Imprecise Prior for Imprecise Inference on Poisson Sampling Model

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Doctor of Philosophy
in the Collaborative Graduate Program of Biostatistics
University of Saskatchewan
Saskatoon

By
Chel Hee Lee

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:

Graduate Chair, Collaborative Biostatistics Program
School of Public Health
University of Saskatchewan
107 Wiggins Road
Saskatoon, Saskatchewan
S7N 5E5 Canada

Abstract

Prevalence is a valuable epidemiological measure about the burden of disease in a community for planning health services; however, true prevalence is typically underestimated and there exists no reliable method of confirming the estimate of this prevalence in question. This thesis studies imprecise priors for the development of a statistical reasoning framework regarding this epidemiological decision making problem. The concept of imprecise probabilities introduced by Walley (1991) is adopted for the construction of this inferential framework in order to model prior ignorance and quantify the degree of imprecision associated with the inferential process. The study is restricted to the standard and zero-truncated Poisson sampling models that give an exponential family with a canonical log-link function because of the mechanism involved with the estimation of population size. A three-parameter exponential family of posteriors which includes the normal and log-gamma as limiting cases is introduced by applying normal priors on the canonical parameter of the Poisson sampling models. The canonical parameters simplify dealing with families of priors as Bayesian updating corresponds to a translation of the family in the canonical hyperparameter space. The canonical link function creates a linear relationship between regression coefficients of explanatory variables and the canonical parameters of the sampling distribution. Thus, normal priors on the regression coefficients induce normal priors on the canonical parameters leading to a higher-dimensional exponential family of posteriors whose limiting cases are again normal or log-gamma. All of these implementations are synthesized to build the `ipeglim` package (Lee and Bickis, 2013) that provides a convenient method for characterizing imprecise probabilities and visualizing their translation, soft-linearity, and focusing behaviours. A characterization strategy for imprecise priors is introduced for instances when there exists a state of complete ignorance. The learning process of an individual intentional unit, the agreement process between several intentional units, and situations concerning prior-data conflict are graphically illustrated. Finally, the methodology is applied for re-analyzing the data collected from the epidemiological disease surveillance of three specific cases – Cholera epidemic (Dahiya and Gross, 1973), Down's syndrome (Zelterman, 1988), and the female users of methamphetamine and heroin (Böhning and van der Heijden, 2009).

Acknowledgements

I would like to express my deep gratitude and appreciation for my supervisor Professor Miķelis Bickis. Professor Bickis has been a source of encouragement to me while working in the doctoral program at the University of Saskatchewan. It has been my great pleasure and privilege to have him as my mentor, to have someone to discuss ideas with; his comments always directed me to dive deeper into the world of statistics. I am very appreciative that Professor Bickis has guided my footsteps and directed my development as a researcher.

I would also like to express my heartfelt appreciation to Professor Tony Kusalik of the department of Computer Science for sharing his experiences and interests from a computational perspective. Professor Bonnie Janzen of the Community Health and Epidemiology, the College of Medicine, has provided epidemiological advice and resources to pursue my interests. A simple thank you does not cover the debt of gratitude I owe to Professor June Hyun-Ja Lim from Community Health and Epidemiology, the College of Medicine. She has assisted me in identifying any areas of misunderstanding in epidemiological and clinical research; her constant encouragement has kept me focused on the successful completion of this doctoral program.

I am truly blessed by my family whose love, support, and prayers are the energy that sustains me. I wish also to express my special gratitude and deepest appreciation to Cheryl Palmer. Although she is not a sister by blood, she always becomes my home in my mind.

I am also grateful to Professor Miķelis Bickis for the financial support through his grant from the Natural Sciences and Engineering Research Council of Canada; the College of Graduate Studies and Research for the PhD University Graduate Scholarship, the Teaching Scholar Doctoral Fellowship, the Student Travel Awards. Lastly I wish to acknowledge my gratitude to Professor Thomas Augustin of the Department of Statistics, Ludwig-Maximilians University Munich (LMU), Germany, for facilitating my attendance at the Fifth International Workshop on Principles and Methods of Statistical Inference with Interval Probability.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Epidemiology is the study about factors that influence the distribution of disease (Gordis, 2009). The term disease here is intended for health-related events of study interest. According to Porta (2008), epidemiology involves the what, when, where, and who of disease in terms of time, space, and personal characteristics. The why and how of that disease can be also described by comparing groups with different risk factors. Hence, epidemiological research ultimately assists public health officials when drafting policies or implementing programs to manage health problems requiring prevention or intervention in a distinct population.

This epidemiological observation is generally made by two important measures of prevalence and incidence. Herein, the prevalence is a measure of the number of total cases of a diseases in a population whereas the incidence is the number of new cases occurring in a population over a specific time interval. These counting measures are derived from the data compiled from surveillance systems. Thacker (2010) noted that they are an ongoing systematic process of collecting, monitoring, analyzing, interpreting, and disseminating information about a disease for the purposes establishing prevention, allocation of resources, and evaluation of policy issues. Various health-related surveillance systems in Canada can be found at the Public Health Agency of Canada website (`http://www.phac-aspc.gc.ca/`).

Health related surveillance systems are differentiated into chronic and infectious disease surveillance. Regarding infectious disease, signs and symptoms typically appear over a short period of time in patients exposed to an identifiable agent of an infectious disease; in this instance the purpose of surveillance is the immediate containment of that agent. Concerning chronic disease, the disease first presents as an acute illness with recognizable signs and

symptoms that persist longer than six months with periods of remission and exacerbation but no definitive cure. The causality of its onset or the interaction between risk factors are not well known; thus, the purpose of chronic disease surveillance is to establish programs of intervention to reduce the frequency of disease or to facilitate the identification of risk factors that can then be modified (Health Canada, 2003).

Regarding the provision of health care for individuals with chronic disease, Brasset-Latulippe et al. (2011) address the problems of high prevalence and high economic burdens as follows:

> It is estimated that nearly 16 million Canadians, almost every other one of us, is living with a chronic condition. There are approximately nine million Canadians living with at least one of seven "high-impact, high-prevalence" chronic illnesses. As the population ages so does the increase in prevalence with a majority of Canadian seniors over the age of 65 reporting at least one chronic illness. Chronic illnesses have become a serious economic burden, with total direct medical costs and indirect productivity losses surpassing $93 billion a year. The importance of addressing these costs cannot be overstated. Even more staggering is the number of lives claimed by chronic illness – nearly three quarters of all deaths in Canada arise from only four types of chronic disease. Our international ranking when it comes to addressing chronic care delivery in primary care is no better–Canada ranked last out of seven countries.

1.2 Does The Count Include All Cases?

There are five major data sources that facilitate chronic disease surveillance: administrative health databases, registries, surveys, vital statistics, and census. Health Canada (2003) provided information detailing available data, purpose, advantages and disadvantages, consideration, and existing initiatives regarding each data source (pp. 10–12), and discussed a number of methodological issues for improving the current process. One of the issues addressed is the quality of data which may result in changes to the epidemiological picture as prevalence and incidence are based on this data.

Significant efforts have been made by the Canadian Institution of Health Information (CIHI) in promoting data quality assurance. In the Data Quality Framework (CIHI, 2009), more than 61 definitions describing data quality can be found. Although the CIHI primarily maintains administrative health databases provided by hospitals, medical expert groups,

governments, and regional and national health care organizations, these definitions can be applied to other data sources as well. Bray and Parkin (2009) also provide a good background reference for evaluating data quality in cancer registries. Among various definitions of data quality, our interest centres on the question of "Does the count include all cases that are intended to be collected?" since the systems are prone to be incomplete (i.e, the number of cases ascertained in the system is smaller than the actual number of cases occurring in the population) in practice. Here are examples showing that databases can be incomplete:

- For registries, the Canadian Cancer Society Steering Committee (2011) recognized that information reported on the death certificate was not available for registry purposes in Newfoundland and Labrador when producing mortality statistics using the Canadian Cancer Registry (p. 111). In practice, primary or secondary causes of death written on the death certificate may not correspond exactly to a particular cancer causing the individual's death. The Canadian Cancer Society Steering Committee (2011) also noted that most provincial and territorial cancer registries do not collect data for skin cancers other than melanoma since a hospitalization for treatment is not required (p. 112). Another example found by the Canadian Cancer Society Steering Committee (2011) is that while the mortality of colorectal cancer has dropped by approximately 10% subsequent to the change of case definition in 2003 while only a small variation was observed in other cancers.

- For administrative health databases, the Canadian Institute for Health Information (2007) noted in their report a relationship between mental health and homelessness:

  > Hospital staff may not always be aware or informed of a patient's current housing status, particularly for subsequent visits, this information may not be up to date and thus may reflect an incomplete count. Information reflects only those homeless individuals presenting for medical attention at participating hospitals. In addition, since there is no comparable count of the total population in Canada, rates cannot be calculated and compared with the total population (p. 20).

- For surveys, the Public Health Agency of Canada (2010b) expressed concern about the fact that "many Canadians at high risk of osteoporosis are not being screened for osteoporosis using a bone density test" regarding the use of the 2009 Canadian

Community Health Survey for the study of osteoporosis. Another concern addressed by the Public Health Agency of Canada (2010a) is that

> Results based on blood pressure measurements from the Canadian Heart Health Surveys in 1986-1992 showed that 42% of the study participants were unaware of their hypertension (47% of men and 35% of women). Likely as the result of intensive efforts to improve the detection and management of hypertension, this has changed. In the 2006 Ontario Survey on the Prevalence and Control of Hypertension a much lower proportion of individuals with hypertension (13.7%) were unaware of their condition. Similarly, national estimates from Cycle 1 of the Canadian Health Measures Survey conducted between 2007 and 2009 indicated that 17% of Canadians with hypertension were unaware (p. 5).

In addition to the examples listed above, "poorly defined criteria for diagnosis, missed diagnosis, poorly designed surveillance systems, lack of health-seeking behaviour by those with the diseases and/or risk factor" are all contributing factors resulting in an incomplete data collecting system (Nanan and White, 1997, p. 144). Consequently, the discrepancy that exists between the true number of cases in a population and the ascertained number of cases in a system may be large; thus, Nanan and White (1997) noted "the number of ascertained cases may greatly be underestimated even though some diseases and their risk factors may have a high prevalence in a population" (p. 144).

Estimates of prevalence or incidence derived from incomplete systems may mislead the true epidemiological structure of a population and ultimately fail when identifying a sub-population most at risk. Recognition of this issue leads us to seek a measure of completeness. The rationale for completeness starts with assuming that the size N of a target population consists of the size n of an observed population and the size n_0 of an unobserved population. It is also assumed that some identification processes are designed with the expectation that all cases occurring in a population are ascertained cases without any misses. If it is possible to estimate n_0, the completeness measure about the system is defined as a ratio of the size n of the ascertained population to the estimated size $\hat{N}$ of the population. The measured completeness ultimately implies a degree of how useful the system is when producing epidemiological measures from data compiled from the system (Nanan and White, 1997). In this sense, the problem of estimating the number n_0 of missed cases is identical to the problem of estimating the size N of a target population since N is the sum of the number n_0

of missed cases and the number n of ascertained cases. An analogous discussion about the measurement of completeness can be found in the studies of Parkin and Bray (2009) and Iron and Manuel (2007, p. 3).

1.3 Capture-Recapture Method

The World Health Organization and UNADIS (2010) provide a practical guideline for epidemiologists conducting studies estimating the size of the HIV population most at risk. Section 3.2 in this guideline describes a number of methods that can be used. The principle of population size estimation, advantages and disadvantages of each method, and working examples are given in general terms. Another good reference is found in the technical briefing written by Walford et al. (2011) for an overview of various methods of producing an estimate of the chronic disease population in Europe. The reader will notice that the capture-recapture (CR) method is covered in both reports.

Extensive literature on the CR method can be found in medicine, epidemiology, and public health for monitoring health-related events: birth defects (Wang et al., 2006), drug misuse (Hay et al., 2009), tuberculosis (Van Hest et al., 2007), multiple sclerosis (Cristiano et al., 2009), cancers (Peragallo et al., 2011), varicella (Goldman, 2003), Down's syndrome (Savva and Morris, 2009), alcohol related problems (Ponzio et al., 2010), various chronic diseases (Zhao et al., 2008), such as hypertension, diabetes, renal disease, ischemic heart disease, etc.

The first application of the CR method is presented in the study by Petersen (1895) dealing with the problem of estimating the size N of the fish population in the pond. The study is conducted as follows: a sample of n_1 fish is taken from the pond; then, the fish are marked and released into the pond. Another sample of n_2 fish is taken from the same pond after allowing adequate time for the marked and unmarked fish to be well mixed. The counts of n_{11} fish captured twice, n_{10} fish captured only in the first sample, n_{01} fish captured only in the second sample are available now. The count of n_{00} fish not captured in either sample is estimated by $n_{10} \times n_{01}/n_{11}$ under four simplifying assumptions: homogeneous population, independent samples, correct matching between samples, and a closed population. Another example is given by Feller (1968, pp. 44–47).

The same analogy can be applied to a human population. Consider a situation where two databases for physician claims and hospital discharges are given. Assume also that a hypothetical population consists of N subjects with a certain disease. The presence or absence of a patient in each database is denoted by 1 and 0, respectively. Patients can be cross-classified in a 2×2 contingency table. The counts of each response profile (01), (10), (11) are made as illustrated in the example of the fish population. Denotation of those counts by n_{01}, n_{10} and n_{11}, and the size of all ascertained cases by $n = n_{10} + n_{01} + n_{11}$. Computation of the estimate $\hat{N}$ of a total population size is then straightforward and can be determined by summing all counts n_{01}, n_{10}, and n_{11}, including the estimated size $\hat{n}_{00}$ of patients who are not ascertained in either database. Subsequently, the degree of completeness is evaluated by $n/\hat{N} \times 100$ in percentage.

One concern regarding the use of the CR method using two data sources is the violation of the independence assumption. In most situations regardless of disease types, patients in hospital databases are also shown in a claim database since they are likely to use medical services at clinics rather than hospitals when their signs and symptoms are not severe. Brenner (1996) shows that $\hat{N}$ is under- or over-estimated, respectively, if two data sources are negatively or positively dependent. No statistical tools are available at this time to test the assumption of independence. For example, a chi-square test for independence requires information of n_{00}. Chao et al. (2001) noted that the independence assumption is the main weakness of the CR method when using two data sources (p. 3129).

The log-linear model proposed by Fienberg (1972) is a useful statistical modelling approach to account for dependence between more than two data sources using the interaction term. Concise mathematical details can be found in Section 6 of Bishop et al. (2007). Another excellent guide for the use of this model in CR studies is provided by Hook and Regal (1995). IWGDMF (1995b) describes the usefulness of a log-linear model as follows: 1. the extension to more than two data sources is straightforward; 2. the study can be easily conducted using standard statistical software (the demo program can be found in `demo/loglinear.R` in the `ipeglim` package accompanying this thesis); 3. various structural features can be examined under a unified framework. However, the log-linear model with a full interaction term be-

tween data sources cannot be studied because the over-specification of the model causes the degree of freedom to be zero (Bishop et al., 2007, p. 238).

Another concern with the CR studies using log-linear models is the assumption of a homogeneous population which means that each subject in the population has an equal probability of being captured in a data source. Papoz et al. (1996) noted that a violation of this assumption has occurred when a capture probability is influenced by covariates such as age, sex, and severity of disease (p. 475). This violated assumption leads the estimated population size to be negatively biased (Plante et al., 1998). In fact, the consequence resulting from heterogeneity is similar to the effect that occurs when the assumption of independence is violated. Many authors noted, from observed data, that it is not possible to identify which violated assumption causes a negative bias on the population size estimate (Chao et al., 2001; Hook and Regal, 1995; IWGDMF, 1995a).

Concerning the presence of heterogeneity, Hook and Regal (1993) suggested to stratify a population based on covariate information and then estimate a population size in each stratum. The total population size is then estimated by pooling all these estimates. This strategy may help to reduce the effect of heterogeneity if covariate information is associated with capture probabilities. However, it does not account for the effect of unobserved heterogeneity which means an unexplained variation by covariate information due to a limitation of our ability to measure some influential factors. Development for modelling of a heterogeneous population occurs in two directions for observed and unobserved heterogeneity.

For observed heterogeneity (i.e., when covariate information is available), Alho (1990) built a logistic regression framework by relating an individual's capture probability to continuous covariate information. The population size is then estimated with a Horvitz-Thompson's estimator. Alho et al. (1993) applied this model to the United State census. For unobserved heterogeneity (i.e., when covariate information is not available), a mixture or latent class model has been considered as an alternative to a regression model. The fundamental idea supporting the latent class model is that a heterogeneous population is divided into a number of homogeneous sub-populations where each one has a class membership of unknown criteria. According to the simulation studies in Dorazio and Royle (2003), the population size is sensitive to the presumed structure of the mixture model. They demonstrated that the

estimated population size fitted by the latent-class model was more biased than that fitted by the beta-binomial model or logistic-normal mixture model.

A further development of latent class modelling that allows for both observed and unobserved heterogeneity is made by Thandrayen and Wang (2010). They studied a binomial latent class model and applied it to the data studied in Bruno et al. (1994) for the prevalence of diabetes. The idea supporting their model is to utilize a latent categorical variable for class membership, and in each latent class, a multinomial logit model proposed by Zwane and van der Heijden (2005) is employed to relate an individual's capture probability to the covariate information. This work thus allows for a conditional dependence between data sources based on the work of Zwane and van der Heijden (2005).

1.4 Zero-Truncation in Statistical Modelling

The zero-truncated Poisson (ZTP) model has been also used for the study of size estimation in a human population. The ZTP model is generally used to describe count outcome in instances where a zero-valued count cannot occur. A typical example is the length of hospital stay since a minimal stay for hospitalization is one day. Winkelmann (2008) noted that a zero-valued count implies the individual in question is healthy in relation to the health care modelling system (p. 173). Since these people are not recorded in the system, the analyst observes only a part of all possible ranges of count data from their frequency distribution. The problem then becomes to infer a probability of a zero-valued count corresponding to the size of a hidden population.

Note that in general a truncation of data is classified as left- and right-truncation. Zero-truncated data is a special case of left-truncated data occurring at the zero value. Note also that zero-truncated data is not censored data since the number exceeding the censoring point is known whereas the number exceeding the truncation point is not known. Principles, estimation methods, and applications about various kinds of truncated and censored distributions are provided in the book by Cohen (1991).

The ZTP model uses a frequency distribution of count data which can be compiled from a single database if the same individual appears multiple times in the same database. This

count data is also regarded as the CR data (Böhning et al., 2005, p. 2) since the count information includes the nature of a repeated capture-recapture process that ascertains a case multiple times. It is also possible to produce such count data from multiple data sources by counting how occurrences of a case are present over multiple data sources (Hook and Regal, 1995, p. 260). For this reason, Vergne et al. (2012) referred to this ZTP model as a unilist CR method (i.e., single databases used) and the models listed in the previous section 1.3 as multilist CR methods (p. 128).

According to Böhning et al. (2005), the first epidemiological application using the zero-truncated count data appears in the study by McKendrick (1925). Dr. McKendrick was interested in the size of households with active but undetected cholera cases during his service in India. Dahiya and Gross (1973) examined this cholera epidemic data and reported the maximum likelihood estimate of the total population size.

Rider (1953) also presents an interesting approach for estimating the frequency of a zero count. He used frequencies of two different counts for estimating the mean parameter of the ZTP model and compared his results with the maximum likelihood estimate. Later, Zelterman (1988) adopted the approach presented in the study by Rider (1953) for proposing his local estimator that uses the frequencies of only the first and second counts for the problem of population size estimation. This local estimator is referred to as the Zelterman's estimator in the study by Böhning (2008).

Vergne et al. (2012) noted two main assumptions of the ZTP model in his study as follows: 1. a homogeneity assumption – all subjects with a certain condition have an equal probability of being ascertained by some identification process, and 2. an independence assumption – each case ascertainment is independent of its previous case ascertainment at the individual level (p. 3). Consequently, the estimated population size is biased if either one of the assumptions is violated. A further development path of this ZTP model is similar to the development history of the multilist CR methods.

Studies can be conducted in numerous ways once heterogeneity is introduced into a homogenous population. In a mixture approach, Zelterman (1988) modelled Poisson data as an average over a gamma or a log-normal distribution. According to Böhning (2008), Zelterman's estimator is robust concerning the occurrence of contamination in the Poisson model

and is also less dependent on the Poisson assumption (p. 413). Continued efforts to deal with unobserved heterogeneity in a population are made using the binomial and Poisson mixture models as shown in the study by Böhning et al. (2005). Using a regression approach, van der Heijden et al. (2003) related the mean parameter of the ZTP model to the linear predictor of covariates with a log-link function in the generalized linear model setup. He also derived the point and interval estimators for population size based on the use of the Horvitz-Thompson's estimator. Since the ZTP regression model accounts for only observed heterogeneity, Cruyff and van der Heijden (2008) became concerned with the problem of the remaining unobserved heterogeneity in a population and extended the work of van der Heijden et al. (2003) by incorporating a dispersion parameter. Hence, the zero-truncated negative binomial model deals with both observed and unobserved heterogeneity.

1.5 Recognition of Ignorance and Uncertainty

The capture-recapture (CR) method provides an inexpensive and fast way for projecting the burden of a disease in the community. The primary purpose supporting the use of the CR method is the introduction of a modelling process of a structural zero term (Bishop et al., 2007, p. 177), which corresponds to incomplete data. Once a probability of this zero term is estimated, the completeness of a surveillance system is also available.

However, it is questionable whether the population size estimated by the CR method measures what it is supposed to measure in the general context. Papoz et al. (1996) noted that there is no way to ensure that the estimate is valid since no information is available concerning these missing cases (p. 477). So should we conclude that the CR method should not be used for further studies? Many authors, including Papoz et al. (1996), have concluded that in the absence of a direct population survey or a complete census we have only the CR method by which to estimate this population size. This is illustrated using the example of an outbreak of the hepatitis A virus (HAV) studied by Chao et al. (2001) which demonstrates that the true size of underreported cases is very closely estimated when using the CR method (p. 3124). Van Hest et al. (2008) also found more plausible estimates using truncated models by re-examining the log-linear estimates in the previous 19 studies.

The author of this thesis believes it sounds more feasible for the question to be revised as "How can we justify the use of the research result as a credible estimate in practice when the unknown quantity of interest cannot be revealed or validated?". More generally, "How do we undertake a reasoning process about unknown quantities of interest when information is insufficient or absent when making a decision?".

This revision originates from questioning the interpretation of a probability that supports the CR models for inferring the unknown quantity of interest. Herein, the CR models are limited to only the models noted in Sections 1.3 and 1.4. Following the definition of probability in Hogg et al. (2013), the standard interpretation of a probability is based on the assumption that an event of interest is observed from an infinite sequence of a well-designed experiments (p. 15). Thus if we are able to observe enough repetitions of a case-ascertainment process for the same person using the same identification method when all else remains unchanged, the probability of a future case ascertainment would be close enough to its actual relative frequency. The interpretation of the underlying probability model does not sound appropriate given the data under examination (Gill, 2008, p. 6). Consider a confidence interval which is reported with the estimate, more generally, a statistic of interest. The problem of interpretation arises again in the same manner since a $(1-\alpha)\%$ confidence interval with a confidence level α implies that a certain percentage $1-\alpha$ of intervals will capture the true value of the unknown parameter assumed to be fixed over many ascertainment processes utilizing an unchanged population.

The attribute of data used for the interpretation of a probability model is also arguable (Gill, 2008, p. 26) since what we are dealing with here is not an outcome produced from an experiment but conceptualized information acquired from a human. Consider a situation where you are seeing a physician in your neighbourhood walk-in clinic since you are sick. Your clinical signs and symptoms will vary depending on the progress of the disease causing your sickness. If it is the initial complaint about your sickness, you may not be able to determine for yourself what disease(s) you may have. The physician uses his or her skills to make a diagnosis about your clinical state for treatment based on your clinical features together with other relevant information. The physician may experience difficulty arriving at a diagnosis of your condition since evidence needed for decision making is not always sufficient in a clinical

setting (Guyatt et al., 2008, pp. 12–13) for various reasons. For example, 1. you may have forgotten to report your unobserved signs and symptoms at the clinic, 2. your clinical features may not be clearly apparent without extensive tests, or 3. your current clinical condition may conflict with the previous clinical observation noted in the medical chart, etc. Hence, a precise diagnosis cannot be stated with certainty so that the diagnostic is generally classified into the one of followings "definite", "probable", "possible", "other disease", or "insufficient information" (Ball et al., 2002, p. 821). Consequently, in the early stages of your disease development, it may be possible that you are in fact receiving inappropriate treatments from your physician. This clinical practice shows that information acquired from a human is a simplified notion of the clinical state inferred from given evidence. Such a cognitive activity for prediction cannot be found using the outcomes produced from an experiment which are also clearly distinct from one and another in a physical manner (Gill, 2008, p. 26).

The epistemic probability which models a degree of personal belief (Walley, 1991, p. 14) meets our concerns with the interpretation of an unknown quantity of interest. The fundamental idea supporting this probability model is that *You* are the intentional unit who assigns some measure that describes how uncertain *your* belief is about the unknown quantity of interest. It is presumed that *your* uncertainty originates from a lack of knowledge; and, this knowledge is the source of the information which is producing *your* belief. If *you* know a lot, *you* may have a strong statement about the uncertain quantity at some point. If *you* know little, *you* make a vague probabilistic statement. Personal experience, expert opinions, literature reviews, and something useful for reducing uncertainty on the inference are the sources of information. Before observing the data, *your* belief is elicited in the form of a probability measure called a prior distribution (also simply called a prior). The Bayes' rule is applied to update the prior by taking the data to produce a probability measure called a posterior. This epistemic probability model thus has an evidential interpretation based on the given data. The Bayesian inference paradigm is built on this subjective perspective of a probability model. The formal mathematical description of Bayesian inference is described in Chapter 2. A good reference on the Bayesian theory is the book by Bernardo and Smith (2000). Another book by Gelman et al. (2004) includes various examples of Bayesian data analysis as well.

One may claim that this Bayesian inference lacks objectivity that research should fulfill since probabilities for personal belief are arbitrarily assigned by *your* own assessment. However, it must be noted that any inferential paradigm has subjective components in modelling uncertainty (Gill, 2008, p. 27). A statistical model is generally understood as an approximation to some truth (Christensen et al., 2011, p. 1). A mathematical language is used as the tool to describe the body of systematic components in this approximation. The results from the model can thus be reproducible under similar circumstances to those in which the study is initially carried out. However, this truth may not be the truth since the description utilizes only the things that can be observed or manipulated. Since researchers as well have their own approaches for the problem, a model has its own assumptions, model specifications, and model selection criteria (Gill, 2008, p. 27) to aid in choosing the most reasonable model to be believed. Hence, the estimate produced from various models are varied despite the use of the same data and the same problem identification.

A general subjectivity involved with the epistemic probability model is defended in the perspective of the modelling process in the previous paragraph; however, the major criticism of Bayesian inference concerns the selection of the most appropriate prior distribution over arbitrary choices (Roberts, 2007, Chapter 3). A number of objective rules are proposed by several authors in order to reduce subjectivity regarding this matter (Berger, 2006). These rules are largely classified into either informative or noninformative priors. Details about these priors are noted in Chapter 2.

However, the question of "how do we precisely characterize our uncertainty in the form of a single probability measure?" is not answered. Walley (1991) referred to this question as "the Bayesian dogma of precision" (p. 3). Herein, the term precision is used for a single point representation in mathematics. A sensitivity analysis can be considered in this case (Roberts, 2007, p. 141). That is, a standard Bayesian inference is applied to each pair of a class of priors and a class of likelihoods, then the robustness of the posterior summary is considered. If a large change of the model assumptions produces a small change on the posterior summary, the data used in the inference is considered to be sufficiently influential.

At this time we return to our research interest again, and assume the perspective of health care professionals who are faced with the problem of making a decision about a population

size for an intervention program. We limit ourselves to only the statistical inference here since the utility assessment required for Bayesian decision theory is difficult and "consequence of choosing a particular conclusion may be wholly indeterminate" at the time when a conclusion is determined (Walley, 1991, p. 21). A broad range of expectations about this size is possible based on a researcher's experience, resources, knowledge of the community, timing, values, etc. It is provident to have a model that produces a range with a lower bound that is larger than the size of ascertained cases since a reasonably excessive allocation of health resources for disease prevention is more prudent than a short fall.

The Bayesian sensitivity analysis appears to be an appropriate potential approach to this problem since the interval measure that contains all plausible point estimates is a mathematical language that describes the range of the researcher's expectations. However, the interpretation of the interval measures produced as a result of arbitrary changes on the model assumptions is not clear and it seems that there exits no formal strategy regarding how to make these changes on the model assumptions.

The imprecise probability theory introduced by Walley (1991) gives insights on how to approach our concerns while making a useful interpretation of the model used in sensitivity analysis. The fundamental premise forming the foundation to his theory is to admit to probability theory the concept of ignorance arising from a lack of information. Based on his theory, a sensitivity analysis is carried out by replacing a precise probability measure characterized in the standard Bayesian inference with a set of probability measures. This set is referred to as an imprecise prior and is not assumed to have a true probability measure in this set of probabilities which is a distinguishable viewpoint of Bayesian sensitivity analysis (Walley, 1991, p. 254). The inferred posterior interval with given data also has an interval form and is referred to as an imprecise posterior. Since the imprecise prior describes our uncertainty due to the lack of information or disagreement between individuals, the imprecision in the posterior describes our indeterminate preference that cannot support an action. The difference between the lower and upper bounds of the imprecise posterior ultimately implies the amount of uncertainty attributed to prior ignorance.

There are two books of interest regarding the general theory of imprecise probabilities; one by Walley (1991) and the other by Weichselberger (2001). Since the book of Weichsel-

berger (2001) is written in German, only the book by Walley (1991) is referenced for this thesis work. The probability model referred to as the imprecise Dirichlet-Multinomial model (Walley, 1996; Bernard, 2005) is frequently used to illustrate the elementary principles of an imprecise inferential framework. However, the mathematical foundation of this theory is still under development and only a few applications have ever been shown in use by other disciplines. Based on our literature search no applications of this imprecise probability with a zero-truncated Poisson sampling model are found in epidemiology and public health. This phenomenon may be attributed to the computational burden on the estimation because of a generalization process of a precise probability measure in the standard Bayesian inference and the corresponding problem of optimization when searching for the minimum and maximum of an imprecise posterior summary, or may be due to unfamiliarity with the concept of imprecise probabilities.

1.6 Thesis Organization

The aim of this thesis work is to build a computationally efficient inferential framework based on the idea of imprecise probabilities for estimating the parameters of a given sampling model. Principal features of the developed inferential framework will be identified. The `R` (R Development Core Team, 2011) package is also provided for potential users of this developed inferential framework. The remainder of this thesis is organized as follows:

In Chapter 2, the essential statistical ingredients necessary for building the proposed inferential framework are briefly reviewed. The principles of Bayesian inference, generalized linear model, exponential family representation, and numerical techniques are covered;

In Chapter 3, a methodology referred to as a canonically parametrized imprecise inferential framework is introduced. A family of log-gamma distributions is utilized to illustrate six primary features of this proposed inferential framework for the problem of estimating the parameters of both standard Poisson and zero-truncated Poisson sampling models. The key functions provided in the `ipeglim` package (Lee and Bickis, 2013) are briefly introduced, and the computational difficulties encountered with the implementation of this proposed inferential framework are discussed.

In Chapter 4, the central part of the proposed inferential framework referred to as the $\mathfrak{B}$-formulation (Lee and Bickis, 2012) is introduced; followed by a new probability distribution resulting from this $\mathfrak{B}$-formulation is discussed for further extension of the proposed inferential framework to the case when 1. a family of normal prior distributions is assigned on a standard Poisson likelihood, 2. either families of log-gamma and normal prior distributions are assigned on a zero-truncated Poisson likelihood, and 3. a family of normal prior distributions is assigned on regression coefficients in the setup of a generalized linear model. The primary features shown in the previous chapter are examined again in order to determine that validity is maintained.

In Chapter 5, simulation studies are carried out under various conditions in response to concerns involving the practical use of the proposed inferential framework. For example, the effects of sampling model misspecification, overdispersion of count data, and the absence of explanatory variables on the estimate produced from the proposed inferential framework are investigated. A further investigation regarding the change of parameter estimates is made against a given correlation structure at both levels of prior variance-covariance matrix and data matrix.

In Chapter 6, the proposed methodology is applied to four real data sets collected from a human population. The case studies of a cholera epidemic (Dahiya and Gross, 1973), Down's syndrome (Zelterman, 1988), and Methamphetamine and Heroin users (Böhning and van der Heijden, 2009) are reanalyzed, and an interpretation of the estimate produced from the proposed methodology is discussed.

Finally, the conclusions and proposals for future work are presented in Chapter 7.

Chapter 2

Essential Statistical Ingredients

The purpose of this chapter is to provide a general overview of the essential statistical ingredients compromising the inferential framework that will be introduced in the next chapter. This chapter is outlined as follows: Section 2.1 gives an introduction to the basic mathematical formulation of the Bayesian inference; Section 2.2 provides brief details regarding the exponential representation of a family of distributions. The three key components of a generalized linear models are also noted pertaining to this exponential representation; Section 2.3 presents five numerical techniques for the purpose of approximating an integral and finding the optimum that maximizes or minimizes a given function; finally, Section 2.4 describes a zero-truncated Poisson regression model introduced by van der Heijden et al. (2003) in the context of population size estimation.

2.1 Bayesian Inference

A statistical inference is a reasoning process that facilitates the drawing of a conclusion from data (Walley, 1991, p. 21). This process is generally described by a model in mathematical language and a number of assumptions are imposed on this description. Since the model is a simplified reflection of what should be described by using only what can be observed and manipulated, the uncertainty that is associated with this simplifying process needs to be reduced by gathering more information when the inference is carried out.

According to Migon and Gamerman (1999, p. 2), there are two main schools of thoughts involving statistical inference called the frequentist paradigm and the Bayesian paradigm (p. 3). One important distinction between the two schools of thought is that, under the frequentist inferential paradigm, the data are random and model parameters are fixed, whereas

when using the Bayesian paradigm, the data are fixed and model parameters are random. Regarding the interpretation of probabilities supporting the two schools of thought, Walley (1991, p. 14) provided a clear explanation such that "a physical property that does not depend on an observer" is a statement of aleatory probability that models the randomness of an event (i.e., the frequentist inferential paradigm) and "the degree of belief of a specific observer" is a statement of epistemic probability that models a degree of that person's belief (i.e. the Bayesian inferential paradigm).

The mathematical framework of Bayesian inference is described as follows. Consider first the data $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ which are the realizations of independently and identically distributed random variables $Y_1, Y_2, \ldots, Y_n$ with a probability distribution $f(\mathbf{y}|\theta)$, having parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$. This probability distribution $f(\mathbf{y}|\theta)$ is referred to as the sampling model of data $\mathbf{y}$, and the parameter space Θ is defined by the set of all possible values of θ. A probability distribution of θ needs to be specified since the model parameter θ is viewed as a random variable for representing the degree of a person's belief. This probability distribution is referred to as a prior distribution $\pi(\theta)$ and is characterized by the parameters $\xi = (\xi_1, \xi_2, \ldots, \xi_k)^T$, so-called hyperparameters. Note that the hyperparameter space Ξ is defined as the set of all possible values of ξ.

The degree of a person's belief represented by a prior density or mass function $\pi(\theta)$ is updated by Bayes' theorem, and this updated degree of his or her belief is referred to as a posterior distribution $p(\theta|\mathbf{y})$ which is given by

$$p(\theta|\mathbf{y}) = \frac{\mathscr{L}(\theta|\mathbf{y})\pi(\theta)}{\int_\Theta \mathscr{L}(\theta|\mathbf{y})\pi(\theta)d\theta} = \frac{p(\theta, \mathbf{y})}{m(\mathbf{y})}, \tag{2.1}$$

where $\mathscr{L}(\theta|\mathbf{y}) = \prod_{i=1}^n f(y_i|\theta)$ is the likelihood that takes the effect of data $\mathbf{y}$, $p(\theta, \mathbf{y}) = \mathscr{L}(\theta|\mathbf{y})\pi(\theta)$ is a joint distribution that describes a relationship between a model parameter θ and the data $\mathbf{y}$, and $m(\mathbf{y}) = \int_\Theta \mathscr{L}(\theta|\mathbf{y})\pi(\theta)d\theta$ is the marginal distribution of $\mathbf{y}$. This marginal distribution of $\mathbf{y}$ promises that (2.1) is a proper probability distribution.

The posterior distribution $p(\theta|y)$ is generally summarized by its first two moments; however, the primary summary used in this thesis is the expected value of some function of θ,

for example $h(\theta)$. By definition, the expected value of $h(\theta)$ is then:

$$E(h(\theta)|\mathbf{y}) = \frac{\int_\Theta h(\theta)\mathscr{L}(\theta|\mathbf{y})\pi(\theta)d\theta}{\int_\Theta \mathscr{L}(\theta|\mathbf{y})\pi(\theta)d\theta}. \tag{2.2}$$

As shown in the above two equations (2.1) and (2.2), a successful Bayesian inference depends on the ability to accurately evaluate the ratio of two integrals (Bernardo and Smith, 2000, p. 340). This evaluation may be challenging when a closed form expression of normalizing constant for (2.1) is not found from some known standard distributions. For the quantification of this integration in this thesis work, three numerical approximation methods are utilized. Details of these approximation methods are described later in Section 2.3.

To complete the Bayesian inference, a prior distribution $\pi(\theta)$ for a model parameter θ needs to be elicited in the form of a probability measure before data $\mathbf{y}$ is examined. Since this elicitation varies from person to person, various rules regarding a choice of a reasonable prior for inference have been proposed. Sections 3.3 and 3.5 in a book by Roberts (2007) are credible sources for these rules. Selected priors from his book are briefly summarized here in two general classes:

1. Noninformative priors – priors in this class have minimal information to explain unknown model parameters θ; thus, only data affect posteriors. A typical example is an unbounded uniform distribution. This prior is also referred to as the Laplace prior since all probabilities are equally assigned. Note that a probability over unbounded space induces an improper prior distribution which means that the integral of the presumed prior is infinite. Because the inference fails when such improper prior distributions are used, some bounds are imposed on the space or avoided in practice. When this uniform prior is considered inappropriate to use, an alternative choice is Jeffrey's prior which is proportional to the square root of the negative expected value of the second derivative of the likelihood $\mathscr{L}(\theta|\mathbf{y})$ (i.e., the square root of Fisher's information). Although Jeffrey's prior is widely accepted in univariate cases, when used in multivariate cases, the results were disappointing. In circumstances where Jeffrey's prior does not work well, one may consider the use of a reference prior which is defined as a function that maximizes some measure of divergence between the prior and the posterior distributions.

Consequently, the effect of the data is maximized on the posterior distribution. The Kullback-Liebler divergence is a commonly used measure for this prior.

2. Informative priors – priors in this class glean specific information from evidence that exists about unknown parameters and utilizes that information as a part of the priors. Examples of such existing evidence would be previously published papers, personal experiences, or the opinions of experts. Because the information delivered to the posterior is ultimately taken into consideration by data, information contained in the posterior is considered more evidential than information derived from the prior. A commonly used informative prior is the conjugate prior which means that the prior and posterior distributions belong to the same family of distributions. This conjugate prior is characterized by a mathematical convenience that offers easy computation of a posterior distribution. Such conjugate prior with a large scale parameter is sometimes referred to as a vague (or flat) prior.

In the above note, prior distributions are grouped into two classes of priors – noninformative and informative priors; however, the author of this thesis questions whether or not a truly noninformative prior exists as discussed in a paper written by Irony and Singpurwalla (1997).

2.2 Generalized Linear Model and Exponential Family

The generalized linear model (GLM) was proposed by Nelder and Wedderburn (1972); McCullagh and Nelder (1989) and Dobson (2001) are comprehensive references for GLM. The central idea behind the GLM is to relate a response variable y_i to a set of explanatory variables $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})^T$. The subscript i is used for the i-th individual of n independently sampled subjects; thus, the response variable y is a vector of independent observations $(y_1, y_2, \ldots, y_n)$ and $\mathbf{x}_i$ is a vector of p explanatory variables corresponding to the i-th row of the design matrix $\mathbf{X}$. To complete the model, the following three components should be defined:

1. a random component – this component specifies the conditional distribution $f(y)$ of a response variable y given the values of the explanatory variables $\mathbf{x}_i$ which need to be a member of an exponential family of distributions. According to Casella and Berger (2001, Section 3.4), an exponential family is a set of probability distributions whose probability density function $f(y|\phi)$ can be written as

$$f(y|\phi) = s(y)t(\phi)\exp\{\sum_{i=1}^{k} a_i(y)b_i(\phi)\}, \tag{2.3}$$

where $s(y)$ and $a_i(y)$ are real-valued functions of y that do not depend on ϕ, and $t(\phi)$ and $b_i(\phi)$ are real-valued functions of ϕ that do not depend on y, and $s(y)$ and $t(\phi)$ are greater than or equal to zero for all y and ϕ. When $k = 1$, $f(y|\phi)$ is a one-parameter exponential family of distributions. Note that if either $a(y) = y$ or $b(\phi) = \phi$, then $f(y|\phi)$ is in the canonical form in terms of y or ϕ, respectively. In the case of a Poisson distribution with mean μ, $a(y) = y$, $b(\phi) = \log\mu$, $s(y) = 1/y!$, and $t(\phi) = e^{-\mu}$. If the canonical parameter $\theta = \log\mu$, the expectation and variance of $a(y)$ can be found by taking the first and second derivatives of the log-normalizer $-\log t^*(\theta)$ such that

$$E[a_i(Y)] = -\frac{\partial}{\partial\theta_i}\log t^*(\theta), \quad \text{and} \quad V[a_i(Y)] = -\frac{\partial^2}{\partial^2\theta_i}\log t^*(\theta). \tag{2.4}$$

2. a systematic component – this component specifies a linear predictor function $\eta = (\eta_1, \eta_2, \ldots, \eta_n)^T$ of explanatory variables $\mathbf{x}_i$ such that

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \tag{2.5}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ is a vector of the regression coefficients.

3. the link function – this function describes how the mean response $E(Y_i) = \mu_i$ in the random component is related to the linear predictor η in the systematic component:

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \tag{2.6}$$

where $g(\cdot)$ is a smooth and invertible link function. In the case of a Poisson sampling model, the choice of canonical link as $\log(\mu)$ ensures that the assumption of a positive Poisson mean μ is met.

2.3 Numerical Techniques

Integration is a key consideration in Bayesian computation for inference. Two sampling-based approaches, called the Metropolis-Hastings and importance sampling, and one analytic approach, called the Laplace approximation, are presented here. This thesis work also considers the problem of optimization which means to identify a parameter that maximizes or minimizes a given function. A particular family of optimization problems called linear programming is described at the end of this section.

2.3.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm is one of the most widely used Markov Chain Monte Carlo methods. The initial concept was presented in Metropolis et al. (1953) and extended later by Hastings (1970). The rudimentary idea supporting this algorithm is to construct a Markov chain whose stationary distribution is the desired posterior distribution of θ. The distribution of the simulated chain approximates the stationary distribution after sufficiently many iterations. A Markov chain means a sequence of dependent random variables $\{\theta_1, \theta_2, \ldots, \theta_t\}$ such that the current state of θ_t conditional on the previous state of θ_{t-1} is independent to all previous states of θ. When performing the actual implementation, this chain starts with an initial value of $\theta^{(0)}$ and a single transition from θ_{t-1} to θ_t is made as follows:

- A candidate θ^* is drawn from some proposal density $q(\theta^*|\theta)$ given the current state θ of the chain (in theory, the proposal distribution can be chosen arbitrarily; however, performance depends on the choice of this proposal distribution);
- Compute the ratio

$$\alpha(\theta, \theta^*) \equiv \min\left\{\frac{p(\theta^*|\mathbf{y})q(\theta|\theta^*)}{p(\theta|\mathbf{y})q(\theta^*|\theta)}, 1\right\}, \tag{2.7}$$

- Accept θ^* as the current state of the Markov chain with a probability α; otherwise, $\theta_t = \theta_{t-1}$.

This update is repeated for m times until the Markov chain appears to stabilize at a distribution (i.e., stationary), and is continued n more times. The first m elements of the full sequence of length $m + n$ is referred to as the burn-in period. After discarding the burn-in period, the sequence of $\{\theta_{m+1}, \theta_{m+2}, \ldots, \theta_{m+n}\}$ is sufficient to approximate the posterior expectation $E(h(\theta)|\mathbf{y})$ in (2.2) such that

$$E(h(\theta)|\mathbf{y}) \approx \frac{1}{n} \sum_{t=m+1}^{m+n} h(\theta_t). \tag{2.8}$$

The main advantage of the MH algorithm is that the normalizing factor (i.e., the marginal distribution of $\mathbf{y}$) is not necessarily evaluated for computing the posterior summary of interest. However, the rate of convergence of a Markov chain to a stationary distribution depends on the choice of proposal distribution; thus, it could take a long time for an arbitrary proposal distribution. Introductory details about the Metropolis-Hastings algorithm and its implementation can be found in Chib and Greenberg (1995).

2.3.2 Importance Sampling

Importance sampling is a useful sampling-based numerical method for approximating an integral in (2.2) when direct sampling is difficult from the posterior distribution $p(\theta|\mathbf{y})$ in (2.1). For purposes of notational convenience, herein, $p(\theta|\mathbf{y})$ is denoted simply by $p(\theta)$. The central idea behind the importance sampling comes from an alternative representation of (2.2) as shown below:

$$E[h(\theta)] = \int h(\theta)p(\theta)d\theta = \int h(\theta)\frac{p(\theta)}{g(\theta)}g(\theta)d\theta = \int h(\theta)w(\theta)g(\theta)d\theta, \tag{2.9}$$

where $g(\theta)$ is the proposal density function whose support is identical to the support of $p(\theta)$, and $w(\theta)$ is a weight function. That is, the implication of (2.9) is that (2.2) can be rewritten as the expectation of $h(\theta)$ with respect to $g(\theta)$ from which it is easy to draw samples.

Once an importance sample of size n is simulated from $g(\theta)$, then the estimate at the posterior expectation is assessed by

$$E[h(\theta)] = \frac{\sum_{i=1}^{n} h(\theta_i)w(\theta_i)}{\sum_{i=1}^{n} w(\theta_i)}. \tag{2.10}$$

One may experience a numeric overflow error which means that the computed value is greater than the maximum allowed by the architecture of the system. To avoid this issue, the posterior expectation $E[h(\theta)]$ in (2.10) is revised as below:

$$E[h(\theta)] \quad = \quad \frac{\sum_{i=1}^{n} h(\theta_i) \exp(\log w(\theta_i) - M)}{\sum_{i=1}^{n} \exp(\log w(\theta_i) - M)}, \tag{2.11}$$

where $M = \arg\max_{\theta} \log(w(\theta_i)) = \arg\max_{\theta} [\log p(\theta_i) - \log g(\theta_i)]$. Another concern on the implementation of importance sampling is that the quality of the importance estimate depends on the choice of a suitable sampling density of $g(\theta)$. A discussion and an exercise on the choice of $g(\theta)$ are noted in Section 3.3.3 in the book by Robert and Casella (2009).

2.3.3 Laplace Approximation

The Laplace method is a well-known analytic approach for evaluating an integral, and the approximation of the posterior expectation in (2.2) can be viewed as its application in Bayesian inference. Details of this approximation are well described in the study by Tierney and Kadane (1986). The fundamental technique required in order to use this approximation method is to rewrite an integral I into a special form $\int \kappa(\theta) \exp[-ng(\theta)]d\theta$ which can be approximated by the first order Taylor approximation at the maximum likelihood estimate $\hat{\theta}$ such that

$$I = \kappa(\hat{\theta}) \exp[-ng(\hat{\theta})] \sqrt{\frac{2\pi}{n|g''(\hat{\theta})|}}, \tag{2.12}$$

where $\kappa(\theta)$ and $g(\theta)$ are three times differentiable smooth functions.

The posterior expectation of $h(\theta)$ in (2.2) can be approximated by setting $-ng_1(\theta) = \log h(\theta) + \log \mathscr{L}(\theta) + \log \pi(\theta)$ for the numerator and $-ng_0(\theta) = \log \mathscr{L}(\theta) + \log \pi(\theta)$ for the denominator. However, it is necessary that $h(\theta)$ be positive. The resulting approximation

for $E[h(\theta|\mathbf{y})]$ is:

$$\hat{E}[h(\theta|\mathbf{y})] = \left(\frac{|\Sigma_1|}{|\Sigma_0|}\right)^{1/2} \exp\left[ng_1(\hat{\theta}_1) - ng_0(\hat{\theta}_0)\right], \tag{2.13}$$

where $\hat{\theta}_1$ and $\hat{\theta}_0$ are the maximum likelihood estimates of $-ng_1(\theta)$ and $-ng_0(\theta)$, respectively, and Σ_1 and Σ_0 are minus the inverse Hessian of $-ng_1(\theta)$ and $-ng_0(\theta)$ at $\hat{\theta}_1$ and $\hat{\theta}_0$.

2.3.4 Linear programming

Linear programming is concerned with the problem of optimization subject to constraints. The objective function $l(\cdot)$ that needs to be minimized (or maximized) must be linear, and constraints also need to be expressed as linear equations or inequalities. Thus, the minimization problem in linear programming is formalized as follows:

$$\min_{\mathbf{x}} l(\mathbf{x}) = a_1x_1 + \cdots + a_kx_k$$

subject to the constraints

$$a_{11}x_1 + \cdots + a_{1k}x_k \geq b_1$$

$$a_{21}x_1 + \cdots + a_{2k}x_k \geq b_2$$

$$\cdots$$

$$a_{m1}x_1 + \cdots + a_{mk}x_k \geq b_m$$

and $x_1 \geq 0, \ldots, x_k \geq 0$. Of primary importance is to test a number of extreme points of a convex polyhedron to see whether or not they are optimal. The polyhedron represents a solution set of a finite number of linear inequalities, and serves as a feasible region for $l(x)$.

More than one linear programming function is available in `R` (R Development Core Team, 2011). The `constrOptim()` function in the `stats` package (R Development Core Team, 2011) is designed for the purpose of examining constrained optimization problems. The general-purpose `optim()` function in the `stats` package provides for a box-constrained optimization. Braun and Murdoch (2008) also noted that "the `lp()` function in the `lpSolve` package (Berkelaar and others, 2013) may be the most stable version currently available" (p. 145). In fact, this linear programming problem can be viewed as a special case of a nonlinear constrained optimization problem. Such problems are known to be very challenging to solve.

2.4 Zero-Truncated Poisson Model

Suppose Y is a standard Poisson random variable with mean $\mu > 0$ whose a probability mass function is given by

$$f(y|\mu) = \frac{1}{y!}e^{-\mu}\mu^y \tag{2.14}$$

for $y = 0, 1, 2, \ldots$. This Poisson random variable Y has the expectation $E(Y) = \mu$ and a variance $V(Y) = \mu$. The probability mass function of a zero-truncated Poisson is obtained as shown in (2.15) by eliminating the possibility of a zero value $f(0) = e^{-\mu}$ from the sample space of the standard Poisson distribution in (2.14) and renormalizing the remaining densities of Y by their sum:

$$f(y|y > 0, \mu) = \frac{e^{-\mu}\mu^y}{y!(1 - e^{-\mu})}, \tag{2.15}$$

for $y = 1, 2, \ldots$ (Cameron and Trivedi, 1998, p. X). The mean $E(Y|Y > 0)$ and variance $V(Y|Y > 0)$ of the zero-truncated Poisson model are:

$$E(Y|Y > 0) = \frac{\mu}{1 - e^{-\mu}}, \quad \text{and} \quad V(Y|Y > 0) = \frac{\mu(1 - e^{-\mu} - \mu e^{-\mu})}{(1 - e^{-\mu})^2}. \tag{2.16}$$

The derivation of the first three central moments using the moment generating function are noted in Appendix B.1. As shown in (2.16), $E(Y|Y > 0)$ is always larger than $E(Y)$ and $V(Y|Y > 0)$ is under-dispersed (i.e., the variance is less than the mean) because of $0 < 1 - e^{-\mu} < 1$.

It is a straightforward extension of the zero-truncated Poisson model in (2.15) to a regression model that accounts for the effect of the explanatory variables. Following the framework described in Section 2.2, a different mean parameter μ_i of the zero-truncated Poisson for i-th individual is linked to a linear predictor of $p+1$ explanatory variable $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})^T$ using the log link function:

$$E(y_i|\mathbf{x}_i) = \mu_i = \exp(\mathbf{x}_i^T\boldsymbol{\beta}), \tag{2.17}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a vector of unknown regression parameters. The log link function ensures that μ_i is strictly positive for a proper distribution. The parameter $\boldsymbol{\beta}$ can be estimated by maximizing the log of a zero-truncated Poisson likelihood using Newton's

method. The required score vector and Hessian matrix for this estimation procedure are noted in Appendix A.3 and also can be found in Gurmu (1991).

Once the regression parameters $\boldsymbol{\beta}$ are estimated by the maximum likelihood method, the Horvitz-Thompson estimator derived by van der Heijden et al. (2003) is applied to the problem of estimating the unknown size N of a target population:

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{1 - e^{-\hat{\mu}_i}}, \tag{2.18}$$

where $\hat{\mu}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ is the mean parameter of a zero-truncated Poisson model that ultimately produces the probability of being ascertained by the identification mechanism for the i-th individual in that population (Böhning and van der Heijden, 2009; Cruyff and van der Heijden, 2008). The confidence interval of the estimated population size is derived based on the asymptotic normal distribution.

According to the simulation study performed by van der Heijden et al. (2003, Section 3), a comprehensive coverage level is shown in either small populations or small capture probabilities. However, their simulation study is based on a homogeneous population using a model having only an intercept. The statistical properties of this estimator has not been explored regarding bias, precision, or asymptotic distribution of the estimated population size. Due to the presence of an unobserved heterogeneity encountered when applying the zero-truncated Poisson regression model to the application of estimating the size of the illegal immigrant population in The Netherlands, this issue of underestimation must be addressed. Note that the Lagrange multiplier test proposed by Gurmu (1991) is used for testing the presence of the unobserved heterogeneity.

Chapter 3

Modelling Complete Prior Ignorance

In this chapter the author provides details of a methodology characterized by a canonically parametrized imprecise inferential framework that has been developed for the quantification of epistemic ignorance on the estimation of a sampling model parameter. The name 'Imprecise inferential framework' arose from both the imprecise probabilities theory presented by Walley (1991), and because the canonical parameter of a sampling model plays a central role in this inferential process.

The methodology presented here is not a newly introduced inferential framework. As noted in Section 1.5, the imprecise inferential framework referred to as the imprecise Beta-Binomial model has been studied by Walley (1991), Coolen (1994) and Walley et al. (1996), and its generalized version called the imprecise Multinomial-Dirichlet model has been also studied by Walley (1996) and Bernard (2005). The PhD thesis work by Quaeghebeur (2009) is also a comprehensive reference that demonstrates a generalization of the imprecise inferential framework to an exponential family of sampling distributions. In particular, a one-parameter exponential family of sampling distributions is well studied in Benavoli and Zaffalon (2012). For the sake of convenience, the inferential framework presented in their studies is referred to as the conventional approach to an imprecise inference and the one presented here is simply referred to as our approach or the proposed methodology in the remainder of this thesis.

A number of distinctions are made in this proposed methodology as compared to the conventional approach of imprecise inference. These distinctions can be elucidated from the detailed description of the proposed methodology provided in Section 3.1. However, this description is limited to Poisson sampling models because of the research interest presented in Chapter 1. These details also serve as a blueprint of the `R` (R Development Core Team, 2011) `ipeglim` package (Lee and Bickis, 2013) that is a collection of functions developed by

the author for producing all numeric and graphic summaries contained this thesis. A brief introduction to the key functions in this package are provided in Section 3.2, and a complete description is available in the documentation of the package which is available from the R-forge. Note also that this thesis written by the `Sweave` system (Leisch, 2002) that allows to embed `R` codes within LaTeX documents. For easy exposition of the proposed methodology, a number of examples are provided in the last Section 3.3 of this chapter. The examples illustrate six major features of the canonically parametrized imprecise inferential framework.

3.1 Canonically Parametrized Imprecise Inferential Framework for Poisson Data

Suppose that the past and future observations y are independently and identically distributed realizations drawn from a standard (untruncated) Poisson sampling model $f(y) = e^{-\mu}\mu^y/y!$ with a mean parameter μ. Assume that *you* are a person who is interested in estimating this Poisson mean parameter μ but *you* have not seen this data yet. (The italic font is used for emphasizing an intentional unit for this inference.)

3.1.1 Exponential Family Representation of Sampling model

The first task for the proposed methodology is to represent a specified sampling model in the form of a natural exponential family of distributions. Since the standard Poisson distribution is an instance of a one-parameter exponential family, the sampling model $f(y)$ is written as follows:

$$f(y|\theta) = h(y)\exp\{y\theta - \mathcal{A}(\theta)\}, \tag{3.1}$$

where $\theta = \log(\mu)$ is the canonical parameter of the Poisson mean parameter μ which is canonically parametrized by a log-link function, $h(y) = 1/y!$ is a base counting measure of y, and $\mathcal{A}(\theta) = e^{\theta}$ is a log-normalizer that integrates the unnormalized densities of (3.1) over the sample space $y \in \mathscr{Y}$ to 1 such that:

$$\mathcal{A}(\theta) = \log \sum h(y)\exp\{y\theta\}. \tag{3.2}$$

The set $\Theta = \{\theta | -\infty < \theta < \infty\}$ is the canonical parameter space of θ, and the corresponding mean parameter space of μ is $\mathscr{M}_{=}\{\mu | 0 < \mu < \infty\}$. Once a sample of size n is observed, the corresponding Poisson likelihood $\mathscr{L}(\theta|\mathbf{y})$ of (3.1) has the form:

$$\mathscr{L}(\theta|y) \propto \exp\{n\bar{\mathbf{y}}\theta - n\mathcal{A}(\theta)\}, \tag{3.3}$$

where $\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is a sample mean of $\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$.

3.1.2 Conjugate Formulation of Prior Measure

The second task of the proposed methodology is to formulate a conjugate prior measure $\pi(\theta)$ to the Poisson likelihood $\mathscr{L}(\theta|\mathbf{y})$ as shown below

$$\pi(\theta|\xi) \propto \exp\{\xi_1\theta - \xi_0\mathcal{A}(\theta)\} \tag{3.4}$$

where $\xi = (\xi_1, \xi_0)^T$ is a vector of hyperparameters. If a normalizing constant that makes a prior measure $\pi(\theta|\xi)$ in (3.4) integrate to one can be found, (3.4) is then a prior probability measure $\pi(\theta)$ on the canonical parameter space Θ. This log-normalizing constant is given by

$$\mathcal{B}(\xi) = \log \int \exp\{\xi_1\theta - \xi_0\mathcal{A}(\theta)\} d\theta. \tag{3.5}$$

The set $\boldsymbol{\Xi} = \{\xi | \mathcal{B}(\xi) < \infty\}$ is thus defined as the canonical hyperparameter space (simply, called hyperparameter space) of the prior probability measure $\pi(\theta|\xi)$.

This conjugate formulation of a prior measure $\pi(\theta)$ also provides a mathematical convenience for the formulation of a posterior distribution $p(\theta|\mathbf{y})$ once a sample of size n is available as shown below:

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\{\xi_1'\theta - \xi_0'\mathcal{A}(\theta)\}, \tag{3.6}$$

where

$$\xi_1' = \xi_1 + n\bar{\mathbf{y}}, \quad \text{and} \quad \xi_0' = \xi_0 + n. \tag{3.7}$$

That is, a posterior distribution $p(\theta|\mathbf{y})$ is closed under the sampling distribution $f(y|\theta)$. If the kernel of this posterior distribution $p(\theta|\mathbf{y})$ in (3.6) has a known parametric form in some standard probability distributions, a simplified form of posterior summaries of interest are easily found. A legitimate numerical approximation of the quantities of interest is also expected when such a simplified form is available since the effect of cumulative rounding errors in computation requiring a long sequence of processing to reach the estimate would be eliminated.

Before moving forward to the next task of this proposed methodology, the role of the hyperparameter space $\boldsymbol{\Xi}$ needs clarification in order to expedite comprehension of subsequent tasks for the reader since the hyperparameter space $\boldsymbol{\Xi}$ serves as the actual space on which the modelling procedure for prior ignorance is carried out in this proposed methodology. Details of this modelling procedure are noted in section 3.1.3; however, the resultant advantage of working with this hyperparameter space $\boldsymbol{\Xi}$ over the mean parameter space $\mathscr{M}$ is briefly discussed in the next paragraph.

Consider the case where the actual prior ignorance is modelled for imprecise inference. Prior ignorance has been modelled by specifying a convex region on the mean parameter space $\mathscr{M}$ in literature on the conventional imprecise inferential approach (Quaeghebeur, 2009; Benavoli and Zaffalon, 2012). For the case of a standard Poisson sampling model, the dimensionality of the mean parameter space $\mathscr{M}$ is a line segment. *You* may have no conceptualization that can delineate some constraint on this line segment other than specifying the upper and lower bounds. However, *you* have more than one choice to give such constraints when working with a two-dimensional hyperparameter space. For instance, a triangle, square, pentagon, or other types of polygons can be formulated. This flexibility of prior ignorance modelling is illustrated throughout the example in Section 3.3.1 .

3.1.3 Characterization Strategy for Prior Ignorance

The third task of the proposed methodology is the characterization of *your* prior ignorance. The term characterization is intended for implying *your* modelling strategy of how to represent *your* prior ignorance in the form of a mathematical language on the hyperparameter space. One may view this characterization task as a prior elicitation process in standard

Bayesian analysis. As briefly noted in the previous section 3.1.2, this mathematical representation of *your* prior ignorance is typically described in the form of a closed and bounded convex region when the hyperparameter space has two dimensions. The number of hyperparameters ξ associated with a prior probability measure $\pi(\theta|\xi)$ is not limited to only two. Three parameter and multi-parameter exponential families of distributions are dealt with in Chapter 4; thus, in this subsection the third task of this proposed methodology is described by assuming that the hyperparameter space $\boldsymbol{\Xi}$ is in p dimensions.

The suggested characterization strategy is to set a convex hull characterized by the extreme points which are found from the solutions of a finite system of k linear inequalities $g_i(\boldsymbol{\xi})$ of hyperparameters $\boldsymbol{\xi}$ such that

$$\begin{aligned} g_1(\boldsymbol{\xi}) = a_{11}\xi_1 + \cdots + a_{1p}\xi_p &\geq c_1 \\ g_2(\boldsymbol{\xi}) = a_{21}\xi_1 + \cdots + a_{2p}\xi_p &\geq c_2 \\ g_k(\boldsymbol{\xi}) = a_{k1}\xi_1 + \cdots + a_{kp}\xi_p &\geq c_k, \end{aligned}$$

where $\boldsymbol{\xi}$ is a p-dimensional vector of hyperparameters $(\xi_1, \xi_2, \ldots, \xi_p)^T$, a_{ij} $(i = 1, \ldots, k$ and $j = 1, \ldots, p)$ is the coefficient of j-th hyperparameter ξ_j in the i-th constraint $g_i(\boldsymbol{\xi})$ in the left hand side of the inequalities, and c_i is the constant of i-th given on the right hand side of the inequalities. For instance, a convex polytope can be characterized as a region $\mathscr{R}$ describing *your* prior ignorance on the hyperparameter space $\boldsymbol{\Xi}$ of ξ_1 and ξ_0 each of which represents a total sum of the count and a number of samples prior to seeing the data. Similarly, a convex polyhedren can be characterized in a three-dimensional hyperparameter space $\boldsymbol{\Xi}$. The number of constraints is arbitrarily determined by *your* known information. This characterization strategy is illustrated together with an example showing the flexibility of prior ignorance modelling in section 3.3.1

Since every single point in the characterized convex hull on the p-dimensional hyperparameter space $\boldsymbol{\Xi}$ maps to a single prior distribution $\pi(\theta|\xi)$ with a one-to-one relationship, a set of points encompassed by line segments connecting the extreme points of a convex hull implies a class of prior distributions that describes *your* prior ignorance. The term of imprecise is used from this point forward to imply a set of distributions; thus, this class of

prior probability distributions will be referred to as an imprecise prior and its corresponding convex hull on the hyperparameter space is denoted by $\mathscr{C}_0$. In the same manner, a set of posterior probability distributions which is updated from an imprecise prior by applying Bayes' theorem after observing new samples will be referred as to an imprecise posterior and its corresponding convex hull on the hyperparameter space is denoted by $\mathscr{C}_n$.

From the equations in (3.7) one may notice that the update of an imprecise prior to an imprecise posterior can be geometrically described by the translation of a convex hull $\mathscr{C}_0$ since all points (ξ_1, ξ_0) in the region (i.e., the convex hull $\mathscr{C}$ in a two dimensional hyperparameter space) moves to the points (ξ_1', ξ_0') by $n\bar{\mathbf{y}}$ and n along with ξ_1 and ξ_0 simultaneously as new data $\mathbf{y}$ is gathered. This translation of a convex hull $\mathscr{C}_0$ is illustrated using the examples in Subsection 3.3.2.

3.1.4 Imprecise Posterior Optimization

The fourth task in the proposed methodology is to quantify the amount of epistemic ignorance lying on the estimation of a sampling model parameter. Walley (1991) referred this quantity as the degree of imprecision and defined this as the difference between two extreme posterior expectations over all possible posterior expectations each of which is evaluated at every single posterior distribution belonging to an imprecise posterior. When two extreme posterior expectations agree (i.e., the difference of two extreme posterior expectations is zero), the parameter of the sampling model is estimated precisely. Since the canonical parameter θ is the quantity of interest to be estimated in the proposed methodology, the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ is defined as

$$\Delta_n(\theta|\mathbf{y}) = \overline{E}_n(\theta|\mathbf{y}) - \underline{E}_n(\theta|\mathbf{y}), \tag{3.8}$$

where $\underline{E}_n(\theta|\mathbf{y})$ is the minimum posterior expectation, $\overline{E}_n(\theta|\mathbf{y})$ is the maximum posterior expectation, and n is the sample size which is used for updating an imprecise prior. The problem is then to search for the extreme posterior distribution that minimizes or maximizes $E_n(\theta|\mathbf{y})$ over all posterior distributions belonging to an imprecise posterior.

In fact, the second and third tasks in the proposed methodology are the preparatory work for this optimization problem. The second task builds a geometric space of hyperparameters in which each single point maps to a single prior distribution with a one-to-one relationship. The third task characterizes a convex polytope on the hyperparameter space over which the extrema of the posterior expectation are to be found. The problem can thus be restated as a typical constrained optimization problem of searching for an optimal solution of hyperparameters $\xi = (\xi_1, \xi_2, \ldots, \xi_p)^T$ that minimizes or maximizes a given posterior expectation $E(\theta|\mathbf{y})$ subject to a set of linear inequality constraints on the hyperparameter space $\boldsymbol{\Xi}$. If the posterior expectation $E(\theta|y)$ is a linear function of hyperparameters ξ, then this optimization becomes simpler since the maximum or minimum posterior expectation occurs at one of the extreme points of the convex polytope (Feiring, 1986, p. 30). This imprecise posterior optimization is illustrated using the examples in Subsection 3.3.5.

3.2 R Package Development

The proposed methodology has been theoretically developed with the hope of being computationally efficient for an imprecise inference. Arriving at an imprecise estimate for a quantity of interest is a complicated procedure due to the complex modelling procedure associated with the aspect of the procedure involving the characterization of a prior ignorance and a long sequence of computations associated with the imprecise posterior optimization procedure. It is important to simplify these two steps in order to explore the major features of the proposed methodology and the behaviours of imprecise estimates.

To resolve these practical concerns when using the proposed methodology, the author of this thesis initiated a project for developing the `R` package entitled `ipeglim` (Lee and Bickis, 2013). This project is hosted at `https://r-forge.r-project.org/projects/ipeglim/`. The key implementations used to produce the numerical results and graphical portrayals presented in this thesis work are explained in this section.

The following eight components are essential to carrying out the proposed canonically parametrized imprecise inferential framework:

1. Defining a sampling model;

2. Defining a family of prior distributions;
3. Solving a set of given linear inequalities;
4. Searching for extreme points of a characterized convex hull;
5. Applying Bayes' theorem at every extreme point;
6. Evaluating a quantity of posterior expectation at every extreme point;
7. Identifying extreme posterior expectations;
8. Summarizing the resulting imprecise estimate numerically and graphically.

The `R` `ipeglim` package provides six primary functions to perform the tasks listed in the above: `model()`, `iprior()`, `update()`, `summary()`, `plot()`, and `pbox()`. From this point forward, all `R` functions used in the reminder of this thesis work are indicated by `teletype font` and the parentheses which are used immediately following the function name.

A help page is provided for every function implemented in the `ipeglim` package. Items documented in the help page are the basic description, usage, a list of all available arguments, details, returning values, references, etc. To see the details of major functions described in this section, please type `help(name)` or simply `?name` in the `R` console. Since the `ipeglim` package is frequently updated as program improvements are made (i.e. more efficient algorithm becomes available, software user convenience is improved, etc.), the details described in the help page may differ slightly from the description provided in this section. Hence, only details necessary for this thesis work are noted.

3.2.1 `model()`

The function `model()` defines a sampling model $f(\mathbf{y})$ for input data $\mathbf{y}$, and returns the class object `imprecise` with a list of response vector, data matrix, statistics produced by the maximum likelihood method, and other information needed for advancing results utilizing the proposed canonically parametrized imprecise inferential framework.

Note that the term class describes a template that produces an object in the context of object-oriented programming. Roughly speaking, this template contains two pieces of information: 1. the value that an object initially has, and 2. the methods which are a predefined implementation of how a function call should be evaluated with that object. Two types

of internal systems called S3 (S version 3) and S4 are offered in R. S4 is a class centered formal (or new) system like Java or C++ while S3 is a function centered informal (or old) system which is less compliant when defining class and methods as compared to S4. For a object-oriented programming style, R also supports a generic function mechanism which means a function whose predefined implementation (i.e., method) depends on the type (i.e., class) of the first argument supplied. This process is also called method dispatch. Hence, the class `imprecise` created by calling the function `model()` facilitates this method dispatch for selecting an appropriate function that performs computations needed to move forward the subsequent steps in the imprecise inferential process. For defining the class `imprecise` and the methods associated with this class, the S3 system was utilized in the development of `ipeglim` package in consideration of a quick change of implementation reflecting the diversity of a methodological development and the efficient adoption of better implementations or techniques intended to extend the utility of the package.

Syntax

```
model(formula, data, dist="poisson", ztrunc=FALSE, verbose=TRUE)
```

The `formula` parameter is assigned a symbolic description of a linear predictor in the context of a generalized linear model for model specification. The basic usage of `formula` follows the R convention. For example, `formula=y~x1+x3` implies the specification of a regression model that has a single response variable `y` and two explanatory variables `x1` and `x3`. By default, an intercept term is included. One exception concerning the use of the `formula` parameter in `model()` is the specification of a model such as `formula=y~0` which means the specified model does not involve any explanatory variables.

The second parameter `data` is the name of the environment where the variables used in the parameter `formula` are located. If the names of variables are not found in the `data`, variables for model specification are taken from the environment where the `model()` is called. The parameter `data` receives a user's data in the form of `vector` or `data.frame` (i.e., a matrix form of data entries for statistical analysis in R).

The third parameter `dist` specifies a family of sampling models which belong to an exponential family of distributions. The option `dist="poisson"` is used for this thesis work. The other values (i.e., `binom`, `exp`, `geo`, `nbinom`, `multinom`) of this option are not available but are defined for the purpose of future development.

The fourth parameters `ztrunc` is an option to change a type of a sampling model in the same family. That is, `ztrunc=TRUE` implies the use of zero-truncated Poisson sampling model and `ztrunc=FALSE` implies the use of a standard Poisson sampling model (without zero-truncation).

3.2.2 `iprior()`

The generic function `iprior()` defines an imprecise prior by obtaining information concerning constraints which are required in order to characterize a convex hull on the hyperparameter space, and returns the class object `imprecise` with a list of information about Cartesian coordinates for every extreme point with its identification number and the object produced from the use of `model()`.

Syntax

```
iprior(obj, eqns=list(lhs, rhs), circle=list(x,y,z,r,len))
```

The argument `obj` passes the object of class `imprecise` produced from `model()` for dispatching the method `iprior`. The methods defined in the function `iprior()` behave differently depending on contemporaneous expression of another/other argument(s). If not found, the default method is to use information passed by the one of following arguments:

- `eqns=list(lhs, rhs)` obtains a set of linear inequality constraints which are to be defined in the form of `lhs %*% xi >= rhs` as shown in (3.8), where `xi` is a vector of hyperparameters. The first element `lhs` of the argument `eqns` is a matrix containing the coefficients of hyperparameters `xi` in the constraints. Each row of the matrix corresponds to a single constraint. The second element `rhs` of the argument `eqns` is a vector containing the constant given on the right-hand side of the constraints. For instances where the dimensionality of a hyperparameter space is greater than or

equal to three, the searching algorithm for extreme points depends on the function `convxhulln()` in the `geometry` package (Barber et al., 2013). If this requirement is not met, a naive searching algorithm implemented by the author is employed. There is no restriction governing the maximum number of constraints.

- `circle=list(x,y,z,r,len)` determines a set of points that lie at a given radius `r` from the center `(x,y)` when a hyperparameter space is defined in a two dimensional plane. A circular object (i.e., sphere) in a three dimensional hyperparameter space can be generated with a given radius `r` from the center defined by `(x,y,z)`. This argument `circle` is used for instances where the number of extreme points is infinite; thus, the last element `len` in this argument determines the number of points needed to characterize a circular object in a two- or three-dimensional hyperparameter space. By default, `x=y=z=r=1` and `len=15`. The author of this thesis work determined by exhaustive experimentation that a default number of 15 extreme points is required to approximate a visual representation of a circular object in either two or three dimensions.

The usage of these options with the function `iprior()` is illustrated in section 3.3.1. Two additional arguments `x` and `mat` are defined in the function `iprior()`. The argument `x` is intended to derive a vector of hyperparameters for specifying a single precise prior distribution. The argument `mat` is intended to have a direct input matrix of extreme points (without a searching process) to create a convex hull. An intense examination of these two arguments `x` and `mat` is currently being performed by the author for extending the current package.

3.2.3 `update()`

The function `update()` applies Bayes' rule to every single prior distribution belonging to an imprecise prior with a given data set, and returns the object of class `imprecise` with a list of the resultant imprecise posterior expectations and the object produced from the previous use of `iprior()`.

Syntax

```
update(obj, ..., [B,] [apriori=c("lgamma", "normal"),]
  [method=c("LA", "MH", "IS", "AQ"),] [control=list(),] [proposal=list()])
```

The argument `obj` secures the object of class `imprecise` produced from the previous use of the function `iprior()`. This `obj` is ultimately passed to the function `cpef()` if the symbolic expression in the argument `formula` does not involve explanatory variables (for the sake of convenience, such models are referred to as non-regression models). If not, the `obj` is passed to the function `cpef2reg()` for regression models. This determination is internally assessed by retrieving information of the data matrix `X` which means a data frame that consists of an intercept and a set of explanatory variables. Note that when the function `model()` is used, the input data passed by the argument `data` is decomposed to a response variable `y` and a data matrix `X` by matching the variable names specified in the argument `formula` to the list of names in the argument `data`.

The second argument `...` (three dots) in the function `update()` passes the following arguments needed to call the functions `cpef()` and `cpef2ref()`:

- The argument `apriori` specifies the name of a family of prior distributions. For this thesis work, two options of `"lgamma"` and `"normal"` are offered for non-regression models. `"lgamma"` and `"normal"` represent families of log-gamma and normal prior distributions, respectively. However, the current implementation supports only option `"normal"` for a p-dimensional multivariate normal distribution for regression models.

- The argument `method` specifies the name of the numerical methods required for the approximation of a quantity of a posterior expectation with a given family of prior distributions and input data. The Metropolis-Hastings algorithm (`"MH"`), the Importance sampler (`"IS"`), and the Laplace approximation (`"LA"`) are offered. The adaptive quadrature (`"AQ"`) is also supported for non-regression models. A brief discussion about the author's experience regarding the use of different numerical methods for this thesis work is given in Section 4.2.

- The argument `control=list()` is the set of variables that permits the user to change the default values that are supplied to each of the numerical methods. For example, when using the Metropolis-Hastings algorithm, the user needs to determine the length of Markov chain and burn-in period, as well as the proposal distribution (the current implementation uses a normal distribution with a mean 0 and a standard deviation that is estimated from input data by default). Instructions detailing how to change the default values on those specifications using the argument `control=list()` are noted in Section 4.2. It is necessary to use the argument `proposal=list()` in order to change the specification of a proposal distribution. These two arguments `control=list()` and `proposal=list()` are also used for the numerical method `"IS"` (Importance sampling) in the same way.

- The argument `B` acquires a p-by-p square matrix that represents the specification of a prior variance-covariance matrix for regression models. If not found, `B` defaults to the p-dimensional identity matrix. The value of p is determined by the number of explanatory variables including an intercept term as specified in the argument `formula` when the function `model()` is used.

3.2.4 `summary()` and `plot()`

The generic functions `summary()` and `plot()` produce numerical and graphical summaries of a resultant imprecise posterior from various model fittings performed with the canonical parametrized imprecise inferential framework.

Syntax

```
summary(obj, HT.est=FALSE, ...)
plot(obj, rm.M0=FALSE, xi, ...)
pbox(obj, pretty=TRUE, ... [, control=list(beta, xtms)])
```

The argument `obj` in the functions `summary()`, `plot()`, and `pbox()` is used to acquire the object of class `imprecise` produced from the previous use of the function `update()`.

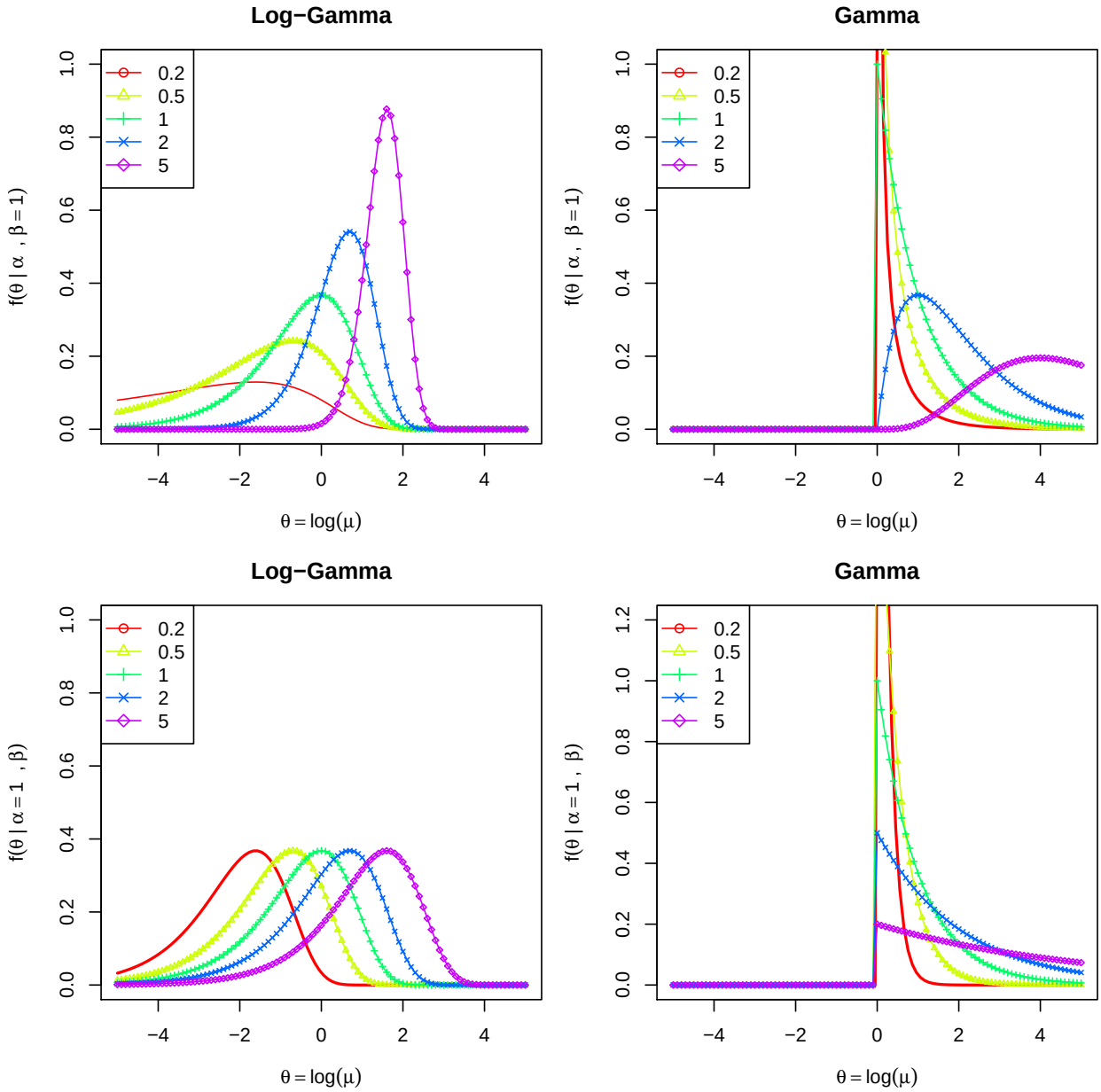
The function `summary()` returns a list of the objects containing all of the information specified during the processing of the proposed methodology for controlling the imprecise inference, the imprecise posterior expectation (complete with the identification numbers of the extreme points), the maximum and minimum imprecise posterior expectations (complete with the identification numbers of the extreme points), and the degree of imprecision, and so on. If the argument `ztrunc` is set by `TRUE` when the function `model()` is used, the optional argument `HT.est=TRUE` computes the Horvitz-Thompson estimator for the population size estimation problem.

The function `plot()` produces various graphical figures depending on the number of hyperparameters used, the type of constraints characterized, and the stage of the imprecise inferential framework. Once the object of class `imprecise` is produced using `iprior()`, the resultant convex hull characterized by the set of given linear inequality constraints is plotted. The reader can examine various shapes of this convex hull from Figures 3.2 and C.1. When the object of class `imprecise` is produced from the use of `update()`, the surface plot of an imprecise posterior expectation over a two-dimensional hyperparameter space is produced for non-regression models. Please see Figure C.2. For regression models, the function `plot()` produces a scatter plot illustrating the change of imprecise probabilities before and after observing `data` as shown in Figure 4.14. The probability box shown in Figure 3.9 is offered by the function `pbox()`. Trellis graphics produced using the function `plot()` are dependent upon the specific `lattice` package (Sarkar, 2008) utilized.

3.3 Example with Log-Gamma Imprecise Prior

Examples are provided in this section to illustrate the main features of the proposed methodology. For this illustration, a family of log-gamma prior distributions is considered for the problem of estimating the canonical parameter of a standard Poisson sampling model. The choice of this family of prior distributions is motivated by the fact that a family of gamma distributions is conjugate to a Poisson likelihood in a standard Bayesian analysis with respect to the mean parameter.

Figure 3.1: Probability densities of a gamma distribution and a log-gamma distribution with the selected values of a shape parameter $\alpha = 0.2, 0.5, 1, 2, 5$ at a given rate parameter $\beta = 1$ on the top left and right, respectively. Probability densities of a gamma distribution and a log-gamma distribution with the selected values of a rate parameter shape parameter $\beta = (0.2, 0.5, 1, 2, 5)$ at a given shape parameter $\alpha = 1$ on the bottom left and right, respectively.



First consider a case where a standard Poisson mean parameter μ is distributed as a gamma distribution with a shape parameter α and a rate parameter β in order to observe the behaviour of a log-gamma distribution. The canonical parameter $\theta = \log(\mu)$ is then a log-gamma distribution of which a probability density function is given by

$$\pi(\theta|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} e^{\alpha\theta - \beta e^{\theta}}, \tag{3.9}$$

where $-\infty < \theta < \infty$, $\alpha > 0$ and $\beta > 0$ are the shape and rate parameters, respectively. The derivation of this log-gamma probability density function of canonical parameter θ from a gamma distribution of mean parameter μ, the moment generating function of (3.9), and the expectation of a canonical parameter θ with respect to $\pi(\theta|\alpha,\beta)$ in (3.9) are noted in Appendix A.1. In this parametrization, the shape α and rate β parameters are the canonical parameters of a log-gamma distribution in (3.9).

The difference between probability densities of log-gamma and gamma distributions are compared in Figure 3.1. Two plots on the top in these figures are produced by varying the values of the shape parameter α in the range of $\{0.2, 0.5, 1, 2, 5\}$ at the fixed value of the rate parameter $\beta = 1$. Two plots on the bottom are produced by varying the values of the rate parameter β in the range of $\{0.2, 0.5, 1, 2, 5\}$ at the fixed value of the shape parameter $\alpha = 1$. Note that the rate parameter β plays a role as a location parameter $\log\beta$ for a family of log-gamma distributions as illustrated on the plot in the bottom left in Figure 3.1. The function `dlgamma()` in the `ipeglim` package is used to compute probability densities of a log-gamma distribution with given values of the shape parameter α and the rate parameter β.

3.3.1 Flexible Prior Ignorance Modelling

The example presented in this subsection illustrates the flexible characterization of a convex hull on the hyperparameter space $\boldsymbol{\Xi}$ when an imprecise prior is elicited using the proposed methodology. Also presented is a demonstration of the usage of function `iprior()` for this characterization. Please note that from this point forward, a convex hull $\mathscr{C}_0$ is simply denoted by a region $\mathscr{R}_0$ since the dimensionality of the hyperparameter space considered in this family of prior distributions is two.

Consider the following four linear inequality constraints:

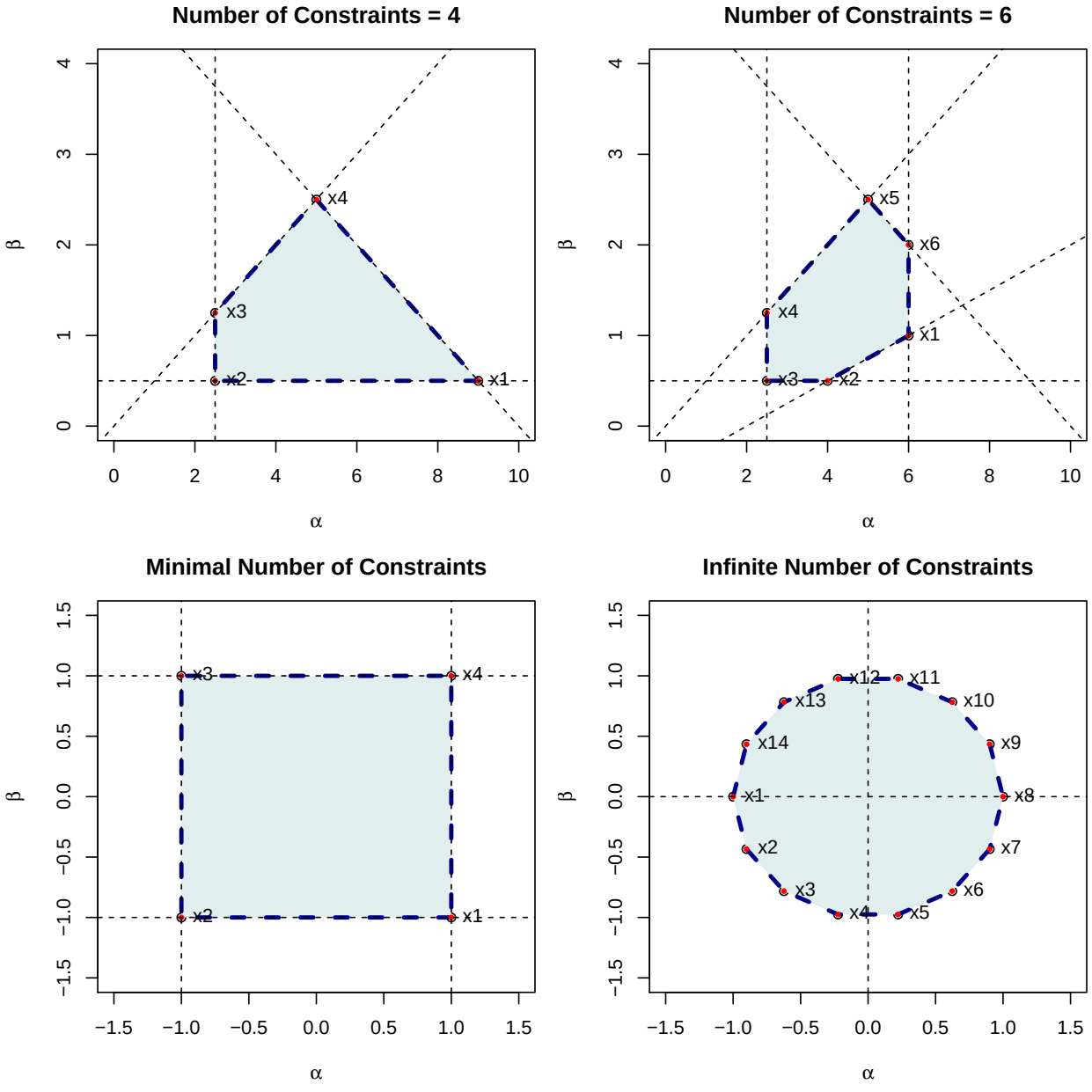
$$\beta \leq -\frac{1}{2}\alpha + 5, \quad \beta \leq \frac{1}{2}\alpha, \quad \alpha \geq 2.5, \quad \beta \geq 0.5, \tag{3.10}$$

where α and β are the shape and rate hyperparameters of a family of log-gamma prior distributions. From this system in (3.10) six points of intersection can be easily found manually but only four intersection points $(9.0, 0.50)$, $(2.5, 0.50)$, $(2.5, 1.25)$, $(5.0, 2.50)$ are required for the characterization of the region $\mathscr{R}_0$. The plot on the top left panel in Figure 3.2 illustrates this characterization. In this figure, the polytope coloured blue depicts the region $\mathscr{R}_0$ describing an imprecise prior that will be ultimately used for a further imprecise inference process. The extreme points located at the vertices of the polygons and identified by the numbers `x1`, `x2`, `x3`, `x4` that are assigned to each of these extreme points in a clockwise direction are shaded in red. These identification numbers are used for tracking each member of the imprecise prior in subsequent inferential steps.

However, it is impractical to search for these extreme points manually from a given system when the number of constraints is large. Consider two additional linear inequalities constraints $\alpha \leq 6$ and $\alpha - 4\beta \geq 2$ on the system given in (3.10). A total of 13 solutions can be found easily from the system of six equalities, but each solution needs to be examined to determine whether or not all inequalities are satisfied for identifying the extreme points. This searching task is conveniently accomplished using the function `iprior()` provided in the `ipeglim` package. In order to use this function, the system in (3.10) needs to be written in the following matrix form first.

$$\begin{bmatrix} -1 & -2 \\ 1 & -2 \\ 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \geq \begin{bmatrix} -10 \\ 0 \\ 2.5 \\ 0.5 \\ -6 \\ -2 \end{bmatrix}. \tag{3.11}$$

Figure 3.2: Different shapes of the region $\mathscr{R}_0$ characterized by linear inequality constraints: 1) using four constraints (on the top left): $\mathscr{R}_0 = \{(\alpha, \beta) | \beta \leq -\frac{1}{2}\alpha + 5, \beta \leq \frac{1}{2}\alpha, \alpha \geq 2.5, \beta \geq 0.5\}$; 2) using six constraints (on the top right): $\mathscr{R}_0 = \{(\alpha, \beta) | \beta \leq -\frac{1}{2}\alpha + 5, \beta \leq \frac{1}{2}\alpha, \alpha \geq 2.5, \beta \geq 0.5, \alpha \leq 6, \alpha - 4\beta \geq 2\}$; 3) using box-type constraints (on the bottom left): $\mathscr{R}_0 = \{(\alpha, \beta) | -1 \leq \alpha \leq 1, -1 \leq \beta \leq 1\}$; 4) using an infinite number of constraints (on the bottom right): $\mathscr{R}_0 = \{(\alpha, \beta) | (\alpha - 0)^2 + (\beta - 0)^2 \leq 1^2\}$.



The matrix on the left hand side and the vector in the right hand side in (3.11) are then used as the input required for the function `iprior()` as shown below:

```
> lc1 <- list(
+   lhs=rbind(c(-1,-2), c(1,-2), c(1,0), c(0,1), c(-1,0), c(-1,4)),
+   rhs=c(-10,0,2.5, 0.5, -6, -2)
+ )
> xtms1 <- iprior(eqns=lc1)
> xtms1

    V1   V2
x1 6.0 1.00
x2 4.0 0.50
x3 2.5 0.50
x4 2.5 1.25
x5 5.0 2.50
x6 6.0 2.00
```

The region $\mathscr{R}_0$ shown on the top right panel in Figure 3.2 is characterized by this generic `x-y` coordinate information of extreme points that is held in the object `xtms1`. For visualizing this characterization, the function `plot()` is used in the following way:

```
> plot(xtms1, xlim=c(0,10), ylim=c(0,5),
+      xlab=expression(alpha), ylab=expression(beta))
```

Two extreme cases of the region $\mathscr{R}_0$ (i.e. a convex hull in a two-dimensional hyperparameter space) are also presented on the bottom in Figure 3.2 based on the number of constraints. The region $\mathscr{R}_0$ shown on the bottom left panel is a convex hull characterized by a minimal number of constraints such that $\{(\alpha, \beta)| -1 \leq \alpha \leq 1, -1 \leq \beta \leq 1\}$, and the region $\mathscr{R}_0$ shown on the bottom right panel is one characterized by an infinite number of constraints such that $\{(\alpha, \beta)|\alpha^2 + \beta^2 \leq 1^2, (\alpha_0, \beta_0) = (0, 0)\}$.

3.3.2 Translation Behaviour of Imprecise Prior

The example presented in this subsection illustrates a translation behaviour of an imprecise prior in the proposed methodology. As noted previously in Section 3.1.2, this behaviour is derived from the formulation of conjugate prior measure. An appealing feature of this behaviour is that an information updating process (or learning process) can be visually demonstrated. Also presented is a characterization strategy that is suggested when the intentional unit is under a state of complete ignorance (Benavoli and Zaffalon, 2012, p. 1974).

In order to illustrate this translation behaviour, consider the past and future observations taken from a standard Poisson distribution with a mean of $\mu = 1$ which corresponds to the canonical parameter $\theta = 0$. The random seed `16979238` is used to simulate these Poisson random variates. Assume that *you* are the person who is interested in estimating the canonical parameter θ of this Poisson sampling model $f(y|\theta)$ using the proposed methodology. Assume also that *you* do not have any information regarding the algorithm nor the random seed for this random number generation. Hence, it is considered that no information is available to *you* at this moment (i.e., before seeing data) regarding the canonical parameter θ to be estimated. Under this circumstance *your* belief regarding the canonical parameter θ may vary from negative infinity and positive infinity because $-\infty < E(\theta) < \infty$. This state is referred to as a state of complete-ignorance in Benavoli and Zaffalon (2012, p. 1974).

In the proposed methodology, the unit square (i.e., $\mathscr{R}_0 = \{(\alpha, \beta)|0 < \alpha \leq 1, 0 < \beta \leq 1\}$) is suggested for use when eliciting an imprecise prior. The rationale for characterizing this region $\mathscr{R}_0$ is that all possible prior expectations $E(\theta)$ of the canonical parameter θ are covered by this unit square. Further discussion regarding this characterization strategy is provided in Subsection 3.3.3. The plot on the top left panel in Figure 3.3 supports this rationale. In that figure, the characterized region $\mathscr{R}_0$ is coloured blue and a level set of a prior expectations $E(\theta) = \psi(\alpha) - \log(\beta)$, where $\psi(\cdot)$ is a digamma function, which is referenced with dashed curves for selected values ranging from -3 to 3 by 0.5. The derivation of this prior expectation $E(\theta)$ is noted in Appendix A.1. Observe that the prior expectation $E(\theta)$ appears to be linear in Figure 3.3, but it is in fact slightly curved near the point of origin $(0, 0)$.

Now suppose that *you* observe a sample $y_1 = 1$. *Your* imprecise prior represented by the region $\mathscr{R}_0$ moves by 1 along α-axis and also moves by 1 along β-axis as shown on the top right panel in the Figure 3.3. Because a family of log-gamma priors is closed under the sampling model, after observing n i.i.d. samples, the posterior expectation $E(\theta|y) = \psi(\alpha + n\bar{y}) - \log(\beta + n)$ has a form identical to the prior expectation $E(\theta)$. The derivation of this posterior expectation $E_n(\theta|\mathbf{y})$ is also noted in Appendix A.1. It is evident from this figure that the range of prior expectations $E(\theta)$ which are covered by the region $\mathscr{R}_1$ representing the imprecise posterior expectation $E(\theta|\mathbf{y})$ after observing the sample y_1 becomes narrower

Figure 3.3: Translation behaviour of the natural imprecise log-gamma prior (i.e., $\mathscr{R}_0 = \{(\alpha, \beta) | 0 < \alpha \leq 1, 0 < \beta \leq 1\}$ coloured blue) as a sample is newly observed in a sequence $y_1 = 1$, $y_2 = 1$, $y_3 = 1$, $y_4 = 0$, $y_5 = 0$, $y_6 = 2$, $y_7 = 1$, $y_8 = 0$, and $y_9 = 2$ which are taken from the standard Poisson sampling model with $\mu = 1$. The dashed curves are the level set of a prior expectation $E(\theta) = \psi(\alpha) - \log(\beta)$, where $\psi(\cdot)$ is a digamma function, for selected values ranged from -3 to 3 by 0.5.

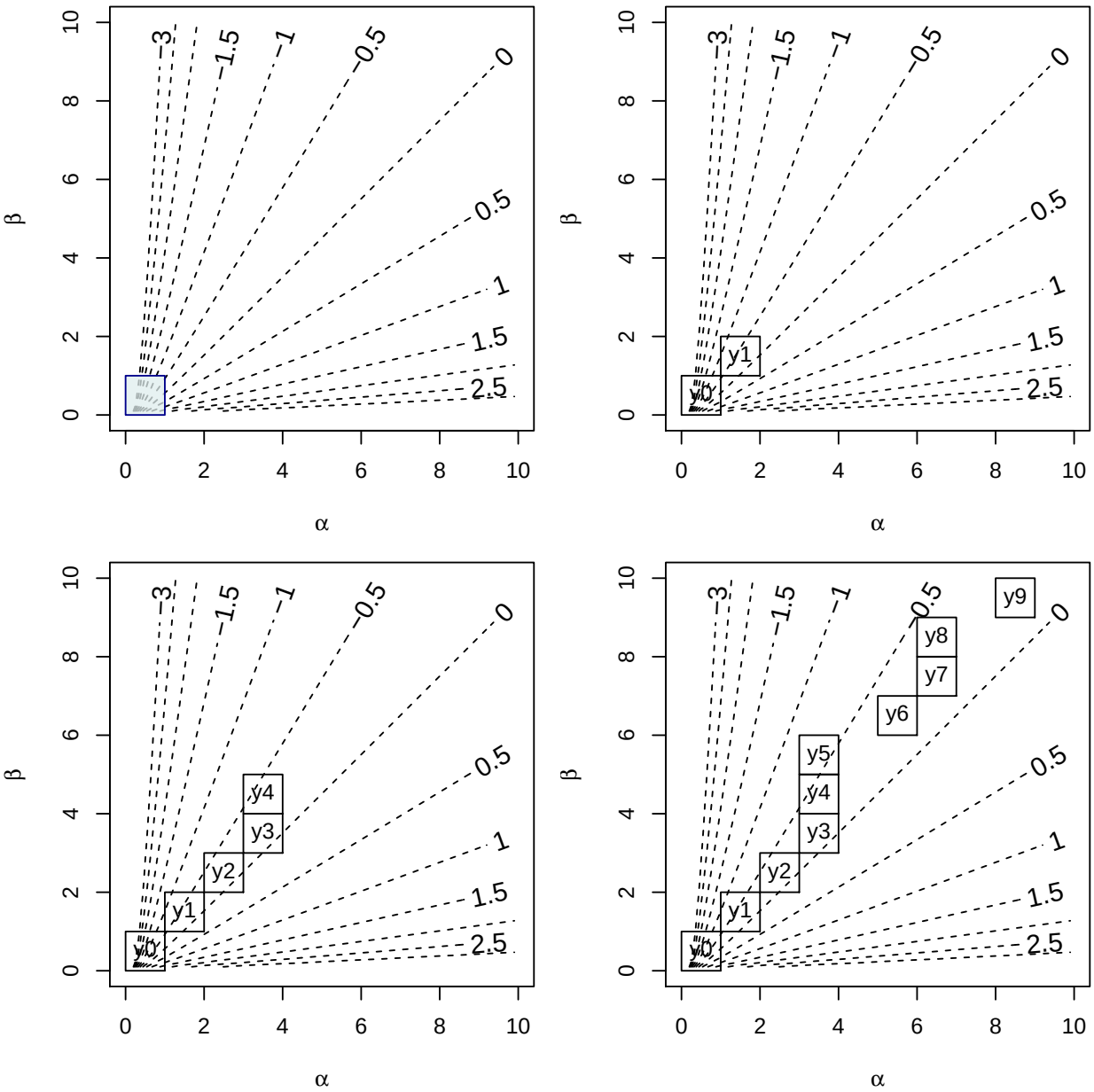
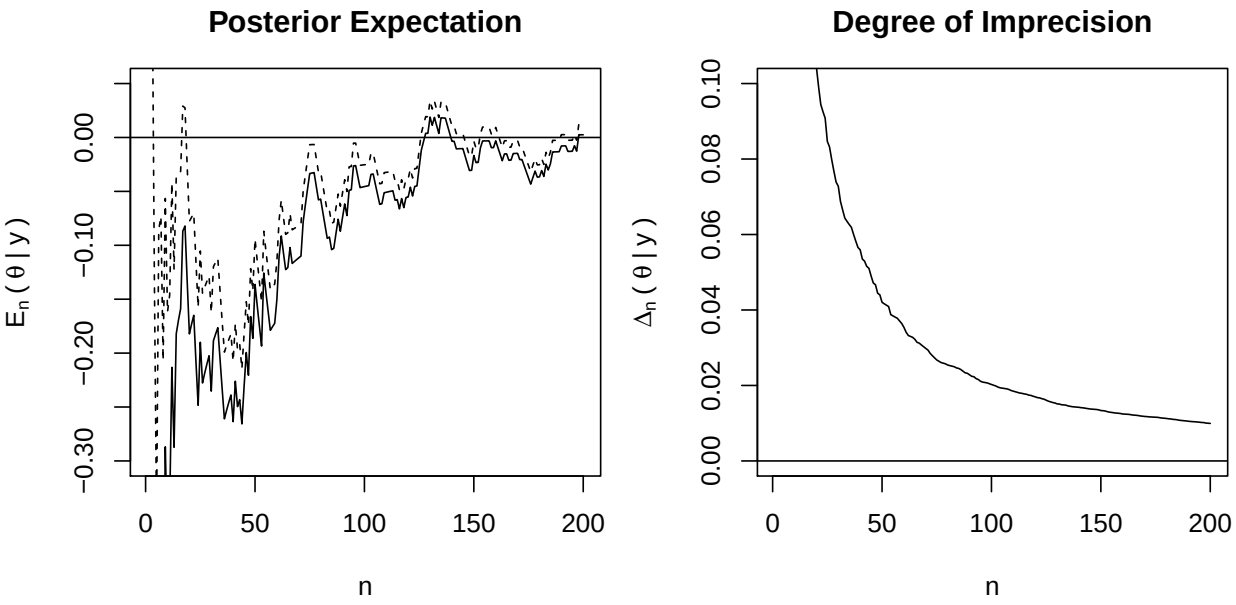


Figure 3.4: The changes of the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$ (dashed line) and the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$ (solid line) over the sample size n are shown in the left panel when the imprecise log-gamma prior (i.e., $\mathscr{R}_0 = \{(\alpha, \beta)|0 < \alpha \leq 1, 0 < \beta \leq 1\}$) is used. 200 samples are sequentially taken from the standard Poisson sampling model with a mean of $\mu = 1$. The change of its corresponding degree of imprecision $\Delta_n(\theta|\mathbf{y})$ is shown in the right panel.



than the range of prior expectations $E(\theta)$ initially covered by $\mathscr{R}_0$ since the space between the levels increases as either of the hyperparameter of α or β is increasing.

The plot on the bottom left panel in Figure 3.3 shows the movement of $\mathscr{R}_n$ individually by sequentially taking a sample of $y_2 = 1$, $y_3 = 1$, $y_4 = 0$, and $y_5 = 0$. The identical movements for the region of $\mathscr{R}_2$ and $\mathscr{R}_3$ are observed; however, the regions $\mathscr{R}_4$ and $\mathscr{R}_5$ move up by one at a time along β-axis but do not have any movement along α-axis. Since zero values with observations y_4 and y_5 do not contribute to the update of information on the α-axis, the regions $\mathscr{R}_4$ and $\mathscr{R}_5$ do not approach the true value of $E(\theta) = 0$. In the same manner, movement on the regions $\mathscr{R}_6$ ($y_6 = 2$), $\mathscr{R}_7$ ($y_7 = 1$), $\mathscr{R}_8$ ($y_8 = 0$), and $\mathscr{R}_9$ ($y_9 = 2$) are shown in the plot on the bottom right in Figure 3.3. Numerical summaries of the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$, the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, and the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ are listed in Table 3.1.

Table 3.1: The maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$, the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, and the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ as an observation y_i is newly taken ($y_1 = 1$, $y_2 = 1$, $y_3 = 1$, $y_4 = 0$, $y_5 = 0$, $y_6 = 2$, $y_7 = 1$, $y_8 = 0$, $y_9 = 2$) when the natural imprecise log-gamma prior is specified.

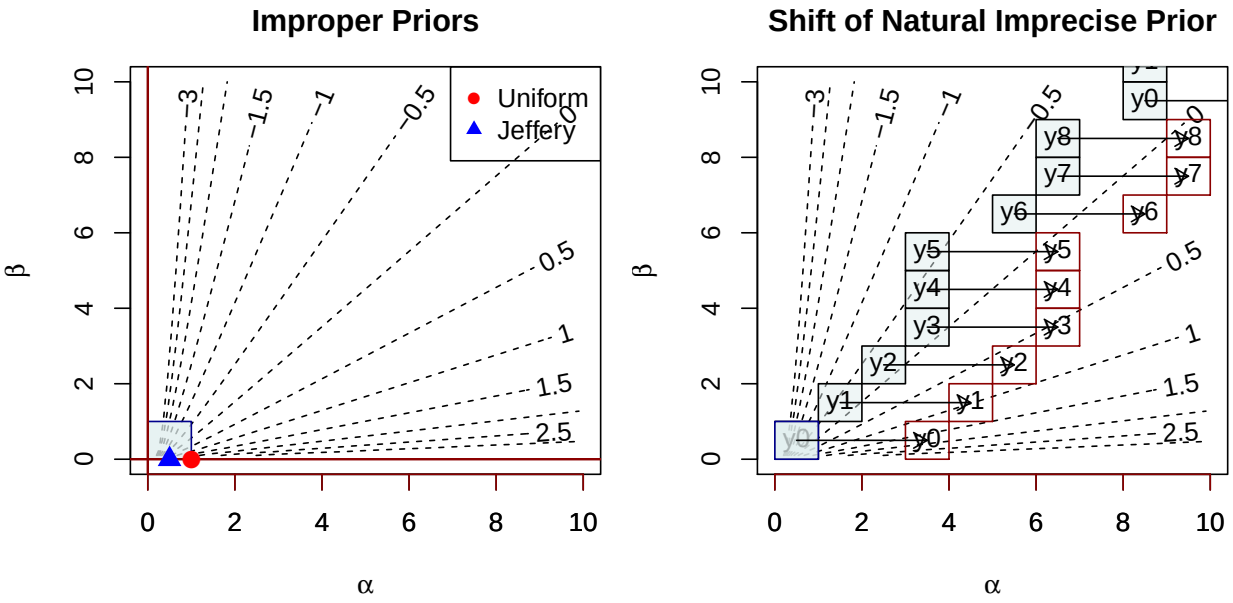
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
$\underline{E}_n(\theta\|\mathbf{y})$	−1.270	−0.676	−0.464	−0.687	−0.869	−0.440	−0.373	−0.491	−0.287
$\overline{E}_n(\theta\|\mathbf{y})$	0.423	0.230	0.158	−0.130	−0.353	−0.086	−0.073	−0.207	−0.057
$\Delta_n(\theta\|\mathbf{y})$	1.693	0.905	0.621	0.556	0.516	0.354	0.300	0.284	0.230

It is observed that the maximum and minimum imprecise posterior expectations $\overline{E}_n(\theta|\mathbf{y})$ and $\underline{E}_n(\theta|\mathbf{y})$ in Table 3.1 correspond to the quantities evaluated at the points on the top left corner and bottom right corner of the region $\mathscr{R}_n$, respectively, in Figure 3.3. It is also confirmed that the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ decreases when a new sample y_i is observed. Although the range of the imprecise posterior expectation does not capture the true value of $\theta = 0$ with a sample of size 9, both the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$ and the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$ are approaching 0. Hence, it is anticipated by the asymptotic properties of Bayes estimator (Roberts, 2007, p. 48) that the two extreme imprecise posterior expectations $\overline{E}_n(\theta|\mathbf{y})$ and $\underline{E}_n(\theta|\mathbf{y})$ will meet in the vicinity of the canonical parameter value $\theta = 0$ when a sample of sufficiently large size n can be observed. Figure 3.4 clearly demonstrates this expectation over 200 Poisson random samples. The solid and dashed lines imply the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$ and the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, respectively.

3.3.3 Less subjective characterization strategy for imprecise prior

In the previous subsection 3.3.2 the unit square $\mathscr{R}_0 = \{(\alpha, \beta)|0 < \alpha \leq 1, 0 < \beta \leq 1\}$ is characterized in order to describe *your* imprecise prior under *your* state of complete ignorance. However, the unit square $\mathscr{R}_0$ is not the only one approach that can be used to elicit an imprecise prior. One could characterize a triangle by formulating five linear inequality constraints of $\{(\alpha, \beta)|0 < \alpha \leq 1, 0 < \beta \leq 1, \alpha + \beta \leq 1\}$) since this newly characterized triangle also covers all possible prior expectations (i.e, $-\infty < E(\theta) < \infty$)) as justified for the case of characterizing the unit square. Another formulation would be the first section of

Figure 3.5: Improper priors on the characterization of the natural imprecise log-gamma prior (i.e., corresponds to $\mathscr{R}_0 = \{(\alpha, \beta)|0 < \alpha \leq 1, 0 < \beta \leq 1\}$) are shown in the left panel. A positive uniform prior is marked by a red coloured circle and Jeffrey's prior is marked by a blue coloured triangle. A shift of the natural imprecise log-gamma prior (i.e., $\mathscr{R}_0 = \{(\alpha, \beta)|3 \leq \alpha \leq 4, 0 \leq \beta \leq 1\}$) is shown on the right panel.



a unit circle such that $\mathscr{R}_0 = \{(\alpha, \beta)|\alpha^2 + \beta^2 \leq 1, \alpha \geq 0, \beta \geq 0\}$. This type of an imprecise prior that covers all possible prior expectations for $E(\theta)$ will be referred to as a natural imprecise prior from this point forward in this proposed methodology since it seems to be a natural and intuitive approach representing an imprecise prior before seeing data. This characterization strategy also permits *you* to maintain *your* objectivity in *your* study as *you* await what the data will reveal; however, *you* must continue to formulate constraints keeping all of *your* prior expectations for $E(\theta)$ in *your* own approach and to determine how large *your* characterization will be. Hence, this characterization strategy for the case of complete ignorance should be viewed as a less subjective approach that represents an imprecise prior but not a fully objective approach.

Before moving to the next example the author draws the reader's attention to an interesting feature of the suggested characterization strategy for describing a natural imprecise prior. Consider a situation in the standard Bayesian analysis where the prior information is

insufficient for the estimation of a standard Poisson mean parameter μ. A positive uniform prior $\pi(\mu) = 1$ and Jeffrey's prior $\pi(\mu) = \mu^{-0.5}$ (Gill, 2008, p. 150) are commonly used as a kind of convention since a posterior expectation can be evaluated despite the knowledge that these two priors are known to be improper. (The derivation of Jeffrey's prior is noted in the Appendix A.1.) Note that a positive uniform prior distribution is a limiting case of a gamma distribution with shape parameter $\alpha = 1$ and rate parameter $\beta = 0$. Jeffrey's prior is also a special case of gamma distribution with a shape parameter $\alpha = 0.5$ and a rate parameter $\beta = 0$. Each of these two priors is represented by a single point lying on the bottom boundary (i.e., $\beta = 0$) of the unit square as shown on the left panel in Figure 3.5. A positive uniform prior is marked by a red coloured circle and Jeffrey's prior is marked by a blue coloured triangle.

Notice that any points lying on either the α- or β- axis ultimately represent improper priors. One may express concern with the failure that results when attempting to evaluate an imprecise posterior expectation using the proposed methodology; however, the geometric representation of the imprecise posterior expectation in Figure 3.5 ensures the existence of the imprecise posterior expectation since the unit square representing the natural imprecise prior translates to another place on the hyperparameter space $\boldsymbol{\Xi}$ once at least a single sample with a non-zero value is observed. Hence, the unit square $\boldsymbol{\mathscr{R}}_0 = \{(\alpha, \beta)|0 < \alpha \leq 1, 0 < \beta \leq 1\}$ suggested in the previous subsection 3.3.2 can be extended to the region $\boldsymbol{\mathscr{R}}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 1, 0 \leq \beta \leq 1\}$ for eliciting the natural imprecise prior by allowing a class of improper prior distributions.

3.3.4 Agreement Between Two Intentional Units

The underlying assumption of the three examples that have been illustrated in the previous three subsections is that *you* are the single intentional unit for inference in *your* study. In other words, the conclusions presented by the author in the previous examples may not be the same as those that the reader concluded since the reader may have a different characterization strategy for eliciting an imprecise prior for various reasons. The example presented in this section is a discussion regarding conclusions that can be drawn from two intentional units having different characterization strategies for the same observations.

Table 3.2: The maximum posterior expectation $\overline{E}_n(\theta|\mathbf{y})$, the minimum posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, and the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ as a sample y_i is newly observed in a sequence ($y_1 = 1$, $y_2 = 1$, $y_3 = 1$, $y_4 = 0$, $y_5 = 0$, $y_6 = 2$, $y_7 = 1$, $y_8 = 0$, $y_9 = 2$) for the second intentional unit *him* ; the imprecise log-gamma prior is specified by characterizing the region $\mathscr{R}_0 = \{(\alpha, \beta)|3 \leq \alpha \leq 4, 0 \leq \beta \leq 1\}$.

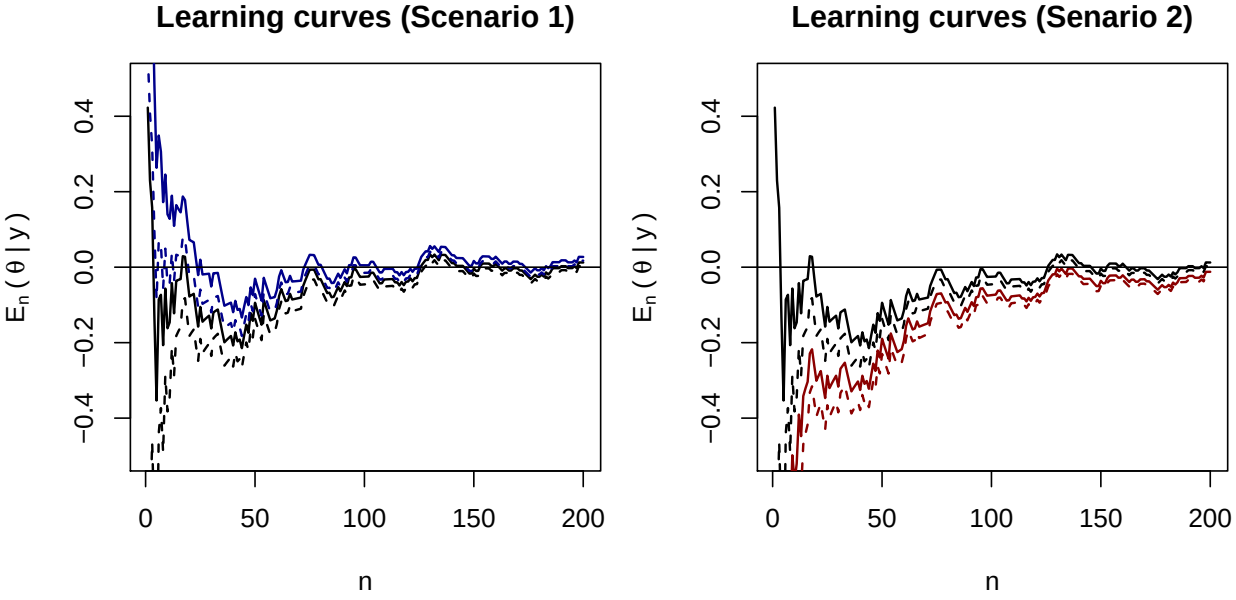
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
$\underline{E}_n(\theta\|\mathbf{y})$	0.563	0.408	0.320	0.097	−0.086	0.070	0.061	−0.057	0.049
$\overline{E}_n(\theta\|\mathbf{y})$	1.506	1.013	0.774	0.486	0.263	0.349	0.306	0.172	0.245
$\Delta_n(\theta\|\mathbf{y})$	0.943	0.605	0.454	0.390	0.349	0.279	0.245	0.229	0.196

To lead this discussion, assume that the first and second intentional units are called by *you* and *him*, respectively. The italic font aids the reader in identifying these intentional units. The samples used in the previous section 3.3.3 are re-used in this subsection. Assume also that *you* continuously use the natural imprecise prior for this inference (i.e., $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 1, 0 \leq \beta \leq 1\}$). Now consider a situation where *he* has some information that already guides *him* to strongly believe that the canonical parameter θ is greater than a value of zero. In *your* case, *you* are still assumed to be under a state of a complete-ignorance. *He* thus may not use the natural imprecise prior as *you* do since *he* is able to characterize some other region such that $\mathscr{R}_0 = \{(\alpha, \beta)|3 \leq \alpha \leq 4, 0 \leq \beta \leq 1\}$ which supports *his* current belief before observing samples.

After *he* observes samples from the first y_1 to the last y_9 one by one in a sequence, *his* imprecise posterior is summarized by the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$, the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, and the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ as shown in Table 3.2. Comparing *his* imprecise posterior summaries to *your* imprecise posterior summaries in Table 3.1 in terms of the size n of a sample observed, *your* degree of imprecision $\Delta_n(\theta|\mathbf{y})$ evaluated with the natural imprecise prior is larger than those evaluated with *his* imprecise prior. This fact represents that *his* imprecise prior is more informative than *your* natural imprecise prior in some way for only *him*.

This fact can be more clearly illustrated by visualizing *your* and *his* imprecise priors on the hyperparameter space as shown on the right panel in Figure 3.5. On may notice that the region $\mathscr{R}_0$ representing *his* imprecise prior shows a shift of the region $\mathscr{R}_0$ representing

Figure 3.6: Two different scenarios of a learning process between two intentional units having different imprecise log-gamma priors on the same observation of a sample of size n. Two curves coloured black on the both left and right panels imply the imprecise posterior expectation of the first intentional unit *you* having the natural imprecise prior (i.e., $\mathscr{R}_0 = \{(\alpha, \beta) | 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1\}$). Two curves coloured dark blue in the left panel imply the imprecise posterior expectation of the second intentional unit *him* having the imprecise prior (i.e., $\mathscr{R}_0 = \{(\alpha, \beta) | 3 \leq \alpha \leq 4, 0 \leq \beta \leq 1\}$). Two curves coloured dark read in the right panel imply the imprecise posterior expectation of someone else having the imprecise prior (i.e., $\mathscr{R}_0 = \{(\alpha, \beta) | 0 \leq \alpha \leq 1, 5 \leq \beta \leq 6\}$). Within the same coloured curves, the solid curve implies the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$ and the dashed curve implies the minimum posterior expectation $\underline{E}_n(\theta|\mathbf{y})$.



your natural imprecise prior by three units in the right direction on α-axis. From that figure, it is also shown that all subsequent regions $\mathscr{R}_n$ representing *his* imprecise posterior are moved from the regions $\mathscr{R}_n$ representing *your* imprecise posteriors by the same amount of units in the same direction on the same axis. The unit squares coloured blue represent *your* imprecise probabilities and the transparent unit squares represent *his* imprecise probabilities. This translation behaviour on the hyperparameter space exemplifies the benefit derived from the conjugate prior measure formulation in the proposed methodology.

Since this imprecise inference is independently performed by each intentional unit, now suppose that *you* and *he* communicate to share the conclusions that *you* draw and that *he*

draws. Both *you* and *he* would be embarrassed since *your* imprecise posterior summaries are not identical to *his* imprecise posterior summaries despite of the use of identical data as shown in the Tables 3.1 and 3.2. The plot on the left panel in Figure 3.6 shows this difference more clearly. To produce this plot, it is assumed that both intentional units *you* and *him* continue to carry out imprecise inferences until 200 sequential samples are observed. The two curves coloured black imply that the imprecise posterior summaries produced from *your* natural imprecise prior, and the other two curves coloured dark blue imply the imprecise posterior summaries produced from *his* imprecise prior. Within the same coloured curves, the solid curve implies the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$ and the dashed line implies the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$. This set of two curves consisting of $\overline{E}_n(\theta|\mathbf{y})$ and $\underline{E}_n(\theta|\mathbf{y})$ is referred to as an imprecise learning curve for a single intentional unit from this point on. The horizontal line at $\theta = 0$ is the reference line implying the true value used for simulating the data. As shown in this plot, *your* imprecise learning curve is much different from *his* imprecise learning curve at the initial learning period (i.e., the period of observing the first few samples). It is also shown that, within each intentional unit, a conflict between prior belief and data (Walley, 1991, p. 222) lasts for a certain period (say, $n = 150$) of the learning process.

The plot on the right panel in Figure 3.6 is produced by assuming that the characterized region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 1, 5 \leq \beta \leq 6\}$ describes the imprecise prior of some other individual who believes that the canonical parameter θ is less than a zero value. In this case, the two curves coloured dark red imply the learning curve of the other individual. It is shown that similar patterns are identified between the two intentional units and within each intentional unit; observation of this can be found in the left panel on the plot.

The two plots in Figure 3.6 show that the pattern of *your* learning curve eventually resembles the pattern of *his* learning curve (in the left panel) or the pattern of someone else's learning curve (in the right panel) since all three intentional units are observing the same data. In other words, a conflict initially observed between learning processes is gradually diminished following a long sequence of observational activities using the same data. Note that a certain gap exists between the two extreme posterior expectations $\overline{E}_n(\theta|\mathbf{y})$ and $\underline{E}_n(\theta|\mathbf{y})$ in each learning curve despite a sufficiently long learning period.

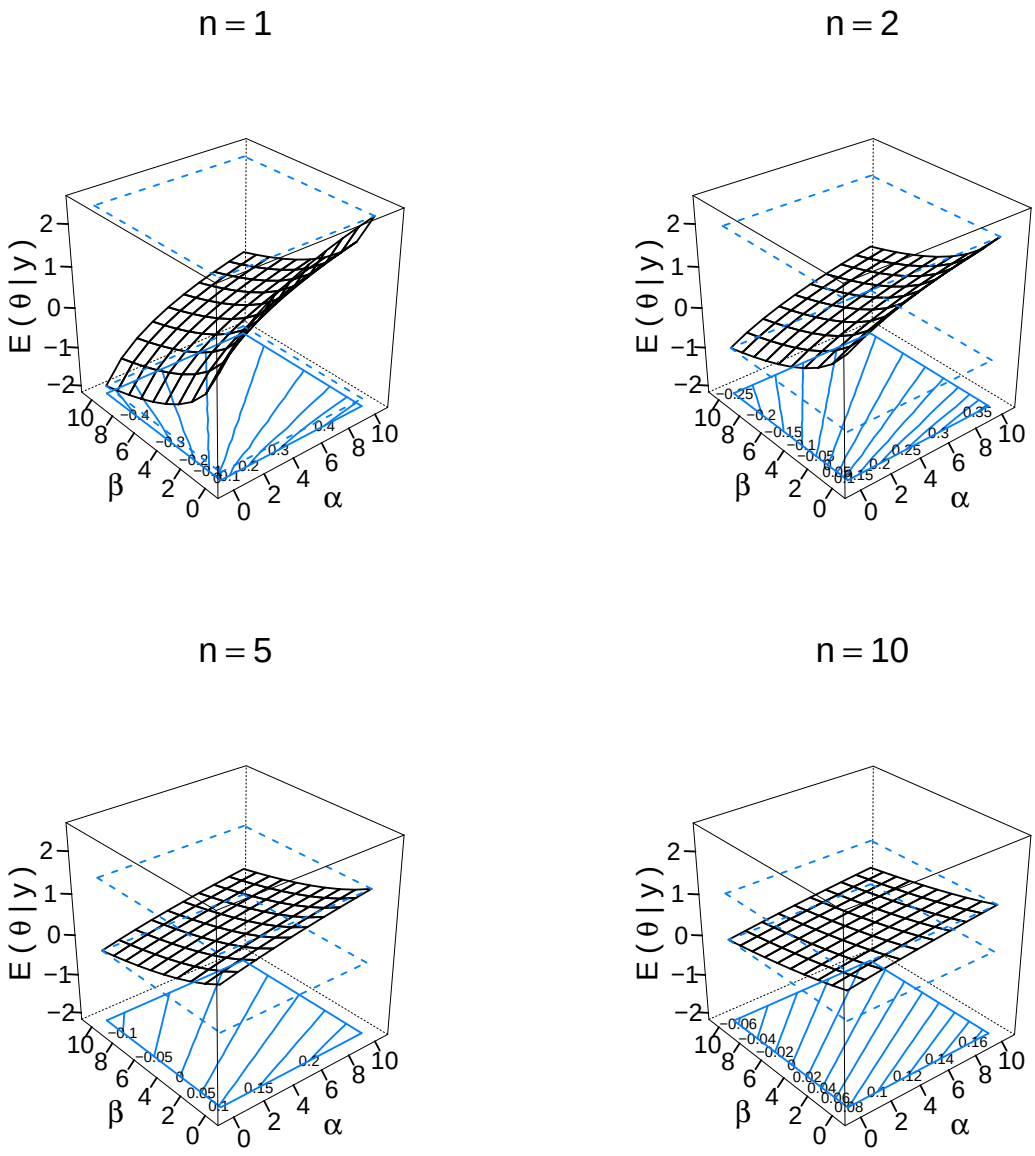
3.3.5 Soft (Weaker) Linear Updating Behaviour of Imprecise Posterior Expectation

The examples presented in this subsection illustrate a soft linear updating behaviour of an imprecise posterior expectation using the proposed methodology. Based on the study of Diaconis and Ylvisaker (1979), a linearity of posterior expectations of the mean parameter μ of a Poisson sampling model $f(y|\mu)$ exists. Johnson (1967), an earlier study of Diaconis and Ylvisaker (1979), also supports this statement. However, this does not hold true in the proposed methodology.

To examine this fact, consider first a natural imprecise prior which is represented by the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 10, 0 \leq \beta \leq 10\}$ on the hyperparameter space $\boldsymbol{\Xi}$ in a family of log-gamma prior distributions. Assume that a sample of size n is drawn from the standard Poisson sampling model with a mean of $\mu = 2$. Different sample sizes of n being $1, 2, 5$, and 10 are considered. The random seed `18372342` is utilized for simulating these four different sizes of Poisson random variates. The proposed methodology is applied to each sample with a predefined natural imprecise prior. For the resulting imprecise posterior with a sample of size n, a surface plot of the imprecise posterior expectation is produced using the function `plot()` provided in the `ipeglim` package in order to investigate whether or not that an imprecise posterior expectation is linearly updated. Figure 3.7 shows four different surface plots of posterior expectation $E(\theta|\mathbf{y})$ each of which is generated from the given imprecise prior $\mathscr{R}_0$ with different sizes n of a sample.

It is observed from the surface plots that all posterior expectations with different sizes of a sample do not lie on the linear plane. Although the surface plot labelled $n = 10$ on the bottom right panel in the Figure 3.7 looks linear, it is not in fact linear since the space between the contour levels on the bottom of that surface plot is not fully parallel. However, this graphically represented diagnostic informs us as to how the surface of an imprecise posterior expectation becomes flattened. Careful attention must be exercised when using the proposed methodology when searching for the upper imprecise posterior expectation $\overline{E}_n(\theta|y)$ and the lower imprecise posterior expectation $\underline{E}_n(\theta|y)$.

Figure 3.7: The surface plots of an imprecise posterior expectation over different sample sizes $n = 1, 2, 5,$ and 10 when the natural imprecise log-gamma prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\alpha, \beta) | 0 \leq \alpha \leq 10, 0 \leq \beta \leq 10\}$) is specified.



Following are the steps describing a further numerical investigation that is planned at this time with regard to this non-linear optimization problem of an imprecise posterior expectation:

1. Find a global optimum $\hat{\xi}$ of hyperparameters ξ using the iterative Newton's method;
2. Evaluate the quantity of a posterior expectation $E_n(\theta|\mathbf{y})$ at the global optimum $\hat{\xi}$ found;
3. Repeat step (2) for all extreme points of the predefined region $\mathscr{R}_0$;
4. Examine if the maximum or minimum of a posterior expectation $E_n(\theta|y)$ occurs at one of the extreme points.

Since the objective function to be optimized here is $E_n(\theta|y) = \psi(\alpha+n\bar{y})-\log(\beta+n)$, where $\psi(\cdot)$ is a digamma function, the score function needed for the iterative Newton's method is given by:

$$\nabla E_\pi(\xi) = (\psi'(\alpha + y), -(\beta + 1)^{-1})^T \tag{3.12}$$

where $\psi'(\alpha)$ is a trigamma function. The Hessian matrix $\mathbb{H}$ is found as detailed below:

$$\mathbb{H}(\xi) = \begin{bmatrix} \psi''(\alpha + y) & 0 \\ 0 & (1 + \beta)^{-2} \end{bmatrix}, \tag{3.13}$$

where $\psi''(\cdot)$ is a polygamm function of order 2 (or, tetragamma function) and its quantity can be obtained using `R` function `psigamma(x, deriv = 2)`. The optimal solution for the objective function can be searched iteratively using the score function in (3.12) and the Hessian matrix in (3.13) as written below:

$$\hat{\xi}^{(t)} \approx \hat{\xi}^{(t-1)} - \nabla E_\pi(\hat{\xi})\mathbb{H}^{-1}(\hat{\xi}). \tag{3.14}$$

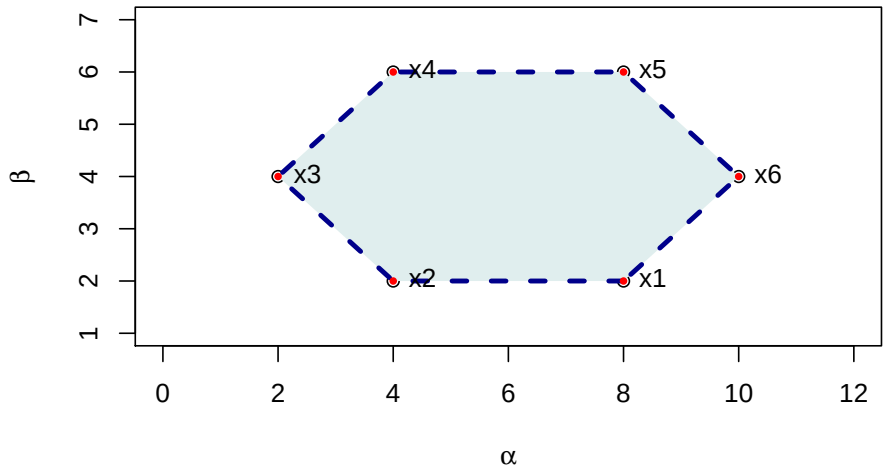
The actual computation of this non-linear optimization problem is done by the `R` function `constrOptim()` instead of using a direct implementation of the iterative Newton's method described in above since the function `constrOptim()` uses an adaptive barrier algorithm which stems from Newton's method (Lange, 1999, p. 185–187). The maximization of a given objective function can be carried out by setting the negative value of the option `fnscale` when the function `constrOptim()` is used.

Table 3.3: Comparison of optimal solutions of hyperparameters (α, β) for the imprecise posterior expectation $E_n(\theta|\mathbf{y})$ with sample sizes $n = 1, 2, 5$, and 10. Optimal solutions are found by using the `R` built-in function `constrOptim()` and using the linear programming (LP) technique adopted in the proposed methodology, respectively. The imprecise log-gamma prior is given by characterizing the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 10, 0 \leq \beta \leq 10\}$.

	Hyperparameters		
	$\hat{\alpha}$	$\hat{\beta}$	$E_n(\theta\|\mathbf{y})$
$n = 1$			
`constrOptim` (min)	0.000	10.000	−1.975
LP Approach	0.000	10.000	−1.975
`constrOptim` (max)	10.000	0.000	2.443
LP Approach	10.000	0.000	2.443
$n = 2$			
`constrOptim` (min)	1.000	8.000	−1.274
LP Approach	1.000	8.000	−1.274
`constrOptim` (max)	5.000	1.000	1.180
LP Approach	5.000	1.000	1.180
$n = 5$			
`constrOptim` (min)	2.000	4.000	−0.353
LP Approach	2.000	4.000	−0.353
`constrOptim` (max)	8.000	2.000	1.153
LP Approach	8.000	2.000	1.153
$n = 10$			
`constrOptim` (min)	0.170	6.294	−1.460
LP Approach	0.170	6.294	−1.460
`constrOptim` (max)	6.294	0.170	1.897
LP Approach	6.294	0.170	1.897

Table 3.3 shows the estimated optimal hyperparameters $\hat{\alpha}$ and $\hat{\beta}$ and the estimated posterior expectation at the estimated optimum of the hyperparameters. It is confirmed that the linear programming technique in the proposed methodology found the exact and identical optimum of hyperparameters as compared to those found when using the function `constrOptim()` for both the minimum and maximum posterior expectations over all different sample sizes of n. Hence, the author of this thesis is confident about employing the linear programming technique to search for the maximum and minimum posterior expectation in this proposed methodology for imprecise inference in favour of considering the Karush-Kuhn-

Figure 3.8: Extreme points of the region $\mathscr{R}_0 = \{(\alpha, \beta) | \beta \geq -\alpha + 3, \beta \leq -\alpha + 7, \beta \geq \alpha - 3, \beta \leq \alpha + 1, \beta \leq 3, \beta \geq 1\}$ characterized on the hyperparameter space $\boldsymbol{\Xi}$ in a family of log-gamma prior distributions.



Turcker (KKT, also known as the Kuhn-Tucker) conditions that uses the Lagrange multipliers method. These numerical results as presented in Table 3.3 are supported by graphical diagnostics using the surface plot of an imprecise posterior expectation. (Please see also the surface plots of an imprecise posterior expectation over four differently characterized regions at the fixed size $n = 1$ of a sample in Appendix C.2).

3.3.6 Focusing Behaviour of Imprecise Posterior

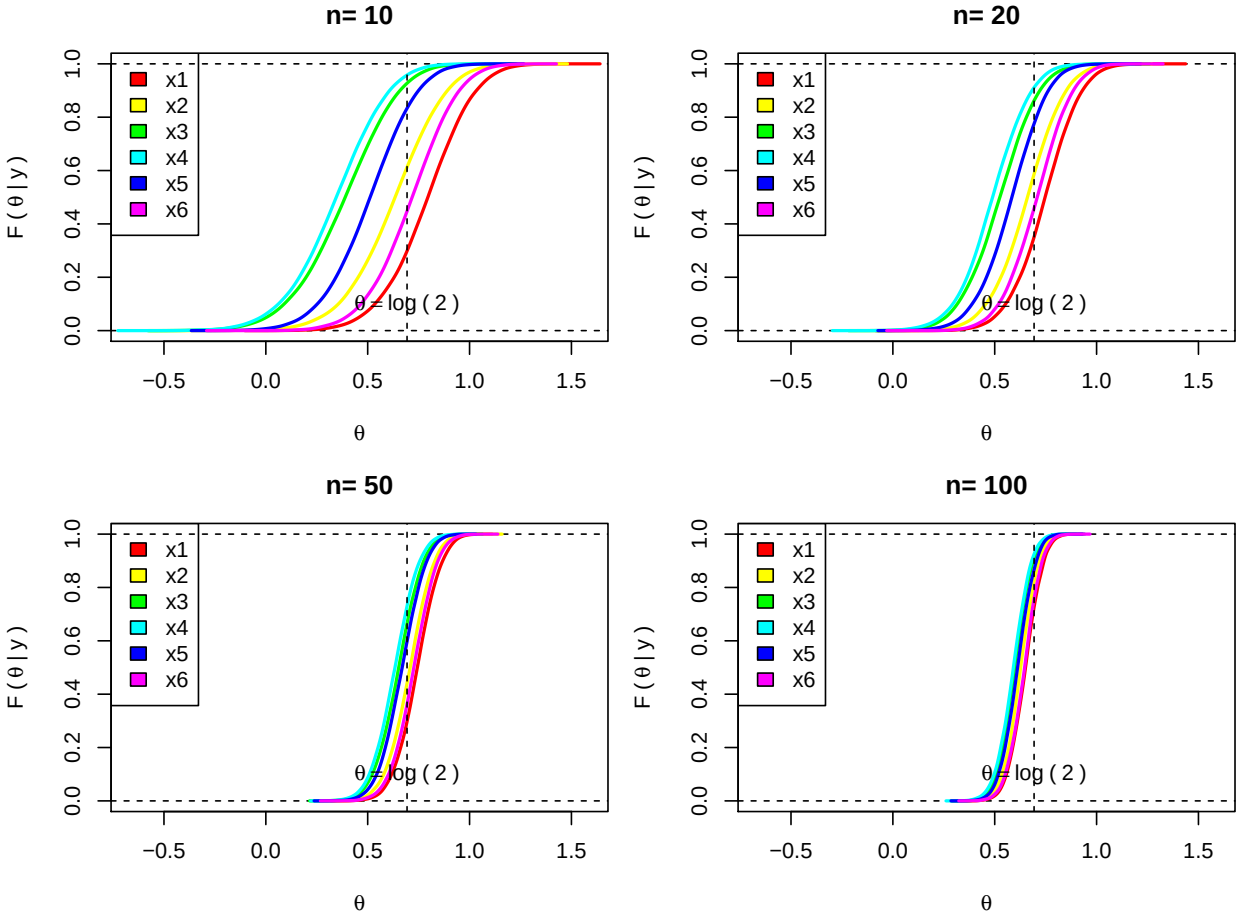
The minimum and maximum posterior expectations are the primary imprecise posterior numerical summaries in this imprecise inference. In the proposed methodology, these two numerical summaries have been used for illustrating the geometric translation behaviour of an imprecise prior in Subsection 3.3.2, the agreement between more than one intentional unit in Subsection 3.3.4, and the soft linear updating behaviour of an imprecise posterior expectation in the previous Subsection 3.3.5. However, it may be difficult for a novice to realize the benefits of applying the imprecise inference to a problem of estimating the parameter of a

sampling model due to an unfamiliarity with the notion of imprecision (Walley, 1991, p. 210). The focusing behaviour of an imprecise posterior presented in this subsection may be one approach that can be used to provide an intuitive approach to the notion of imprecision.

To present this focusing behaviour, a probability box (or, simply p-box) presented in Ferson et al. (2007) is used. The principle of the p-box is that all possible cumulative distributions are enclosed by the upper (left) and the lower (right) bounds on the probabilities. The enclosing probabilities are in fact matched to the distributions producing the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$ and the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$. Since these extreme posterior expectations have occurred at one of the extreme points in the proposed methodology, the p-box can be efficiently constructed using the optimal solutions of hyperparameters. The analytical proof of this suggestion is related to Troffaes and Destercke (2011) and is left for further research.

Consider the region $\mathscr{R}_0$ characterized by six linear inequality constraints of $\{(\alpha, \beta)|\beta \geq -\alpha + 3, \beta \leq -\alpha + 7, \beta \geq \alpha - 3, \beta \leq \alpha + 1, \beta \leq 3, \beta \geq 1\}$ on the hyperparameter space in a family of log-gamma prior distributions. The characterized region $\mathscr{R}_0$ is shown in Figure 3.8. The extreme points are coloured red, and an identification number is assigned to each of these extreme points. A sample of size n is drawn from the standard Poisson sampling model with a mean of $\mu = 2$. Four different sizes of n which are 10, 20, 50, and 100 are selected. The function `pbox()` provided in the `ipeglim` package generates a sampling distribution of the canonical parameter θ at each of the extreme points of the characterized region $\mathscr{R}_0$ by using the Metropolis-Hastings algorithm with a given sample of size n. The empirical cumulative probabilities computed from the generated sampling distribution are plotted against the ordered samples of the canonical parameter θ as shown in Figure 3.9.

Figure 3.9: Focusing behaviour of an imprecise posterior with different sample sizes $n = 10, 20, 50$, and 100; Samples are taken from the standard Poisson sampling model with a mean of $\mu = 2$; The imprecise log-gamma prior illustrated in Figure 3.8 is used.



It is confirmed that the posterior distributions with the identification numbers `x4` and `x1` lie at the most left and right sides, respectively, and all posterior distributions generated at the extreme points are enclosed by these two posterior distributions for all cases of the four different samples. Note also that the probability band enclosed but not overlapping each other by these two posterior distributions of `x4` and `x1` is becoming focused on the true value of the canonical parameter $\theta = \log(2)$ which was used for simulating the samples as the size n of a sample increases. In fact, the author of this thesis tested repeatedly to determine whether or not any member of the resulting imprecise posteriors are enclosed by these two posterior distributions at different sizes of a sample using at least 100 hyperparameters inside or lying on the boundaries of the characterized region $\mathscr{R}_0$. These test results are not presented here

Table 3.4: The posterior expectation $E_n(\theta|\mathbf{y})$ estimated at the every extreme point `x1`, `x2`, `x3`, `x4`, `x5`, and `x6` of the region $\mathscr{R}_0 = \{(\alpha, \beta)|\beta \geq -\alpha + 3, \beta \leq -\alpha + 7, \beta \geq \alpha - 3, \beta \leq \alpha + 1, \beta \leq 3, \beta \geq 1\}$ which was used for constructing the probability boxes in Figure 3.9 for sample sizes $n = 10, 20, 50$, and 100.

	$n = 10$	$n = 20$	$n = 50$	$n = 100$
`x1`	0.792	0.748	0.745	0.651
`x2`	0.629	0.658	0.707	0.630
`x3`	0.381	0.523	0.651	0.600
`x4`	0.341	0.491	0.633	0.591
`x5`	0.505	0.581	0.671	0.612
`x6`	0.711	0.704	0.725	0.641
$\Delta_n(\theta\|\mathbf{y})$	0.451	0.257	0.111	0.059
Area of Band	0.450	0.256	0.110	0.059

since the use of posterior distributions generated at the extreme points is sufficient to present this focusing behaviour of an imprecise posterior as shown in the Figure 3.9 .

To ensure that the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ induced from this probability box is equal to the one found by a difference between two extreme posterior expectations, the area enclosed by two posterior distributions with `x4` and `x1` is evaluated. This quantity is reported in the last row on Table 3.4 for each of four different samples, and the posterior expectations $E(\theta|\mathbf{y})$ estimated at every extreme point are also listed. The difference noted between the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ computed by using the maximum and minimum expectations versus the area of the probability band, compared in Table 3.4,is apparent at third decimal place and is due to the fact that the probability band is constructed using the Metropolis-Hastings algorithm. The exact quantity of this imprecision $\Delta_n(\theta|\mathbf{y})$ will be found in future research.

Chapter 4

𝔅-Formulation

A family of log-gamma prior distributions used in the previous Chapter 3 is a conjugate prior probability measure for a canonically parametrized standard Poisson likelihood. The question to be examined is how to lead an imprecise inferential framework when a family of prior distributions is no longer conjugate for a given sampling model. The purpose of this chapter is to introduce a formulation strategy, the author has named the 𝔅-formulation , for structuring a conjugate prior probability measure for such a case. The principle of the 𝔅-formulation is to construct a new family of probability distributions that accounts for both prior and posterior distributions using a higher-dimensional exponential family representation (Lee and Bickis, 2012).

At this time, the author of this thesis takes this opportunity to provide a background regarding the origin of the name 𝔅-formulation. The author has been privileged to have Professor Miķelis Bickis supervising this thesis work. It was from a discussion with him that the idea of this 𝔅-formulation arose in July, 2012, and its results ultimately culminate in the outline of this thesis work. The author thus refers this canonically parametrized conjugate prior measure formulation to as 𝔅-formulation for Professor Miķelis Bickis in appreciation of his statistical insight.

The 𝔅-formulation is discussed in the following sections for three different cases. In Section 4.1, a family of normal prior distributions is examined for the instance of lacking a conjugate prior measure due to the use of other families of prior distributions for a canonically parametrized standard Poisson likelihood. The three primary behaviours of the proposed methodology are re-examined to ensure that 𝔅-formulation produces the anticipated results which in this instance is the construction of a conjugate prior probability measure. In keeping with the research interest as presented in Chapter 1, the 𝔅-formulation presented in

Section 4.1 is applied to the problem of inferring the canonical parameter of a zero-truncated Poisson sampling model when either a family of log-gamma or normal prior distributions is considered for an imprecise inference in Section 4.2. A comparative investigation into imprecise estimates of the canonical parameter of standard Poisson and zero-truncated Poisson sampling models is made. An examination of the $\mathfrak{B}$-formulation in fact aids in determining whether or not the proposed methodology can be extended to further regression analysis in the context of a generalized linear model. In the last section 4.3 of this chapter, the proposed methodology is applied to the problem of estimating the regression parameters for both the standard and the zero-truncated Poisson regression models.

4.1 $\mathfrak{B}$-Formulation: Three-Parameter Exponential Family Representation

A family of log-normal distributions is often considered to be a mixing distribution for explaining over-dispersed count data (Winkelmann, 2008). The underlying probability theory of this mixture modelling is not the same as the probability theory supporting the Bayesian paradigm; however, both mixture modelling and the Bayesian paradigm share the same mathematical framework for marginalizing the model parameter. Note that a normalizing constant of the mixed Poisson-log-normal distribution is not known (Weems and Smith, 2004, p. 192).

Suppose that the mean parameter μ of a standard Poisson sampling model $f(y|\mu)$ is log-normally distributed with a mean of $\nu \in (-\infty, \infty)$ and a variance of $\tau^2 > 0$. The canonical parameter $\theta = \log(\mu)$ then has a normal distribution for which the probability density function is given by

$$\pi(\theta|\nu, \tau^2) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2\tau^2}(\theta - \nu)^2\right). \tag{4.1}$$

Since the underlying assumption of the proposed methodology is that an imprecise prior and an imprecise posterior are conjugate, the functional form of a family of normal prior distributions $\pi(\theta)$ in (4.1) needs to be compared with a family of posterior distributions

$p(\theta|\mathbf{y})$ after observing n i.i.d. Possion samples such that

$$p(\theta|\mathbf{y}) \propto \exp(n\bar{\mathbf{y}}\theta - ne^{\theta}) \exp\left(-\frac{1}{2\tau^2}\theta^2 + \frac{\nu}{\tau^2}\theta - \frac{\nu^2}{2\tau^2}\right) \tag{4.2}$$

$$\propto \exp\left(-\frac{1}{2\tau^2}\theta^2 + (\frac{\nu}{\tau^2} + n\bar{\mathbf{y}})\theta - ne^{\theta}\right), \tag{4.3}$$

where $\bar{\mathbf{y}} = \frac{1}{n}\sum_i^n y_i$. Since the functional form of the unnormalized posterior distribution $p(\theta|\mathbf{y})$ in (4.3) is not identical to the one in (4.1), it becomes apparent that theses two families of distributions are not conjugate.

However, the unnormalized posterior distribution $p(\theta|\mathbf{y})$ in (4.3) provides another perspective for the formulation of a conjugate relationship between two families of distributions by viewing it as the representation of an exponential family of distributions that involves three random quantities of θ^2, θ, and e^{θ} each of which is a function of the canonical parameter θ such that

$$p(\theta|\mathbf{y}) = \exp(-\xi_2\theta^2 + \xi_1\theta - \xi_0 e^{\theta} - \mathcal{A}(\xi_2, \xi_1, \xi_0)) \tag{4.4}$$

where $\mathcal{A}(\xi_2, \xi_1, \xi_0)$ is the log-normalizer given by

$$\mathcal{A}(\xi_2, \xi_1, \xi_0) = \log \int_{-\infty}^{\infty} \exp(-\xi_2\theta^2 + \xi_1\theta - \xi_0 e^{\theta}) d\theta < 0, \tag{4.5}$$

and ξ_2, ξ_1, and ξ_0 are the hyperparameters defined as

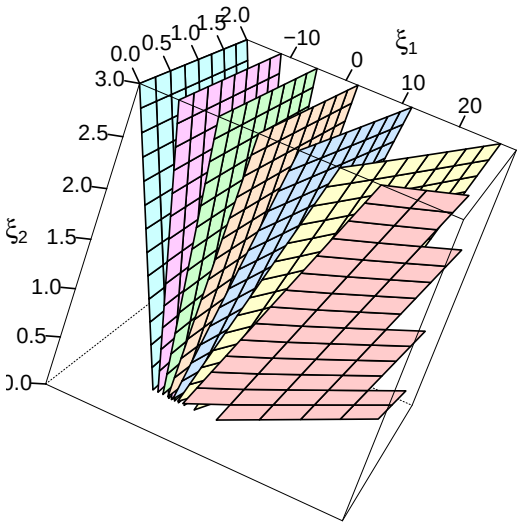
$$\xi_2 = \frac{1}{2\tau^2}, \quad \xi_1 = \frac{\nu}{\tau^2} + n\bar{\mathbf{y}}, \quad \xi_0 = n. \tag{4.6}$$

The hyperparameter space $\boldsymbol{\Xi}$ of this newly defined distribution in (4.4) is found by

$$\boldsymbol{\Xi} = \{(\xi_2, \xi_1, \xi_0) | \xi_2 > 0, \quad -\infty < \xi_1 < \infty, \quad \xi_0 \geq 0\}. \tag{4.7}$$

Note that if the parameter ξ_0 associated with e^{θ} is set by 0 (i.e., no observations), then (4.4) corresponds to a family of normal prior distributions $\pi(\theta)$ in (4.1) since the term e^{θ} contributed by the standard Poisson samples $\mathbf{y}$ are eliminated. Similarly, if the parameter ξ_2 associated with θ^2 is set by 0, then (4.4) corresponds to a family of log-gamma prior

Figure 4.1: The level surfaces of a prior expectation $E(\theta)$ of a newly defined probability distribution in (4.4) referenced at the values of $\{-3, -2, -1, 0, 1, 2, 3\}$ starting from the left to the right using different colours on the hyperparameter space $\boldsymbol{\Xi} = \{(\xi_2, \xi_1, \xi_0) | \xi_2 > 0, -\infty < \xi_1 < \infty, \xi_0 \geq 0\}$.



distributions in (3.9). This newly defined family of distributions in (4.4) ultimately includes both families of log-gamma and normal prior distributions.

This canonical parametrization-based conjugate prior measure formulation presented above needs to be examined prior to use with the proposed methodology. If the newly defined family of distributions in (4.4) accounts for a conjugate relationship between two families of prior and posterior distributions for a Poisson likelihood, the three major behaviours of the proposed methodology would be reproduced as shown in Subsections 3.3.2, 3.3.5, and 3.3.6.

In order to examine the translation behaviour of an imprecise prior, a posterior expectation $E(\theta|\mathbf{y})$ needs to be computed numerically by evaluating the ratio of two integrals:

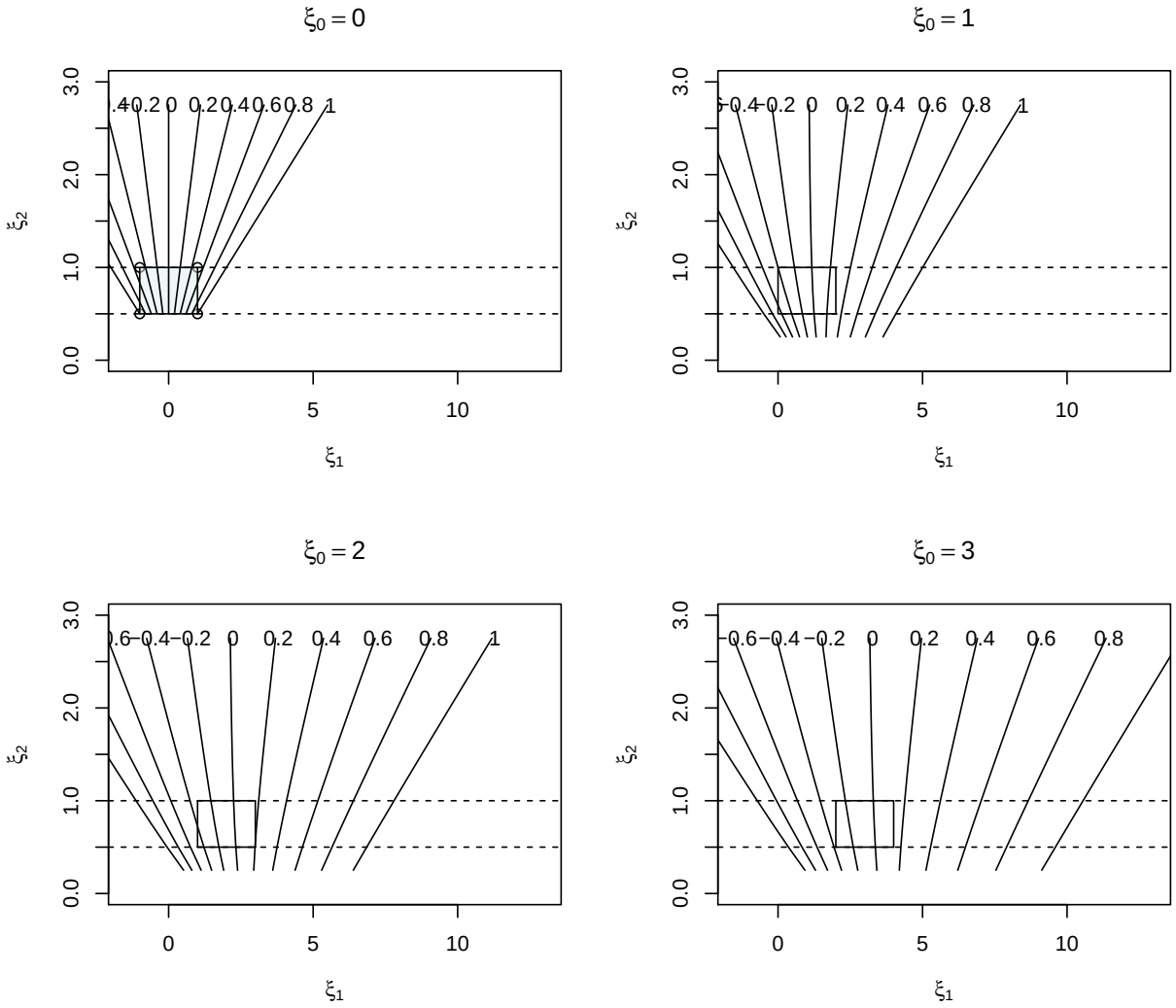
$$E(\theta|\mathbf{y}) = \frac{\int_{-\infty}^{\infty} \theta \exp(-\xi_2\theta^2 + \xi_1\theta - \xi_0 e^{\theta}) d\theta}{\int_{-\infty}^{\infty} \exp(-\xi_2\theta^2 + \xi_1\theta - \xi_0 e^{\theta}) d\theta}, \tag{4.8}$$

and all possible prior expectations of $E(\theta)$ have to be evaluated over the entire hyperparameter space $\boldsymbol{\Xi}$ since the state of complete-ignorance is most accurately portrayed by a convex hull $\mathscr{C}$ that geometrically captures the prior expectation $E(\theta)$ as well as possible.

Figure 4.1 illustrates the range of this prior expectations of $E(\theta)$. The level surfaces are referenced at the values of the prior expectations of $E(\theta)$ from -3 to 3 by 1 beginning from the left to the right using different colours for visual clarity. Each level surface appears as an almost linear surface over the three-dimensional hyperparameter space $\boldsymbol{\Xi}$. Note that each surface curves slightly near the point of origin $(0, 0, 0)$. (Each level surface in this figure appears to be a linear plane due to the matters of scale related to presenting this set on a wide range of the hyperparameter space.) From this figure it is observed that the space between level surfaces becomes wider as the value of the hyperparameter ξ_0 increases. Since the hyperparameter ξ_0 in (4.6) implies the size n of a sample observed, a natural imprecise prior is defined by characterizing some region on the plane of two hyperparameters ξ_2 and ξ_1 which is the hyperparameter space $\boldsymbol{\Xi}$ sliced at the zero value of the hyperparameter ξ_0. One potential approach to characterize this region is to utilize four linear inequality constraints such that $\boldsymbol{\mathscr{R}}_0 = \{(\xi_2, \xi_1) | 0.5 \leq \xi_2 \leq 1, -1 \leq \xi_1 \leq 1\}$. Using this analogy, the translation behaviour of an imprecise prior can be investigated as a new sample is observed.

For illustrating this translation behaviour, three samples $y_1 = y_2 = y_3 = 1$ are taken from the standard Poisson sampling model with a mean of $\mu = 1$ (the random seed `16979238` is used). Assume that *you* are the intentional unit who is leading the imprecise inference for estimating the canonical parameter θ of this standard Poisson sampling model from which samples are drawn. The natural imprecise prior is used as given above, and presented in the plot on the top left in Figure 4.2. The almost straight lines on the hyperparameter space represented by ξ_1 and ξ_2 are the reference lines of the prior expectation $E(\theta)$ of which values range from -1 to 1 by 0.2. The plots in Figure 4.2 are labelled as $n = 1$ (i.e., the first sample $y_1 = 1$ is observed), $n = 2$ (the second sample $y = 2$), and $n = 3$ (the third sample $y_3 = 1$). Since the hyperparameter $\xi_2 = 1/\tau^2$ is a constant, the region $\boldsymbol{\mathscr{R}}_0$ moves only to the right along the ξ_1-axis by the quantity of each sample $y_1 = 1$, $y_2 = 1$, and $y_3 = 1$. Concurrently, the reference curves of the prior expectation $E(\theta)$ are extended outward so that the range of the prior expectation $E(\theta)$ covered by the natural imprecise prior narrows. This movement of the natural imprecise prior illustrates that the amount of uncertainty due to *your* prior ignorance is being reduced by learning from data.

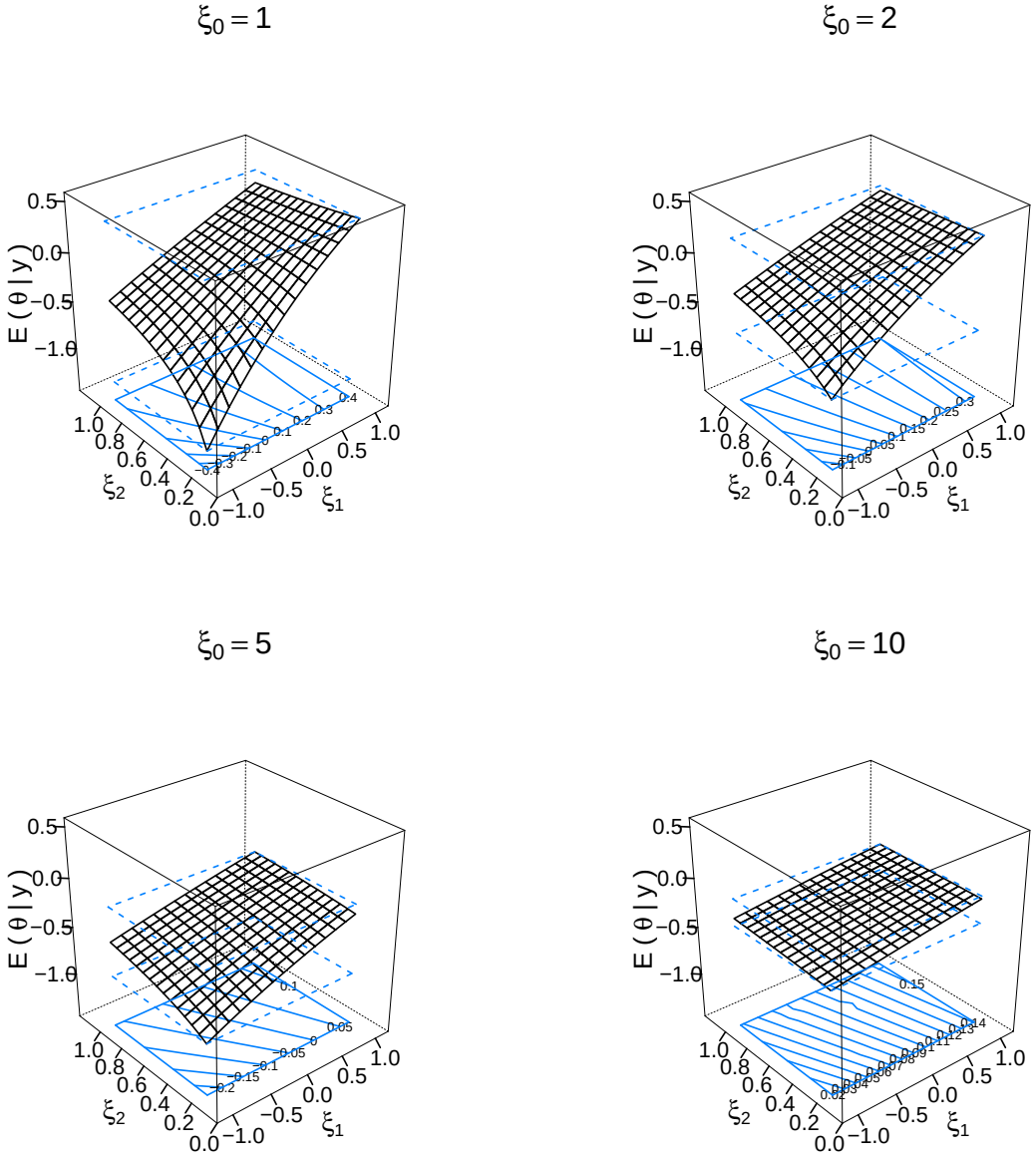
Figure 4.2: Translation behaviour of the imprecise normal prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\xi_2, \xi_1) | 0.5 \leq \xi_2 \leq 1, -1 \leq \xi_1 \leq 1\}$) on the plane of ξ_1 and ξ_2 which is the hyperparameter space $\boldsymbol{\Xi}$ sliced at $\xi_0 = 0, 1, 2$, and 3.



For an examination of the soft linear updating behaviour of an imprecise posterior expectation, a sample of size 10 is newly drawn from the standard Poisson sampling model with a mean of $\mu = 2$ ($\mathbf{y} = \{2, 1, 2, 0, 1, 3, 2, 1, 3, 1\}$). The random seed `16979238` is utilized for these Poisson random variates, and the identical natural imprecise prior is employed. The surface plot for the imprecise posterior expectation is produced over the region $\mathscr{R}_0$ by assuming that *you* observe the first n samples sequentially. These plots in Figure 4.3 are labelled $n = 1$ (i.e., after observing the first sample $y_1 = 2$), $n = 2$ (i.e., the first two samples), $n = 5$ (i.e., the first five samples), and $n = 10$ (all samples). The contour plot on the bottom of each surface

plot is used as a graphical diagnostic tool to examine the linearity of the imprecise posterior expectation. For example, the imprecise posterior expectation is linearly updated over the region $\mathscr{R}_0$ in the two-dimensional hyperparameter space if the contour lines are parallel and equally spaced.

Figure 4.3: Soft-linearity behaviour of a posterior expectation $E_n(\theta|\mathbf{y})$ for the case when an imprecise normal prior is used. The surface plots of an imprecise posterior expectation with different sample sizes $\xi_0 = 1, 2, 5$, and 10 over the imprecise normal prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\xi_2, \xi_1)|0.5 \leq \xi_2 \leq 1, -1 \leq \xi_1 \leq 1\}$).



The imprecise posterior expectations in the plots labelled $n = 1$ and $n = 2$ are not linear while those in the plots labelled with $n = 5$ and $n = 10$ seem to be nearly linear. Consequently, it is then anticipated that the linear programming technique employed in the proposed methodology can efficiently search for the optimal solution of hyperparameters ξ_2 and ξ_1 that either maximizes or minimizes the posterior expectation of $E_n(\theta|\mathbf{y})$ over the imprecise posterior. The two dashed horizontal squares in the surface plot are the reference boxes that show the extreme posterior expectations over the region $\mathscr{R}_0$ in the two-dimensional hyperparameter space of ξ_2 and ξ_1. The maximum and the minimum are found at $(\xi_1, \xi_2) = (1.0, 0.1)$ and $(\xi_1, \xi_2) = (-1.0, 0.1)$, respectively, for all different sample sizes from the surface plots. These optimal hyperparameters are also ascertained by examining whether or not the solutions found by the linear programming technique coincide with those searched by the `R` function `constrOptim()` (as studied in Subsection 3.3.5). This numerical comparison is presented in Table 4.1. The degree of imprecision $\Delta_n(\theta|\mathbf{y})$ is graphically represented by the gap between the two dashed squares on the vertical axis of the surface plot, and the corresponding numerical quantities are also reported in Table 4.1. A decrease on the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ is observed as the size n of a sample increases.

Table 4.1: Comparison of optimal solutions of hyperparameters (ξ_1, ξ_2) for the imprecise posterior expectation $E_n(\theta|\mathbf{y})$ with different sample sizes $\xi_0 = 1, 2, 5$, and 10. Optimal solutions are found by using the `R` built-in function `constrOptim()` and using the linear programming (LP) approach adopted in the proposed methodology, respectively. The imprecise normal prior is described by characterizing the region $\mathscr{R}_0 = \{(\xi_2, \xi_1)|0.1 \leq \xi_2 \leq 1, -1 \leq \xi_1 \leq 1\}$.

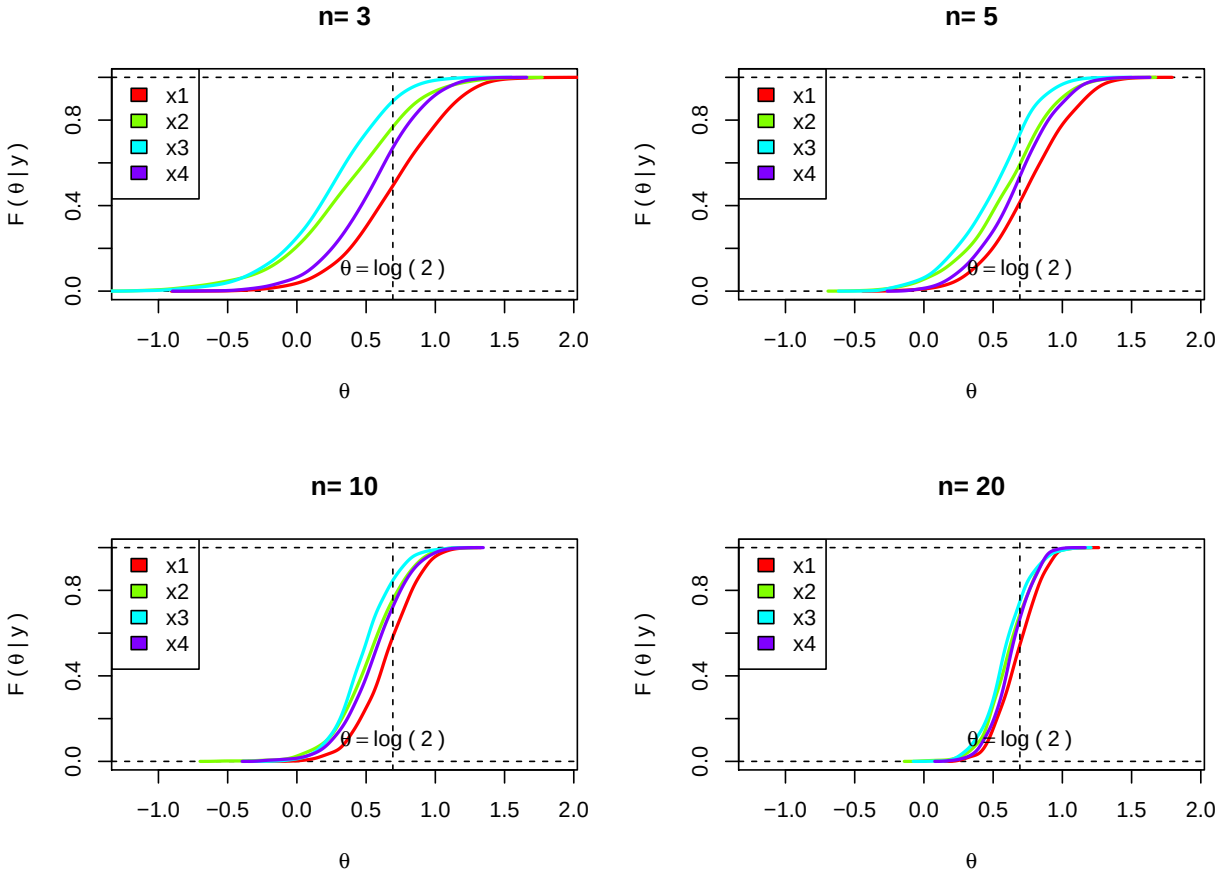
	Optimum		
	$\hat{\xi}_1$	$\hat{\xi}_2$	$E_n(\theta\|\mathbf{y})$
$\xi_0 = 1$			
`constrOptim` (min)	-1	0.1	-0.342
LP Approach	-1	0.1	-0.343
`constrOptim` (max)	1	0.1	0.873
LP Approach	1	0.1	0.868

Table 4.1: *(continued)*

	Optimum		
	$\hat{\xi}_1$	$\hat{\xi}_2$	$E_n(\theta\|\mathbf{y})$
$\xi_0 = 2$			
`constrOptim` (min)	-1	0.1	-0.204
LP Approach	-1	0.1	-0.212
`constrOptim` (max)	1	0.1	0.543
LP Approach	1	0.1	0.540
$\xi_0 = 5$			
`constrOptim` (min)	-1	0.1	-0.092
LP Approach	-1	0.1	-0.095
`constrOptim` (max)	1	0.1	0.259
LP Approach	1	0.1	0.258
$\xi_0 = 10$			
`constrOptim` (min)	-1	1.0	0.330
LP Approach	-1	1.0	0.330
`constrOptim` (max)	1	0.1	0.496
LP Approach	1	0.1	0.495

To examine the focusing behaviour of an imprecise posterior, a sample of size 20 is taken from the standard Poisson sampling model with a mean of $\mu = 2$ (The random seed `18372342` is used for this generation). The imprecise inference is carried out sequentially with the first 3, 5, 10, and 20 samples using the proposed methodology. A probability box is constructed using the imprecise posterior resulting from each imprecise inferential process, and labelled $n = 3$, $n = 5$, $n = 10$, and $n = 20$ as shown in Figure 4.4. The labels `x1`, `x2`, `x3`, and `x4` associated the imprecise posterior are the identification numbers of extreme points characterizing the region $\mathscr{R}$ on the hyperparameter space of ξ_1 and ξ_2. Note that the probability band enclosed by the two posterior distributions labelled `x1` (corresponds to the maximum posterior expectation) and `x3` (corresponds to the minimum posterior expectation) are becoming focused upon the vertical line which lies at $\theta = \log(2)$ as a result of observing a greater number of samples. $\log(2)$ is the value which was used for simulating the Poisson random variates in the log scale of the canonical parameter θ.

Figure 4.4: Probability boxes of an imprecise normal prior after observing a sample of size $n = 3, 5, 10$, and 20 which are simulated from the standard Poisson sampling model with a mean of $\mu = 2$ (i.e., corresponds to the canonical parameter $\theta = \log 2$).



Notes in Numerical Computation

From the examination of the three primary behaviours in the proposed methodology, the $\mathfrak{B}$-formulation is a useful formulation strategy when a prior distribution is no longer conjugate for the canonically parametrized Poisson likelihood. However, one of the goals in this thesis work is seeking to provide a reliable and efficient numerical method for the proposed methodology. It is prudent at this time to discuss a computational difficulty associated with $\mathfrak{B}$-formulation due to the fact the integrals needed for computing the quantity of a posterior expectation in (4.8) cannot be analytically evaluated.

To ensure the correctness of a numerically evaluated quantity, the quantities produced from two (or three) numerical methods are compared to three decimal places. The most commonly used numerical methods used throughout the examples in this section are the Metropolis-Hastings algorithm (Hastings, 1970) and the Laplace approximation (Tierney and Kadane, 1986). Importance sampling or the adaptive quadrature method is selectively used when a discrepancy occurs between the results obtained when using the Metropolis-Hastings and the Laplace approximation. Table 4.2 exemplifies this numerical comparison.

The results in this table are taken from the example that was used to illustrate the focusing behaviour of an imprecise posterior. The quantities of the posterior expectation evaluated at each of the extreme points `x1`, `x2`, `x3`, and `x4` are summarized using numerical methods over different sample sizes. The missing values in this table imply a failure of evaluation because of the non-finite value returned from the function `integrate()`. The computing time (in seconds, s) needed for evaluating a single posterior expectation (i.e., at a single extreme point) is reported as well in the last column of this table, and the function `proc.time()` is used for measuring the single-thread run-time processed in `R`. Note that the author's computing environment is Ubuntu Linux 13.10 (64 bits) on the Intel quadcore i5-3230M CPU (2.60 GHz) with 7.5 GB RAM.

Table 4.2: Comparison of the quantities of the posterior expectation $E_n(\theta|\mathbf{y})$ evaluated by the three different numerical methods (the adaptive quadrature, the Metropolis-Hastings algorithm, and the Laplace approximation) at each extreme point `x1`, `x2`, `x3` and `x4` of the region $\mathscr{R}_0 = \{(\xi_2, \xi_1)|0.5 \leq \xi_2 \leq 1, -1 \leq \xi_1 \leq 1\}$ that describes the imprecise normal prior. A sample of size $n = 3, 5, 10$, and 20 which are drawn from the standard Poisson sampling model with a mean of $\mu = 1$ is used for this quantification, and the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ is computed. The computation time (in seconds, s) elapsed for evaluating the posterior expectation $E_n(\theta|\mathbf{y})$ over all extreme points is listed in the last column.

	`x1`	`x2`	`x3`	`x4`	$\Delta_n(\theta\|\mathbf{y})$	Time (s)
$\xi_0 = 3$						
Adaptive Quadrature						0.007
Laplace Approximation	0.677	0.349	0.267	0.530	0.411	0.010
Metropolis-Hastings	0.697	0.350	0.239	0.537	0.458	0.420

Table 4.2: *(continued)*

	x1	x2	x3	x4	$\Delta_n(\theta\|\mathbf{y})$	Time (s)
$\xi_0 = 5$						
Adaptive Quadrature						0.003
Laplace Approximation	0.768	0.585	0.490	0.657	0.278	0.004
Metropolis-Hastings	0.757	0.588	0.494	0.659	0.264	0.426
$\xi_0 = 10$						
Adaptive Quadrature	0.636	0.531	0.479	0.579	0.157	0.002
Laplace Approximation	0.636	0.531	0.479	0.579	0.157	0.004
Metropolis-Hastings	0.645	0.522	0.479	0.554	0.166	0.436
$\xi_0 = 20$						
Adaptive Quadrature	0.664	0.613	0.582	0.632	0.082	0.003
Laplace Approximation	0.664	0.613	0.582	0.632	0.082	0.005
Metropolis-Hastings	0.673	0.611	0.587	0.630	0.086	0.421

Regarding the use of the Metropolis-Hastings algorithm, the kernel $\exp(-\xi_2\theta^2+\xi_1\theta-\xi_0 e^{\theta})$ of a newly defined family of distributions in (4.4) is used as an objective function from which samples are needed to be drawn. The normal distribution with a mean of μ and a variance of σ^2 is employed for a proposal distribution, and the mean μ and variance σ^2 defaults to the sample mean $\bar{\mathbf{y}}$ and sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{\mathbf{y}})^2$. When a single sample (i.e., $n = 1$) is given, the variance σ^2 defaults to 1. The length of the Markov chain and the burn-in period are set to 2000 (or 10000) and 500 by default, respectively. Winkelmann (2008) noted these values are commonly used in practice (p. 244). Experience with this algorithm reveals that a Markov chain of $10,000$ is preferred; however, it is also found that a length $2,000$ is sufficient for use in most of the illustrations. Thus for reasons of economy, $2,000$ is the default for the `ipeglim` package. These default values can be changed by the `control` option when the function `update()` is used. The following command illustrates how to change the default values of the lengths of Markov chain and the burn-in period

```
> update(obj=cmfit, method="MH", apriori="lgamma",
+        control=list(len.chain=1e4, len.burnin=1e3))
```

and changes the default values of mean μ and standard deviation s of a proposal distribution

```
> update(obj=cmfit, method="MH", apriori="lgamma",
+        proposal=list(mean=mean(y), sd=sd(y)))
```

Regarding the use of the Laplace approximation, the equation (4.8) is used as an objective function for direct evaluation. The score function and the Hessian matrix needed for implementing this Laplace approximation are noted in Appendix A.3. Note that the condition required for the standard Laplace approximation is that the objective function to be estimated should be positive as noted in Section 2.3.3; however, the working objective function in this proposed methodology does not meet this requirement since the canonical parameter θ ranges from minus infinity to infinity. According to the study done by Tierney et al. (1989), one resolution when working with a non-positive objective function is to add a large constant, and then subtract that constant after completion of the estimation (p. 713). However, no additional information has been found regarding determining the exact size of that constant. Through repeated test results, the constants `exp(1e1)` and `exp(2e1)` provide reliable estimates. As illustrated in the case of the Metropolis-Hastings algorithm, these default values can be changed by the option `control` when the function `update()` is used as below:

```
> update(obj=cmfit, method="LA", apriori="lgamma",
+        control=list(const=2e1))
```

During the testing of the `ipeglim` package, the following observations were made regarding the use of the different numerical methods. First, the Metropolis-Hastings algorithm is considerably slower than the Laplace approximation method to reach an estimate. Second, both the Metropolis-Hastings algorithm and the Laplace approximation are generally successful in achieving estimates despite the small sample size (roughly less than 10); however, the adaptive quadrature method fails when examining cases characterized by a comparatively small sample size. Finally, estimates produced using an importance sampling method are not stable at two or three decimal places when compared to the estimates produced using other numerical methods. Because of these unstable numerical results, the importance sampling procedure is not fully automated in the `ipeglim` package and is currently under examination. It may be due to the use of a normal distribution with a sample mean and a sample variance

as a proposal sampling density function; thus, further efforts are needed to employ other families of proposal distributions.

4.2 Zero-Truncated Poisson Sampling Model

The previous section presented a conjugate relationship between two families of prior and posterior distributions that is built by the newly introduced formulation strategy called the $\mathfrak{B}$-formulation for the instance where a family of normal distributions is considered as a prior probability measure for a standard Poisson likelihood which is canonically parametrized. Another instance of lacking a typical conjugate prior probability measure can be found when a standard Poisson sampling model is truncated at zero. The question then is as whether or not the $\mathfrak{B}$-formulation can be applied to build a conjugate relationship for this case as shown in the previous section.

Suppose that a standard Poisson sampling model $f(y|\mu) = e^{-\mu}\mu^y/y!$ having the mean parameter μ is given. For performing the proposed imprecise inferential framework, the given sampling model $f(y|\mu)$ needs to be canonically parametrized and examined first to determine if it is a member of an exponential family. Since a probability of zero count is $f(0|\mu) = e^{-\mu}$, a probability density function of a zero-truncated Poisson sampling model $f(y|y > 0, \theta)$ can be written as

$$f(y|y > 0, \theta) = \frac{1}{y!}\exp(y\theta - e^{\theta} - \log(1 - e^{-\exp(\theta)})), \tag{4.9}$$

which continues to belong to an one-parameter exponential family of distributions having the same canonical parameter $\theta = \log(\mu)$ as a standard Poisson sampling model $f(y|\mu)$. Note that the log-normalizing constant for the zero-truncated Poisson sampling model $f(y|y > 0)$ in (4.9) is $e^{\theta} + \log(1 - e^{-\exp(\theta)})$.

Consider now a situation where n i.i.d. samples drawn from the zero-truncated Poisson sampling model $f(y|y > 0)$ is observed. Assume that a family of log-gamma prior distributions given in (3.9) is considered to address the problem of estimating the canonical parameter θ of this sampling model $f(y|y > 0)$. For the convenience of readers, note that the kernel of

a log-gamma distribution in (3.9) is $\exp(\alpha\theta - \beta e^{\theta})$ and α and β are the hyperparameters. The kernel of the resultant family of posterior distributions $p(\theta|\mathbf{y})$ is then given by

$$p(\theta|\mathbf{y}) \propto \exp\left\{(n\bar{\mathbf{y}} + \alpha)\theta - (\beta + n)e^{\theta} - n\log(1 - e^{-\exp(\theta)})\right\}, \tag{4.10}$$

which differs from the one of a family of log-gamma prior distributions in (3.9). However, it is noticed that (4.10) is a family of distributions that has been derived by the $\mathfrak{B}$-formulation as shown in (4.4). That is, the unnormalized posterior distribution in (4.10) can be represented by a three-parameter exponential family of distributions such that

$$p(\theta|\mathbf{y}) \propto \exp\left\{\xi_2\theta - \xi_1 e^{\theta} - \xi_0\log(1 - e^{-\exp(\theta)})\right\}, \tag{4.11}$$

where ξ_2, ξ_1, and ξ_0 are hyperparameters given by

$$\xi_2 = n\bar{\mathbf{y}} + \alpha, \quad \xi_1 = \beta + n, \quad \xi_0 = n. \tag{4.12}$$

Note that, if there is no observation (i.e., $n = 0$), then the kernel of the distribution defined in (4.11) becomes a family of log-gamma distributions.

The exact same analogy is applied to the case where a family of normal distributions is considered as a prior probability measure for the zero-truncated Poisson sampling model $f(y|y > 0)$. Since the framework for deriving the kernel of a posterior distribution using the given functional form of a normal prior distribution with a zero-truncated Poisson sampling model $f(y|y > 0)$ is identical to the framework described in the previous section 4.1 when using a standard Poisson sampling model $f(y)$ is used, only the key result is noted here. Following the $\mathfrak{B}$-formulation , the kernel of a canonically parametrized posterior distribution $p(\theta|\mathbf{y})$ is found as

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left\{-\xi_2\theta^2 + \xi_1\theta - \xi_0\left[e^{\theta} + \log(1 - e^{-\exp(\theta)})\right]\right\}, \tag{4.13}$$

where ξ_2, ξ_1, and ξ_0 are the hyperparameters given by

$$\xi_2 = \frac{1}{2\tau^2}, \quad \xi_1 = \left(n\bar{y} + \frac{\nu}{\tau^2}\right), \quad \xi_0 = n, \tag{4.14}$$

which are identical to those in (4.6). The difference between the kernel in (4.4) and the kernel in (4.13) is the presence of a renormalizing constant $\log(1 - e^{-\exp(\theta)})$ which is contributed by the truncation of a standard Poisson sampling model $f(y)$ at zero. It is so concluded that the $\mathfrak{B}$-formulation is a generalized approach for formulating a conjugate prior measure for both the standard Poisson $f(y)$ and the zero-truncated Poisson $f(y|y > 0)$ sampling models when either a family of log-gamma or normal prior distributions is considered for an imprecise inference using the proposed methodology.

The `ipeglim` package provides a convenient approach for inferring the canonical parameter of either a standard Poisson or a zero-truncated Poisson sampling model with a given family of either log-gamma or normal prior distributions. The `R` code template is provided here for the reader. Please be aware that the option `ztrunc` in the function `model()` must be set to `TRUE` when a zero-truncated Poisson samples is used.

```
> library(ipeglim)
> lc0 <- list(lhs = rbind(diag(2), -diag(2)), rhs = c(0,
+     0, -4, -4))
> y <- "put data in the form of a 'vector'"
> m2fit <- model(formula = y ~ 0, ztrunc = TRUE, dist = "poisson")
> cmfit <- iprior(obj = m2fit, eqns = lc0)
> plot(cmfit)
> op <- update(obj = cmfit, method = "LA", apriori = "lgamma")
> sop <- summary(op)
> plot(sop)
> pbox(sop, pretty = TRUE)
```

Note that the `R` built-in function `ppois()`, which computes a standard Poisson cumulative function, is used on the computation of a renormalizing constant $\log(1 - e^{-\exp(\theta)})$, in the `ipeglim` package, in order to reduce the computational problems associated with the rounding error and the numerical overflow. For example, the direct computation of $\log(1 - e^{-\exp(\theta)})$ gives 0, 0, -4.10782519111309e-15, -36.7368005696771, -36.7368005696771, and -Inf when $\theta = 4.5, 4, 3.5, -36, -36.5, -37$, and -38 are evaluated while the use of `ppois()` returns -8.0548340897409e-40, -1.94233760495641e-24, -4.15089692010905e-15, -36, -36.5, -37, and -38

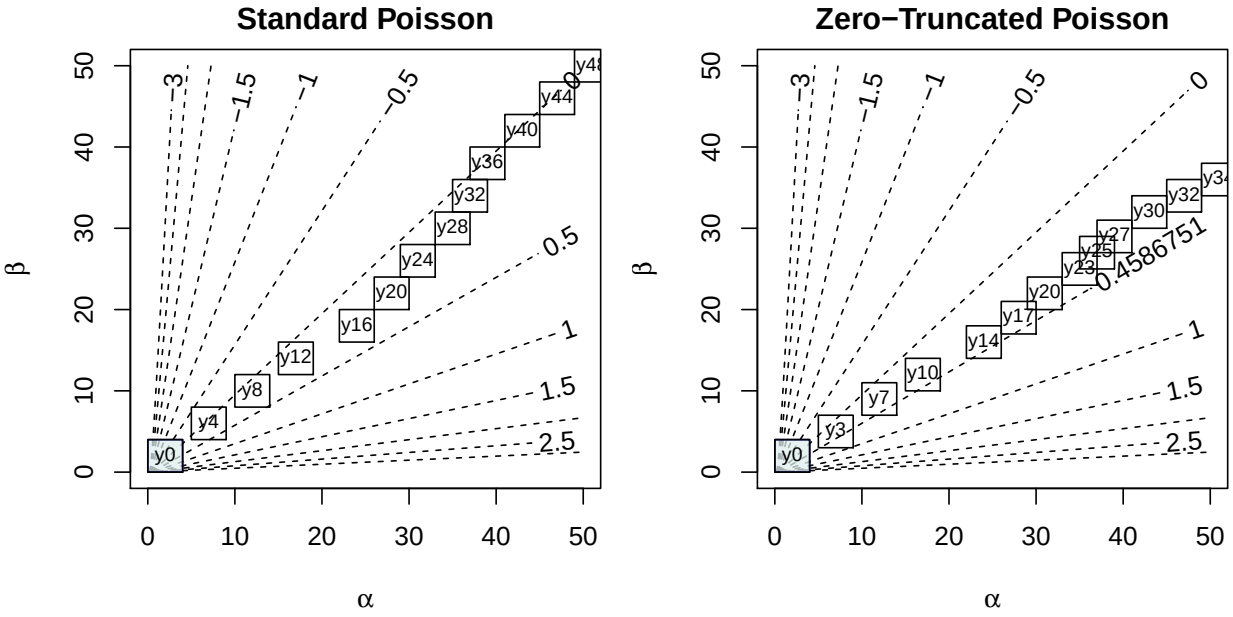
which are the anticipated results. Hence, it is hypothesized that the estimate of a posterior expectation produced using the `ipeglim` package would be more accurate and produce less numerical failure when compared to an estimate produced using direct computation.

From this point on three primary behaviours of the proposed methodology are re-examined using zero-truncated Poisson samples. Please observe that the numeric and graphic summaries of the resultant imprecise estimate are reported briefly since a detailed description demonstrating how to lead an imprecise inference and construct three primary behaviours of the proposed methodology are provided multiple times.

Suppose that 48 samples are taken from a standard Poisson sampling model $f(y|\mu)$ with a mean of $\mu = 1$. Then, the zero-truncated Poisson samples are obtained by removing the value zero from the generated samples. The corresponding mean parameter μ_T of a zero-truncated Poisson sampling is then $E(Y|Y > 0) = \mu_T = \mu/(1 - e^{-\mu})$ =1.582 and the canonical parameter $\theta_T = \log(\mu_T)$ for μ_T has the value of 0.4587. (The probability mass function, the moment generating function, and the first three moments of a zero-truncated Poisson distribution $f(y|y > 0)$ are noted in Appendix B.1.) For inferring the canonical parameter θ, natural imprecise prior is defined by characterizing the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$. The only task that remains is a numerical computation to search for the extremes of a resulting imprecise posterior using the `ipeglim` package.

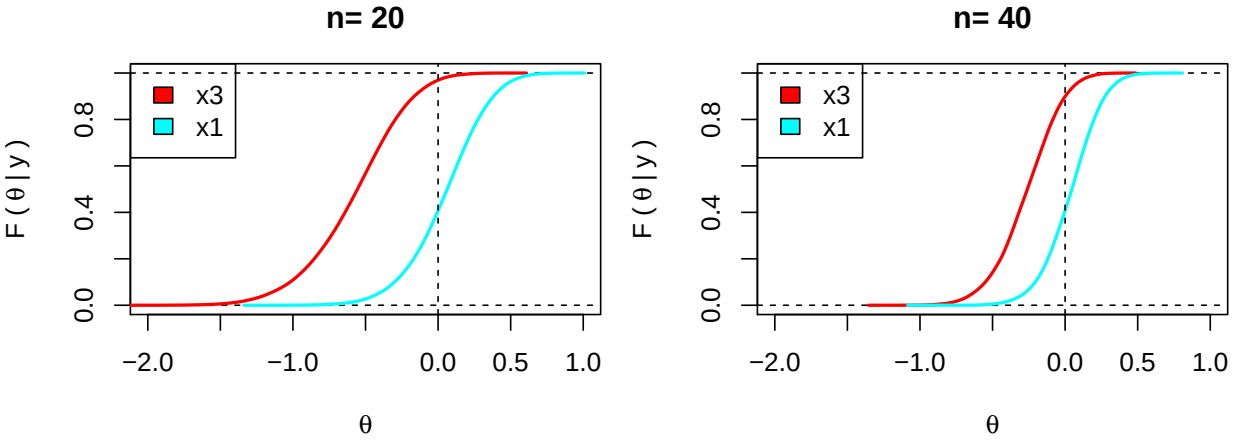
The translation behaviour of *your* natural imprecise prior is geometrically represented for both cases of where standard Poisson samples and zero-truncated Poisson samples are used for imprecise inference as shown on the left and right panels, respectively, in Figure 4.5. The level set of a prior expectation $E(\theta)$ is presented beginning from -3 to 3 by 0.5 using dashed curves in each plot. However, for the use of zero-truncated Poisson samples, the curve labelled with the zero value corresponding to the canonical parameter θ of the standard Poisson mean parameter μ is replaced by one with the value 0.4587 that corresponds to the canonical parameter θ_T of the zero-truncated Poisson mean parameter μ_T=1.582. It shows that *your* natural imprecise prior moves consistently with the reference line (i.e., $E(\theta) = 0$ for the use of standard Poisson samples and $E(\theta_T)$ =0.459 for the use of zero-truncated Poisson samples).

Figure 4.5: Translation behaviour of the natural imprecise log-gamma prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$) for cases of when using a standard Poisson and a zero-truncated Poisson samples are used for the imprecise inference. Samples are simulated from the standard Poisson sampling model with a mean of $\mu = 1$, and zero values are removed for zero-truncated samples. The level set of a prior expectation $E(\theta)$ is referenced beginning from -3 to 3 by 0.5 using dashed curves. For the use of zero-truncated Poisson samples (in the right panel), the curve labelled with a zero value corresponding to the canonical parameter θ of the standard Poisson mean parameter μ is replaced by one with the value 0.4587 that corresponds to the canonical parameter θ_T of the zero-truncated Poisson mean parameter μ_T=1.582.



The focusing behaviour of an imprecise posterior is presented by two probability boxes listed on the left and right panels in Figure 4.6. The labels $n = 20$ and $n = 40$ on each probability box imply that 20 and 40 zero-truncated samples are used for inferring the canonical parameter θ of a standard Poisson sampling model $f(y)$. It shows that the probability band enclosed by two posterior distributions labelled `x1` and `x3` is becoming focused upon the dashed reference line located at $\theta = 0$. To clarify the parameter to be estimated in this imprecise inference, note that the zero-truncated Poisson sampling model $f(y|y > 0)$ is a scaled standard Poisson sampling model $f(y)$ as shown in (4.9). Hence, although zero-truncated Poisson random variates are used, the canonical parameter θ of a standard Poisson sampling model $f(y)$ is estimated.

Figure 4.6: Focusing behaviour of an imprecise posterior for the case when the natural imprecise log-gamma prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$) is used with zero-truncated samples for imprecise inference. Probability boxes of the imprecise posterior are constructed using two different samples of size $n = 20$ and $n = 40$ as shown in the left and right panels, respectively. Samples are taken from the standard Poisson sampling model with a mean of $\mu = 1$ (i.e., $\theta = \log 1 = 0$).



The results in Table 4.3 support that the enclosing posterior distributions labelled `x1` and `x3` of each probability band have the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$ and the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, respectively. The results also convey the coincidence of the imprecise estimate produced using different numerical methods over different sample sizes. The missing values shown in Table 4.3 imply a failure of estimation when using the numerical method of adaptive quadrature.

Table 4.3: Comparison of the quantities of the posterior expectation $E_n(\theta|\mathbf{y})$ evaluated by the different numerical methods (the Metropolis-Hastings algorithm, the Laplace approximation, and the adaptive quadrature) at each extreme point `x1`, `x2`, `x3` and `x4` of the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$ that describes the imprecise log-gamma prior. Samples are taken from the standard Poisson sampling model with a mean of $\mu = 1$ and the zero values removed. The first n samples are used in a sequence for this quantification ($n = 4, 20$, and 40). Missing values imply a failure of numerical quantification.

	x1	x2	x3	x4
$n = 4$				
Laplace Approximation	0.716	−0.398	−1.364	−0.137
Metropolis-Hastings	0.717	−0.399	−1.347	−0.129
Adaptive Quadrature				−0.139

Table 4.3: *(continued)*

	x1	x2	x3	x4
$n = 20$				
Laplace Approximation	0.055	−0.288	−0.569	−0.210
Metropolis-Hastings	0.052	−0.297	−0.574	−0.210
Adaptive Quadrature	0.055	−0.289	−0.570	−0.210
$n = 40$				
Laplace Approximation	−0.007	−0.159	−0.293	−0.137
Metropolis-Hastings	−0.009	−0.160	−0.291	−0.136
Adaptive Quadrature	−0.008	−0.159	−0.293	−0.137

Table 4.4 also shows the decrease in the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ as the size n of a sample increases.

Table 4.4: The maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$, the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, and the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ as a zero-truncated Poisson sample is taken sequentially. The natural imprecise log-gamma prior (i.e., corresponds to the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$) is used and the samples are taken from the standard Poisson sampling model with a mean of 1 (i.e., $\theta = \log 1 = 0$). The summary is reported at the sample sizes $n = 4, 12, 20, 28, 36$, and 40.

	$n = 4$	$n = 12$	$n = 20$	$n = 28$	$n = 36$	$n = 40$
$\underline{E}_n(\theta\|\mathbf{y})$	−1.364	−0.652	−0.569	−0.705	−0.326	−0.293
$\overline{E}_n(\theta\|\mathbf{y})$	0.716	0.258	0.055	−0.173	0.018	−0.007
$\Delta_n(\theta\|\mathbf{y})$	2.080	0.910	0.624	0.532	0.344	0.285

It is intriguing to consider how the imprecise estimate is affected by zero-truncation in a standard Poisson sampling model. To examine this question the following study was designed.

1. Draw a sample of size 15 from a standard Poisson sampling model with a mean of μ;
2. Choose the first n samples from the drawn sample;
3. Apply the proposed methodology to the chosen sample;
4. Remove the zero values from the chosen sample;
5. Apply again the proposed methodology to the zero-truncated sample;
6. Repeat the steps (2)–(5) for $n = 5, 10$, and 15;

7. Repeat the steps (1)–(6) for $\mu = 0.5, 1$, and 2;

The surface plot of the imprecise posterior expectations are listed by the size n of a sample and whether or not zero values are truncated from the chosen sample over different mean parameters throughout Figures 4.7, 4.8, 4.9 when a family of log-gamma prior distributions is used. The identical graphical summary is provided for Figures 4.10, 4.11, 4.12 when a family of normal prior distributions is used.

Three general patterns become apparent from each of the six figures regardless which family of distributions is employed in the imprecise inference. First, the acute angle of the surface of the posterior expectation associated with the zero-truncated Poisson sampling model is greater than the one associated with a standard Poisson sampling model over all different sample sizes. Hence, the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ associated with the zero-truncated Poisson sampling model is larger when compared to the one associated with standard Poisson sampling model. Secondly, the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ for both the standard and zero-truncated Poisson sampling models decreases as the size of a sample increases. Lastly, the surface of the posterior expectation becomes linear even though the size n of a sample is small when the Poisson mean parameter is large.

Figure 4.7: Comparison of the imprecise posterior expectation for cases of where standard Poisson and zero-truncated Poisson samples are used on the given natural imprecise log-gamma prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$). Samples are generated from the standard Poisson sampling model with a mean of 0.5 (i.e., $\theta = \log 0.5$ = -0.693), and different samples sizes $n = 5, 10$, and 15 are selected.

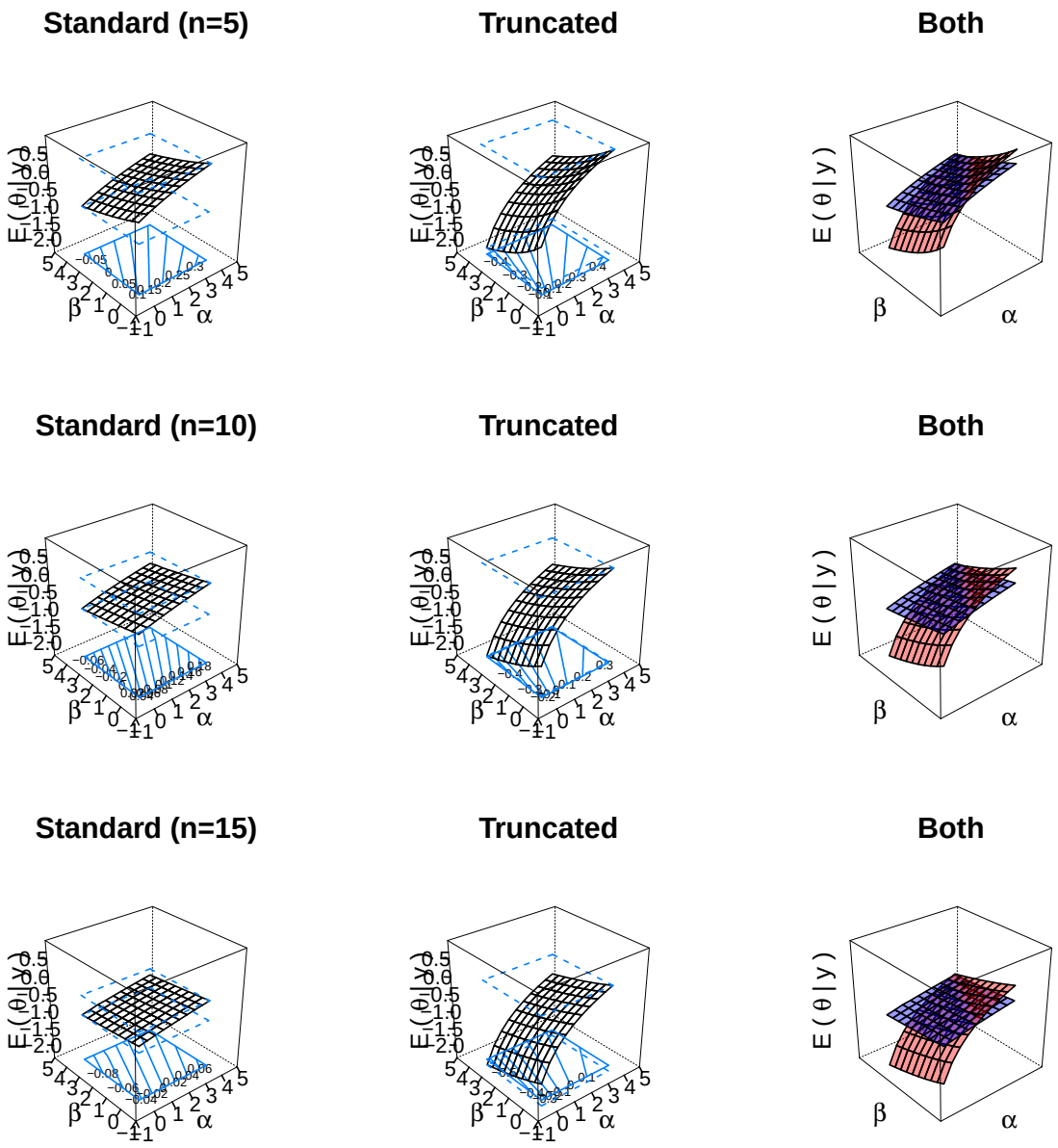


Figure 4.8: Comparison of the imprecise posterior expectation for cases of where standard Poisson and zero-truncated Poisson samples are used on the given natural imprecise log-gamma prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$). Samples are generated from the standard Poisson sampling model with a mean of 1 (i.e., $\theta = \log 1.0 = 0$), and different samples sizes $n = 5, 10$, and 15 are selected.

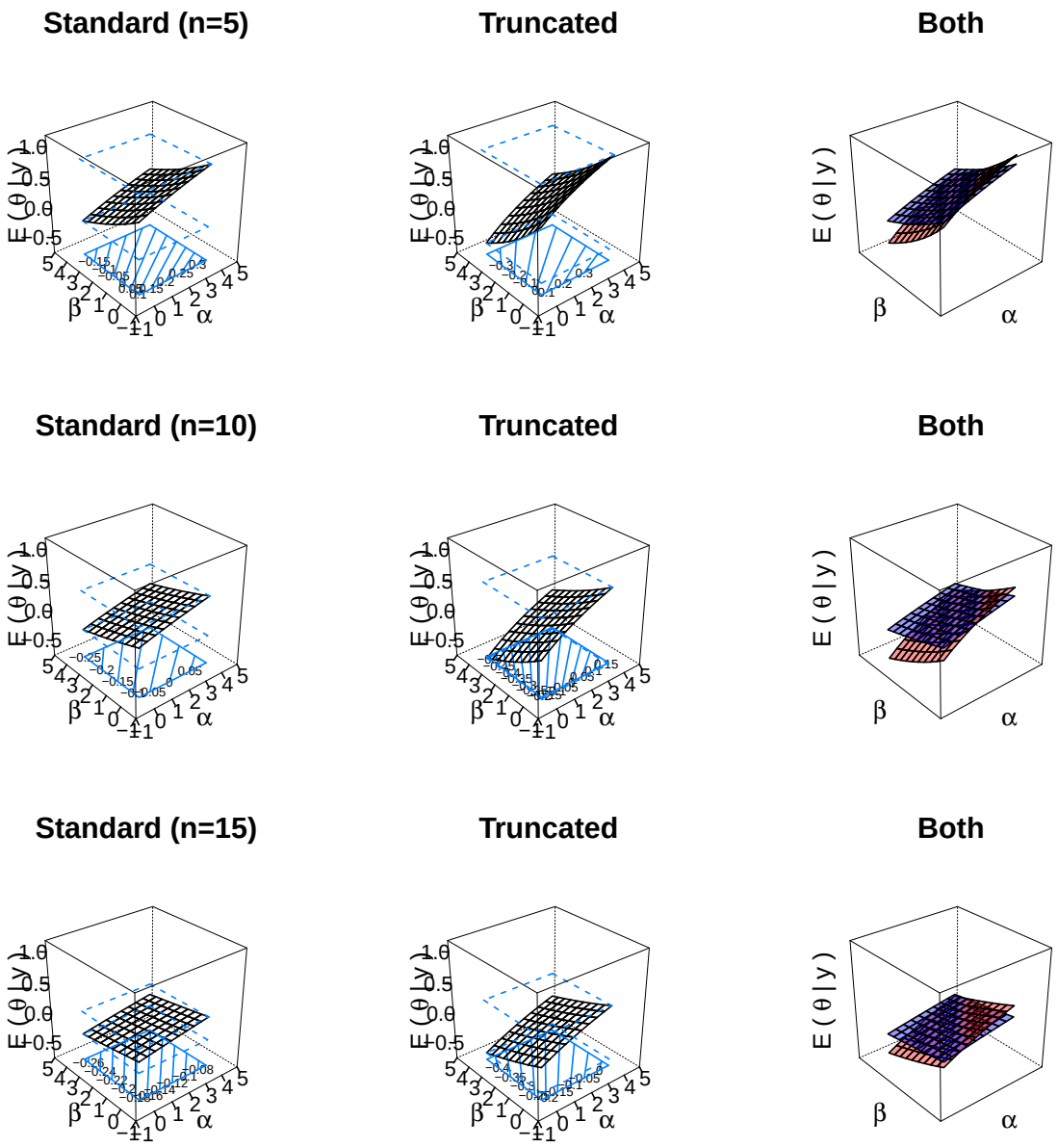


Figure 4.9: Comparison of the imprecise posterior expectation for cases of where standard Poisson and zero-truncated Poisson samples are used on the given natural imprecise log-gamma prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 4, 0 \leq \beta \leq 4\}$). Samples are generated from the standard Poisson sampling model with a mean of 2 (i.e., $\theta = \log 2.0 = 0.693$), and different samples sizes $n = 5, 10$, and 15 are selected.

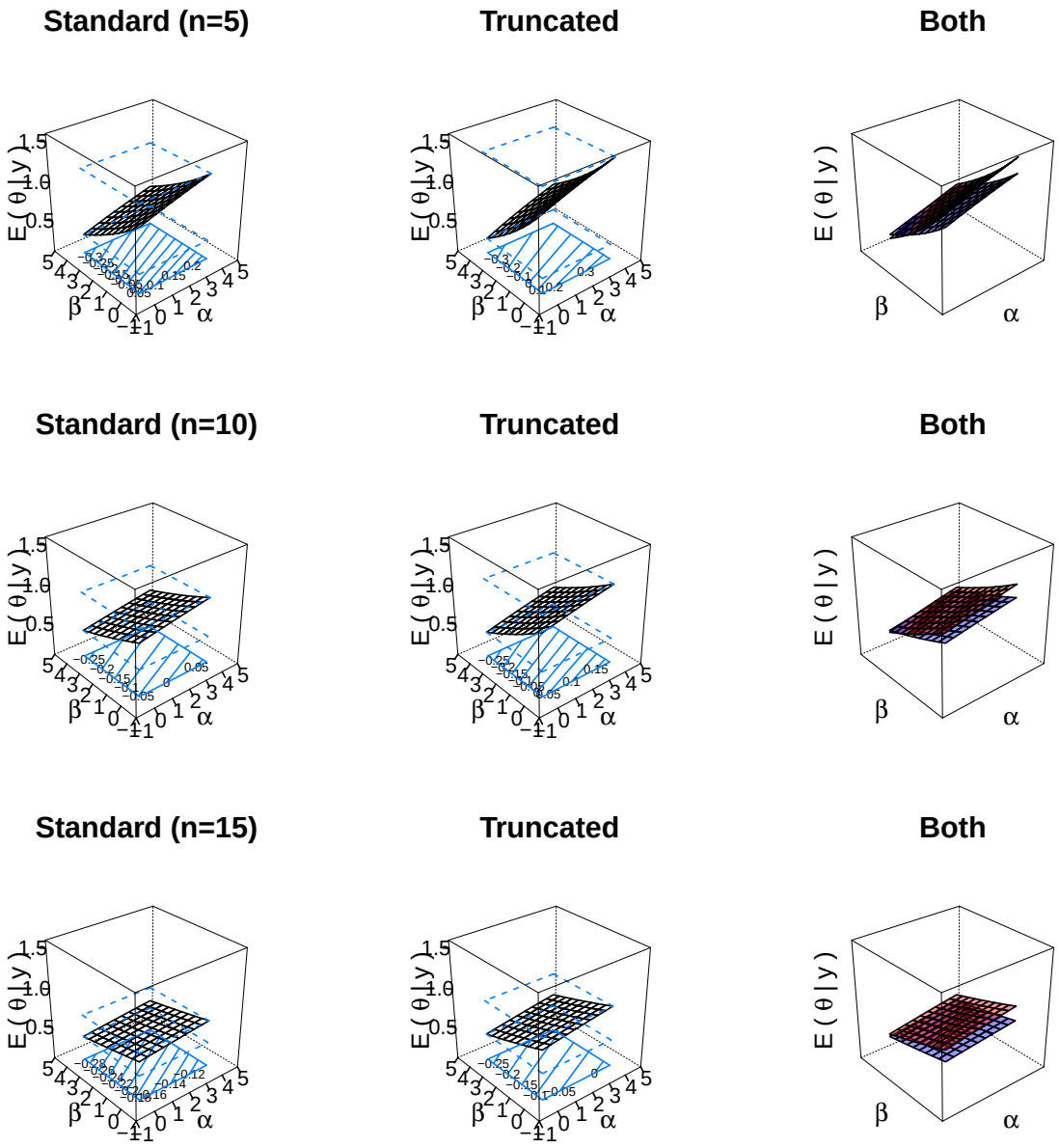


Figure 4.10: Comparison of the imprecise posterior expectation for cases of where standard Poisson and zero-truncated Poisson samples are used on the given natural imprecise normal prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\xi_2, \xi_1) | -1 \leq \xi_2 \leq 1, 0.15 \leq \xi_1 \leq 1\}$). Samples are generated from the standard Poisson sampling model with a mean of 0.5 (i.e., $\theta = \log 0.5 =$ -0.693), and different samples sizes $n = 5, 10,$ and 15 are selected.

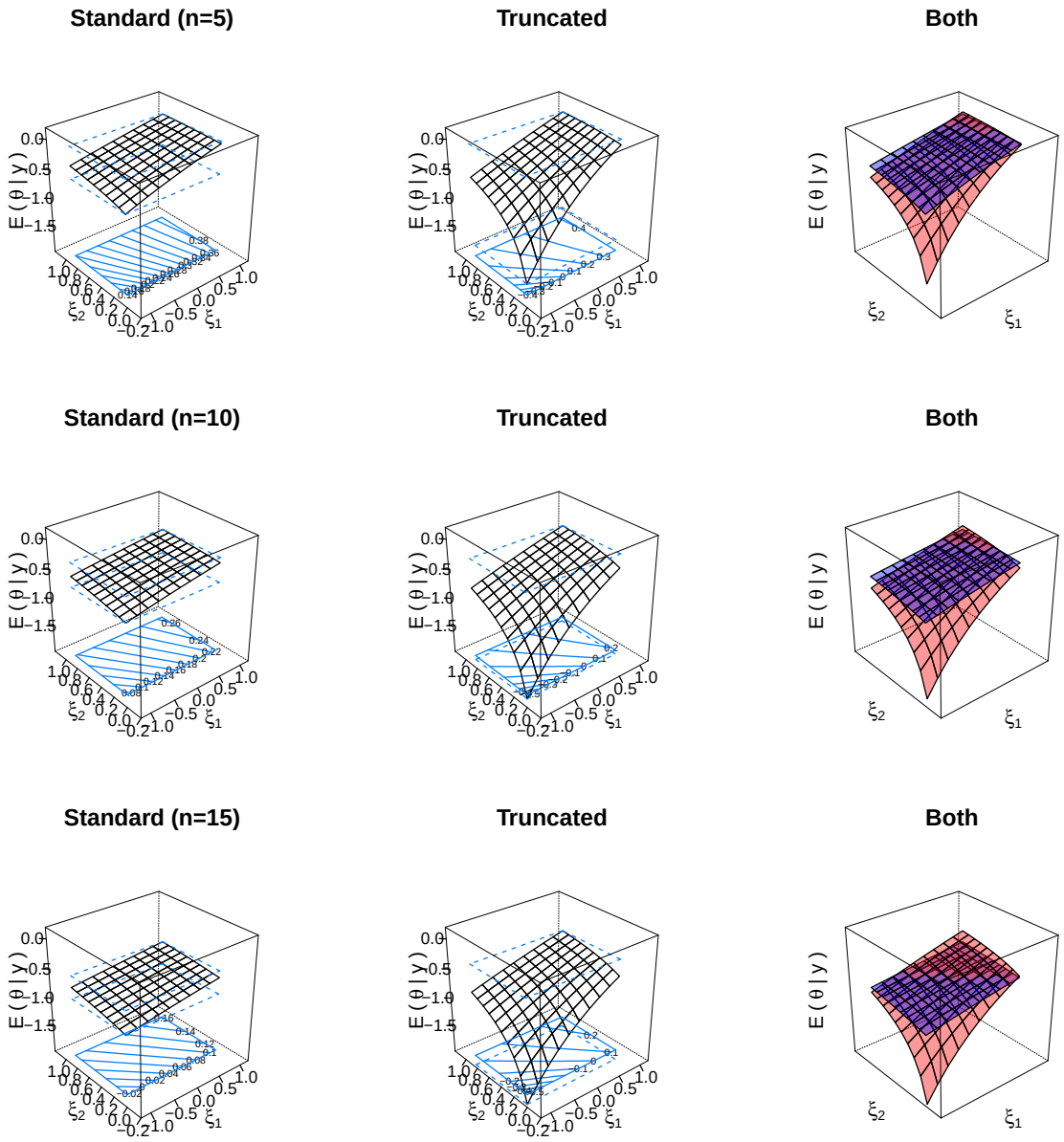


Figure 4.11: Comparison of the imprecise posterior expectation for cases of where standard Poisson and zero-truncated Poisson samples are used on the given natural imprecise normal prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\xi_2, \xi_1) | -1 \leq \xi_2 \leq 1, 0.15 \leq \xi_1 \leq 1\}$). Samples are generated from the standard Poisson sampling model with a mean of 1 (i.e., $\theta = \log 1.0 = 0$), and different samples sizes $n = 5, 10$, and 15 are selected.

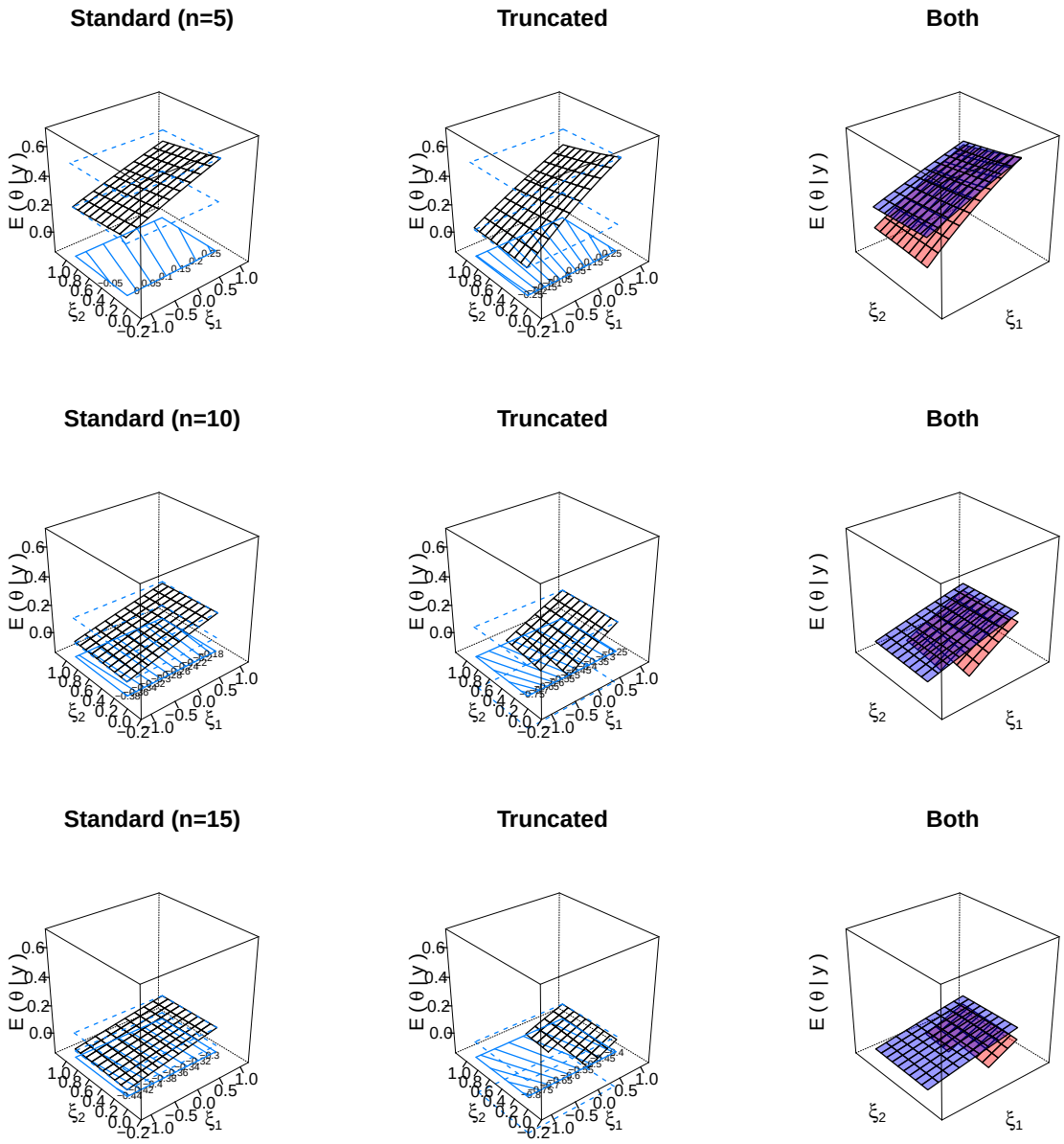
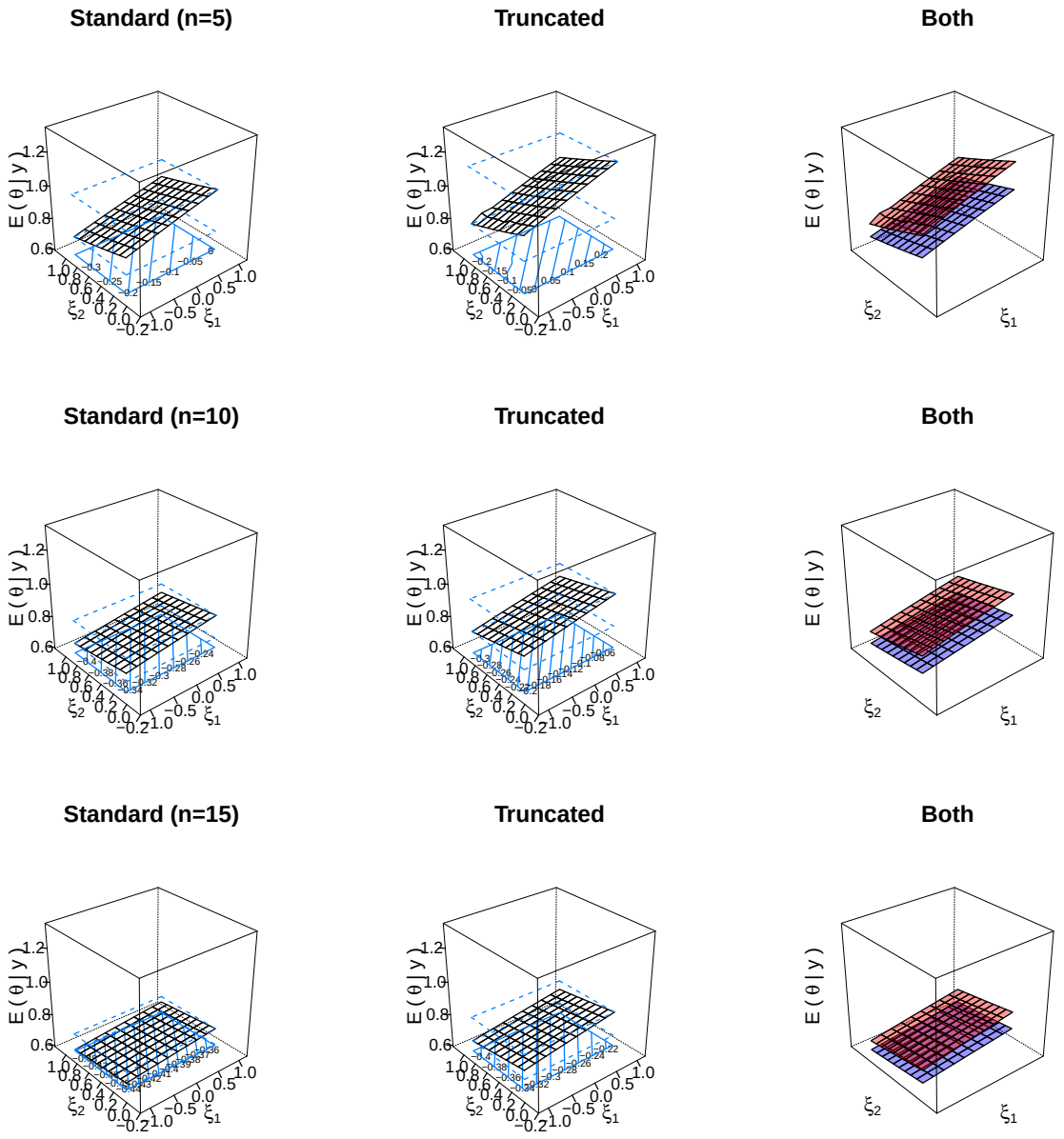


Figure 4.12: Comparison of the imprecise posterior expectation for cases of where standard Poisson and zero-truncated Poisson samples are used on the given natural imprecise normal prior (i.e., corresponding to the region $\mathscr{R}_0 = \{(\xi_2, \xi_1) | -1 \leq \xi_2 \leq 1, 0.15 \leq \xi_1 \leq 1\}$). Samples are generated from the standard Poisson sampling model with a mean of 2 (i.e., $\theta = \log 2.0 = 0.693$), and different samples sizes $n = 5, 10$, and 15 are selected.



4.3 Regression Model

Regression analysis has been widely used to describe the relationship between a response variable and one or more explanatory variables since the generalized linear model (GLM) was introduced by Nelder and Wedderburn (1972). In this section, the proposed imprecise inferential framework is examined under the GLM setup to ensure that the $\mathfrak{B}$-formulation functions appropriately for inferring regression parameters.

Suppose that the mean parameter μ_i of a standard Poisson sampling model is linked to the linear predictor $\mathbf{x}_i'\boldsymbol{\beta}$ by the log-link function as below:

$$\theta_i = \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} \tag{4.15}$$

where θ_i is the canonical parameter of μ, $\mathbf{x}_i = (1, x_{1i}, x_{2i}, \ldots, x_{ip})^T$ is a vector of $(p+1)$ explanatory variables including the intercept term 1 of i-th individual, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a vector of $(p+1)$ regression parameters associated with $\mathbf{x}_i$. The subscript i is the index for the i-th individual of n independently sampled subjects; thus, the response variable $\mathbf{y}$ is a vector of independent observations $(y_1, y_2, \ldots, y_n)$. The Poisson log-likelihood $\log \mathscr{L}(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$ is then given by

$$\log \mathscr{L}(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) \propto \sum_{i=1}^{n} y_i\theta_i - e^{\theta_i} = \sum_{i=1}^{n} \left[y_i \mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta})\right]. \tag{4.16}$$

Consider now a family of $p+1$ dimensional multivariate normal prior distributions $\pi(\boldsymbol{\beta})$ with a mean of b_0 and a variance-covariance matrix B_0 given by

$$\pi(\boldsymbol{\beta}) \;\; = \;\; \exp\left(-\frac{1}{2}\mathrm{tr}(B_0^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^T) + (B_0^{-1}b_0)^T\boldsymbol{\beta} - \mathcal{A}(b_0, B_0)\right), \tag{4.17}$$

where $\mathcal{A}(b_0, B_0) = \frac{1}{2}b_0^T B_0^{-1} b_0$ is the log-normalizer, and B_0^{-1} is a prior precision matrix. In the form of canonical parameters $\Lambda = B_0^{-1}$ and $\eta = B_0^{-1}b_0$,

$$\pi(\boldsymbol{\beta}) \;\; = \;\; \exp\left(-\frac{1}{2}\mathrm{tr}(\Lambda\boldsymbol{\beta}\boldsymbol{\beta}^T) + \eta^T\boldsymbol{\beta} - \mathcal{A}(\eta, \Lambda)\right), \tag{4.18}$$

where $\mathcal{A}(\eta, \Lambda) = \frac{1}{2}\eta^T \Lambda \eta$ is the log-normalizer. That is, a family of normal prior distributions on the regression parameters $MVN(b_0, B_0)$ induces a family of normal prior distributions $MVN(\eta = B_0^{-1}b_0, \Lambda = B_0^{-1})$ having canonical parameters η and Λ. The number of canonical parameters associated with this prior probability measure in (4.18) is $p + 1 + p(p+1)/2$. A family of posterior distributions induced from (4.18) is then

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) \quad \propto \quad \exp\left(-\frac{1}{2}\text{tr}(B_0^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^T) + (B_0^{-1}b_0)^T\boldsymbol{\beta} + \sum_{i=1}^{n}\left[y_i\mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta})\right]\right). \quad (4.19)$$

which is a multiparameter exponential family of distributions. Since it appears that a closed form expression for a normalizing constant of (4.19) does not exist, a posterior expectation $E(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$ needs to be computed numerically.

The exact same analogy is applied for inferring the parameters of a zero-truncated Poisson regression model. The difference between the two Poisson regression models is the presence of a renormalizing constant $\log(1 - e^{-\exp(\theta)})$ in a Poisson likelihood in (4.16). Hence, the unnormalizing posterior distribution $p(\boldsymbol{\beta}|\mathbf{y})$ in (4.19) can be rewritten in a more general form such that

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) \quad = \quad \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - b_0)^T B_0^{-1}(\boldsymbol{\beta} - b_0) + \log\mathscr{L}(\boldsymbol{\beta})\right), \quad (4.20)$$

where $\log\mathscr{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[y_i\mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta})\right]$ for the standard Poisson regression model and $\log\mathscr{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[y_i\mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta}) - \log(1 - e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})})\right]$ for the zero-truncated Poisson regression model. This general form of a posterior distribution $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$ in (4.20) is adopted with the implementation of an estimation algorithm in the `ipeglim` package.

In order to understand the behaviour of the imprecise estimate of regression parameters, a heterogeneous population of size n is considered. The explanatory variable $x_i \sim NID(0, 1)$ is used to generate observations under the model given by

$$\log(\mu_i) = \beta_0 + \beta_1 x_i \quad (4.21)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)^T = (-0.5, 1.0)^T$, $\mathbf{x}_i = (1, x_i)^T$, and $i = 1, 2, \ldots, n$. Count data for a response variable y_i is generated from the standard Poisson sampling model with a mean

of μ_i (for this study, the random seed `18372342` is utilized). Zero-truncated count data are obtained by removing the zero values from the generated samples.

The parameters of the prior variance-covariance matrix B_0 are structured as below:

$$V(\boldsymbol{\beta}) = B_0 = \begin{bmatrix} \sigma_1^{\pi} & 0 \\ 0 & \sigma_2^{\pi} \end{bmatrix} \begin{bmatrix} 1 & \rho_{\pi} \\ \rho_{\pi} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^{\pi} & 0 \\ 0 & \sigma_2^{\pi} \end{bmatrix} \tag{4.22}$$

where σ_1^{π} and σ_2^{π} are the standard deviations of the regression parameters β_0 and β_1, respectively, and ρ_{π} is a correlation coefficient between the regression parameters β_0 and β_1. It is assumed that $\sigma_1^{\pi} = \sigma_2^{\pi} = \sigma_{\pi}$. The change of the imprecise estimate of regression parameters is investigated in this given structure by changing the values of three components: 1. variances $\sigma_{\pi}^2 = \{0.01, 0.05, 0.1\}$ (degree of uncertainty concerning the location of the regression parameters); 2. correlation coefficient $\rho_{\pi} = \{0.0, 0.4, 0.8\}$ (how these parameters are associated); and, 3. sample size $n = \{20, 50, 100\}$ (how influential is a given data to this inference). The imprecise normal prior is represented by characterizing the region $\mathscr{R}_0 = \{(b_0, b_1) | (b_0 - 0)^2 + (b_1 - 0)^2 \leq 1\}$ on the hyperparameter space $\boldsymbol{\Xi} = \{(b_0, b_1) | -\infty < b_0, b_1 < \infty\}$. Thirty extreme points are employed for characterizing this oval object. Certain select cases are graphically reported here to illustrate the general pattern of the imprecise estimate of the regression parameters β_0 and β_1.

The simultaneous change of imprecise posterior expectations $E(\beta_0|\mathbf{y})$ and $E(\beta_1|\mathbf{y})$ for regression parameter β_0 and β_1 are visualized in Figure 4.13, 4.14, and 4.15 in terms of the sample size n, correlation coefficient ρ_{π}, and variance σ_{π}^2. The function `plot()` in the `ipeglim` package is used for producing all plots in these figures. Two oval objects are shown in each plot. The circle coloured blue implies the characterized region $\mathscr{R}_0$ representing the imprecise prior expectation. The other oval object coloured yellow represents the imprecise posterior expectation.

Before creating this oval object, the author tested that all points enclosed by the circle coloured blue translate into other points that form a convex polygon. Connecting the extreme points of that convex polygon ultimately produces this oval object. This fact is annotated in the plot such that the extreme point identification number on the boundary of the circle coloured blue translates to the red points which lie on the boundary of the oval object. Since

Figure 4.13: Translation and focusing behaviours of the imprecise estimate for the regression coefficients of β_0 and β_1 in the standard Poisson and zero-truncated Poisson regression models by changing the sample size $n = 20, 50,$ and 100 with the fixed correlation coefficient $\rho = 0$ and standard deviation $\sigma^2 = 0.01$. The circle coloured blue (i.e., $\mathscr{R}_0 = \{(b_0, b_1)|b_0^2 + b_1^2 \leq 1\}$) is the imprecise normal prior for β_0 and β_1. The ellipsoid coloured yellow is the imprecise posterior expectation for β_0 and β_1. The star symbol is the maximum likelihood estimates for β_0 and β_1. The two solid black lines implies the imprecise posterior predicted response using the first row of the data matrix.

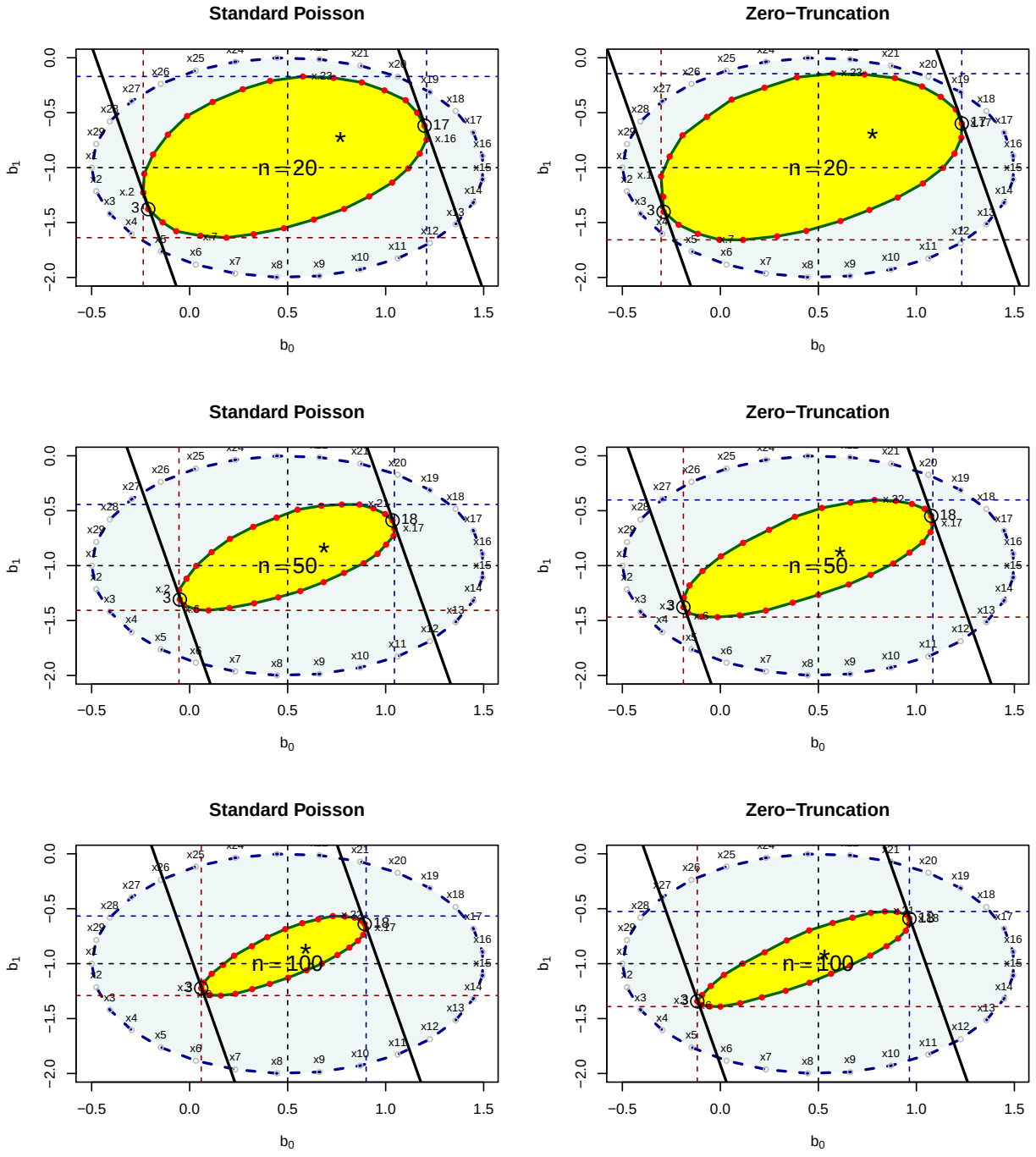


Figure 4.14: Translation and focusing behaviours of the imprecise estimate for the regression coefficients of β_0 and β_1 in the standard Poisson and zero-truncated Poisson regression models by changing the correlation coefficient $\rho = 0.0$, 0.4, and 0.8 at the fixed sample size $n = 50$ and standard deviation $\sigma^2 = 0.01$. The circle coloured blue (i.e., $\mathscr{R}_0 = \{(b_0, b_1)|b_0^2 + b_1^2 \leq 1\}$) is the imprecise normal prior for β_0 and β_1. The ellipsoid coloured yellow is the imprecise posterior expectation for β_0 and β_1. The star symbol is the maximum likelihood estimates for β_0 and β_1. The two solid black lines implies the imprecise posterior predicted response using the first row of the data matrix.

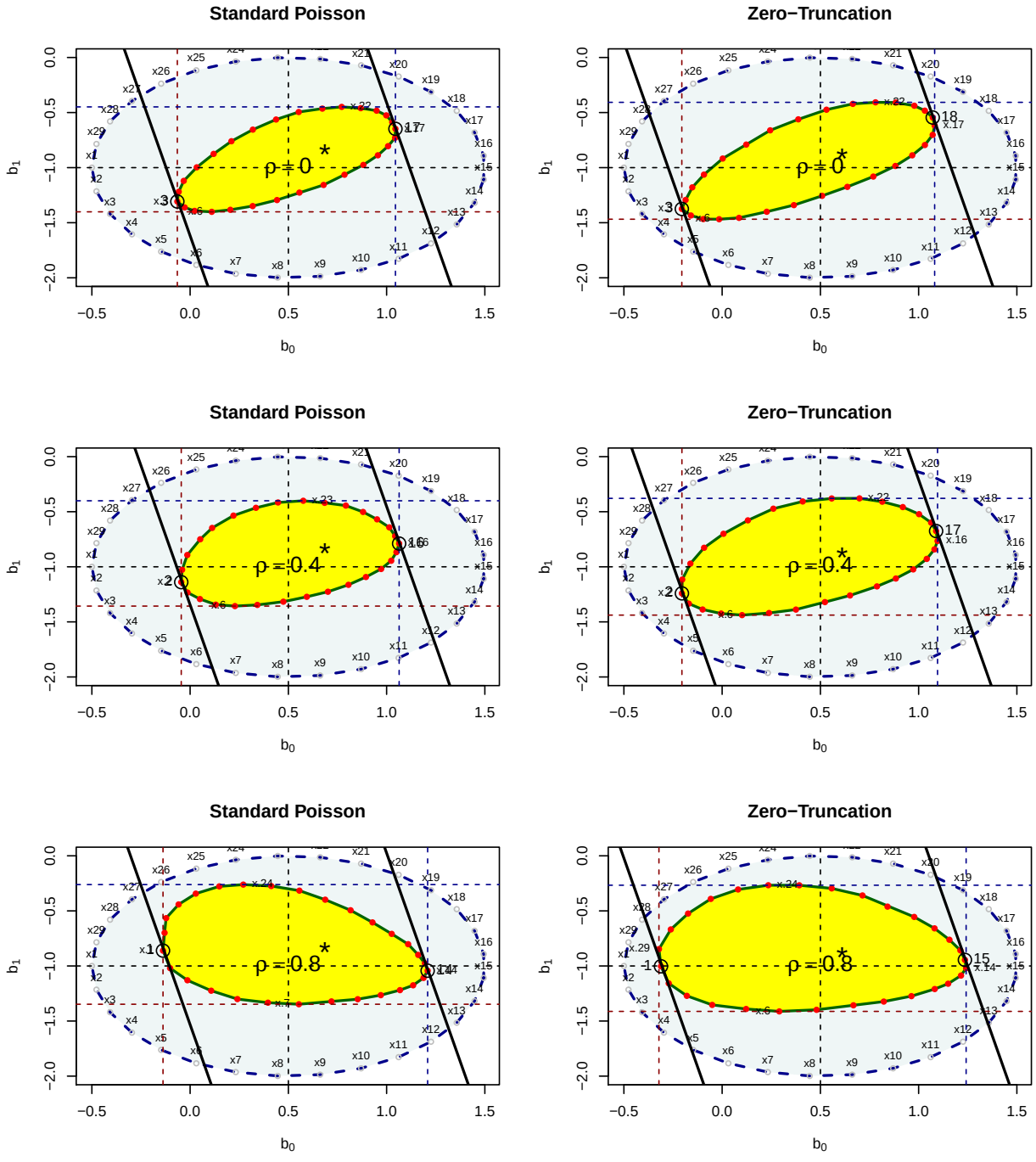
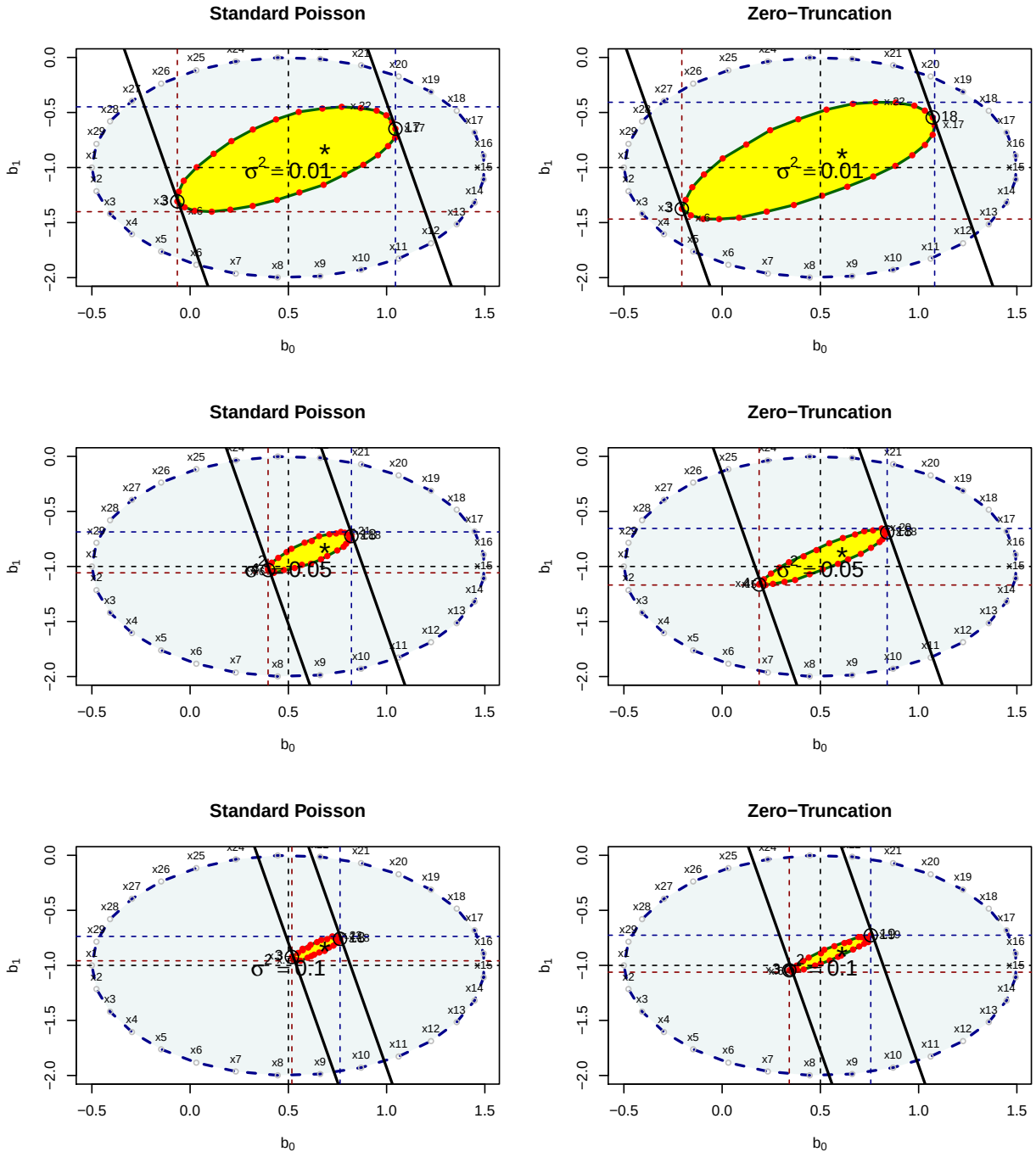


Figure 4.15: Translation and focusing behaviours of the imprecise estimate for the regression coefficients of β_0 and β_1 in the standard Poisson and zero-truncated Poisson regression models by changing the value of variance $\sigma^2 = 0.01$, 0.05, and 0.1 at the fixed sample size $n = 50$ and correlation coefficient $\rho = 0$. The circle coloured blue (i.e., $\mathscr{R}_0 = \{(b_0, b_1)|b_0^2 + b_1^2 \leq 1\}$) is the imprecise normal prior for β_0 and β_1. The ellipsoid coloured yellow is the imprecise posterior expectation for β_0 and β_1. The star symbol is the maximum likelihood estimates for β_0 and β_1. The two solid black lines implies the imprecise posterior predicted response using the first row of the data matrix.



it is certain that all points are enclosed by the circle coloured blue are translated to all points enclosed by the oval object coloured yellow, the function `plot()` in the `ipeglim` package by default prints only the extreme points on the boundary of this oval object as presented here. The function `set.grid()` can be used to validate that the internal points in the circle coloured blue translate to the points in the oval object coloured yellow. Thus, the comparison of two oval objects in the plot aids in the comprehension of the translation behaviour of the imprecise prior after observing samples under controlled conditions. The two black solid lines tangential to the oval object are the points that identify which of the extreme points produce either the maximum or the minimum of the posterior predicted response at a given value of x_1. In this simulation study, the value of the explanatory variable x from the first observation is used for x_1.

Other annotated symbols in the plot are provided here. The central point at which two dashed black lines intersect is the true regression parameter value of (β_0, β_1). The two vertical dashed lines represent the extreme posterior expectations of the regression parameter β_0, and the two horizontal dashed lines represent the minimum and maximum posterior expectations of the regression parameter β_1. The black star shaped symbol portrays the estimate obtained using the maximum likelihood method.

The plots are organized as below in each figure. On the left side, the plots produced from the imprecise fit of the standard Poisson regression model to the untruncated count data are listed. On the right side, the reader can find the plots produced from the imprecise fit of the zero-truncated Poisson regression model to the zero-truncated count data.

To elaborate, the plots in Figure 4.13 are produced by changing the sample size $n = 20, 50$ and 100 under the conditions of a correlation coefficient $\rho = 0$ and a variance $\sigma^2 = 0.01$. It is observed that both degrees of imprecision for the regression parameter β_0 and β_1 decrease as the size n of a sample increases. The plots in Figure 4.14 are produced by changing the correlation coefficient $\rho = 0, 0.4$, and 0.8 under the condition that the sample size $n = 50$ and the variance $\sigma^2 = 0.01$. Both imprecise estimates of the regression parameters β_0 and β_1 become increasingly extended as the correlation coefficient ρ increases. The plots in Figure 4.15 are produced by changing the variance $\sigma^2 = 0.01, 0.05$, and 0.1 under the condition that the sample size $n = 50$ and the correlation coefficient $\rho = 0$. Both imprecise estimates of the

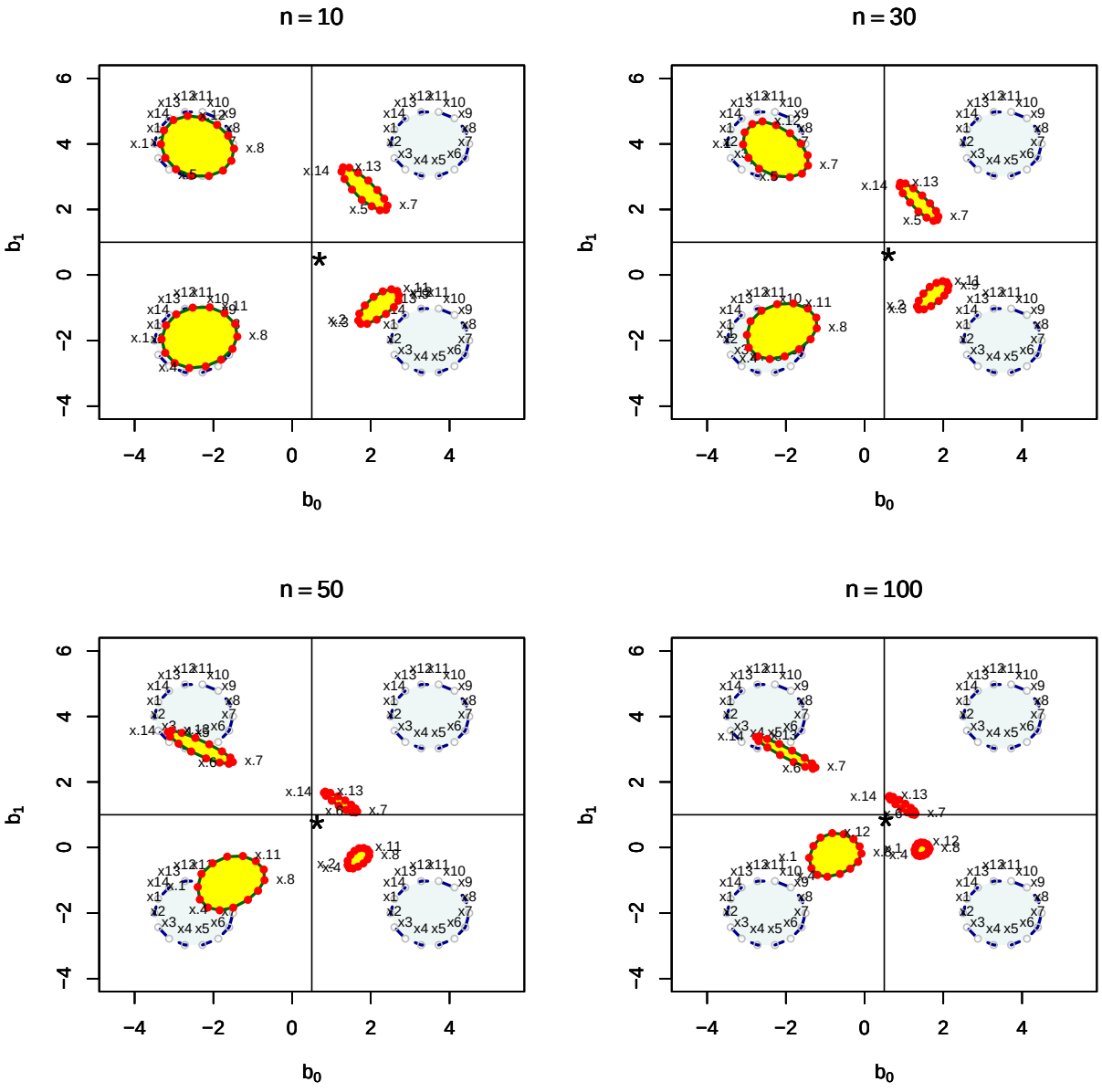
regression parameters β_0 and β_1 decrease as the values of the variances decrease. Based on an observation using three figures, we conclude that the focusing behaviour can be explained by the translation behaviour in the proposed methodology that demonstrates that the oval object becomes increasingly focused on a central point as the quantity of known data increases (i.e., the case of a large sample size, small variance, and independent structure).

Two assumptions underlie the case study presented above. The first assumption is that the true parameter values of the regression parameters (β_0, β_1) are captured by an imprecise prior. The second one is that *you* are the only intentional unit who is involved in this inference. These two assumptions are too idealistic since in practice, every intentional unit has different experiences and expertise.

Consider now the situation where there are four different intentional units and each of these units has their own imprecise prior that does not capture the true parameter values of the regression parameters. For this illustration, the true parameter values of the regression parameters (β_0, β_1) are assumed to be $(0.5, 1.0)$. Observations are generated using the mechanism described in the above case study. Four samples are then selected from the observation; 10, 30, 50, and 100. Four different imprecise priors are characterized as follows: $\mathscr{R}_{01} = \{(b_0, b_1) | (b_0 + 3)^2 + (b_1 + 3)^2 \leq 1\}$ in the first quadrant on the hyperparameter space $\boldsymbol{\Xi}$, $\mathscr{R}_{02} = \{(b_0, b_1) | (b_0 - 3)^2 + (b_1 - 3)^2 \leq 1\}$ in the second quadrant, $\mathscr{R}_{03} = \{(b_0, b_1) | (b_0 - 3)^2 + (b_1 - 3)^2 \leq 1\}$ in the third quadrant, and $\mathscr{R}_{04} = \{(b_0, b_1) | (b_0 + 3)^2 + (b_1 - 3)^2 \leq 1\}$ in the fourth quadrant. In Figure 4.16 the imprecise prior expectation regarding the regression parameters (β_0, β_1) are represented by blue circles and yellow circles represent the imprecise posterior expectations of (β_0, β_1) as described in the above case study.

A noteworthy feature of Figure 4.16 depicts how all of the different intentional units arrive at the central point, which constitutes the unknown truth, by exposure to the same data. Despite different learning speeds (i.e., how quickly his or her imprecise prior reaches the central point), all of the imprecise posterior expectations approach the point of convergence. It is expected by the asymptotic properties of Bayes estimator (Roberts, 2007, p. 48) that all of the imprecise posterior expectations will meet at that central point if there is a sufficiency of data and that the same data are observed.

Figure 4.16: The agreement process of four intentional units having different imprecise normal priors for the regression coefficients of β_0 and β_1 in the zero-truncated Poisson regression models on the same observation. Four imprecise normal priors for β_0 and β_1 are: 1) $\mathscr{R}_{01} = \{(b_0, b_1)|(b_0+3)^2 + (b_1+3)^2 \leq 1\}$ in the first quadrant, 2) $\mathscr{R}_{02} = \{(b_0, b_1)|(b_0-3)^2 + (b_1-3)^2 \leq 1\}$ in the second quadrant, 3) $\mathscr{R}_{03} = \{(b_0, b_1)|(b_0-3)^2 + (b_1-3)^2 \leq 1\}$ in the third quadrant, and 4) $\mathscr{R}_{04} = \{(b_0, b_1)|(b_0+3)^2 + (b_1-3)^2 \leq 1\}$ in the fourth quadrant. Four different zero-truncated Poisson sample sizes $n = 10, 30, 50,$ and 100 are selected from the observation. The circle coloured blue is the imprecise normal prior for β_0 and β_1. The ellipsoid coloured yellow is the imprecise posterior expectation for β_0 and β_1. The star symbol is the maximum likelihood estimates for β_0 and β_1.



However, it is also noted that two imprecise priors lying on the second and third quadrants are characterized by a slower approach to the central point and that their degrees of imprecision are not effectively decreasing. The reason for this fact is that the observed samples were originally generated with the true parameters of the regression coefficients $(0.5, 1.0)$. Hence, this fact needs to be interpreted such that two intentional units who assigned their imprecise prior on the second and third quadrants have a conflicting belief against the observation. This fact also supports the foundation of the proposed methodology such that an imprecise posterior expectation is a weighted mean of prior knowledge and observation. Therefore, this example demonstrates that the proposed methodology is a useful tool with which to explain prior-data conflict, the individual learning processes, and the agreement process between several intentional units. It is important to consider, however, that the methodology is highly dependent on the size of the observation.

The proposed methodology has been concerned with working on a two-dimensional hyperparameter space since most investigations involve two families of log-gamma and normal prior distributions. However, the number of hyperparameters is not limited to only two when using the proposed methodology in a regression analysis since the dimensionality of the hyperparameter space depends on the number of explanatory variables. The $\mathfrak{B}$-formulation theoretically ensures that the two families of prior and posterior distributions shown as (4.19) are conjugate, analogous to (4.6); however, the dimensionality of the hyperparameter space makes the proposed methodology impractical since the process required for performing the proposed methodology is computationally very expensive without dealing with the two problems of accurately evaluating the quantity of a normalizing constant and searching for the extreme points of a convex hull.

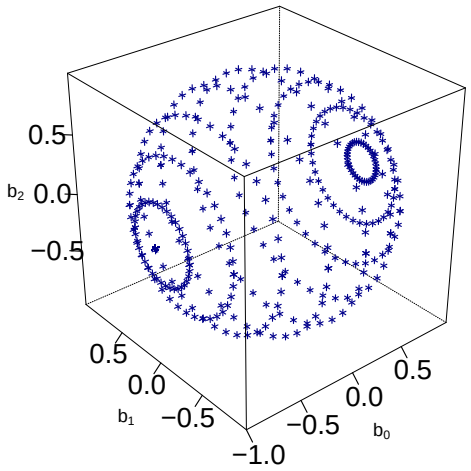
Before ending this chapter, the model having two explanatory variables is explained. For this example, the linear predictor in (4.21) is changed as below by incorporating an additional explanatory variable x_2:

$$\log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \tag{4.23}$$

where

$$E(\mathbf{x}_i) = \begin{bmatrix} x_1 = 2 \\ x_2 = 1 \end{bmatrix}, \quad V(\mathbf{x}_i) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.24}$$

Figure 4.17: A graphical representation of the imprecise normal prior (i.e., $\mathscr{C} = \{(b_0, b_1, b_2)|b_0^2 + b_1^2 + b_2^2 \leq 1\}$) that will be used for estimating the regression parameters β_0, β_1, and β_2 in the standard Poisson and zero-truncated Poisson regression models.



and

$$E(\boldsymbol{\beta}) = \begin{bmatrix} \beta_0 = 0 \\ \beta_1 = 0 \\ \beta_2 = 0 \end{bmatrix}, \quad V(\boldsymbol{\beta}) = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \tag{4.25}$$

For the specification of this variance-covariance matrix B_0, the identity matrix I is recommended in order to express a vague state of prior ignorance. The same suggestion can be found in Winkelmann (2008, p. 243).

The true parameter values of $(\beta_0, \beta_1, \beta_2)$ are assumed to be $(0.5, 0.5, -0.5)$, and the imprecise prior is characterized by specifying the convex hull $\mathscr{C}_0 = \{(b_0, b_1, b_2)|b_0^2 + b_1^2 + b_2^3 \leq 1\}$. Figure 4.17 shows this convex hull (i.e., sphere) for representing an imprecise prior in a three-dimensional hyperparameter space $\boldsymbol{\Xi}$. The imprecise posterior is produced with the option `method="MH"` (i.e, the Meteropolis-Hastings algorithm) when the function `update()` was used. The imprecise posterior expectation is graphically summarized in Figure 4.18. The scatter plot of the imprecise posterior expectation shows the ellipsoid on the top of that figure

and the scatterplot matrix shows the projection of this ellipsoid on each of three dimensions β_0, β_1, and β_2. The probability box of the imprecise posterior is shown in Figure 4.19. To generate these figures the functions `plot()` and `pbox()` in the `ipeglim` package are used.

Figure 4.18: The imprecise posterior expectation of the regression parameters β_0, β_1, β_2 after observing 100 standard Poisson (on the left panel) and zero-truncated Poisson (on the right panel) samples, respectively, using the imprecise normal prior show in Figure 4.17.

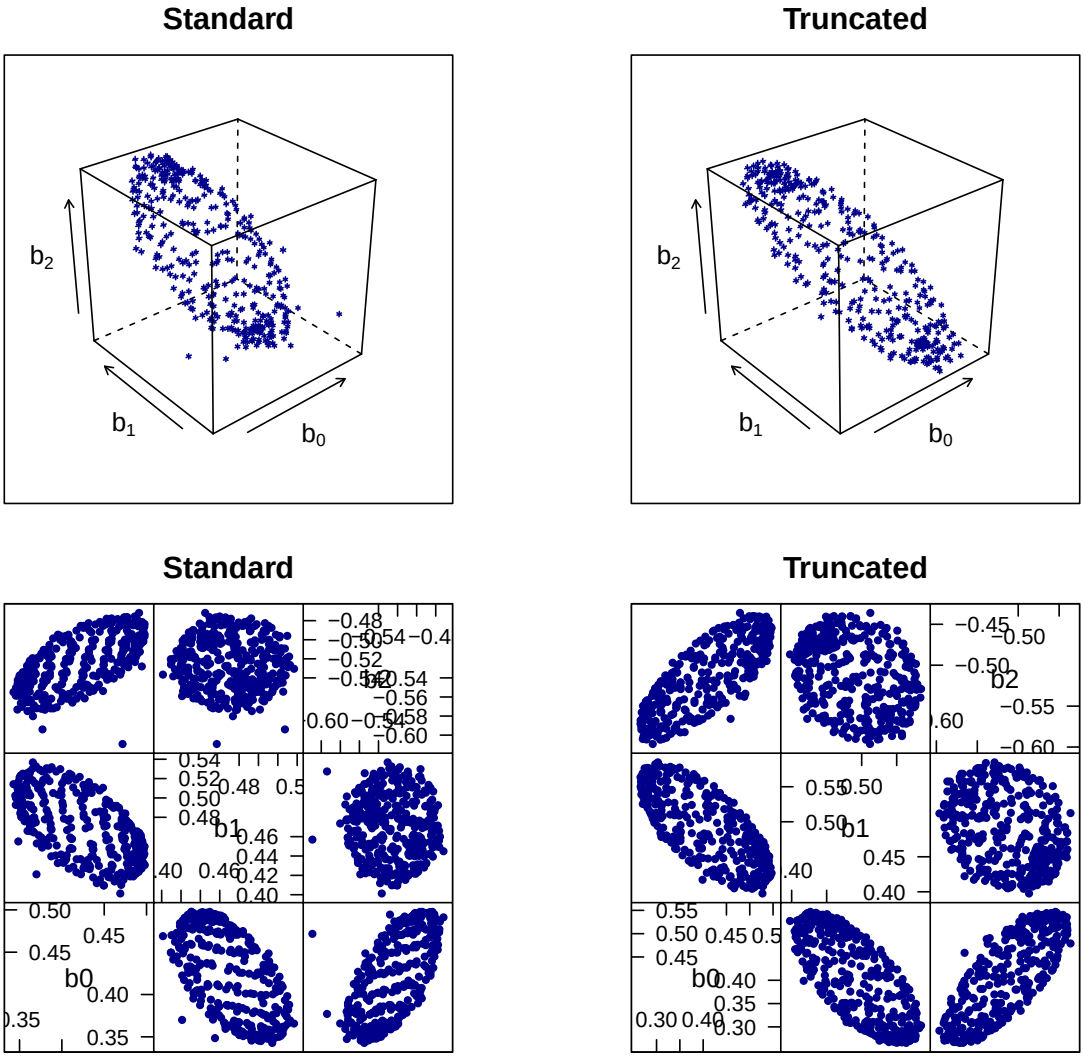
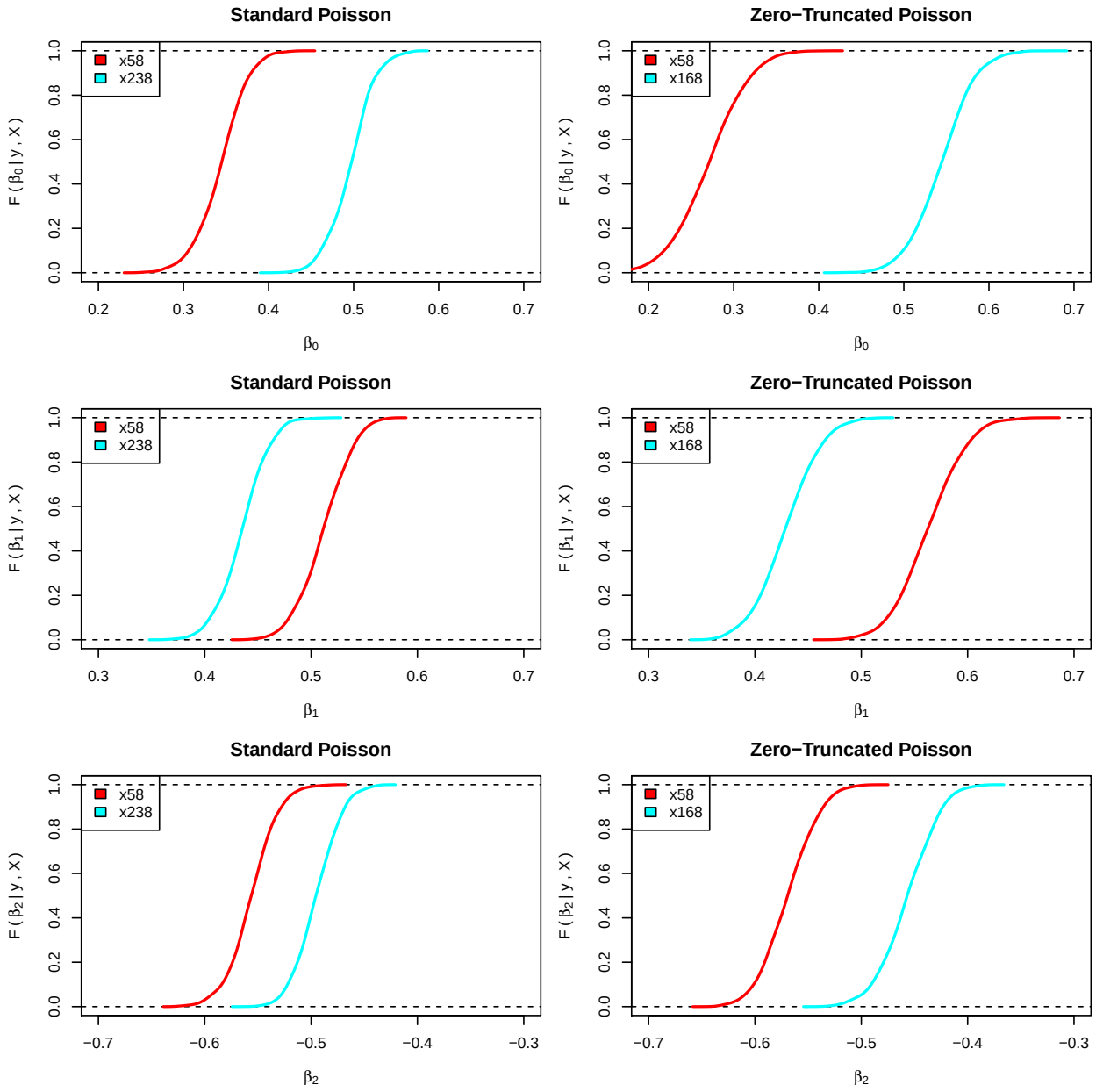


Figure 4.19: Probability boxes of the imprecise posterior for the regression parameters β_0, β_1, and β_2 after observing 100 standard (on the left panel) and zero-truncated Poisson (on the right panel) samples, respectively, using the imprecise normal prior show in Figure 4.17.



Chapter 5

Simulation Study

The proposed methodology demonstrated using several examples of a family of log-gamma prior distributions, and extended by the $\mathfrak{B}$-formulation to cases where 1. a family of normal prior distributions is assigned on a standard Poisson likelihood, 2. either families of log-gamma and normal prior distributions are assigned on a zero-truncated Poisson likelihood, and 3. a family of normal prior distributions is assigned on regression coefficients in the setup of a generalized linear model. Three primary behaviours of the proposed methodology are also clearly illustrated. Briefly, as a new sample is observed, 1. the convex hull representing an imprecise prior moves on the hyperparameter space (i.e., translation behaviour); 2. an imprecise posterior is stochastically ordered and squeezed by the upper and lower posterior distributions (i.e., focusing behaviour); and 3. the surface of an imprecise posterior expectation becomes flat (i.e., soft-linearity behaviour).

These behaviours may be blurred if a distribution of frequencies of the observed count is not a Poisson distribution which is the primary assumption of the proposed methodology. A typical instance of this blurring is a case when over-dispersion is present. The absence of explanatory variables that are influential to the Poisson mean parameter would be another instance in a regression analysis. One may question the effect of a correlation between two or more explanatory variables on the estimation of Poisson mean parameter. All these instances are attributed to the concerns regarding the use of the proposed methodology in practice. Hence, the aim of this chapter is to investigate the effect of such factors on the imprecise estimate of the parameter of interest through simulation studies under various conditions. Note that, for this investigation, an imprecise inference is made on the estimation of the parameter of a zero-truncated Poisson sampling rather than the parameter of a standard sampling model as it is an appropriate fit with the author's research interest.

5.1 Sampling model misspecification

A potential error that could be made when using the proposed methodology on count data would be a misspecification of a sampling model. For example, a standard Poisson sampling model is mistakenly chosen for describing zero-truncated count data, and vice versa. The focus in this section is the consequence that arises from this sampling model misspecification. The simulation study is designed as follows.

1. A sample of size N is taken from a standard Poisson sampling model with mean μ and zero values are removed from that sample. The size of the zero-truncated sample is denoted by n hereafter;

2. The imprecise log-gamma prior is described by characterizing the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 2, 0 \leq \beta \leq 2\}$ on the hyperparameters space $\boldsymbol{\Xi}$ of α and β;

3. An imprecise inference is carried out with the proposed methodology by assuming that the underlying sampling mechanism of this simulated zero-truncated sample is a standard Poisson sampling model (i.e., misspecified sampling model);

4. The above procedure in the step (3) is carried out again by assuming that the underlying sampling mechanism of the simulated zero-truncated sample is a zero-truncated Poisson sampling model (i.e., correctly specified sampling model). With this correct sampling model specification, the size N which is the sample size before truncating the zero values is estimated. From this point forward, this size N is understood as a population of size N in the context of the population size estimation problem;

5. The above four steps (1)–(4) are repeated 1,000 times for each pair of sample sizes $N = \{100, 300, 500\}$ and Poisson mean parameters $\mu = \{0.5, 0.75, 1, 1.25, 1.5\}$, i.e., corresponding to the canonical parameters $\tilde{\theta} = \log(\mu) = \{-0.693, -0.288, 0.000, 0.223, 0.405\}$; Note that the symbol $\tilde{\theta}$ is used for distinguishing the meta-parameter that controls the simulation condition from the canonical parameter θ to be inferred.

The size n of the zero-truncated sample, the minimum imprecise posterior expectation $\underline{E}_n(\theta|\mathbf{y})$, the maximum imprecise posterior expectation $\overline{E}_n(\theta|\mathbf{y})$, the degree of imprecision $\Delta_n(\theta|\mathbf{y})$, the minimum population size estimate $\underline{N}_n$, and the maximum population size estimate $\overline{N}_n$ are kept in every single performance of the imprecise inference. The results from one thousand repetitions, the mean size $\bar{n}$ of the zero-truncated sample, the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\theta|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$, the mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$, the mean minimum population size estimate $\underline{\underline{N}}_n$, and the mean maximum population size estimate $\overline{\overline{N}}_n$ are evaluated on each pair of two meta-parameters $\tilde{\theta}$ and N in this simulation. Four patterns are found from the results in Table 5.1.

1. The rate of the mean zero-truncation (i.e., $1 - \bar{n}/N$) decreases as the meta-parameter $\tilde{\theta}$ increases;

2. All imprecise estimates (i.e., $\overline{\overline{E}}_n(\theta|\mathbf{y})$, $\underline{\underline{E}}_n(\theta|\mathbf{y})$, $\overline{\Delta}_n(\theta|\mathbf{y})$) produced from the misspecified sampling model (i.e., the use of a standard Poisson sampling model) are greater than the meta-parameter $\tilde{\theta}$, which is considered as the true value of the canonical parameter θ in this simulation, over all different values of meta-parameter N. For the correctly specified sampling model (i.e., the use of a zero-truncated Poisson sampling model), the value of meta-parameter $\tilde{\theta}$ is captured within the range of the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\theta|\mathbf{y})$ and the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$ over all different values of the meta-parameter N;

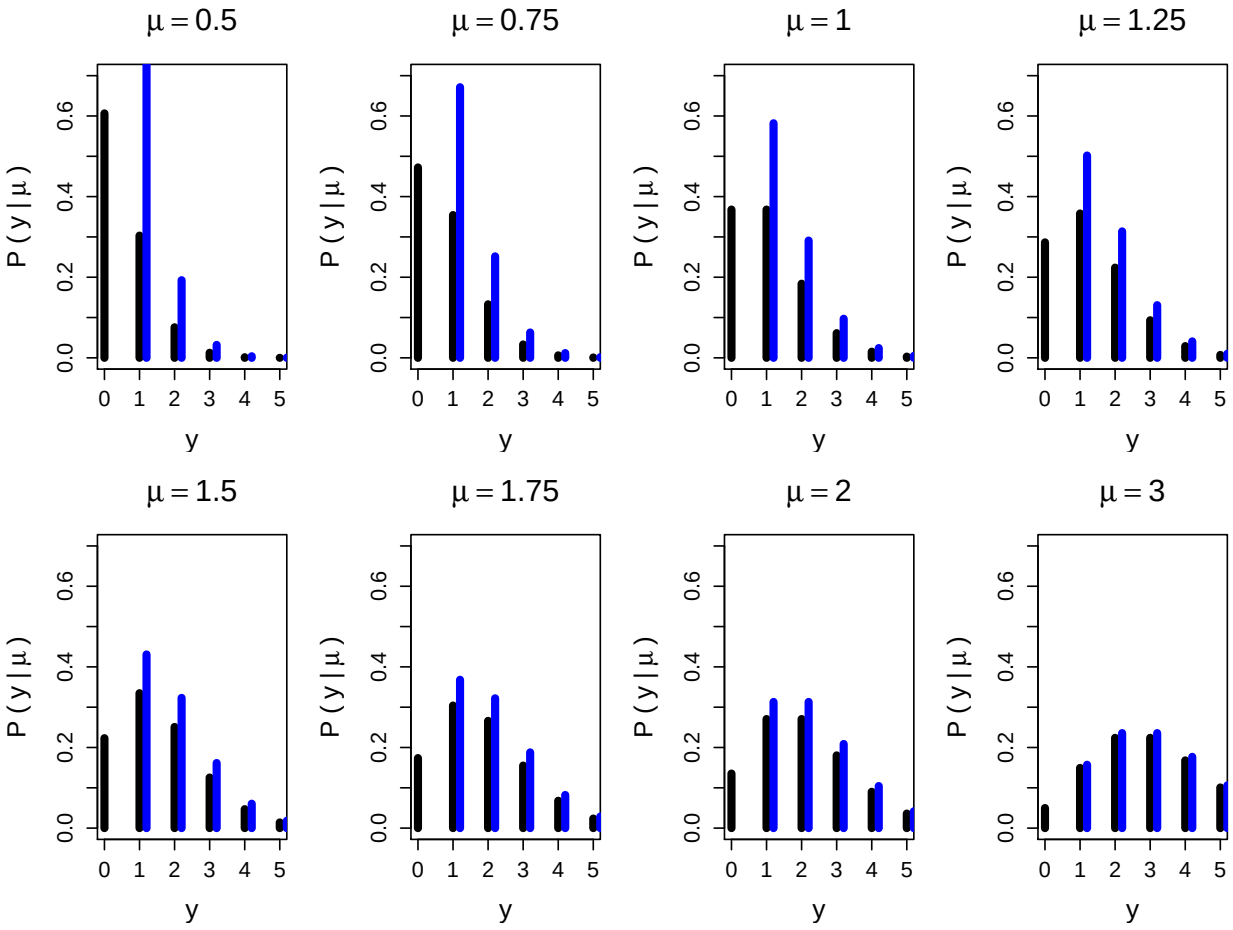
3. The mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$ decreases as the mean zero-truncated sample size $\bar{n}$ (or, the meta-parameter $\tilde{\theta}$) increases regardless whether or not a sampling model is correctly specified.

4. The interval between the mean minimum population size estimate $\underline{\underline{N}}_n$ and the mean maximum population size estimate $\overline{\overline{N}}_n$ captures the value of the meta-parameter $\tilde{\theta}$ over all different values of the meta-parameter N; however, it is shown that the length of this interval becomes wider as the meta-parameter $\tilde{\theta}$ decreases.

Table 5.1: The mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\theta|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$ over the different values of the meta-parameter $\tilde{\theta} = \{0.693, -0.288, 0, 0.233, 0.405\}$, i.e., the mean parameter $\mu = \{0.50, 0.75, 1.00, 1.25, 1.50\}$, with different sample sizes $N = \{100, 300, 500\}$ when a standard Poisson sampling model is specified for describing zero-truncated count data (i.e., sampling model misspecification). The zero-truncated Poisson sampling model is also used for describing zero-truncated count data (i.e., correct sampling model specification). $\bar{n}$ is the mean size of zero-truncated samples; $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

N		**Model Misspecification**			**Correct Model Spec.**			**Estimated Pop. Size**		
	$\tilde{\theta}$	$\overline{\overline{E}}_n(\theta\|\mathbf{y})$	$\underline{\underline{E}}_n(\theta\|\mathbf{y})$	$\overline{\Delta}_n(\theta\|\mathbf{y})$	$\overline{\overline{E}}_n(\theta\|\mathbf{y})$	$\underline{\underline{E}}_n(\theta\|\mathbf{y})$	$\overline{\Delta}_n(\theta\|\mathbf{y})$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
100										
	−0.693	0.139	0.322	0.184	−1.177	−0.518	0.660	39.782	91.650	163.960
	−0.288	0.278	0.409	0.131	−0.492	−0.189	0.303	52.653	94.777	117.214
	0.000	0.402	0.506	0.104	−0.125	0.073	0.198	63.079	96.439	108.577
	0.223	0.508	0.597	0.089	0.122	0.271	0.149	71.140	97.875	105.728
	0.405	0.613	0.692	0.078	0.330	0.449	0.119	77.876	98.796	104.115
300										
	−0.693	0.206	0.266	0.061	−0.818	−0.636	0.182	118.216	290.824	335.724
	−0.288	0.329	0.372	0.043	−0.347	−0.251	0.096	158.207	294.108	313.941
	0.000	0.439	0.474	0.034	−0.042	0.022	0.064	190.138	297.793	309.209
	0.223	0.544	0.573	0.029	0.191	0.240	0.049	213.669	297.483	304.879
	0.405	0.642	0.668	0.026	0.378	0.417	0.039	233.084	298.867	304.027
500										
	−0.693	0.219	0.255	0.036	−0.768	−0.661	0.107	197.093	491.429	534.231
	−0.288	0.337	0.363	0.026	−0.325	−0.268	0.057	264.403	496.058	515.573
	0.000	0.449	0.470	0.021	−0.020	0.018	0.038	315.813	495.037	506.184
	0.223	0.550	0.568	0.018	0.204	0.233	0.029	356.468	497.769	505.083
	0.405	0.649	0.665	0.016	0.391	0.414	0.023	388.481	498.512	503.589

Figure 5.1: Probability densities $P(y|\mu)$ of a standard Poisson distribution (coloured black) with a mean parameter of $\mu = \{0.50, 0.75, 1.00, 1.25, 1.25, 2.00, 3.00\}$ for count data y and probabilities densities of a zero-truncated Poisson distribution (coloured blue) with the same mean parameter μ.



From this simulation study, an overestimation of the imprecise inferential framework is expected when a standard Poisson sampling model is used for describing zero-truncated count data. Note that the result that is produced when an imprecise normal prior is utilized on the inference is not reported here since the patterns presented are similar. This consequence resulting from a sampling model misspecification is in fact attributed to the difference between the two probability densities of a standard Poisson and a zero-truncated Poisson distributions. The plots in Figure 5.1 are generated at some selected values of a standard Poisson mean parameter μ of $0.5, 0.75, 1, 1.25, 1.5., 2, 3$ for presenting this difference. The bars coloured black and blue are the densities of the standard Poisson and the zero-truncated Poisson

densities, respectively. From the plots in this figure, it is apparent that the shapes of the two Poisson probability densities are becoming less distinct from each other as the Poisson mean parameter value μ is becoming larger; hence, it is anticipated that no discernible difference can be identified using either of the two Poisson distributions when the mean parameter is large enough, for example a value greater than 3.

5.2 Over-dispersed count data

Overdispersion is a commonly discussed condition which means that the variability in the data is greater than the variability expected from a given sampling model. The question is what result will be shown on the imprecise estimate of the sampling model parameter when overdispersion is present. One may consider that this study is similar to the study in the previous section (i.e., sampling model misspecification). However, the zero-truncated count data in the previous section belongs to a family of Poisson sampling models whereas the count data being dealt with in this section does not belong to a family of Poisson sampling models.

A negative binomial distribution is often considered for cases of over-dispersed count data. In the context of mixture modelling, it is viewed as a mixed Poisson distribution of which the Poisson mean parameter is distributed as a gamma distribution with a shape parameter α and a scale parameter β. An alternative parametrization can arise by setting the Poisson mean $\mu = \alpha\beta$ and the dispersion parameter α^{-1} (i.e., a reciprocal of the shape parameter α of a gamma distribution). Note that a geometric distribution is a special case of this negative binomial distribution with the shape parameter $\alpha = 1$. This parametrization is useful to investigate the effect of overdispersion on the imprecise estimate of a model parameter using the fact that the negative binomial distribution converges to a Poisson distribution with a mean of μ as the shape parameter α goes to ∞ (or the dispersion parameter α^{-1} goes to 0). The simulation is designed as follows:

1. A sample of size $N = 300$ is taken from a standard negative binomial sampling model with the Poisson mean parameter μ and the shape parameter α, then the zero values are removed from that sample. The size of this zero-truncated sample is denoted by n from this point on;

2. The imprecise log-gamma prior is described by characterizing the region $\mathscr{R}_0 = \{(\alpha, \beta)|0 \leq \alpha \leq 5, 0 \leq \beta \leq 5\}$ on the hyperparameter space of α and β;

3. An imprecise inference is carried out with the proposed methodology using the imprecise log-gamma prior defined in the step (2);

4. Using the identical sample, an imprecise inference is carried out again with the imprecise normal prior described by the region $\mathscr{R}_0 = \{(\xi_2, \xi_1)|0 \leq \xi_2 \leq 2, -2 \leq \xi_1 \leq 2\}$ on the hyperparameter space of ξ_2 and ξ_1;

5. The above steps (1)-(4) are repeated 1,000 times for each pair of different Poisson mean parameter values $\mu = \{0.5, 0.75, 1.00, 1.25, 1.50\}$, which correspond to the canonical meta-parameters $\tilde{\theta} = \log(\mu) = \{-0.693, -0.288, 0.000, 0.223, 0.405\}$, and to the different shape parameter values $\alpha = \{1, 2, 5, 10, 200\}$.

Note that the shape parameter value $\alpha = 200$ is chosen for the reference distribution (i.e., Poisson distribution). The choice of this value is based on the fact that two probability densities of a negative binomial and a Poisson distribution are sufficiently similar but not identical to two decimal places. Hence, the sample simulated from a negative binomial distribution with shape parameter $\alpha = 200$ and Poisson mean parameter μ is considered to be a sample simulated from a standard Poisson distribution with the same mean parameter μ in this simulation study. As noted earlier section 5.1, the mean size $\bar{n}$ of a zero-truncated sample, the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\theta|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$, the mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$, the mean minimum population size estimate $\underline{\underline{N}}_n$, and the mean maximum population size estimate $\overline{\overline{N}}_n$ are computed using 1,000 repetitions. The results are summarized in Table 5.2 when using the imprecise log-gamma prior and Table 5.3 when using the imprecise normal prior.

Four interesting patterns are found from the results in the tables.

1. The rate of zero-truncation (i.e., $1 - \bar{n}/N$) increases as the meta-parameter shape α decreases (i.e., as data becomes increasingly over-dispersed);

2. The mean imprecise interval between the mean maximum imprecise posterior expectation ($\overline{\overline{E}}_n(\theta|\mathbf{y})$ and the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$) captures the true value of the meta-parameter $\tilde{\theta}$ when the meta-parameter shape $\alpha = 200$

Table 5.2: Mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\theta|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$ over the meta-parameter $\tilde{\theta} = \{-0.693, -0.288, 0.000, 0.233, 0.405\}$, i.e., $\mu = \{0.50, 0.75, 1.00, 1.25, 1.50\}$, with the different shape parameter values $\alpha = \{1, 2, 5, 10, 200\}$ for inferring the canonical parameter θ of a zero-truncated Poisson sampling model with the imprecise log-gamma prior ($\mathscr{R}_0 = \underline{\underline{\{}}(\alpha, \beta)|0 \leq \alpha \leq 5, 0 \leq \beta \leq 5\}$). $\bar{n}$ is the mean size of zero-truncated samples. $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

$\tilde{\theta}$			**Imprecise Estimate**			**Estimated Pop. Size**	
	α	$\bar{n}$	$\overline{\overline{E}}_n(\theta\|\mathbf{y})$	$\underline{\underline{E}}_n(\theta\|\mathbf{y})$	$\overline{\Delta}_n(\theta\|\mathbf{y})$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
-0.693							
	1	100.222	-0.337	0.008	0.345	158.969	198.707
	2	108.024	-0.583	-0.203	0.381	195.199	256.286
	5	113.628	-0.786	-0.368	0.418	229.342	315.509
	10	115.798	-0.876	-0.439	0.437	245.874	346.390
	200	118.155	-0.968	-0.510	0.458	264.274	382.074
-0.288							
	1	128.571	0.101	0.306	0.205	173.753	193.087
	2	141.309	-0.107	0.109	0.215	211.155	239.649
	5	150.703	-0.281	-0.053	0.229	247.086	286.218
	10	154.596	-0.352	-0.118	0.234	263.722	308.062
	200	157.895	-0.425	-0.184	0.242	280.803	331.206
0.000							
	1	149.639	0.381	0.529	0.148	183.703	195.331
	2	166.952	0.197	0.347	0.150	221.116	237.870
	5	179.701	0.037	0.192	0.156	256.497	279.223
	10	184.330	-0.017	0.140	0.157	270.297	295.402
	200	189.711	-0.087	0.074	0.161	288.490	317.079
0.223							
	1	167.300	0.588	0.705	0.117	193.135	200.834
	2	186.399	0.411	0.529	0.117	228.647	239.844
	5	201.722	0.268	0.387	0.119	262.227	277.171
	10	208.201	0.216	0.335	0.120	276.916	293.443
	200	214.322	0.156	0.277	0.121	293.079	311.661
0.405							
	1	179.678	0.747	0.846	0.099	199.310	204.783
	2	201.946	0.585	0.682	0.097	234.679	242.570
	5	218.845	0.456	0.553	0.098	265.804	276.173
	10	226.168	0.406	0.503	0.098	279.961	291.459
	200	232.820	0.356	0.454	0.098	293.926	306.638

Table 5.3: Mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\theta|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$ over the meta-parameter $\tilde{\theta} = \{-0.693, -0.288, 0.000, 0.233, 0.405\}$, i.e., $\mu = \{0.50, 0.75, 1.00, 1.25, 1.50\}$, with the different shape parameter values $\alpha = \{1, 2, 5, 10, 200\}$ for inferring the canonical parameter θ of a zero-truncated Poisson sampling model the imprecise normal prior ($\mathscr{R}_0 = \{(\xi_2, \xi_1)|0 \leq \xi_2 \leq 2, -2 \leq \xi_1 \leq 2\}$). $\bar{n}$ is the mean size of zero-truncated samples. $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

$\tilde{\theta}$			**Imprecise Estimate**			**Estimated Pop. Size**	
	α	$\bar{n}$	$\overline{\overline{E}}_n(\theta\|\mathbf{y})$	$\underline{\underline{E}}_n(\theta\|\mathbf{y})$	$\overline{\Delta}_n(\theta\|\mathbf{y})$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
-0.693							
	1	100.222	-0.199	-0.112	0.086	170.794	180.884
	2	108.024	-0.433	-0.312	0.121	209.737	229.024
	5	113.628	-0.622	-0.466	0.156	245.563	277.003
	10	115.798	-0.705	-0.531	0.174	262.551	301.329
	200	118.155	-0.790	-0.596	0.194	281.154	329.000
-0.288							
	1	128.571	0.180	0.223	0.043	180.983	185.061
	2	141.309	-0.021	0.026	0.046	221.224	227.403
	5	150.703	-0.190	-0.133	0.057	259.504	269.447
	10	154.596	-0.260	-0.196	0.064	277.039	289.144
	200	157.895	-0.330	-0.259	0.071	294.993	309.795
0.000							
	1	149.639	0.433	0.466	0.033	188.302	190.899
	2	166.952	0.253	0.285	0.032	227.608	231.220
	5	179.701	0.097	0.130	0.032	265.088	269.819
	10	184.330	0.045	0.078	0.033	279.683	284.894
	200	189.711	-0.023	0.011	0.034	298.918	305.021
0.223							
	1	167.300	0.628	0.654	0.026	196.279	198.022
	2	186.399	0.453	0.479	0.026	233.147	235.660
	5	201.722	0.311	0.337	0.026	268.136	271.418
	10	208.201	0.260	0.286	0.026	283.413	287.003
	200	214.322	0.201	0.227	0.026	300.328	304.296
0.405							
	1	179.678	0.781	0.803	0.022	201.592	202.820
	2	201.946	0.618	0.640	0.022	237.930	239.717
	5	218.845	0.490	0.512	0.022	270.025	272.363
	10	226.168	0.440	0.462	0.022	284.613	287.201
	200	232.820	0.391	0.413	0.022	299.045	301.897

(i.e., Poisson samples), but misses that true value of the meta-parameter $\tilde{\theta}$ for other values of meta-parameter shape α (i.e., overdispersion is present on Poisson samples).

3. Both the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\theta|\mathbf{y})$ and the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\theta|\mathbf{y})$ are increasing as the meta-parameter shape α is decreasing (i.e., data are increasingly over-dispersed).

4. The mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$ decreases as the meta-parameter shape α decreases when the meta-parameter θ is less than or equal to zero (i.e., $\mu \leq 1$). However, the magnitude of this decrement on the mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$ decreases as the meta-parameter $\tilde{\theta}$ is greater than zero (i.e., $\mu > 1$). This feature is more clearly shown when using an imprecise normal prior compared to when using an imprecise log-gamma prior. More interestingly, the mean degree of imprecision $\overline{\Delta}_n(\theta|\mathbf{y})$ does not change even though the change of the meta-parameter shape α increases when the meta-parameter $\tilde{\theta}$ is greater than or equal to zero.

5. The length of the interval between the mean minimum population size estimate $\underline{\underline{N}}_n$ and the mean maximum population size estimate $\overline{\overline{N}}_n$ becomes wider as the meta-parameter $\tilde{\theta}$ decreases, and this length becomes narrower as the meta-parameter α decreases at the given level of the meta-parameter $\tilde{\theta}$.

5.3 Structured Prior Variance-Covariance Matrix

When the proposed methodology is applied to the problem of inferring regression parameters in the context of a generalized linear model, a $p+1$ dimensional multivariate normal distribution $MVN(b_0, B_0)$ with mean vector b_0 and variance-covariance matrix B_0 is assumed to be a prior distribution $\pi(\boldsymbol{\beta})$ for regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$. In the current implementation of the proposed methodology, the identity matrix I_{p+1} (i.e., regression parameters $\boldsymbol{\beta}$ are independent structures) defaults to the prior variance-covariance matrix B_0. However, it is not necessarily limited to this independence structure since one may have a strong inclination to impose his or her own structure on this prior variance-covariance matrix B_0.

In the following two subsections, two simulation studies are designed as follows for investigating the change of the imprecise estimate of the regression parameters $\boldsymbol{\beta}$. Depending on the number of explanatory variables in the model, the studies are differentiated such that a single explanatory variable x_1 is involved with the model for the first study and two explanatory variables x_1 and x_2 are involved with the model for the second study. Because of the presence of an intercept term in the model, a prior distribution $\pi(\boldsymbol{\beta})$ for regression coefficients is assumed to be 2- and 3-dimensional multivariate normal distribution for the first and second studies, respectively. The first order auto-correlation structure is then imposed on the prior variance-covariance matrix B_0, and the value of a correlation coefficient ρ_π is controlled.

5.3.1 A single explanatory variable

Suppose that a linear predictor is defined as below:

$$\log \mu_i = \beta_0 + \beta_1 x_{1i} \tag{5.1}$$

where μ_i is the Poisson mean parameter for i-th individual and x_{1i} is a realization of i-th individual generated from a normal distribution with a mean of μ_X and a standard deviation of σ_X. The prior variance-covariance matrix B_0 for regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ is structured in the following way:

$$B_0 = V_\pi(\boldsymbol{\beta}) = D\Lambda_\pi D^T = \begin{bmatrix} \sigma_1^\pi & 0 \\ 0 & \sigma_2^\pi \end{bmatrix} \begin{bmatrix} 1 & \rho_\pi \\ \rho_\pi & 1 \end{bmatrix} \begin{bmatrix} \sigma_0^\pi & 0 \\ 0 & \sigma_1^\pi \end{bmatrix}, \tag{5.2}$$

where σ_1^π and σ_2^π are the standard deviations of β_0 and β_1, respectively, D is the diagonal matrix the elements of which are σ_1^π and σ_2^π, and ρ_π is a correlation coefficient between the regression parameters β_0 and β_1. The simulation is carried out as follows:

1. A sample of size $N = 300$ is taken from a normal distribution with a mean of μ_X and a standard deviation of σ_X;

2. The Poisson mean parameter μ_i for i-th individual is evaluated from the presumed linear predictor in (5.1) with predetermined values of the regression coefficients $\beta_0 = -0.5$ and $\beta_1 = 1.0$ for $i = 1, 2, \ldots, N$;

3. Individual level count data of the response variable y_i is simulated from a Poisson distribution with the mean value μ_i;

4. Both a response variable y_i and an explanatory variable x_i are removed if the simulated count of a response variable y_i has a zero value. The number of observations remaining in the sample after this zero-truncation is denoted by n;

5. The imprecise normal prior for regression parameters $\boldsymbol{\beta}$ is described by characterizing the region $\mathscr{R}_0 = \{(b_0, b_1)| -1.5 \leq b_0 \leq 0.5, 0 \leq b_1 \leq 2\}$ on the hyperparameter space of b_0 and b_1;

6. The proposed methodology is carried out for inferring regression parameters $\boldsymbol{\beta}$ using the zero-truncated sample of size n;

7. The above steps (1)–(6) are repeated 100 times to each pair of mean $\mu_X = \{-1, 0, 1\}$ and standard deviation $\sigma_X = \{0.5, 1.0, 2.0\}$ over different values of a correlation coefficient $\rho_\pi = \{0.0, 0.4, 0.8\}$.

Note that the standard deviations σ_1^π and σ_2^π associated with the prior variance-covariance matrix B_0 are assumed to be both 0.1. The values of mean μ_X and standard deviation σ_X are determined by the first two moments of a symmetric two-point design using Fisher's information matrix which measures the amount of information that the simulated samples for the explanatory variable x_1 carry about the regression parameter $\boldsymbol{\beta}$. The symmetric two-point design is performed as follows. η_1 and η_2 are considered as samples of the explanatory variable x_1. With the given two samples of η_1 and η_2, the minimum and maximum of the determinant of the Fisher's information is found. The derivation of this Fisher's information matrix and the optimal sampling design with two samples are noted in Appendix A.2. The minimum value of the determinant of Fisher's information matrix is found when $\eta_1 = \eta_2$ (i.e., there is no information about the regression parameters $\boldsymbol{\beta}$ is delivered when the values of two samples η_1 and η_2 are identical.) and the maximum value is not found since the upper

bound of this determinant does not exist. Hence, the different values are arbitrarily but symmetrically chosen such that $(\eta_1, \eta_2) = (-1.0, 1.0)$. Since the first two moments of these two samples are a mean of $\mu_X = (\eta_1 + \eta_2)/2 = 0$ and a variance of $sigma_X = (\eta_1 - \eta_2)/2 = 1$, samples for the explanatory variable x_1 are simulated from a normal distribution with a mean of 0 and a standard deviation of 1.

Table 5.4: Mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$ of the zero-truncated Poisson regression parameter β_i, where $i = 0$ and 1, over the different values of a correlation coefficient $\rho_\pi = \{0.0, 0.4, 0.8\}$ associated with the prior variance-covariance matrix B_0 and different values of the standard deviation $\sigma_X = \{0.5, 1.0, 2.0\}$ of the explanatory variable x_1. The predetermined regression coefficients are $(\beta_0, \beta_1) = (-0.5, 1.0)$. The imprecise normal prior is defined by characterizing the region $\mathscr{R}_0 = \{(b_0, b_1)| -1.0 \leq b_0 \leq 0.0, 0.5 \leq b_1 \leq 1.5\}$. $\bar{n}$ is the mean size of zero-truncated samples. $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

$N = 300$			$\beta_0 = -0.5$			$\beta_1 = 1.0$			**Estimated Pop. Size**		
ρ_π	σ_X	μ_X	$\overline{\overline{E}}(\beta_0\|\mathbf{y})$	$\underline{\underline{E}}(\beta_0\|\mathbf{y})$	$\overline{\Delta}(\beta_0\|\mathbf{y})$	$\overline{\overline{E}}(\beta_1\|\mathbf{y})$	$\underline{\underline{E}}(\beta_1\|\mathbf{y})$	$\overline{\Delta}(\beta_1\|\mathbf{y})$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
0.0	0.5	−1	−0.902	−0.151	0.751	0.526	1.464	0.939	140.33	236.918	429.636
		0	−0.902	−0.151	0.751	0.525	1.470	0.944	140.44	236.239	424.186
		1	−0.901	−0.151	0.750	0.531	1.476	0.945	140.62	237.093	428.770
	1.0	−1	−0.954	−0.103	0.850	0.730	1.275	0.545	145.92	226.569	451.128
		0	−0.939	−0.093	0.846	0.733	1.267	0.534	146.26	226.331	448.740
		1	−0.942	−0.097	0.846	0.731	1.267	0.536	146.07	226.235	448.648
	2.0	−1	−0.759	−0.260	0.499	0.940	1.061	0.121	150.14	256.659	373.256
		0	−0.772	−0.269	0.502	0.943	1.065	0.122	151.22	256.897	376.000
		1	−0.765	−0.257	0.508	0.940	1.066	0.125	150.21	256.153	390.433
0.4	0.5	−1	−0.909	−0.130	0.779	0.486	1.513	1.027	141.55	235.861	436.771
		0	−0.909	−0.127	0.782	0.486	1.510	1.024	139.85	232.594	432.390
		1	−0.920	−0.135	0.785	0.486	1.505	1.019	141.39	236.485	443.800

Table 5.4: *(continued)*

$N = 300$			$\beta_0 = -0.5$			$\beta_1 = 1.0$			**Estimated Pop. Size**		
ρ_π	σ_X	μ_X	$\overline{\overline{E}}(\beta_0\|\mathbf{y})$	$\underline{\underline{E}}(\beta_0\|\mathbf{y})$	$\overline{\Delta}(\beta_0\|\mathbf{y})$	$\overline{\overline{E}}(\beta_1\|\mathbf{y})$	$\underline{\underline{E}}(\beta_1\|\mathbf{y})$	$\overline{\Delta}(\beta_1\|\mathbf{y})$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
	1.0	-1	-1.020	-0.041	0.980	0.690	1.304	0.614	147.20	220.927	488.945
		0	-1.009	-0.033	0.976	0.688	1.305	0.617	147.54	220.739	486.458
		1	-1.014	-0.035	0.979	0.687	1.296	0.609	145.76	218.747	489.499
	2.0	-1	-0.861	-0.154	0.707	0.914	1.084	0.171	150.67	238.801	409.584
		0	-0.890	-0.151	0.739	0.910	1.095	0.184	150.42	234.531	407.090
		1	-0.861	-0.175	0.686	0.921	1.082	0.161	150.71	236.216	388.820
0.8	0.5	-1	-0.937	-0.114	0.823	0.426	1.536	1.110	139.37	229.830	437.299
		0	-0.930	-0.113	0.817	0.429	1.544	1.115	140.33	232.171	447.378
		1	-0.928	-0.106	0.822	0.436	1.546	1.111	140.12	230.101	439.527
	1.0	-1	-1.093	0.047	1.140	0.637	1.341	0.704	145.74	209.188	531.424
		0	-1.096	0.043	1.139	0.628	1.338	0.710	146.36	210.687	542.956
		1	-1.087	0.054	1.141	0.647	1.347	0.700	145.57	207.303	519.785
	2.0	-1	-1.088	0.038	1.125	0.869	1.134	0.264	150.76	214.110	522.689
		0	-1.095	0.046	1.141	0.864	1.139	0.275	151.51	214.267	516.106
		1	-1.087	0.041	1.129	0.868	1.135	0.267	150.34	213.487	499.560

The mean size $\bar{n}$ of the zero-truncated sample, the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$, for $i = 0$ and 1, the mean maximum population size estimate $\overline{\overline{N}}_n$, and the mean minimum population size estimate $\underline{\underline{N}}_n$ are computed from the results of 100 repetitions. The results are summarized in Table 5.4.

1. The mean imprecise interval estimate ($\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$) of regression parameters β_i, for $i = 0$ and 1, is insensitive to the change of the meta-parameter μ_X when the other meta-parameters σ_X and ρ_π are controlled;

2. The mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ for the regression parameter β_0 is larger at the meta-parameter $\sigma_X = 1$ than when compared to the one at the meta-parameter $\rho_\pi = 0$ (i.e., independent structure) and $\rho_\pi = 0.4$ (i.e., mild correlation), but this differentiation almost disappears at the meta-parameter $\rho_\pi = 0.8$ (i.e., strong correlation);

3. The mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ for the regression parameter β_1 decreases as the value of the meta-parameter σ_X increases at a given value of the meta-parameter ρ_π. In addition, the mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ seems to be inflated over the values of the meta-parameter ρ_π at a given value of the meta-parameter σ_X.

4. The interval between the mean minimum population size estimate $\underline{\underline{N}}_n$ and the mean maximum population size estimate $\overline{\overline{N}}_n$ captures the value of the meta-parameter N over all different values of the meta-parameters ρ_X, σ_X, and μ_X; however, the mean maximum population size estimate $\overline{\overline{N}}_n$ increases and the mean minimum population size estimate $\underline{\underline{N}}_n$ decreases as the meta-parameter ρ_π increases when the other meta-parameters σ_X and μ_X are controlled. In addition, this interval has the widest length at the meta-parameter $\sigma_X = 1$ when the other meta-parameters π_X and μ_X are controlled. It is also shown that the deviance of the mean minimum population size estimate $\underline{\underline{N}}_n$ from the meta-parameter N is shorter than the deviance of the mean maximum population size estimate $\overline{\overline{N}}_n$ from the meta-parameter N for all conditions of meta-parameters ρ_π, σ_X, and μ_X.

5.3.2 Two explanatory variables

Consider now the linear predictor defined as follow:

$$\log \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \tag{5.3}$$

where x_1 and x_2 are explanatory variables, and $i = 1, 2, \ldots, N$. In this simulation study, the regression coefficients $\boldsymbol{\beta}$ is assumed to be $(\beta_0, \beta_1, \beta_2)^T = (0.5, -0.5, 0.5)^T$. The imprecise prior for the regression parameters $\boldsymbol{\beta}$ is described by characterizing the convex hull $\mathscr{C}_0 = \{(b_0, b_1, b_2) | -0.5 \leq b_0 \leq 1.5, -1.5 \leq b_1 \leq 0.5, -0.5 \leq b_2 \leq 1.5\}$ on the three-dimensional hyperparameter space (b_0, b_1, b_2). The prior variance-covariance matrix B_0 is structured as below:

$$B_0 = V(\boldsymbol{\beta}) = D\Lambda_\pi D^T = \begin{bmatrix} \sigma_1^\pi & 0 & 0 \\ 0 & \sigma_2^\pi & 0 \\ 0 & 0 & \sigma_3^\pi \end{bmatrix} \begin{bmatrix} 1 & \rho_\pi & \rho_\pi^2 \\ \rho_\pi & 1 & \rho_\pi \\ \rho_\pi^2 & \rho_\pi & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^\pi & 0 & 0 \\ 0 & \sigma_2^\pi & 0 \\ 0 & 0 & \sigma_3^\pi \end{bmatrix}, \tag{5.4}$$

where σ_0^π, σ_1^π, and σ_2^π are the standard deviation of the regression parameters β_0, β_1, and β_2, respectively, and D is the diagonal matrix the elements of which are σ_0^π, σ_1^π, and σ_2^π. The values of a correlation coefficient ρ_π are varied in the range of $\{0, 0.2, 0.4, 0.6, 0.8\}$. Random variates for two explanatory variables x_1 and x_2 are simulated from each of a combinatorial pair of $\sigma_1^X = \{0.5, 1.0, 2.0\}$ and $\sigma_2^X = \{0.5, 1.0, 2.0\}$ at the centre $\mu_X = (0, 0)$. Note that no correlation ρ_X between two explanatory variables x_1 and x_2 is assumed in this study; however, this assumption is released and the effect of a correlation ρ_X between the explanatory variables are separately investigated later in section 5.4. Based on this design scheme, $N = 300$ Poisson samples are simulated at each repetition and the identical simulation process described in the above subsection 5.3.1 is carried out. The mean size $\bar{n}$ of a zero-truncated sample, the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$, for $i = 0, 1$ and 2, the mean maximum population size estimate $\overline{\overline{N}}_n$, and the mean minimum population size estimate $\underline{\underline{N}}_n$ are computed from the results of 100 repetitions.

The results are tabulated in Table 5.5, and the following patterns are found:

1. The mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ for the regression parameter β_1 decreases as either the meta-parameter σ_1^X or σ_2^X increases at the meta-parameter $\rho_\pi = 0$ (i.e., independent correlation between regression parameters) while this mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ increases as the meta-parameter ρ_π increases when the other conditions of the meta-parameters σ_1^X and σ_2^X are controlled. Analogous patterns are also found for the case of the regression parameter β_2.

2. The mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ for the regression parameter β_0 does not have a notable increment or decrement on the change of either the meta-parameter σ_1^X or σ_2^X at the meta-parameter $\rho_\pi = 0$; however, the mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ increases as either meta-parameter σ_1^X or σ_2^X increases at the given value of the meta-parameter $\rho_\pi = 0.4$ or $\rho_\pi = 0.8$. In addition, the mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ increases as the meta-parameter ρ_π increases when other conditions of the meta-parameters σ_1^X and σ_2^X are controlled.

3. The interval between the mean minimum population size estimate $\underline{\underline{N}}_n$ and the mean maximum population size estimate $\overline{\overline{N}}_n$ captures the value of meta-parameter N over all different conditions of meta-parameters ρ_π, σ_1^X, and σ_2^X. This phenomenon is due to the fact that the mean maximum population size estimate $\overline{\overline{N}}_n$ increases as the meta-parameter ρ_π increases while the mean minimum population size estimate $\underline{\underline{N}}_n$ does not have a notable change over the change of the meta-parameters ρ_π. This observation also accounts for the fact that the deviance of the mean minimum population size estimate $\underline{\underline{N}}_n$ from the meta-parameter N is shorter than the deviance of the mean maximum population size estimate $\overline{\overline{N}}_n$ from the meta-parameter N for all conditions of meta-parameters ρ_π, σ_X, and μ_X. That is, this skewness (to the right) is becoming overt as the meta-parameter ρ_π increases.

Table 5.5: The mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$ for the regression parameter β_i, where $i = 0, 1$, and 2, over different values of a correlation coefficient $\rho_\pi = \{0.0, 0.4, 0.8\}$ associated with the prior variance-covariance matrix B_0 and different values of standard deviation $\sigma_1^X = \sigma_2^X = \{0.5, 1.0, 2.0\}$ of the explanatory variables x_1 and x_2. The predetermined values of the regression parameters are $(\beta_0, \beta_1, \beta_2) = (0.5, -0.5, 0.5)$. The imprecise prior is defined by characterizing the convex hull $\mathscr{C}_0 = \{(b_0, b_1, b_2)|0.0 \leq b_0 \leq 1.0, -1.0 \leq b_1 \leq 0.0, 0.0 \leq b_2 \leq 1.0\}$. Note that $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, $\overline{\Delta}_n(\beta_i|\mathbf{y})$ are simply denoted by $\overline{\overline{E}}(\beta_i)$, $\underline{\underline{E}}(\beta_i)$, $\overline{\Delta}(\beta_i)$ for better presentation of results in the table. $\bar{n}$ is the mean size of the zero-truncated samples. $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

(σ_1^X, σ_2^X)	$\beta_0 = 0.5$			$\beta_1 = -0.5$			$\beta_2 = 0.5$			**Estimated Pop. Size**		
	$\overline{\overline{E}}(\beta_0)$	$\underline{\underline{E}}(\beta_0)$	$\overline{\Delta}(\beta_0)$	$\overline{\overline{E}}(\beta_1)$	$\underline{\underline{E}}(\beta_1)$	$\overline{\Delta}(\beta_1)$	$\overline{\overline{E}}(\beta_2)$	$\underline{\underline{E}}(\beta_2)$	$\overline{\Delta}(\beta_2)$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
$\rho_\pi = 0.0$												
(0.5,0.5)	0.269	0.679	0.410	−0.825	−0.194	0.631	0.189	0.821	0.632	238.35	277.179	347.838
(1.0,0.5)	0.243	0.702	0.459	−0.669	−0.347	0.322	0.229	0.814	0.585	233.19	271.862	355.216
(2.0,0.5)	0.271	0.698	0.427	−0.567	−0.436	0.131	0.289	0.727	0.438	221.58	271.618	352.131
$\rho_\pi = 0.0$												
(0.5,1.0)	0.248	0.705	0.456	−0.802	−0.218	0.584	0.343	0.669	0.326	233.86	272.432	355.387
(1.0,1.0)	0.224	0.709	0.486	−0.653	−0.360	0.293	0.367	0.655	0.288	229.67	269.752	361.876
(2.0,1.0)	0.271	0.689	0.419	−0.557	−0.446	0.111	0.407	0.597	0.190	219.53	272.363	350.974
$\rho_\pi = 0.0$												
(0.5,2.0)	0.278	0.696	0.418	−0.725	−0.291	0.434	0.440	0.567	0.127	221.19	271.016	347.773
(1.0,2.0)	0.277	0.692	0.415	−0.597	−0.412	0.185	0.448	0.557	0.109	219.07	271.825	349.380
(2.0,2.0)	0.317	0.666	0.349	−0.540	−0.462	0.078	0.466	0.537	0.072	211.93	273.350	335.988
$\rho_\pi = 0.4$												
(0.5,0.5)	0.261	0.673	0.412	−0.895	−0.072	0.823	0.140	0.870	0.731	238.51	277.799	351.606
(1.0,0.5)	0.246	0.684	0.437	−0.654	−0.353	0.301	0.135	0.874	0.738	234.09	274.449	359.149
(2.0,0.5)	0.277	0.682	0.405	−0.559	−0.449	0.110	0.199	0.797	0.598	221.50	274.306	348.610

Table 5.5: *(continued)*

(σ_1^X, σ_2^X)	$\beta_0 = 0.5$			$\beta_1 = -0.5$			$\beta_2 = 0.5$			**Estimated Pop. Size**		
	$\overline{\overline{E}}(\beta_0)$	$\underline{\underline{E}}(\beta_0)$	$\overline{\Delta}(\beta_0)$	$\overline{\overline{E}}(\beta_1)$	$\underline{\underline{E}}(\beta_1)$	$\overline{\Delta}(\beta_1)$	$\overline{\overline{E}}(\beta_2)$	$\underline{\underline{E}}(\beta_2)$	$\overline{\Delta}(\beta_2)$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
$\rho_\pi = 0.4$												
(0.5,1.0)	0.228	0.712	0.483	−0.906	−0.034	0.873	0.321	0.690	0.369	234.93	272.364	364.657
(1.0,1.0)	0.227	0.719	0.492	−0.667	−0.352	0.315	0.337	0.673	0.336	229.89	268.326	364.817
(2.0,1.0)	0.268	0.686	0.418	−0.554	−0.453	0.101	0.384	0.614	0.230	219.37	273.263	350.475
$\rho_\pi = 0.4$												
(0.5,2.0)	0.244	0.693	0.449	−0.836	−0.138	0.698	0.432	0.580	0.148	222.36	272.847	359.620
(1.0,2.0)	0.257	0.695	0.438	−0.622	−0.397	0.224	0.435	0.572	0.136	219.29	270.389	352.970
(2.0,2.0)	0.298	0.678	0.380	−0.531	−0.467	0.064	0.459	0.545	0.087	212.72	274.638	340.345
$\rho_\pi = 0.8$												
(0.5,0.5)	0.204	0.709	0.505	−1.182	0.181	1.364	−0.016	0.993	1.009	238.47	274.797	368.026
(1.0,0.5)	0.174	0.760	0.585	−0.733	−0.204	0.530	−0.059	1.038	1.097	233.86	271.594	378.830
(2.0,0.5)	0.028	0.873	0.845	−0.595	−0.426	0.170	−0.095	1.033	1.128	221.12	256.579	403.050
$\rho_\pi = 0.8$												
(0.5,1.0)	0.096	0.799	0.703	−1.251	0.260	1.511	0.230	0.783	0.553	234.29	263.186	409.652
(1.0,1.0)	0.091	0.801	0.710	−0.805	−0.090	0.715	0.213	0.756	0.543	230.12	258.563	419.960
(2.0,1.0)	0.038	0.868	0.830	−0.584	−0.421	0.164	0.270	0.723	0.453	219.62	257.037	404.447
$\rho_\pi = 0.8$												
(0.5,2.0)	0.081	0.838	0.757	−1.173	0.185	1.359	0.383	0.630	0.247	221.18	257.117	412.595
(1.0,2.0)	0.060	0.852	0.792	−0.760	−0.181	0.580	0.388	0.624	0.236	218.86	255.979	406.287
(2.0,2.0)	0.047	0.860	0.813	−0.564	−0.439	0.125	0.421	0.592	0.171	212.28	255.099	411.136

5.4 Correlated explanatory variables

The effect of the correlation ρ_X between two explanatory variables x_1 and x_2 on the imprecise estimate of the regression parameters $\boldsymbol{\beta}$ is considered under identical conditions as those controlled in the previous study (i.e., the effect of a correlation ρ_π associated with the prior variance-covariance matrix B_0 in the previous section 5.3).

For this study, the variance-covariance matrix of two explanatory variables x_1 and x_2 is structured as below:

$$V(\mathbf{x}) = D\Lambda_X D^T = \begin{bmatrix} \sigma_1^X & 0 \\ 0 & \sigma_2^X \end{bmatrix} \begin{bmatrix} 1 & \rho_X \\ \rho_X & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^X & 0 \\ 0 & \sigma_2^X \end{bmatrix}, \tag{5.5}$$

where σ_1^X and σ_2^X are the standard deviations of the explanatory variables x_1 and x_2, respectively, D is the diagonal matrix elements of which are σ_1^X and σ_2^X, and ρ_X is the correlation coefficient of x_1 and x_2. The `R` function `rmvnorm()` in the `mvtnorm` package (Genz and Bretz, 2009) is used for generating correlated random variates of two explanatory variables x_1 and x_2 for the given structure of a variance-covariance matrix in (5.5). When the proposed methodology is carried out for an imprecise inference, the default prior distribution $\pi(\boldsymbol{\beta})$ for the regression parameters $\boldsymbol{\beta}$ is specified (i.e., a three-dimensional multivariate normal distribution with a mean of $\mathbf{0}$ and a variance-covariance matrix I_3).

The mean size $\bar{n}$ of a zero-truncated sample, the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$, for $i = 0, 1$ and 2, the mean maximum population size estimate $\overline{\overline{N}}_n$, and the mean minimum population size estimate $\underline{\underline{N}}_n$ are computed using 100 repetitions. The results are tabulated in Table 5.6.

1. The mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ for the regression parameter β_1 decreases as either the meta-parameter σ_1^X or σ_2^X increases at a given value of the meta-parameter ρ_X while this mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ increases as the meta-parameter ρ_X increases when the other conditions of meta-parameters σ_1^X and σ_2^X are controlled. The analogous patterns are also found for the case of the regression parameter β_2.

2. The mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ for the regression parameter β_0 increases as either the meta-parameter σ_1^X or σ_2^X increases at the meta-parameter ρ_X while this mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ decreases as the meta-parameter ρ_X increases when the other conditions of the meta-parameters σ_1^X and σ_2^X are controlled.

3. The interval between the mean minimum population size estimate $\underline{\underline{N}}_n$ and the mean maximum population size estimate $\overline{\overline{N}}_n$ captures the value of the meta-parameter N over all different conditions of the meta-parameters ρ_X, σ_1^X, and σ_2^X. The notable change is not shown in the mean minimum population size estimate $\underline{\underline{N}}_n$, whereas the mean maximum population size estimate $\overline{\overline{N}}_n$ increases as the meta-parameter ρ_X increases when the other meta-parameters σ_1^X and σ_2^X are controlled.

Table 5.6: The mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$ for the regression parameter β_i, where $i = 0, 1$, and 2, over different values of a correlation coefficient $\rho_X = \{0.0, 0.4, 0.8\}$ between two explanatory variables x_1 and x_2 under the controlled conditions of the standard deviations $\sigma_1^X = \sigma_2^X = \{0.5, 1.0, 2.0\}$ of two explanatory variables x_1 and x_2. The predetermined values of the regression parameters are $(\beta_0, \beta_1, \beta_2) = (0.5, -0.5, 0.5)$. The imprecise prior is defined by characterizing the convex hull $\mathscr{C}_0 = \{(b_0, b_1, b_2)|0.0 \leq b_0 \leq 1.0, -1.0 \leq b_1 \leq 0.0, 0.0 \leq b_2 \leq 1.0\}$. Note that $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, $\overline{\Delta}_n(\beta_i|\mathbf{y})$ are simply denoted by $\overline{\overline{E}}(\beta_i)$, $\underline{\underline{E}}(\beta_i)$, $\overline{\Delta}(\beta_i)$ for better presentation of results in the table. $\bar{n}$ is the mean size of the zero-truncated samples. $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

(ρ_X, σ_1^X)	$\beta_0 = 0.5$			$\beta_1 = -0.5$			$\beta_2 = 0.5$			**Estimated Pop. Size**		
	$\overline{\overline{E}}(\beta_0)$	$\underline{\underline{E}}(\beta_0)$	$\overline{\Delta}(\beta_0)$	$\overline{\overline{E}}(\beta_1)$	$\underline{\underline{E}}(\beta_1)$	$\overline{\Delta}(\beta_1)$	$\overline{\overline{E}}(\beta_2)$	$\underline{\underline{E}}(\beta_2)$	$\overline{\Delta}(\beta_2)$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
$\sigma_2^X = 0.5$												
(0.0,0.5)	0.267	0.678	0.410	−0.827	−0.192	0.635	0.191	0.821	0.630	238.39	277.297	348.013
(0.4,0.5)	0.282	0.652	0.371	−0.878	−0.147	0.731	0.152	0.883	0.730	240.18	281.692	343.446
(0.8,0.5)	0.323	0.628	0.304	−0.955	−0.060	0.895	0.067	0.962	0.895	240.77	284.918	329.102
$\sigma_2^X = 0.5$												
(0.0,1.0)	0.234	0.695	0.461	−0.676	−0.345	0.332	0.223	0.808	0.585	234.23	273.588	359.690
(0.4,1.0)	0.257	0.684	0.427	−0.710	−0.300	0.410	0.168	0.853	0.685	237.22	277.121	354.649
(0.8,1.0)	0.286	0.651	0.365	−0.818	−0.220	0.599	0.027	0.968	0.941	239.44	281.991	342.341
$\sigma_2^X = 0.5$												
(0.0,2.0)	0.272	0.685	0.413	−0.566	−0.443	0.123	0.293	0.721	0.428	221.88	274.865	352.734
(0.4,2.0)	0.276	0.700	0.424	−0.587	−0.415	0.172	0.244	0.775	0.531	224.39	272.618	350.919
(0.8,2.0)	0.286	0.688	0.402	−0.642	−0.362	0.281	0.108	0.917	0.809	227.36	273.709	345.245
$\sigma_2^X = 1.0$												
(0.0,0.5)	0.238	0.698	0.460	−0.815	−0.227	0.587	0.349	0.676	0.327	234.29	273.728	360.097
(0.4,0.5)	0.258	0.685	0.427	−0.852	−0.157	0.695	0.292	0.708	0.415	236.12	275.412	351.092
(0.8,0.5)	0.297	0.658	0.361	−0.979	−0.036	0.943	0.209	0.804	0.595	238.22	279.762	338.143

Table 5.6: *(continued)*

(ρ_X, σ_1^X)	$\beta_0 = 0.5$			$\beta_1 = -0.5$			$\beta_2 = 0.5$			**Estimated Pop. Size**		
	$\overline{\overline{E}}(\beta_0)$	$\underline{\underline{E}}(\beta_0)$	$\overline{\Delta}(\beta_0)$	$\overline{\overline{E}}(\beta_1)$	$\underline{\underline{E}}(\beta_1)$	$\overline{\Delta}(\beta_1)$	$\overline{\overline{E}}(\beta_2)$	$\underline{\underline{E}}(\beta_2)$	$\overline{\Delta}(\beta_2)$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
$\sigma_2^X = 1.0$												
(0.0,1.0)	0.241	0.720	0.478	−0.647	−0.358	0.290	0.363	0.649	0.286	231.26	271.090	360.678
(0.4,1.0)	0.233	0.706	0.473	−0.701	−0.314	0.387	0.313	0.706	0.394	233.75	271.418	357.439
(0.8,1.0)	0.261	0.661	0.400	−0.843	−0.179	0.663	0.186	0.852	0.666	238.36	278.839	347.092
$\sigma_2^X = 1.0$												
(0.0,2.0)	0.279	0.700	0.421	−0.557	−0.444	0.112	0.407	0.602	0.194	219.96	271.706	349.983
(0.4,2.0)	0.251	0.707	0.456	−0.589	−0.417	0.172	0.366	0.645	0.279	224.11	270.387	357.291
(0.8,2.0)	0.253	0.693	0.441	−0.673	−0.343	0.330	0.230	0.780	0.550	231.95	275.043	356.986
$\sigma_2^X = 2$												
(0.0,0.5)	0.270	0.692	0.422	−0.730	−0.296	0.434	0.437	0.564	0.127	220.84	271.390	349.660
(0.4,0.5)	0.264	0.686	0.422	−0.775	−0.247	0.529	0.420	0.589	0.169	224.57	273.914	352.312
(0.8,0.5)	0.273	0.682	0.408	−0.918	−0.101	0.818	0.363	0.647	0.284	226.59	273.688	347.521
$\sigma_2^X = 2$												
(0.0,1.0)	0.273	0.687	0.414	−0.606	−0.418	0.188	0.446	0.555	0.109	219.62	271.931	347.368
(0.4,1.0)	0.254	0.705	0.451	−0.648	−0.373	0.275	0.420	0.585	0.166	224.29	271.443	356.970
(0.8,1.0)	0.243	0.690	0.447	−0.791	−0.229	0.563	0.338	0.677	0.339	231.10	273.607	357.230
$\sigma_2^X = 2$												
(0.0,2.0)	0.312	0.668	0.356	−0.541	−0.466	0.075	0.463	0.538	0.075	212.76	276.472	343.414
(0.4,2.0)	0.282	0.703	0.421	−0.563	−0.438	0.125	0.437	0.562	0.126	219.54	270.131	347.541
(0.8,2.0)	0.233	0.715	0.482	−0.675	−0.345	0.330	0.346	0.676	0.330	232.55	270.957	361.115

Two Binary Explanatory Variables

In the above simulation study the scale of two explanatory variables is limited to a continuous measurement although various types of scale are available for explanatory variables. One of the commonly used scales is a dichotomous outcome that has only two categories such as dead or alive. One additional simulation study is proposed at this time regarding the effect of correlation for the case where two explanatory variables x_1 and x_2 has both binary measurement scales under the identical conditions as those controlled for the previous study when two explanatory variables are both assumed to be continuous.

For this simulation study, the `R` function `rmvbin()` in the `bindata` package (Leisch et al., 2012) is used for generating correlated binary random variates. The marginal probability p_1 for the first binary explanatory variable x_1 is assumed to have the values of $\{0.3, 0.5, 0.7\}$, and the marginal probability p_2 for the second binary explanatory variable x_2 is fixed at 0.5. The values of the correlation coefficient ρ_X are varied in the range of $\{0.0, 0.25, 0.5\}$. The results are enumerated in Table 5.6.

1. The mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ for the regression parameter β_1 increases as the meta-parameter ρ_X increases at the given value of the meta-parameter p_1; however, this mean degree of imprecision $\overline{\Delta}_n(\beta_1|\mathbf{y})$ increases as the meta-parameter p_1 increases when the meta-parameter ρ_X is controlled. An analogous pattern is also found for the case of the regression parameter β_2.

2. The mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ for the regression parameter β_0 decreases as the meta-parameter ρ_X at the given value of the meta-parameter p_1; however, this mean degree of imprecision $\overline{\Delta}_n(\beta_0|\mathbf{y})$ increases as the meta-parameter p_1 increases when the meta-parameter ρ_X is controlled.

3. No notable change is shown in the mean minimum population size estimate $\underline{\underline{N}}_n$; however, the mean maximum population size estimate $\overline{\overline{N}}_n$ increases as the meta-parameter ρ_X increases when the meta-parameter p_1 is controlled.

Table 5.7: The mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$ for the regression parameter β_i, where $i = 0, 1$, and 2, over different values of a correlation coefficient $\rho_X = \{0.0, 0.25, 0.50\}$ between the two binary explanatory variables x_1 and x_2 under the controlled conditions of $p_1 = \{0.3, 0.5, 0.7\}$ and $p_2 = 0.5$. The predetermined values of the regression parameters are $(\beta_0, \beta_1, \beta_2) = (0.5, -0.5, 0.5)$. The imprecise prior is defined by characterizing the convex hull $\mathscr{C}_0 = \{(b_0, b_1, b_2)|0.0 \leq b_0 \leq 1.0, -1.0 \leq b_1 \leq 0.0, 0.0 \leq b_2 \leq 1.0\}$. Note that $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, $\overline{\Delta}_n(\beta_i|\mathbf{y})$ are simply denoted by $\overline{\overline{E}}(\beta_i)$, $\underline{\underline{E}}(\beta_i)$, $\overline{\Delta}(\beta_i)$ for better presentation of results in the table. $\bar{n}$ is the mean size of the zero-truncated samples. $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

p_1		$\beta_0 = 0.5$			$\beta_1 = -0.5$			$\beta_2 = 0.5$			**Estimated Pop. Size**		
	ρ_X	$\overline{\overline{E}}(\beta_0)$	$\underline{\underline{E}}(\beta_0)$	$\overline{\Delta}(\beta_0)$	$\overline{\overline{E}}(\beta_1)$	$\underline{\underline{E}}(\beta_1)$	$\overline{\Delta}(\beta_1)$	$\overline{\overline{E}}(\beta_2)$	$\underline{\underline{E}}(\beta_2)$	$\overline{\Delta}(\beta_2)$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
0.3													
	0.00	0.142	0.793	0.651	−0.880	−0.133	0.747	0.158	0.880	0.722	247.98	283.051	343.087
	0.25	0.163	0.783	0.620	−0.881	−0.123	0.758	0.131	0.885	0.754	248.23	284.295	335.430
	0.50	0.174	0.766	0.592	−0.889	−0.103	0.786	0.108	0.906	0.798	249.82	287.148	329.916
0.5													
	0.00	0.098	0.835	0.738	−0.854	−0.135	0.719	0.132	0.905	0.773	238.32	278.982	349.350
	0.25	0.115	0.812	0.696	−0.876	−0.114	0.762	0.106	0.919	0.813	239.23	281.677	344.240
	0.50	0.132	0.798	0.665	−0.903	−0.071	0.832	0.067	0.942	0.875	239.62	283.288	335.831
0.7													
	0.00	0.046	0.891	0.845	−0.866	−0.092	0.775	0.099	0.931	0.832	229.63	277.287	354.584
	0.25	0.065	0.869	0.805	−0.909	−0.055	0.854	0.073	0.953	0.880	230.75	279.748	350.355
	0.50	0.075	0.846	0.770	−0.973	−0.001	0.972	0.034	0.993	0.960	231.88	282.196	344.095

5.5 Absence of explanatory variable

One strong assumption in a regression analysis is that all explanatory variables defined in the linear predictor are influential to a response variable; however, it is unlikely to have all such explanatory variables in practice. The question is then what impact is shown on the imprecise estimate when the intentional unit is not aware that some influential explanatory variables needed for a study are absent.

To investigate this question, the simulation design that was used for investigating the effect of a correlation ρ_X between two explanatory variables x_1 and x_2 on the imprecise estimate of regression parameters is slightly modified as follows. The linear predictor $\log \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, where $(\beta_0, \beta_1, \beta_2)^T = (0.5, -0.5, 0.5)$, is used for simulating Poisson random variates as the true sampling mechanism; however, the model to be fitted to the data is defined as $\log \mu_i = \beta_0 + \beta_1 x_{1i}$ has a only single explanatory variable x_1. Since there are only two regression parameters β_0 and β_1 in the model, the imprecise prior for the regression parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ is represented by characterizing the region $\mathscr{R}_0 = \{(b_0, b_1) | 0 \leq b_0 \leq 1, -1 \leq b_1 \leq 0\}$ on the hyperparameter space of b_0 and b_1. A correlation coefficient ρ_π associated with the prior variance-covariance matrix B_0 in (5.2) and a correlation ρ_X between two explanatory variables x_1 and x_2 in (5.5) are continuously employed. This simulation is carried out with different sample size n which is varied in the range of $\{100, 300, 500\}$. The mean size $\bar{n}$ of the zero-truncated sample, the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$, for $i = 0$ and 1, the mean maximum population size estimate $\overline{\overline{N}}_n$, and the mean minimum population size estimate $\underline{\underline{N}}_n$ are computed using 100 repetitions. The results are summarized in Table 5.8.

1. Both the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_0|\mathbf{y})$ and the minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_0|\mathbf{y})$ for the regression parameter β_0 decrease as the meta-parameter ρ_X increases; however, the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_1|\mathbf{y})$ and the minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_1|\mathbf{y})$ for the regression parameter β_1 increases as the meta-parameter ρ_X increases;

2. The mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$ has a smaller value when the meta-parameter $\rho_\pi = 0.4$ (i.e., mild correlation) compared to when the meta-parameter $\rho_\pi = 0.0$ (i.e., independence) or $\rho_\pi = 0.8$ (i.e., strong correlation);

3. The interval between the mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$ and the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$ is likely to capture the regression parameters $\boldsymbol{\beta} = (\beta_0 = 0.5, \beta_1 = -0.5)^T$ when the meta-parameter $N = 100$ or $\rho_\pi = 0.8$.

4. The interval between the mean minimum population size estimate $\underline{\underline{N}}_n$ and the mean maximum population size estimate $\overline{\overline{N}}_n$ does not capture the given value of the meta-parameter N when the meta-parameters $\sigma_X = 0.0$ (or 0.4) and $N = 500$ for all conditions of the meta-parameter ρ_π. For all conditions of the meta-parameter N, the mean minimum and maximum population size estimates $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ increase as the meta-parameter σ_X increases. However, the mean maximum population size estimate $\overline{\overline{N}}_n$ decreases as the meta-parameter ρ_π increases when the meta-parameter $N = 100$ when the other meta-parameter σ_X is controlled.

Table 5.8: The mean maximum imprecise posterior expectation $\overline{\overline{E}}_n(\beta_i|\mathbf{y})$, the mean minimum imprecise posterior expectation $\underline{\underline{E}}_n(\beta_i|\mathbf{y})$, and the mean degree of imprecision $\overline{\Delta}_n(\beta_i|\mathbf{y})$ for the regression parameter β_i, where $i = 0, 1$, over different values of a correlation coefficient $\rho_\pi = \{0.0, 0.4, 0.8\}$ associated with the prior variance-covariance matrix B_0 and different values of standard deviations $\sigma_1^X = \sigma_2^X = \sigma_X = \{0.0, 0.4, 0.8\}$ of two explanatory variables x_1 and x_2 on different sample sizes $N = \{100, 300, 500\}$. The linear predictor $\log(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ is used for generating Poisson random variates. The predetermined values of the regression parameters are $(\beta_0, \beta_1, \beta_2) = (0.5, -0.5, 0.5)$. The linear predictor $\log(\mu_i) = \beta_0 + \beta_1 x_1$ is used for inferring the regression parameters using the proposed methodology. The imprecise prior is defined by characterizing the region $\mathscr{R}_0 = \{(b_0, b_1)|0.0 \leq b_0 \leq 1.0, -1.0 \leq b_1 \leq 0.0\}$. $\bar{n}$ is the mean size of the zero-truncated samples. $\underline{\underline{N}}_n$ and $\overline{\overline{N}}_n$ are the mean minimum population size estimate and the mean maximum population size estimate, respectively.

			$\beta_0 = 0.5$			$\beta_1 = -0.5$			**Estimated Pop. Size**		
N	σ_X	ρ_π	$\overline{\overline{E}}(\beta_0\|\mathbf{y})$	$\underline{\underline{E}}(\beta_0\|\mathbf{y})$	$\overline{\Delta}(\beta_0\|\mathbf{y})$	$\overline{\overline{E}}(\beta_1\|\mathbf{y})$	$\underline{\underline{E}}(\beta_1\|\mathbf{y})$	$\overline{\Delta}(\beta_1\|\mathbf{y})$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
100	0.0	0.0	0.264	0.903	0.639	−0.809	−0.261	0.548	76.91	84.681	119.815
		0.4	0.308	0.856	0.548	−0.748	−0.265	0.483	76.55	85.287	114.527
		0.8	0.175	0.977	0.802	−0.703	−0.129	0.573	76.00	85.158	111.727
	0.4	0.0	0.231	0.848	0.617	−0.696	−0.137	0.559	78.44	87.212	123.405
		0.4	0.296	0.824	0.527	−0.643	−0.130	0.513	78.29	87.770	116.332
		0.8	0.209	0.990	0.780	−0.616	0.010	0.625	78.35	87.680	112.667
	0.8	0.0	0.174	0.769	0.595	−0.598	−0.043	0.555	79.63	90.319	127.942
		0.4	0.250	0.788	0.539	−0.556	0.001	0.557	79.11	89.692	118.040
		0.8	0.191	0.966	0.775	−0.593	0.088	0.681	79.97	90.206	115.911
300	0.0	0.0	0.476	0.825	0.349	−0.643	−0.358	0.285	229.56	259.266	307.154
		0.4	0.523	0.792	0.269	−0.597	−0.372	0.226	229.78	262.162	298.169
		0.8	0.382	0.884	0.501	−0.571	−0.277	0.294	230.61	263.721	302.177

Table 5.8: *(continued)*

			$\beta_0 = 0.5$			$\beta_1 = -0.5$			**Estimated Pop. Size**		
N	σ_X	ρ_π	$\overline{\overline{E}}(\beta_0\|\mathbf{y})$	$\underline{\underline{E}}(\beta_0\|\mathbf{y})$	$\overline{\Delta}(\beta_0\|\mathbf{y})$	$\overline{\overline{E}}(\beta_1\|\mathbf{y})$	$\underline{\underline{E}}(\beta_1\|\mathbf{y})$	$\overline{\Delta}(\beta_1\|\mathbf{y})$	$\bar{n}$	$\underline{\underline{N}}_n$	$\overline{\overline{N}}_n$
	0.4	0.0	0.456	0.774	0.318	−0.485	−0.201	0.284	234.86	266.734	310.606
		0.4	0.510	0.760	0.250	−0.450	−0.219	0.231	235.22	268.933	301.569
		0.8	0.385	0.892	0.508	−0.467	−0.072	0.394	233.44	264.977	303.703
	0.8	0.0	0.391	0.685	0.294	−0.346	−0.069	0.277	240.31	279.232	321.512
		0.4	0.409	0.696	0.286	−0.321	−0.046	0.275	239.44	279.703	311.923
		0.8	0.330	0.865	0.535	−0.403	0.077	0.481	240.47	274.155	322.047
500	0.0	0.0	0.557	0.792	0.235	−0.586	−0.396	0.190	384.12	438.946	489.386
		0.4	0.600	0.776	0.176	−0.549	−0.402	0.147	383.24	440.485	477.341
		0.8	0.482	0.839	0.357	−0.545	−0.361	0.183	384.41	443.331	485.892
	0.4	0.0	0.543	0.754	0.211	−0.413	−0.226	0.187	390.78	446.670	490.670
		0.4	0.581	0.742	0.162	−0.381	−0.235	0.145	389.94	448.054	479.772
		0.8	0.479	0.847	0.368	−0.420	−0.133	0.287	390.38	444.679	488.080
	0.8	0.0	0.452	0.645	0.193	−0.258	−0.071	0.186	399.58	470.131	512.849
		0.4	0.467	0.664	0.197	−0.248	−0.059	0.189	400.70	472.532	505.157
		0.8	0.388	0.801	0.413	−0.344	0.035	0.380	398.96	458.613	518.879

Chapter 6

Case Studies

This chapter presents an application of the proposed methodology. As noted in Section 1.3, the affected population size estimation is a significant decision making problem in epidemiology since the reasonable estimate of prevalence is not only used for addressing the burden of a disease in a community but also essential when constructing a fiscal funding plan. The difficulty associated with this research is that there exists no reliable method of confirming the true size of the population in question. Development of the proposed methodology is motivated by this epidemiological decision making problem, and intended to address circumstances when the intentional unit who is concerned with how to lead his or her inferential problem is under the uncertainty due to the lack of information or disagreement between individuals.

The proposed methodology leads an inferential process to an unknown truth using the modelled intentional unit's prior ignorance for purposes of decision making in a particular situation for a given problem as illustrated using the examples in the preceding chapters. However, the illustrated inferential process may not function as expected with real data since the examination has been done with simulated data. More than one problem associated with the proposed methodology may be identified in practice.

Four real data sets are collected from an epidemiological literature search in a disease surveillance – Cholera epidemic (Dahiya and Gross, 1973), Down's syndrome (Zelterman, 1988), and the female users of methamphetamine and the female users of heroin (Böhning and van der Heijden, 2009). These data are re-analyzed using the proposed methodology.

6.1 Cholera Epidemic in India

This example is taken from Table 1 in the study of Dahiya and Gross (1973). This data is described as follows. An outbreak of cholera occurred in a village in India comprised of 223 households. Cholera is an infectious disease that is caused by a specific agent. It is assumed that the causative agent for developing cholera resided in the water supply servicing that village and also assumed that anyone who used drinking water supplied from the contaminated well was affected. If a household has at least one diagnosed case, then all members of the household are considered to be affected, whereas households with no diagnosed cases might still be affected. According to Böhning et al. (2005), these data were originally collected by McKendrick (1925) and presented to the Edinburgh Mathematical Society.

The data are shown in Table 6.1. y is the number of cholera cases in a household and n_y is the number of households with y cases of cholera. A total of 55 households in that village had at least one case of cholera. The question of interest is now to estimate the number n_0 of households that were infected but had undetected cases of cholera.

Table 6.1: Distribution of cholera cases by household in a village in India (McKendrick, 1925)

y	1	2	3	4	Total
n_y	32	16	6	1	55

Dahiya and Gross (1973) studied this data using a conditional maximum likelihood estimation of the zero-truncated Poisson model, and concluded that the number n_0 of households with active but undetected cases of cholera totals 34 out of 168 households with a 95% confidence interval of $(11, 57)$. Hence, the estimated size N of the population affected by the cholera epidemic in the village is 89 with a 95% confidence interval of $(66, 112)$. This example has also been examined by Blumenthal et al. (1978), Scollnik (1997), and Böhning et al. (2005). Blumenthal et al. (1978) followed the study of Dahiya and Gross (1973) and showed that the unconditional maximum likelihood estimate of n_0 is 32 (i.e., the estimated

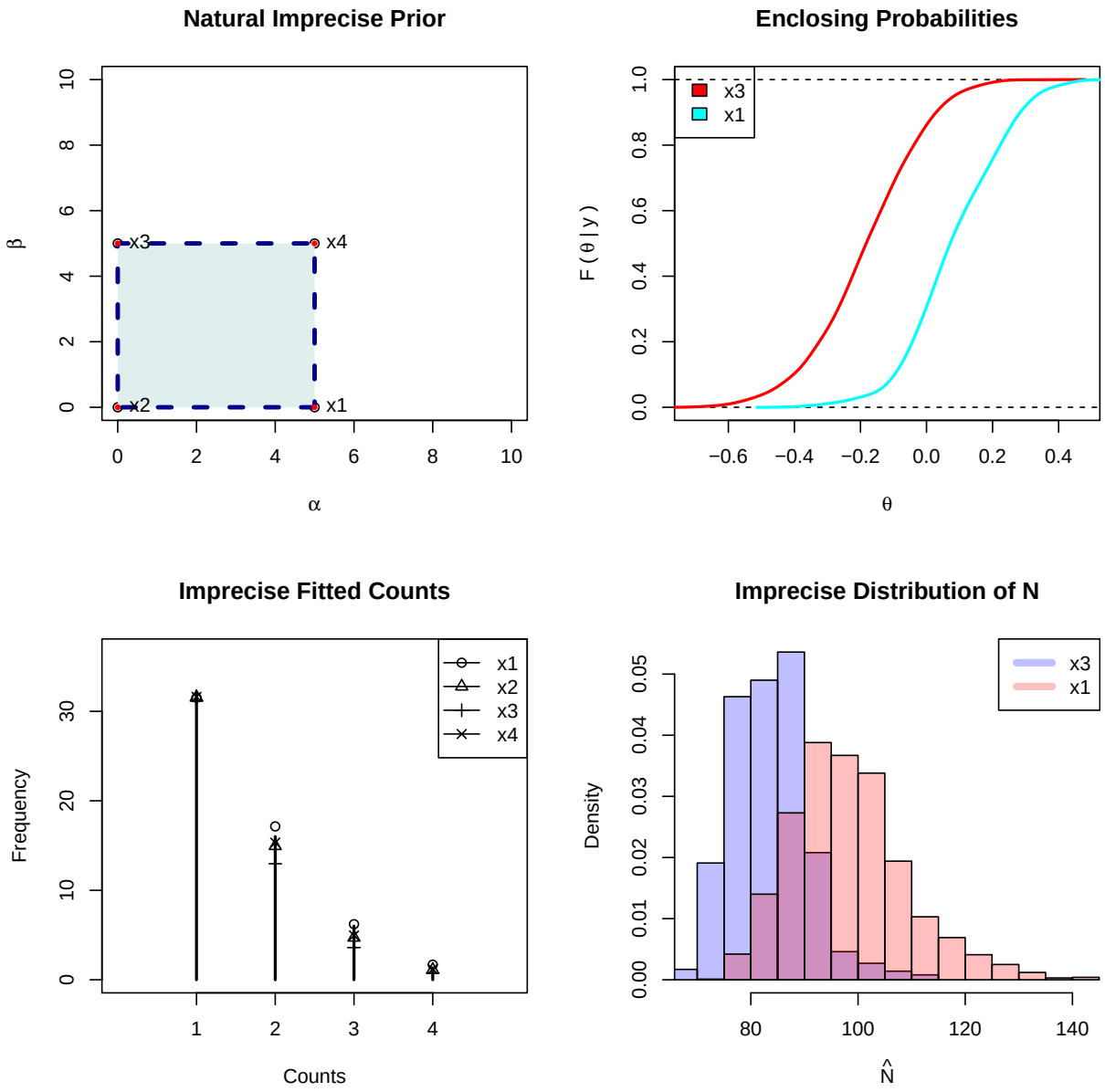
population size is 87). Scollnik (1997) reanalyzed this data from the Bayesian perspective. He demonstrated that the number n_0 of households is 34.795 (i.e., $\hat{N} \approx 89$), 31.367 ($\hat{N} \approx 86$), 35.880 ($\hat{N} \approx 90$), and 32.014 ($\hat{N} \approx 87$). Each of his estimates is produced using a different specification of the prior distributions. More details about prior specifications and the resulting posterior summaries of his data analysis can be found in Tables II, III, IV, and V in his study. McKendrick (1925) also determined the number n_0 of these households using a binomial expansion of the Poisson model. His estimate of N is 93 (p. 101). It is an interesting fact that the estimated population size $\hat{N}$ varies over different studies despite the use of the identical cholera outbreak data.

From this point forward the author of this thesis is regarded as the intentional unit who carries out the proposed imprecise inferential framework to reach the estimate. The family of log-gamma prior distributions in (3.9) is chosen for the zero-truncated Poisson sampling model for reasons of mathematical convenience. The author's state of complete-ignorance regarding the parameter of this sampling model needs to be represented by an imprecise prior. The region $\mathscr{R}_0 = \{(\alpha, \beta) | 0 \leq \alpha \leq 5, 0 \leq \beta \leq 5\}$ is characterized on the hyperparameter space Ξ to express the author's natural imprecise prior as shown in the plot on the top left panel in Figure 6.1. `x1`, `x2`, `x3`, and `x4` on that plot are the identification numbers of the extreme points of the characterized region $\mathscr{R}_0$. One may question how the author arrived at these characterization rules since the shape and size can be arbitrarily chosen. The primary reason for this characterization is that all possible prior expectations are captured in this region $\mathscr{R}_0$ before seeing the data. Please see Figure 3.4 presented in Chapter 3.

Although various polygons can be characterized in this hyperparameter space Ξ, the square is a shape that can be conveniently formed by four linear inequality constraints which requires only two upper and lower bounds of each dimension α and β on the hyperparameter space Ξ. This square also has a minimal number of constraints required to form the shape of a polygon rather than the number of constraints required for any other shape. A triangle shape may be considered since only one additional constraint is needed from the set of constraints used to form the shape of a square; however, the author does not have sufficient justification to rationalize how to assign a constraint that describes a relationship between two hyperparameters α and β. The rectangular shape is also a candidate for this characterization

since the number of constraints needed for forming a rectangle are identical to those needed for the square. The question is then a determination of the length of each hyperparameter α and β. Again, the author does not have a sufficient region to determine these lengths so that both lengths are set as equal. It remains now to determine the quantification of the length since the length of the characterized square is five times greater than the one of a unit square. The choice of length is in fact arbitrary; however, the author has chosen a length sufficiently large to be feasible for the study.

Figure 6.1: Imprecise prior, probability box of imprecise posterior, imprecise fitted counts, distribution of imprecise estimate of N in cholera epidemic data (McKendrick, 1925) analysis.



Since the natural imprecise prior is determined for this data analysis, all remaining tasks required to perform an imprecise inference are done by the `ipeglim` package (Lee and Bickis, 2013). The author now observes the data and executes the `R` functions (Please see the `R` code template provided in Section 4.2). Numerical summaries of the resulting imprecise posterior

are reported in Table 6.2. The first four rows represent the imprecise estimate of the canonical parameter $\theta = \log(\mu)$ where μ is the mean parameter of the standard Poisson sampling model which are listed in the next four rows.

Table 6.2: Imprecise posterior summary statistics at the extreme points of `x1`, `x2`, `x3`, and `x4` of the posterior hyperparameter set for the Cholera epidemics in India

		Quantile				
	Est.	0%	25%	50%	75%	100%
$E(\theta\|\mathbf{y})$						
`x1`	0.084	−0.425	−0.022	0.070	0.200	0.477
`x2`	−0.055	−0.702	−0.147	−0.047	0.059	0.461
`x3`	−0.183	−0.715	−0.284	−0.185	−0.066	0.386
`x4`	−0.031	−0.494	−0.135	−0.029	0.078	0.392
$\exp(\hat{\theta})$						
`x1`	1.088	0.654	0.978	1.072	1.221	1.612
`x2`	0.947	0.496	0.863	0.954	1.061	1.585
`x3`	0.833	0.489	0.753	0.831	0.936	1.472
`x4`	0.969	0.610	0.874	0.971	1.081	1.480
$\hat{N}$						
`x1`	83.580	68.714	78.010	83.627	88.130	114.618
`x2`	90.743	69.180	84.110	89.461	95.143	140.704
`x3`	98.311	71.383	90.492	97.469	103.955	142.181
`x4`	89.318	71.205	83.223	88.519	94.404	120.377

The last four rows represent the estimate of population size N. Based on the results in Table 6.2, the author concludes that the expected number of households with undetected cases of cholera ranges from 83.58 to 98.311. Interestingly, this range includes all estimates presented in the studies of McKendrick (1925), Dahiya and Gross (1973), Blumenthal et al. (1978), and Scollnik (1997). The author interprets this range as an indeterminate number of households due to the author's prior ignorance regarding the cholera epidemic in the village in India.

The plot on the top right panel in Figure 6.1 shows the probability box of the canonical parameter θ. Based on this probability box, the maximum and minimum posterior expectation $E(\theta|\mathbf{y})$ are produced from the posterior probability distributions labelled `x1` and `x3`, respectively. The area enclosed by two cumulative posterior probability functions $F(\theta|y)$ of

the canonical parameter θ is the degree of imprecision $\Delta_n(\theta|\mathbf{y})$ =0.267. The plot on the bottom left panel shows how imprecisely the presumed zero-truncated Poisson sampling model is fitted to the cholera epidemic data. The plot on the bottom right panel shows how imprecisely the estimated population size $\hat{N}$ is distributed. The histogram coloured light blue and light red are the distributions of the estimated population size $\hat{N}$ generated at the extreme points labelled `x1` and `x3`, respectively. The area coloured light purple is the overlap of two histograms. The two plots on the bottom of Figure 6.1 demonstrate the fact that although the presumed sampling model seems to be well fitted (by different intentional units) a discrepancy on the distribution of the prevalence estimation regarding the cholera epidemic remains.

However, it must be noted that the results presented in this section are based on the author's natural imprecise prior which is vaguely characterized. One may have more or less information than what the author has. If he or she has more information, his or her imprecise prior would be less vague by characterizing a smaller region $\mathscr{R}_0$ on the hyperparameter space Ξ. In contrast, if he or she has less information, a larger region $\mathscr{R}_0$ would be characterized so that his or her imprecise estimate would be wider than the one the author has reported here.

6.2 Down's Syndrome Data

In this section the proposed methodology is applied to the Down's syndrome data for performing an imprecise inference. Fienberg (1972), Hook and Regal (1982), and Zelterman (1988) have examined this data set. Based on the data description noted in the study by Fienberg (1972), the results of a survey on Down's syndrome births in Massachusetts during the period from January 1, 1955 to December 31, 1959, who were alive on December 31, 1966 are examined. Diagnostics of "specific", "relatively common", "congenital anomaly" are considered as a positive case of Down's syndrome. Five data sources of obstetric records, miscellaneous hospital records, Massachusetts Department of Health records, Massachusetts Department of Mental Health records and school records are used to generate this data set (p. 600). Positive cases are cross-classified in a five-way contingency table in terms of the

data sources. This contingency table can be found on page 601 in the study by Fienberg (1972).

Fienberg (1972) viewed this Down's syndrome data as multilist capture-recapture experiment data since a single missing cell in the cross-classified contingency table in terms of the presence or absence of a disease condition corresponds to the number of individuals who are not ascertained on any of the five sets of data sources. After examining various fits of log-linear models, Fienberg (1972) concluded that the total number of children with Down's syndrome during the period from 1955 to 1959 is 635 with a 95% confidence interval of $(598, 670)$ (p. 601). Regarding this estimate, Zelterman (1988) reasoned that the log-linear model missed the feature of children who are not shown in all data sources is similar to the children who are listed in one or two data sources (p. 234). He reorganized the Down's syndrome data structured in the contingency table into a distribution of frequencies by counting the number f_y of children listed on only y data sources as shown in Table 6.3.

Table 6.3: The number f_y of children with Down's syndrome listed on only y data sources (Zelterman, 1988).

y	1	2	3	4	5	Total
f_y	248	188	81	18	2	537

Zelterman (1988) reported two population size estimates. The maximum likelihood method is applied under a zero-truncated Poisson sampling model to estimate the population size. The reported number of children with Down's syndrome is 746 with a 95% confidence interval of $(714, 780)$. The second estimate reported in his study is obtained from the local estimator developed in his study. In a later study of Böhning (2008), this estimator is referred to as the Zelterman's estimator. The estimated number of children is 688 with a 95% confidence interval of $(632, 755)$. It is noticed that the first estimate has a 95% confidence interval which is narrower than the interval found by Fienberg (1972), and these

two intervals do not overlap each other. He also noted that his local estimate has a 95% confidence interval which partially overlaps on the interval found by Fienberg (1972).

Now consider the proposed methodology for inferring the parameter of a zero-truncated Poisson sampling model that describes the Down's syndrome data in Table 6.3. A family of log-gamma prior distributions is employed again for the same reason noted in the previous section 6.1. On the determination of an imprecise prior the author attempts to use information on the prevalence of Down's syndrome in newborns noted in the fact sheet provided by the National Association for Down Syndrome (2013) and the annual birth statistics in Massachusetts from 1980 to 1998 in Table 1 on page 14 in the report provided by the Division of Research and Epidemiology, Bureau of Health Information, Statistics, Research, and Evaluation, Massachusetts Department of Public Health (2012). From the fact sheet, it is noticed that a positive case of Down's syndrome is identified in approximately one in every 800 births. When considering an average annual total of births in the state of Massachusetts is approximately 80,000, the number of newborns with Down's syndrome in Massachusetts is approximately 100.

Note that the author speculates that many environmental factors, cultural considerations, and technologies regarding Down's syndrome have changed since the time when the original survey was carried out; however, it is believed that the rough estimates derived from this data is informative in some way notwithstanding that the data may not be reliable. The author multiplies this value by 4 to set the upper bound of the hyperparameter α for characterizing an imprecise prior. The value of the multiplier could be larger than 4 since it is not known how many patients with Down's syndrome are in the population. However, since the survey was conducted over a period of four years, it is reasonable to expect that there were at least 400 patients with Down's syndrome in the state of Massachusetts before seeing the data. Hence, the author's imprecise prior is represented by characterizing the region $\mathscr{R}_0 = \{(\alpha, \beta) | 0 \leq \alpha \leq 400, 0 \leq \beta \leq 1\}$ on the hyperparameter space $\boldsymbol{\Xi}$.

The numeric and graphic summaries of the imprecise posterior are presented in Table 6.4 and Figure 6.2, respectively, in the same manner as the results derived from the case of the Cholera epidemic data analysis in the previous section 6.1. Based on the author's imprecise prior, it is concluded that the number of children with Down's syndrome in the state of

Figure 6.2: Imprecise prior, probability box of imprecise posterior, imprecise fitted counts, imprecise estimate of N from the Down's Syndrome Data (Zelterman, 1988).

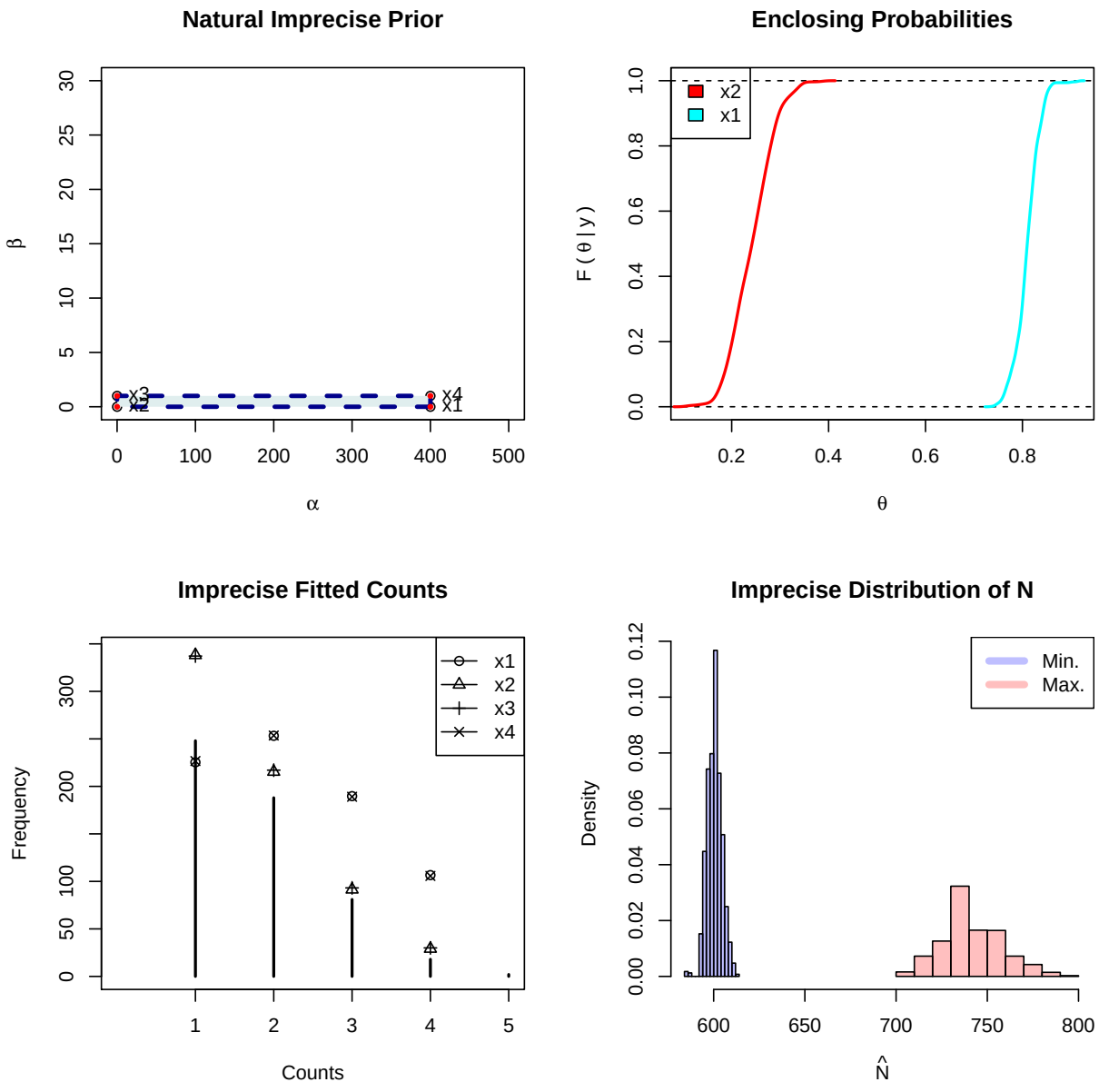


Table 6.4: Imprecise posterior summary statistics at the extreme points of `x1`, `x2`, `x3`, and `x4` of the posterior hyperparameter set for the Down's Syndrome Data (Zelterman, 1988)

		Quantile				
	Est.	0%	25%	50%	75%	100%
$E(\theta\|\mathbf{y})$						
`x1`	0.810	0.735	0.798	0.810	0.826	0.915
`x2`	0.242	0.107	0.210	0.242	0.274	0.388
`x3`	0.253	0.108	0.217	0.259	0.282	0.371
`x4`	0.805	0.718	0.781	0.804	0.831	0.888
$\exp(\hat{\theta})$						
`x1`	2.248	2.086	2.220	2.248	2.284	2.497
`x2`	1.274	1.113	1.233	1.274	1.315	1.473
`x3`	1.288	1.114	1.243	1.296	1.326	1.450
`x4`	2.237	2.050	2.183	2.234	2.296	2.429
$\hat{N}$						
`x1`	600.522	585.170	597.923	600.384	602.405	613.147
`x2`	746.049	696.621	734.093	745.637	757.787	799.834
`x3`	742.123	701.628	731.116	739.223	754.847	799.439
`x4`	601.340	588.888	597.078	601.425	605.207	616.312

Massachusetts during the period of the survey ranges from 600.522 to 746.049. The lower bound of this range approximates the lower bound of the 95% confidence interval found by Fienberg (1972) and the upper bound resembles the point estimate obtained by the maximum likelihood method. This range also overlaps most of the 95% confidence interval found by Zelterman's local estimator.

6.3 Heroin and Methamphetamine Users in Bangkok

The data regarding female heroin and methamphetamine users studied in Böhning and van der Heijden (2009) is re-analyzed in this section. In their study this data is presented in two tables (pp. 603–604). Each table shows the distribution of contact counts to the treatment institution by age. This illegal drug users data was originally part of a larger database

collected by the Office of the Narcotics Control Board, Ministry of the Prime Minister in Bangkok during the period from October 1 to December 31 in 2001. Details about data sources, reporting mechanisms and the regulation of the admission to a drug dependence treatment program, and the identification of heroin and methamphetamine drug users can be found in Section 2 (Böhning et al., 2004, p. 1076–1077). Since data is presented in the form of a distribution of frequencies, the author recreated the individual level data from the table. This data can be loaded in R for analysis by typing the following commands:

```
> library(ipeglim)
> data(heroin)
> data(methamphetamine)
```

Böhning and van der Heijden (2009) estimated the size of female heroin and methamphetamine populations using Zelterman's estimator by incorporating it with and without the explanatory variable age. When this age variable is not incorporated, the estimated female methamphetamine user population is $\hat{N} = 3714$ with a 95% confidence interval of $(1417, 6011)$ and the estimated female heroin user population is $\hat{N} = 504$ with a 95% confidence interval of $(389, 628)$. When this age variable is incorporated, the estimated female methamphetamine user population is $\hat{N} = 3772$ with a 95% confidence interval of $(1376, 6169)$ and the female heroin user population is $\hat{N} = 505$ with a 95% confidence interval of $(379, 630)$. When comparing the results produced using a regression model to those produced from using a non-regression model (i.e., no incorporation with the age variable), it is noticed that the age variable influences the estimation of the illegal drug user population size.

The author conducted a zero-truncated Poisson regression model as presented by van der Heijden et al. (2003) prior to carrying out an imprecise inference with the proposed methodology since Böhning and van der Heijden (2009) reported only the results produced by Zelterman's estimator. Note also that Böhning and van der Heijden (2009) compared the estimates of the regression parameters for the full model produced using Zelterman's regression model to those produced using the zero-truncated Poisson regression model for the data when van der Heijden et al. (2003) was studied for estimating the number of illegal immigrants in The Netherlands. Please see this numerical comparison in Table 6 on page 606 in Böhning and van der Heijden (2009). The individual level raw data of the van der Heijden et al. (2003)

study is available at http://stat.uibk.ac.at/SMIJ/. To ensure that the zero-truncated Poisson regression model is correctly implemented, the author reproduced all of the study results of van der Heijden et al. (2003) using the `ipeglim` package. Please see Appendix C.1 for these reproduced results. The following `R` codes are used for re-analyzing the heroin and methamphetamine data. The results are then summarized in Table 6.5.

```
> rm(list=ls())
> library(ipeglim)
> data(heroin)
> sfith0 <- summary(ztpreg(visits ~ 1, data=heroin, dist="poisson",
+                   ztrunc=TRUE), HT.est=TRUE)
> sfith1 <- summary(ztpreg(visits ~ age, data=heroin, dist="poisson",
+                   ztrunc=TRUE), HT.est=TRUE)
> data(methamphetamine)
> sfitm0 <- summary(ztpreg(visits ~ 1, data=methamphetamine,
+                   dist="poisson", ztrunc=TRUE), HT.est=TRUE)
> sfitm1 <- summary(ztpreg(visits ~ age, data=methamphetamine,
+                   dist="poisson", ztrunc=TRUE), HT.est=TRUE)
```

Table 6.5: The maximum likelihood estimates of the model parameters and Horvitz-Thompson's estimates of the heroin and methamphetamine population sizes (with their 95% confidence interval) using the zero-truncated Poisson regression model studied by van der Heijden et al. (2003); METH (Methamphetamine).

	Zero-Truncated Poisson				**Horvitz-Thompson**		
	Est.	S.E.	Z-score	$Pr(>\|z\|)$	$\hat{N}$	Lower	Upper
Heroin							
(Intercept)	0.655	0.050	13.188	0.000	313.770	296.258	331.282
METH.							
(Intercept)	−2.107	0.240	−8.772	0.000	2392.643	1300.464	3484.821
Heroin							
(Intercept)	1.036	0.194	5.341	0.000	315.456	297.222	333.689
age	−0.013	0.007	−1.996	0.046			
METH.							
(Intercept)	−3.688	0.919	−4.013	0.000	2751.446	1245.889	4257.002
age	0.070	0.037	1.892	0.059			

The proposed inferential framework is preformed under the setup of a zero-truncated Poisson regression model using the age variable. Since the proposed methodology has been

illustrated numerous times for cases of non-regression models in the previous Sections 6.1 and 6.2, only the regression model is studied in this section. The imprecise prior for regression parameters of the intercept of the slope in a simple regression model is represented by characterizing $\mathscr{R} = \{(b_0, b_1)| -1 \leq b_0 \leq 1, -1 \leq b_1 \leq 1\}$ on a two-dimensional hyperparameter space $\boldsymbol{\Xi} = \{(b_0, b_1)| -\infty < b_0 < \infty, -\infty < b_1 < \infty\}$. The identity matrix is used for the prior variance-covariance matrix of the regression parameters; however, the variances are controlled by varying the values 0.01, 0.1, and 5. Graphical summaries of the imprecise estimates of the regression parameter β_0 and β_1 are presented in Figure 6.3 for the female heroin user population and Figure 6.4 for the female methamphetamine user, respectively.

The author draws the reader's attention to an interesting fact that the maximum likelihood estimates obtained using the zero-truncated Poisson regression model are substantially lower than the Zelterman's estimate regardless whether or not the age variable is incorporated as shown in Table 6.5. This phenomenon is also shown on Table 6 in the study by Böhning and van der Heijden (2009). Another fact determined using the proposed methodology is that the imprecise estimate moves toward the maximum likelihood estimates of the zero-truncated Poisson regression model when a variance of larger value is given to the identity matrix as illustrated in Figures 6.3 and 6.4. This graphical feature provides an insight for interpreting the imprecise posterior of the proposed methodology. For the population of female heroin users (i.e., Figure 6.3), the graphic feature prompts the intentional unit (i.e., the author of this thesis) to have greater assurance that the two distributions of the population size estimate (two histograms coloured light blue and red on the right panel) are more directly related to each other if more data is available to the intentional unit. This feature may be useful to describe someone's action such that "My guess was pretty close since the outcome has occurred within the range that I expected before seeing the data!" since the imprecise estimate is located within the characterized imprecise prior. On the other hand, for the methamphetamine user population (i.e., Figure 6.4), the imprecise posterior does not escape the imprecise prior but is also not completely enclosed by the imprecise prior with some remaining amount of imprecision. This appearance of an imprecise estimate may also explain the situation such that "It does not seem to be right since some outcome has occurred

Figure 6.3: The imprecise estimate for the female heroin user population in Bangkok (Böhning and van der Heijden, 2009) produced from the characterization $\mathscr{R} = \{(b_0, b_1) | -1 \leq b_0 \leq 1, -1 \leq b_1 \leq 1\}$.

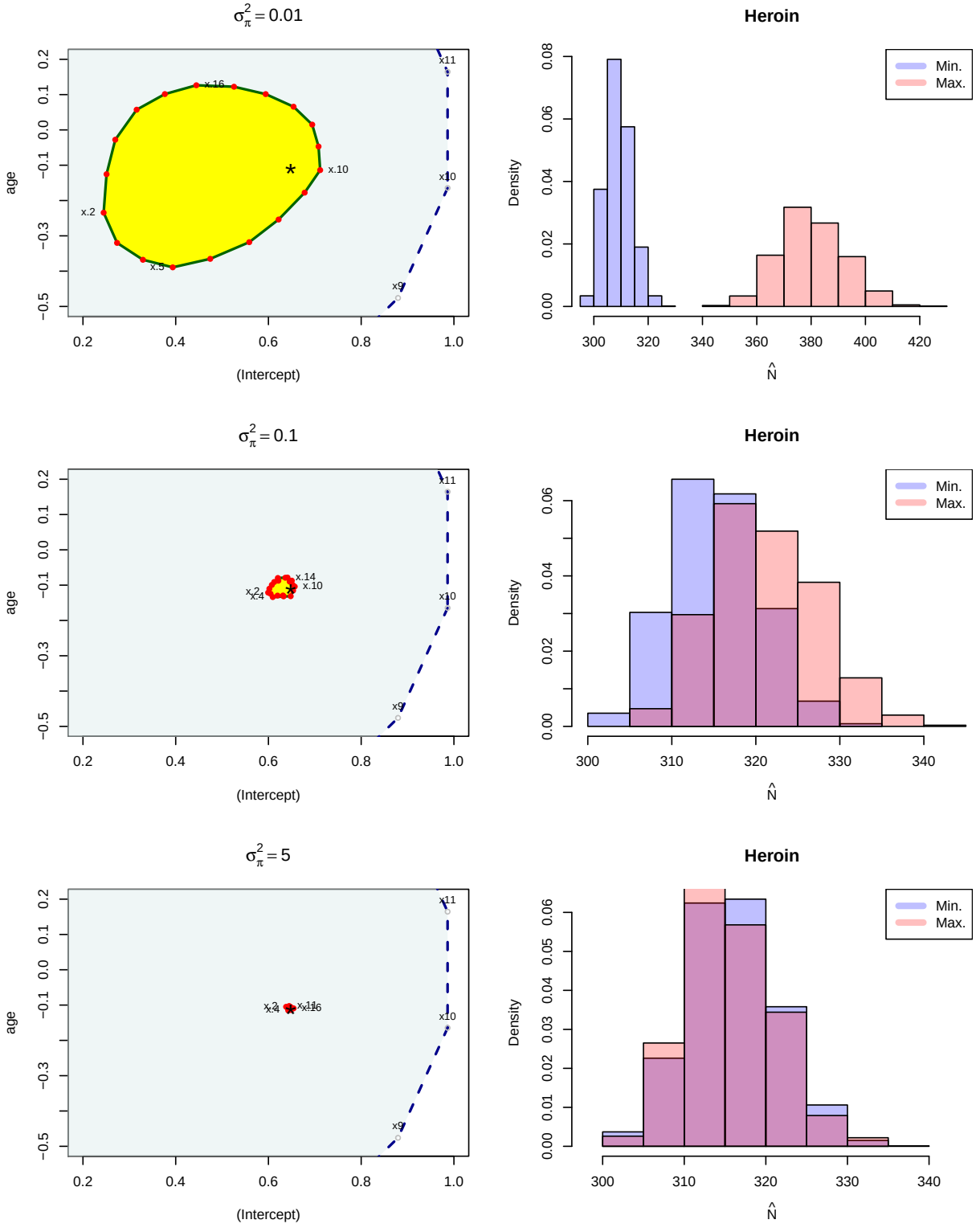
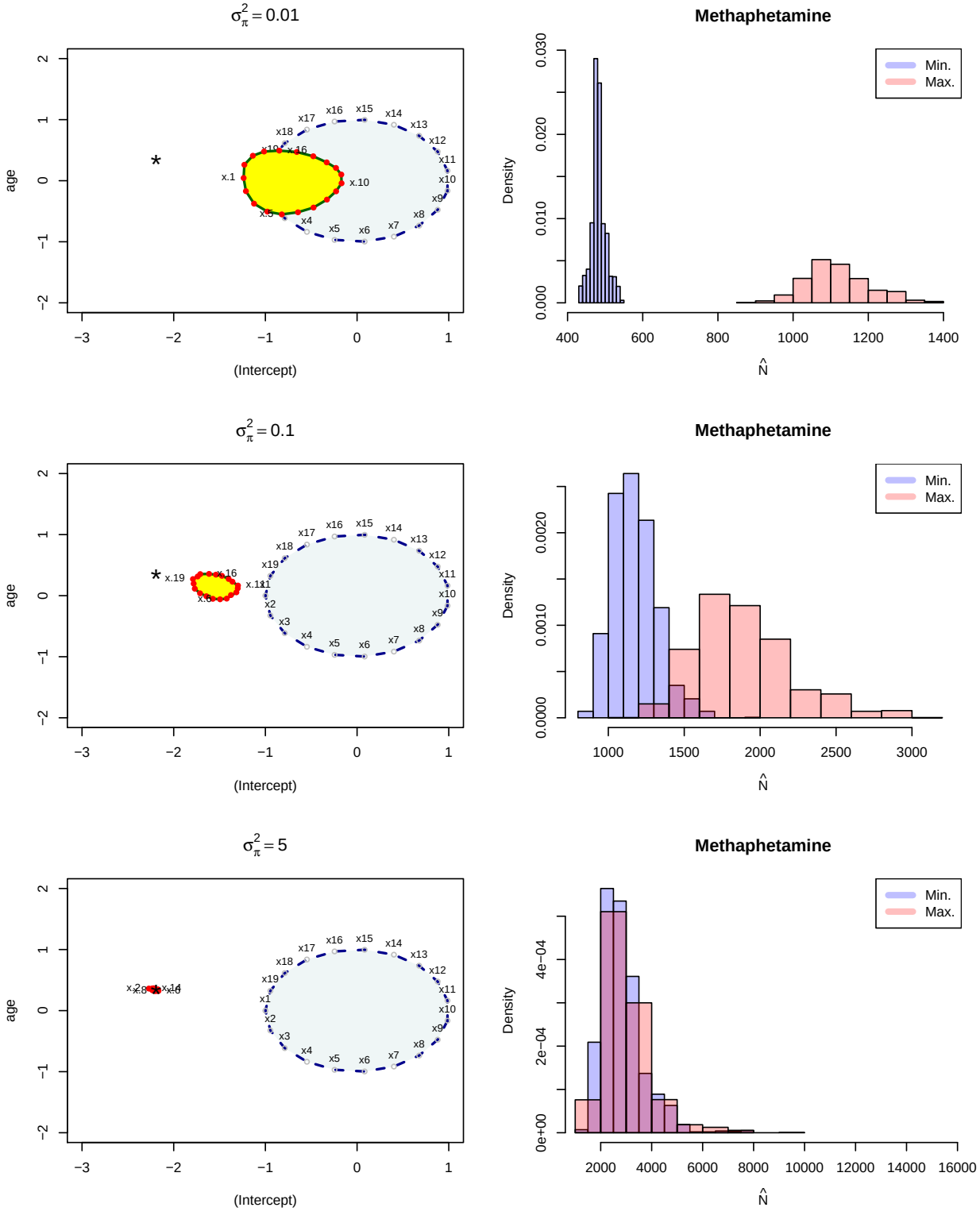


Figure 6.4: The imprecise estimate for the female methamphetamine user population in Bangkok (Böhning and van der Heijden, 2009) produced from the characterization $\mathscr{R} = \{(b_0, b_1) | -1 \leq b_0 \leq 1, -1 \leq b_1 \leq 1\}$.



that is outside of the range that I expected prior to seeing the data". As expected from this phenomenon, the two histograms are located far away from each other as shown in the plot on the top right in that figure. Hence, the author could not conclude the methamphetamine user population size and is seeking additional information at this time.

Chapter 7

Conclusion and Further Studies

This thesis is concerned with the problem of epidemiological decision making of estimating the size of a target population. It is a fact that true prevalence is typically underestimated and there exists no reliable method of confirming this estimate of prevalence. The primary objective of this thesis work is to develop a statistical reasoning framework that enables the modelling of prior ignorance and the quantification of the degree of epistemic imprecision associated with the inferential process by adopting the concept of imprecise probabilities introduced by Walley (1991).

For this methodological development, this thesis starts with providing a procedural description of the proposed methodology called the canonically parametrized imprecise inferential framework. Firstly, a sampling model is canonically parametrized in the form of an exponential family of distributions. Secondly, a conjugate prior measure of the canonical parameter is formulated. Thirdly, a set of linear inequality constraints is utilized for characterizing the convex hull that represents an imprecise prior on the hyperparameter space of the canonical parameter. Lastly, the degree of an epistemic imprecision is evaluated throughout the imprecise posterior optimization process.

All these components are synthesized to build the `ipeglim` package (Lee and Bickis, 2013) that is a collection of functions developed by the author for producing all numeric and graphic summaries contained in this thesis since arriving at an imprecise estimate for a quantity of interest is a complicated procedure associated with the characterization of a prior ignorance and a long sequence of computations associated with the process of optimizing the imprecise posterior. A brief description to the key functions in this package is summarized in Table 7.1 (a complete description is available in the documentation of the `ipeglim` package).

Table 7.1: Key functions in the `ipeglim` package and their description. A complete list of functions is available at the project site http://ipeglim.r-forge.r-project.org

Functions	Description
`model()`	Specifying a sampling model
`iprior()`	Modelling prior ignorance
	Defining a family of prior distributions
	Solving a set of given linear inequality constraints
	Searching for extreme points of a characterized convex hull
`update()`	Applying Bayes' theorem at every extreme point
	Evaluating a quantity of posterior expectation at every extreme point
`summary()`	Identifying extreme posterior expectations
`plot()`	Summarizing the resulting imprecise estimates graphically
`pbox()`	Producing the probability box

By using a family of log-gamma prior distributions, three major behaviours of the proposed inferential framework are found as a new sample is observed:

1. the translation behaviour of an imprecise prior – the convex hull representing an imprecise prior moves on the hyperparameter space,
2. a soft linear updating behaviour of an imprecise posterior expectation – the surface of an imprecise posterior expectation becomes flat, and
3. the focusing behaviour of an imprecise posterior – an imprecise posterior is stochastically ordered and squeezed by the upper and lower posterior distributions.

The imprecise learning curve generated from the proposed inferential framework was useful to graphically illustrate that a conflict between prior belief and data lasts for a certain period of the initial learning period (i.e., the period of observing the first few samples); however, a certain gap remains between the two extreme posterior expectations in this learning curve despite a sufficiently long learning period. This imprecise learning curve effectively started a discussion regarding conclusions that can be drawn from two or more intentional units having different characterization strategies for the same observations. That is, the agreement process between different individuals was illustrated by describing that a conflict initially observed between learning processes diminishes gradually following a long sequence of observational activities on the same data.

The central part of the proposed inferential framework referred to as the $\mathfrak{B}$-formulation(Lee and Bickis, 2012) introduced in Chapter 4 was integral for extending the proposed inferential framework to a case when

1. a family of normal prior distributions is considered for the instance of lacking a conjugate prior measure due to the use of other families of prior distributions for a standard Poisson likelihood,
2. either families of log-gamma and normal prior distributions are considered on a zero-truncated Poisson likelihood, and
3. a family of multivariate normal prior distributions is considered on the regression coefficients of both the standard and the zero-truncated Poisson regression models.

Simulation studies carried out under various conditions in response to concerns involving the practical use of the proposed inferential framework showed the following: when a standard Poisson sampling model is used for describing zero-truncated count data, an overestimation of the imprecise inferential framework is expected; as data are increasingly over-dispersed, both the extremes of the imprecise estimate are increasing; lastly, the imprecise estimates of regression parameters become increasingly extended as the size of a sample decreases, the correlation coefficient of a prior variance-covariance matrix increases, and the values of the variances decrease.

Throughout the case studies using four real data sets collected from an epidemiological literature in a disease surveillance, the proposed inferential framework gives insights on how to interpret the imprecise estimate. Since the imprecise prior describes our uncertainty due to the lack of information, the imprecise posterior describes our indeterminate preference that cannot support an action. The graphic features produced using the `ipeglim` package also supports someone's action in saying either "My guess was close since the outcome has occurred within the range that I expected before seeing the data" when the imprecise estimate is located within the imprecise prior or "It does not seem that the right outcome has occurred since it is outside of the range that I expected prior to seeing the data" when the imprecise posterior does not escape the imprecise prior but is also not completely enclosed by the imprecise prior. It was also found that the imprecise estimate moves toward the maximum likelihood estimates of regression coefficients when a variance of larger value is given to the identity matrix.

During the development of this proposed methodology a number of questions have arisen that require further research and investigation. These questions addressed are classified into three categories: 1. computational development – "How can be the current `ipeglim` package be extended?" and "How can be the required computation be accelerated?"; 2. model extensions – "Is this proposed methodology limited to only a Poisson sampling model?" and "Can this proposed methodology be extended to multivariate Poisson sampling model?"; and 3. alternative approaches – "Are there other approaches that can be used to estimate the size of a population with a certain condition?". The aim of this last chapter in this thesis is to present these questions and discuss potential approaches for subsequent studies.

Computational Development

The computational efficiency of the proposed methodology is attributed to the adoption of a linear programming technique that searches for the extreme posterior expectations from a set of given linear inequality constraints. However, a considerable amount of computation time is still required for the evaluation process of quantifying a normalizing constant that induces a probability measure. When zero-truncation in a sampling model is considered, this evaluation process becomes more difficult as discussed in Section 4.2. Another concern with this evaluation process is the dimensionality of the hyperparameter space which is dependent on the number of explanatory variables in the model. Although the Metropolis-Hastings algorithm is a useful numerical method as noted in Section 4.1, a sufficiently long length of Markov chain may be required in order to reach a stable quantity of interest. Indeed, this dimensionality is not only associated with the evaluation of the normalizing constant but is also associated with the number of extreme points to be searched which are directly proportional to the computing time required by the proposed imprecise inferential framework. To illustrate, please consider the application briefly mentioned in Section 6.3 regarding the illegal immigrants in The Netherlands (van der Heijden et al., 2003). The imprecise estimate of the regression parameters for various models are summarized in Tables C.1, C.2, C.3, and C.4 in Appendix C.1. For the models with nine and eight binary explanatory variables in Tables C.3 and C.4, respectively, a total of 4548 seconds ($\approx$ 75 minutes) and 2482 seconds

($\approx$ 41 minutes) were needed. The equipment used for measuring this computing time is noted on page 74. The number of extreme points associated with these models are 512 ($= 2^9$) and 256 ($= 2^8$), respectively, and the number of observation in that data is $1,880$. This computation time is measured when the Metropolis-Hasting algorithm is used for numerical approximation. Although the numerical results are successfully obtained, the computational expense is considered costly.

One possibility to consider for the reduction of the current computing time is to write C modules for the evaluation process of a normalizing constant and interface them with R in order to replace the current implementation in the `ipeglim` package.

Further Model Extensions

The $\mathfrak{B}$-formulation is the primary contributing factor that allows the proposed methodology to be extended since a family of distributions newly defined by this formulation accounts for a conjugate relationship between the imprecise prior and posterior. Only Poisson sampling models are the focus in this study; however, the principle of the $\mathfrak{B}$-formulation is to use the natural conjugate exponential family representation of a given functional form. Therefore, the proposed methodology is expected to be applied to other sampling models which are members of an exponential family of distributions. Furthermore, rigorous proofs of both focusing behaviour and soft-linearity behaviour should be explored.

A binomial sampling model would be the optimal candidate with which to examine a one-parameter exponential family of distribution as well as a special case of a multinomial sampling distribution. It is also known that a family of beta distributions is conjugate to this sampling distribution (for the multinomial sampling model, a family of Dirichlet distributions is conjugate). Although the imprecise Beta-Binomial model is well studied by Coolen (1994) and Walley (1996), it has not been examined using the proposed methodology. Zero-truncation in a binomial sampling model is also of interest for further study along with the study of a binomial sampling model since a zero-truncated binomial sampling model is a member of an exponential family of distributions. However, the author anticipates encountering the same computational issue as studied in the case of the Poisson sampling

models because there exists no closed form expression of the posterior expectation that results from the renormalizing constant of a zero-truncated binomial sampling model.

Another candidate would be a bivariate Poisson sampling model. According to Lindsey (1996), this bivariate Poisson model is a special case of a generalized exponential family of distributions (p. 38). The book written by Kocherlakota and Kocherlakota in 1992 is a good reference for studying the properties of this model. The hypothesized advantage to studying this bivariate Poisson sampling model would be the study of the correlation between two Poisson variables which are not considered in the current thesis work. Indeed, the author of this thesis had implemented the Metropolis-within-Gibbs sampling algorithm to estimate the regression parameters of a zero-truncated bivariate Poisson regression model using a trivariate reduction technique (Kocherlakota and Kocherlakota, 1992) in the Bayesian paradigm. The rationale supporting the choice of the Metropolis-within-Gibbs sampling algorithm is the complexity of functional form in the posterior expectation. It seems that a significant amount of computation time is an inevitable consequence when using this sampling algorithm since the computing time spent finding the posterior expectation at a single extreme point is was approximately twenty-five minutes which renders this study impractical without access to better equipment. The author is keenly interested in exploring the possibility of other numerical algorithms that may reduce this substantial amount of processing time for the computation of this zero-truncated bivariate Poisson regression model.

Alternative Approaches

The Horvitz-Thompson estimator is utilized in this thesis work for the problem of population size estimation following the study of van der Heijden et al. (2003). The rationale behind using this estimator is as follows: 1. the size N of a population is assumed to be consistent with the size n of observed and the size n_0 of unobserved sub-populations; 2. a probability p_0 of being in an unobserved sub-population estimated by a zero-truncated Poisson sampling model; 3. a probability $1 - p_0$ of being in an observed sub-population (i.e., an inclusion probability) is induced utilizing the Horvitz-Thompson estimator.

However, the studies of Scollnik (1997) and Puza et al. (2008) estimated the size of a target population from the Bayesian perspective instead of using the Horvitz-Thompson's estimator as it is more reasonable to set the quantity of interest as a random variable. The author is highly interested in following their approach for this population size estimation problem. In order to incorporate the proposed imprecise inferential framework with their approach, two different families of prior distributions are required to be assigned on the binomial sampling for the identification mechanism and the zero-truncated Poisson sampling model for the distribution of observed frequencies. The author is considering a future study that will apply the $\mathfrak{B}$-formulation to the combined hyperparameter space of two different families of prior distributions.

References

Alho, J. M. (1990). Logistic regression in capture-recapture models. *Biometrics*, 46(3):623–635.

Alho, J. M., Mulry, M. H., Wurdeman, K., and Kim, J. (1993). Estimating heterogeneity in the probabilities of enumeration for dual-system estimation. *Journal of the American Statistical Association*, 88(423):1130–1136.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

Ball, R., Halsey, N., Braun, M. M., Moulton, L. H., Gale, A. D., Rammohan, K., Wiznitzer, M., Johnson, R., and Salive, M. E. (2002). Development of case definitions for acute encephalopathy, encephalitis, and multiple sclerosis reports to the vaccine adverse event reporting system. *Journal of Clinical Epidemiology*, 55(8):819–824.

Barber, C. B., Habel, K., Grasman, R., Gramacy, R. B., Stahel, A., and Sterratt, D. C. (2013). *geometry: Mesh generation and surface tesselation*. R package version 0.3-3 available from http://CRAN.R-project.org/package=geometry.

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(7):1960–1979.

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

Berkelaar, M. and others (2013). *lpSolve: Interface to Lp_solve v. 5.5 to solve linear/integer programs*. R package version 5.6.7 available from http://CRAN.R-project.org/package=lpSolve.

Bernard, J.-M. (2005). An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2-3):123–150.

Bernardo, J. and Smith, A. (2000). *Bayesian Theory*. John Wiley & Sons, West Sussex, England.

Bishop, Y., Fienberg, S., and Holland, P. (2007). *Discrete Multivariate Analysis Theory and Practice*. Springer New York.

Blumenthal, S., Dahiya, R. C., and Gross, A. J. (1978). Estimating the complete sample size from an incomplete Poisson sample. *Journal of the American Statistical Association*, 73(361):182–187.

Böhning, D. (2008). A simple variance formula for population size estimators by conditioning. *Statistical Methodology*, 5(5):410–423.

Böhning, D., Dietz, E., Kuhnert, R., and Schön, D. (2005). Mixture models for capture-recapture count data. *Statistical Methods & Applications*, 14:29–43.

Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W., and Viwatwongkasem, C. (2004). Estimating the number of drug users in Bangkok 2001: A capture-recapture approach using repeated entries in one list. *European journal of epidemiology*, 19(12):1075–1083.

Böhning, D. and van der Heijden, P. G. M. (2009). A covariate adjustment for zero-truncated approaches to estimating the size of hidden and elusive populations. *The Annals of Applied Statistics*, 3(2):595–610.

Brasset-Latulippe, A., Verma, J., Mulvale, G., and Barclay, K. (2011). Transformation lessons from disease-based strategies: An environmental scan. http://dev.cfhi-fcass.ca/PublicationsAndResources/ResearchReports/articleview/11-12-17/9946107e-f35f-44d8-9068-bc09aa8feae5.aspx.

Braun, W. J. and Murdoch, D. J. (2008). *A First Course in Statistical Programming with R*. Cambridge University Press, New York, NY, USA, 1 edition.

Bray, F. and Parkin, D. M. (2009). Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *European journal of cancer (Oxford, England : 1990)*, 45(5):747–755.

Brenner, H. (1996). Effects of misdiagnoses on disease monitoring with capture—Recapture methods. *Journal of clinical epidemiology*, 49(11):1303–1307.

Bruno, G., LaPorte, R. E., Merletti, F., Biggeri, A., McCarty, D., and Pagano, G. (1994). National diabetes programs. Application of capture-recapture to count diabetes? *Diabetes care*, 17(6):548–556.

Cameron, A. and Trivedi, P. (1998). *Regression Analysis of Count Data*. Econometric Society Monographs. Cambridge University Press.

Canadian Cancer Society Steering Committee (2011). *Canadian Cancer Statistics 2011*. Canadian Cancer Society, Toronto, ON.

Canadian Institute for Health Information (2007). *Improving the Health of Canadians: Mental Health and Homelessness*. CIHI, Ottawa.

Canadian Institute for Health Information (2009). *The CIHI Data Quality Framework 2009*. CIHI, Ottawa: Ontario.

Casella, G. and Berger, R. (2001). *Statistical Inference*. Duxbury Resource Center.

Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y., and Chao, D. Y. (2001). The applications of capture-recapture models to epidemiological data. *Stat Med*, 20(20):3123–3157.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):pp. 327–335.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Cohen, A. (1991). *Truncated and Censored Samples: Theory and Applications*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis.

Coolen, F. P. A. (1994). On Bernoulli experiments with imprecise prior probabilities. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43(1):155–167.

Cormack, R. M. (1999). Problems with using capture-recapture in epidemiology: An example of a measles epidemic. *Journal of clinical epidemiology*, 52(10):909–914.

Cristiano, E., Patrucco, L., Rojas, J. I., Cáceres, F., Carrá, A., Correale, J., Garcea, O., Gold, L., Tessler, J., and Kremenchutzky, M. (2009). Prevalence of multiple sclerosis in Buenos Aires, Argentina using the capture-recapture method. *European Journal of Neurology*, 16(2):183–187.

Cruyff, M. J. L. F. and van der Heijden, P. G. M. (2008). Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Biometrical Journal*, 50(6):1035–1050.

Dahiya, R. C. and Gross, A. J. (1973). Estimating the zero class from a truncated Poisson sample. *Journal of the American Statistical Association*, 68(343):731–733.

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

Division of Research and Epidemiology, Bureau of Health Information, Statistics, Research, and Evaluation, Massachusetts Department of Public Health (2012). Massachusetts Births 2010. http://www.mass.gov/eohhs/docs/dph/research-epi/birth-report-2010.pdf.

Dobson, A. J. (2001). *An Introduction to Generalized Linear Models, Second Edition*. Chapman & Hall/CRC, 2 edition.

Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, 59(2):351–364.

Feiring, B. R. (1986). *Linear programming :an introduction*, volume 60. Sage Publications, Beverly Hills.

Feller, W. (1968). *An Introduction to Probability Theory and Its Application.*, volume 1. John Wiley & Sons Inc, New York, Wiley.

Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., and Ginzburg, L. (2007). *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*. SAND (Series) (Albuquerque, N.M.). Sandia National Laboratories.

Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete 2k contingency tables. *Biometrika*, 59(3):591–603.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg.

Gill, J. (2008). *Bayesian Methods :A Social and Behavioral Sciences Approach*. Chapman & Hall/CRC, Boca Raton, 2 edition.

Goldman, G. S. (2003). Using capture–recapture methods to assess varicella incidence in a community under active surveillance. *Vaccine*, 21(27-30):4250–4255.

Gordis, L. (2009.). *Epidemiology*. Elsevier/Saunders, Philadelphia.

Gurmu, S. (1991). Tests for detecting overdispersion in the positive Poisson regression model. *Journal of Business & Economic Statistics*, 9(2):215–222.

Gutiérrez-Peña, E. and Smith, A. F. M. (1995). Conjugate Parameterizations for Natural Exponential Families. *Journal of the American Statistical Association*, 90(432):pp. 1347–1356.

Guyatt, G., Rennie, D., Meade, M., and Cook, D. (2008). *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, Second Edition (Jama & Archives Journals)*. McGraw-Hill Professional, 2 edition.

Hastings, W. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57(1):97–109.

Hay, G., Gannon, M., MacDougall, J., Eastwood, C., Williams, K., and Millar, T. (2009). Capture–recapture and anchored prevalence estimation of injecting drug users in England: national and regional estimates. *Statistical methods in medical research*, 18(4):323–339.

Health Canada (2003). *Chronic Disease Surveillance in Canada :A Background Paper*. Health Canada, Ottawa: Centre for Surveillance Coordination, Health Surveillance Coordination Division.

Hogg, R. V., McKean, J. W., and Craig, A. T. (2013). *Introduction to Mathematical Statistics*. Pearson, Boston, 7th edition.

Hook, E. B. and Regal, R. A. (1993). Effect of variation in probability of ascertainment by sources (variable catchability) upon capture-recapture estimates of prevalence. *American Journal of Epidemiology*, 137(10):1148–1166.

Hook, E. B. and Regal, R. R. (1982). Validity of Bernoulli census, log-linear, and truncated Binomial models for correcting for underestimates in prevalence studies. *American Journal of Epidemiology*, 116(1):168–176.

Hook, E. B. and Regal, R. R. (1995). Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Reviews*, 17(2):243–264.

Hook, E. B. and Regal, R. R. (1999). Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. *Journal of Clinical Epidemiology*, 52(10):917–926.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

International Working Group for Disease Monitoring and Forecasting (1995a). Capture-recapture and multiple-record systems estimation I: History and theoretical development. *American Journal of Epidemiology*, 142(10):1047–1058.

International Working Group for Disease Monitoring and Forecasting (1995b). Capture-recapture and multiple-record systems estimation II: Applications in human diseases. *American Journal of Epidemiology*, 142(10):1059–1068.

Iron, K. and Manuel, D. G. (2007). Quality assessment of administrative data (QuAAD): an opportunity for enhancing Ontario's health data. Technical report, Institute for Clinical Evaluative Sciences, Toronto, ON. ICES Investigative Report.

Irony, T. and Singpurwalla, N. (1997). Noninformative priors do not exist: a Discussion with Jose M. Bernardo. *Journal of Statistical Inference and Planning*, 65:159–189.

Johnson, N., Kemp, A., and Kotz, S. (2005). *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley.

Johnson, N. L. (1967). Note on a uniqueness relation in certain accident proneness models. *Journal of the American Statistical Association*, 62(317):288–289.

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

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. Statistics: A Series of Textbooks and Monogrphs. Marcel Dekker.

Lange, K. (1999). *Numerical Analysis for Statisticians*. Springer, New York.

Lee, C. H. and Bickis, M. (2012). Imprecise probability estimates for GLIM. Fifth Workshop on Principles and Methods of Statistical Inference with Interval Probabilit, Department of Statistics, Ludwig-Maximilians University Munich (LMU), Munich, Germany, September 11, 2012.

Lee, C. H. and Bickis, M. (2013). *ipeglim: Imprecise Inferential Framework on Statistical Reasoning*. R package version 0.3.54 available from http://r-forge.r-project.org/projects/ipeglim.

Lee, P. M. (1989). *Bayesian Statistics: An Introduction*. Oxford, Oxford, England.

Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In Härdle, W. and Rönz, B., editors, *Compstat 2002 — Proceedings in Computational Statistics*, pages 575–580. Physica Verlag, Heidelberg. ISBN 3-7908-1517-9.

Leisch, F., Weingessel, A., and Hornik, K. (2012). *bindata: Generation of Artificial Binary Data*. R package version 0.9-19 available from http://CRAN.R-project.org/package=bindata.

Lindsey, J. K. (1996). *Parametric Statistical Inference*. Oxford University Press, New York.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Monographs on Statistics and Applied Probability. Taylor & Francis.

McKendrick, A. G. (1925). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130.

Metropolis, N., Rosenbluth, A., M.N., R., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Migon, H. d. S. and Gamerman, D. (1999). *Statistical Inference :An Integrated Approach*. Arnold, London; New York.

Nanan, D. J. and White, F. (1997). Capture-recapture: Reconnaissance of a demographic technique in epidemiology. *Chrnoic Dis Can.*, 18(4):144–148.

National Association for Down Syndrome (2013). Down Syndrome Facts. http://www.nads.org/docs/DS_Facts.pdf.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.

Papoz, L., Balkau, B., and Lellouch, J. (1996). Case counting in epidemiology: Limitations of methods based on multiple data sources. *International journal of epidemiology*, 25(3):474–478.

Parkin, D. M. and Bray, F. (2009). Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. *European journal of cancer (Oxford, England : 1990)*, 45(5):756–764.

Peragallo, M. S., Urbano, F., Lista, F., Sarnicola, G., and Vecchione, A. (2011). Evaluation of cancer surveillance completeness among the Italian army personnel, by capture-recapture methodology. *Cancer Epidemiology*, 35(2):132–138.

Petersen, C. G. J. (1895). The yearly immigration of young plaice into the Limfjord from the German sea. *Report of the Danish Biological Station*, 6:5–84.

Plante, N., Rivest, L.-P., and Tremblay, G. (1998). Stratified capture-recapture estimation of the size of a closed population. *Biometrics*, 54(1):47–60.

Ponzio, M., Perotti, P. G., Monti, M. C., Montomoli, C., Bartolomeo, P. S., Iannello, G., and Mariani, S. (2010). Prevalence estimates of alcohol related problems in an area of northern Italy using the capture–recapture method. *The European Journal of Public Health*, 20(5):576–581.

Porta, M. S. (2008.). *A Dictionary of Epidemiology*. Oxford University Press, Oxford.

Public Health Agency of Canada (2010a). Report from the Canadian Chronic Disease Surveillance System: Hypertension in Canada, 2010.

Public Health Agency of Canada (2010b). What is the impact of osteoporosis in Canada and what are Canadians doing to maintain healthy bones? Available from: http://www.phac-aspc.gc.ca/cd-mc/osteoporosis-osteoporose/index-eng.php.

Puza, B. D., Johnson, H. L., O'Neill, T. J., and Barry, S. C. (2008). Bayesian truncated Poisson regression with application to Dutch illegal immigrant data. *Communications in Statistics - Simulation and Computation*, 37(8):1565–1577.

Quaeghebeur, E. (2009). *Learning from samples using coherent lower previsions*. PhD thesis, Ghent University.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rider, P. R. (1953). Truncated Poisson distributions. *Journal of the American Statistical Association*, 48(264):826–830.

Robert, C. P. and Casella, G. (2009). *Introducing Monte Carlo Methods with R*. Springer Verlag, 1 edition.

Roberts, C. (2007). *The Bayesian Choice*. Springer, Paris, France, 2nd edition.

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5.

Savva, G. M. and Morris, J. K. (2009). Ascertainment and accuracy of Down syndrome cases reported in congenital anomaly registers in England and Wales. *Archives of disease in childhood.Fetal and neonatal edition*, 94(1):F23–7.

Scollnik, D. P. M. (1997). Inference concerning the size of the zero class from an incomplete poisson sample. *Communications in Statistics - Theory and Methods*, 26(1):221–236.

Thacker, S. (2010). *Historical Development*, chapter 1, pages 1–16. Principles and Practice of Public Health Surveillance. Oxford University Press, New York, 3 edition.

Thandrayen, J. and Wang, Y. (2010). Capture–recapture analysis with a latent class model allowing for local dependence and observed heterogeneity. *Biometrical Journal*, 52(4):552–561.

Tierney, L. and Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.

Tierney, L., Kass, R., and Kadane, J. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84(407):710–716.

Troffaes, M. and Destercke, S. (2011). Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52(6):767–791.

UNAIDS/WHO (2010). Guidelines on estimating the size of populations most at risk to HIV. Technical report.

van der Heijden, P. G., Bustami, R., Cruyff, M. J., Engbersen, G., and van Houwelingen, H. C. (2003). Point and interval estimation of the population size using the truncated Poisson regression model. *Statistical Modelling*, 3(4):305–322.

Van Hest, N. A. H., Grant, A. D., Smit, F., Story, A., and Richardus, J. H. (2008). Estimating infectious diseases incidence: validity of capture-recapture analysis and truncated models for incomplete count data. *Epidemiology and infection*, 136(1):14–22.

Van Hest, N. A. H., Smit, F., Baars, H. W. M., Vries, G. D., Haas, P. E. W. D., Westenend, P. J., Nagelkerke, N. J. D., and Richardus, J. H. (2007). Completeness of notification of Tuberculosis in The Netherlands: How reliable is record-linkage and capture-recapture analysis? *Epidemiology and infection*, 135(6):1021–1029.

Vergne, T., Calavas, D., Cazeau, G., Durand, B., Dufour, B., and Grosbois, V. (2012). A Bayesian zero-truncated approach for analysing capture–recapture count data from classical scrapie surveillance in France. *Preventive Veterinary Medicine*, 105(1–2):127–135.

Walford, H., Kearns, B., and Barron, S. (2011). Technical Briefing 8: Prevalence Modelling. Technical report, Association of Public Health Observatories.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Monographs on statistics and applied probability. Chapman and Hall.

Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):3–57.

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

Wang, Y., Druschel, C. M., Cross, P. K., Hwang, S.-A., and Gensburg, L. J. (2006). Problems in using birth certificate files in the capture-recapture model to estimate the completeness of case ascertainment in a population-based birth defects registry in New York State. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 76(11):772–777.

Weems, K. and Smith, P. (2004). On robustness of maximum likelihood estimates for Poisson-lognormal models. *Statistics & Probability Letters*, 66(2):189–196.

Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept*. Physica-Verlag HD.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Springer.

Zelterman, D. (1988). Robust estimation in truncated discrete distributions with application to capture-recapture experiments. *Journal of Statistical Planning and Inference*, 18(2):225–237.

Zhao, Y., Connors, C., Wright, J., Guthridge, S., and Bailie, R. (2008). Estimating chronic disease prevalence among the remote aboriginal population of the Northern Territory using multiple data sources. *Australian and New Zealand Journal of Public Health*, 32(4):307–313.

Zwane, E. and van der Heijden, P. (2005). Population estimation using the multiple system estimator in the presence of continuous covariates. *Statistical Modelling*, 5(1):39–52.

Appendix A

Derivations

A.1 Log-Gamma Distribution

Probability Density Function: By transforming a random variable $\theta = \log(\mu)$,

$$\begin{aligned}\pi(\theta|\alpha,\beta) &= \pi_\mu(e^\theta)\left|\frac{\partial}{\partial\theta}e^\theta\right| = \frac{\beta^\alpha}{\Gamma(\alpha)}(e^\theta)^{\alpha-1}e^{-\beta(e^\theta)}e^\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}e^{\alpha\theta-\beta e^\theta} \equiv \text{Log-Gamma}(\alpha,\beta),\end{aligned}$$

where $-\infty < \theta < \infty$, $\alpha > 0$ and $\beta > 0$. The function `dlgamma(x,shape,rate,scale=1/rate)` in the `ipeglim` package computes the probability density of a log-gamma distribution with `shape=a` and `rate=b`.

Moment Generating Function (m.g.f.):

$$\begin{aligned}M_\theta(t) &= E(e^{t\theta}) = \int_{-\infty}^{\infty} e^{t\theta}\frac{\beta^\alpha}{\Gamma(\alpha)}e^{\alpha\theta-\beta e^\theta}d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_{-\infty}^{\infty} e^{(\alpha+t)\theta-\beta e^\theta}d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+t)}{\beta^{\alpha+t}} \\ &= \frac{\Gamma(\alpha+t)}{\Gamma(\alpha)}\beta^{-t},\end{aligned}$$

Expectation Value of θ: By taking the first derivative of the m.g.f. $M_\theta(t)$ at $t = 0$,

$$\begin{aligned}\frac{\partial}{\partial t}M_\theta(0) &= \frac{1}{\Gamma(\alpha)}\left\{\left(\frac{\partial}{\partial t}\Gamma(\alpha+t)\right)\beta^{-t} + \Gamma(\alpha+t)\left(\frac{\partial}{\partial t}e^{-t\log(\beta)}\right)\right\}\Bigg|_{t=0} \\ &= \frac{1}{\Gamma(\alpha)}\left\{\Gamma'(\alpha+t)\beta^{-t} + e^{-t\log(\beta)}\log(\beta)(-1)\right\}\Bigg|_{t=0} \\ &= \frac{\Gamma(\alpha+t)}{\Gamma(\alpha)}\beta^{-t}\left\{\psi(\alpha+t) - \log(\beta)\right\}\Bigg|_{t=0} \\ &= \psi(\alpha) - \log(\beta),\end{aligned}$$

where $\psi(\alpha)$ is a `digamma` function which is defined as

$$\psi(\alpha) = \frac{\partial}{\partial\alpha}\log\Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}.$$

Posterior Expectation $E(\theta|\mathbf{y})$**:** Consider $\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$ samples which are identically and independently drawn from a Poisson sampling model $f(y|\mu) = 1/y!e^{-\mu}\mu^y$. When a family of log-gamma prior distributions $\pi(\theta)$ with a shape parameter α and a rate β is assumed on the natural parameter $\theta = \log(\mu)$, the resultant family of posterior distributions is then

$$p(\theta|\mathbf{y}) \quad \propto \quad e^{(n\bar{\mathbf{y}}\theta - e^{\theta})} e^{(\alpha\theta - \beta e^{\theta})} = e^{(\alpha + n\bar{\mathbf{y}})\theta - (\beta+1)e^{\theta}},$$

which is a family of a log-gamma distributions with a shape parameter $\alpha' = \alpha + n\bar{\mathbf{y}}$ and a rate $\beta' = \beta + n$, where $\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Hence, a family of log-gamma posterior expectations is $E_{\pi}(\theta|y) = \psi(\alpha + n\bar{\mathbf{y}}) - \log(\beta + n)$.

A.2 Standard Poisson Regression Model

The Poisson log-likelihood is given by

$$\log \mathscr{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \mathbf{x}_i' \boldsymbol{\beta} - e^{\mathbf{x}_i' \boldsymbol{\beta}} - \log \Gamma(y_i + 1) \right] \tag{A.1}$$

The score function of the Poisson log-likelihood is obtained by taking the first derivative in terms of β

$$\frac{d}{d\beta} \log \mathscr{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i - e^{\mathbf{x}_i' \boldsymbol{\beta}} \right] \mathbf{x}_i \tag{A.2}$$

The Hessian matrix of regression parameters β is

$$\frac{d^2}{d\boldsymbol{\beta} d\boldsymbol{\beta}'} \log \mathscr{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} - \left(e^{\mathbf{x}_i' \boldsymbol{\beta}} \right) \mathbf{x}_i \mathbf{x}_i' \tag{A.3}$$

Fisher's information matrix is an expectation of the negative Hessian matrix over the sampling distribution

$$\frac{d^2}{d\mathbf{x}_i d\mathbf{x}_i'} \left[\frac{d^2}{d\boldsymbol{\beta} d\boldsymbol{\beta}'} \log \mathscr{L}(\boldsymbol{\beta}) \right] \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} = \frac{d^2}{d\mathbf{x}_i d\mathbf{x}_i'} \left[\sum_{i=1}^{n} \left(e^{\mathbf{x}_i' \boldsymbol{\beta}_0} \right) \mathbf{x}_i \mathbf{x}_i' \right] \tag{A.4}$$

Consider the simple linear regression $\log(\mu_i) = b_0 + b_1 x_i$. Assume that x_1 and x_2 are the realizations of x_i; then, $\mathbf{x}_1 = (1, x_1)^T$ and $\mathbf{x}_2 = (1, x_2)^T$. The above Fisher's Information matrix can be explicitly written as:

$$\frac{d^2}{d\mathbf{x}_i d\mathbf{x}_i'} \left[e^{b_0 + b_1 x_1} \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 \end{bmatrix} + e^{b_0 + b_1 x_2} \begin{bmatrix} 1 & x_2 \\ x_2 & x_2^2 \end{bmatrix} \right] \tag{A.5}$$

The optima of x_1 and x_2 can be found by minimizing the determinant of (A.5):

$$D(\mathbf{x}) \;=\; e^{(\mathbf{x}_1'+\mathbf{x}_2')\mathbf{b}}(x_1-x_2)^2$$

The minimum value of determinant is found when $x_1 = x_2$.

A.3 Zero-Truncated Poisson Log-Likelihood

The log-likelihood of a zero-truncated Poisson regression model is given by

$$\log \mathscr{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \mathbf{x}_i'\boldsymbol{\beta} - e^{\mathbf{x}_i'\boldsymbol{\beta}} - \log\Gamma(y_i+1) - \log(1-e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}) \right] \tag{A.6}$$

The score function of the Poisson log-likelihood is obtained by taking the first derivative in terms of β

$$\begin{aligned}
\frac{\partial}{\partial\beta}\mathscr{L}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \left[y_i\mathbf{x}_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i - \frac{e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i}{1-e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}} \right] && \text{(A.7)}\\
&= \sum_{i=1}^{n} \left[y_i - \mu_i - \frac{e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1-e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}} \cdot \frac{/e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}}{/e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}} \right] \mathbf{x}_i && \text{(A.8)}\\
&= \sum_{i=1}^{n} \left[y_i - \mu_i - \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}-1} \right] \mathbf{x}_i && \text{(A.9)}
\end{aligned}$$

The Hessian matrix of regression parameters β is

$$\begin{aligned}
\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\log\mathscr{L}(\boldsymbol{\beta}) &= \frac{d}{d\boldsymbol{\beta}'}\sum_{i=1}^{n}\left[y_i - e^{\mathbf{x}_i'\boldsymbol{\beta}} - \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}-1} \right]\mathbf{x}_i && \text{(A.10)}\\
&= \sum_{i=1}^{n}\left[-e^{\mathbf{x}_i'\boldsymbol{\beta}}\mathbf{x}_i' - \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}\mathbf{x}_i'(e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}-1) - e^{\mathbf{x}_i'\boldsymbol{\beta}}e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}e^{\mathbf{x}_i'\boldsymbol{\beta}}\mathbf{x}_i'}{(e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}-1)^2} \right]\mathbf{x}_i && \text{(A.11)}\\
&= \sum_{i=1}^{n}\left[-e^{\mathbf{x}_i'\boldsymbol{\beta}} - \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}(e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}-1) - e^{\mathbf{x}_i'\boldsymbol{\beta}}e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}e^{\mathbf{x}_i'\boldsymbol{\beta}}}{(e^{\exp(\mathbf{x}_i'\boldsymbol{\beta})}-1)^2} \right]\mathbf{x}_i\mathbf{x}_i' && \text{(A.12)}\\
&= \sum_{i=1}^{n}\left[-\mu_i - \frac{\mu_i e^{\mu_i} - \mu_i - \mu_i^2 e^{\mu_i}}{(e^{\mu_i}-1)^2} \right]\mathbf{x}_i\mathbf{x}_i' && \text{(A.13)}\\
&= -\sum_{i=1}^{n}\mu_i\left[1 - \frac{\mu_i e^{\mu_i} - e^{\mu_i} + 1}{(e^{\mu_i}-1)^2} \right]\mathbf{x}_i\mathbf{x}_i' && \text{(A.14)}
\end{aligned}$$

A.4 Prior Measure Formulation by Diaconis and Ylvisaker (1979)

The Poisson sampling model discussed in this thesis is an instance of a one-parameter natural exponential family model in the form of

$$f(y|\theta) = h(y)\exp\{y\theta - \mathcal{A}(\theta)\}, \quad \theta \in \Theta. \tag{A.15}$$

where $h(y)$ is a base measure, θ is a natural parameter, Θ is a natural parameter space, $\mathcal{A}(\cdot)$ is a log-normalizer. Prior distributions $\pi_{n_0,m_0}(\theta)$ conjugate to $f(y|\theta)$ are defined by

$$\pi_{n_0,m_0}(\theta) = k(n_0, m_0)\exp\{n_0 m_0\theta - n_0\mathcal{A}(\theta)\} \tag{A.16}$$

where n_0 and m_0 are the hyperparameters that represent a prior strength and a prior expectation, respectively. The normalizing constant $k(n_0, m_0)$ is defined in terms of θ on Θ in order to make a probability distribution such that

$$k(n_0, \mu_0)^{-1} = \int \exp\{n_0 m_0\theta - n_0\mathcal{A}(\theta)\}d\theta \tag{A.17}$$

When n i.i.d. samples are taken from $f(y|\theta)$, the likelihood of θ is given by

$$L(\theta|y) = \prod_{i=1}^{n} f(y_i|\theta) = \frac{1}{\prod_{i=1}^{n} y_i!}\exp\{\theta n\bar{y} - n\mathcal{A}(\theta)\} \tag{A.18}$$

The posterior of θ is found as

$$\begin{aligned} p(\theta|y) &= k(n_1, m_1)\exp\{(n_0 m_0 + n\bar{y})\theta - (n_0 + n)\mathcal{A}(\theta)\} && \text{(A.19)} \\ &= k(n_1, m_1)\exp\{n_1 m_1\theta - n_1\mathcal{A}(\theta)\} && \text{(A.20)} \end{aligned}$$

where

$$n_1 = n_0 + n, \qquad m_1 = \frac{n_0 m_0 + n\bar{y}}{n_1}. \tag{A.21}$$

and

$$k(n_1, m_1) = \int \exp\{n_1 m_1\theta - n_1\mathcal{A}(\theta)\}\, d\theta. \tag{A.22}$$

Appendix B

Functions and Distributions

B.1 Moments of Zero-Truncated Poisson

Probability mass function (pmf), moment generating function (mgf), and the first three moments of the zero-truncated Poisson (ZTP) for mean, variance, and skewness are listed. The same representations of pmf, mgf, and the first three moments are also obtained using the form of expoential family of distributions.

PMF The pmf $f_T(y)$ of ZTP is obtained by conditioning an (untruncated) standard Poisson $f(y)$ on $y > 0$ as below:

$$f_T(y|y>0) = \frac{f(y, y>0)}{f(y>0)} = \frac{f(y)}{f(y>0)} = \left[\frac{1}{1-f(0)}\right]\left[\frac{e^{-\mu}\mu^y}{\Gamma(y+1)}\right] = \frac{1}{1-e^{-\mu}}f(y), \quad \text{(B.1)}$$

which is a scaled pmf of $f(y)$, where $y = 1, 2, \ldots$.

MGF The mgf $M_T(t)$ of $f_T(y)$ is obtained by the use of $e^x = \sum_{y=0}^{\infty} y^n/n!$.

$$\begin{aligned} M_T(t) &= E(e^{ty}) = \sum_{y=1}^{\infty} e^{ty}\frac{e^{-\mu}\mu^y}{y!(1-e^{-\mu})} = \frac{e^{-\mu}}{1-e^{-\mu}}\sum_{y=1}^{\infty}\frac{(e^t\mu)^y}{y!} && \text{(B.2)} \\ &= \frac{e^{-\mu}}{1-e^{-\mu}}[e^{e^t\mu}-1] = \frac{1}{1-e^{-\mu}}(e^{(e^t-1)\mu}-1) && \text{(B.3)} \\ &= [M(t)-1]/(1-f(0)), && \text{(B.4)} \end{aligned}$$

where $M(t) = e^{\mu(e^t-1)}$ is the mgf of $f(y)$ and $f(0) = e^{-\mu}$.

First Moment for Mean The first central moment $E(Y)$ of $f_T(y)$ is obtained using $M_T'(t) = M'(t)/(1-e^{-\mu})$, where $M'(t) = M(t)e^t\mu$,

$$E(Y) = M_T'(t)|_{t=0} = \mu/(1-e^{-\mu}), \quad \text{(B.5)}$$

which is the maximum likelihood estimate $\hat{\mu}$ of $f_T(\mu)$.

Second Moment for Variance The second moment $E(Y^2)$ is obtained for $V(Y) = E(Y^2) - [E(Y)]^2$ using $M_T''(t) = M''(t)/(1-e^{-\mu})$, where $M''(t) = M(t)e^{2t}\mu^2 + M'(t)$:

$$E(Y^2) = M_T''(t)|_{t=0} = \mu(\mu+1)/(1-e^{-\mu}) = (\mu+1)E(Y) \quad \text{(B.6)}$$

Thus,

$$V(Y) = (\mu+1)E(Y) - [E(Y)]^2 = (\mu+1-E(Y))E(Y) \quad \text{(B.7)}$$
$$= \mu(1-e^{-\mu}-\mu e^{-\mu})/(1-e^{-\mu})^2 \quad \text{(B.8)}$$

Third Moment for Skewness The third moment $E(Y^3)$ is obtained for skewness γ_1, which is defined as a ratio of the third central moment $E([Y-E(Y)]^3)$ to a quantity of $[V(Y)]^{3/2}$, using $M_T'''(t) = M'''(t)/(1-e^{-\mu})$, where $M'''(t) = M(t)e^{2t}(e^t\mu+2)\mu^2 + M''(t)$.

$$E(Y^3) = M_T'''(t)|_{t=0} = (\mu^2+3\mu+1)\mu/(1-e^{-\mu}) = (\mu^2+3\mu+1)E(Y). \quad \text{(B.9)}$$

Thus,

$$\gamma_1 = \frac{E([Y-E(Y)]^3)}{V(Y)^{3/2}}$$
$$= \frac{E(Y^3) - 3E(Y^2)E(Y) + 3E(Y)[E(Y)]^2 - [E(Y)]^3}{V(Y)^{3/2}}$$
$$= \frac{(\mu^2+3\mu+1)E(Y) - 3(\mu+1)[E(Y)]^2 + 2[E(Y)]^3}{[V(Y)]^{3/2}}$$

B.2 Moments of Zero-Truncated Negative Binomial

PMF The pmf of ZTNB is obtained by conditioning an (untruncated) standard negative binomial distribution $f(y)$ on $y > 0$. Since $f(y=0|r,p) = p^r$,

$$f_T(y) = \frac{1}{1-p^r}\binom{y+r-1}{r-1}p^r(1-p)^y \quad \text{(B.10)}$$

where $y = 1, 2, \ldots, r$ is a shape parameter (in the mixture of Gamma-Poisson model) and and a Poisson mean parameter $\mu = rp/(1-p)$. With our parametrization, $p = r/(r+\mu)$, and $f(y)$ goes to the Poisson, as $r \to \infty$.

MGF The mgf is obtained using $(1-p)^{-r} = \sum_{x=0}^{\infty}\binom{y+r-1}{r-1}p^x$ as below:

$$M_T(t) = E(e^{ty}) = \sum_{y=1}^{\infty} e^{ty}\frac{1}{(1-p^r)}\binom{y+r-1}{r-1}p^r(1-p)^y \quad \text{(B.11)}$$
$$= \frac{p^r}{(1-p^r)}[(1-e^tq)^{-r}-1] \quad \text{(B.12)}$$
$$= k[M(t)-p^r], \quad \text{(B.13)}$$

where $k = 1/(1-p^r)$, $q = 1-p$, and $M(t) = p^r/(1-e^t q)^r$. The first three moments are listed for computing the mean, variance, and skewness.

$$\begin{aligned} M'_T(t) &= kp^r qre^t(1-qe^t)^{-r-1} && \text{(B.14)} \\ M''_T(t) &= M'_T(t) - kp^r q^2(-r-1)r(1-qe^t)^{-r-2}e^{2t} && \text{(B.15)} \\ M'''_T(t) &= kp^r q^3(-r-2)(-r-1)r(1-qe^t)^{-r-3}e^{3t} - 3M''(t) + 2M'(t) && \text{(B.16)} \\ & && \text{(B.17)} \end{aligned}$$

First Moment for Mean

$$E_T(Y) = kqr/p = kr(\mu/r) = k\mu \tag{B.18}$$

Second Moment for Variance

$$E(Y^2) = kr\left(\frac{\mu}{r} + \frac{\mu^2}{r}(r+1)\right) \tag{B.19}$$

Third Moment for Skewness

$$E(Y^3) = kr\left[\frac{\mu^3}{r}(r+1)(r+2) + 3\frac{\mu^2}{r}(r+1) + \frac{\mu}{r}\right] \tag{B.20}$$

Appendix C

Illustration of using IPEGLIM

C.1 Illegal Immigrants in Netherlands

```
> library(ipeglim)
> rm(list=ls())
> data(IINEE)
> iinee <- IINEE
> head(iinee)
```

```
  capture gender age nation1 nation2 nation3 nation4 nation5 reason
1       1      1   1       0       1       0       0       0      0
2       1      1   1       0       1       0       0       0      0
3       1      1   1       0       1       0       0       0      0
4       1      1   1       0       0       0       0       1      0
5       1      1   1       0       0       0       0       1      0
6       2      1   1       0       1       0       0       0      0
```

```
> ## table 2 (p.309)
> iinee$nation0 <- with(iinee, ifelse(
+   nation1==0 & nation2==0 & nation3==0 & nation4==0 & nation5==0,
+   1, 0))
> tb <- with(iinee, list(
+   xtabs(~age+capture),
+   xtabs(~gender+capture),
+   xtabs(~nation1+capture)[1,],
+   xtabs(~nation2+capture)[1,],
+   xtabs(~nation3+capture)[1,],
+   xtabs(~nation4+capture)[1,],
+   xtabs(~nation5+capture)[1,],
+   xtabs(~nation0+capture)[2,],
+   xtabs(~reason+capture)
+   ))
> tb2 <- as.table(do.call(rbind, tb))
> colnames(tb2) <- paste("f", 1:6, sep="")
> rownames(tb2) <- c(">40 years", "<40 years", "Female", "Male",
+   "Turkey", "North Africa", "Rest of Africa", "Surinam", "Asia",
+   "America, America", "Being illegal", "Other reason")
> tb2
```

```
                  f1   f2   f3   f4   f5   f6
>40 years        105    6    0    0    0    0
<40 years       1540  177   37   13    1    1
Female           366   24    6    1    1    0
Male            1279  159   31   12    0    1
Turkey          1555  180   37   13    1    1
North Africa     807   37    9    4    0    0
```

```
Rest of Africa   1416  172   34   13    1    1
Surinam          1582  182   37   13    1    1
Asia             1373  174   36   11    1    1
America, America  153   13    5    2    0    0
Being illegal    1421  154   32   12    1    1
Other reason      224   29    5    1    0    0


> ## table 4. (p.318)
> fit <- ztpreg(formula=capture ~ gender + age + nation1 + nation2
+                       + nation3 + nation4 + nation5 + reason,
+               data=iinee, dist="poisson", ztrunc=TRUE)
> tb4 <- summary(fit, HT.est=TRUE, LM.test=TRUE)
> tb4


The model is  successfully converged
Optimization method BFGS is used
Number of iterations in optimization is  45

Coefficients for zero-truncated  Poisson  model with log link

             Estimate        SE z-score  Pr(>|z|)
(Intercept) -2.317185  0.449371 -5.1565 2.516e-07 ***
gender       0.397373  0.163047  2.4372 0.0148029 *
age          0.974439  0.408204  2.3871 0.0169801 *
nation1     -1.674381  0.602882 -2.7773 0.0054814 **
nation2      0.190023  0.194003  0.9795 0.3273385
nation3     -0.911244  0.300968 -3.0277 0.0024641 **
nation4     -2.337257  1.013891 -2.3052 0.0211534 *
nation5     -1.092308  0.301634 -3.6213 0.0002931 ***
reason       0.010969  0.161527  0.0679 0.9458606
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-likelihood =  -848.4481  on  9
AIC =  1714.896

Horvitz-Thompson Estimator for N
N = 12692.28 (se = 2811.218 )
95 % CI of N = [ 7182.395 , 18202.17 ]

Lagrange multiplier Test for over-dispersion
Chi-square test-statistic =  54.98507 with df=1
Pr(chi2 >= 0.05)= 3.841459


> ## table 5. (p.319)
> fit0 <- summary(ztpreg(formula=capture ~ 1,
+   data=iinee, dist="poisson", ztrunc=TRUE), HT.est=TRUE, LM.test=TRUE)
> fit1 <- summary(ztpreg(formula=capture ~ gender,
+   data=iinee, dist="poisson", ztrunc=TRUE), HT.est=TRUE, LM.test=TRUE)
> fit2 <- summary(ztpreg(formula=capture ~ gender + age,
+   data=iinee, dist="poisson", ztrunc=TRUE), HT.est=TRUE, LM.test=TRUE)
> fit3 <- summary(ztpreg(formula=capture ~ gender + age + nation1
+   + nation2 + nation3 + nation4 + nation5,
+   data=iinee, dist="poisson", ztrunc=TRUE), HT.est=TRUE, LM.test=TRUE)
> fit4 <- summary(ztpreg(formula=capture ~ gender + age + nation1
```

```
+   + nation2 + nation3 + nation4 + nation5 + reason,
+   data=iinee, dist="poisson", ztrunc=TRUE), HT.est=TRUE, LM.test=TRUE)
> fit <- list(fit0, fit1, fit2, fit3, fit4)
> aic <- do.call(c, lapply(fit, with, aic))
> k <- do.call(c, lapply(fit, with, df)) # aic = 2*k - 2llk
> G2 <- c(NA, -(diff(aic)-2*diff(k)))
> df <- c(NA, diff(k))
> pstar <- pchisq(q=G2, df=df, lower.tail=FALSE)
> chi2 <- do.call(c, lapply(fit, with, LM.chisq))
> Nhat <- do.call(c, lapply(fit, with, N))
> cil <- do.call(c, lapply(fit, with, cil))
> ciu <- do.call(c, lapply(fit, with, ciu))
> tb5 <- cbind(aic, G2, df, pstar, chi2, Nhat, cil, ciu)
> colnames(tb5) <- c("AIC", "G2", "df", "P*", "chi2", "N", "CIL", "CIU")
> rownames(tb5) <- c("Null", "G", "G+A", "G+A+N", "G+A+N+R")
> round(tb5,3)
```

	AIC	G2	df	P*	chi2	N	CIL	CIU
Null	1805.904	NA	NA	NA	105.996	7079.926	6363.067	7796.785
G	1798.278	9.626	1	0.002	99.666	7319.302	6503.997	8134.606
G+A	1789.043	11.235	1	0.001	93.661	7807.097	6636.797	8977.397
G+A+N	1712.901	86.142	5	0.000	54.959	12687.935	7190.614	18185.256
G+A+N+R	1714.896	0.005	1	0.946	54.985	12692.281	7182.395	18202.168

```
> # table 6. (p.319)
> mu <- fit3$mu
> freq <- table(iinee$capture)
> k <- as.numeric(names(freq))
> expected <- numeric(length(k))
> for(i in k) expected[i] <- sum(dztpois(i, lambda=mu))
> obs <- c(0, freq)
> est <- c(fit3$N-sum(expected), expected)
> res <- c(NA, (freq-expected)/sqrt(expected))
> tb6 <- cbind(obs, est, res)
> colnames(tb6) <- c("Observed", "Estimated", "Residuals")
> rownames(tb6) <- seq(0,6,1)
> tb6
```

	Observed	Estimated	Residuals
0	0	1.080794e+04	NA
1	1645	1.612589e+03	0.8071111
2	183	2.337196e+02	-3.3176320
3	37	3.013454e+01	1.2506548
4	13	3.242503e+00	5.4187391
5	1	2.908711e-01	1.3148460
6	1	2.214732e-02	6.5707180

```
> # table 7. (p.320)
> m3 <- ztpreg(formula=capture ~ gender + age + nation1 + nation2
+   + nation3 + nation4 + nation5,
+   data=iinee, dist="poisson", ztrunc=TRUE)
```

```
> X <- model.matrix(obj=capture ~ gender + age + nation1 + nation2
+   + nation3 + nation4 + nation5, data=iinee)
> m3$X <- X[which(iinee$gender==1),]
> sub3.11 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$gender==0),]
> sub3.10 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$age==1),]
> sub3.21 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$age==0),]
> sub3.20 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$nation1==1),]
> sub3.31 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$nation2==1),]
> sub3.41 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$nation3==1),]
> sub3.51 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$nation4==1),]
> sub3.61 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> m3$X <- X[which(iinee$nation5==1),]
> sub3.71 <- summary(m3, HT.est=TRUE, LM.test=FALSE)
> sub <- list(sub3.11, sub3.10, sub3.21, sub3.20, sub3.31, sub3.41,
+   sub3.51, sub3.61, sub3.71)
> n <- do.call(rbind, lapply(sub, with, n))
> N <- do.call(rbind, lapply(sub, with, N))
> cil <- do.call(rbind, lapply(sub, with, cil))
> ciu <- do.call(rbind, lapply(sub, with, ciu))
> tb7 <- cbind(n, N, cil, ciu, n/N)
> colnames(tb7) <- c("Observed", "Expected", "cil", "ciu", "Rate")
> rownames(tb7) <- c("Male", "Female", "Age<40", "Age>40", "Turkey",
+   "N.Africa", "R.Africa", "Surinam", "Asia")
> tb7
```

	Observed	Expected	cil	ciu	Rate
Male	1482	8877.859	5330.01035	12425.707	0.16693214
Female	398	3810.076	1550.97527	6069.177	0.10445986
Age<40	1769	10505.584	6556.30563	14454.862	0.16838664
Age>40	111	2182.351	-34.99246	4399.694	0.05086258
Turkey	93	1740.370	-236.54401	3717.284	0.05343691
N.Africa	1023	3055.122	2660.75633	3449.488	0.33484750
R.Africa	243	2057.790	1091.28679	3024.293	0.11808786
Surinam	64	2383.743	-2273.45454	7040.940	0.02684853
Asia	284	2742.533	1458.41691	4026.648	0.10355392

```
>
```

Table C.1: Imprecise estimate of the zero-truncated Poisson regression model `capture~ gender` in illegal immigrants in The Netherlands.

		Posterior Quantiles				
	Estimate	0%	25%	50%	75%	100%
$E_n(\beta_0\|\mathbf{y})$						
`min - x3`	−1.486	−1.709	−1.531	−1.490	−1.437	−1.313
`max - x1`	−0.241	−0.399	−0.276	−0.237	−0.209	−0.086
$E_n(\beta_1\|\mathbf{y})$						
`min - x1`	−0.890	−0.399	−0.276	−0.237	−0.209	−0.086
`max - x3`	0.545	−1.709	−1.531	−1.490	−1.437	−1.313
$\hat{N}$						
`min - x4`	5046.233	4670.886	4919.754	5041.629	5164.267	5520.951
`max - x2`	7529.342	6491.433	7292.727	7501.711	7709.158	8686.917

Table C.2: Imprecise estimate of the zero-truncated Poisson regression model `capture~ gender+age` in illegal immigrants in The Netherlands.

		Posterior Quantiles				
	Estimate	0%	25%	50%	75%	100%
$E_n(\beta_0\|\mathbf{y})$						
`min - x4`	−1.692	−1.918	−1.743	−1.692	−1.634	−1.491
`max - x5`	0.429	0.249	0.375	0.430	0.472	0.632
$E_n(\beta_1\|\mathbf{y})$						
`min - x6`	−0.897	−0.443	−0.253	−0.205	−0.156	0.005
`max - x3`	0.768	−1.342	−1.179	−1.128	−1.063	−0.938
$E_n(\beta_2\|\mathbf{y})$						
`min - x7`	−1.263	−0.193	−0.057	−0.006	0.056	0.240
`max - x2`	0.711	−1.576	−1.382	−1.318	−1.253	−1.121
$\hat{N}$						
`min - x8`	5135.563	4735.787	4989.529	5121.298	5254.853	5661.656
`max - x1`	8148.298	7147.954	7872.585	8154.362	8407.660	9225.808

Table C.3: Imprecise estimate of the zero-truncated Poisson regression model `capture ~ gender+age+nationality` in illegal immigrants in The Netherlands.

		Posterior Quantiles				
	Estimate	0%	25%	50%	75%	100%
$E_n(\beta_0\|\mathbf{y})$						
`min - x1`	−2.172	−2.232	−2.232	−2.201	−2.131	−1.895
`max - x129`	0.784	0.546	0.777	0.831	0.842	0.842
$E_n(\beta_1\|\mathbf{y})$						
`min - x205`	−1.095	−0.602	−0.517	−0.474	−0.447	−0.265
`max - x48`	0.939	−1.088	−0.970	−0.970	−0.953	−0.869
$E_n(\beta_2\|\mathbf{y})$						
`min - x187`	−1.537	−0.325	−0.171	−0.171	−0.171	−0.080
`max - x81`	0.942	−1.341	−1.107	−1.093	−0.946	−0.946
$E_n(\beta_3\|\mathbf{y})$						
`min - x137`	−1.234	−0.536	−0.218	−0.218	−0.178	−0.157
`max - x104`	1.086	−1.295	−1.138	−1.131	−1.131	−0.875
$E_n(\beta_4\|\mathbf{y})$						
`min - x139`	−0.842	−0.298	−0.298	−0.298	−0.298	0.037
`max - x104`	1.388	−1.295	−1.138	−1.131	−1.131	−0.875
$E_n(\beta_5\|\mathbf{y})$						
`min - x238`	−1.278	−0.228	−0.126	−0.126	−0.126	0.028
`max - x104`	0.870	−1.295	−1.138	−1.131	−1.131	−0.875
$E_n(\beta_6\|\mathbf{y})$						
`min - x246`	−1.264	−0.434	−0.337	−0.337	−0.217	−0.217
`max - x68`	1.088	−1.206	−1.061	−1.061	−1.061	−0.930
$E_n(\beta_7\|\mathbf{y})$						
`min - x256`	−1.387	−0.152	0.007	0.041	0.155	0.235
`max - x104`	0.871	−1.295	−1.138	−1.131	−1.131	−0.875
$\hat{N}$						
`min - x210`	5154.804	4921.349	4921.349	4921.349	5538.104	5538.104
`max - x76`	15182.243	12635.749	15077.927	15077.927	15077.927	19138.944

Table C.4: Imprecise estimate of the zero-truncated Poisson regression model `capture ~ gender+age+nationality+reason` in illegal immigrants in The Netherlands.

		Posterior Quantiles				
	Estimate	0%	25%	50%	75%	100%
$E_n(\beta_0\|\mathbf{y})$						
`min - x209`	−2.093	−2.438	−2.170	−2.085	−2.085	−1.751
`max - x130`	0.789	0.639	0.750	0.779	0.822	0.916
$E_n(\beta_1\|\mathbf{y})$						
`min - x154`	−1.131	−0.662	−0.491	−0.384	−0.356	−0.206
`max - x261`	1.066	−0.984	−0.984	−0.984	−0.908	−0.908
$E_n(\beta_2\|\mathbf{y})$						
`min - x56`	−1.522	−0.431	−0.263	−0.263	−0.263	−0.220
`max - x356`	1.074	−1.081	−1.081	−1.078	−0.970	−0.940
$E_n(\beta_3\|\mathbf{y})$						
`min - x136`	−1.272	0.649	0.649	0.649	0.649	0.845
`max - x240`	1.038	−1.109	−1.093	−1.093	−1.073	−0.963
$E_n(\beta_4\|\mathbf{y})$						
`min - x506`	−0.978	−0.437	−0.218	0.004	0.021	0.021
`max - x240`	1.438	−1.109	−1.093	−1.093	−1.073	−0.963
$E_n(\beta_5\|\mathbf{y})$						
`min - x506`	−1.324	−0.437	−0.218	0.004	0.021	0.021
`max - x227`	0.946	−1.891	−1.847	−1.831	−1.795	−1.793
$E_n(\beta_6\|\mathbf{y})$						
`min - x144`	−1.318	0.274	0.588	0.588	0.588	0.609
`max - x309`	1.042	−1.028	−1.028	−1.028	−1.016	−0.988
$E_n(\beta_7\|\mathbf{y})$						
`min - x502`	−1.377	−0.332	−0.332	−0.332	−0.330	−0.175
`max - x407`	0.897	−1.158	−1.132	−1.109	−0.932	−0.869
$E_n(\beta_8\|\mathbf{y})$						
`min - x431`	−1.065	−1.502	−1.394	−1.394	−1.394	−1.236
`max - x317`	0.829	−0.603	−0.603	−0.517	−0.497	−0.464
$\hat{N}$						
`min - x22`	4981.663	4778.723	4778.723	5012.304	5136.512	5293.380
`max - x397`	16032.996	14560.128	16048.193	16048.193	16364.474	16364.474

C.2 Various Shapes of Imprecise Probabilities

Figure C.1: Four regions $\mathscr{R}_{01}$, $\mathscr{R}_{02}$, $\mathscr{R}_{03}$, $\mathscr{R}_{04}$ differently characterized for examining the linearity of an imprecise posterior expectation of the natural parameter of a standard Poisson sampling model.

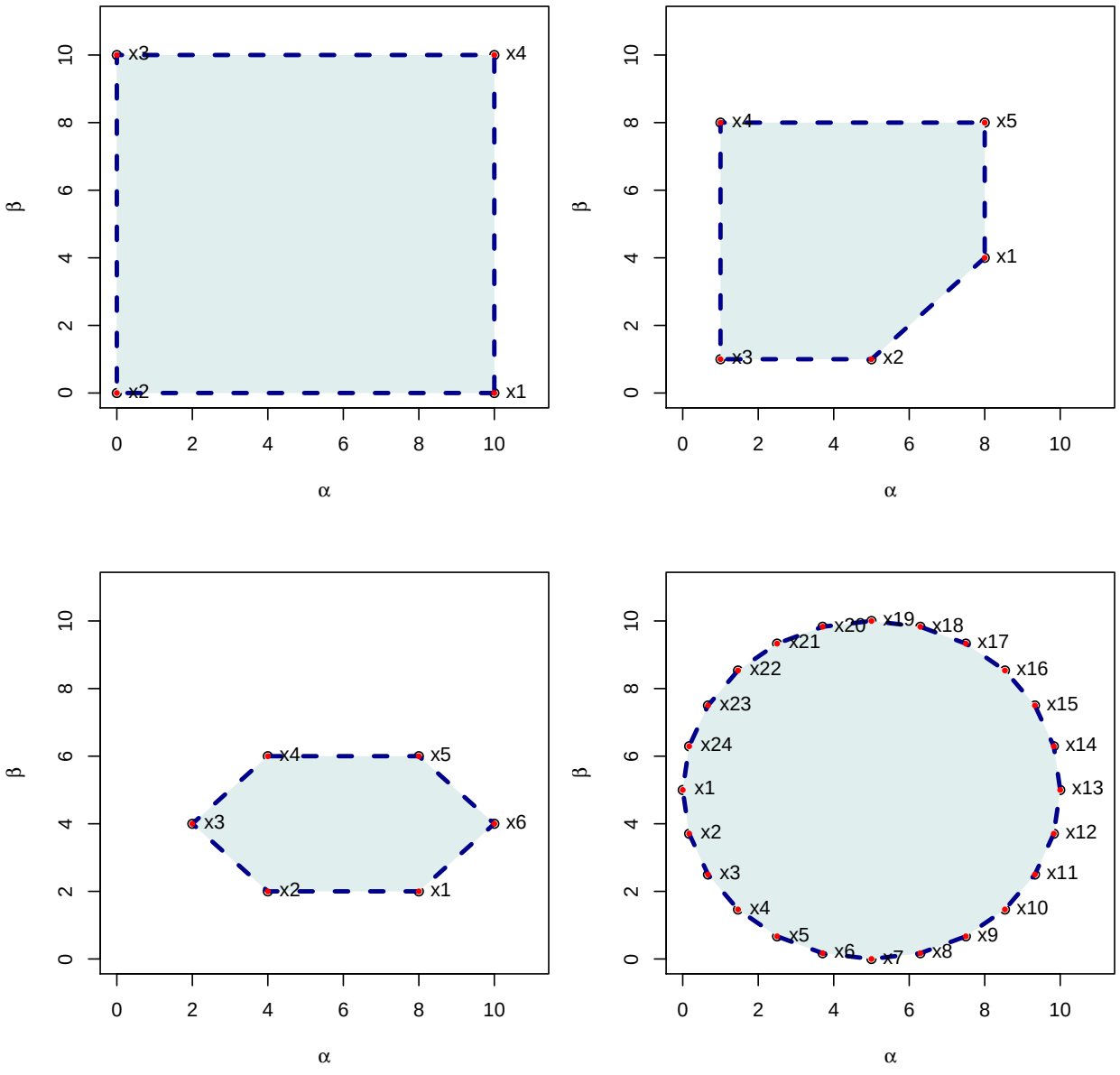
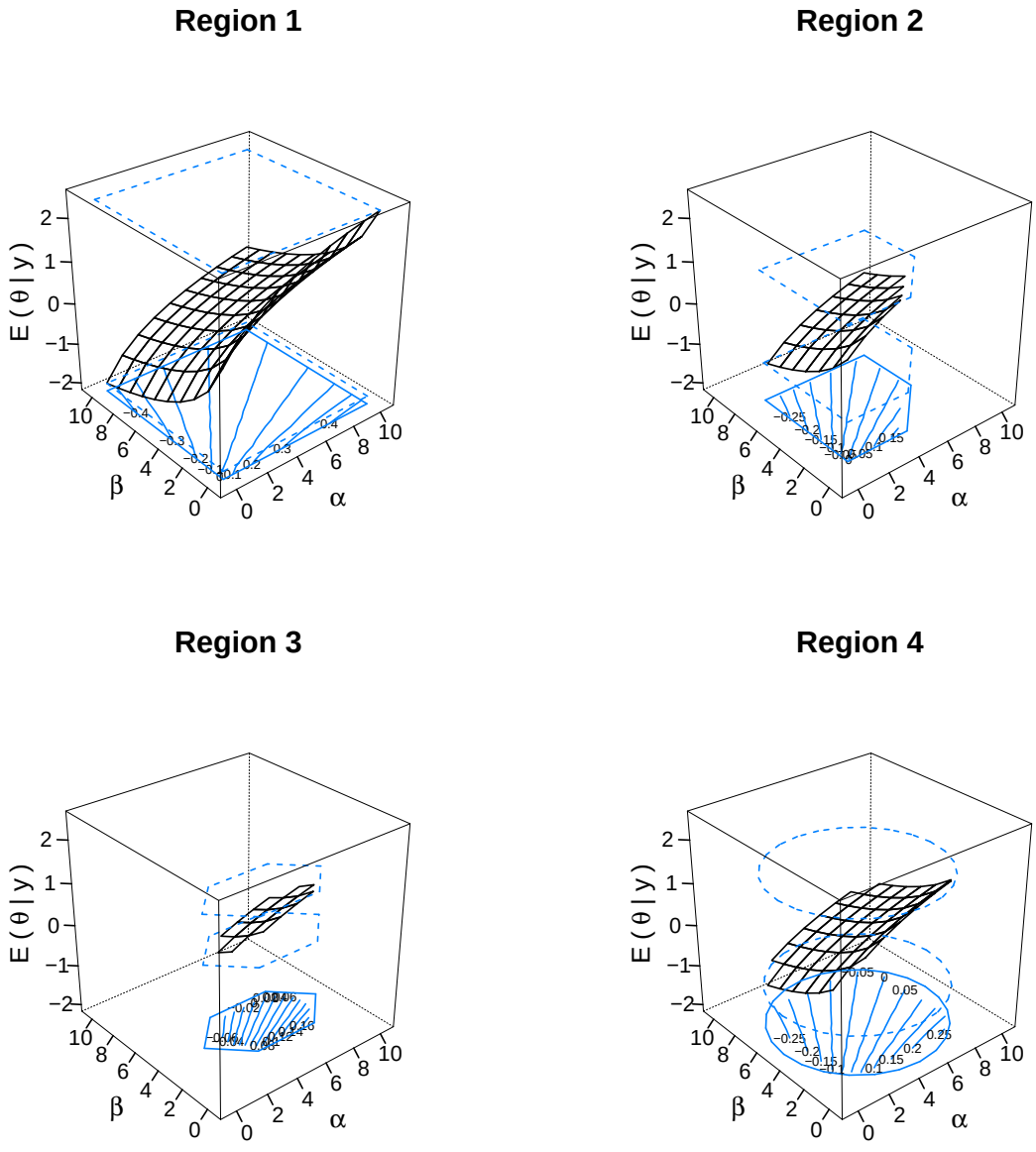


Figure C.2: The surface plot of an imprecise posterior expectation with differently characterized regions $\mathscr{R}_{01} = \{(\alpha, \beta)|0 \leq \alpha \leq 10, 0 \leq \beta \leq 10\}$, $\mathscr{R}_{02} = \{(\alpha, \beta)|1 \leq \alpha \leq 8, 1 \leq \beta \leq 8, \beta \geq \alpha - 4\}$, $\mathscr{R}_{03} = \{(\alpha, \beta)|\beta \geq -\alpha + 3, \beta \leq -\alpha + 7, \beta \geq \alpha - 3, \beta \leq \alpha + 1, \beta \leq 3, \beta \geq 1\}$, $\mathscr{R}_{04} = \{(\alpha, \beta)|(\alpha - 5)^2 + (\beta - 5)^2 \leq 5^2\}$.



Index