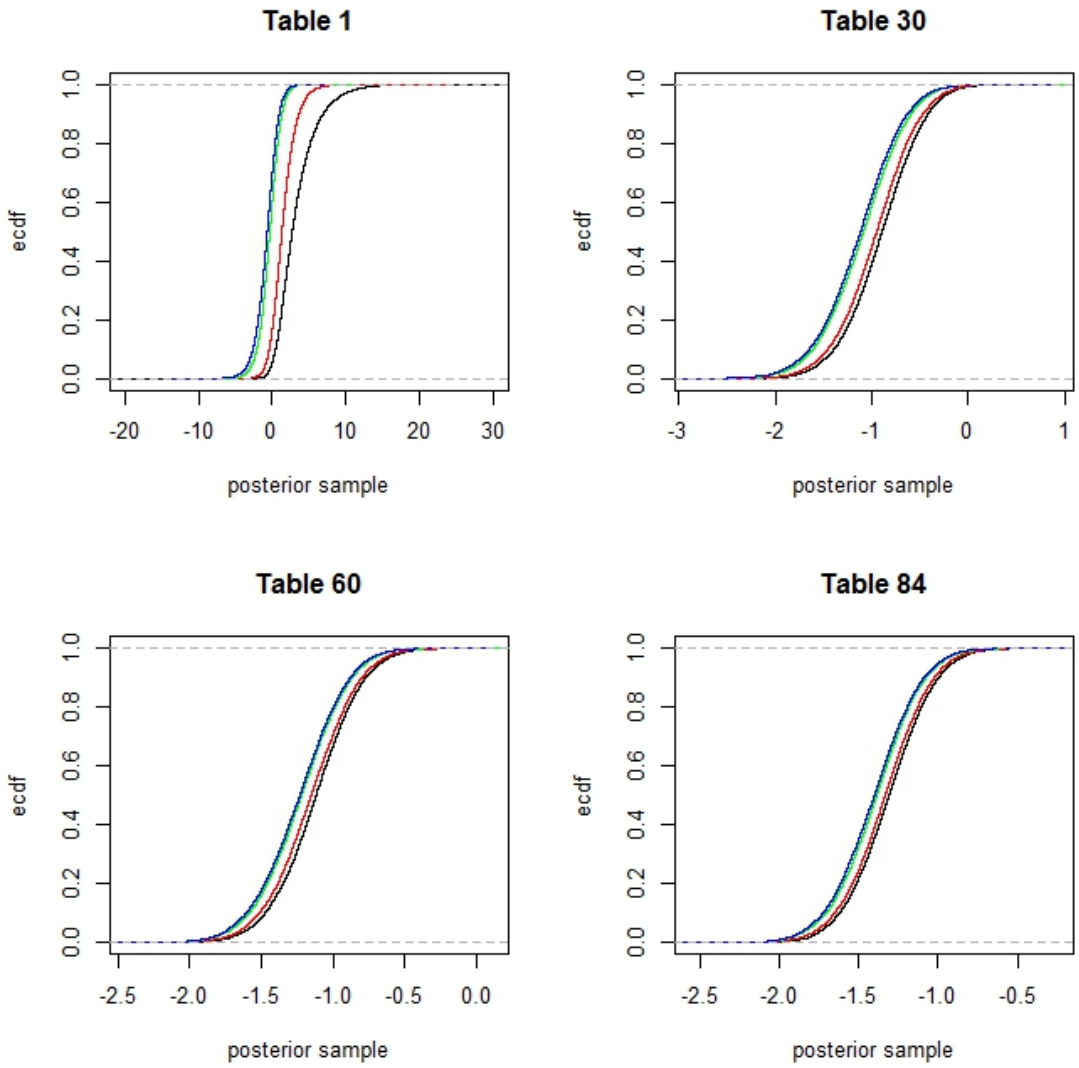


Figure A.23: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 12.

Combination 13:

$n = 100$, $\lambda_c = 0.1$, $\theta = 0$, and $r > 1$.

Table A.14: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 13.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	199.986	-1.189	201.175
	30	0.404	0.161	0.242
	60	-0.395	-0.488	0.092
	77	-0.284	-0.369	0.085
Beta	1	-0.189	-16.312	16.123
	30	-0.226	-0.458	0.231
	60	0.047	-0.090	0.138
	75	-0.026	-0.135	0.108

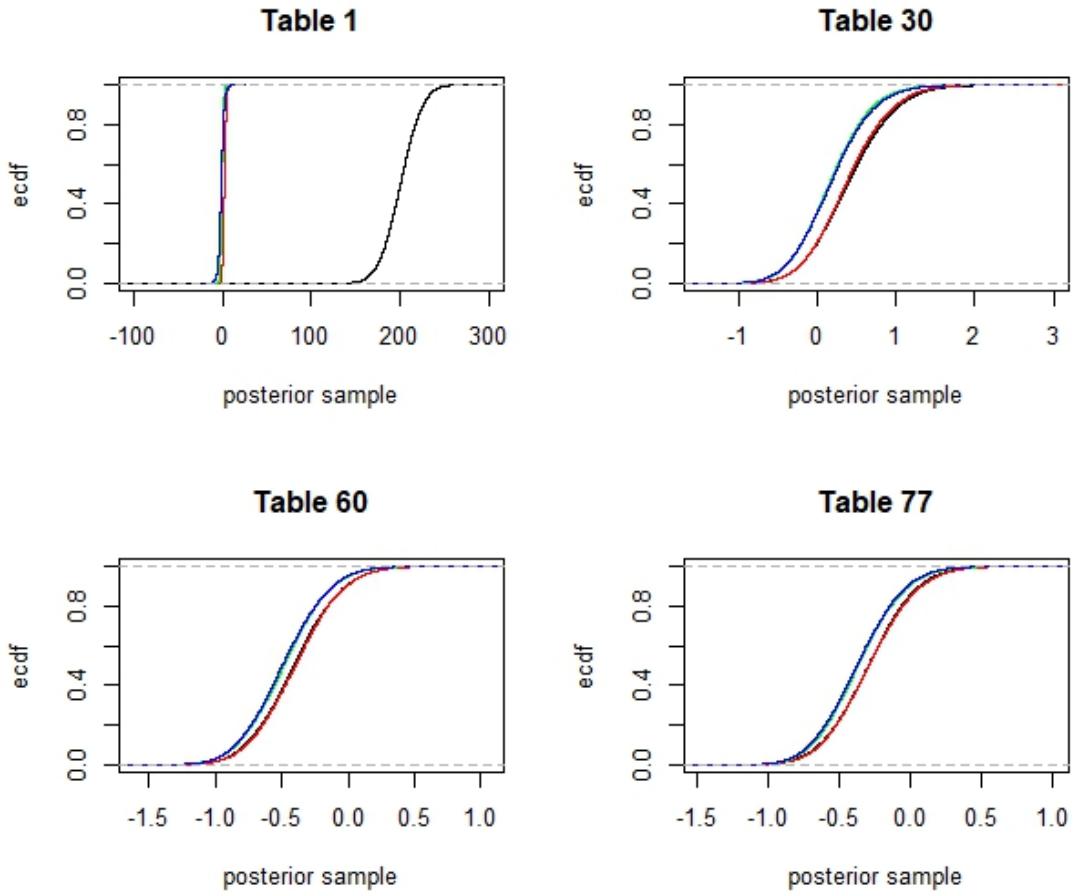


Figure A.24: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 13.

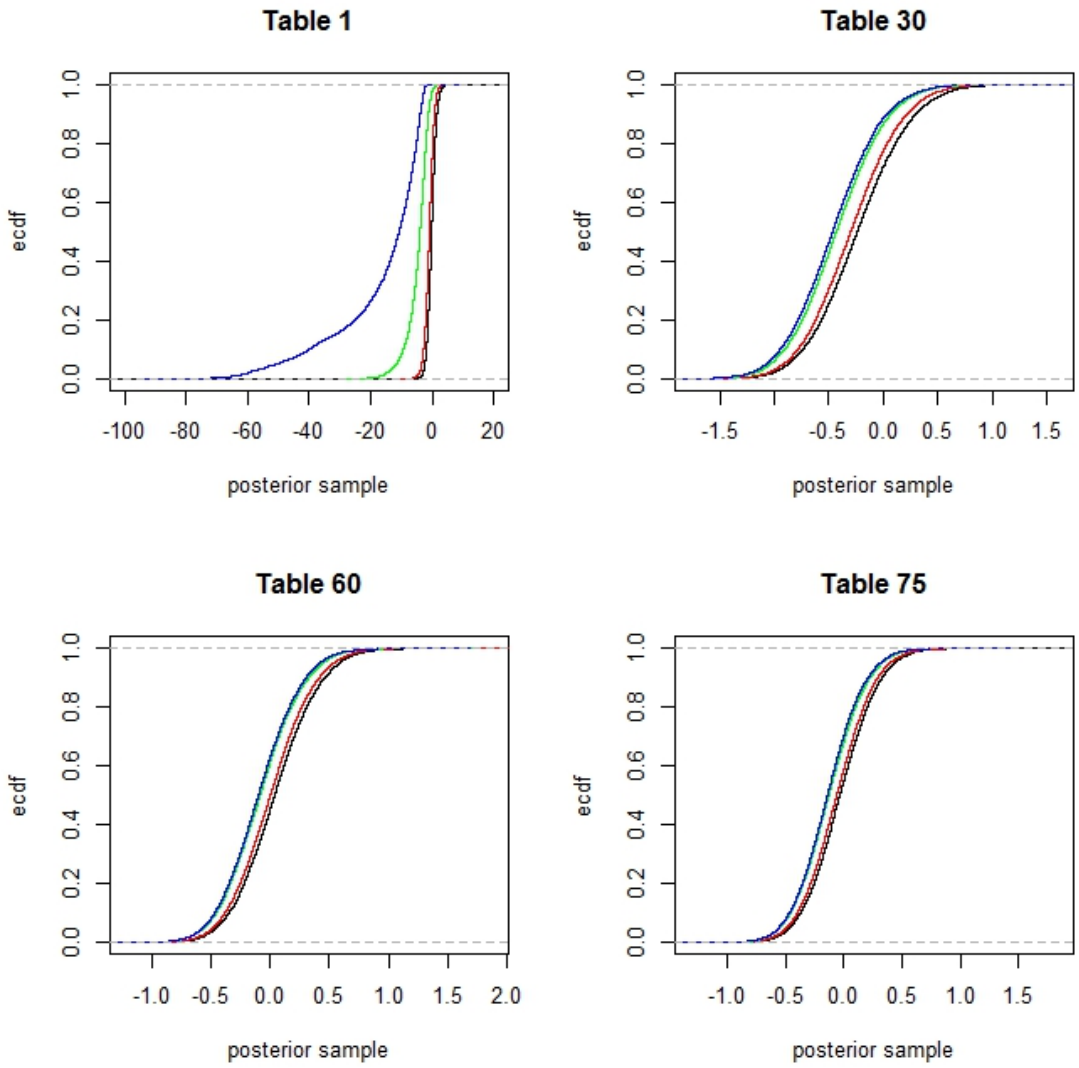


Figure A.25: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 13.

Combination 14:

$n = 40$, $\lambda_c = 0.1$, $\theta = -0.6$, and $r = 1$.

Table A.15: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 14.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	199.903	-0.015	199.919
	30	-0.950	-1.142	0.192
	60	-0.650	-0.740	0.089
	80	-0.608	-0.670	0.062
Beta	1	3.371	-0.782	4.153
	30	-0.380	-0.582	0.201
	60	-0.751	-0.862	0.111
	82	-0.616	-0.696	0.080

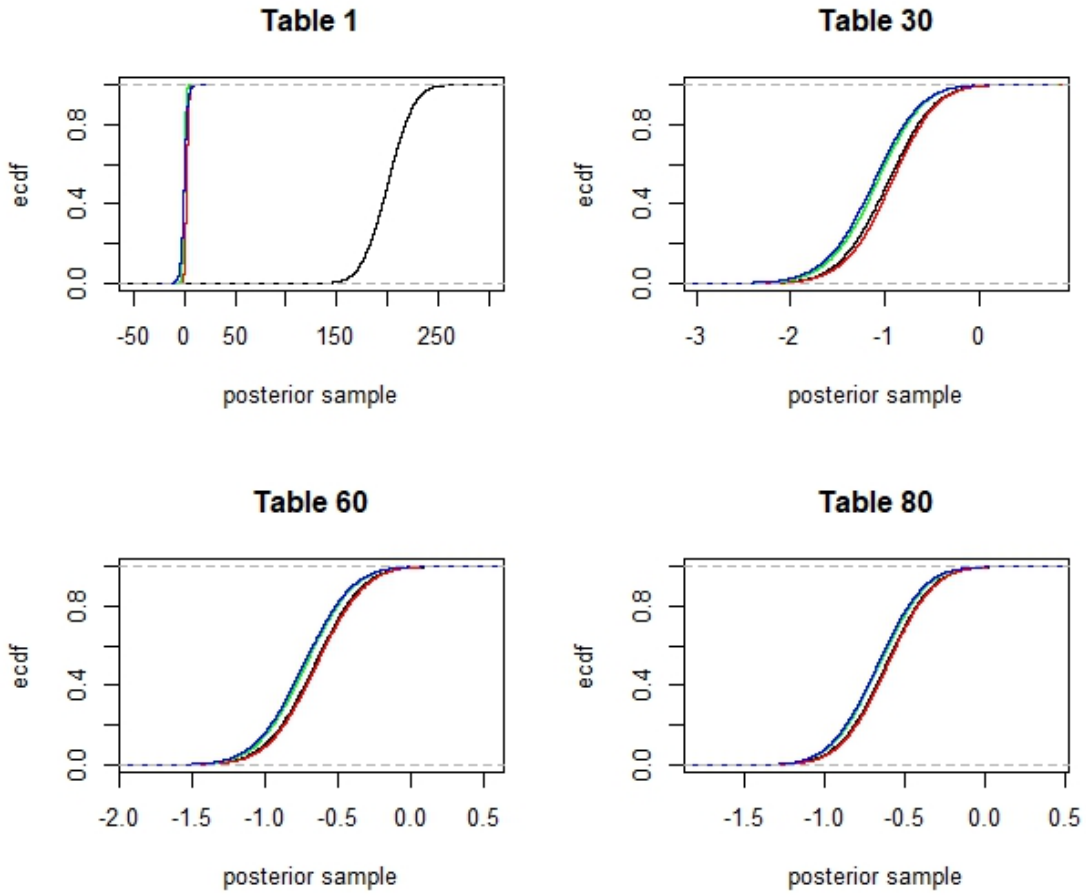


Figure A.26: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 14.

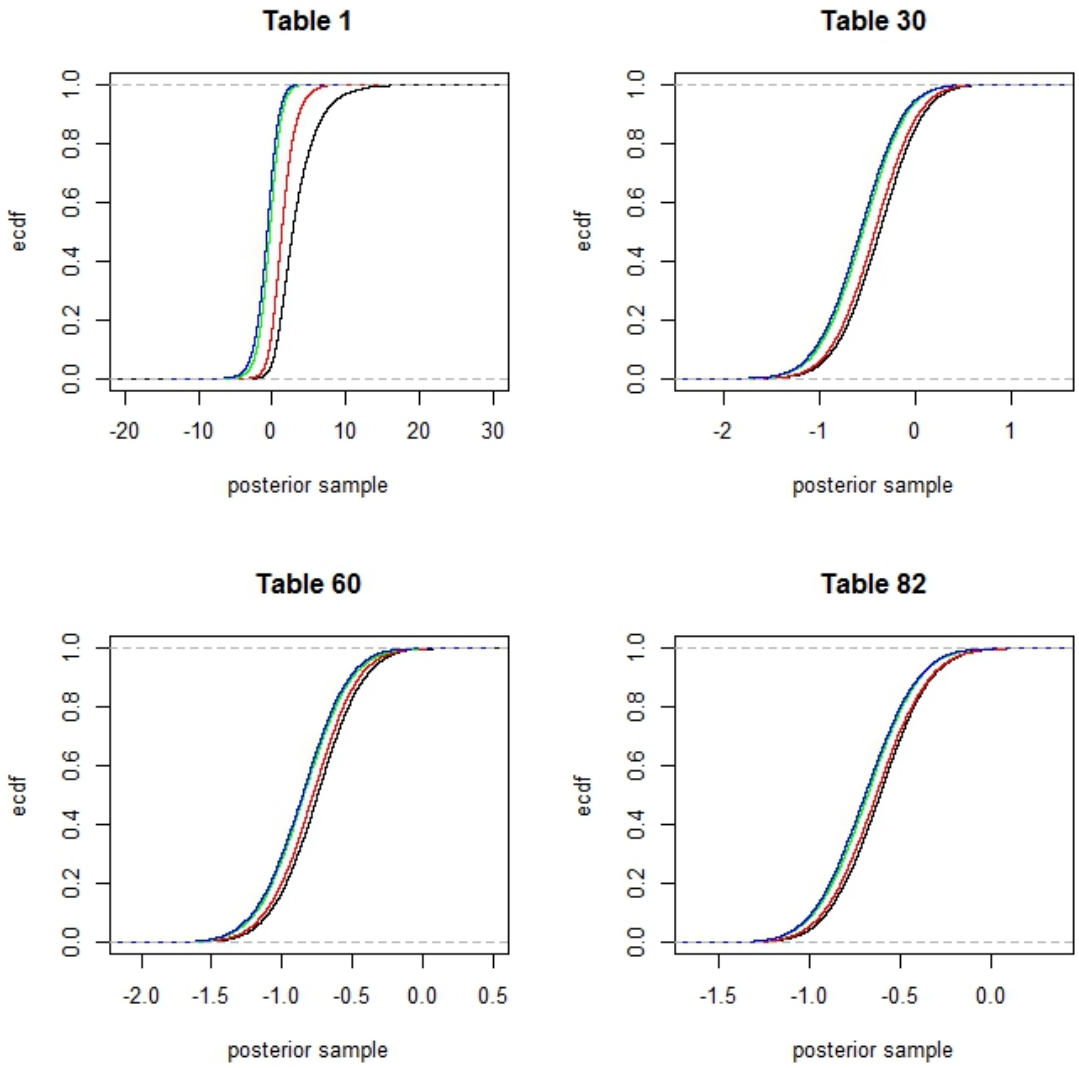


Figure A.27: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 14.

Combination 15:

$n = 100$, $\lambda_c = 0.1$, $\theta = -1.2$, and $r < 1$.

Table A.16: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 15.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	1.202	-199.972	201.175
	30	-1.400	-2.302	0.901
	60	-1.271	-1.538	0.266
	75	-1.203	-1.364	0.161
Beta	1	0.550	-10.350	10.900
	30	-0.992	-1.508	0.516
	50	-1.100	-1.369	0.268
	75	-1.100	-1.249	0.149

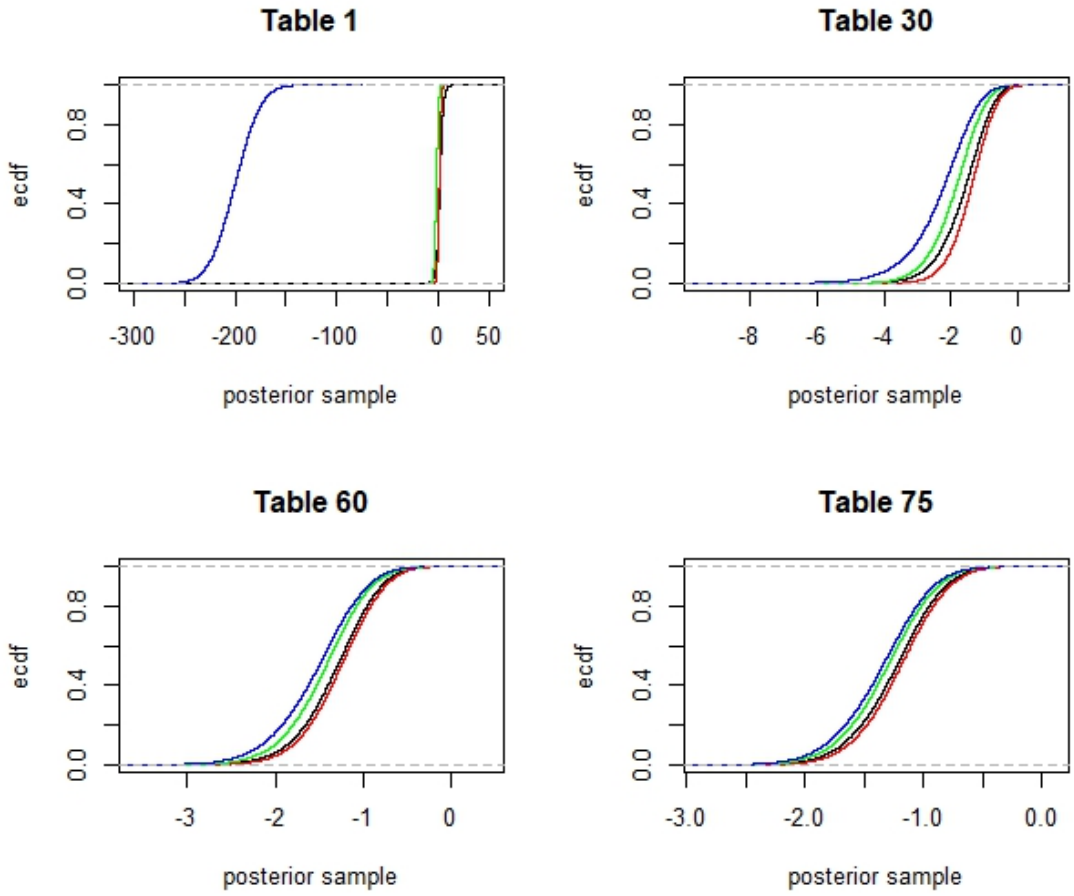


Figure A.28: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 15.

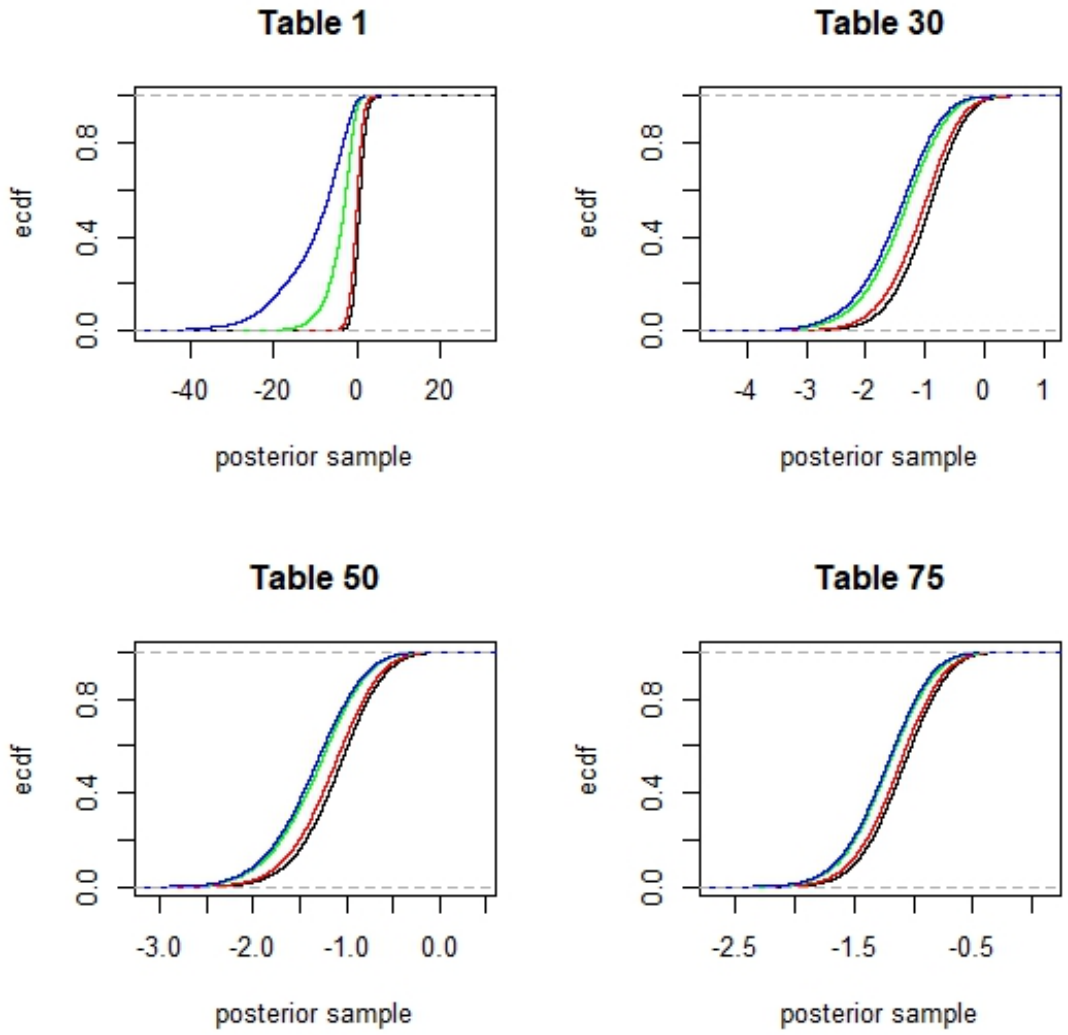


Figure A.29: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 15.

Combination 16:

$n = 100$, $\lambda_c = 0.5$, $\theta = 0$, and $r = 1$.

Table A.17: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 16.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	200.136	-0.019	200.156
	30	-0.409	-0.570	0.160
	50	0.010	-0.071	0.081
	54	0.018	-0.058	0.076
Beta	1	3.350	-0.779	4.130
	20	0.934	0.606	0.328
	40	0.497	0.344	0.152
	42	0.346	0.204	0.142

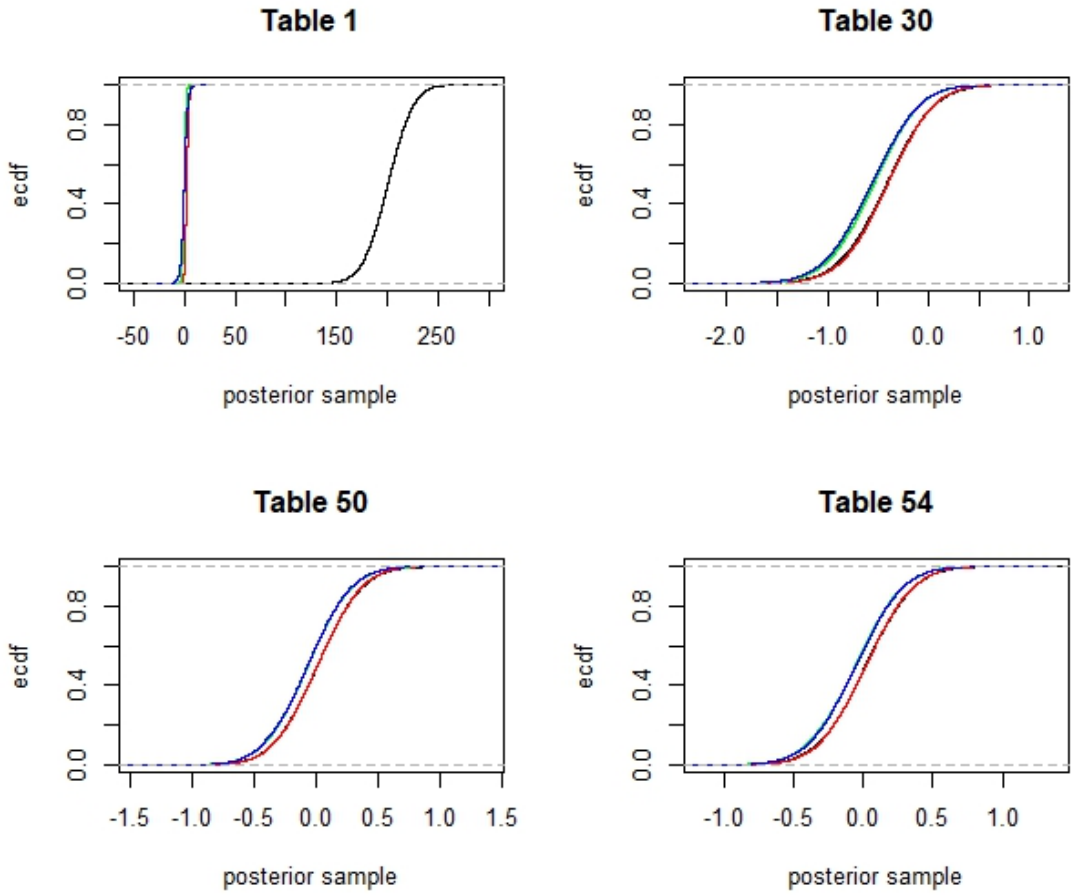


Figure A.30: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 16.

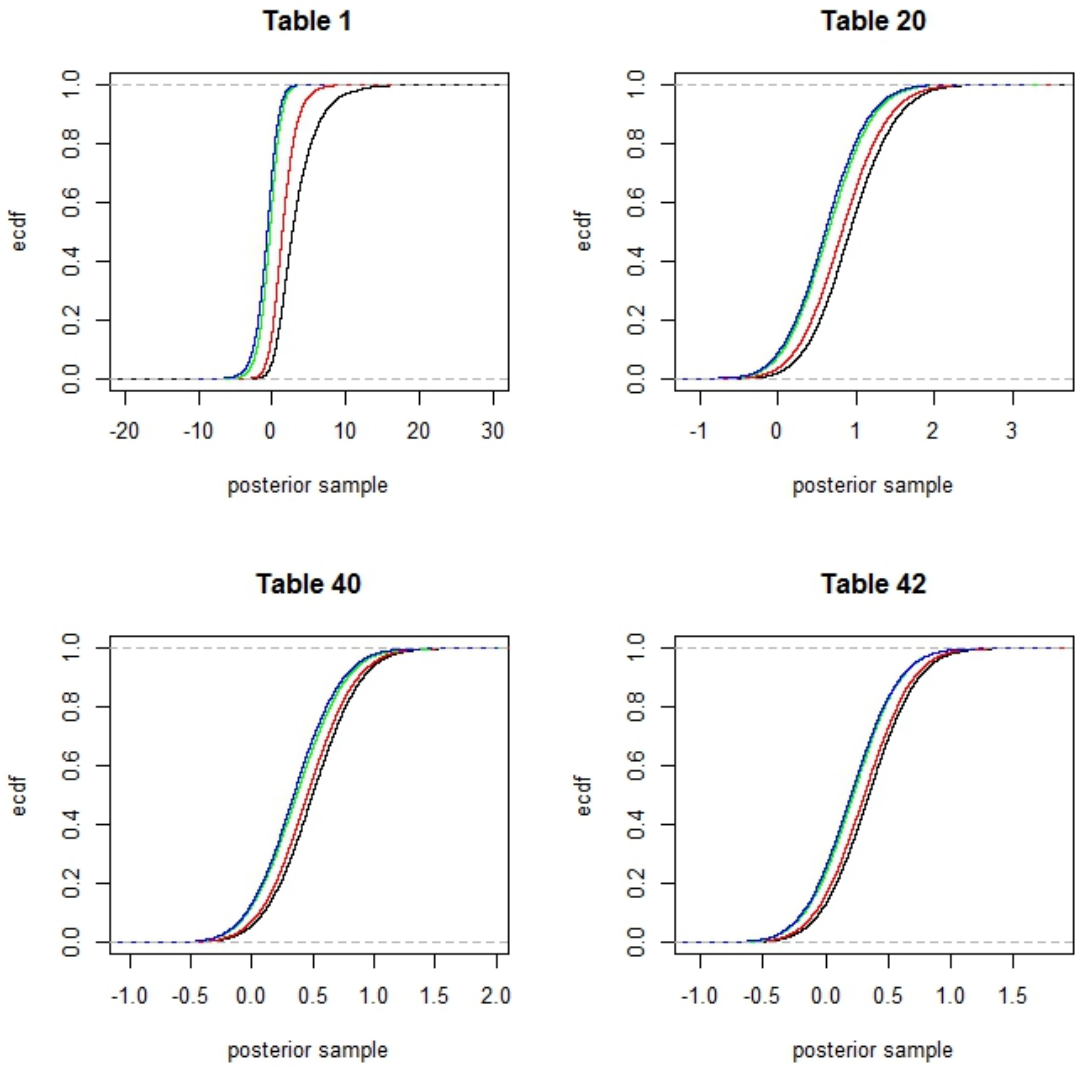


Figure A.31: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 16.

Combination 17:

$n = 40$, $\lambda_c = 0.5$, $\theta = -0.6$, and $r < 1$.

Table A.18: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 17.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	1.187	-200.075	201.262
	30	-1.276	-2.141	0.864
	40	-0.860	-1.163	0.302
	44	-0.841	-1.099	0.257
Beta	1	0.633	-9.234	9.867
	20	-0.864	-1.611	0.746
	30	-0.783	-1.196	0.412
	44	-0.655	-0.876	0.220

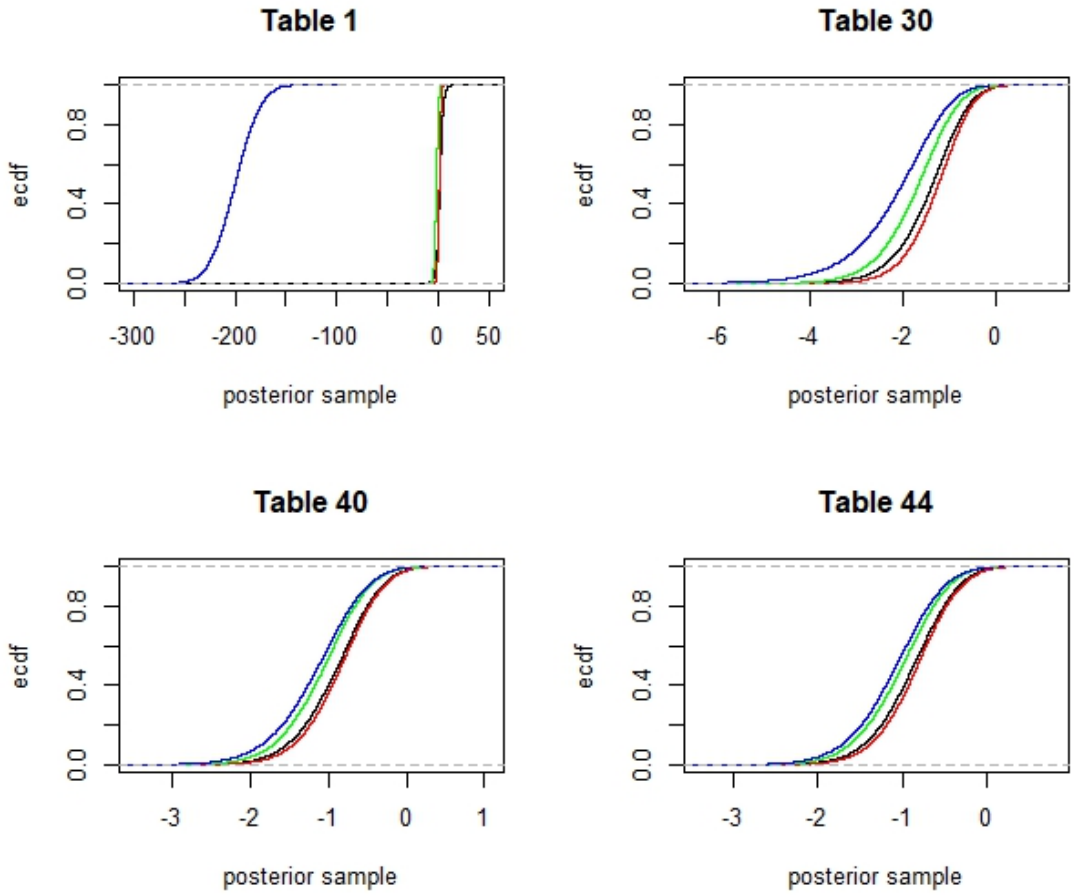


Figure A.32: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 17.

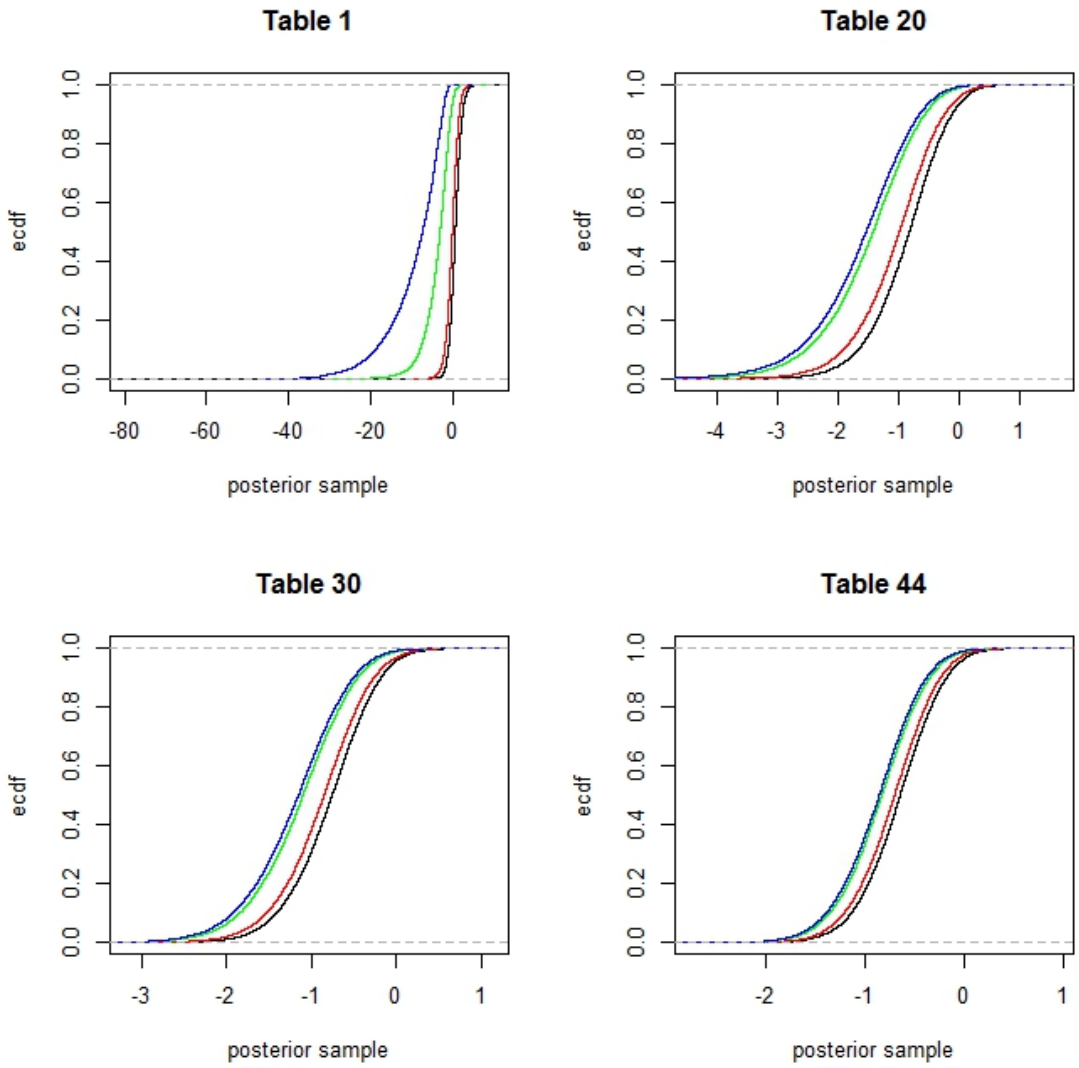


Figure A.33: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 17.

Combination 18:

$n = 100$, $\lambda_c = 0.5$, $\theta = -1.2$, and $r > 1$.

Table A.19: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 18.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	199.638	-1.186	200.824
	15	-0.392	-0.746	0.354
	20	-0.818	-1.091	0.273
	26	-1.123	-1.328	0.204
Beta	1	-0.195	-12.889	12.694
	15	-1.281	-1.656	0.374
	20	-1.299	-1.601	0.301
	25	-1.418	-1.66	0.245

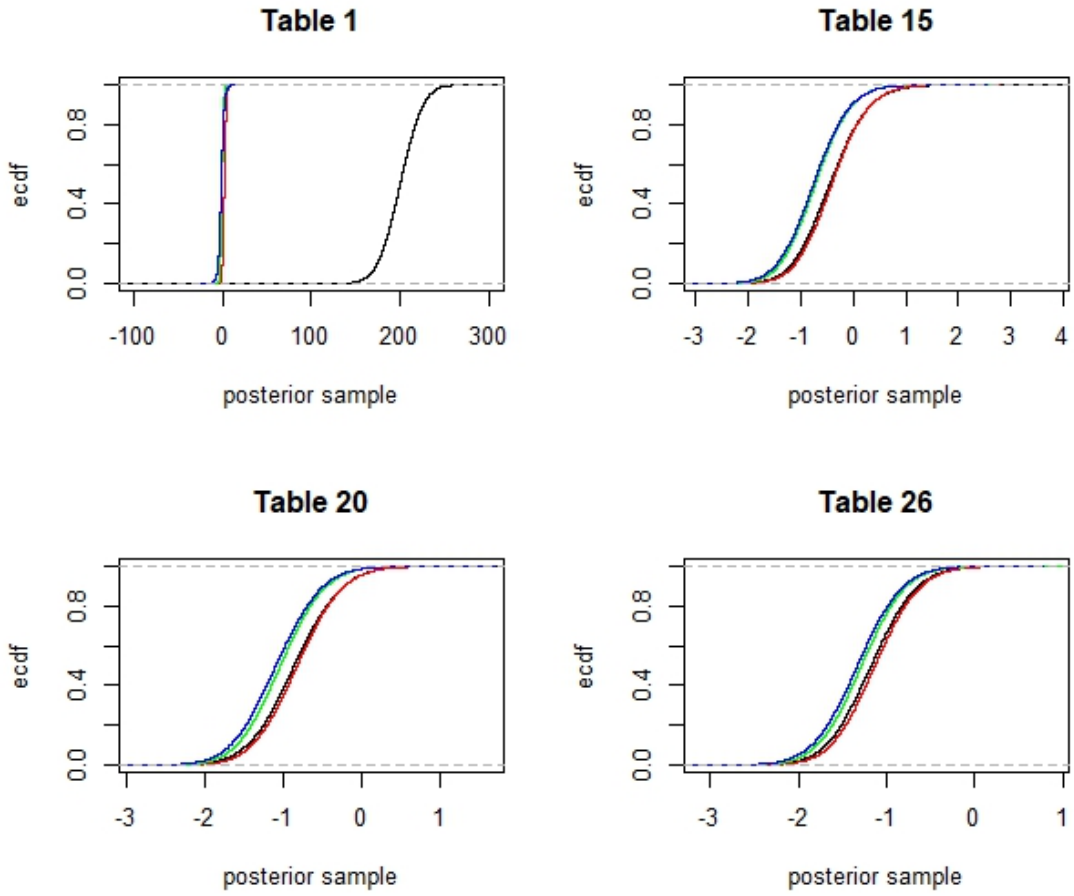


Figure A.34: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 18.

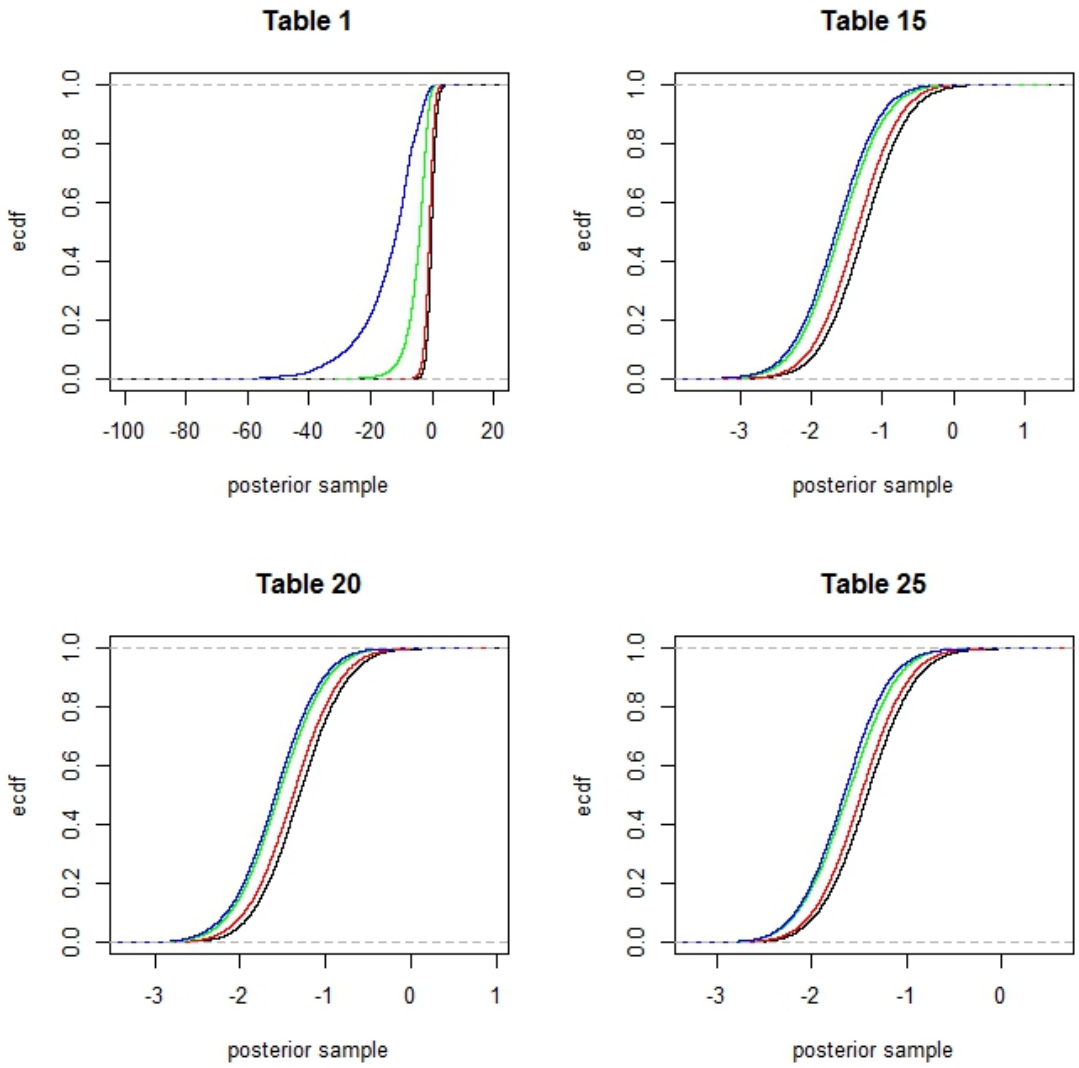


Figure A.35: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 18.

Combination 19:

$n = 200$, $\lambda_c = 0.001$, $\theta = 0$, and $r > 1$.

Table A.20: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 19.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	199.757	-0.933	200.691
	50	0.222	0.104	0.118
	100	0.126	0.067	0.058
	193	-0.030	-0.057	0.027
Beta	1	3.053	-1.167	4.220
	50	0.221	0.060	0.161
	100	0.128	0.051	0.076
	193	-0.028	-0.067	0.038

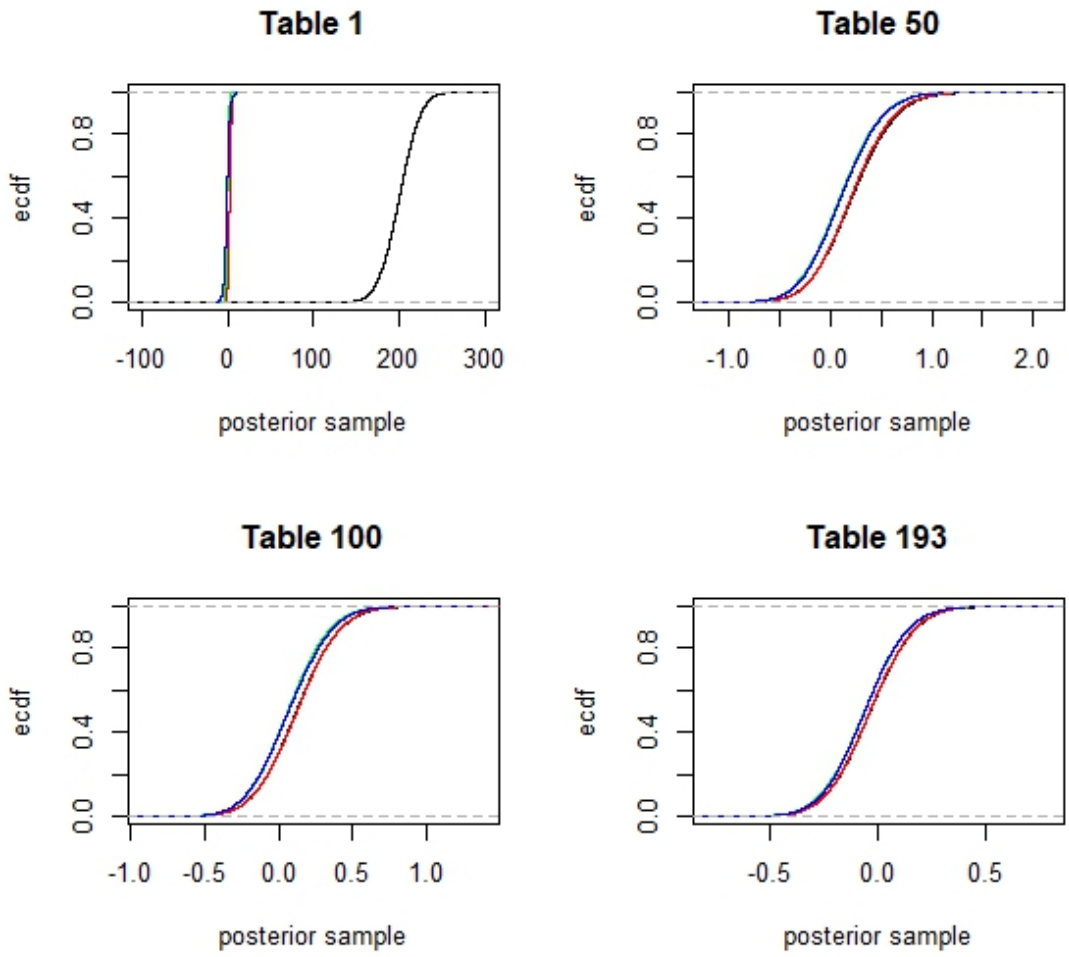


Figure A.36: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 19.

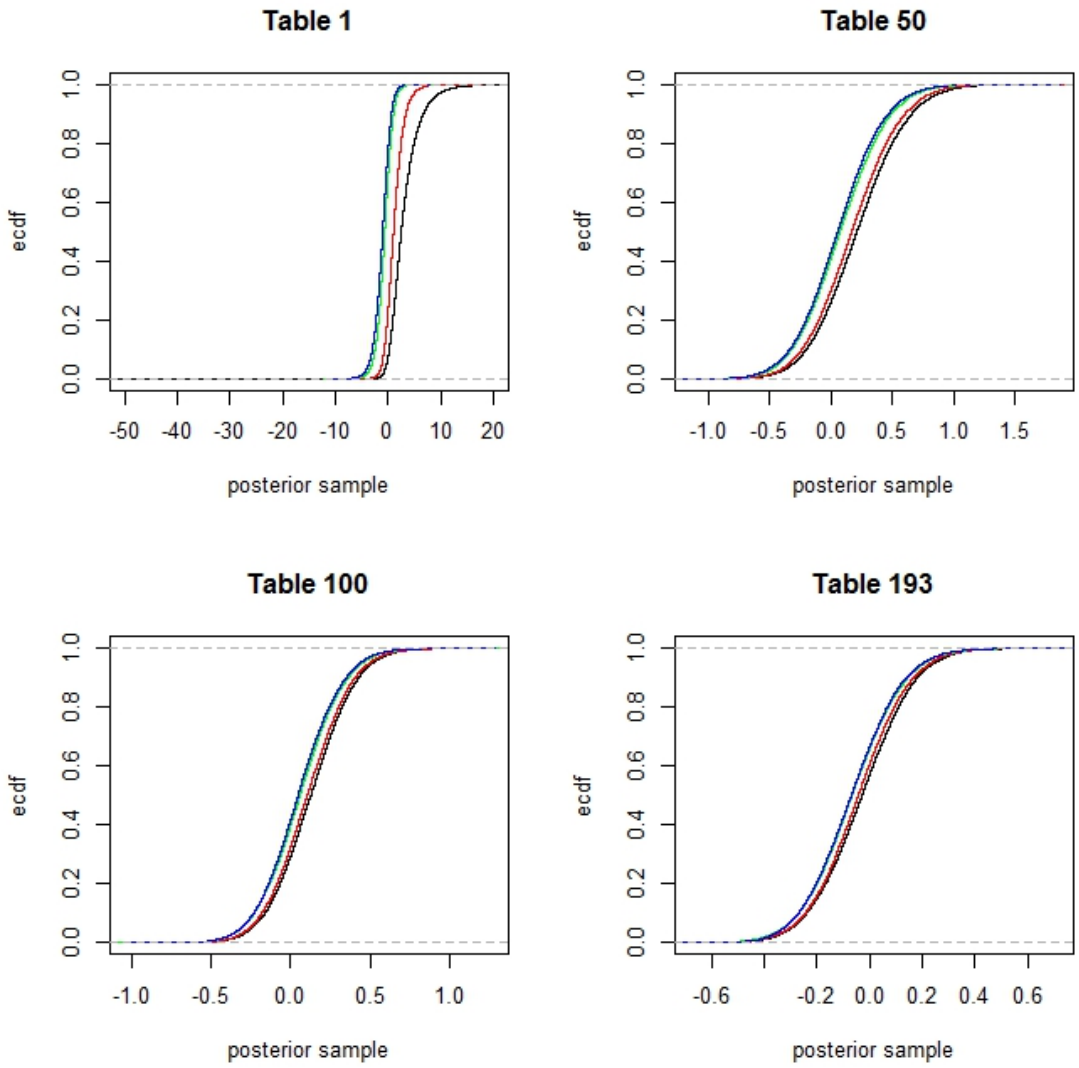


Figure A.37: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 19.

Combination 20:

$n = 200$, $\lambda_c = 0.001$, $\theta = -0.6$, and $r = 1$.

Table A.21: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 20.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	0.020	-199.983	200.003
	50	-0.528	-0.627	0.099
	100	-0.504	-0.551	0.047
	194	-0.665	-0.692	0.026
Beta	1	0.166	-10.688	10.854
	50	-0.782	-0.917	0.134
	100	-0.561	-0.617	0.056
	192	-0.595	-0.627	0.031

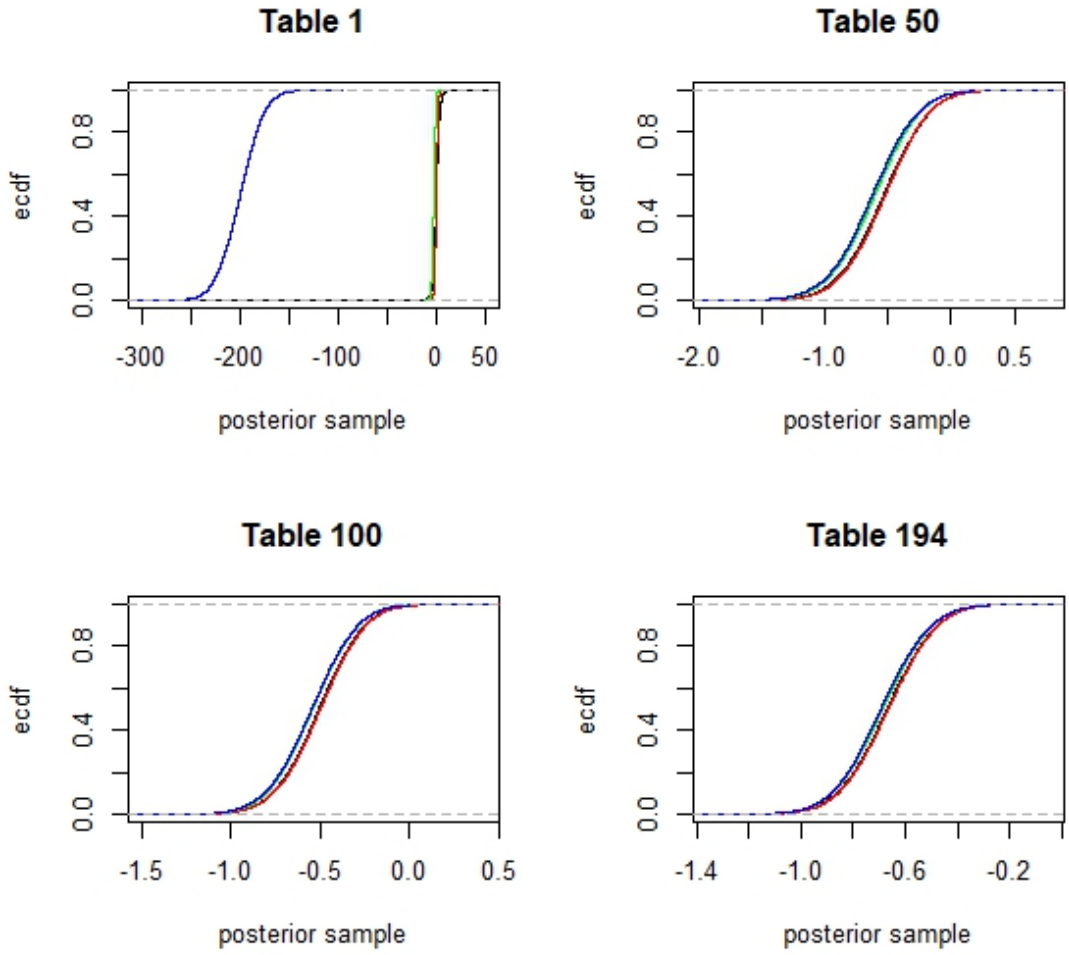


Figure A.38: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 20.

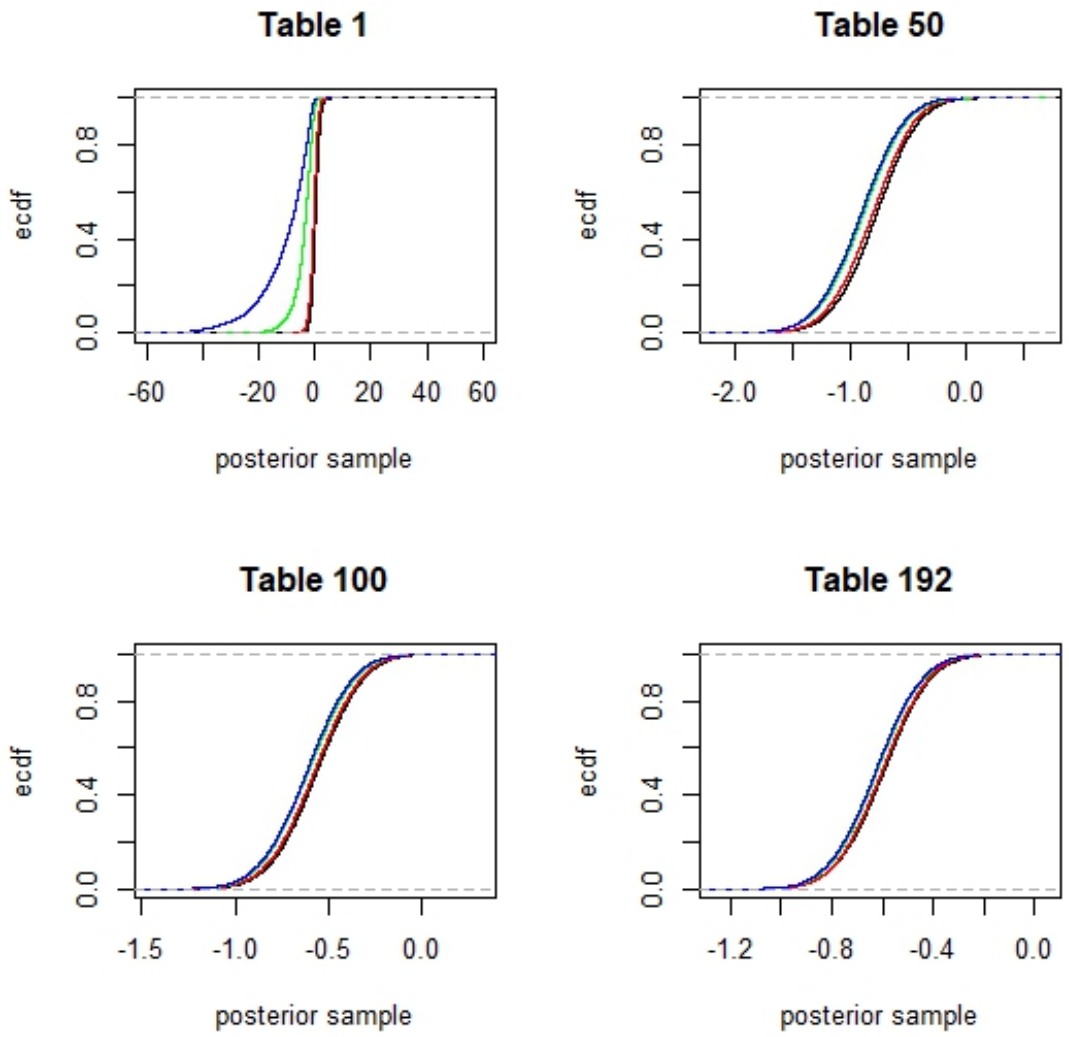


Figure A.39: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 20.

Combination 21:

$n = 200$, $\lambda_c = 0.001$, $\theta = -1.2$, and $r < 1$.

Table A.22: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 21.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	1.011	-200.004	201.015
	50	-1.426	-1.828	0.401
	100	-1.419	-1.568	0.149
	190	-1.212	-1.260	0.047
Beta	1	0.646	-12.525	13.171
	50	-1.310	-1.814	0.504
	100	-1.461	-1.685	0.224
	193	-1.250	-1.309	0.058

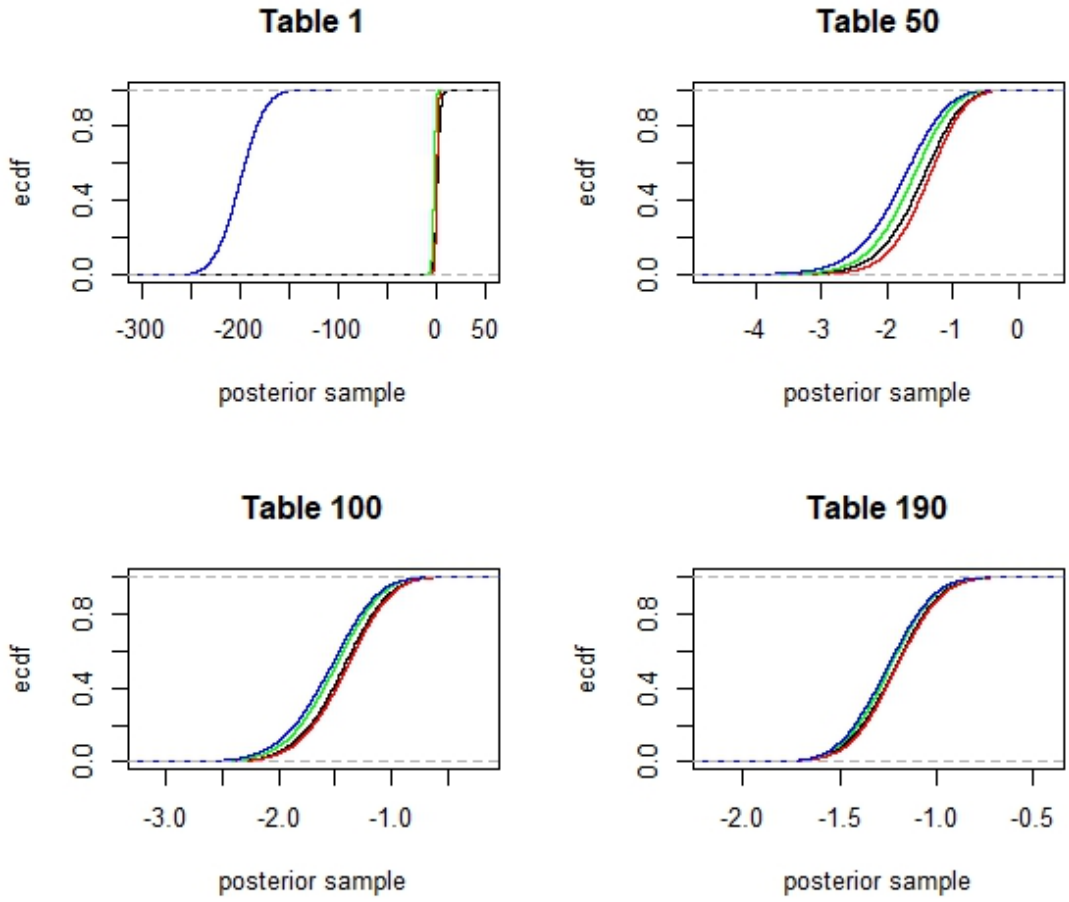


Figure A.40: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 21.

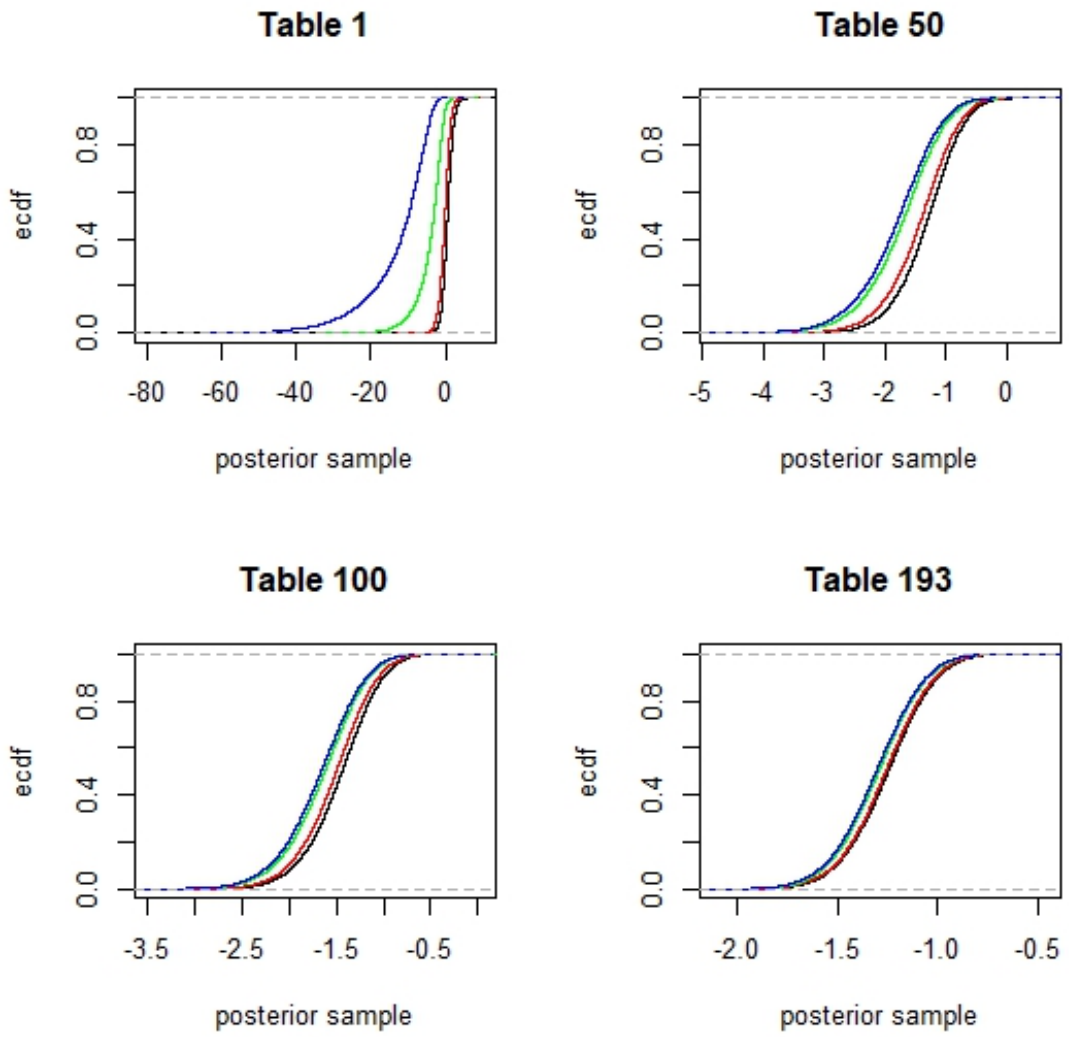


Figure A.41: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 21.

Combination 22:

$n = 200$, $\lambda_c = 0.1$, $\theta = 0$, and $r = 1$.

Table A.23: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 22.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	-0.002	-200.277	200.275
	50	0.121	0.035	0.086
	100	0.191	0.148	0.043
	158	-0.003	-0.031	0.028
Beta	1	0.193	-11.714	11.907
	50	0.148	0.031	0.117
	100	0.029	-0.029	0.058
	153	-0.097	-0.136	0.038

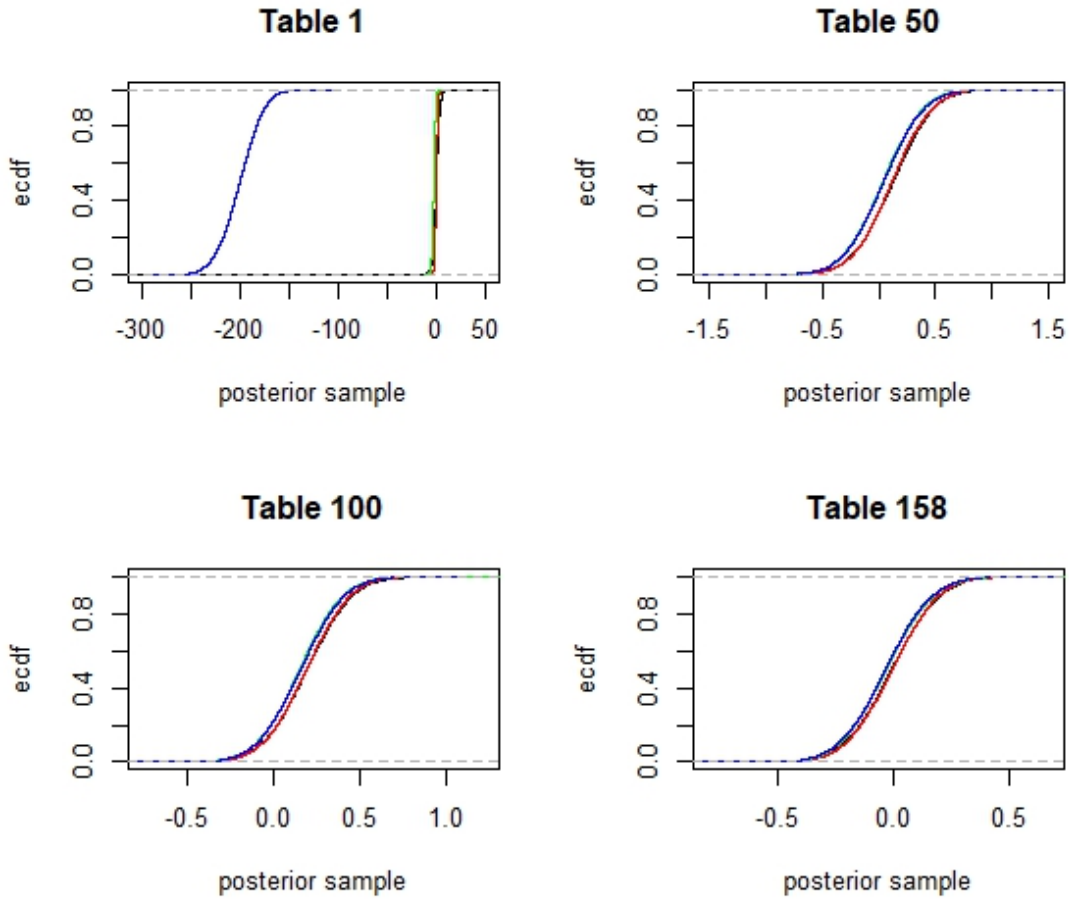


Figure A.42: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 22.

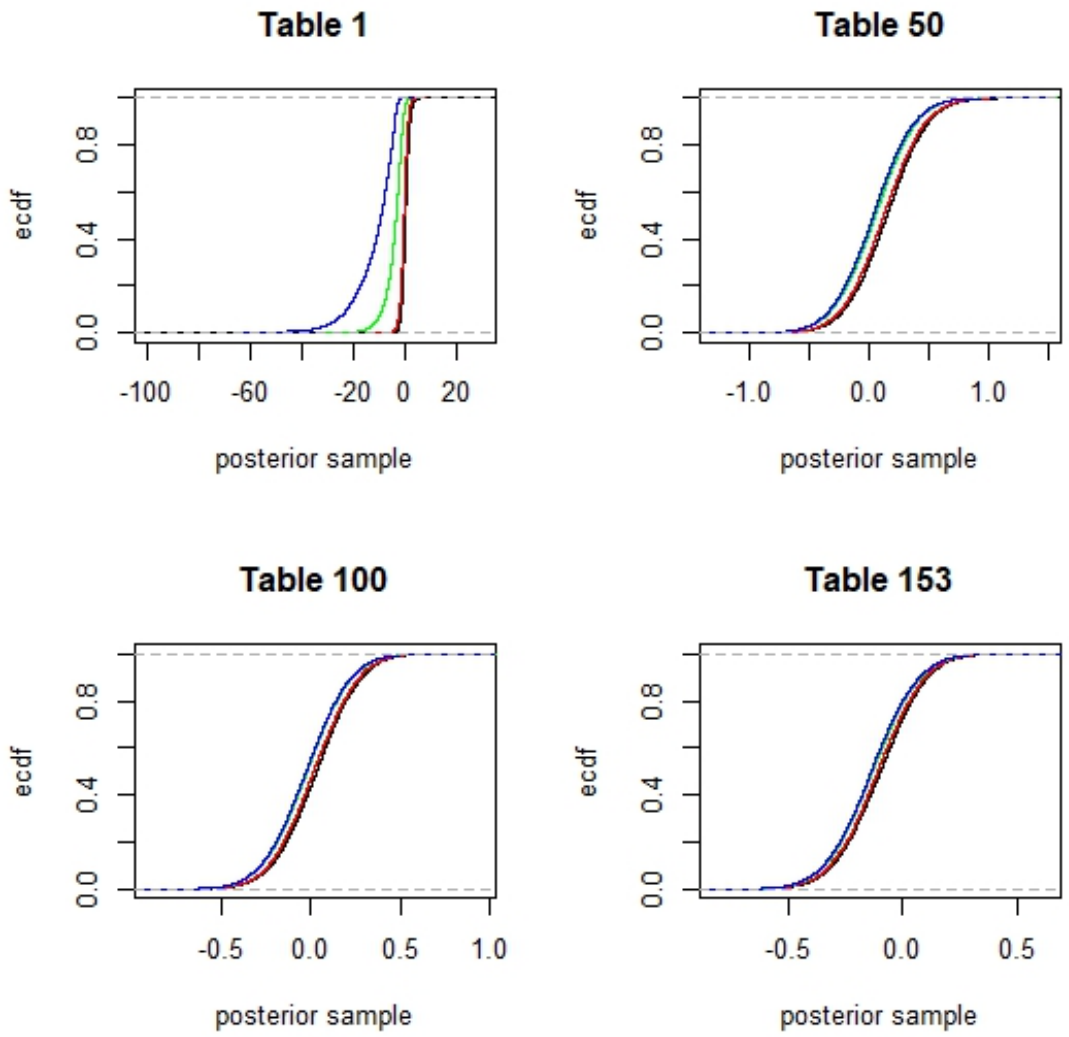


Figure A.43: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 22.

Combination 23:

$n = 200$, $\lambda_c = 0.1$, $\theta = -0.6$, and $r < 1$.

Table A.24: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 23.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	1.014	-199.730	200.744
	50	-0.798	-1.017	0.219
	100	-0.631	-0.708	0.076
	158	-0.645	-0.693	0.047
Beta	1	0.467	-11.399	11.867
	50	-0.440	-0.623	0.182
	100	-0.328	-0.408	0.080
	155	-0.443	-0.496	0.053

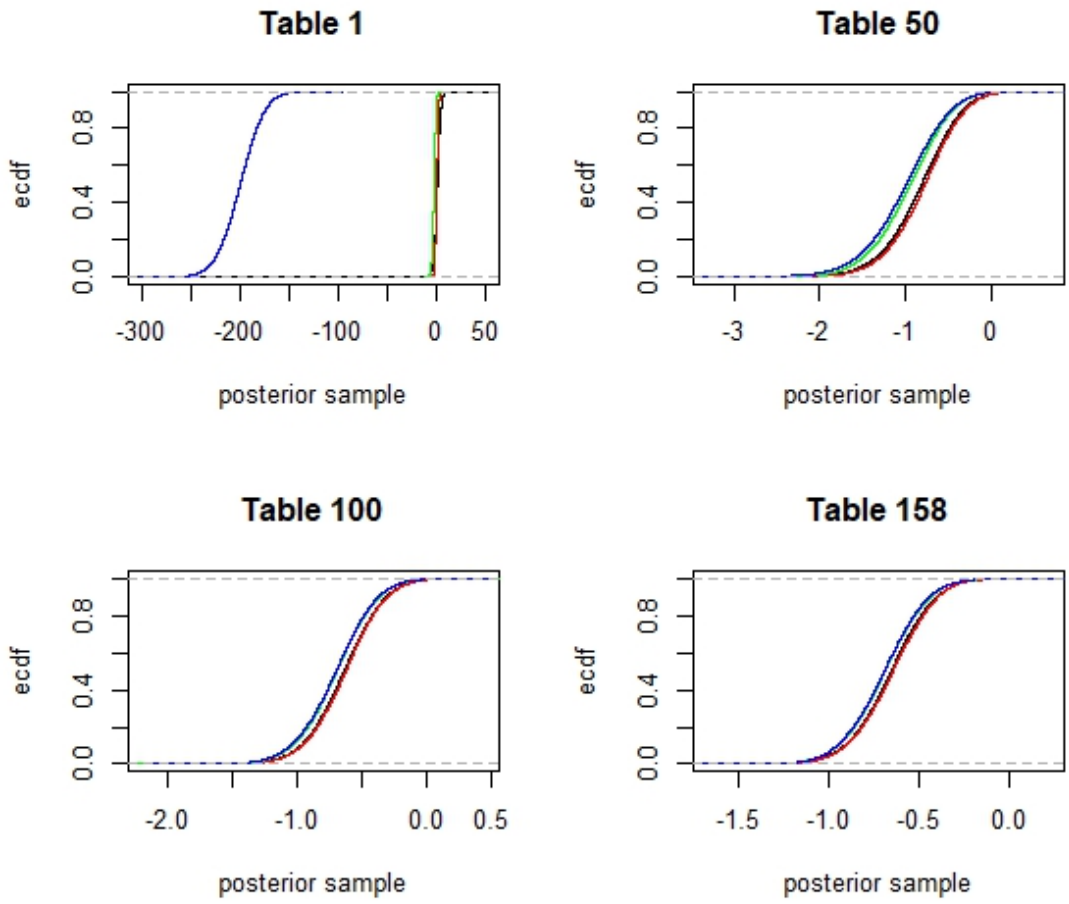


Figure A.44: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 23.

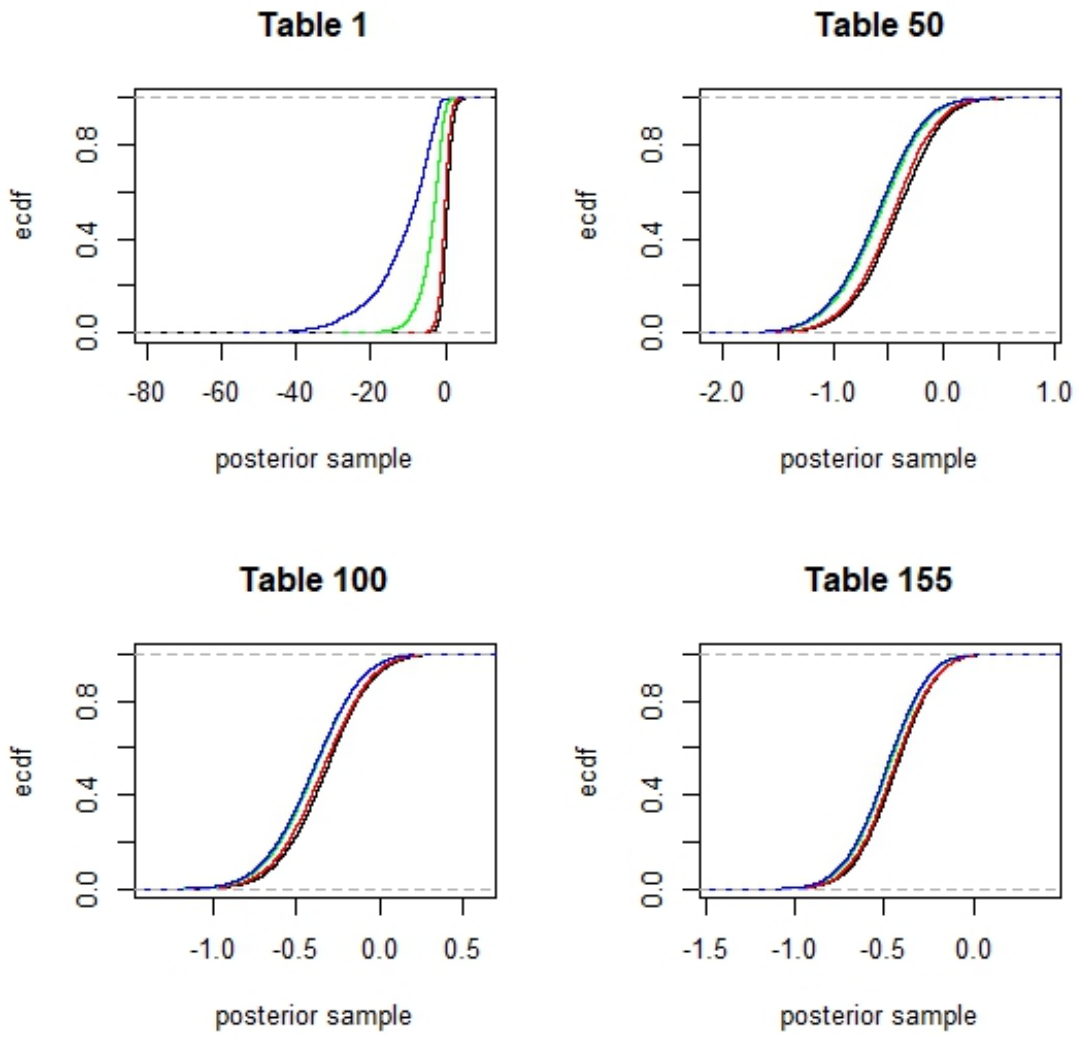


Figure A.45: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 23.

Combination 24:

$n = 200$, $\lambda_c = 0.1$, $\theta = -1.2$, and $r > 1$.

Table A.25: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 24.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	200.238	-1.075	201.314
	50	-0.877	-0.974	0.097
	100	-1.010	-1.065	0.055
	111	-1.074	-1.128	0.053
Beta	1	3.062	-1.239	4.301
	50	-1.625	-1.745	0.119
	100	-1.299	-1.368	0.069
	123	-1.253	-1.309	0.055

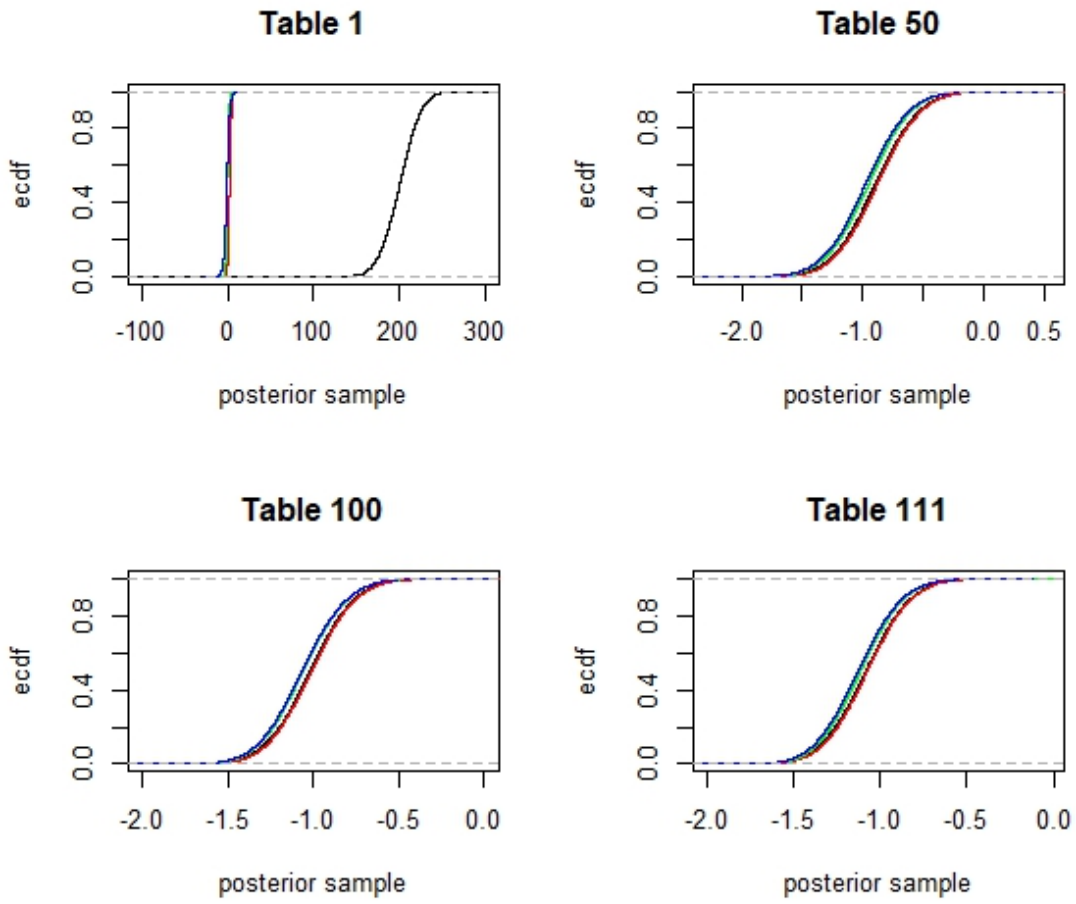


Figure A.46: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 24.

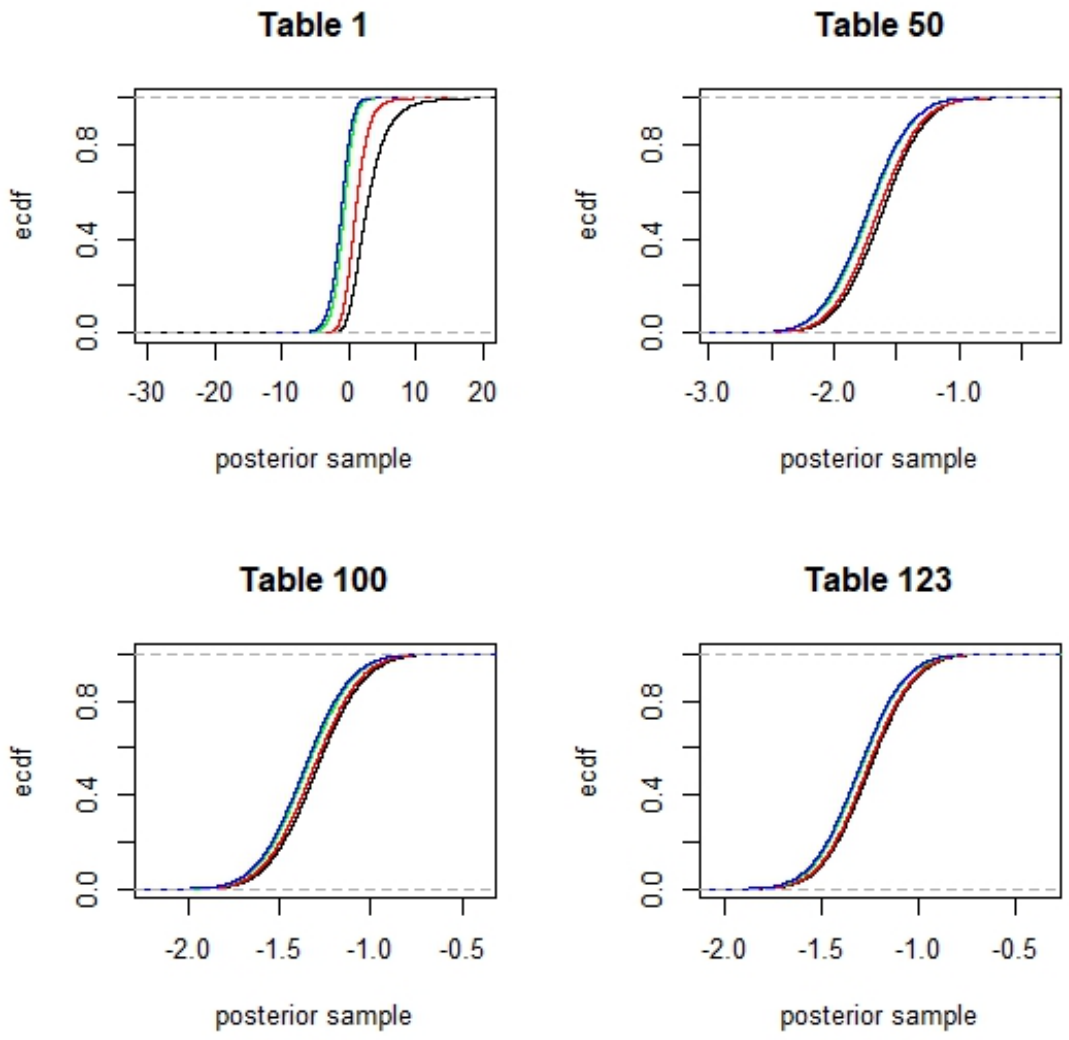


Figure A.47: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 24.

Combination 25:

$n = 200$, $\lambda_c = 0.5$, $\theta = 0$, and $r < 1$.

Table A.26: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 25.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	1.037	-200.147	201.185
	50	-0.568	-0.749	0.180
	80	-0.103	-0.174	0.071
	102	-0.160	-0.217	0.057
Beta	1	0.467	-14.458	14.925
	50	0.344	0.210	0.133
	80	0.191	0.106	0.085
	92	0.110	0.032	0.077

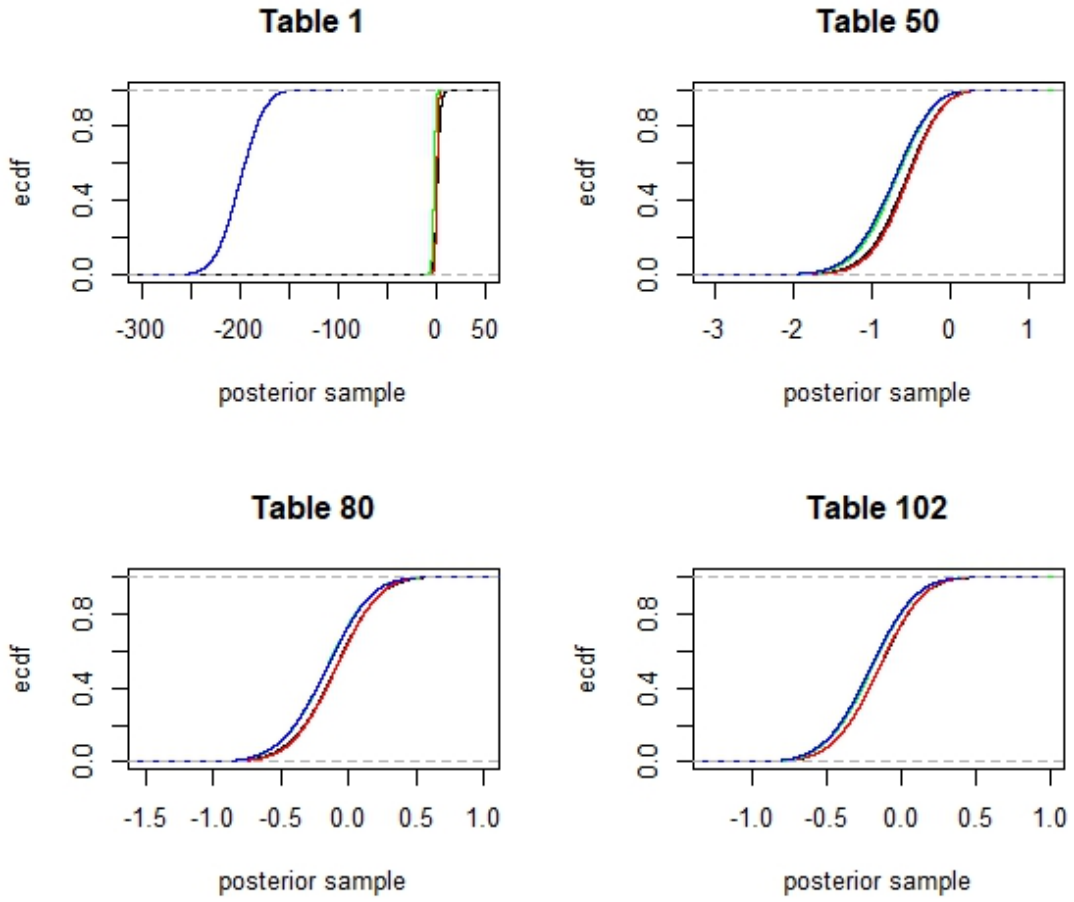


Figure A.48: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 25.

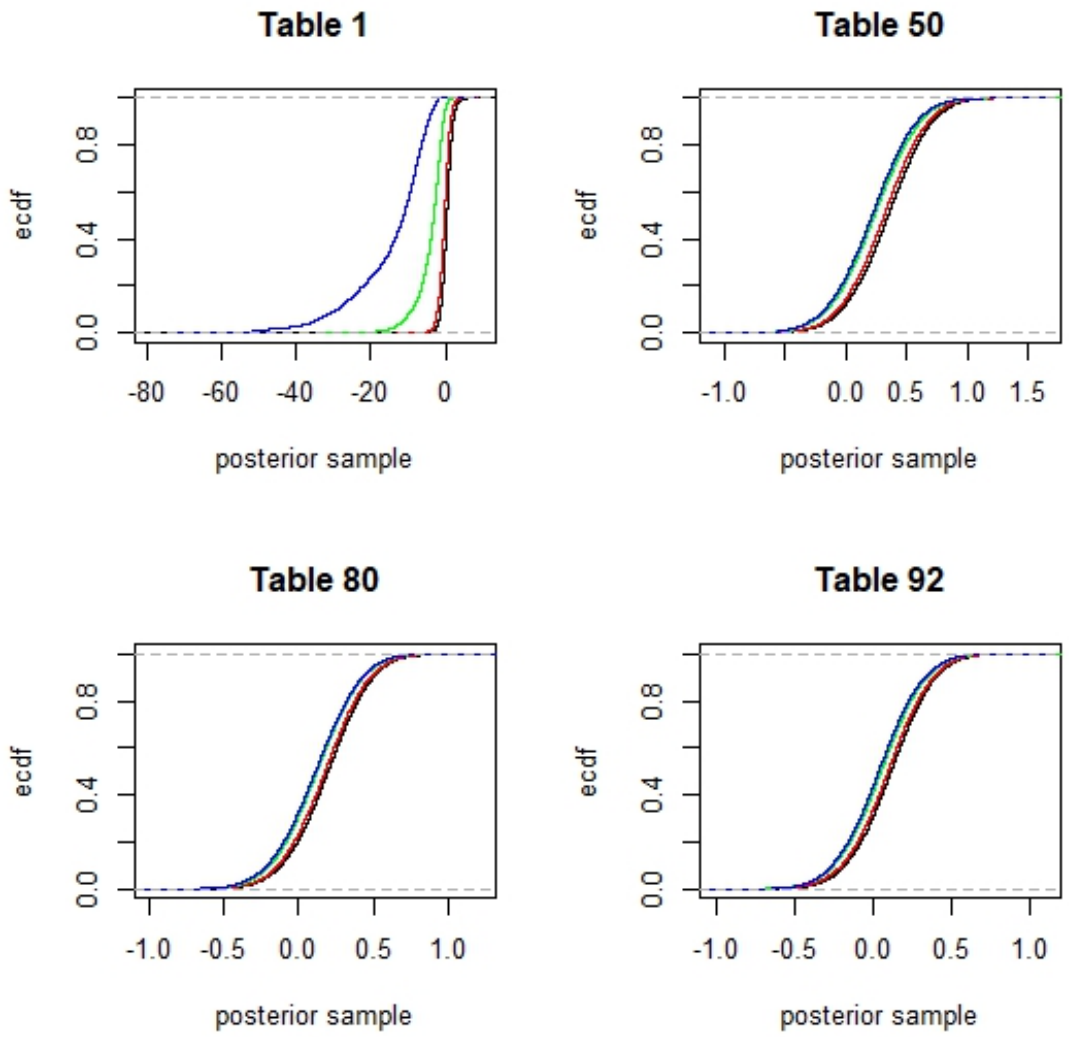


Figure A.49: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 25.

Combination 26:

$n = 200$, $\lambda_c = 0.5$, $\theta = -0.6$, and $r > 1$.

Table A.27: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 26.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	200.003	-1.128	201.131
	30	0.053	-0.119	0.173
	60	-0.399	-0.489	0.089
	71	-0.354	-0.436	0.081
Beta	1	-0.184	-9.824	9.639
	30	-0.311	-0.531	0.219
	50	-0.531	-0.659	0.128
	88	-0.603	-0.685	0.081

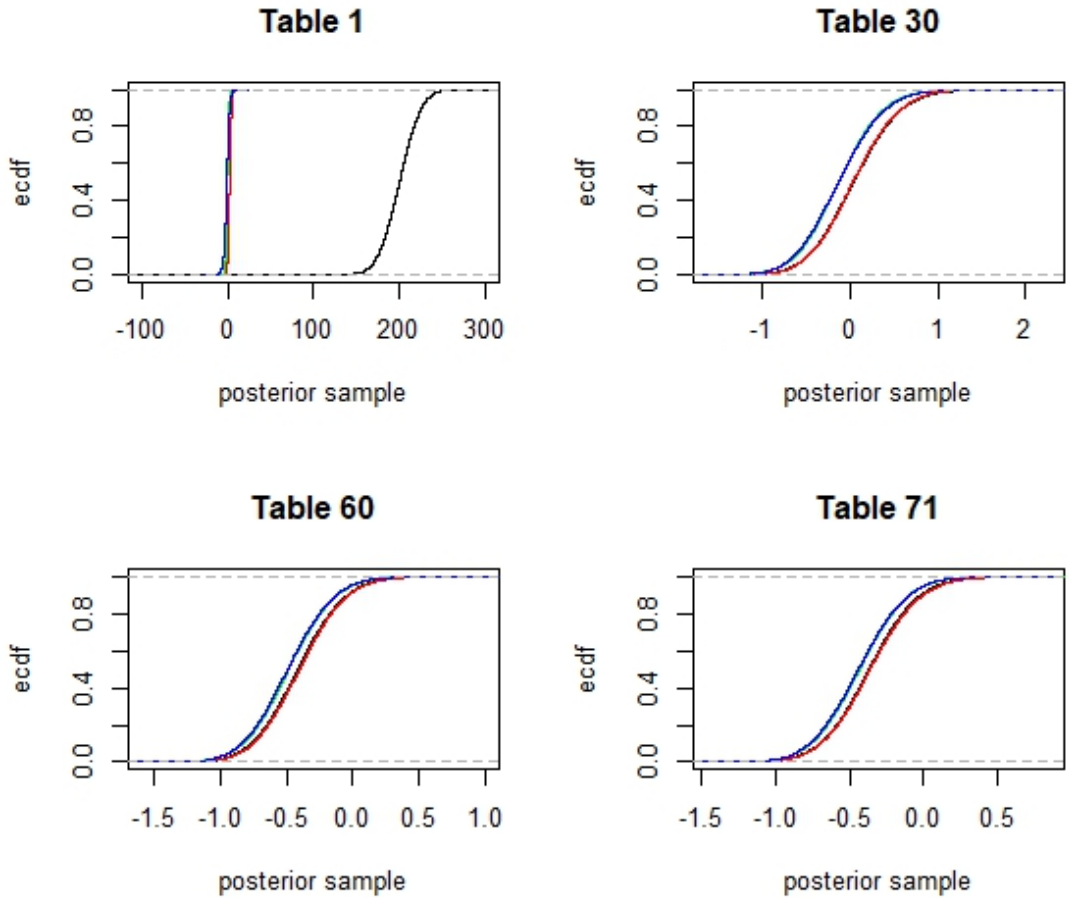


Figure A.50: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 26.

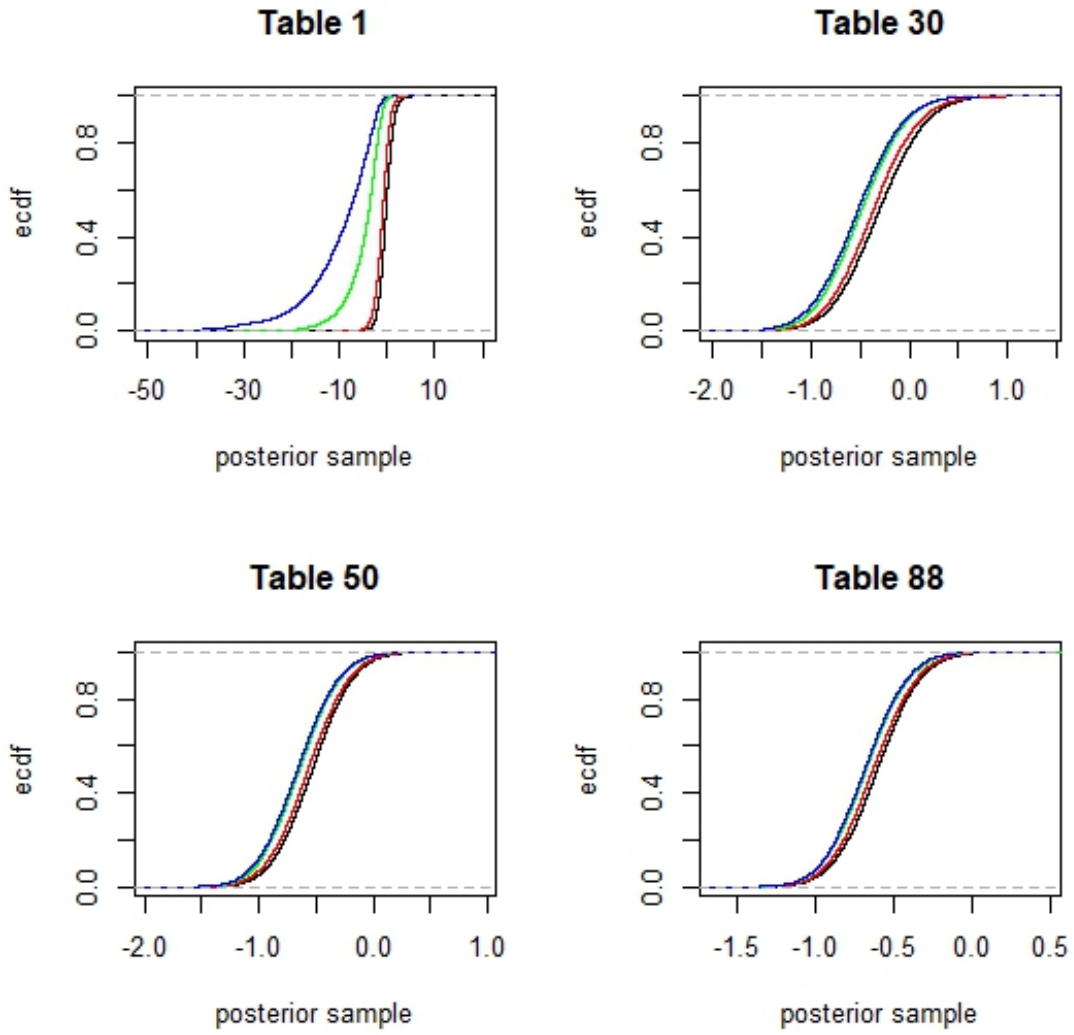


Figure A.51: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 26.

Combination 27:

$n = 200$, $\lambda_c = 0.5$, $\theta = -1.2$, and $r = 1$.

Table A.28: Lower and upper posterior expectations and the degree of imprecision of log-odds ratio of combination 27.

Prior	Table	$\overline{E}(\theta\|y)$	$\underline{E}(\theta\|y)$	$\Delta_{\theta\|y}$
Normal	1	0.008	-200.148	200.156
	20	-1.412	-1.861	0.449
	40	-1.077	-1.229	0.152
	66	-1.074	-1.162	0.088
Beta	1	0.183	-13.281	13.464
	30	-0.929	-1.182	0.253
	50	-0.879	-1.017	0.137
	78	-1.120	-1.210	0.089

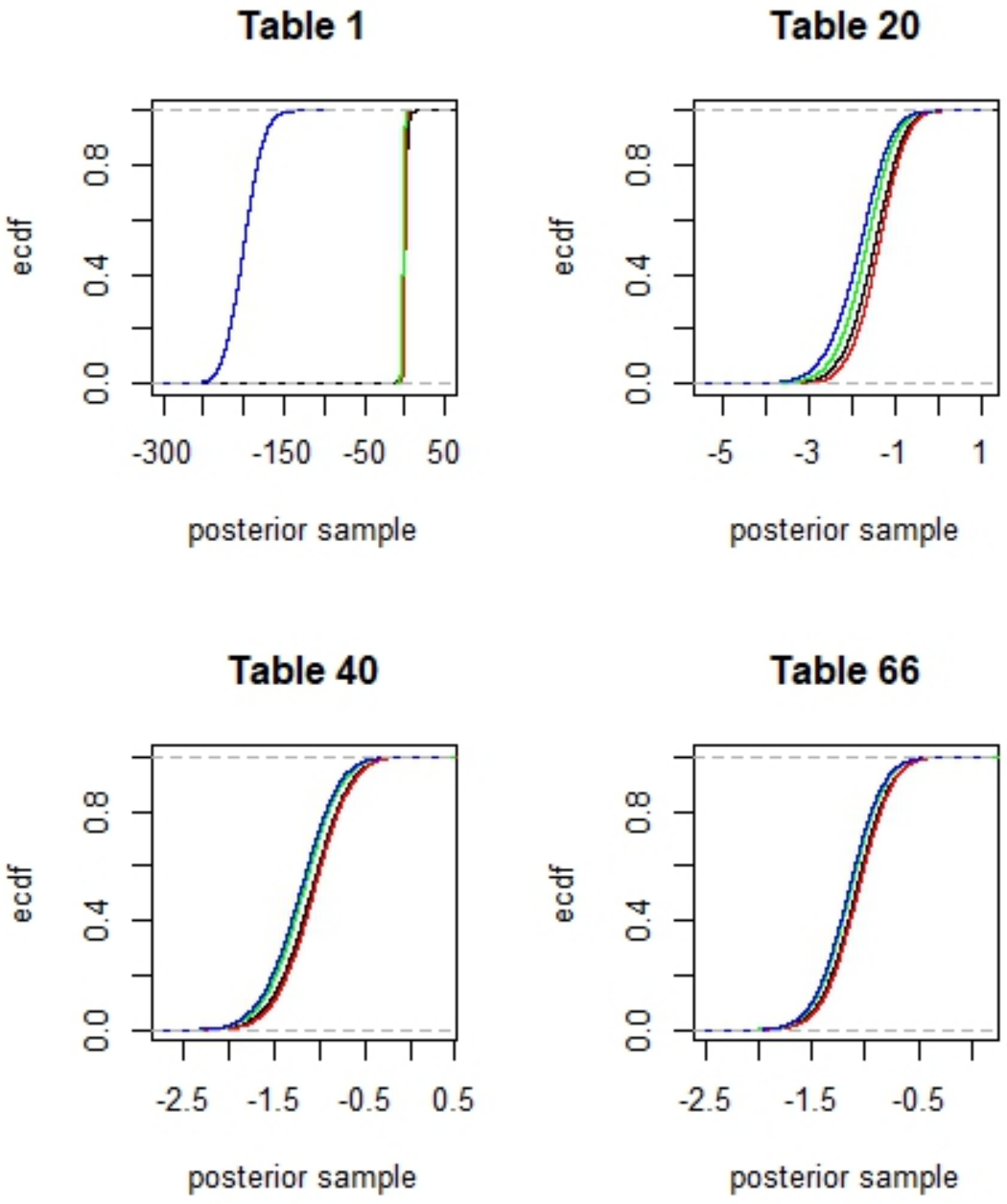


Figure A.52: Empirical CDFs of the posterior samples for number of tables considering a family of normal priors, combination 27.

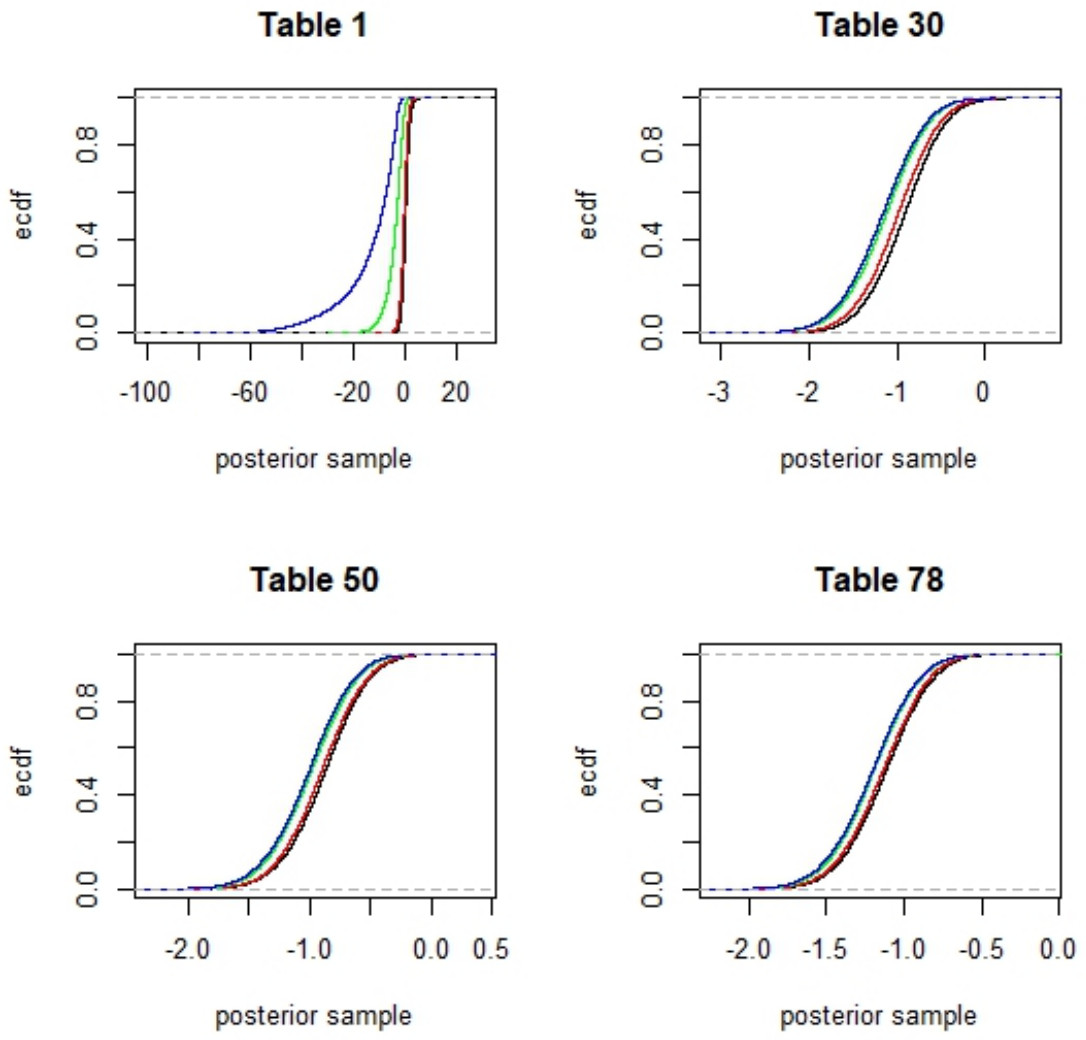


Figure A.53: Empirical CDFs of the posterior samples for number of tables considering a family of beta priors, combination 27.

Appendix B

R Codes

```
##########################################
# MCMC simulations
# Normal Prior for Theta with mu and sigma=1/tau, tau=0.001
##########################################################

## sim.surv is a function for generating Right-censored Survival Data
# n is the total number of observations
# lambdaC  is the rate parameter of exp. dist to gen. censoring obs.
#####################################################

sim.surv<-function(n=40,lambda=0.5,lambdaC=0.001,beta=0){

  # two groups 0 = control, 1=test(treated)
  set.seed(466)
  x1<-sample(c(rep(0,0.3*n),rep(1,0.1*n)),size=n,replace=TRUE)

  # survival times of CPH model with a single binary covariate x1
  # using exponential distribution
  set.seed(477)
  u<-runif(n)
  Time<-(-log(u) / (lambda*exp(x1*beta)))

  # generate censoring
  set.seed(488)
  cen1 <- rexp(n, rate = lambdaC)

  # follow-up times and event indicators
  stime <- pmin(Time, cen1)
  status<-rep(0,length(stime))
  status[stime==Time]<-1
  return(list(stime=stime,status=status,x1=x1))
}
sim.survdata<-sim.surv()

#################################################

library(survival)
km <-survfit(Surv(sim.survdata$stime, sim.survdata$status)~ sim.survdata$x1)
```

```
plot(km, col=3:4, mark.time=TRUE, xlab="Survival Time",
ylab="Estimated Survivor Function",
main="Kaplan-Meier Curves")
legend("topright",legend = c("Control","Test"), col = c("green3","blue")
,lty=1)
#################################################################
# sdata is list of observed times, non-censored indicator,
# and treatment (assumed 0 or 1)
sdata<-data.frame(Time=sim.survdata$stime, status=sim.survdata$status,
treat=sim.survdata$x1)

# con.table is a function to convert right-censored survival data
# to k 2x2 tables
#Constructs 2x2 tables from survival data
con.table<-function(stime=stime,status=status,x1=x1){
# Extract observed deaths
tobs<-sdata[as.logical(sdata[,2]),1]
tall<-sdata[,1] # Extract all times
#survivors at each observed time
obs.survivors<-outer(tobs,tall,match.fun("<"))
#deaths at each time
obs.deaths<-outer(tobs,tall,match.fun("=="))
 # Create data frame for data
stratified<-data.frame(time=tobs,treat=rep(sdata[,3],each=length(tobs)),
alive=as.vector(obs.survivors),dead=as.vector(obs.deaths))
#Filter out who are not at risk
stratified<-stratified[stratified$alive|stratified$dead,]
#Create tables
tw<-table(stratified$treat,stratified$alive,stratified$time,
dnn=c("Treatment","Alive","Time"))
# Remove tables with empty rows
tw[,,apply(apply(tw,c(1,3),sum),2,min)>0]

}
con.table1<-con.table(sdata)

#####################################################
library(coda)
library(rjags)
library(runjags)
#####################################################

# The model in JAGS language (Normal prior on paramete Theta)
model<-"
model{
```

```
c <- 10000 # this just has to be large enough to ensure all phi[i]'s > 0
zeros ~ dpois(phi)
phi <- -L+c
L<-log(prod(((r*exp(theta))^y)/(1+r*exp(theta))))
theta~dnorm(mu,tau)
}
"

############ The Function to run MCMC from R with JAGS

LogOddsSurv<-function(N=10000,C=2,con.table1=con.table1){

  y<-con.table1[2,1, ]
  # y is a binary varable(1 if death and 0 if censoring) and
  #we have only one death in each table.

  # the row  totals in a 2x2 table
  n0<-con.table1[1,1,]+con.table1[1,2,]
  n1<-con.table1[2,1,]+con.table1[2,2,]
  r<-n1/n0
  k<-length(r)

  # To extract posterior sample in an array k*10000*2
  post.sample<-array(dim=c(k,4,10000,2))

  ##run RJAGS using runjags package

  for(i in 1:k){# this loop for k tables

    # Hyperparameters mu and tau for a set of Normal priors
    mu<-c(-200,-2,2,200)
    #var<-c(400,4,4,400)
    tau<-c(0.0025,0.25,0.25,0.0025)

    for(j in 1:length(mu)){#this loop for a set of normal priors
      #as a vecotr of mu's and vector of tau's

      dat<-list(y=y[1:i],r=r[1:i],mu=mu[j],tau=tau[j],zeros=0)
      initfunction <- function(chain)
        return(switch(chain,
                      "1"=list(theta=3), "2"=list(theta=-3)))


      monitor<-c("theta")
      n.chain<-C
```

```
        iter<-N
        burn<-20000
        #thin<-10000

        results <- run.jags(model=model, monitor=monitor,
        data=dat, n.chains=n.chain,sample=iter,
        burnin=burn, method="rjags", inits=initfunction)

        post.sample[i,j,,]<-as.matrix(results$mcmc)

      }
    }
    return(list(results=results,post.sample=post.sample))
}

output<-LogOddsSurv(N=10000,C=2,con.table1=con.table1)
results<-output$results
plot(output$results)
#################################################
# posterior sample and means
post.sample<-output$post.sample # posterior sample observations
###
post.mean<-apply(post.sample,1:2,mean) # posterior means
##### Upper and Lower posterior Expectations of some chosen tables
UE1<-max(post.mean[1,], na.rm = FALSE) #(upper expectation)
LE1<-min(post.mean[1,], na.rm = FALSE) # (lower expectation)
imp1<-UE1-LE1# degree of imprecision
UE1
LE1
imp1
##
UE2<-max(post.mean[50,], na.rm = FALSE)
LE2<-min(post.mean[50,], na.rm = FALSE)
imp2<-UE2-LE2
UE2
LE2
imp2
##
UE3<-max(post.mean[80,], na.rm = FALSE)
LE3<-min(post.mean[80,], na.rm = FALSE)
imp3<-UE3-LE3
UE3
LE3
imp3
##
```

```
UE4<-max(post.mean[93,], na.rm = FALSE)
LE4<-min(post.mean[93,], na.rm = FALSE)
imp4<-UE4-LE4
UE4
LE4
imp4
#################################################################

#imprecise crdible intervals
l1<-quantile(post.sample[37,1,,],0.025)
l2<-quantile(post.sample[37,2,,],0.025)
l3<-quantile(post.sample[33,3,,],0.025)
l4<-quantile(post.sample[37,4,,],0.025)
L<-min(l1,l2,l3,l4)
L

u1<-quantile(post.sample[37,1,,],0.975)
u2<-quantile(post.sample[37,2,,],0.975)
u3<-quantile(post.sample[37,3,,],0.975)
u4<-quantile(post.sample[37,4,,],0.975)
U<-max(u1,u2,u3,u4)
U
#################################################################
#################################################################
##### ECDF plots ################################################

plot1<-function(){ # for table number 1
plot(ecdf(post.sample[1,4,,]),xlim=c(-300,50),
col="black",xlab="posterior sample", ylab="ecdf",main="Table 1")
 lines(ecdf(post.sample[1,3,,]),col="red")
 lines(ecdf(post.sample[1,2,,]),col="green")
 lines(ecdf(post.sample[1,1,,]),col="blue")
}
plot2<-function(){# for table number 50
  plot(ecdf(post.sample[50,4,,]),
  col="black",xlab="posterior sample", ylab="ecdf",main="Table 50")
  lines(ecdf(post.sample[50,3,,]),col="red")
  lines(ecdf(post.sample[50,2,,]),col="green")
  lines(ecdf(post.sample[50,1,,]),col="blue")
}
plot3<-function(){# for table number 80
  plot(ecdf(post.sample[80,4,,]),
  col="black",xlab="posterior sample", ylab="ecdf",main="Table 80")
  lines(ecdf(post.sample[80,3,,]),col="red")
  lines(ecdf(post.sample[80,2,,]),col="green")
```

```
  lines(ecdf(post.sample[80,1,,]),col="blue")
}
plot4<-function(){# for table number 93
  plot(ecdf(post.sample[93,4,,]),
  col="black",xlab="posterior sample", ylab="ecdf",main="Table 93")
  lines(ecdf(post.sample[93,3,,]),col="red")
  lines(ecdf(post.sample[93,2,,]),col="green")
  lines(ecdf(post.sample[93,1,,]),col="blue")
}
##########

par(mfrow=c(2,2))
plot1()
plot2()
plot3()
plot4()
#################################################

# MCMC simulations
# Using Imprecise Beta prior on p with parameters a and b.
# theta = log(p/(1-p))
#################################################
library(coda)
library(rjags)
library(runjags)
#################################################

# The model in JAGS language
model2<-"
model{
c <- 10000 # this just has to be large enough to ensure all phi[i]'s > 0
zeros ~ dpois(phi)
phi <- -L+c
L<-log(prod(((r*p)^y)*((1-p)^(1-y))/(1+(r-1)*p)))
p~dbeta(a,b)
theta<-log(p/(1-p))
}
"

############################## The Function

LogOddsSurv2<-function(n.itr=10000,C=2,con.table1=con.table1){

  y<-con.table1[2,1,]
  # y is a binary varable(1 if death and 0 if censoring) and
```

```
#we have only one death in each table.

# the row  totals in a 2x2 table
n0<-con.table1[1,1,]+con.table1[1,2,]
n1<-con.table1[2,1,]+con.table1[2,2,]
r<-n1/n0
k<-length(r)

# To extract posterior sample in an array k*10000*2
post.sample2<-array(dim=c(k,4,10000,2))

##run RJAGS using runjags package

for(i in 1:k){# this loop for k tables

  # Hyperparameters alpha=a and beta for beta=b priors
  a<-c(0.1,0.3,1.2,1.6)
  b<-c(1.9,1.7,0.8,0.4)

  for(j in 1:length(a)){#this loop for a set of beta priors


    dat<-list(y=y[1:i],r=r[1:i],a=a[j],b=b[j],zeros=0)
    initfunction <- function(chain)
      return(switch(chain,
                    "1"=list(p=0.2), "2"=list(p=0.8)))


    # .RNG.seed <- function(chain)
    #   return(switch(chain, "1" = 1, "2" = 2))
    # .RNG.name <- function(chain)
    #   return(switch(chain, "1" = "base::Super-Duper",
    #                 "2" = "base::Wichmann-Hill"))
    #
    monitor<-c("theta")
    n.chain<-C
    iter<-n.itr
    burn<-20000
    #thin<-10000

    results2 <- run.jags(model=model2, monitor=monitor,
     data=dat, n.chains=n.chain,sample=iter,burnin=burn,
      method="rjags", inits=initfunction)

    post.sample2[i,j,,]<-as.matrix(results2$mcmc)
```

```
    }
  }
  return(list(results2=results2 ,post.sample2=post.sample2))
}

output2<-LogOddsSurv2(n.itr=10000,C=2,con.table1=con.table1)
set.seed(20)
results2<-output2$results2
plot(output2$results2)
################################
```