ABSTRACT

Bayesian Inference for Bivariate Poisson Data with Zero-Inflation

Madeline L. Drevets, Ph.D.

Chairperson: John W. Seaman, Jr.

Multivariate count data with zero-inflation is common throughout pure and applied science. Such count data often includes excess zeros. Zero-inflated Poisson regression models have been used in several applications to model bivariate count data with excess zeros. In this dissertation, we explore a Bayesian approach to bivariate Poisson models where either one or both counts is zero-inflated, with a primary focus on informative prior structures for these models. Bayesian treatments of zero-inflated Poisson models have focused on diffuse prior structures for model parameters. Nevertheless, we demonstrate that such an approach can be problematic with respect to convergence. We offer an informative prior approach, and propose methods of prior elicitation from a subject-matter expert. This includes exploration of methods for informative prior construction for an association parameter, and a multivariate distribution. We demonstrate our proposed methods within the context of a clinical example.

Bayesian Inference for Bivariate Poisson Data with Zero-Inflation

by

Madeline L. Drevets, B.S.

A Dissertation

Approved by the Department of Statistical Science

_____

Jack Tubbs, Ph.D., Chairperson

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

Approved by the Dissertation Committee

_____

John W. Seaman, Jr., Ph.D., Chairperson

_____

Steve Green, Ph.D.

_____

David Kahle, Ph.D.

_____

James Stamey, Ph.D.

_____

Jack Tubbs, Ph.D.

Accepted by the Graduate School
August 2017

_____

J. Larry Lyon, Ph.D., Dean

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

Copyright © 2017 by Madeline L. Drevets
All rights reserved

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First off, I want to thank God, for it is He that deserves all of the glory and praise. His plan for my life is bigger than ever imagined. He has always been with me, is with me and will continue to be with me.

To Dr. John Seaman, for your guidance, dedication, thoughtfulness and passion throughout this process. The detail with which you provide feedback and your dedication to helping your mentees produce the best possible product is admirable. You have helped me grow not only as a statistician, but as a person. You always made yourself available. The advice and insight that you provided me throughout the process was invaluable. I am truly blessed to have had such an incredible mentor.

To Dr. Amy Maddox, for realizing my potential in Statistics. Your persistent encouragement to pursue my studies in Statistics is the reason I am here today. Your passion and exceptional teaching ability provided me with a solid foundation for statistical principles that remain with me to this day.

To Dr. Jeanne Hill, for your mentorship as an undergraduate. You are an exceptional teacher and your passion for statistics is contagious. I greatly appreciate the countless hours spent to ensure my preparation for graduate school.

To Dr. Dean Young, for pushing me farther than what I thought I was capable. You showed me that with hard work, dedication, and God, there are no limits. I will never forget the first week. I questioned my ability to complete the program. Nevertheless, as a result of your constant encouragement, support and determination to see me succeed-I made strides every day. You were always there to offer words of reassurance when I needed it most. Thank you for providing me the resources and tools needed to succeed beyond the classroom. You are responsible for so much of my growth over the past few years.

To Dr. James Stamey, I am fortunate for the opportunity to be a part of this program and for the opportunities provided to me throughout the program. Your enthusiasm for statistics is contagious. Thank you for your guidance and for the genuine care with which you demonstrate for each of the students in the program.

To Dr. Jack Tubbs, for the program that you have established and for the many opportunities that you have provided throughout my time at Baylor. You are a great teacher. You taught me to the importance of skepticism and the importance of critical thinking.

To my parents, Wayne and Michele, for their constant support, guidance and encouragement. God has blessed me with such caring parents. I am thankful for your overflowing love and your desire to see me succeed. You have been my biggest fans throughout this process. I appreciate the sacrifices that you have made to ensure the best possible path for my success in every aspect of my life. You have always believed in my potential and it is because of you both that I have been able to reach my potential thus far. Thank you for shaping me into the person that I am today. The woman that I have become is a reflection of you both. Most importantly, you taught me that my faith in God serves the foundation for all aspects of my life. He is my Cornerstone. I love you both so much and look forward to sharing the next step in my career with you.

To my brother, Marcus, for always putting a smile on my face. You have grown in to a handsome, loving, funny, caring young man. Your laughter is contagious. It has been incredible to see your progression over the past few years. You are a constant reminder to appreciate every day that God gives us. You have been such a blessing to my life and have impacted my life more than you will ever know.

And again, most importantly to God, the Alpha and Omega. I am here to make much of You, and I am ready to embark on the next steps you have planned for my life.

DEDICATION

To my parents, Wayne and Michele, my brother, Marcus, and my faith,

"For nothing is impossible with God."

CHAPTER ONE

Introduction

The analysis of count data is essential to research throughout pure and applied science. Count data outcomes are common in medical, healthcare, environmental, and ecological studies. For example, in clinical trials, such count data may be considered as markers for predictions of clinical outcomes such as safety and/or efficacy. These counts might include the number of occurrence of adverse events in a drug safety study. In psychological and behavioral studies outcomes recording the frequencies of behavioral outcomes such as the number of occurrences of substance abuse or the number of suicide attempts. Typically such counts include a larger number of zeros than expected under standard count models, such as the Poisson or negative binomial distributions. Count models in the literature generally distinguish between two types (sources) of zeros: structural and sampling. It is assumed that the process that generates the structural zeros always yields a zero count whereas sampling zeros occur by chance and are assumed to be generated from a count distribution such as the Poisson. Thus, we can think of structural zeros as the "always zero group", and we can think of sampling zeros as the "not always zero group."

There has been increased attention to analysis of count data containing excess zeros. Common models used for handling excess zeros in count data include zero-inflated models and hurdle models. Zero-inflated models are mixture models that consist of a distribution degenerate at zero and a standard count distribution such as the Poisson. Examples of zero-inflated models include the zero-inflated Poisson (ZIP) model and the zero-inflated negative binomial (ZINB) model. Both ZIP and ZINB models assume (account for) two sources of zeros: structural and sam-

pling. The ZIP model is often used for fitting purely zero-inflated data, whereas the ZINB model is often used for fitting data that exhibits both zero-inflation and overdispersion. Hurdle models assume that the zero counts are generated from a different process than the positive counts. In particular, a hurdle model assumes that all zero counts are structural and assumes that all positive data come from a truncated count distribution (i.e. truncated Poisson or truncated negative binomial distribution). Thus, whereas a zero-inflated model assumes zero inflation due to both structural and sampling zeros, a hurdle model assumes only structural zeros.

As a result of hierarchical Bayesian modeling and the availability of software to implement methods such as MCMC, Bayesian inference for univariate and multivariate zero-inflated models has become more prevalent in the literature. Applications of Bayesian multivariate zero-inflated Poisson models in the literature include occupational health data to assess significance of intervention on the reduction in the number of musculoskeletal and non-mukculosketal injuries (Wang et al. (2003)), outpatient psychiatric use data (Neelon et al. (2010)), plant population count data (Majumdar and Gries (2010)), automobile insurance claims for three different types of claims (Bermdez and Karlis (2011)), analysis of safety crash data at intersections Dong et al. (2014), and the joint modeling of the number of blood donation and the number of blood deferral (Mohammadi et al. (2016)). All of these depend on diffuse priors for model parameters. Hence, despite the increased prevalence of Bayesian inference for bivariate zero-inflated Poisson models, use of informative prior structures applied to these models has not been explored in detail.

The focus of this dissertation is Bayesian inference for bivariate Poisson data with zero-inflation. In Chapter Two we introduce a bivariate partial zero-inflated Poisson model, which assumes that one outcome is zero-inflated and one outcome is not zero-inflated. This model has not been covered in the literature. We discuss diffuse priors for this model and demonstrate situations in which use of diffuse priors

can be problematic with respect to convergence. We discuss informative prior structures, with a primary focus on methods of prior elicitation from a subject matter expert within the context of a contextual adverse event drug safety trial. In Chapter Three, we present a conditional representation of the bivariate partial zero-inflated Poisson model and demonstrate how this alternate representation provides a plausible route of prior specification for the association parameter of our bivariate model. In Chapter Four, we consider Bayesian inference for a bivariate zero-inflated Poisson model, which assume both outcomes are zero-inflated. In this chapter, we extend the ideas proposed in Chapters Two and Three to the bivariate zero-inflated Poisson model. We discuss method for prior elicitation from a subject-matter expert for the multinomial zero-inflation parameters, which is often a difficult task. We apply our methods within the context of a study to investigate the safety and efficacy of a hypothetical new drug. Throughout, we demonstrate potential uses of both the prior predictive and posterior predictive distributions in the analysis of our hypothetical clinical studies. These predictive distributions have not been covered in detail in the literature for bivariate Poisson models with zero-inflation.

CHAPTER TWO

A Bayesian Bivariate Partial Zero-Inflated Poisson Model

2.1 Introduction to Zero-Inflated Models

The analysis of count data is essential to research throughout pure and applied science. Such count data often includes excess zeros or overdispersion. Often the number of zeros in a sample is underestimated standard count models. "Excess zeros" refers to additional zeros that are present in data that cannot be accommodated by standard count models. For example, a Poisson model assumes that the conditional variance of the dependent variable is equal to the conditional mean. Overdispersion in the analysis of discrete data occurs when the variability in the data is larger than expected under the assumed standard count distribution. If overdispersion is not taken into account in the modeling process, it can lead to underestimated variance and incorrect inference.

There has been increased attention to analysis of count data containing excess zeros, and the analysis of such data is a primary focus of this dissertation. In clinical trials, such count data may be considered as markers for predictions of clinical outcomes such as safety and/or efficacy. For example, we might be interested in the number of occurrence of adverse events in a drug safety study. Count data outcomes are also common in prevention or intervention trials. For example, the frequency of substance abuse or risky behaviors in psychosocial and behavioral studies. Typically such counts include a larger number of zeros than expected under standard count models, such as the Poisson or negative binomial distributions. Count models in the literature generally distinguish between two types (sources) of zeros: structural and sampling. It is assumed that the process that generates the structural zeros always yields a zero count whereas sampling zeros occur by chance and are assumed

to be generated from a count distribution such as the Poisson. Thus, we can think of structural zeros as the "always zero group", and we can think of sampling zeros as the "not always zero group." It is the structural zeros that are the "excess zeros", whereas sampling zeros are due to sampling variability. For example, suppose we are modeling the number of children born to a woman. Some women might not have children because they (or their partner) are physically unable to bear children (structural zero) whereas some women might be physically able to bear children, but do not have children, (sampling zero). Neelon et al. (2010) distinguishes between structural and sampling zeros by referring to structural zeros as the "not-at-risk" group and sampling zero as the "at-risk" group. We adopt this interpretation throughout the dissertation.

Common models used for handeling excess zeros in count data include zero-inflated models and hurdle models. Zero-inflated models are mixture models that consist of a distribution degenerate at zero and a standard count distribution such as the Poisson. Examples of zero-inflated models include the zero-inflated Poisson (ZIP) model and the zero-inflated negative binomial (ZINB) model. Both ZIP and ZINB models assume (account for) two sources of zeros: structural and sampling. The ZIP model is often used for fitting purely zero-inflated data, whereas the ZINB model is often used for fitting data that exhibits both zero-inflation and overdispersion. Thus, they consist of two latent classes of observations; that is, whether or not a zero is a structural or sampling zero is unobserved. Hurdle models assume that the zero counts are generated from a different process than the positive counts. In particular, a hurdle model assumes that all zero counts are structural and assumes that all positive data come from a truncated count distribution (i.e. truncated Poisson or truncated negative binomial distribution). Thus, whereas a zero-inflated model assumes zero inflation due to both structural and sampling zeros, a hurdle model assumes only structural zeros. In practice, it is this distinguishing feature

that often guides the appropriateness of zero-inflated versus hurdle models in data analysis (Rose et al. (2006)). The organization of this chapter is as follows: In Section 2.1.1 we introduce common count models in the literature for zero-inflated data, with a particular focus on the univariate zero-inflated Poisson distribution. In Section 2.2 we introduce a bivariate partial zero-inflated Poisson distribution. We give a Bayesian development which includes discussion of diffuse priors and posterior inference for this model. In Section 2.3 we introduce a drug safety adverse event study that will serve as a vehicle for which we apply our proposed methods throughout this chapter. In Section 2.4 we demonstrate potential problems that can arise in posterior inference for our Bayesian bivariate partial zero-inflated Poisson model which leads to a discssion of nonidentified models in Section 2.5. In Section 2.6 we propose methods for informative prior construction for a Bayesian partial zero-inflated Poisson model for both the non-regression and regression case. The implementation of our propose methods of prior construction is illustrated within the context of our clinical example in Section 2.7. Concluding comments are given in Section 2.8.

2.1.1 Univariate Zero-Inflated Poisson Model

The literature is rich with analysis and applications of univariate ZIP models (McCullagh and Nelder (1989), Lambert (1992), Johnson et al. (1997), Dagne (2004), Ntzoufras and Karlis (2005),Rose et al. (2006), Baughman (2007), He et al. (2014)). A univariate zero-inflated Poisson (ZIP) distribution (Cohen (1963)) is constructed as a mixture of a Poisson distribution and a distribution that is degenerate at zero. Let Y denote a random variable with probability mass function given by

$$\Pr(Y = y) = \begin{cases} p + (1-p)e^{-\lambda} & \text{for } y = 0 \\ (1-p)\frac{e^{-\lambda}\lambda^{y}}{y!} & \text{for } y = 1, 2, \ldots \end{cases} \tag{2.1}$$

where $0 \leq p \leq 1$ is the probability of excess zeros (or zero-inflation probability), and λ is the mean of the Poisson distribution. We write $Y \sim \text{ZIP}(p, \lambda)$. In (2.1) the model assumes two sources of zeros: a structural zero is observed with probability p and a sampling zero is observed with probability $1 - p$. In a ZIP model, it is often of interest to estimate the proportion of structural zeros, p, and to estimate the Poisson rate, in the "not always 0" group (Dagne (2004)).

The ZIP model in (2.1) represents sampling from a mixture of two subpopulations: one subpopulation is considered "not at-risk" and yields a response count of zero with probability one (structural zero), the other subpopulation is considered "at-risk" and the responses follow a Poisson distribution. Sampling zeros may arise from the "at-risk" population with probability $\exp(-\lambda)$. However, if the response is zero, the membership (structural or sampling) is unobserved.

It can be shown that the moments of the univariate zero-inflated Poisson model are

$$\text{E}(Y) = (1 - p)\lambda$$

and

$$\text{Var}(Y) = (1 - p)\lambda + p\,(1 - p)\,\lambda^2.$$

Note that the mean of a ZIP model is smaller than that of a standard Poisson, since $0 < p < 1$. Furthermore, $\text{Var}(Y) > \text{E}(Y)$.

2.1.2 Univariate Zero-Inflated Regression Model

Consider $Y_i \overset{ind}{\sim} \text{ZIP}(p_i, \lambda_i)$, $i = 1, \ldots, n$, where n denotes the number of observations. We can incorporate covariates into the model through canonical link functions, in generalized linear models (Lambert (1992)). In particular, for the ith observation, we use a logistic regression model for the zero-inflation probability,

$$\text{logit}(p_i) = \mathbf{x}_i'\boldsymbol{\beta},$$

where $\mathbf{x}$ and $\boldsymbol{\beta}$ are vectors of covariates and corresponding coefficients, respectively. We use a log-linear model for the mean rate of the Poisson state,

$$\log(\lambda_i) = \mathbf{w}_i'\boldsymbol{\gamma},$$

where $\mathbf{w}$ and $\boldsymbol{\gamma}$ are vectors of covariates and corresponding coefficients, respectively. This representation allows the Poisson rate and zero-inflation probability to depend on the same or different covariates. In the literature, the zero-inflation probability, p, is often assumed constant across observations (see Ntzoufras and Karlis (2005), Liu and Tian (2015)).

2.1.3 Bayesian Univariate Zero-Inflated Poisson Models

Bayesian treatments of the univariate ZIP model have been discussed in the literature. For example, Ghosh et al. (2006) developed Bayesian ZIP models for cross-sectional data, using Markov chain Monte Carlo (MCMC) methods with data augmentation to obtain posterior samples. Bayesian inference for zero-inflated Poisson models found in the literature rely on relatively non-informative prior structures. The univariate case is not the focus of this dissertation and thus for further discussion of applications and treatments of this model we refer the reader to Ntzoufras and Karlis (2005), Ghosh et al. (2006), Dagne (2004) and references therein.

2.2 Bivariate Partial Zero-Inflated Poisson Model

In this section we extend the univariate ZIP to a bivariate count response vector. We introduce a bivariate partial ZIP model, which assumes one outcome is zero-inflated and the other outcome is not zero-inflated. This case has not been covered in detail in the literature, but is worth exploring as this situation is commonly encountered in medical applications.[1]

[1] Personal acknowledgement to Dr. Ding-Geng (Din) Chen, Clinical Professor in the Department of Biostatistics at the University of North Carolina Gillings School of Global Health, for suggestion to explore this model.

Suppose we have two Poisson counts, denoted Y_1 and Y_2, such that (Y_1, Y_2) denotes a bivariate count vector. Further, suppose that Y_1 is zero-inflated and Y_2 is not zero-inflated. That is, we assume two sources of zeros for Y_1, structural and sampling, and assume only sampling zeros for Y_2. We call this a bivariate partial zero-inflated Poisson (BPZIP). It can be constructed from a mixture of a degenerate distribution at 0 and a univariate Poisson distribution with parameter μ_2, and a bivariate Poisson (BP) distribution (see Appendix A for details of this distribution) with parameters $(\lambda_0, \lambda_1, \lambda_2)$ as follows:

$$(Y_1, Y_2) \sim \begin{cases} (0, \text{Poisson}(\mu_2)), & \text{with probability } p \\ \text{BP}(\lambda_0, \lambda_1, \lambda_2), & \text{with probability } 1-p, \end{cases} \tag{2.2}$$

where $\lambda_k > 0$, $k = 0, 1, 2$, $\mu_2 = \lambda_2 + \lambda_0$, and $0 < p < 1$ denotes the zero-inflation (or excess proportion/probability of zeros than expected under a standard bivariate Poisson count distribution) parameter such that $p + (1-p) = 1$. In particular, p represents the excess proportion of zeros for Y_1 as this model assumes that the zero-inflation is attributed to only Y_1. This is in contrast to a bivariate zero-inflated Poisson model which assumes both bivariate counts, Y_1 and Y_2, contribute the observed excess zeros. We consider the bivariate ZIP model Chapter Four.

The BPZIP model in (2.2) consists of a zero-inflated Poisson and an ordinary bivariate Poisson. For the bivariate Poisson component we use the trivariate reduction representation introduced by Johnson et al. (1997). This representation is most commonly used in the literature (we refer the reader to Appendix A for more details on the bivariate Poisson distribution). The trivariate reduction representation of the BP distribution assumes that the bivariate responses (Y_1, Y_2) are positively correlated. Thus, we propose the bivariate partial zero-inflated Poisson model in (2.2) as a possible model to fit positively correlated bivariate count Poisson data where one count is zero-inflated and the other count is not zero-inflated.

Let $(Y_1, Y_2) \sim \text{BPZIP}(p, \boldsymbol{\lambda})$ where $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$. Then, for $y_1, y_2 \in \{0, 1, 2, \ldots\}$, the bivariate joint probability mass function, $f_{Y_1,Y_2}(y_1, y_2 \,|\, p, \boldsymbol{\lambda}) \equiv f_{Y_1,Y_2}(y_1, y_2)$, is given by

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} p\, e^{-(\lambda_2+\lambda_0)} + (1-p)\, e^{-(\lambda_1+\lambda_2+\lambda_0)}, & y_1 = 0, y_2 = 0, \\ (1-p)\, e^{-(\lambda_1+\lambda_2+\lambda_0)} \frac{\lambda_1^{y_1}}{y_1!}, & y_1 \neq 0, y_2 = 0, \\ p\, \frac{(\lambda_2+\lambda_0)^{y_2}\, e^{-(\lambda_2+\lambda_0)}}{y_2!} + (1-p)\, \frac{\lambda_2^{y_2}}{y_2!}\, e^{-(\lambda_1+\lambda_2+\lambda_0)}, & y_1 = 0, y_2 \neq 0, \\ (1-p)\, e^{-(\lambda_1+\lambda_2+\lambda_0)} \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_2^{y_2}}{y_2!} \sum_{m=0}^{\min(y_1,y_2)} \binom{y_1}{m}\binom{y_2}{m} m! \left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^m, & y_1 \neq 0, y_2 \neq 0. \end{cases} \tag{2.3}$$

In the applications we consider in this dissertation we typically take $f_{(Y_1,Y_2)}(0,0)$ and $f_{(Y_1,Y_2)}(y_1, 0)$, $y_1 \neq 0$, to be small (compared to the bivariate ZIP model), particularly when the rate of Y_2 is small (i.e. close to zero).

Let $\mathbf{y}_i = (y_{1i}, y_{2i})$, $i = 1, \ldots, n$, denote the observed bivariate outcomes. The corresponding likelihood function is

$$\begin{aligned} \ell(p_i, \boldsymbol{\lambda}_i \,|\, \mathbf{y}_i) = &\prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = 0 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i1}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = 0 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i2}} \\ &\times \prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = y_2 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i3}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = y_2 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i4}}, \end{aligned} \tag{2.4}$$

where $\boldsymbol{\lambda}_i = (\lambda_{0i}, \lambda_{1i}, \lambda_{2i})$ and I_{ik} is an indicator function defined as

$$\begin{aligned} I_{i1} = 1: &\text{ if } y_{1i} = 0, y_{2i} = 0, \\ I_{i2} = 1: &\text{ if } y_{1i} \neq 0, y_{2i} = 0, \\ I_{i3} = 1: &\text{ if } y_{1i} = 0, y_{2i} \neq 0, \end{aligned} \tag{2.5}$$

and

$$I_{i4} = 1: \text{ if } y_{1i} \neq 0, y_{2i} \neq 0.$$

It can be shown that the marginal distribution of Y_1 is a univariate zero-inflated Poisson distribution, denoted $Y_1 \sim \text{ZIP}(p, \mu_1)$, with corresponding marginal probability mass function

$$f_{Y_1}(y_1 \,|\, p, \mu_1) \equiv \Pr(Y_1 = y_1) = \begin{cases} p + (1-p)\ e^{-\mu_1}, & y_1 = 0 \\ (1-p)\frac{\mu_1^{y_1}\, e^{-\mu_1}}{y_1!}, & y_1 = 1, 2, 3 \ldots \end{cases} \tag{2.6}$$

where p denotes the zero-inflation probability for Y_1, $(p+(1-p) = 1)$ and $\mu_1 = \lambda_1 + \lambda_0$. Thus, Y_1 can be a structural zero with probability p and a sampling zero with probability $1-p$. The marginal rate associated with count Y_1 among those "at-risk" for Y_1 is μ_1. It follows that the mean and variance of Y_1 are given by

$$\text{E}(Y_1) = (1-p)\mu_1, \tag{2.7}$$

and

$$\text{Var}(Y_1) = (1-p)\mu_1 \left[1 + p\,\mu_1\right], \tag{2.8}$$

respectively.

Similarly, it can be shown that the marginal distribution of Y_2 is a Poisson distribution with mean, μ_2, denoted $Y_2 \sim \text{Poisson}(\mu_2)$. The marginal probability mass function is given by

$$f_{Y_2}(y_2 \,|\mu_2) \equiv \Pr(Y_2 = y_2) = \frac{\mu_2^{y_2} e^{-\mu_2}}{y_2!}, \quad y_2 = 0, 1, 2, \tag{2.9}$$

where $\mu_2 = \lambda_2 + \lambda_0$. It follows that $\text{E}(Y_2) = \text{Var}(Y_2) = \mu_2$.

2.2.1 Bivariate Partial Zero-inflated Poisson Regression Model

Let $(y_{1i}, y_{2i}) \sim \text{BPZIP}(p_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i})$, for $i = 1, \ldots, n$ where n denotes the number of observations. We can represent the BPZIP model in such a way that the Poisson parameters, λ_k, and zero-inflation parameter, p, depend on covariates through canonical link, generalized linear models. In particular, we express the Poisson parameters, λ_k as function of covariates via the logarithmic link and the

zero-inflation parameter is related to covariates via the logit link. That is, for the ith individual we have that

$$\log(\lambda_{k,i}) = \mathbf{x}_{k,i}^T \boldsymbol{\gamma}_k, \tag{2.10}$$

for $k = 0, 1, 2$, and

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{w}_i^T \boldsymbol{\beta}, \tag{2.11}$$

where $\mathbf{x}_{k,i}$ and $\boldsymbol{\gamma}_k$ are vectors of covariates and corresponding regresssion coefficients, respectively, associated with λ_k, and $\mathbf{w}_i$ and $\boldsymbol{\beta}$ are vectors of covariates and corresponding regresssion coefficients, respectively, associated with the zero-inflation parameter. Let q and r denote the number of covariates corresponding to λ_k and p, respectively. Specifically, $\boldsymbol{\gamma}_k$ is a $(q+1)$ vector of regression coefficients and $\boldsymbol{\beta}$ is a $(r+1)$ vector of regression coefficents. This parameterization allows the same or different explanatory variables to affect the Poisson rates and zero-inflation probability. In addition, this representation allows the Poisson rates to depend on different explanatory variables, which extends the use of this model to a wide range of applications (Mohammadi et al. (2016)).

Let $\mathbf{y}_i = (y_{1i}, y_{2i})$, $i = 1, \ldots, n$. The BPZIP regression likelihood function is given by

$$\begin{aligned}\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}_k | \mathbf{y}_i) = &\prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = 0 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i1}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = 0 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i2}} \\ &\times \prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = y_2 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i3}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = y_2 \,|\, p, \boldsymbol{\lambda})\right]^{I_{i4}},\end{aligned} \tag{2.12}$$

where

$$p_i = \frac{\exp(\mathbf{w}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{w}_i^T \boldsymbol{\beta})}$$

and for $k = 0, 1, 2$,

$$\lambda_{k,i} = \exp(\mathbf{x}_{k,i}^T \boldsymbol{\gamma}_k),$$

and I_{it}, $t = 1, \ldots, 4$, is an indicator function defined in (2.5).

2.2.2 The Bayesian Partial Bivariate Zero-Inflated Poisson Model

Bayesian analysis of models of other zero-inflated models have appeared in the literature. For example, Ghosh et al. (2006) relies on diffuse priors for inference of a univariate zero-inflated Poisson model applied to manufacturing defects, Majumdar and Gries (2010) and Mohammadi et al. (2016) rely on diffuse priors for inference of a bivariate zero-inflated Poisson model applied to ecological outcomes and healthcare outcomes, respectively. In this section, we discuss a Bayesian approach to our partial bivariate zero-inflated Poisson model, which has not been addressed in the literature.

2.2.3 Diffuse Prior Structure: Bayesian BPZIP Non-regression Model

We first consider the BPZIP model in the absence of covariates as outlined in Section 2.2. Consistent with the common approach in the literature on Bayesian inference for univiariate and bivariate zero-inflated Poisson (BZIP) models (i.e. Ghosh et al. (2006) and Majumdar and Gries (2010)), we assume that prior distributions for the zero-inflation parameter, p, and Poisson parameters, $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$ are independent. We will use the following conditional conjugate priors:

$$p \sim \text{Beta}(a, b),$$

and for $k = 0, 1, 2$,

$$\lambda_k \sim \text{Gamma}(c_k, d_k),$$

where a, b, c_k, and d_k are considered hyperparameters. In the absence of prior information regarding the zero-inflation probability, uniform ($a = b = 1$) or Jeffreys ($a = b = 0.5$) are commonly used. The latter can be problematic, however, particularly for values of p close to zero or one. This is further discussed in Section 2.4.1. In the absence of prior information regarding the Poisson parameters, small values of c_k and d_k for the prior on the λ_k's result in a diffuse prior for λ_k with large variance. For example, $c_k = 1$ and $d_k = 0.01$ for $k = 0, 1, 2$, where c_k and d_k are the shape and rate, respectively. However, as we shall demonstrate, this prior distribution on

λ_k can be problematic within the context of a bivariate partial zero-inflated Poisson model.

2.2.4 Diffuse Prior Structure: Bayesian BPZIP Regression Model

Consider the BPZIP regression model outlined in Section 2.2.1, where p and $\boldsymbol{\lambda}$ both depend on covariates. As in the non-regression case, we assume that the prior distributions for all parameters are independent. Let $\boldsymbol{\gamma}_k = (\gamma_{k,0}, \ldots, \gamma_{k,q})$ be a $1 \times (q+1)$ vector of regression parameters consisting of an intercept and q covariates corresponding to λ_k, $k = 0, 1, 2$. Further, let $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_r)$ be a $1 \times (r+1)$ vector of regression parameters consisting of an intercept and r covariates corresponding to the zero-inflation probability. We assume the elements of $\boldsymbol{\gamma}_k$ are mutually independent and independent of the elements of $\boldsymbol{\beta}$. Thus, the joint prior distribution is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}_k) = \prod_{k=0}^{2} \prod_{v=0}^{q} \pi(\gamma_{k,v}) \prod_{v=0}^{r} \pi(\beta_v).$$

For example, for one covariate, the joint prior distribution is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}_k) = \pi(\beta_0)\pi(\beta_1) \prod_{k=0}^{2} \pi(\gamma_{k,0})\pi(\gamma_{k,1}). \tag{2.13}$$

A typical prior structure places independent diffuse normal priors on both sets of regression coefficients. That is,

$$\boldsymbol{\beta} \sim \mathrm{N}_r(\mathbf{0}, \boldsymbol{\sigma}^2_{\boldsymbol{\beta}} \mathbf{I}_r)$$

and

$$\boldsymbol{\gamma}_k \sim \mathrm{N}_q(\mathbf{0}, \boldsymbol{\sigma}^2_{\boldsymbol{\gamma}} \mathbf{I}_q),$$

for $k = 0, 1, 2$, and where $\sigma^2_{\boldsymbol{\gamma}}$ and $\sigma^2_{\boldsymbol{\beta}}$ are chosen to be large (e.g. 10^3) to express absence of prior information. Here $N_d(\boldsymbol{\varphi}, \boldsymbol{\Lambda})$ denotes a d-variate normal distribution with mean vector $\boldsymbol{\varphi}$ and covariance matrix, $\boldsymbol{\Lambda}$. Alternatively, we can place inverse-gamma priors on $\sigma^2_{\boldsymbol{\gamma}}$ and $\sigma^2_{\boldsymbol{\beta}}$, or proper uniform prior distributions on $\sigma_{\boldsymbol{\gamma}}$ and $\sigma_{\boldsymbol{\beta}}$.

2.2.5 Posterior Inference for BPZIP Model

Suppose we have n bivariate outcomes where the ith observation is represented by $\mathbf{y}_i = (y_{1i}, y_{2i})$ for $i = 1, \ldots, n$ and $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ represents the vector of the observed bivariate responses. Consider the non-regression case where $\ell(p, \boldsymbol{\lambda} \,|\, \mathbf{y})$ denotes the likelihood given by (2.4). Again, we assume that the prior distributions for all parameters are independent. Let $\pi(p)$ denote the prior distribution for the zero-inflation parameter, and let $\pi(\lambda_k)$ denote the prior for λ_k, $k = 0, 1, 2$. Denote the joint prior

$$\pi(\boldsymbol{\lambda}) = \pi(\lambda_0)\pi(\lambda_1)\pi(\lambda_2).$$

Then, the joint posterior distribution of $(p, \boldsymbol{\lambda})$ is given by

$$\pi(p, \boldsymbol{\lambda}|\mathbf{y}) \propto \ell(p, \boldsymbol{\lambda} \,|\, \mathbf{y})\, \pi(p)\, \pi(\boldsymbol{\lambda}),$$

which has a nonstandard density. Thus, Markov chain Monte Carlo (MCMC) methods, such as Gibbs sampling, are used to sample from the posterior distribution. Gibbs sampling draws iteratively from the full conditional distributions of the model parameters.[2]

We adapt a data augmentation method similar to that applied in Ghosh et al. (2006) and Majumdar and Gries (2010).[3] This method eases implementation of the Gibbs sampler to generate samples from the posterior distribution of parameters of interest. Inference for the BPZIP model relies on representing Y_1 and Y_2 in terms of latent variables. From the model in (2.2) we have that (Y_1, Y_2) consists of two underlying sub-populations;

---

[2] The full conditionals for univariate zero-inflated models and hurdle models also do not have closed forms (Neelon et al. (2010)).

[3] Tanner and Wong (1987) propose data augmentation schemes to ease the computation of posterior computations in models such that the data can be augmented in such a way that eases sampling from the posterior distribution. This is commonly applied to missing value problems.

(1) Subjects not at-risk for Y_1 and at-risk for Y_2

(2) Subjects at-risk for Y_1 and Y_2,

where "not at-risk" represents the case where a structural zero is always observed, and "at-risk" represents the case where either a non-zero count or sampling zero is observed. We do not actually observe which sub-population each bivariate observation is from. Furthermore, from the standard BP distribution described in Appendix A we have that

$$Y_1 = X_1 + X_0$$

and

$$Y_2 = X_2 + X_0,$$

where X_1, X_2, and X_0 are independent Poisson random variables with means λ_1, λ_2, and λ_0, respectively, such that $\lambda_i > 0$, $i = 0, 1, 2$. The observed data are the counts Y_1 and Y_2 and the unobserved data are the counts X_1, X_2 and X_0. It follows that we can represent the BPZIP random variables in terms of latent (unobserved) variables. Define the random variables Y_1 and Y_2 in terms of latent variables U and $\mathbf{X} = (X_1, X_2, X_0)$ such that

$$Y_1 = (1 - U_i)(X_1 + X_0), \tag{2.14}$$

and

$$Y_2 = X_2 + X_0, \tag{2.15}$$

where $U_i \sim \text{Bernoulli}(p)$, and p is the zero-inflation probability for Y_1. That is,

$$U_i = \begin{cases} 1, & \text{if } y_{1i} = 0 \\ 0, & \text{if } y_{1i} > 0. \end{cases}$$

Further, the bivariate Poisson latent variables X_1, X_2, and X_0 are independent Poisson random variables with means λ_1, λ_2, and λ_0, respectively. Note that, per (2.14)

and (2.15), Y_1 depends on three latent variables where as Y_2 depends on two latent variables. Denote the model parameters by $\boldsymbol{\theta} = (p, \lambda_1, \lambda_2, \lambda_0)$, the observed data by $\mathbf{Y} = (Y_1, Y_2)$ and the latent (unobserved) data by $\mathbf{Z} = (U, X_1, X_2, X_0)$. We are interested in the posterior distribution, $\pi(\boldsymbol{\theta}|\mathbf{Y})$, however, this is difficult to compute directly. To ease computation, we obtain the posterior, $\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Z})$ (often referred to as the augmented data posterior), which is more straightforward to compute. Instead of sampling directly from the posterior, $\pi(\boldsymbol{\theta} \,|\, \mathbf{Y})$, we sample from the posterior, $\pi(\boldsymbol{\theta}, \mathbf{Z} \,|\, \mathbf{Y})$. In order to implement this method within a Gibbs sampling framework, we must be able to sample from two conditional distributions, namely the posterior distribution of augmented data, $\pi(\boldsymbol{\theta} \,|\, \mathbf{Y}, \mathbf{Z})$, and $\pi(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y})$.

This can be extended to the regression model. The joint posterior distribution is given by

$$\pi(U, \mathbf{X}, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_k, \boldsymbol{\sigma}^2_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}^2_{\boldsymbol{\beta}} \,|\, \mathbf{Y}) \propto \ell(U, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\gamma}_k|\mathbf{Y}) \prod_{k=0}^{2} \left\{ \prod_{v=0}^{q} \pi(\boldsymbol{\gamma}_{k,v}) \right\} \prod_{v=0}^{r} \pi(\boldsymbol{\beta}_v),$$

which again has no closed form. Thus, for posterior inference we use MCMC methods, such as Gibbs sampling, to sample from the full conditionals, namely, $\pi(\boldsymbol{\theta} \,|\, \mathbf{Y}, \mathbf{Z})$ and $\pi(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y})$, where here $\boldsymbol{\theta} = (\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\beta})$. This algorithm can be readily implemented in software such as WinBUGS, OpenBUGS or JAGS (for both the non-regression case and regression case), which use MCMC algorithms to generate samples from the posterior distribution of parameters (Plummer (2003), Sturtz et al. (2005)). For this dissertation, inference for the BPZIP model was carried out using JAGS through the R package `rjags`.

2.2.6 BPZIP Prior and Posterior Predictive Distributions

The prior and posterior predictive distributions are commonly used in the implementation of Bayesian analysis for prediction. By using both of these joint distributions, we can model our uncertainty completely. We make use of both the prior and posterior predictive distributions in Section 2.7.1 and Section 2.7.4, respectively.

The prior predictive distribution is the expected value of the likelihood with respect to the prior. Often we are interested in predicting the "next" observation or observations (e.g. gauging the prospects for a future sample). That is, suppose we want to predict future bivariate observation(s) denoted $\tilde{\mathbf{y}}$ (assumed independent from our data $\mathbf{y}$). We can make use of the posterior predictive distribution, which is defined as the expected value of the BPZIP likelihood, evaluated at $\tilde{\mathbf{y}}$, with respect to the BPZIP posterior distribution given the data. The posterior predictive distribution for the BPZIP model has the form

$$\tilde{\pi}(\tilde{\mathbf{y}}|\mathbf{y}, \tilde{\mathbf{x}}) = \int_{\Theta} \ell(\tilde{\mathbf{y}} \,|\, \boldsymbol{\beta}, \boldsymbol{\gamma}_k, \mathbf{y}, \mathbf{x}) \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}_k \,|\, \mathbf{y}, \mathbf{x}) d\boldsymbol{\theta}, \tag{2.16}$$

where $\mathbf{x}$ is the vector of covariates for the current sample, and $\tilde{\mathbf{x}}$ is the vector of covariates corresponding to the future observations, and $\boldsymbol{\theta}$ is the vector of parameters defined on Θ.

2.3 Application: Drug Safety Adverse Event Study

We consider a hypothetical study to investigate the safety of a new drug. Suppose we want to study the safety of a new drug with respect to two side effects, common to similar medications currently on the market. Specifically, suppose 100 subjects enroll in a six-month study to track the two adverse events. Throughout the course of the six month study, the subjects were asked to recall the number of occurrences of these two adverse events, which are known to be related. We assume that observed person-time is the same for each individual subject. In addition, we assume that not all subjects involved in this particular study are at-risk to experience a migraine, but that all subjects are at-risk to experience a nausea episode. Moreover, previous studies suggest positive association between the number of migraines and the number of nausea episodes for those at-risk to the former. Thus, a BPZIP model is appropriate in this scenario.

Let (Y_{1i}, Y_{2i}) be a bivariate response count for the i^{th} subject such that

$$Y_{1i} = \text{Number of migraines experienced during study,}$$

and

$$Y_{2i} = \text{Number of nausea episodes experienced during study,}$$

for $i = 1, \ldots, n$, where n is the number of subjects in the study. We assume $(Y_1, Y_2) \sim \text{BPZIP}(p, \boldsymbol{\lambda})$. From (2.2) we have that the data can arise from one of two distributions:

(1) $(Y_{1i}, Y_{2i}) \sim (0, \text{Poisson}(\mu_2))$: the subject is not at-risk to experience a migraine but is at-risk to experience a nausea episode (with probability p).

(2) $(Y_{1i}, Y_{2i}) \sim \text{BP}(\lambda_0, \lambda_1, \lambda_2)$: The subject is at-risk to experience a migraine and is at-risk to experience a nausea episode (with probability $1 - p$).

Table 2.1 contains the interpretations of the bivariate partial zero-inflated Poisson model parameters within the context of this hypothetical study.

This hypothetical study will be referred to throughout this chapter to facilitate the illustration of several proposed methods of prior construction for a Bayesian bivariate partial zero-inflated Poisson model. In addition, this hypothetical study will be used to demonstrate plausible methods of prior elicitation and how they can be applied in practice to scenarios such as the study described above as well as several other applications.

2.3.1 JAGS Specifications for Posterior Inference

Posterior inference was carried out with JAGS using the R package `rJAGS`. Inference was done with two chains. Initial values for parameters were randomly generated from the corresponding prior distributions. We ran 160,000 iterations and used the first 10,000 iterations as a burn-in. For the remaining 150,000 iterations we

Table 2.1: Interpretation of parameters for bivariate partial zero-inflated Poisson model for hypothetical adverse event study.

Parameter	Interpretation
p	The zero inflation parameter. This is the proportion of excess zeros for Y_1. It can also be thought of as the probability that a subject is not at-risk to experience Y_1 and is at-risk to experience Y_2.
$1-p$	The proportion of individuals that are at risk to experience both a migraine and a nausea episode.
λ_0	The rate associated with simultaneously experiencing both adverse events; this is an association parameter, represents a measure of dependency between outcomes Y_1 and Y_2.
λ_1	The mean of experiencing just migraines; $\lambda_1 + \lambda_0$ is the rate of experiencing migraines among those that are at-risk for experiencing migraines.
λ_2	The rate of experiencing just a nausea episode; $\lambda_2 + \lambda_0$ represents the mean of experiencing just a nausea episode.

sampled every 10th value to reduce autocorrelation. Accordingly, 30,000 parameter values were retained for each chain.

2.3.2 Posterior Inference for Adverse Event Study

Suppose we want to evaluate the safety of this new drug as defined by some pre-specified threshold that the number of migraines and the number of nausea episodes does not exceed. Specifically, for critical values, c_1 and c_2, we are interested in the posterior predictive probability that

$$\Pr(Y_1 < c_1 \text{ and } Y_2 < c_2 \,|\, \text{data}) \geq \delta, \tag{2.17}$$

for some probability δ. For our example, we generated the data depicted in Figure 2.1, including a slight positive association between the two adverse events.

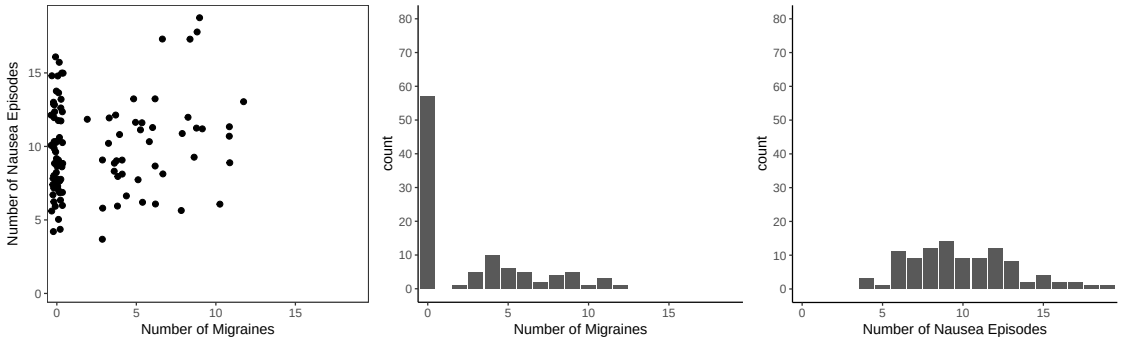

Figure 2.1: Scatterplot and marginal count data for number of migraines and number of nausea episodes for $n = 100$ observations. True values of parameters are $\boldsymbol{\lambda} = (2, 4, 8)$ and $p = 0.60$.

That is, as the number of migraines experienced by a subject increases the number of nausea episodes also increases. Later we extend this example to account for age as a covariate.

We consider relatively informative priors for model parameters, as indicated by the red, dashed densites in Figure 2.2. Discussion of methods for prior construction of model parameters within the context of the BPZIP model will be discussed in detail in subsequent sections. Posterior inference was carried out with the JAGS specifications described in Section 2.3.1. Standard diagnostics based on trace plots and the Gelman-Rubin statistic indicate no problems with convergence (Gelman and Rubin (1992), Brooks and Gelman (1998)). The resulting posterior densities are shown in Figure 2.2.

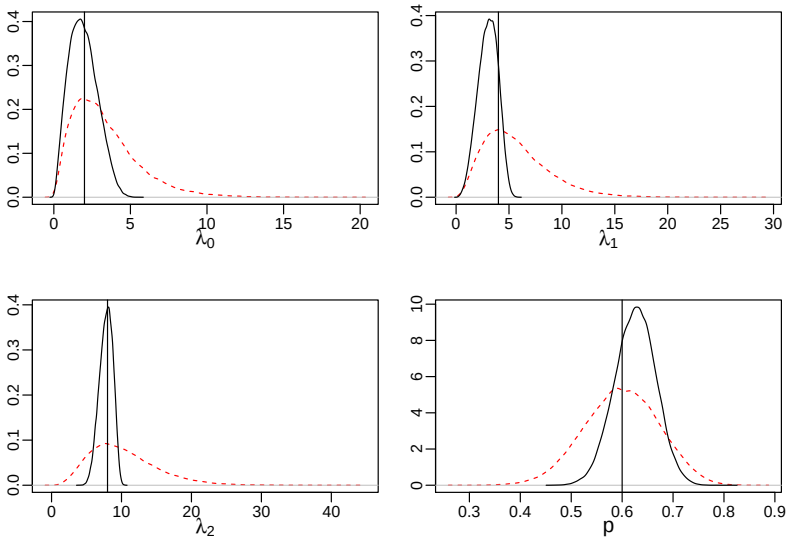

Figure 2.2: Priors (red, dashed) and posterior (black, solid) densities for BZIP model parameters.

Table 2.2 provides the posterior estimates and corresponding 95% credible intervals for model parameters.

Table 2.2: Posterior estimates and 95% credible intervals for BPZIP model parameters.

Parameter	Truth	Mean	SD	50%	2.5%	97.5%
λ_0	2	1.931	0.909	1.860	0.412	3.823
λ_1	4	3.018	0.933	3.060	1.153	4.698
λ_2	8	7.792	0.958	7.854	5.817	9.456
p	0.60	0.626	0.040	0.626	0.546	0.703

In Section 2.4 we present several examples within the context of the drug safety study in which inference for model parameters is problematic. We now evaluate the safety of this new drug. Suppose in (2.17) we set $c_1 = 10$, $c_2 = 18$ and $\delta = 0.95$. That is, given the data from this study, we require the posterior predictive probability that, given the data in Figure 2.1, the number of migraines is less than 8 and the

number of nausea episodes is less than 18 is at least 0.95. That is,

$$\Pr(Y_1 < 10 \text{ and } Y_2 < 18 \,|\, \text{data}) \geq 0.95.$$

This probability is easily computed with JAGS. Figure 2.3 shows a scatterplot for the posterior predictive distribution for Y_1 and Y_2

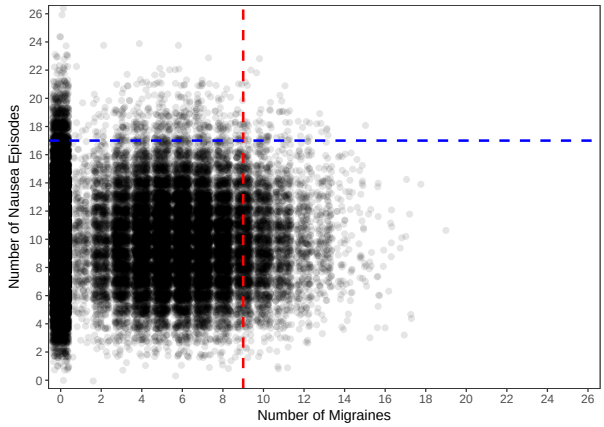

Figure 2.3: Scatterplot of the posterior predictive distribution for the number of migraines and number of nausea episodes. The red dashed line indicates the prespecified critical safety limit for the number of migraines (c_1), and the blue dashed line indicates the prespecified critical safety limit for the number of nausea episodes (c_2).

In particular, using the posterior predictive distribution, we have that

$$\Pr(Y_1 < 10 \text{ and } Y_2 < 18 \,|\, \text{data}) = 0.98.$$

Hence, given the data from the current study, it appears that this drug is in fact safe by this criterion. The FDA guidance on Bayesian methods (FDA (2010)) requires that prior probabilities of such success criteria be considerably less than the requisite success probability, in this case δ. This prior predictive probability of success is

$$\Pr(Y_1 < 10 \text{ and } Y_2 < 18 \,|\, \text{prior}) = 0.61,$$

which is indeed much less than $\delta = 0.95$. The prior structure does not unreasonably favor the desired threshold values.

2.4 BPZIP Examples

In this section, we present some problems that can arise in posterior inference for BPZIP model parameters, particularly in a diffuse prior setting. We begin by demonstrating a situation in which we have poor estimation of the zero-inflation parameter, p. We follow this with an example exhibiting a lack of convergence for the Poisson parameters λ_k, $k = 0, 1, 2$. For these examples, we did not consider covariates.

For the examples in this section, we consider diffuse priors for model parameters. Specifically, for the Poisson parameters we assume $\lambda_k \sim \text{Gamma}(1, 0.01)$ as the prior and for the zero-inflation parameter we consider $p \sim \text{Beta}(0.5, 0.5)$ as the prior.[4] These priors are depicted in Figure 2.4.

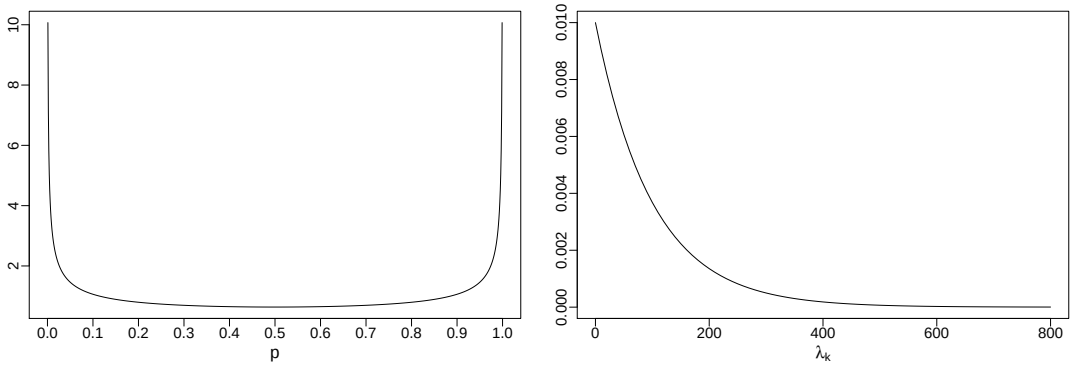

Figure 2.4: Diffuse priors on λ_k, $k = 0, 1, 2$ and p.

Posterior inference for all examples in this section was implemented in JAGS using the specifications outlined in Section 2.3.1.

2.4.1 BPZIP Example: Potential Problems with Inference for Zero-inflation Probabilty

In this example, we illustrate parameter values that can be problematic for estimation of the zero-inflation probability. Specifically, we demonstrate that, for p

[4] Here 1 and 0.01 denote the shape and rate parameter, respectively.

close to 0.5 and small λ_k, poor estimates obtain for p. Suppose the data we generated to simulate this hypothetical study are as depicted in Figures 2.5.

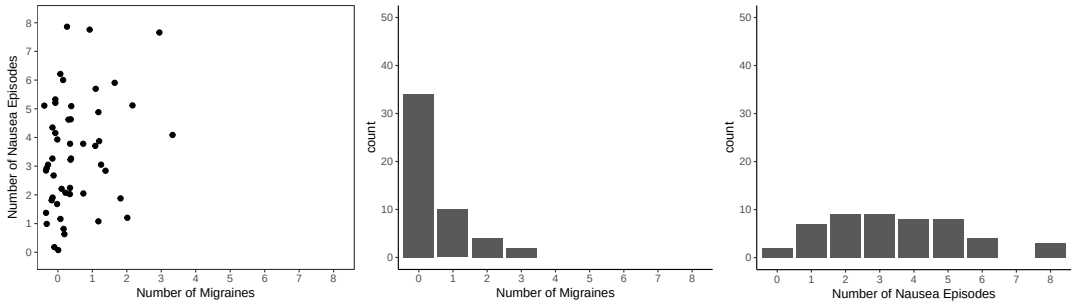

Figure 2.5: Data for $n = 50$ observations generated to simulate the hypothetical adverse event study. True values of parameters are $\boldsymbol{\lambda} = (0.5, 0.5, 3)$ and $p = 0.5$.

We consider diffuse priors on model parameters as presented in Section 2.4 and posterior inference was carried out with JAGS using the same specifications as described in Section 2.3.1. The posterior distribution for p is shown in Figure 2.6.

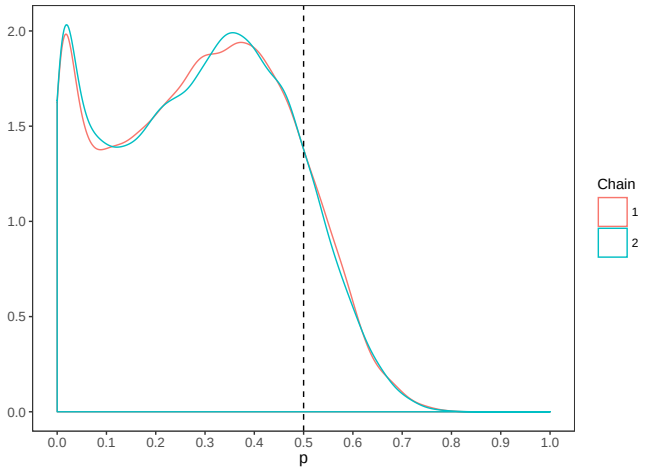

Figure 2.6: Posterior distribution for two chains (indicated by the red and blue density) for zero-inflation probability. The true value of the zero-inflation probability is indicated by the vertical dashed line.

The bimodality seen in the posterior for p suggests a lack of convergence of the parameter p (this was further indicated by standard diagnostic tests such as

trace plots and the Gelman-Rubin statistic). In addition, the posterior mode for p is underestimated compared to the true value. It is possible that this can be attributed to the fact that the model struggles to identify whether an observed zero is structural or sampling. Increasing the initial burn-in length, chain length and thinning rate yielded comparable results as shown here. This example demonstrates a situation in which diffuse prior may not be appropriate. A possible remedy to improve estimation of the zero-inflation probability could be to is to increase the sample size to say $n = 200$ or 300. Alternatively, we could consider informative priors for the zero-inflation probability. Construction of such is discussed in detail in subsequent sections.

2.4.2 BPZIP Example: Potential Problems with Inference for Poisson Parameters

In this example, we demonstrate the lack of convergence for the Poisson parameters, $\boldsymbol{\lambda}$. Consider again our adverse event study and suppose the data we generated to simulate the hypothetical drug study are as depicted in Figure 2.7. Specifically, we generated $n = 100$ observations from the model $(Y_1, Y_2) \sim \text{BPZIP}(p, \boldsymbol{\lambda})$, with true values of the parameters are $\boldsymbol{\lambda} = (2, 3, 8)$ and $p = 0.8$. That is, the data were generated in such a way that 80% of the population in our hypothetical adverse event study is not at-risk to experience migraines.

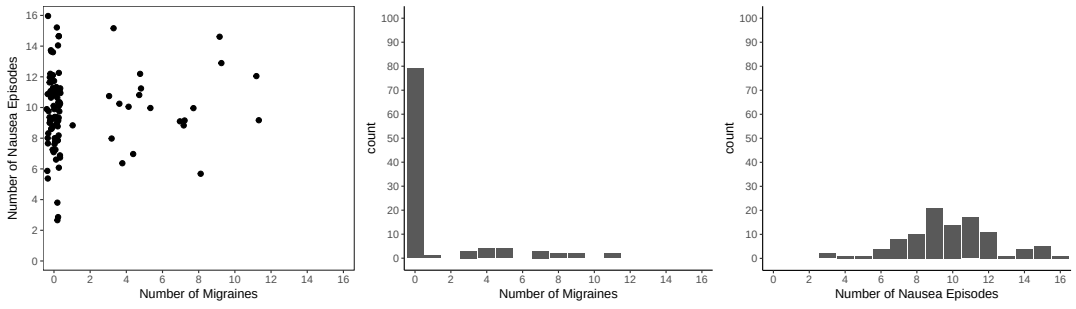

Figure 2.7: Data for $n = 100$ randomly generated observations from a BPZIP with true values $\boldsymbol{\lambda} = (2, 3, 8)$ and $p = 0.8$.

We again consider the diffuse priors shown in Figure 2.4. Posterior inference was carried out in JAGS with the specifications described in Section 2.3.1. Figure 2.8 shows the posterior densities for λ_k's.

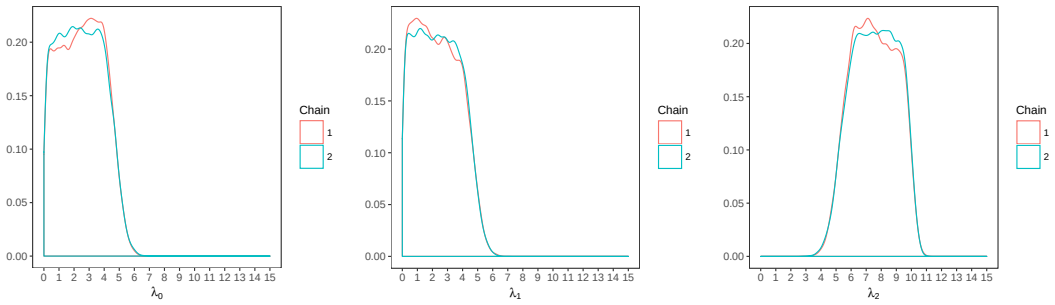

Figure 2.8: Posterior results for two MCMC chains (represented by the red and blue densities) for the λ_k's with diffuse priors. True values are $\boldsymbol{\lambda} = (2, 3, 8)$.

The irregular posteriors densities in Figure 2.8 suggest a lack of convergence of the λ_k's. Standard diagnostics based on trace plots, autocorrelation plots, and the Gelman-Rubin statistic further suggest that convergence is questionable. We increased length of initial burn-in, number of chains, number of iterations for each chain, and thinning rate. Nevertheless, this did not observe convergence.

Moreover, there is a lack of posterior updating of the λ_k's. Again, increasing burn-in length, chain length and thinning rate did not improve posterior updating. This lack of posterior updating is a common feature in nonidentifiable models. Specifically, the BPZIP model parameters, λ_0, λ_1, and λ_2 appear to be unidentified.

Figure 2.9 shows the posterior densities for $\mu_1 = \lambda_1 + \lambda_0$ and $\mu_2 = \lambda_2 + \lambda_0$. In contrast to the irregular posterior densities in Figure 2.8, the posterior densities for μ_1 and μ_2 are smooth and unimodal. In addition, there is more updating *a posteriori* compared to that seen in the posteriors for the λ_k's. This is not surprising as the observed data informs the model about the sums $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$. The data does not inform the model about the individual summands λ_1, λ_2, and λ_0. This contrast between the posterior densities for the individual summands, λ_k, and the

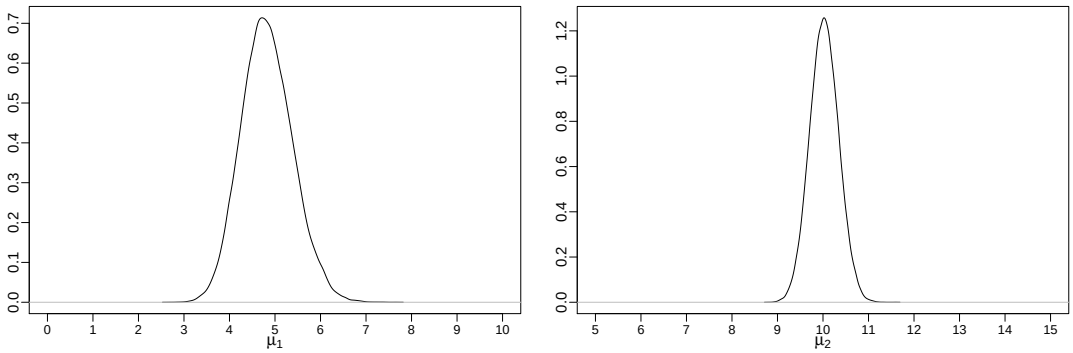

Figure 2.9: Posterior densities for marginal rates μ_1 and μ_2.

sums μ_1 and μ_2 demonstrates the nonidentifiability of the model parameters λ_k and the identifiability of the sums μ_1 and μ_2.

Now suppose we place a highly informative prior on λ_0, while keeping the priors on λ_1 and λ_2 diffuse. Namely, we assume

$$\lambda_0 \sim \text{Gamma}(2, 1),$$

where 2 and 1 denote the shape and scale parameter, respectively. This prior is depicted in Figure 2.10.

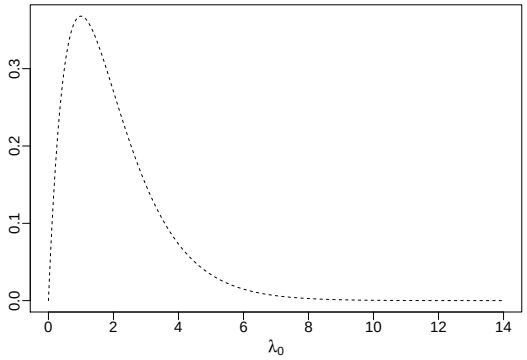

Figure 2.10: Informative gamma prior on λ_0.

We repeat posterior inference with the same JAGS specifications as in Section 2.3.1. Figure 2.11 and Table 2.3 shows the resulting posterior densities for the λ_k's.

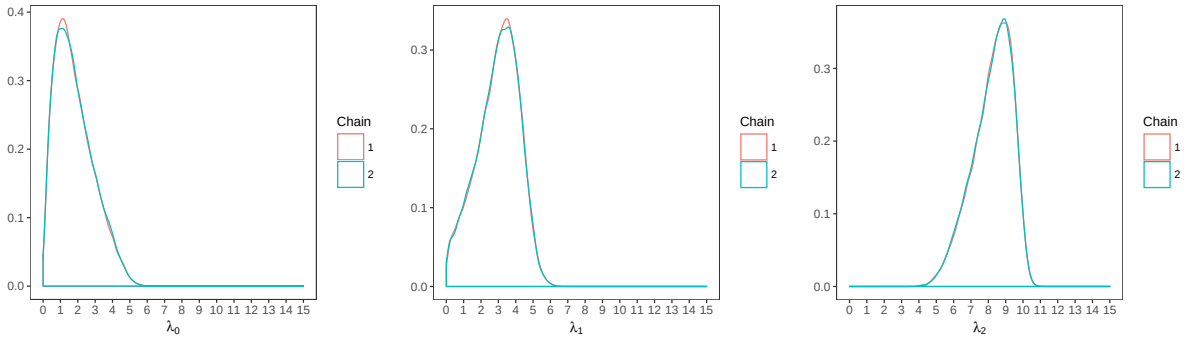

Figure 2.11: Posterior densities for two MCMC chains (represented by the red and blue densities) for Poisson parameters with informative prior on λ_0 and diffuse priors on λ_1 and λ_2.

Table 2.3: Posterior estimates and corresponding credible intervals for λ_k's when consider an informative prior on λ_0.

Parameter	Truth	Mean	SD	50%	2.5%	97.5%
λ_0	2	1.813	1.096	1.621	0.240	4.294
λ_1	3	2.967	1.200	3.094	0.419	5.008
λ_2	8	8.215	1.140	8.391	5.670	9.960

The smooth, unimodal posterior densities and standard diagnostics now suggest convergence of the sampler. This example suggests informative priors as a remedy for problems that nonidentifiability can cause. This is explored in greater detail in subsequent sections.

2.4.3 Additional Remarks on Posterior Inference for BPZIP Parameters

We constructed a variety of other examples leading to the following tentative conclusions with respect to inference for BPZIP model parameters. In several examples, we observed under-or overestimation for p for true values of p close to zero or one and small n. The estimation of the zero-inflation probability seems to improve with increasing sample size (e.g. $n = 200, 300$). Moreover, we consistently observed

that diffuse priors on p, such as the Jeffreys or uniform, can be problematic, particular with small sample sizes.[5]

With respect to inference for the λ_k's, in several examples, we observed that the effects of nonidentifiability are more evident for values of λ_k close to zero and close to each other (e.g. $\boldsymbol{\lambda} = (1, 1.5, 2)$). As the values of λ_k become farther from zero, and more distinguished from each other (e.g., $\boldsymbol{\lambda} = (5, 10, 20)$), the effects of nonidentifiability are less evident. In general, increasing the sample size does not improve estimation of the λ_k's (another observation that suggests nonidentifiability of the λ_k's). With respect to posterior inference for the sums, $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$, diagnostics suggest convergence and reasonable posterior estimates were obtained for all values of the parameter space explored. Hence, the one-off examples using simulated data indicate that identifiability of sums of parameters need not require identifiable summands, as expected.

In general, we found that the diffuse prior approach in Section 2.2.3 can be problematic with respect to convergence of the Gibbs sampler and thus problematic for posterior inference of model parameters. These findings motivated the exploration of informative priors for model parameters.

2.5 The Allure of Identifiability

Inference for nonidentifiable models can be problematic. This is the case whether using frequentist methods or Bayesian methods. The identifiability problems of the bivariate partial zero-inflated Poisson model can be mitigated somewhat by the use of Bayesian methods with informative priors. In Section 2.4.2 we demonstrated that the Poisson parameters λ_0, λ_1 and λ_2 are non-identifiable, but that the sums $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$ are identifiable. Throughout the dissertation, we

[5] We conclude that the results of several examples with respect to posterior inference for p in the BPZIP model were similar to those documented in Ghosh et al. (2006) in the context of the univariate ZIP model. This is what we would expect as our BPZIP model consists of a univariate ZIP model.

assume that estimation of the individual Poisson components is of interest.[6] Methods in the literature have been proposed to achieve identifiability of other models such as the standard bivariate Poisson model and the bivariate zero-inflated Poisson model. Theoretically, these methods could also be applied to our bivariate partial zero-inflated Poisson model in (2.2). For example, identifiability can be obtained by setting $\lambda_1 = \lambda_2$ (Yuen et al. (2015) do this in the context of a bivariate zero-inflated Poisson). In the context of a BPZIP model, this constraint implies that the the counts, Y_1 and Y_2, have equal rates among those "at-risk" to experience Y_1. Karlis and Ntzoufras (2003) achieve identifiability for the standard bivariate Poisson model by assuming a standard set of constraints, such as sum to zero constraints. Since the proposed BPZIP model consists of a standard BP distribution, we might also consider such constraints. However, such a simplifying assumption is very informative and does not seem reasonable in practice. Another option to obtain identifiability could be to assume a fixed value as the prior for λ_0. Again, this is extremely informative and not reasonable in practice. Mohammadi et al. (2016) alleviates the effects of nonidentifiability for a bivariate zero-inflated Poisson model by assuming that different covariates effect the individual Poisson rates. Such an assumption could also be applied to the BPZIP model, however, it would seem that this approach has limited use in practice.

An alternative, as used in other types of models with unmeasured confounding components or measurement error models is to add constraints (Robert (1994),Gustafson (2004)). For example, we could assume a strict ordering of the Poisson rates, $0 < \lambda_0 < \lambda_1 < \lambda_2$. Another option is to apply constraints to the sums $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$. However, determining conceivable constraints on these sums would be a difficult task. A potentially less extreme approach to mitigate problems non-

[6] Note that, even if the goal is to provide inference for $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$, despite reasonable posterior inference for these quantities, it is important to note the lack of model convergence which can compromise the results.

identifiability can cause is using moderately informative priors. In particular, in the context of the BPZIP regression model, an alternative is to place priors on the regression coefficents such that the induced priors on the BPZIP model parameters are relatively informative. The problems that nonidentifiability can cause in the estimation of parameters for the bivariate partial zero-inflated Poisson model makes the Bayesian approach to these models desirable as provided we supply a legitimate probability distribution as the prior, Bayes theorem will output a legitimate probability distribution as the posterior distribution. (Gustafson (2015)). This is the case whether or not the model is identifiable. Moreover, an aspect of the Bayesian paradigm in nonidentifiable settings is that the prior distribution can be used as a tool to *identify* parts of the parameter space that are not covered by the likelihood (i.e. in this case λ_1, λ_2 and λ_0), even though the choice of prior may affect the identifiable part (Robert (1994)).

Bayesian inference does not require indentifiability in models. That is, provided we supply a legitimate prior distribution, Bayesian inference will supply a legitimate posterior distribution. However, no statistical methodology is immune from the consequences of nonidentifiable models. In Bayesian inference, nonidentifiability typically presents convergence issues with associated Markov chain Monte Carlo (MCMC) methods. Moreover, in the absence of MCMC convergence problems, identifiability issues manifest themselves in the failure of a prior to update *a posteriori*. That is, there is little or no updating seen for the prior distribution of the unidentified parameter compared identified parameters. Gustafson (2015) remarks that

> One intuitive way of thinking about Bayesian inference in the absence of parameter identifiability is that the prior distributions play more of a role than usual (than in identifiable models) in determining the posterior belief about the parameters having seen the data.

In what follows we propose plausible methods of informative prior construction for the bivariate partial zero-inflated Poisson model. We begin with a discussion for informative prior construction for the simpler case without covariates. We use this to introduce prior elicitation tasks, which subsequently serve as a basis for prior construction and elicitation in the case with covariates. We return to the drug safety example as a vehicle for this development.

2.6 Informative Prior Structure for BPZIP Model

Bayesian methods for incorporating prior information in the analysis of zero-inflated models has not been covered in detail in the literature. Nevertheless, such methods seem desirable and can mitigate the effects of nonidentifiability. We offer several relatively informative prior structures for model parameters and plausible methods of prior elicitation from subject-matter experts for a Bayesian bivariate partial zero-inflated Poisson model with and without covariates. Although, less applicable in practice we use the non-regression case to introduce prior elicitation tasks, which will serve as a basis for prior construction and prior elicitation in the more complex case with covariates, which will be the primary focus.

2.6.1 Informative Priors on Poisson Parameters: Non-regression Case

We begin by discussing plausible methods of informative prior construction for the Poisson parameters, $\boldsymbol{\lambda}$, in the case of no covariates. We extend this to regression models in Section 2.6. We assume *a priori* independence of λ_0, λ_1, and λ_2 in (2.2). That is,

$$\pi(\boldsymbol{\lambda}) = \pi(\lambda_0) \times \pi(\lambda_1) \times \pi(\lambda_2).$$

Consider a subset of the bivariate set of outcomes from the hypothetical drug safety trial described in Section 2.3 such that all subjects are similar in age. That is, for the sake of illustration, we assume age does not affect the Poisson rates or zero-inflation probability. For reference, the interpretations of the Poisson rates in the context of

the adverse event drug safety trial are presented in Table 2.1. This information will be essential in the assessment tasks for the prior elicitation process.

In Section 2.4.2 we demonstrated the extent of informativeness on the rate of simultaneously experiencing both Y_1 and Y_2, denoted by λ_0, is critical in obtaining reasonable posterior estimates, particularly when the prior distributions on λ_1 and λ_2 are diffuse. This might suggest eliciting prior information from a subject matter expert on λ_0 as a solution to mitigate the deleterious effects of nonidentifiability. However, the Poisson parameter, λ_0, does not have an operational interpretation, making prior elicitation from a subject matter expert implausible.[7]

Accordingly, we consider several options for prior construction, including prior elicitation methods from a subject-matter expert for the Poisson rates λ_1 and λ_2 and provide plausible routes for elicitation on these quantities. One option is to impose a constraint on the ordering of parameters λ_1 and λ_2. For example, in the context of the adverse events example described in Section 2.3, suppose we have prior knowledge that the counts associated with migraines (Y_1) are often less than the counts associated with nausea events (Y_2). Given this information, we might impose the constraint that the rate of experiencing migraines is less than the rate of experiencing nausea events. That is, $\lambda_1 < \lambda_2$. The approach of adding constraints is widely used in other types of models such as models with unmeasured confounding components or measurement error models. Note that care should be taken when imposing informative constraints that are not inherent in the likelihood.

A potentially less extreme approach is to incorporate a *stochastic* ordering of the parameters λ_1 and λ_2 in the prior structure. Suppose again that, *a priori*, we have reason to believe that the rate of experiencing just Y_1 is less than the rate of experiencing just Y_2, $\lambda_1 < \lambda_2$, with some probability. In our example, suppose a subject matter expert believes that the rate of experiencing just migraines (λ_1)

[7] An alternative might be to use historical information to construct a prior for λ_0 if available.

is less than the rate of experiencing just nausea (λ_2). That is, $\Pr(\lambda_1 < \lambda_2) = 0.8$. One option to incorporate this stochastic ordering constraint in the prior structure is with copulas. We will not explore this option further, and for more details of this method we direct the reader to Nikoloulopoulos and Karlis (2009) and references therein.

A more plausible approach is to elicit expert judgement on λ_1 and λ_2 as they are conceivably observable quantities. The elictation of a gamma distribution is well documented in the literature. We use the mode-percentile method of elicitation to construct gamma distributions to reflect expert knowledge about the Poisson parameters, λ_1 and λ_2. The assessment tasks involved in the elicititation of prior information about the parameters λ_1 and λ_2 in the context of the hypothetical adverse event study are as follows: To assess the expert's knowledge and uncertainty about λ_1 the expert might be asked (recall that we are assuming age does not effect the Poisson parameters), "Among those at-risk to experience migraines, how many do you expect a subject to experience over the course of the study?" In response, the expert relays a most likely value (mode) and most extreme value (upper percentile). Their response is then translated into the parameters for a gamma distribution using numerical methods to reflect this information (O'Hagan et al. (2006)). To assess the expert's knowledge and uncertainty about λ_2 the expert is asked, "How many nausea episodes do you expect a subject to experience over the course of the study?" In response, the expert relays a most likely value (mode) and most extreme value (upper percentile) which is then translated into the parameters of a gamma distribution.

A limitation of this approach is that it involves independent elicitation of the modal value and percentile value (upper or lower) for constructing the prior distributions of λ_1 and λ_2,

$$\lambda_1 \sim \text{Gamma}(c_1, d_1)$$

and

$$\lambda_2 \sim \text{Gamma}(c_2, d_2).$$

Consequently, this approach is less reasonable as λ_1 and λ_2 are more dependent. Again, copula methods might be used to remedy this deficiency. Conditional beta distributions might also be used along with log transformations to $[0, \infty]$. For more on the latter approach, see Arnold et al. (2004)).

We now consider a simple method for modeling some degree of dependence among λ_1 and λ_2. In the hypothetical adverse event example, suppose the subject-matter expert believes that among subjects that experience both migraines and nausea, the rate of experiencing just nausea (λ_2) is *always* some percent greater (or λ_1 is always some percent smaller than λ_2) than the rate of experiencing just migraines (λ_1). This dependence can be represented by

$$\lambda_1 = \omega\lambda_2, \tag{2.18}$$

where ω represents a proportionality constant. Construction of a prior for λ_1 that reflects the relationship described by (2.18) requires elicitation of a prior on the rate of experiencing just Y_2, λ_2, and on the proportionality parameter, ω. Again, we use the mode-percentile method of elicitation to construct a gamma prior distribution on the rate of experiencing just nausea, λ_2. This involves prompting the expert to relay a most likely value (mode) along with an upper or lower bound (percentile) to represent uncertainty. Again, in the context of the hypothetical adverse event study the expert is asked, "How many nausea episodes do you expect a subject to experience over the course of the study?" In response, the expert relays a most likely value (mode) and most extreme value (upper percentile). Their response is then translated into the parameters for a gamma distribution to reflect this information using numerical methods (O'Hagan et al. (2006)). For example, suppose the expert believes the most likely value for the number of nausea espisodes over the study

is $\lambda_2 = 5$ and that the most extreme value is $\lambda_2 = 8$ (the 80th percentile). The corresponding prior distribution on the rate of experiencing just nausea is

$$\lambda_2 \sim \text{Gamma}(c_2 = 5.78, d_2 = 1.05),$$

where c_2 and d_2 denote the shape parameter and scale parameter, respectively.

Next, we construct a prior for the proportionality parameter in (2.18). If the expert believes that $\lambda_1 < \lambda_2$ with probability one, then $0 < \omega < 1$ and an appropriate choice for a prior on ω that reflects this is a beta distribution or a truncated scaled four parameter beta distribution. However, if it is possible that λ_1 could be greater than λ_2, and hence possible for $\omega > 1$, then a truncated Normal distribution, Gamma distribution, or a half-t distribution (less informative option compared to a truncated normal distribution) are appropriate choices as a prior distribution for ω. Note that each of the aforementioned possible prior distributions for ω ensure that the range of the prior for λ_1 is positive. For purposes of illustration, we assume that it is possible that $\omega > 1$ and assume a truncated normal distribution as a prior for ω. In particular, we can elicit information about ω from a subject-matter expert as follows

(1) The expert is asked "Consider a subset of 100 subjects. For what percent of these subjects, do you think, the average number of nausea episodes will be greater than the average number of migraines (λ_2)?" Their response is translated into the modal value of the truncated normal.

(2) The expert is then asked "What is the largest (smallest) percent that this can be?" Their response is translated into a upper (lower) percentile.

As an example, suppose the expert says that the number of migraines will be less than the number of nausea episodes in 80% of the subjects and that the smallest this percentage can be is 65%. We construct a truncated normal distribution on ω using 80% as the modal value and 65% as the lower 20th percentile. This information

yields the following prior on ω:

$$\omega \sim \text{Truncated-Normal}_{(0,\infty)}(\mu = 0.8, \sigma = 0.18).$$

We now use the elicited prior distribution on λ_2 and α to construct a prior on the rate of nausea episodes, λ_1. From (2.18), we have that

$$\pi(\lambda_1) \propto \pi(\omega)\pi(\lambda_2),$$

where $\pi(\lambda_2)$ and $\pi(\omega)$ denote the elicited gamma and truncated-normal densities, respectively.

The resulting prior distributions based on the information obtained from the expert are shown in Figure 2.12. Furthermore, Table 2.4 summarizes the mode and 25th percentile of the induced prior on λ_1, as well as the mode and 25th percentile of the gamma distribution based on the information collected from the expert on λ_2.

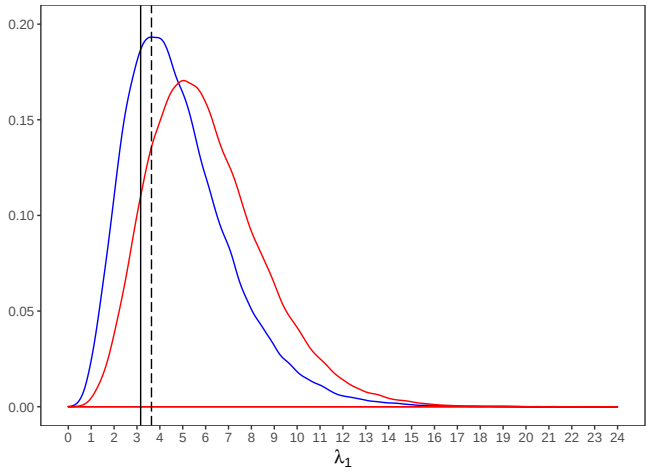

Figure 2.12: Induced prior on λ_1 is in blue, and elicited gamma prior on λ_2 is in red. The mode and 25th percentile of the induced prior on λ_1 is represented by the dashed line and solid line, respectively.

Table 2.4: Prior distributions on λ_1 and λ_2 based on the information elicited from the expert.

Parameter	Mode	25th Percentile
λ_1	3.72	3.17
λ_2	5.16	4.24

Providing feedback to the subject-matter expert is an essential part of the prior elicitation process as it allows the expert to evaluate their assessed prior distributions. We elaborate more on possible ways to provide feedback to the expert to show them the consequences/wisdom of their priors in Section 2.7.1. For our purposes, we assume that the expert is satisfied with these priors for λ_1 and λ_2 accurately reflect their prior knowledge.

2.6.2 Informative Priors on Zero-inflation Parameter: Non-regression Case

The posterior behavior of the zero-inflation parameter p in the bivariate partial zero-inflated Poisson case is similar to that of a univariate zero-inflated Poisson (see Ghosh et al. (2006)). Via several one-off examples using simulated data, we observed that diffuse priors for the the zero-inflation parameter can be problematic, particularly for small n ($n = 50, 100$) and small values of λ_k in which case the model has trouble distinguishing between structural and sampling zeros (see Section 2.4.1).

Recall that p is the zero-inflation probability and represents the proportion of individuals that are not at-risk to experience migraines (in the context of the hypothetical adverse event study). This is a conceivably observable quantity about which to elicit subject-matter expert opinion. There are several methods for eliciting a beta distribution found in the literature. We adapt the method proposed by Elfadaly and Garthwaite (2013a). In particular, Elfadaly and Garthwaite (2013a) suggest eliciting a median, lower bound (25th percentile) and upper bound (75th percentile) to determine parameters of a beta distribution that reflect the expert's

belief within a general context. We modify the general assessment tasks proposed by Elfadaly and Garthwaite (2013a) to apply the the BPZIP model. Consider a subset of the hypothetical adverse event example. The assessment tasks are as follows (again here we assume that age does not effect the zero-inflation parameter):

(1) "Suppose we have 100 subjects from the study population. What, do you think, is the percentage of these subjects that are *not* at-risk to experience migraines?" Their response is taken to be the median.

(2) Next the expert is asked "Suppose the percentage of these subjects not at-risk to experience migraines is actually higher than your initial assessment. Given this information, what do you now think is the percentage of these subjects not at-risk to experience migraines?" Their response is taken to be the 75th percentile.

(3) Finally, the expert is asked "Suppose the percentage of these subjects not at-risk to experience migraines is actually lower than your initial assessment. Given this information, what do you think now think is the percentage of these subjects not at-risk to experience migraines?" Their response is taken to be the 25th percentile.

As an example, suppose the expert believes that the 25th percentile, median and 75th percentile are 0.35, 0.40, and 0.45, respectively. This information can then be translated in to the parameters of a beta distribution using numerical methods (Garthwaite et al. (2005), Elfadaly and Garthwaite (2013a)). Specifically,

$$p \sim \text{Beta}(17.6, 26.3). \tag{2.19}$$

In practice, this prior would be shown to the expert to verify whether or not this distribution accurately reflects their prior beliefs. Modifications are made until satisfaction obtains.

Prior effective sample size (ESS) is an important part of the feedback process. Prior ESS (this is further illustrated in Section 2.7.1) is particularly important when applying Bayesian methods in settings with a small to moderate sample size (Morita et al. (2008)). The beta prior distribution in Figure 2.19 has a prior equivalent sample size of about $n = 44$ observations, which is half of our study sample size and, therefore, highly informative. In practice, we can relay this information to the expert in order to see if they want to reconsider their prior choice. Although this is a considerably informative prior assessment the data (rather than the prior) will still dominate the posterior. We can also provide feedback by using the prior predictive distribution to generate observations as plausible data that might result based on the expert's judgment. This provides feedback to the expert in a meaningful scale. We demonstrate use of the prior predictive distribution as a tool to provide feedback in Section 2.7.1.

2.6.3 Conditional Means Priors

Inference for generalized linear models often relies on diffuse priors for regression coefficients as outlined in Section 2.2.4. We showed that placing diffuse priors on regression coefficients in the BPZIP model can be problematic as often result in convergence issues. If we have prior data, we can use it to construct the prior (power priors, mixture priors etc.). If not, we can elicit subject-matter expert opinion to construct the prior structure. Bedrick et al. (1996) developed an approach by which priors for regression parameters can be specified via indirect elicitation. The priors produced by this method (namely, the induced priors on regression coefficients) are referred to as conditional means priors, or BCJ priors, and are commonly used on regression parameters in generalized linear models. In particular the conditional means prior (CMP) method involves eliciting prior information about expectations of mean responses corresponding to observables with fixed covariates.

It is difficult to elicit prior information directly on regression coefficents because it is often hard for subject-matter experts to think in terms of intercepts and partial slopes. The CMP approach addresses this problem by instead eliciting information from experts on average response values at specific covariate values, a quantity that is in the scale of the model's observables, and therefore thus has a more meaningful interpretation. This information is subsequently used to induce priors on the regression coefficients.

The general procedure for specifying CMPs is as follows. Specific values of predictor variables are selected for which prior information will be obtained. For r covariates, we choose K r-dimensional $(1 \times r)$ vectors of covariate values. Priors are then elicited for the mean of the responses for each of the K covariate vectors, hence the name "conditional means prior", as these means are conditioned on the covariate values. These priors are then used to induce a prior on the regression parameters.

Consider the general form of a generalized linear model. For $i = 1, \ldots, n$, let y_i have density $g(y_i \,|\, \mu_i, \boldsymbol{\phi})$ with $\mu_i = f(\mathbf{x}_i'\boldsymbol{\beta})$, where $\mathbf{x}_i'$ represents a $r \times 1$ vector of covariates and $\boldsymbol{\beta}$ is the corresponding $r-$dimensional vector of regression coefficients. That is, $y_i \sim g(y_i \,|\, \mu_i, \boldsymbol{\phi})$, $\mathrm{E}(Y_i) = \mu_i$, and $f^{-1}(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$. In this context, f^{-1} represents the link function (e.g. for binomial, f^{-1} is the logit(μ_i)) and $\boldsymbol{\phi}$ represents a vector of nuisance parameters. The goal is to induce a prior distribution on the regression coefficients, $\boldsymbol{\beta}$, based on priors elicited on the mean vector, $\boldsymbol{\mu}$, at various covariate configurations of $\mathbf{x}_i$ through link function, $f^{-1}(\cdot)$. (h is the inverse link function here)

Suppose we have K covariate vector configurations. We define the $K \times r$ design matrix

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \widetilde{\mathbf{x}}_1 \\ \vdots \\ \widetilde{\mathbf{x}}_K \end{bmatrix},$$

where the rows represent K distinct values of covariate vector $\mathbf{x}_i$ (i.e. if $r = 2$, $\widetilde{\mathbf{x}}_1 = (1, 25), \widetilde{\mathbf{x}}_2 = (1, 65)$). Note that it is convenient for $\widetilde{\mathbf{X}}$ to be nonsingular and thus, we require $K = r$. Moreover, it should be emphasized that the $\widetilde{\mathbf{x}}_i$, $i = 1, \ldots, K$ values are hypothetical values, not actual observed data values. Denote the covariate configurations, $\widetilde{\mathbf{x}}_h$ for $h = 1, \ldots, K$. For each covariate configuration, $\widetilde{\mathbf{x}}_h$, a prior is elicited for the corresponding mean response value $\tilde{\boldsymbol{\mu}}_h = h(\widetilde{\mathbf{x}}_h\boldsymbol{\beta})$. This informative prior is elicited from an expert and denoted as G_h. Note that all the G_h's are assumed to be independent which in turn indicates that the covariate configurations $\widetilde{\mathbf{x}}_h$ are assumed to be sufficiently distinct.

The general prior structure for the conditional means prior approach can be derived as follows. Define the matrix

$$\tilde{\boldsymbol{\mu}} = \begin{bmatrix} \tilde{\mu}_1 \\ \vdots \\ \tilde{\mu}_K \end{bmatrix} = \begin{bmatrix} f(\widetilde{\mathbf{x}}_1'\boldsymbol{\beta}) \\ \vdots \\ f(\widetilde{\mathbf{x}}_K'\boldsymbol{\beta}) \end{bmatrix} \equiv f(\widetilde{\mathbf{X}}\boldsymbol{\beta}).$$

Note that f corresponds to some monotonic increasing and invertible function. Common selections for f are the logistic, probit and complementary log-log functions. Since we construct $\widetilde{\mathbf{X}}$ in such a way that it is invertible, we can solve for $\boldsymbol{\beta}$ and obtain:

$$\boldsymbol{\beta} = \widetilde{\mathbf{X}}^{-1} f^{-1}(\tilde{\boldsymbol{\mu}}),$$

where, for example, if f is the logistic transform, then $f^{-1}(\tilde{\boldsymbol{\mu}}) = \text{logit}(\tilde{\boldsymbol{\mu}})$. Thus, the induced priors on the regression coefficients, $\boldsymbol{\beta}$ are defined as

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} \sim \widetilde{\mathbf{X}}^{-1} f^{-1} \left(\begin{bmatrix} G_1 \\ \vdots \\ G_K \end{bmatrix} \right).$$

This induced prior often has no closed form. This is not problematic as posterior computations can be readily handled by Monte Carlo methods.

2.6.4 Conditional Means Priors for a Bayesian BPZIP Regression Model

We can represent a BPZIP model in terms of a generalized linear model such that the zero-inflation parameter and Poisson parameters depend on covariates through canonical link generalized linear models as in (2.11) and (2.10), respectively. Thus, the BPZIP model lends itself well to the conditional means prior approach for model parameters. Consider our adverse event drug safety study in Section 2.3. The data we generated to simulate this study for $n = 100$ subjects between the age of 20 and 80 are depicted in Figures 2.13 and 2.14.

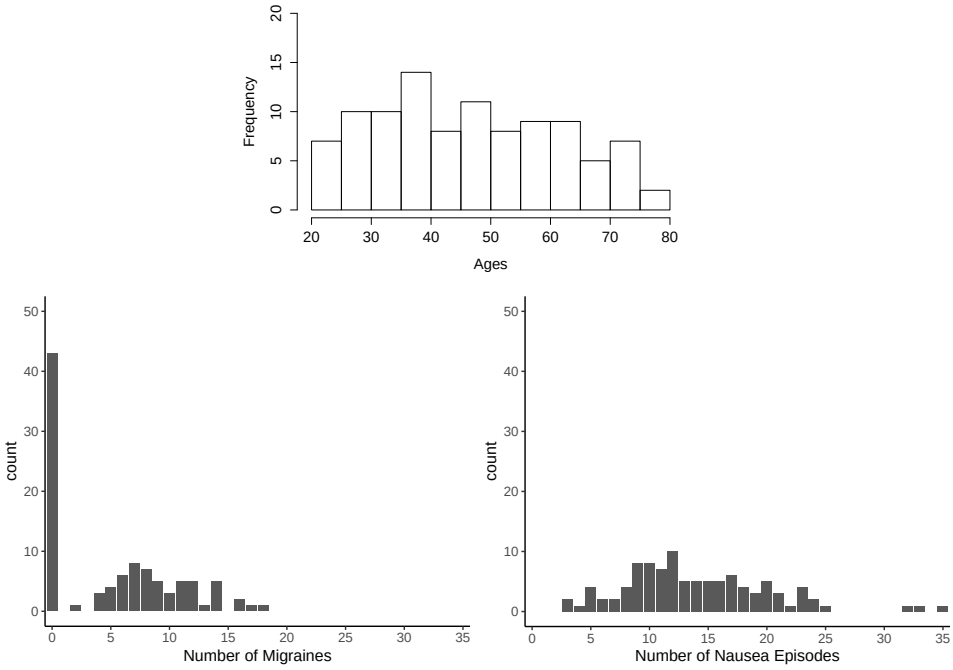

Figure 2.13: Ages and marginal data for counts Y_1 and Y_2 for $n = 100$ subjects.

Figure 2.14 suggests a positive association. As the number of migraines experienced by a subject increases the number of nausea episodes also increases.

Table 2.5 provides the model parameter summary for subject i and adverse event k, $i = 1, \ldots, n$, $k = 0, 1, 2$ ($k = 0$ denotes both adverse events).

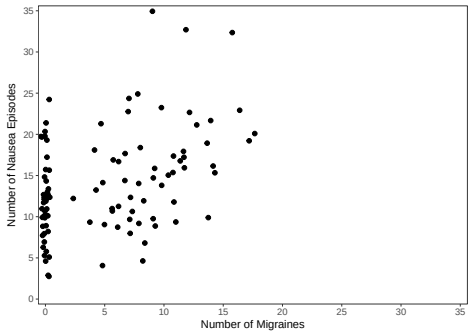

Figure 2.14: Scatterplot illustrating the joint association between the number of migraines and the number of nausea episodes.

Table 2.5: Model parameter summary for subject i.

Parameter	Interpretation
p_i	The probability that the ith subject age x_i years-old is *not* at-risk to experience migraines.
λ_{1i}	The mean rate of experiencing just a migraine (among those at risk to experience migraines) for a subject age x_i years old.
λ_{2i}	The mean rate of experiencing just a nausea episode for a subject age x_i years old.
λ_{0i}	The mean rate associated with simultaneously experiencing both migraines and nausea for a x_i-year-old subject.
β_0	If covariates are centered, the log odds that a subject aged $\bar{x} = 46$ is not at-risk to experience migraines.
β_1	The log odds that a subject aged x_i is not at-risk to experience migraines; e^{β_1} represents the increased (or decreased) odds that a subject is not at-risk for migraines for a one year increase in age.
$\gamma_{k,0}$	If covariates are centered, average log adverse event rate for a subject aged $\bar{x} = 46$.
$\gamma_{k,1}$	Average change in log adverse event rate for a one year increase in age; For a one-year increase in age, the expected number of adverse event k increases (decreases) by a factor of $e^{\gamma_{k,1}}$.

Figure 2.15 provides a diagram illustrating the conditional means prior as applied to the hypothetical adverse event study.

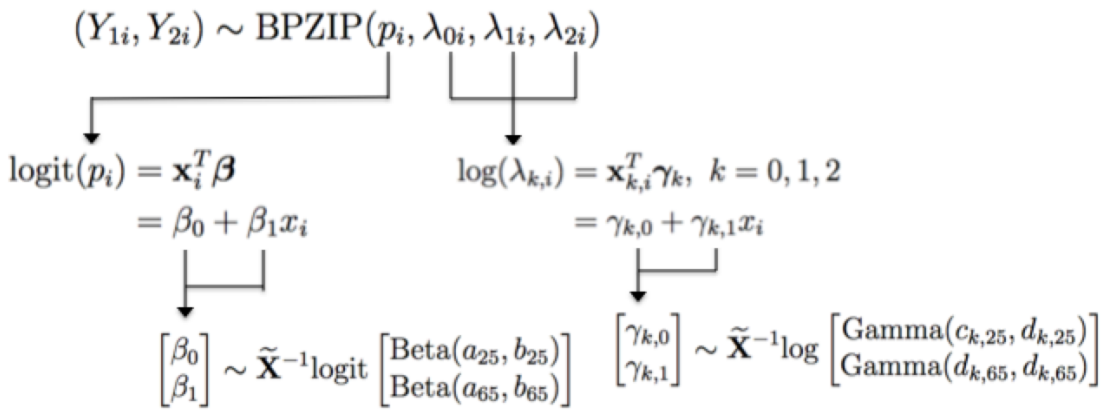

Figure 2.15: Bayesian BPZIP model with conditional means priors in context of hypothetical adverse event study.

We first consider elicitation of a conditional means prior for the logistic regression parameters used to model the zero-inflation probability. Within the context of a hypothetical adverse event study, p_i denotes the probability that the ith subject is not at-risk to experience migraines. In particular, for the ith subject we have that

$$\begin{aligned}\text{logit}(p_i) &= \mathbf{x}_i'\boldsymbol{\beta} \\ &= \beta_0 + \beta_1\, x_i,\end{aligned}$$

where x_i represents the single covariate, age. Since there are two regression coefficients, β_0 and β_1, the design matrix, $\widetilde{\mathbf{X}}$, will consist of two covariate configurations. Bedrick et al. (1996) discuss guidelines for value selection for covariate configurations. For example, we require that the values of $\widetilde{\mathbf{x}}_i$ and $\widetilde{\mathbf{x}}_j$, $i \neq j$ are sufficiently "far apart" so that knowledge of elicitation at covariate configuration $\widetilde{\mathbf{x}}_i$ does not influence elicitation at covariate configuration $\widetilde{\mathbf{x}}_j$. For our hypothetical adverse event example, we assume that the difference between age 25 and age 65 represents sufficiently distinct responses. Since we have two regression coefficients, we consider two

covariate configurations for the design matrix, $\widetilde{\mathbf{X}}$. In particular, let $\tilde{x}_1 = 25$ years old and $\tilde{x}_2 = 65$ years old.

Applying the conditional means prior approach, we elicit information about the mean zero-inflation probability at the two choosen covariate configurations, age 25 and age 65, and subsequently induce priors on $\boldsymbol{\beta}$. There are several methods discussed in the literature on elicitation of parameters for a beta distribution (see Hughes and Madden (2002), O'Hagan et al. (2006)). We use a method proposed by Elfadaly and Garthwaite (2013a) to elicit univariate beta distributions as priors on $\tilde{\mathbf{p}} = (\tilde{p}_{25}, \tilde{p}_{65})$, which involves assessments of a median value, and two quartiles (upper 75th percentile and lower 25th percentile) as a measure of the expert's uncertainty. Numerical methods are then used to solve for the parameters of a beta distribution that reflects the information collected from the subject-matter expert. For further detail of the numerical methods used to solve for the beta parameters see Elfadaly and Garthwaite (2013a).

Consistent with the assessment tasks outlined in Elfadaly and Garthwaite (2013a), the elicitation process with a subject-matter expert with respect to their judgment about the zero-inflation probability is conducted as follows. At both age 25 years old and 65 years old, the subject-matter expert is asked a series of three questions (one for assessment of a median value, one for assessment of the lower quartile and one for assessment of a upper quartile) in order to represent the expert's knowledge and uncertainty.

(1) To assess the median, the expert is asked "Suppose we have 100 subjects from the study population that are x_i years old. What, do you think, is the percentage of these x_i year old subjects that are not at-risk for experiencing migraines? Their response is taken to be the median value.

(2) Next, to assess a lower quartile for the percentage of subjects not at-risk for experiencing migraines, the expert is asked,"Suppose the percentage of

x_i year-old subjects that are not at-risk for experiencing migraines is actually *less than* your initial assessment (that is, their median assessment is too high). Given this information, what would you now estimate as the percentage?" Their response is taken to be the 25th percentile.

(3) Finally, to obtain an upper quartile the expert is asked "Suppose the percentage of x_i year-old subjects not at-risk for experiencing migraines is actually *greater than* your initial assessment (that is, their median assessment is too low). Given this information, what would you now estimate as the percentage?" Their response is taken to be the 75th percentile.

A primary goal in the prior elicitation process is to represent knowledge and uncertainty of a subject matter expert on conceivably observable quantities familiar to the expert. Note that in the exchange described above, we asked the expert to consider a group of 100 subjects that are 25 years old and a group of 100 subjects that are 65 years old. We did not ask the expert to assess *one* subject that is 25 years old or *one* subject that is 65 years old. This is an important distinction. We want to prompt the expert to think about a group of x_i year old subjects, not an individual subject that is x_i years old as an individual's assessment is superior for an aggregate rather than an individual.

Following the conditional means priors approach outlined in Section 2.6.3, we induce priors on $\boldsymbol{\beta}$ as follows. We have that $\tilde{\mathbf{p}} = f(\mathbf{x}_i'\boldsymbol{\beta})$, where here f is the logistic model, and the inverse, f^{-1}, is the logit link function. Thus, the resulting prior at $x_i = 25$ years old is

$$p_{x_i=25} \equiv \text{logit}^{-1}\left(\beta_0 + \beta_1(25)\right) \sim \text{Beta}(10.40, 4.60)$$

and the resulting prior at $x_i = 65$ is

$$p_{x_i=65} \equiv \text{logit}^{-1}\left(\beta_0 + \beta_1(65)\right) \sim \text{Beta}(4.60, 10.40).$$

Suppose we obtain the information in Table 2.6 corresponding to the expert from the assessment tasks above.

Table 2.6: Prior information collected from expert about the zero-inflation probability.

Parameter	25th percentile	Median	75th percentile
p_{25}	0.62	0.70	0.78
p_{65}	0.22	0.30	0.38

The resulting beta prior density plots are shown in Figure 2.16. In practice, these densities are presented to the subject-matter expert for feedback and may be adjusted if the elicited densities do not accurately reflect their beliefs.

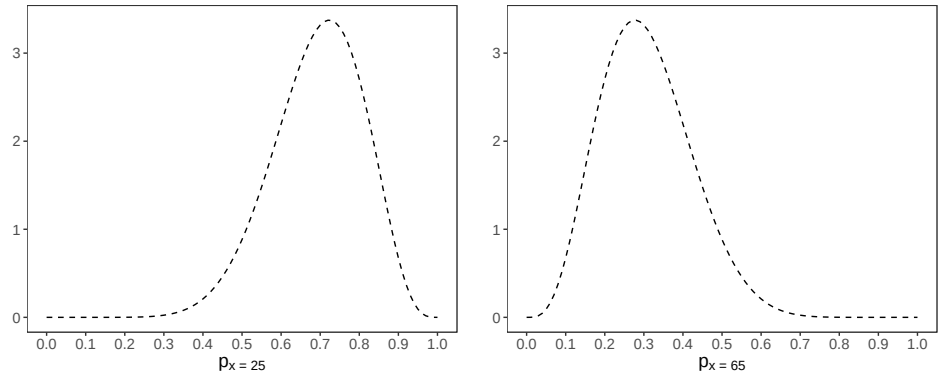

Figure 2.16: Informative beta priors for p at age 25 and 65.

The conditional means priors for β_0 and β_1 are the resulting induced priors given by

$$\boldsymbol{\beta} = \widetilde{\mathbf{X}}^{-1}\text{logit}(\tilde{\mathbf{p}}).$$

For this example, we have that

$$\begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix} \sim \widetilde{\mathbf{X}}^{-1}\text{logit} \begin{bmatrix} \text{Beta}(10.40, 4.60) \\ \text{Beta}(4.60, 10.40) \end{bmatrix},$$

where

$$\widetilde{\mathbf{X}} = \begin{bmatrix} 1 & 25 \\ 1 & 65 \end{bmatrix}.$$

The induced priors on $\boldsymbol{\beta}$ have no closed form but can easily be simulated as shown in Figure 2.17.

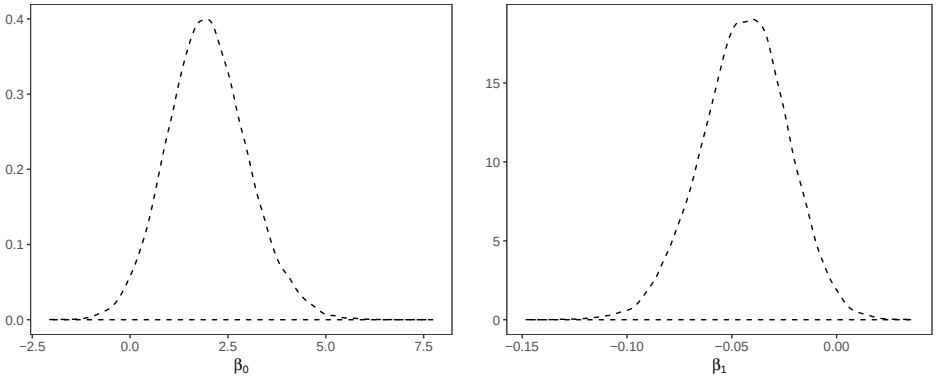

Figure 2.17: Simulated density plots for the induced priors on β_0 and β_1.

Similarly, we can elicit conditional means priors for the Poisson regression used to model the number of adverse events represented by the bivariate response vector (Y_1, Y_2). In particular, for the ith subject we have

$$\begin{aligned} \log(\lambda_{k,i}) &= \mathbf{x}_i' \boldsymbol{\gamma}_k \\ &= \gamma_{k,0} + \gamma_{k,1}\, x_i, \end{aligned}$$

where $\gamma_{k,0}$ and $\gamma_{k,1}$ denote the intercept and slope, respectively, corresponding to $\lambda_{k,i}$. Here the log link function replaces the logistic function. That is, $f(\mathbf{x}_i'\boldsymbol{\gamma}_k) = \log(\mathbf{x}_i'\boldsymbol{\gamma}_k)$ and the inverse, f^{-1} is the exponential function. The covariate configurations are again $\tilde{x}_i = 25$ years old and $x_i = 65$ years old where we assume that this difference in age renders sufficiently distinct responses.

For the Poisson parameters, $\lambda_{k,i}$, $k = 0, 1, 2$ the assessment tasks for the elicitation of a modal value and quartile (upper or lower bound) for a gamma distribution process are as follows. Note that the choice of whether to elicit an upper bound or

a lower bound is problem dependent and left to the judgement of the researcher as to which they feel is more natural and best gauges the expert's uncertainty. For our purposes we elicit an upper bound from the expert when evaluating their knowledge about the rates of occurrence of the adverse events of interest in the study population. In general, we obtain a value for the upper bound by asking the expert to relay a most extreme value. O'Hagan et al. (2006) discusses the choice of percentile to represent the expert's most extreme value. We choose the 80th percentile to represent the expert-relayed most extreme value.[8] First, we prompt the expert about expectations for λ_1, the rate of experiencing just migraines. For $x_i = 25$ and $x_i = 65$, the expert is asked a series of questions:

(1) "Suppose we have subjects from the study population aged x_i years-old that are at-risk to experience migraines. What, do you think, is the most likely value for the number of migraines experienced by these subjects over the course of the study?" Their response is taken to be the mode.

(2) Next, we prompt the expert to suggest an upper bound to assess their uncertainty. The expert is asked "What, do you think, is the largest the number of migraines can be among subjects aged x_i?" Their response is taken to be the 80th percentile.

Next, we elicit information about λ_2 by assessing the expert's judgment about the number of nausea episodes. Similarly, we elicit information about the rate, λ_0, of experiencing both adverse events.

We assume that we have prior information to construct a moderately informative prior on λ_0. Elicitation about λ_0 may be difficult in practice. A subject-matter

[8] The literature on prior elicitiation suggests that the choice of the most extreme value relayed by the expert to represent the 67-80 percentile is far better than setting the value as the 90-95 percentile as even the most circumspect expert tends to be over-confident in their ability to make assessments (O'Hagan et al. (2006)).

expert may have difficulty assessing the simultaneous rate of two events. We propose a method of prior elicitation to obtain an informative prior for λ_0 in Chapter Three.

Table 2.7 shows the information we assume has been collected from the expert about the rates of the adverse events in our example.

Table 2.7: Prior information collected from expert about the Poisson parameters.

Parameter	Mode	80th percentile
$\lambda_{1,x_i=25}$	4	7
$\lambda_{1,x_i=65}$	8	11
$\lambda_{2,x_i=25}$	6	9
$\lambda_{2,x_i=65}$	16	19
$\lambda_{0,x_i=25}$	2	4
$\lambda_{0\,x_i=65}$	3	5

The information in Table 2.7 is translated in to the parameters of a gamma distribution to get the following gamma priors for λ_1:

$$\lambda_{1,x_i=25} \equiv \exp\left(\gamma_{1,0} + \gamma_{1,1}(25)\right) \sim \text{Gamma}(4.49, 1.15),$$

$$\lambda_{1,x_i=65} \equiv \exp\left(\gamma_{1,0} + \gamma_{1,1}(65)\right) \sim \text{Gamma}(10.59, 0.83),$$

and the following gamma priors for λ_2

$$\lambda_{2,x_i=25} \equiv \exp\left(\gamma_{2,0} + \gamma_{2,1}(25)\right) \sim \text{Gamma}(7.22, 0.96),$$

and

$$\lambda_{2,x_i=65} \equiv \exp\left(\gamma_{2,0} + \gamma_{2,1}(65)\right) \sim \text{Gamma}(30.44, 0.54).$$

Finally, suppose we have the following gamma priors for λ_0:

$$\lambda_{0,x_i=25} \equiv \exp\left(\gamma_{0,0} + \gamma_{1,0}(25)\right) \sim \text{Gamma}(3.37, 0.84),$$

and

$$\lambda_{0,x_i=65} \equiv \exp\left(\gamma_{0,0} + \gamma_{1,0}(65)\right) \sim \text{Gamma}(5.11, 0.73).$$

Prior density plots are shown in Figure 2.18. In practice, we would obtain feedback from the subject-matter expert and may adjust the prior distributions accordingly. We assume the expert is satisfied with resulting priors.

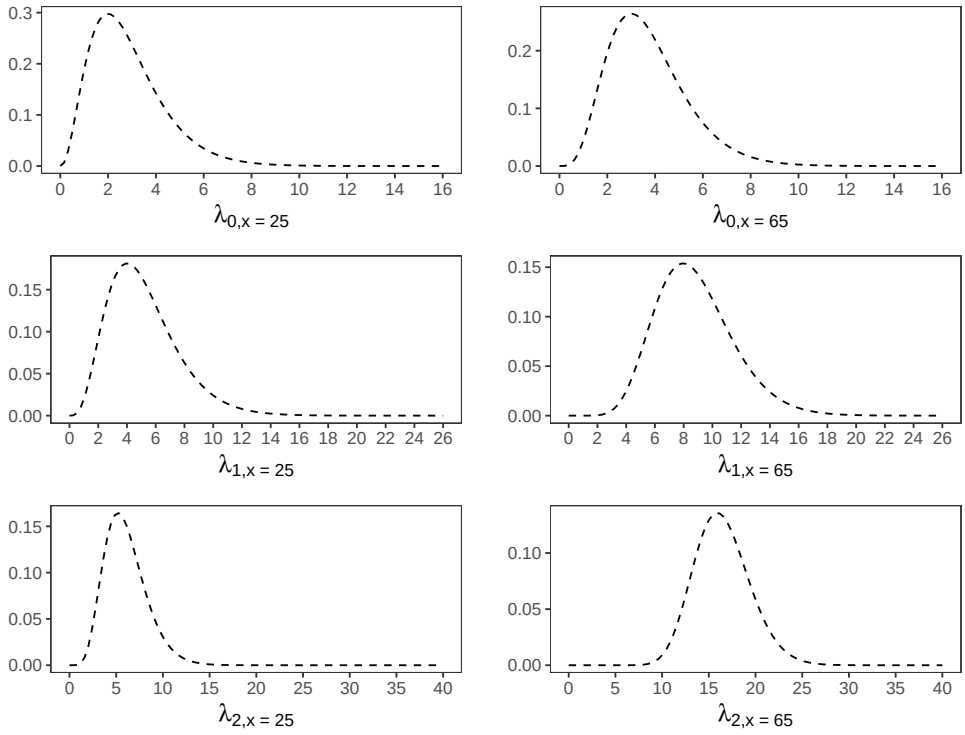

Figure 2.18: Informative gamma distributions for λ_k, $k = 0, 1, 2$, at age 25 and 65.

The conditional means priors for the regression coefficients, $\gamma_{k,0}$ and $\gamma_{k,1}$, $k = 0, 1, 2$, are given by

$$\begin{bmatrix} \gamma_{k,0} \\ \gamma_{k,1} \end{bmatrix} \sim \widetilde{\mathbf{X}}^{-1} \log \begin{bmatrix} \text{Gamma}(c_{1k}, d_{1k}) \\ \text{Gamma}(c_{2k}, d_{2k}) \end{bmatrix},$$

where

$$\widetilde{\mathbf{X}} = \begin{bmatrix} 1 & 25 \\ 1 & 65 \end{bmatrix}.$$

Again, these induced priors have no closed form but can be simulated and are shown in Figure 2.19.

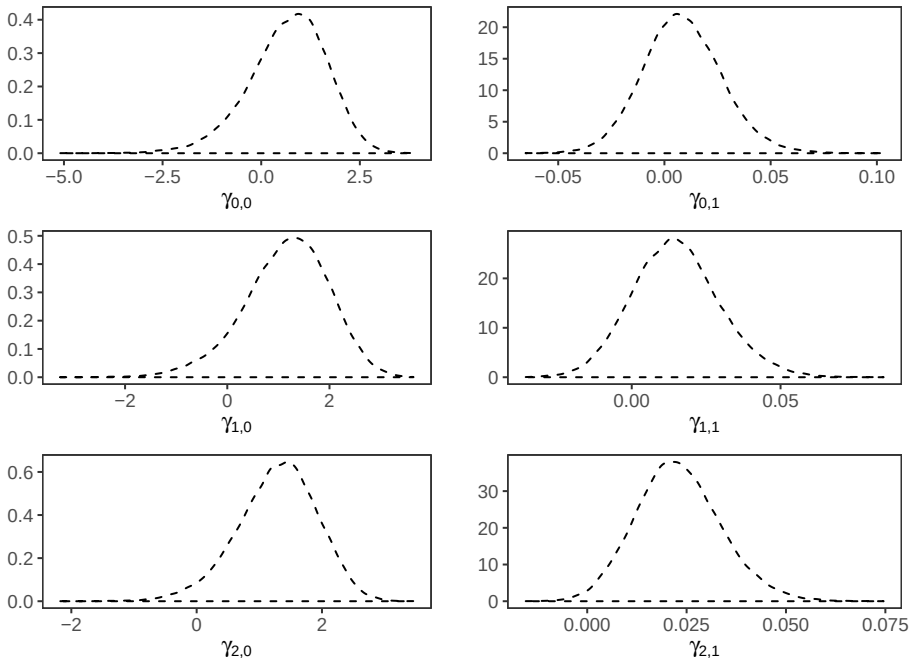

Figure 2.19: Simulated density plots for induced priors on $\gamma_{k,0}$ and $\gamma_{k,1}$.

Although we consider a single covariate, the conditional means prior approach described above can be extended to the case with multiple covariates. Moreover, we assume that the Poisson parameters and zero-inflation probability depend on the same covariate (age). The conditional means prior approach to a Bayesian BPZIP model can be extended to the case where the Poisson parameters and zero-inflation parameter depend on different covariates.

2.7 Implementing CMP Approach for a Bayesian BPZIP Regression Model

Within the context of the hypothetical adverse event study, suppose we have a well-informed expert. That is, the expert is reasonably accurate in his/her assessments of the modal value but also has some uncertainty in his/her assessments. Furthermore, the expert is equally confident in their assessments made for subjects in the range of 20-30 years old as they are in their assessments for subjects in the range of 60-70 years old.

For convenience we again present the information collected from the expert in Table 2.8 using the assessment tasks described in Section 2.6.4. For example, the expert believes that a subject aged 25 will most likely experience 6 nausea episodes and no more than 9. The information obtained about each of the Poisson parameters is translated into the parameters of a gamma distribution and, using numerical methods, the information obtained about the zero-inflated proportion is translated into the parameters of a beta distribution. Then, using the CMP approach, we induce priors on the regression coefficients. These densities are shown in Figure 2.17 and Figure 2.19.

Table 2.8: Summary of prior information collected from the expert.

Parameter	Mode	80th percentile	
$\lambda_{1,x_i=25}$	4	7	
$\lambda_{1,x_i=65}$	8	11	
$\lambda_{2,x_i=25}$	6	9	
$\lambda_{2,x_i=65}$	16	19	
$\lambda_{0,x_i=25}$	2	4	
$\lambda_{0,x_i=65}$	3	5	
Parameter	25th percentile	Median	75th percentile
p_{25}	0.62	0.70	0.78
p_{65}	0.22	0.30	0.38

2.7.1 Providing Feedback to Expert

Providing feedback to the expert about the implications of their prior assessments is an essential part of the prior elicitation process. This allows the expert to confirm whether or not the resulting prior distributions accurately reflect their prior beliefs and if necessary allows the opportunity for modifications of the prior structure. For example, we might show the expert the resulting prior densities shown in Figure 2.16 and Figure 2.18. Alternatively, we can use the prior predictive distribution as a tool to provide the feedback in a scale that is perhaps more meaningful

to the expert. Specifically, we can use the prior predictive distribution to generate hypothetical data that might result given their prior assessments. This can easily be implemented in OpenBUGS or JAGS. We might, for example, generate data from the prior predictive distribution corresponding to ages 25, 45, and 65, with a sample size of $n = 100$. Table 2.9 summarizes possible data that might result.

Table 2.9: Summary of hypothetical dataset generated from the prior predictive distribution for ages 25, 45, and 65 based on the expert's prior assessments.

Variable	Age	Mean	SD	Variance	Median	2.5%	97.5%
Y_1	25	2.43	4.31	18.57	0	0	14
Y_1	45	4.75	5.48	29.99	0	0	16
Y_1	65	8.78	6.93	48.10	10	0	22
Y_2	25	9.79	4.42	19.54	9	3	20
Y_2	45	13.60	4.48	20.08	13	6	23
Y_2	65	20.21	5.71	32.57	20	10	32

Figure 2.20 provides a scatterplot and histogram for plausible data that could result based on the expert's prior judgment. All three scatter plots suggest positive association between the number of migraines and number of nausea episodes. The bar plots for the number of migraines (Y_1) suggest a decrease in the proportion of subjects not at-risk to experience migraines as age increases. In addition, the bar plots indicate that the expert's assessments suggest an increase in the number of occurrences of migraines and nausea episodes as age increases. These hypothetical data allows the expert to evaluate and modify their prior assessments in a meaningful scale.

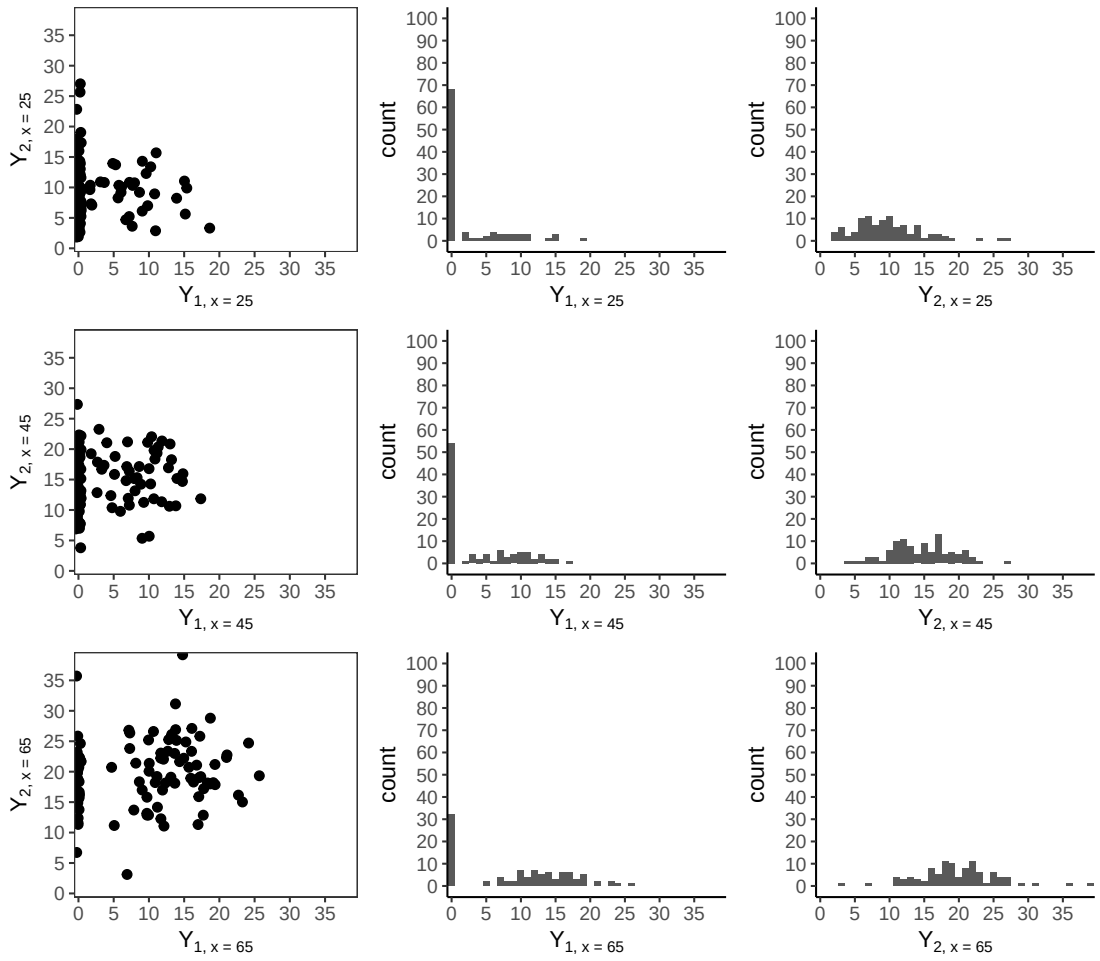

Figure 2.20: Simulated data based on prior predictive distribution for number of migraines and number of nausea episodes for $n = 100$ subjects age 25, 45, and 65 based on information collected from expert.

We can also provide the expert with the induced priors at several ages as another way to gauge the consequences of their prior assessments. The interval widths at priors for age 25 and age 65 in Figure 2.21 reflect that the expert is equally confident in his/her assessments at these ages. Figure 2.21 is also consistent with the histograms in Figure 2.20, which indicate that the expert believes the rate of experiencing adverse events increases with increasing age, while the probability of not being at risk to experience migraines decreases with increasing age.

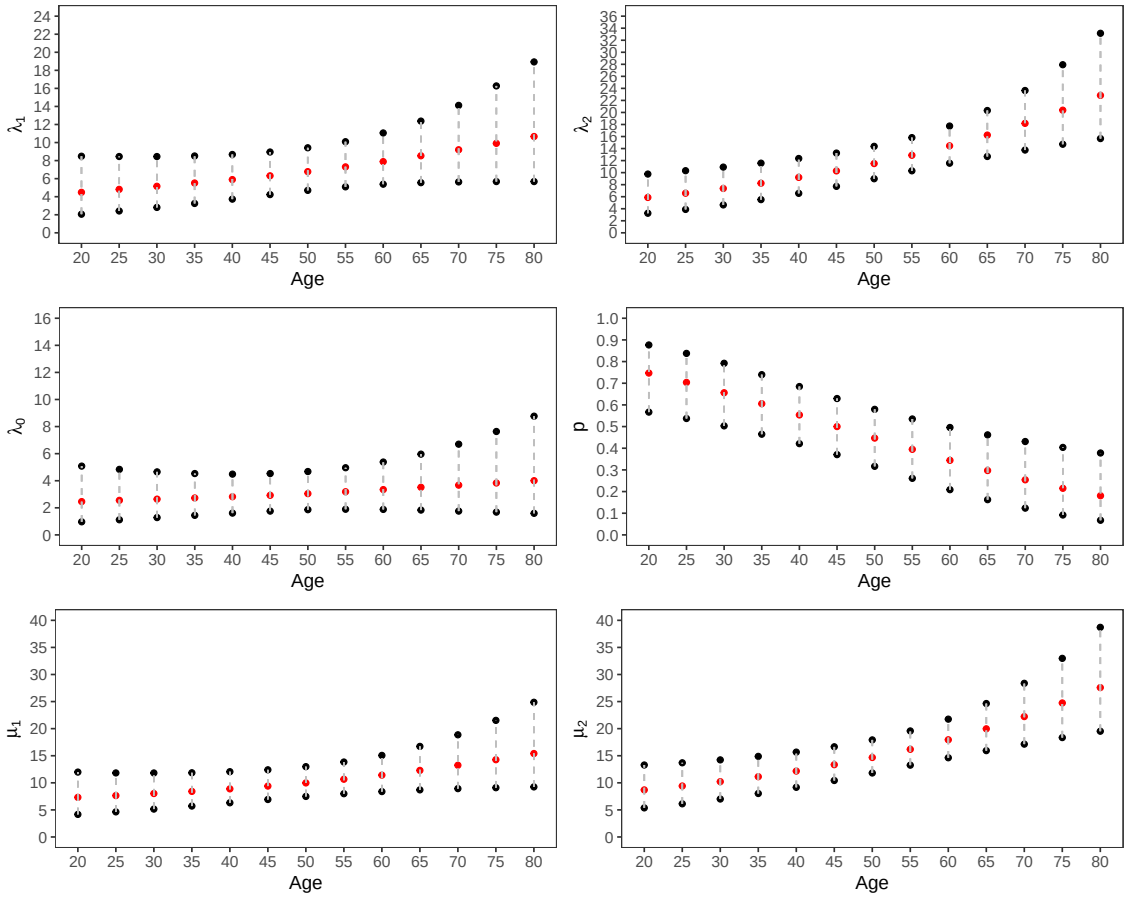

Figure 2.21: Prior median (red dot), 10th percentile and 90th percentile (black dots) based on experts knowledge.

2.7.2 Prior Effective Sample Size

Another way to assess the prior distribution is to understand the prior equivalent sample size (ESS) implied by our choice of prior. Morita et al. (2008) discuss the importance of the prior ESS. Denote n_e as the prior ESS. The information in the prior is equivalent to an independent sample of size n_e, that is independent of the *actual* sample of size n. This allows one to gauge the realism of the prior. For example, if n_e is inordinately large, one should be concerned about the prior having too great an influence on the posterior, relative to the observed data. This question is one that is sure to be asked in this context, particularly for the posteriors for quantities related to λ_0, λ_1, and λ_2 as the priors for these parameters will not

be updated as much compared to the priors of identified parameters. That is, the elicited priors for λ_0, λ_1, and λ_2 will dictate the resulting posteriors more so than the elicited priors for p. The prior effective sample size for the priors in Figure 2.16 and Figure 2.18 are provided in Table 2.10. All of these prior ESS values are quite small compared to the sample size of $n = 100$.

Table 2.10: Prior equivalent sample size.

Parameter	Prior ESS
$\lambda_{0,25}$	1
$\lambda_{0,65}$	1
$\lambda_{1,25}$	1
$\lambda_{1,65}$	1
$\lambda_{2,25}$	1
$\lambda_{2,65}$	1
p_{25}	15
p_{65}	15

2.7.3 Posterior Inference for Adverse Event Study

Suppose that the expert feels the prior distributions accurately reflect his/her beliefs. In this section we provide posterior inference for model parameters. Figure 2.22 provides the induced priors (specified indirectly via the CMP approach) and resulting posterior densities for the regression coefficients, as constructed in Section 2.6.4. The JAGS specifications are as described in Section 2.3.1. Standard diagnostics based on trace plots and the Gelman-Rubin statistic indicate no problems with convergence.

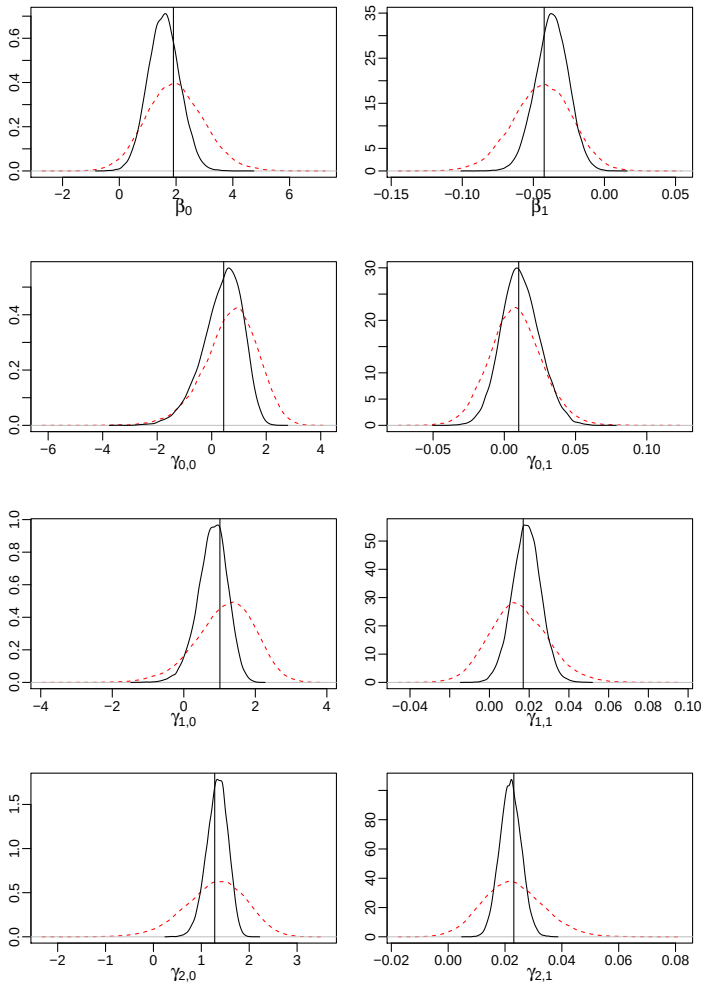

Figure 2.22: Prior (red, dashed) and posterior (black, solid) for regression coefficients. Here prior precision is independent of age. True values are indicated by vertical solid black line.

Table 2.11: Posterior results for regression coefficients with well-informed expert across all ages basic CMP.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
β_0	1.906	1.576	0.559	1.571	0.506	2.682	2.175
β_1	-0.042	-0.037	0.012	-0.036	-0.060	-0.015	0.045
$\gamma_{0,0}$	0.439	0.385	0.735	0.471	-1.272	1.580	2.851
$\gamma_{0,1}$	0.010	0.011	0.013	0.010	-0.013	0.039	0.053
$\gamma_{1,0}$	1.009	0.799	0.415	0.824	-0.088	1.542	1.631
$\gamma_{1,1}$	0.017	0.018	0.007	0.018	0.004	0.032	0.028
$\gamma_{2,0}$	1.279	1.328	0.220	1.340	0.867	1.723	0.855
$\gamma_{2,1}$	0.023	0.021	0.003	0.021	0.014	0.029	0.014

Note that the true value for each regression parameter is contained within their respective 95% credible interval. The posterior 95% credible interval for β_1 suggests that the odds that a subject is at-risk to experience migraines is between $e^{20(0.0149)} = 1.35$ and $e^{20(0.0600)} = 3.32$ times greater for a 20-year increase in age. The posterior density and corresponding 95% credible interval for $\gamma_{1,1}$ suggest that for a 20-year increase in age, the number of migraines increases by a factor between $e^{20(0.0044)} = 1.09$ and $e^{20(0.0328)} = 1.93$, with probability 0.95. Similarly, the posterior density and corresponding 95% credible interval for $\gamma_{2,1}$ suggest an that for a 20-year increase in age, the number of nausea episodes increases by a factor between 1.34 and 1.80, with probability 0.95.

Figure 2.23 shows the priors and posteriors for the zero-inflation probability and Poisson parameters at age 25, 45 and 65.

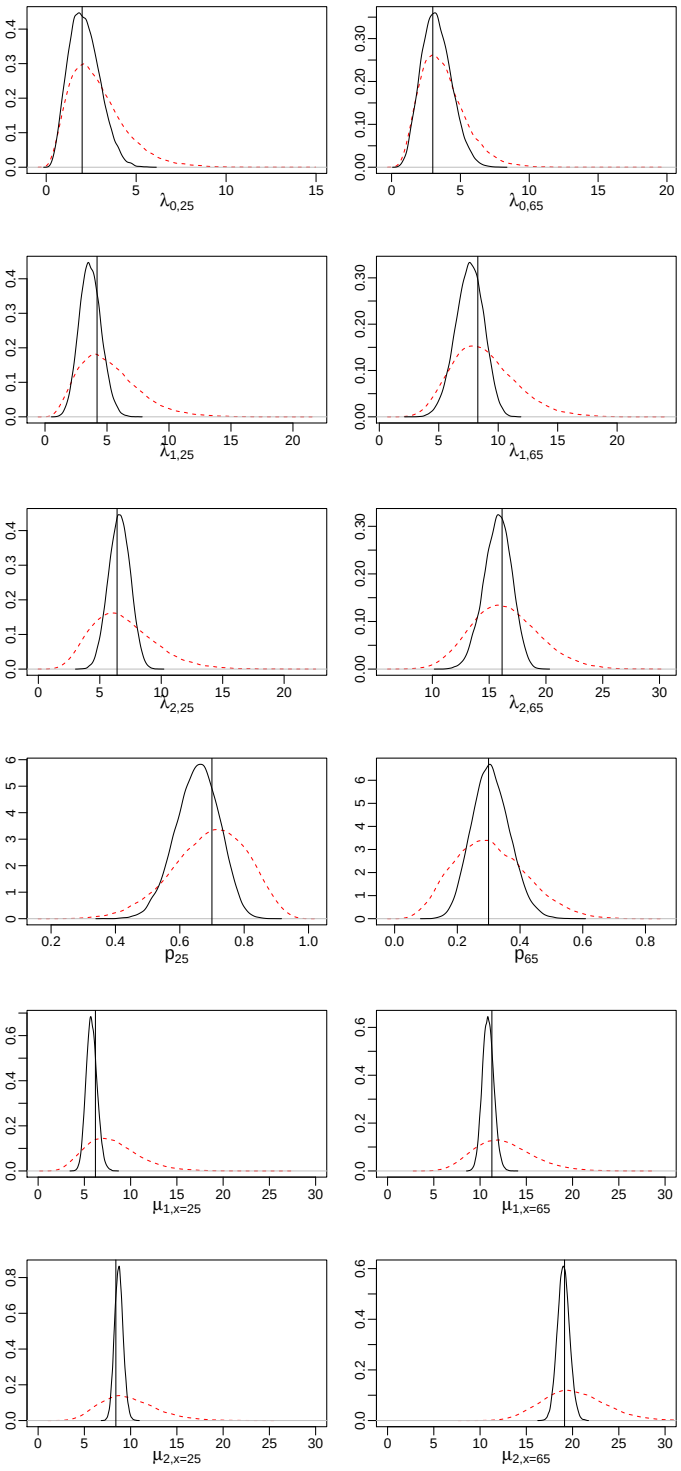

Figure 2.23: Prior (red, dashed) and posterior (black, solid) densities for BPZIP model parameters. Here prior precision is independent of age.

Table 2.12: Posterior results for model parameters with well-informed expert across all ages basic CMP.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
$\lambda_{0,25}$	1.998	2.137	0.850	2.068	0.706	3.957	3.251
$\lambda_{0,45}$	2.445	2.563	0.721	2.528	1.262	4.063	2.801
$\lambda_{0,65}$	2.992	3.279	1.085	3.211	1.375	5.578	4.202
$\lambda_{1,25}$	4.197	3.660	0.881	3.628	2.029	5.459	3.430
$\lambda_{1,45}$	5.896	5.220	0.786	5.235	3.635	6.709	3.074
$\lambda_{1,65}$	8.284	7.591	1.178	7.623	5.184	9.793	4.609
$\lambda_{2,25}$	6.405	6.589	0.868	6.600	4.878	8.253	3.375
$\lambda_{2,45}$	10.167	10.155	0.811	10.188	8.478	11.659	3.181
$\lambda_{2,65}$	16.138	15.738	1.220	15.785	13.245	18.000	4.754
$\mu_{1,25}$	6.195	5.798	0.592	5.771	4.711	7.017	2.306
$\mu_{1,45}$	8.342	7.784	0.425	7.776	6.967	8.634	1.666
$\mu_{1,65}$	11.277	10.870	0.612	10.855	9.710	12.111	2.400
$\mu_{2,25}$	8.403	8.727	0.468	8.719	7.847	9.681	1.834
$\mu_{2,45}$	12.612	12.719	0.384	12.721	11.966	13.483	1.517
$\mu_{2,65}$	19.131	19.017	0.654	19.011	17.751	20.332	2.580
p_{25}	0.699	0.654	0.066	0.656	0.517	0.777	0.260
p_{45}	0.499	0.478	0.045	0.478	0.390	0.568	0.177
p_{65}	0.299	0.307	0.060	0.305	0.197	0.431	0.234

The posterior densities shown in Figure 2.23 show more updating *a posteriori* for μ_1 and μ_2 compared to λ_0, λ_1 and λ_2. This is not surprising as the data informs the model about μ_1 and μ_2, but does not inform the model about the individual summands, λ_0, λ_1 and λ_2. The nonoverlapping 95% credible intervals for μ_1 and μ_2 for $x_i = 25, 45$, and 65 suggest that, among those at-risk for migraines, the rate of experiencing migraines increases. This is similarly seen for the rate of experiencing nausea episodes. All of these inferences are consistent with the model we used to generate the data.

2.7.4 Critical Safety Factor

In practice, it is often of interest to compare a new treatment to the current standard of care. Suppose previous studies have shown that, for the current standard of care,

$$\Pr\left(Y_1 + Y_2 > 25 | \text{ Experience both Events}\right) \leq 0.30,$$

where $Y_1 + Y_2$ denotes total number of migraines and nausea episodes experienced by a subject. We can use the posterior predictive distribution to assess this probability for the hypothetical drug.

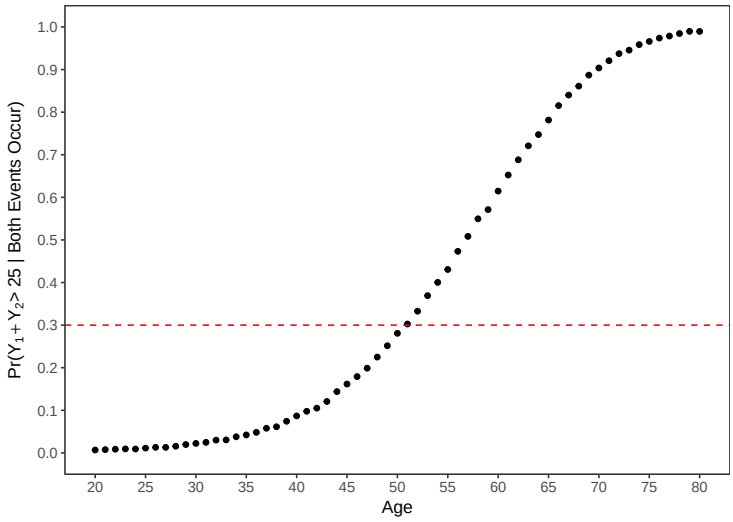

Figure 2.24: Posterior predictive distribution of the probability that the number of adverse events for a subject age x_i exceeds 25, given the subject experiences both events, given the data for the current study.

Figure 2.24 is the probability that the number of adverse events for a subject age x_i exceeds 25, given the subject experiences both adverse events with respect to the posterior predictive distribution given by 2.16. Figure 2.24 suggests that the hypothetical drug might be superior to current standard of care with respect to this metric for subjects less than or equal to 50 years-old, but that the current standard of care is superior for ages greater than 50.

2.7.5 *Another Illustration: Posterior Inference Based on an Expert Unfamilar with Older Ages*

As another illustration, suppose we ask the expert for the ages about which they are *most* comfortable making assessments about and for the ages about which they are *least* comfortable making assessments about. In response, the expert says that they are most familiar with individuals age 20-30 years-old and least familiar with individuals over the age of 60. That is, the expert is increasingly uncertain with increasing age. Suppose elicitation from this expert yields the information in Table 2.13.

Table 2.13: Prior information collected from expert that is more uncertain in assessments at age 65 compared to assessments at age 25.

Parameter	Mode	80th Percentile	
$\lambda_{1,x_i=25}$	4	7	
$\lambda_{1,x_i=65}$	8	16	
$\lambda_{2,x_i=25}$	6	9	
$\lambda_{2,x_i=65}$	15	25	
$\lambda_{0,x_i=25}$	2	4	
$\lambda_{0,x_i=65}$	3	6	
Parameter	25th Percentile	Median	75th Percentile
p_{25}	0.62	0.70	0.78
p_{65}	0.15	0.30	0.45

Note that the elicited percentiles at age 65 are wider than those in Table 2.8, representing the expert's increased uncertainty at older ages. This information collected from the expert suggests the prior information at certain ages shown in Figure 2.25. The increased prior interval widths at older ages in Figure 2.25 further illustrate the expert's increased uncertainty with increasing age. Moreover, the prior densities in Figure 2.26 demonstrate the effect of the expert's increasing uncertainty with age on the induced priors specified on the regression coefficients. Based on the information collected from the expert we again make use of the prior predictive

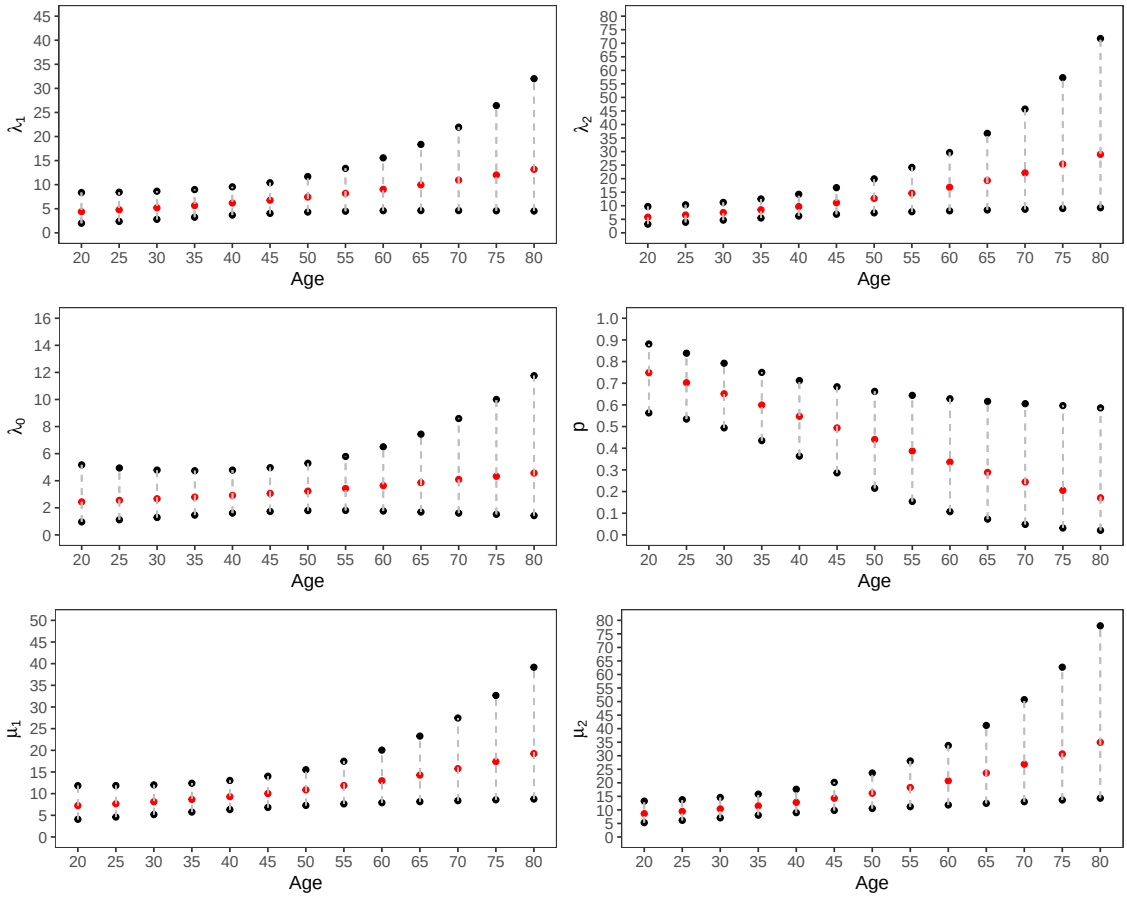

Figure 2.25: Prior median (red dot) and 10th percentile and 90th percentile (black dots) based on information collected from an expert.

distribution to generate a possible data set that could result based on the expert's judgment. For our purposes we assume we have presented feedback as in Section 2.7.1 and that the expert is satisfied with the resulting priors. Posterior inference was implemented in JAGS with the specifications outlined in Section 2.3.1. Standard diagnostic tests based on trace plots and the Gelman-Rubin statistic indicate convergence for all parameters.

Note that true values of parameters are again contained within their respective 95% credible intervals. Furthermore, the increased uncertainty of the expert at age 65 is reflected in the increased width of the 95% posterior credible intervals at age 65 in Table 2.15 compared to the respective credible intervals in Table 2.12. However,

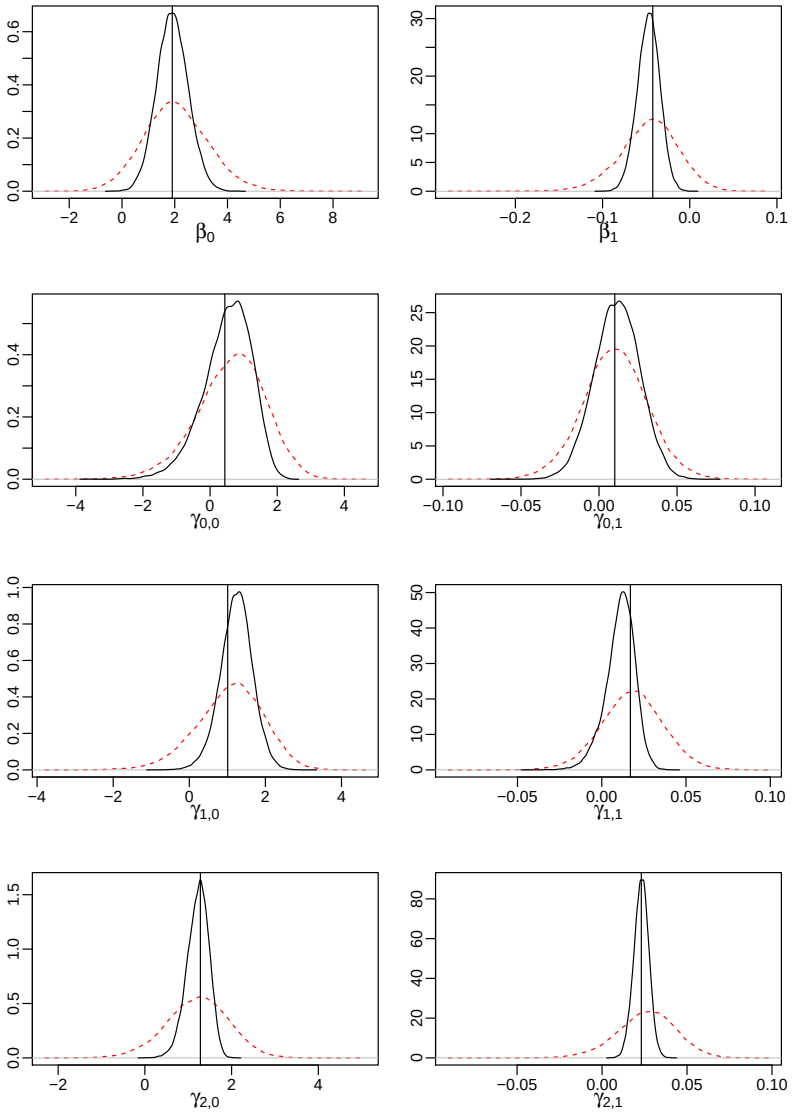

Figure 2.26: Prior (red, dashed) and posterior (black, solid) densities for regression coefficients. Here prior precision is not independent of age.

we did not observe increased interval width for μ_1 or μ_2, as expected.

2.8 Summary

We have discussed a Bayesian approach to a bivariate partial zero-inflated Poisson model. We have shown how relatively informative priors can mitigate the effects nonidentifiability can cause. We have demonstrated how in nonidentifiable settings the prior distribution can be used as a tool to identify parts of the parameter

Table 2.14: Posterior results for regression coefficients based on priors obtained from an expert that is less confident making assessments for older ages.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
β_0	1.906	2.087	0.603	2.077	0.949	3.301	2.351
β_1	-0.042	-0.039	0.012	-0.038	-0.064	-0.015	0.048
$\gamma_{0,0}$	0.439	0.591	0.680	0.653	-0.914	1.733	2.647
$\gamma_{0,1}$	0.010	0.010	0.013	0.011	-0.017	0.035	0.053
$\gamma_{1,0}$	1.009	1.319	0.398	1.335	0.481	2.058	1.576
$\gamma_{1,1}$	0.017	0.011	0.008	0.011	-0.007	0.025	0.032
$\gamma_{2,0}$	1.279	1.366	0.246	1.383	0.837	1.803	0.966
$\gamma_{2,1}$	0.023	0.019	0.004	0.019	0.010	0.028	0.017

Table 2.15: Posterior results for model parameters based on priors obtained from an expert that is less confident making assessments for older ages.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
$\lambda_{0,25}$	1.998	2.377	0.899	2.308	0.834	4.289	3.455
$\lambda_{0,45}$	2.445	2.971	0.975	2.908	1.273	5.039	3.766
$\lambda_{0,65}$	2.992	3.945	1.629	3.796	1.233	7.488	6.255
$\lambda_{1,25}$	4.197	4.945	1.075	4.922	2.908	7.084	4.175
$\lambda_{1,45}$	5.896	6.247	1.052	6.289	4.097	8.173	4.075
$\lambda_{1,65}$	8.284	8.058	1.730	8.168	4.433	11.119	6.686
$\lambda_{2,25}$	6.405	5.786	0.890	5.808	4.000	7.468	3.468
$\lambda_{2,45}$	10.167	9.436	1.026	9.490	7.285	11.277	3.991
$\lambda_{2,65}$	16.138	15.492	1.745	15.601	11.779	18.520	6.740
p_{25}	0.699	0.682	0.058	0.684	0.561	0.790	0.229
p_{45}	0.499	0.517	0.045	0.517	0.427	0.605	0.177
p_{65}	0.299	0.348	0.067	0.346	0.223	0.487	0.264

space that are not covered by the likelihood, namely for the unobserved Poisson parameters λ_0, λ_1 and λ_2. We also demonstrated that use of such informative priors alleviates convergence issues often present when using a diffuse prior structure for model parameters. We proposed and described methods of prior elicitation for a BPZIP model within the context of a hypothetical adverse event drug safety study. Finally, we propose a tool by which to provide feedback to the expert to illustrate the the implications of their prior structure. Namely, in Section 2.7.1, we demonstrated

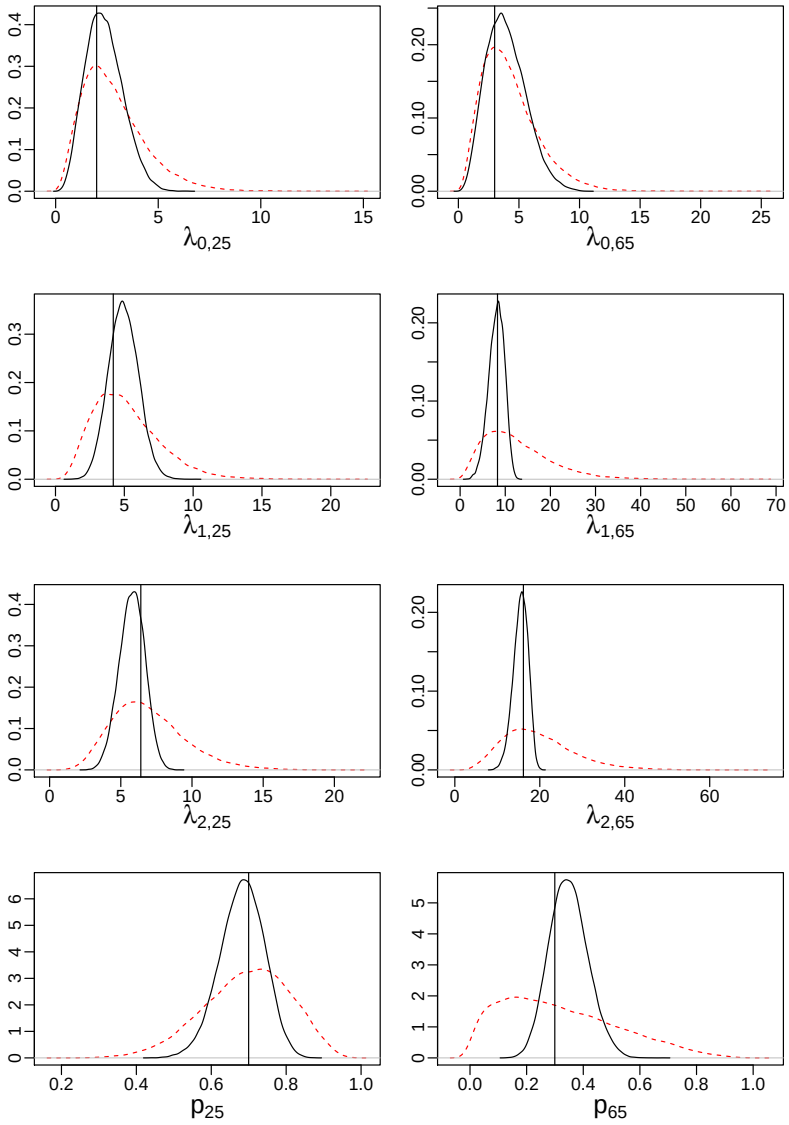

Figure 2.27: Prior (red, dashed) and posterior (black, solid) for BPZIP model parameters. Here prior precision is not independent of age.

how we can use the prior predictive distribution as a tool to provide feedback to the expert by showing them the implications of their prior assessments in a meaningful scale.

We applied the conditional means prior approach to a BPZIP regression model and described methods of prior elicitation from a subject-matter expert within the context of a hypothetical adverse event drug safety study. We discussed prior as-

sessment questions for subject-matter experts about operational quantities, such as the rate of experiencing just one or the other adverse event and the proportion of subjects not at-risk to experience an adverse event (specifically, the adverse event denoted by Y_1) at different covariate values.

Throughout this discussion, we assumed a highly informative prior on the parameter λ_0. The parameter, λ_0 is an association parameter, and represents a measure of dependency between the two bivariate outcomes. Within the context of the hypothetical adverse event example, prior elicitation for this parameter requires prompting an expert to think about the simultaneous rate of experiencing both adverse events, a quantity that is difficult for an expert to make assessments about. We return to this difficult issue in Chapter Three.

CHAPTER THREE

Prior Elicitation for Bayesian Conditional Partial Zero-inflated Poisson Models with Conditional Probabilities

In Chapter Two, we introduced a bivariate partial zero-inflated Poisson (BPZIP) model which assumes one count is zero-inflated and one count is not zero-inflated (see (2.2) in Section 2.2). We demonstrated that diffuse priors can be problematic and offer nonidentifiability as the culprit for lack of convergence for the Poisson parameters. We proposed informative prior structures to mitigate the effects caused by nonidentifiability. Finally, we suggested plausible methods of prior elicitation for BPZIP model parameters, with the exception of the association parameter λ_0. Prior information on association parameters can be difficult to elicit directly because they are rarely on a scale familiar to the expert. For example, within the context of the adverse event example described in Section 2.3, prior elicitation for this parameter requires prompting an expert to think about the simultaneous rate of experiencing both adverse events. The focus of this chapter is to provide a plausible route of prior specification for the association parameter λ_0.

A critical aspect of prior elicitation is that the assessment tasks involve prompting the expert about their expectations for conceivably observable quantities. One alternative is to specify a power prior on λ_0 based on historical data. Another option could be to design and conduct a study that would provide information about the simultaneous occurrence of Y_1 and Y_2. For example, within the context of the adverse event study in Section 2.3, if we observed the count for the number of migraines accompanied by a nausea episode for the ith subject, estimation for λ_0 would be straightforward. Recall the standard bivariate Poisson distribution,

$(Y_1, Y_2) \sim \text{BP}(\lambda_1, \lambda_2, \lambda_0)$, such that

$$Y_1 = X_1 + X_0, \tag{3.1}$$

and

$$Y_2 = X_2 + X_0, \tag{3.2}$$

where X_1, X_2 and X_0 are independent Poisson random variables with positive means λ_1, λ_2, and λ_0, respectively. In context of the adverse event study, X_1 represents the count for *exclusive* migraines, X_2 represents the count for *exclusive* nausea episodes and X_0 represents the count for having a migraine and being nauseated. These counts, X_0, X_1 and X_2, are not observed. The underlying subject-level contingency table shown in Table 3.1 shows the data that we would design the study to collect.

Table 3.1: Underlying counts for the ith subject for a BPZIP model.

(Y_1, Y_2)	$Y_2 = 0$	$Y_2 > 0$	$f_{Y_1}(y_1)$
$Y_1 = 0$	n_{00}	n_{02}	$n_{00} + n_{02}$
$Y_1 > 0$	n_{10}	n_{12}	$n_{10} + n_{12}$
$f_{Y_2}(y_2)$	$n_{00} + n_{10}$	$n_{02} + n_{12}$	N_i

Here, N_i denotes the total count for subject i. From this table, n_{12} is the count that represents the number of times subject i had a migraine and was also nauseated. It is this count that could be used to specify a prior on λ_0 for a subsequent trial.[1] In the absence of such prior data, informative priors must be constructed using expert opinion. Thus, in what follows we propose an alternative prior structure that allows for direct elicitation about a conditional probability, a quantity that is in a scale familiar to the expert. The expert's knowledge about this quantity is then used to specify a prior on λ_0.

[1] If we in fact had this individual level data then it would make estimation of the parameters λ_1, λ_2 and λ_0 straightforward. For example, the count n_{10} would inform X_1 in the bivariate Poisson (BP) (see Appendix A), the count n_{02} would inform X_2 in the BP and the count n_{12} would inform X_0. However, we assume throughout that we only observe $Y_1 = X_1 + X_0$ and $Y_2 = X_2 + X_0$.

The organization of this chapter is as follows: In Section 3.1 we develop a conditional representation of the joint probability mass function for a Bayesian bivariate partial zero-inflated Poisson model. We extend this representation to the regression case in Section 3.2. In Section 3.3 we discuss informative prior construction for the conditional representation of the BPZIP model, with an emphasis on prior construction for an association parameter. In Section 3.4, we discuss a conditional means prior approach for the conditional representation of the BPZIP model. We describe prior elicitation from a subject-matter expert within the context of an adverse event drug safety study. We offer use of the prior predictive distribution as a tool to provide feedback to the subject matter expert, and propose the posterior predictive distribution as a tool for a clinician. Concluding comments are given in Section 3.5.

3.1 A Bayesian Conditional Bivariate Partial Zero-inflated Poisson Model

AlMuhayfith et al. (2015) propose a conditional method for estimating the parameters of standard bivariate Poisson and zero-inflated bivariate Poisson regression models within a frequentist framework. This method involves the use of conditional probability theory to represent the joint probability mass function, $f_{Y_1,Y_2}(y_1, y_2)$, as the product of the marginal and conditional distribution. We can, of course, write

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_2 \,|\, Y_1}(y_2 \,|\, y_1) \times f_{Y_1}(y_1) \tag{3.3}$$

or

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1 \,|\, Y_2}(y_1 \,|\, y_2) \times f_{Y_2}(y_2). \tag{3.4}$$

AlMuhayfith et al. (2015) show that for both a standard bivariate Poisson (BP) and a bivariate zero-inflated Poisson (BZIP) regression model, the conditional method yields almost identical model performance (based on Akaike Information Criterion) and parameter estimates compared to the standard method of using the joint distribution. For their purposes, AlMuhayfith et al. (2015) recommend the conditional

method for inference of a standard BP or BZIP regression model as the conditional method is less computationally intensive compared to the standard method.

We extend the conditional representation to the Bayesian bivariate partial zero-inflated Poisson model. Namely, we rewrite the joint probability mass function of (Y_1, Y_2) as a product of the conditional distribution and the marginal distribution. Our motivation for representing the BPZIP regression model as a product of a conditional distribution and a marginal distribution is that it provides a plausible route of elicitation to specify a prior on λ_0 via direct elicitation of an conceivably observable quantity.

Suppose we have a bivariate count vector $(Y_1, Y_2) \sim \text{BPZIP}(p, \lambda_0, \lambda_1, \lambda_2)$, where Y_1 is zero-inflated and Y_2 is not zero-inflated as in (2.2). Then the joint probability mass function is given by (2.3). The conditional distribution of Y_2 given Y_1 for the BP part of (2.2) is[2]

$$f_{Y_2 \,|\, Y_1}(y_2 \,|\, y_1) = \sum_{m=0}^{\min(y_1, y_2)} \binom{y_1}{m} \theta^m (1-\theta)^{y_1 - m} \frac{e^{-\lambda_2} \lambda_2^{y_2 - m}}{(y_2 - m)!}, \tag{3.5}$$

where $\theta = \frac{\lambda_0}{\lambda_1 + \lambda_0}$. It can be shown that (Kocherlakota and Kocherlakota (1992)) the conditional distribution of $Y_2 \,|\, Y_1$ is a convolution of a Poisson random variable with parameter λ_2 and a binomial random variable with parameters (y_1, θ). Namely, let $W \equiv X_0 | Y_1$ and $T \equiv X_2$. Then $Z = W + T$ is the sum of two independent discrete random variables with corresponding probability distribution

$$\begin{aligned} \Pr(W + T = z) = f_Z(z) &= \sum_{m=0}^{z} f_W(m) f_T(z - x) \\ &= \sum_{m=0}^{\min(y_1, y_2)} \underbrace{\binom{y_1}{m} \left(\frac{\lambda_0}{\lambda_1 + \lambda_0}\right)^m \left(\frac{\lambda_1}{\lambda_1 + \lambda_0}\right)^{y_1 - m}}_{\text{Binomial}(y_1, \theta)} \underbrace{e^{-\lambda_2} \frac{\lambda_2^{y_2 - m}}{(y_2 - m)!}}_{\text{Poisson}(\lambda_2)}. \end{aligned}$$

Thus, Z is the convolution of a binomial random variable, $X_0 | Y_1$, and a Poisson random variable, X_2, where X_0 and X_2 are latent variables from the BP representation

[2] Note that this conditional distribution applies to observations that are at-risk for Y_1 as a subject cannot experience Y_1 if they are not at-risk.

in (3.1) and (3.2). The conditional mean and variance are given by

$$\mathrm{E}(Y_2 \,|\, Y_1) = \lambda_2 + \theta y_1 \tag{3.6}$$

and

$$\mathrm{Var}(Y_2 \,|\, Y_1) = \lambda_2 + \theta(1-\theta) y_1. \tag{3.7}$$

Following (3.3), the conditional representation of the BP joint probability mass function is

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{Y_2 \,|\, Y_1}(y_2 \,|\, y_1) \times f_{Y_1}(y_1) \\ &= \sum_{m=0}^{\min(y_1,y_2)} \theta^m (1-\theta)^{y_1-m} \frac{e^{-\lambda_2} \lambda_2^{y_2-m}}{(y_2-m)!} \times \frac{e^{\mu_1} \mu_1^{y_1}}{y_1!} \\ &= f_{\mathrm{CBP}}(y_1, y_2 \,|\, \theta, \mu_1, \lambda_2), \end{aligned} \tag{3.8}$$

where $\mu_1 = \lambda_1 + \lambda_0$ and $\theta = \frac{\lambda_0}{\mu_1}$. Thus, the conditional representation of the BPZIP model in (2.2), $f_{\mathrm{CBPZIP}}(y_1, y_2 \,|\, p, \theta, \mu_1, \lambda_2) \equiv f_{\mathrm{CBPZIP}}(y_1, y_2)$, is given by

$$f_{\mathrm{CBPZIP}}(y_1, y_2) = \begin{cases} p + (1-p) f_{\mathrm{CBP}}(y_1 = 0, y_2 = 0 \,|\, \theta, \mu_1, \lambda_2), & y_1 = 0, y_2 = 0 \\ (1-p) f_{\mathrm{CBP}}(y_1 = y_1, y_2 = 0 \,|\, \theta, \mu_1, \lambda_2), & y_1 \neq 0, y_2 = 0, \\ p + (1-p) f_{\mathrm{CBP}}(y_1 = 0, y_2 = y_2 \,|\, \theta, \mu_1, \lambda_2), & y_1 = 0, y_2 \neq 0 \\ (1-p) f_{\mathrm{CBP}}(y_1 = y_1, y_2 = y_2 \,|\, \theta, \mu_1, \lambda_2), & y_1 \neq 0, y_2 \neq 0. \end{cases} \tag{3.9}$$

We use this representation of the CBPZIP throughout this chapter.[3] We write $(Y_1, Y_2) \sim \mathrm{CBPZIP}(p, \theta, \mu_1, \lambda_2)$.

One way in which the conditional representation could be used to construct a prior for λ_0 is via the conditional expectation in (3.6). In particular, we could artificially make y_1 a covariate and prompt an expert to suggest a most likely value

[3] Note that we could have also represented the BPZIP model using (3.4). For our purposes, we use (3.9) throughout.

(as well as a most extreme value) for the conditional probability, $Y_2|Y_1$, for a specific value of Y_1. As an illustration, within the context of the adverse event study described in Section 2.3, we might prompt an expert about their expectations for the probability that given a subject experienced 6 migraines, the subject also experienced nausea, $\Pr(Y_2|Y_1 = 6)$. To do this we would ask the expert "Suppose we have 100 subjects that have experienced $y_1 = 6$ migraines. What, do you think, is the percentage of these subjects that also experienced nausea?" Their response would be taken as the modal value. We then would ask for a most extreme value to assess the the expert's uncertainty. In practice, however, it is not ideal to ask an expert about their expectations for a specific value of the *response* (or dependent) variable.[4]

Another option, which is the focus of what follows, is use of the distribution $X_0 \,|\, Y_1 \sim \text{Binomial}(y_1, \theta)$. In particular, this representation suggests a plausible method in which we can use expert judgment about an elementary event to indirectly specify a prior on λ_0. By definition,

$$Y_1 \overset{iid}{\sim} \text{Bernoulli}(\theta). \tag{3.10}$$

Thus, θ is the probability associated with each independent Bernoulli event. Within the context of the adverse event drug safety trial, θ represents the conditional probability that, given a subject has a migraine, the subject is also nauseated. This suggests how we can indirectly specify a prior for λ_0 via direct elicitation on a conditional probability (which is in a scale familiar to an expert). From (3.10) we have that for each independent event, (subject has a migraine), the probability that the individual is also nauseated is θ. That is, θ represents the probability:

$$\Pr(\text{subject is nauseated} \,|\, \text{subject has a migraine}).$$

[4] Note this is *not* what is done in the CMP construction. There we ask the expert about their expectations for specific values of *independent* variables.

In this way, we have described an operational quantity about which we can directly elicit information. Perhaps most importantly, we have introduced a plausible way in which we can use the expectations of the expert about θ to inform λ_0.[5]

3.2 CBPZIP Regression Model

Let $(Y_{1i}, Y_{2i}) \sim \text{BPZIP}(p_i, \lambda_{0i}, \lambda_{1i}, \lambda_{2i})$ for $i = 1, \ldots, n$. We denote the conditional bivariate partial zero-inflated Poisson (CBPZIP) as $(Y_{1i}, Y_{2i}) \sim \text{CBPZIP}(p_i, \theta_i, \mu_{1i}, \lambda_{2i})$, where μ_{1i} denotes the marginal rate of experiencing Y_1 among those "at-risk" to experience Y_1, λ_{2i} is the rate associated with just Y_2, θ_i is the conditional probability that, given the ith subject experiences Y_1, they simultaneously experience Y_2, and p_i is the zero-inflation probability (associated with Y_1).

We can represent the CBPZIP model in such a way that the parameters depend on covariates in a generalized linear model. In particular, for the ith subject we have that

$$\begin{aligned} \log(\mu_{1,i}) &= \mathbf{x}_i^T \boldsymbol{\phi}, \\ \log(\lambda_{2,i}) &= \mathbf{z}_i^T \boldsymbol{\gamma}_2, \\ \text{logit}(\theta_i) &= \mathbf{v}_i^T \boldsymbol{\alpha}, \end{aligned} \tag{3.11}$$

and

$$\text{logit}(p_i) = \mathbf{w}_i^T \boldsymbol{\beta}.$$

where $\mathbf{x}_i$ and $\boldsymbol{\phi}$ are vectors of covariates and corresponding regression coefficients, respectively, for the marginal rate of Y_1, $\mathbf{z}_i$ and $\boldsymbol{\gamma}$ are vectors of covariates and corresponding regression coefficients, respectively, for the rate of just Y_2, $\mathbf{v}_i$ and $\boldsymbol{\alpha}$ are vectors of covariates and corresponding regression coefficients, respectively, for the conditional probability that, given the ith subject experiences Y_1, the ith subject

[5] We could easily have used the conditional distribution of $Y_1|Y_2$ to get at an informative prior for λ_0. In this case, we would elicit information about the probability that given a subject is nauseated, the subject also has a migraine (i.e. $\Pr(X_0|Y_2)$).

also is experiencing Y_2. Finally, $\mathbf{w}_i$ and $\boldsymbol{\beta}$ are vectors of covariates and corresponding regression coefficients for the zero-inflation parameter. This representation allows for different parameters to depend on different covariates. It follows that,

$$\lambda_{0,i} = \theta_i \times \mu_{1,i}, \tag{3.12}$$

$$\lambda_{1,i} = \mu_{1,i} - \lambda_{0,i}, \tag{3.13}$$

and

$$\mu_{2,i} = \lambda_{2,i} + \lambda_{0,i}. \tag{3.14}$$

We provide the a summary of the model parameters within the context of our adverse event example in Table 3.3 in Section 3.4. Let $\mathbf{y}_i = (y_{1i}, y_{2i})$, $i = 1, \ldots, n$. The CBPZIP regression likelihood function is given by

$$\begin{aligned}
\ell(\boldsymbol{\phi}, \boldsymbol{\gamma}_2, \boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}_i) = & \prod_{i=1}^{n} \left[f_{\text{CBPZIP}}(y_{1i} = 0, y_{2i} = 0 \,|\, p_i, \theta_i, \mu_{1,i}, \lambda_{2,i}) \right]^{I_{i1}} \\
& \times \prod_{i=1}^{n} \left[f_{\text{CBPZIP}}(y_{1i} = y_1, y_{2i} = 0 \,|\, p_i, \theta_i, \mu_{1,i}, \lambda_{2,i}) \right]^{I_{i2}} \\
& \times \prod_{i=1}^{n} \left[f_{\text{CBPZIP}}(y_{1i} = 0, y_{2i} = y_2 \mid p_i, \theta_i, \mu_{1,i}, \lambda_{2,i}) \right]^{I_{i3}} \\
& \times \prod_{i=1}^{n} \left[f_{\text{CBPZIP}}(y_{1i} = y_1, y_{2i} = y_2 \,|\, p_i, \theta_i, \mu_{1,i}, \lambda_{2,i}) \right]^{I_{i4}},
\end{aligned} \tag{3.15}$$

where

$$\begin{aligned}
p_i &= \frac{\exp(\mathbf{w}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{w}_i^T \boldsymbol{\beta})}, \\
\lambda_{2,i} &= \exp(\mathbf{z}_i^T \boldsymbol{\gamma}_2), \\
\theta_i &= \frac{\exp(\mathbf{v}_i^T \alpha)}{1 + \exp(\mathbf{v}_i^T \boldsymbol{\alpha})}, \\
\mu_{1,i} &= \exp(\mathbf{x}_i^T \boldsymbol{\phi}),
\end{aligned}$$

and I_{it}, $t = 1, \ldots, 4$, is an indicator function defined in (2.5). For our purposes we use the conditional representation as an alternative prior structure for data generated the BPZIP model introduced in Chapter Two.

The CBPZIP regression model assumes that the Poisson parameters $\lambda_{2,i}$ and $\mu_{1,i}$ depend on covariates in a generalized linear model, and inference for the Poisson parameters $\lambda_{0,i}$ and $\lambda_{1,i}$ are induced via (3.12) and (3.13). This is different from the BPZIP regression model in Section 2.2.1, which assumes the Poisson parameters $\lambda_{k,i}$, $k = 0, 1, 2$ depend on covariates in a generalized linear model. Thus, when comparing inference for the two representations of the BPZIP model, it is reasonable to compare the regression coefficients for λ_2 and p.

3.3 An Alternate Prior Structure for a Bayesian BPZIP Model

The conditional representation of the BPZIP model provides an alternative method of prior construction for data that are assumed to follow a bivariate partial zero-inflated Poisson distribution (see (2.2)). Note that the conditional representation does not introduce constraints that make the Poisson parameters, λ_k, $k = 0, 1, 2$ identifiable. Unsurprisingly, our conditional formulation does not render the λ_k's identifiable. Hence, based on what we observed the BPZIP model, informative priors are still needed.

3.3.1 Application: Adverse Event Drug Safety Study

Consider again the hypothetical adverse event study introduced in Section 2.3. Let (Y_{1i}, Y_{2i}) be a bivariate response count for the i^{th} subject such that

$$Y_{1i} = \text{Number of migraines experienced during study}$$

$$Y_{2i} = \text{Number of nausea episodes experienced during study,}$$

for $i = 1, \ldots, n$, where n is the number of subjects in the study. A scatterplot and histogram of the bivariate outcomes are presented again in Figure 3.1.

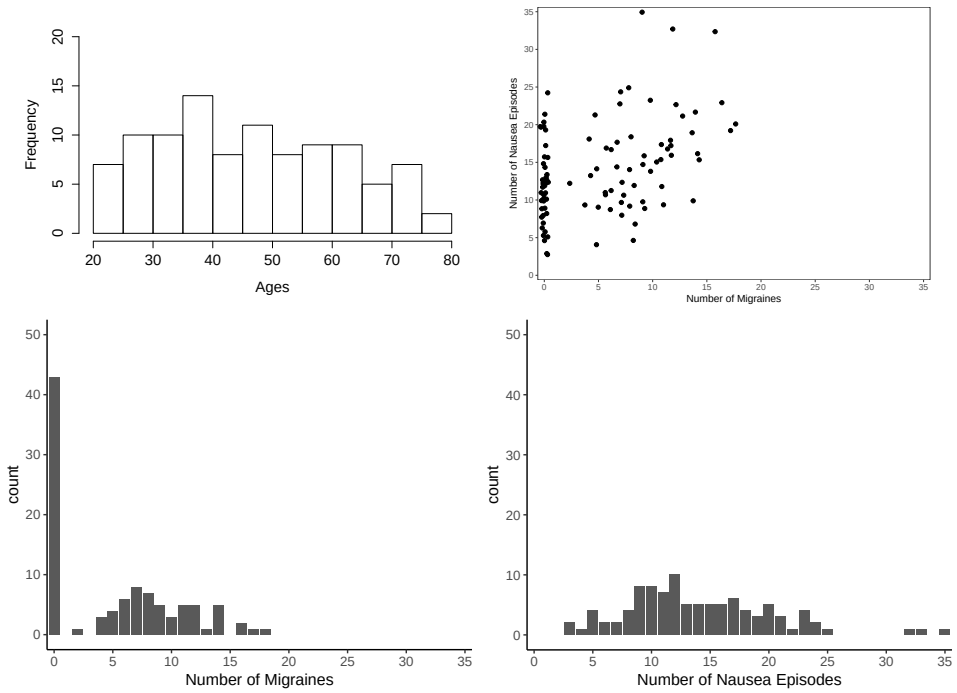

Figure 3.1: Ages and marginal data for counts Y_1 and Y_2 for $n = 100$ subjects.

3.3.2 Prior Construction for CBPZIP Model: Non-regression Case

As in Chapter Two we first describe the assessment tasks for prior construction for the CBPZIP model in the non-regression case and then extend to the regression case in Section 3.4. The conditional representation of the BPZIP model require priors for μ_1, λ_2, θ, and p. Priors for λ_0, λ_1, and μ_2 are subsequently induced using (3.12), (3.13) and (3.16), respectively. Within the context of the hypothetical adverse event study, μ_1 represents the rate of experiencing migraines among those at-risk for migraines, λ_2 represents the rate of experiencing just nausea, p represents the proportion of subjects not at-risk to experience migraines (the zero-inflation parameter), and θ is the conditional probability that given a subject has a migraine, the subject is also (simultaneously) nauseated. For purposes of illustration, we first assume that model parameters do not depend on covariates.

We use the mode-percentile method of elicitation to translate information obtained from an expert into informative prior distributions on μ_{Y_1} and λ_2. To elicit information about μ_1, the expert is prompted to suggest a most likely value and most extreme value for the number of migraines experienced by subjects at-risk for migraines over the course of the study. Similarly, to elicit information about λ_2, the expert is prompted to suggest a most likely value and most extreme value for the number of exclusive nausea episodes. This information is then translated into the parameters of a gamma distribution for μ_1 and λ_2. Next, to assess the expert's judgement about p, the proportion of subjects not at-risk to experience migraines, the expert is asked a series of three questions to obtain a median, upper and lower bound as described in Chapter Two. This information is then translated into the parameters of a beta distribution that represents the expert's knowledge and uncertainty about the zero-inflation probability. Finally, we adapt the assessment tasks proposed in Elfadaly and Garthwaite (2013a) to assess the expert's expectations about θ, the conditional probability that given a subject has a migraine, the subject is also nauseated. Using numerical methods, this information is translated into the parameters of a beta distribution on θ.

Finally, we induce a prior on λ_0, λ_1 and μ_2 using (3.12), (3.13), and (3.14), respectively. Note that because we are assuming informative priors for model parameters, theoretically, we need not be concerned about the induced priors on λ_0, λ_1 and μ_2. This is in contrast to a diffuse prior setting, in which we need be concerned with the priors induced on λ_0, λ_1 and μ_2.

The above information is summarized in Table 3.2, which includes parameter interpretations within the context of the hypothetical adverse event drug safety study, whether the prior is elicited or induced and the corresponding prior.

Table 3.2: Summary of model parameters and prior construction.

Parameter	Interpretation	Elicit/Induce	Prior
μ_1	The marginal rate of experiencing migraines (Y_1) among those at-risk to experience migraines	Elicit	Gamma(c_1, d_1)
μ_2	The marginal rate that a subject experiences nausea (Y_2).	Induce	$\mu_2 = \lambda_2 + \lambda_0$
θ	The conditional probability that given a subject experiences migraines (Y_1), the subject also experiences nausea (Y_2).	Elicit	Beta(a_θ, b_θ)
λ_0	The rate of simultaneously experiencing migraines (Y_1) and nausea (Y_2).	Induce	$\lambda_0 = \theta \times \mu_1$
λ_1	The rate of just experiencing migraines (Y_1)	Induce	$\lambda_1 = \mu_1 - \lambda_0$
λ_2	The rate of just experiencing nausea (Y_2)	Elicit	Gamma(c_2, d_2)
p	Proportion of subjects not at-risk to experience migraines	Elicit	Beta(a, b)

3.3.3 Example: Specification of Prior for λ_0 Via Direct Elicitation on Conditional Probability

In this section we demonstrate how prior information obtained from an expert about θ can be used to indirectly specify a prior on λ_0. Consider the hypothetical adverse event drug safety study described in Section 2.3. To assess the expert's expectations about θ, we adapt the assessment tasks proposed in Elfadaly and Garthwaite (2013a) to translate information obtained from an expert into informative distributions for θ. In particular, the expert is asked a series of three questions:

(1) To assess the median, the expert is asked "Suppose we have 200 at-risk subjects from the study population that have a migraine.[6] What, do you think, is the percentage of these subjects that are also nauseated?" Their response is taken to be the median.

(2) Next, to assess the lower quartile, the expert is asked "Suppose the percentage of subjects that have experienced a migraine and are also nauseated is actually *less than* your initial assessment. Given this information, what would you now estimate as the percentage?" Their response is taken to be the 25th percentile.

(3) Finally, to obtain an upper quartile, the expert is asked "Suppose the percentage of subjects that have a migraine and are also nauseated is actually *greater than* your initial assessment. Given this information, what would you now estimate as the percentage?" Their response is taken to be the 75th percentile.

Suppose we the expert relays 0.45, 0.50, and 0.55 as the 25th, 50th, and 75th percentile, respectively. That is, the expert believes that, 50% of subjects that have a migraine are also nauseated. Using numerical methods, this information collected from the expert is translated into the parameters of a beta distribution.

Next, we elicit information about μ_1, the marginal rate of experiencing migraines among subjects at-risk for migraines. We do this by prompting the expert to suggest a most likely value for the number of migraines experienced by subjects over the course of the study, as well as a most extreme value to represent their uncertainty. Similarly, we elicit information about λ_2 by assessing the expert's judgment about the number of nausea episodes. Finally, we elicit information about the pro-

[6] In this context,"at-risk" subjects refers to subjects that are at-risk to experience migraines, and hence are at-risk to experience both adverse events

portion of subjects not "at-risk" to experience migraines using the same assessment tasks outlined in Section 2.6.4.

The information directly obtained from the expert about θ (conditional probability of a subject experiencing a migraine and simultaneously being nauseated) and μ_1 can then be used to indirectly specify a prior on λ_0 via

$$\lambda_0 = \theta \times \mu_1.$$

Suppose the expert believes that the most likely value for the number of migraines among those at-risk for migraines is 6 (mode) and no more than 10 (80th percentile). The information collected from the expert about θ and μ_1 can be translated into the parameters of a beta and gamma distribution shown in Figure 3.2.

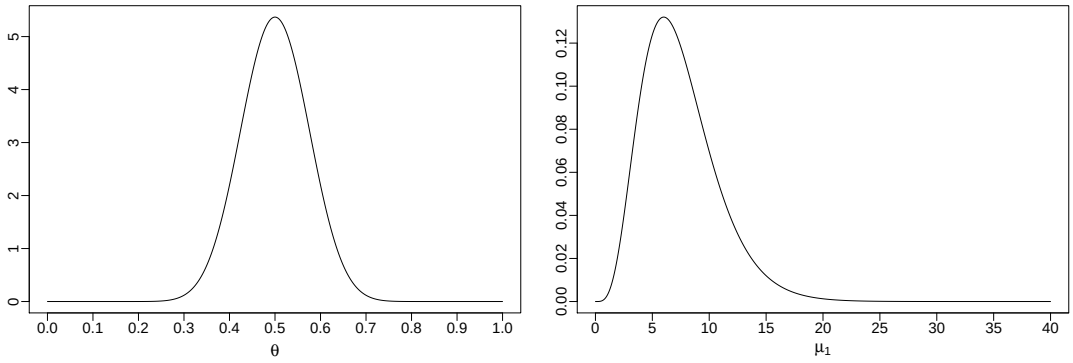

Figure 3.2: Prior distributions on θ and μ_1 based on information collected from expert. Namely, $\theta \sim \text{Beta}(22.88, 22.88)$ and $\mu_1 \sim \text{Gamma}(5.11, 1.46)$.

Using (3.12) we obtain the prior for λ_0 shown in Figure 3.3. The distribution in Figure 3.3 suggests that the expert believes the most likely value for the number of simultaneously experiencing both adverse events is about 3 and is most likely not less than 1 and no more than 7.[7]

[7] The prior distribution shown in Figure 3.3 is simulated. Neverthless, this distribution does have a closed form. Specifically, the prior for λ_0 is the product of a beta random variable and a gamma random variable. The distribution of this product has a closed form and is derived in Nadarajah and Kotz (2005).

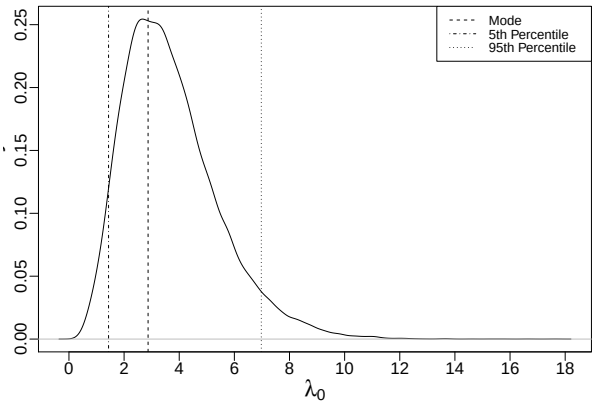

Figure 3.3: A prior distribution on λ_0.

3.4 Conditional Means Prior Approach for a Conditional BPZIP Model

We now consider informative prior construction for the regression case. We can represent the conditional bivariate partial zero-inflated Poisson model in such a way that the model parameters depend on covariates in a generalized linear model as in (3.11). Thus, the conditional representation of the BPZIP model introduced in this chapter lends itself to the conditional means prior approach introduced in Section 2.6.3. Per Section 2.6.3, the conditional means prior approach involves eliciting information from a subject-matter expert about conceivably observable quantities at covariate configurations that render sufficiently distinct responses. Priors for regression coefficients are then specified via indirect elicitation.

In the hypothetical adverse event drug safety study, we are interested in the joint modeling of two adverse events (migraines and nausea episodes) experienced by subjects enrolled in a study for a new drug. We assume that model parameters depend on a single covariate, age, x_i. In particular, for the ith subject we have that

$$\log(\mu_{1,i}) = \phi_0 + \phi_1 x_i,$$

$$\log(\lambda_{2,i}) = \gamma_{2,0} + \gamma_{2,1} x_i,$$

$$\text{logit}(\theta_i) = \alpha_0 + \alpha_1 x_i,$$

and

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i. \tag{3.16}$$

For our purposes, we choose $x_i = 25$ years old and $x_i = 65$ years old as the covariate configuration for which we elicit information. Here we are assuming that this age difference yields sufficiently distinct responses. Accordingly, the prior assessment tasks for parameters μ_1, λ_2, p and θ involve prompting the expert about expectations for subjects aged $x_i = 25$ and $x_i = 65$. Table 3.3 provides the model parameter summary for subject i and adverse event $j = 1, 2$. The conditional means prior approach for the conditional representation of the CBPZIP model involves eliciting information from subject-matter experts about μ_1, λ_2, θ, and p. Priors for regression coefficients are subsequently specified via indirect elicitation. Additionally, we induce priors on the Poisson parameters λ_0, λ_1 and μ_2 for the ith subject age x_i years old via (3.12), (3.13) and (3.14), respectively. Figure 3.4 provides a diagram illustrating the conditional means prior for a CBPZIP model as applied to the hypothetical adverse event drug safety study. Note differences from the diagram in Figure 2.15, which illustrates the CMP approach for the standard joint representation of the BPZIP model. Namely, the CMP approach in Figure 2.15 involves prior elicitation about the Poisson rates λ_1, λ_2 and λ_0, and p, and indirect specification of priors on regression coefficients for these parameters via the CMP approach. Priors are then induced on μ_1 and μ_2. Figure (3.4) suggests that we might potentially observe more posterior updating of λ_0 compared to that seen in Chapter Two as λ_0 is updated by μ_1 (a parameter the data informs the model about).[8]

3.4.1 CMP Approach to Alternate Prior Structure: Prior Assessment Tasks

The prior assessment tasks involve prompting the expert about expectations for subjects aged $x_i = 25$ and $x_i = 65$. We assume we have a well informed expert across all ages. We first consider elicitation of a conditional means prior for the

[8] See Appendix F for simulation results.

Table 3.3: Model parameter summary for the ith subject.

Parameter	Interpretation
p_i	The probability that the ith subject age x_i years-old is *not* at-risk to experience migraines.
$\mu_{1,i}$	The mean rate of experiencing migraines (among those at risk to experience migraines) for a subject age x_i years old.
$\mu_{2,i}$	The mean rate of experiencing a nausea episode for a subject age x_i years old.
θ_i	The conditional probability that among subjects at risk to experience migraines, given subject i has a migraine, subject i is also nauseated.
$\lambda_{1,i}$	The mean rate of of experiencing *just* a migraine (among those at risk to experience migraines) for a subject age x_i years old.
$\lambda_{2,i}$	The mean rate of experiencing *just* a nausea episode for a subject age x_i years old.
$\lambda_{0,i}$	The mean rate associated with simultaneously experiencing both migraines and nausea for a x_i-year-old subject.
α_0	If covariates are centered, the log odds that a subject aged $\bar{x} = 46$ that is at risk to experience migraines, has a migraine and is nauseated (simultaneously experiences both adverse events).
α_1	The log odds that a subject aged x_i that is at-risk to experience migraines, has a migraine and is nauseated; e^{α_1} represents the increased (or decreased) odds that a subject at-risk for migraines experiences a migraine and is nauseated for a one year increase in age.
β_0	If covariates are centered, the log odds that a subject aged $\bar{x} = 46$ is not at-risk to experience migraines.
β_1	The log odds that a subject aged x_i is not at-risk to experience migraines; e^{β_1} represents the increased (or decreased) odds that a subject is not at-risk for migraines for a one year increase in age.
ϕ_0	Average log adverse event rate for experiencing migraines for a subject aged $\bar{x} = 46$.
ϕ_1	Average change in log adverse event rate for experiencing migraines (among those at-risk to experience migraines) for a one-year increase in age; For a one-year increase in age, the expected number of migraines increases (decreases) by a factor of e^{ϕ_1}.
$\gamma_{2,0}$	Average log adverse event rate for just experiencing nausea for a subject aged $\bar{x} = 46$.
$\gamma_{2,1}$	Average change in log adverse event rate for experiencing just nausea for a one year increase in age; For a one-year increase in age, the expected number of exclusive nausea episodes increases (decreases) by a factor of $e^{\gamma_{2,1}}$.

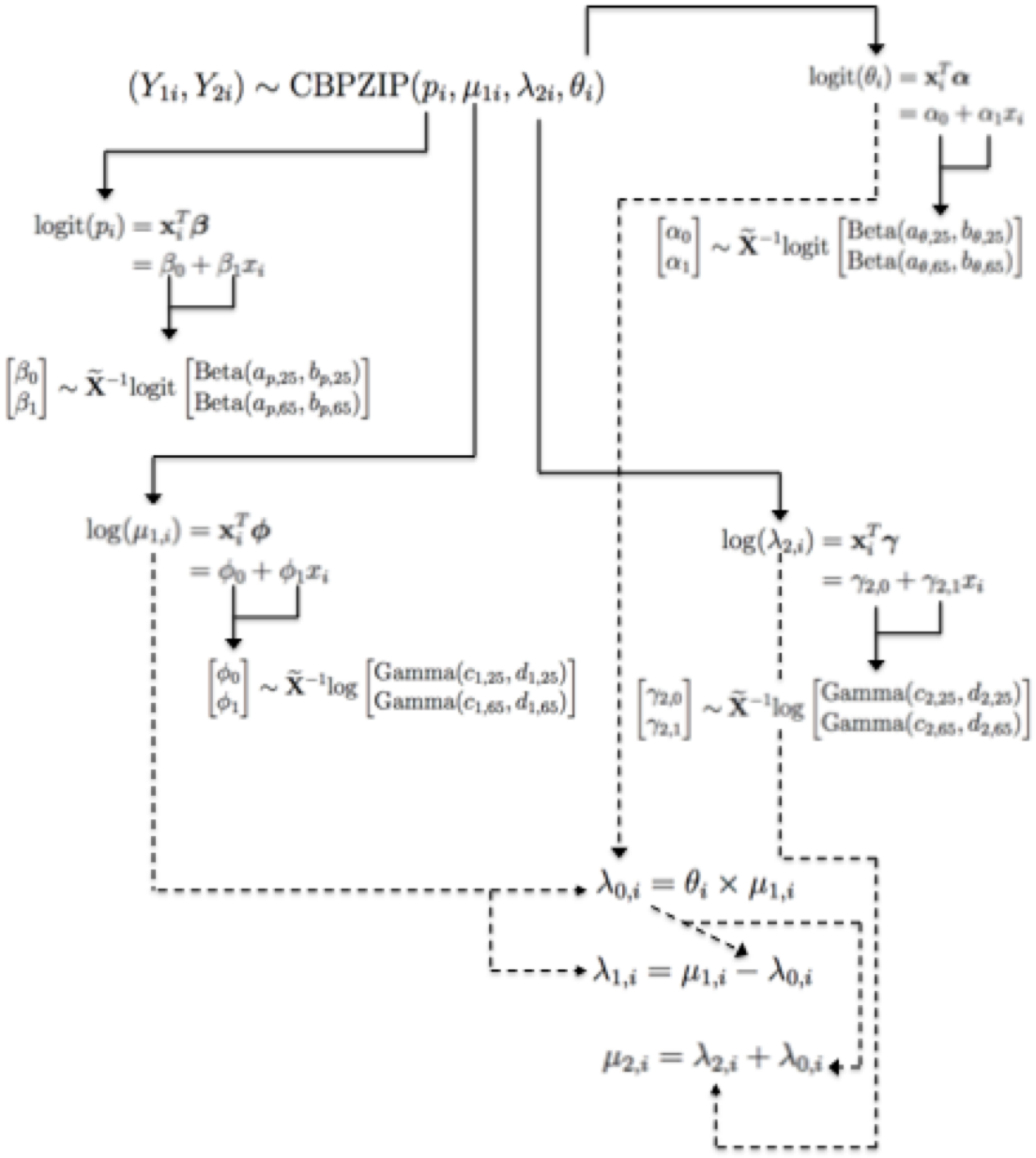

Figure 3.4: Bayesian conditional representation for a bivariate partial zero-inflated Poisson with conditional means priors in the context of the hypothetical adverse event study. Dashed lines to λ_0, λ_1 and μ_2 indicate that priors for these quantities are induced based on information elicited from the expert for other quantities.

Poisson regression used to model the number of adverse events represented by the bivariate response vector (Y_1, Y_2). To elicit information about μ_1, the rate of experiencing migraines, among those at-risk for migraines, at age $x_i = 25$ and $x_i = 65$, the expert is asked a series of questions to assess their knowledge and uncertainty:

(1) "Suppose we have subjects from the study population aged x_i that are at-risk to experience migraines. What, do you think, is the most likely value for the number of migraines experienced by these subjects over the course of the study?" Their response is taken to be the mode.

(2) Next, we prompt the expert to suggest an upper bound to assess their uncertainty. The expert is asked, "What, do you think, is the largest number the number of migraines can be among these x_i-year-old subjects?" Their response is taken to be the 80th percentile.

Next, we elicit information about λ_2 by assessing the expert's judgment about the number of nausea episodes. This is done using the same assessment tasks as in Section 2.6.4. For $x_i = 25$ and $x_i = 65$, the expert is asked a series of questions:

(1) "Suppose we have subjects from the study population aged x_i years-old. What, do you think, is the most likely value for the number of just (exclusive) nausea episodes experienced by subjects aged x_i over the course of the study?" Their response is taken to be the mode.

(2) Next, we prompt the expert to suggest an upper bound to represent their uncertainty. The expert is asked "What, do you think, is the largest the number of (exclusive) nausea episodes can be among subjects aged x_i?" Their response is taken to be the 80th percentile.

The information collected from the expert for the Poisson parameters is summarized in Table 3.4.

Table 3.4: Expert elicited information on Poisson parameters from well-informed expert across all ages.

Parameter	Mode	80th percentile
$\mu_{1,25}$	5	10
$\mu_{1,65}$	10	15
$\lambda_{2,25}$	6	10
$\lambda_{2,65}$	15	20

This information is translated into the parameters of a gamma distribution. In particular,

$$\mu_{1,25} \equiv \exp\left(\phi_0 + \phi_1(25)\right) \sim \text{Gamma}(c_1, d_1),$$

$$\mu_{1,65} \equiv \exp\left(\phi_0 + \phi_1(65)\right) \sim \text{Gamma}(c_2, d_2),$$

and

$$\lambda_{2,25} \equiv \exp\left(\gamma_{2,0} + \gamma_{2,1}(25)\right) \sim \text{Gamma}(c_3, d_3),$$

$$\lambda_{2,65} \equiv \exp\left(\gamma_{2,0} + \gamma_{2,1}(65)\right) \sim \text{Gamma}(c_4, d_4),$$

where c and d denote the shape and scale hyperparameters for the gamma distributions. The resulting prior density plots are shown in Figure 3.5.

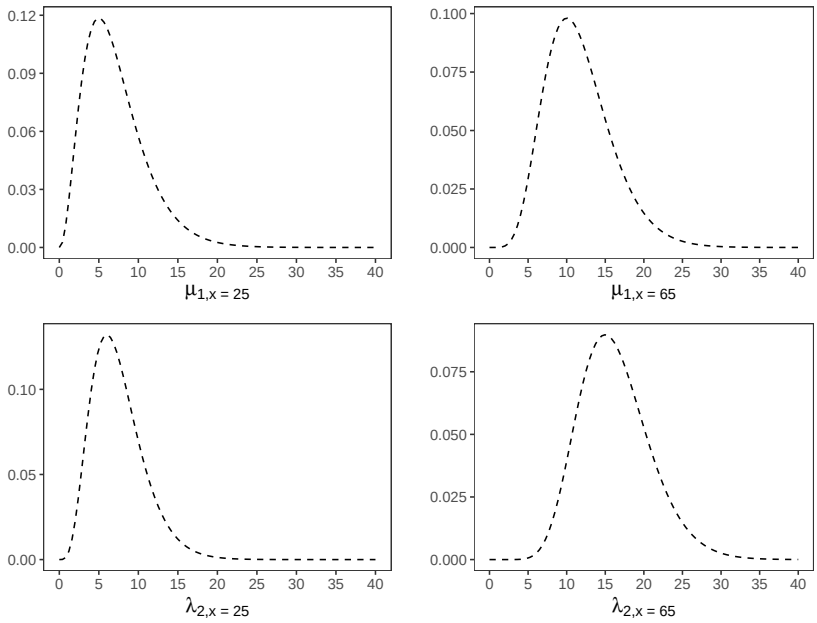

Figure 3.5: Priors on μ_1 and λ_2 based on information collected from expert.

In practice, these prior distributions are presented to the expert for feedback on whether or not the distributions accurately reflect their prior belief. Modifications are made until satisfactory. Suppose the above expert is satisfied with these resulting prior distributions. Then, the conditional means priors for the regression coefficients, ϕ_0 and ϕ_1 are given by

$$\begin{bmatrix} \phi_0 \\ \phi_1 \end{bmatrix} \sim \widetilde{\mathbf{X}}^{-1} \log \begin{bmatrix} \text{Gamma}(3.37, 2.10) \\ \text{Gamma}(7.22, 1.61) \end{bmatrix},$$

and the conditional means priors for the regression coefficients, $\gamma_{2,0}$ and $\gamma_{2,1}$ are given by

$$\begin{bmatrix} \gamma_{2,0} \\ \gamma_{2,1} \end{bmatrix} \sim \widetilde{\mathbf{X}}^{-1} \log \begin{bmatrix} \text{Gamma}(5.11, 1.46) \\ \text{Gamma}(12.52, 1.30) \end{bmatrix},$$

where

$$\widetilde{\mathbf{X}} = \begin{bmatrix} 1 & 25 \\ 1 & 65 \end{bmatrix}.$$

These induced priors have no closed form but can be simulated and are shown in Figure 3.6.

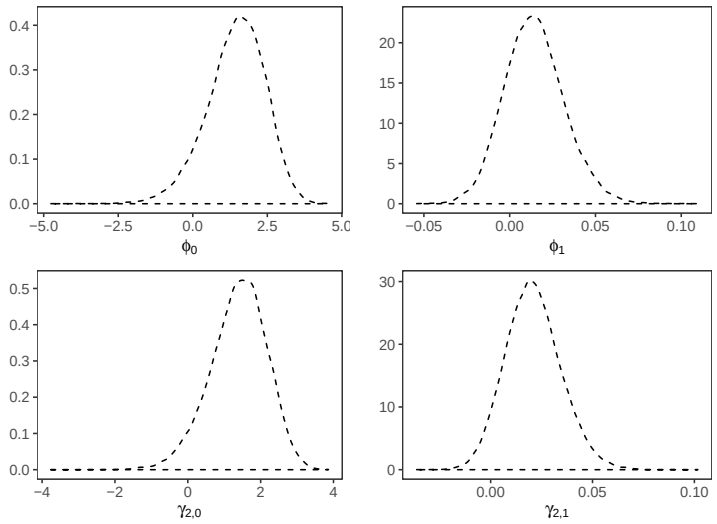

Figure 3.6: Simulated density plots for the induced priors on the regression coefficients corresponding to μ_1 and λ_2.

Next, we consider elicitation of a conditional means prior for the logistic regression parameters used to model the conditional probability, θ and the zero-inflation parameter, p. To elicit information the conditional probability that given a subject in the study has a migraine, the subject is also nauseated, the expert is asked a series of three questions. At age $x_i = 25$ and $x_i = 65$:

(1) To assess the median, the expert is asked"Suppose we have 100 at-risk subjects from the study population that are x_i years old that have a migraine.[9] What, do you think, is the percentage of these x_i year old subjects that are also nauseated?" Their response is taken to be the median.

(2) Next, to assess the lower quartile for the percentage of subjects that are nauseated given they have a migraine, the expert is asked "Suppose the percentage of x_i year-old subjects that are nauseated given they have a migraine is actually *less than* your initial assessment. Given this information, what

[9] In this context,"at-risk" subjects refers to subjects that are at-risk to experience migraines, and hence are at-risk to experience both adverse events.

would you now estimate as the percentage?" Their response is taken to be the 25th percentile.

(3) Finally, to obtain an upper quartile, the expert is asked "Suppose the percentage of x_i year-old subjects that are nauseated given they have a migraine is actually *greater than* your initial assessment. Given this information, what would you now estimate as the percentage?" Their response is taken to be the 75th percentile.

We elicit information about the zero-inflation probability, p, by asking the expert about their expectations about the proportion of individuals not at-risk to experience migraines. These assessment tasks are identical to those in Section 2.6.4. At both age $x_i = 25$ years old and $x_i = 65$ years old,

(1) To assess the median, the expert is asked "Suppose we have 100 subjects from the study population that are x_i years old. What, do you think, is the percentage of these x_i year old subjects that are not at-risk for experiencing migraines? Their response is taken to be the median value.

(2) Next, to assess a lower quartile for the percentage of subjects not at-risk for experiencing migraines, the expert is asked,"Suppose the percentage of x_i year-old subjects that are not at-risk for experiencing migraines is actually *less than* your initial assessment (that is, their median assessment is too high). Given this information, what would you now estimate as the percentage?" Their response is taken to be 25th percentile.

(3) Finally, to obtain an upper quartile the expert is asked "Suppose the percentage of x_i year-old subjects not at-risk for experiencing migraines is actually *greater than* your initial assessment (that is, their median assessment is too low). Given this information, what would you now estimate as the percentage?" Their response is taken to be the 75th percentile.

Suppose the information collected from the expert is that summarized in Table 3.5.

Table 3.5: Expert elicited information on conditional probability and zero-inflation parameter assuming a well-informed expert across all ages.

Parameter	25th Percentile	Median	75th Percentile
θ_{25}	0.25	0.32	0.39
θ_{65}	0.20	0.27	0.34
p_{25}	0.65	0.70	0.75
p_{65}	0.25	0.30	0.35

This information in obtained about θ and p in Table 3.5 is translated into parameters of a beta distribution. The resulting priors for age $x_i = 25$ and $x_i = 65$ are

$$\theta_{25} \equiv \text{logit}^{-1}\left(\alpha_0 + \alpha_1(25)\right) \sim \text{Beta}(a_{\theta,25}, b_{\theta,65})$$

and

$$\theta_{60} \equiv \text{logit}^{-1}\left(\alpha_0 + \alpha_1(65)\right) \sim \text{Beta}(a_{\theta,65}, b_{\theta,65}).$$

Similarly, for p we have that the resulting priors at $x_i = 25$ and $x_i = 65$ years old are

$$p_{25} \equiv \text{logit}^{-1}\left(\beta_0 + \beta_1(25)\right) \sim \text{Beta}(a_{p,25}, b_{p,25})$$

and

$$p_{65} \equiv \text{logit}^{-1}\left(\beta_0 + \beta_1(65)\right) \sim \text{Beta}(a_{p,65}, b_{p,65}).$$

These resulting beta prior densities for p and θ based on the expert's expectations at age $x_i = 25$ and $x_i = 65$ are shown in Figure 3.7.

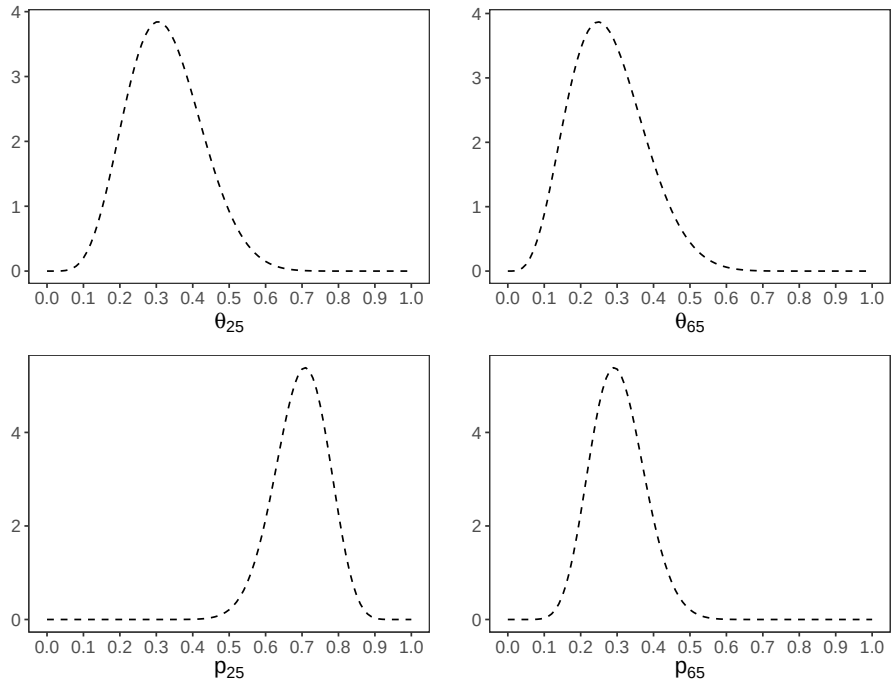

Figure 3.7: Priors based on information collected from expert about θ and p at age $x_i = 25$ and $x_i = 65$.

The conditional means priors for $\boldsymbol{\alpha}$ are the resulting induced priors given by

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \sim \widetilde{\mathbf{X}}^{-1}\text{logit} \begin{bmatrix} \text{Beta}(6.59, 13.73) \\ \text{Beta}(5.05, 13.25) \end{bmatrix},$$

and the conditional means priors for $\boldsymbol{\beta}$ are the resulting induced priors given by

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \sim \widetilde{\mathbf{X}}^{-1}\text{logit} \begin{bmatrix} \text{Beta}(26.71, 11.59) \\ \text{Beta}(11.59, 26.71) \end{bmatrix},$$

where

$$\widetilde{\mathbf{X}} = \begin{bmatrix} 1 & 25 \\ 1 & 65 \end{bmatrix}.$$

The induced priors on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have no closed form but can easily be simulated as shown in Figure 3.8.

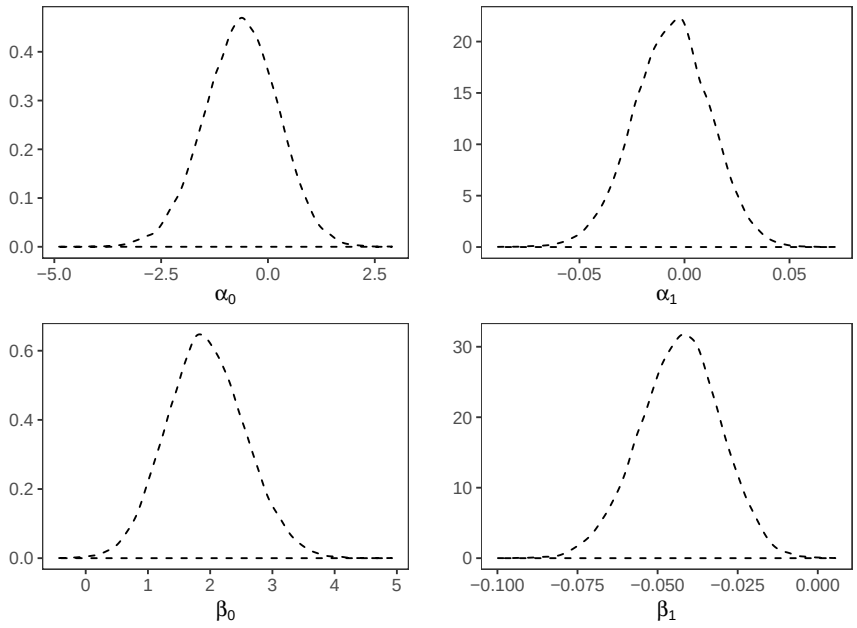

Figure 3.8: Simulated density plots for the induced priors on the regression coefficients corresponding to conditional probability, θ, and zero-inflation probability, p assuming a well-informed expert across all ages.

Finally, we obtain priors for λ_0, λ_1 and μ_2 at age $x_i = 25$ and $x_i = 65$ using (3.12), (3.13) and (3.14). From (3.12) we have that the prior for λ_0 is the product of a beta distribution and a gamma distribution. This is a nonstandard density, but has a closed form and is derived in Nadarajah and Kotz (2005). It follows that the priors for (3.13) and (3.14) is the difference and the sum, respectively, of a gamma distribution and the product of a beta distribution and a gamma distribution. Simulated density plots for these induced priors are shown in Figure 3.9.

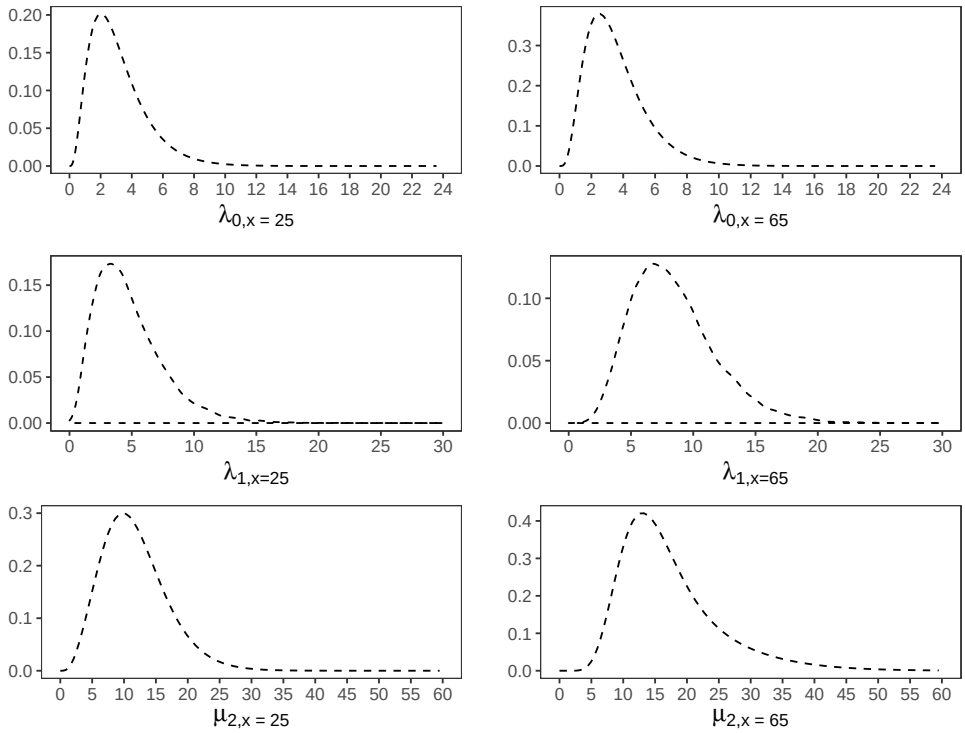

Figure 3.9: Density plots for the induced priors for λ_0, λ_1 and μ_2 at $x_i = 25$ and 65 based on information collected from the well-informed expert.

3.4.2 Using Prior Predictive as a Tool to Provide Feedback to Expert about λ_0.

A primary objective of this chapter is inference for the association parameter, λ_0, which represents the dependence between the joint events Y_1 and Y_2. In Chapter Two, we used the prior predictive distribution as a tool to provide feedback to the expert about the implications of their prior assessments in a meaningful scale. In this section, we demonstrate how the prior predictive distribution can be used as a tool to explore the implications of the expert's knowledge on the simultaneous occurrence of both adverse events. Recall the information collected from the expert from the assessment tasks outlined in Section 3.4.1:

Table 3.6: Prior information collected from a well informed expert across all ages.

Parameter	Mode	80th percentile	
$\mu_{1,25}$	5	10	
$\mu_{1,65}$	10	15	
Parameter	25th percentile	Median	75th percentile
θ_{25}	0.25	0.32	0.39
θ_{65}	0.20	0.25	0.34
p_{25}	0.65	0.70	0.75
p_{65}	0.25	0.30	0.35

This information is then translated into the priors shown in Figure 3.5 (top two graphs) and Figure 3.7.

In the introduction of this chapter we note that if subject-level counts for the number of migraines with nausea are available then posterior inference for λ_0 is straightforward. Similarly, if we have subject-level counts for the number of exclusive migraines and the number of exclusive nausea episodes, posterior inference for λ_1 and λ_2, respectively, is straightforward. Although, the observed data in Figure 3.16 still does not provide these counts, we can use prior predictive distribution as a tool to generate plausible data for these unobserved counts (particularly, X_1, X_2 and X_0 from the bivariate Poisson distribution in (3.1) and (3.2)) based on the information collected from the expert using the alternate prior elicitation tasks in Section 3.4.1.

As an example, we generate data from the prior predictive distribution corresponding to age 25, with $n = 200$ to demonstrate the implications of the prior assessments in Table 3.6. Based on the expert's expectation's about μ_1 and p at $x_i = 25$, we first use the prior predictive distribution to generate plausible data for Y_1, the number of migraines experienced over the study. This is shown in top left plot in Figure 3.10. Now consider only subject that experienced a migraine. This subset is shown in the top right plot in Figure 3.10. For illustration purposes, assume that $\theta_{25} = 0.32$ (the median value elicited from the expert). This suggests that the

expert believes that, for a subject age 25 experiencing a migraine, the probability they will also experience nausea is 0.32. That is, for each migraine event (among subjects age 25) we have $Y_{1i} \sim \text{Bernoulli}(0.32)$.

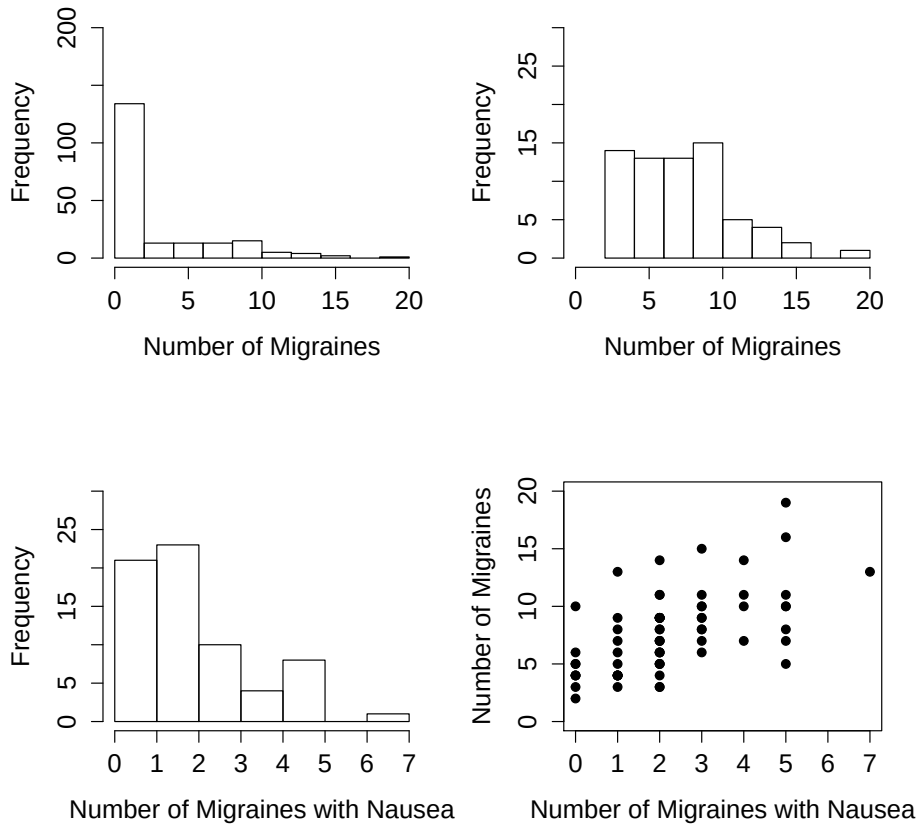

Figure 3.10: Prior predictive distribution for number of migraines for $n = 200$ subjects aged 25 (left) and prior predictive for number of migraines given a subject experienced migraines (right). The bottom left plot is the possible frequencies for the number of migraines among subjects age 25 that were accompanied by nausea. Assuming $\theta = 0.32$.

We can use this information to generate possible counts for the number of migraines accompanied by nausea for the ith subject given the number of migraines they experienced over the course of the study (the data represented in the top right histogram in Figure 3.10). These counts can be thought of as the unobserved counts, X_0 in Table 3.1 and are depicted in the bottom left histogram in Figure 3.10. It is

this count that we could then be used to construct a prior on λ_0. For example, we could use the mode and 80th percentile of the counts to construct a gamma prior distribution for λ_0 at $x_i = 25$. The bottom right plot shows the relationship between the number of migraines with nausea and the number of (just/exclusive) migraines based on this plausible data set generated from the prior predictive distribution and assuming $\theta = 0.32$. For example, this plot suggests that among subjects age 25 (at-risk to experience migraines) that experienced 5 migraines, the number of those migraines that were accompanied with nausea ranges from about 0 to 5. The Bayesian approach allows us to account for the uncertainty in θ_{25}. Based on the expert's assessments, take $\theta_{25} \sim \text{Beta}(6.59, 13.73)$. We select several values for θ_{25} across the prior parameter space to obtain a possible prior for λ_0 for subjects age 25, as shown in Figure 3.11.

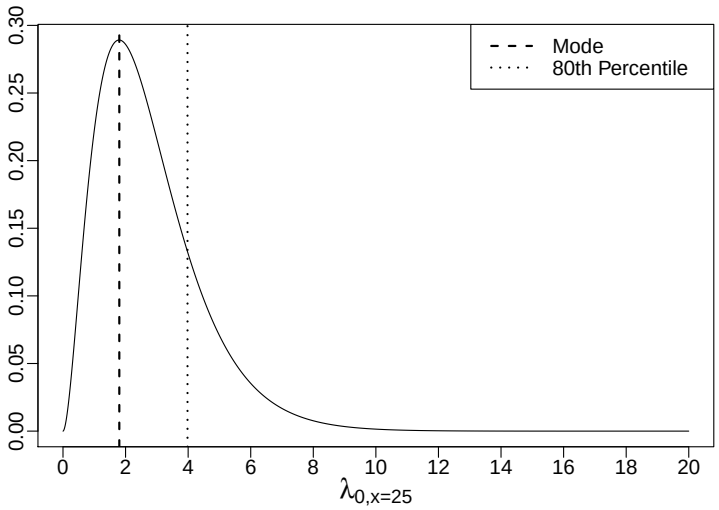

Figure 3.11: A prior distribution for λ_0 for subjects of age 25. Namely, $\lambda_{0,25} \sim \text{Gamma}(2.85, 0.97)$.

Similarly, we can obtain a plausible prior for λ_0 at $x_i = 65$ based on the expert's assessments. In a previous section, we acknowledged that designing a study to provide information about the unobserved counts X_0, X_1, and X_2 may not be a

viable option. In this section we have demonstrated how we can make use of the prior predictive distribution to visualize these unobserved counts.

3.4.3 Providing Feedback to the Expert

Providing feedback to the expert about the implications of their prior assessments is an essential aspect of the prior elicitation process. This allows the expert the opportunity for modifications to their assessments if the expert feels the resulting priors do not accurately reflect their prior beliefs. As in Section 2.7.1 we use the prior predictive distribution as a tool to provide feedback to the expert in a meaningful scale. For example, we might generate data from the prior predictive distribution corresponding to ages 25, 45, and 65, with a sample size of $n = 100$. For our purposes we assume that we show the expert a scatterplot and histograms for a hypothetical dataset based on their assessments and that the expert is satisfied. As another visual representation to show the expert the consequences of their prior assessments, we can provide the expert with information about the induced priors at certain ages as presented in Figure 3.12.

For example, Figure 3.12 suggests that the expert believes that the probability that a subject has a migraine and is nauseated is slightly less for older ages compared to that for younger ages. Similarly, we can present the expert a visual representation of the implications of their assessments on parameters that we do not directly elicit information about. For example, the induced priors for λ_0, λ_1 and μ_2 for certain ages are shown in Figure 3.13.

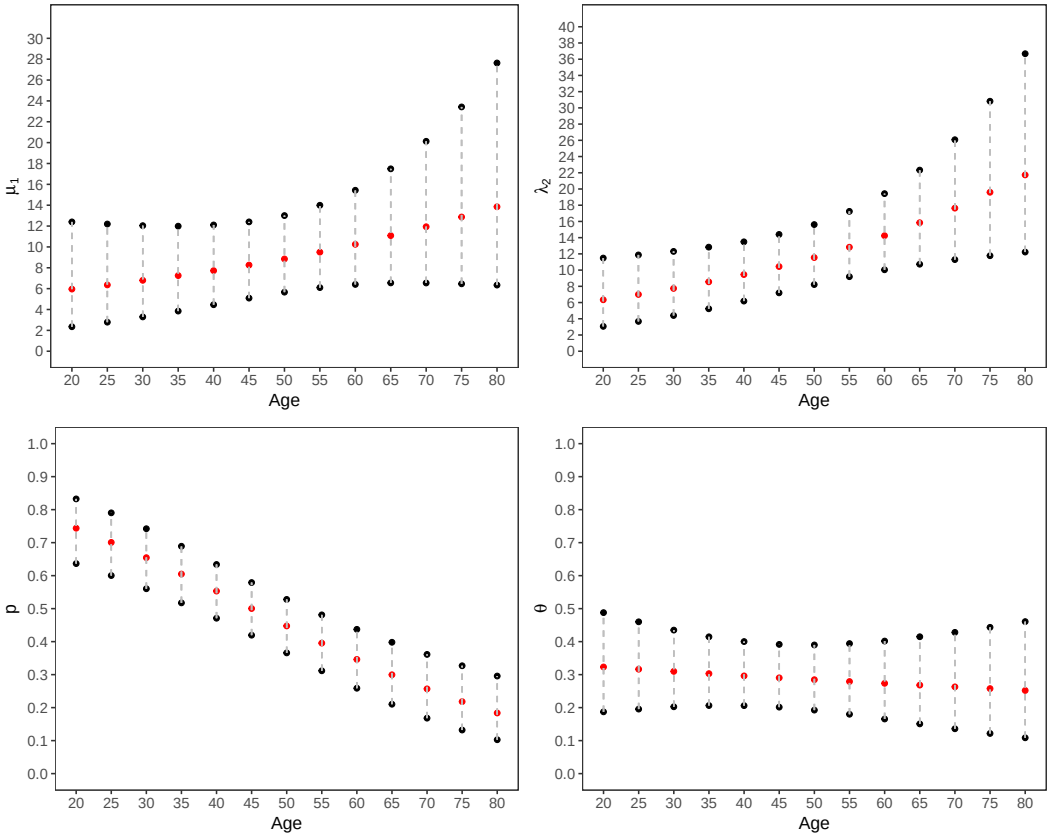

Figure 3.12: Induced prior median (red dot), 10th percentile and 90th percentile (black dots) based on a well-informed expert's expectations of these quantities.

Moreover, Figure 3.13 indicates that the expert's direct assessment of other quantities, suggests a slight increase in the number of migraines accompanied by nausea for subjects that are older in age. Moreover, Figure 3.13 suggests that the expert expects that the number of exclusive migraines and the number of nausea episodes will be greater in older subjects compared to younger subjects.

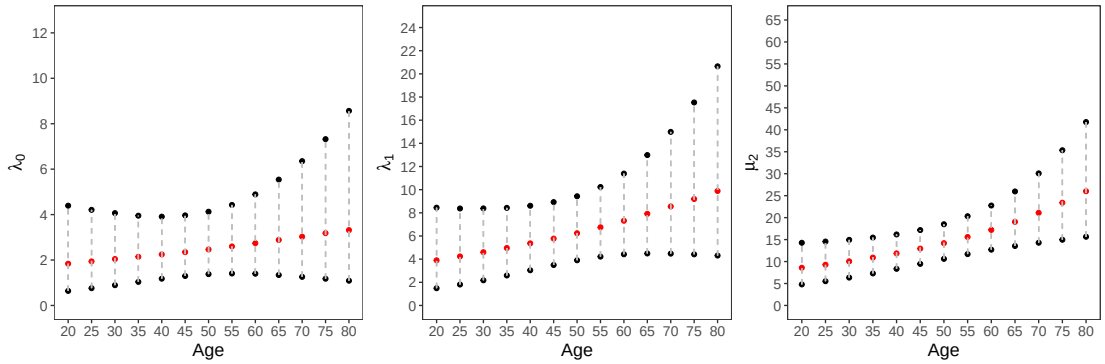

Figure 3.13: Induced prior median (red), 10th percentile and 90th percentile (black dots) based on expert's expectations of other operational quantities.

3.4.4 Posterior Inference Based on Information Obtained Via Alternate Elicitation

We assume that, based on the feedback in Section 3.4.3, the expert is satisfied with the prior distributions in Section 3.4.1. Posterior inference was carried out with the same specifications as in Section 2.3.1. We repeat them here for convienence. Namely, posterior inference was carried out in JAGS with two chains. Initial values for parameters were randomly generated from the corresponding prior distributions. We ran 160,000 iterations and used the first 10,000 iterations as a burn-in. For the remaining 150,000 iterations we sampled every 10th value to reduce autocorrelation. Accordingly, 30,000 parameter values were retained for each chain. Standard diagnostic methods based on trace plots and the Gelman-Rubin statistic revealed no convergence issues. The resulting posterior distributions for the regression coefficients are shown in Figure 3.14.

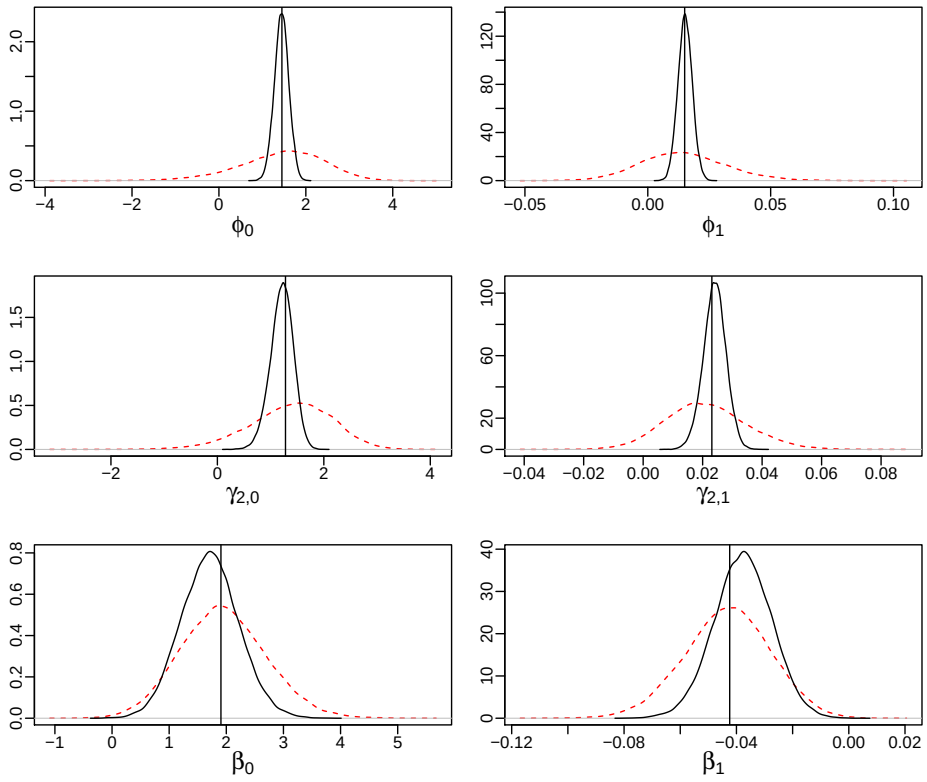

Figure 3.14: Prior (red, dashed) and posterior (black, solid) densities for BPZIP regression coefficients. True value of coefficient is indicated with vertical line. Here prior precision is independent of age.

Table 3.7: Posterior inference for regression coefficients. Priors specified indirectly via the CMP approach based on knowledge from a well-informed expert across all ages.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
ϕ_0	1.449	1.441	0.165	1.443	1.111	1.763	0.651
ϕ_1	0.015	0.015	0.003	0.015	0.009	0.021	0.012
$\gamma_{2,0}$	1.279	1.209	0.214	1.219	0.765	1.603	0.838
$\gamma_{2,1}$	0.023	0.024	0.003	0.024	0.016	0.031	0.015
β_0	1.906	1.713	0.506	1.711	0.736	2.729	1.993
β_1	-0.042	-0.037	0.010	-0.037	-0.058	-0.018	0.040

Posterior inference for $\boldsymbol{\gamma}_2$ and $\boldsymbol{\beta}$ is comparable to that presented in Section 2.7.3 using the standard representation of the BPZIP model. Moreover, the 95% credible intervals in Table 3.7 suggest that age is a significant factor in the estimation of p, λ_2 and μ_1. For example, the 95% credible interval for β_1 suggests that the odds of being at-risk to experience a migraine is between about $e^{20*0.0178} = 1.43$ and $e^{20*0.0583} = 3.21$ times greater for each additional 20-year increase in age with probability 0.95. Similarly, this suggests that the odds of being at-risk for a migraine is 4.52 times greater for subjects age 65 compared to subjects age 25 (this applies to any 40 year difference in age). Posterior inference for $\gamma_{2,1}$ suggests that for a 20-year increase in age, we expect to see the number of nausea episodes to increase by a factor between 1.38 and 1.87, with probability 0.95. Similarly, the 95% credible interval for ϕ_1 suggests that for a 20-year increase in age, we expect to see the number of migraines increase by a factor between 1.21 and 1.52. These results are consistent with the data we generated to simulate this example.

Similar to the posterior inference presented in Section 2.7.3, in Figure 3.15, we see more posterior updating of μ_1 and μ_2 compared to that seen for λ_0, λ_1 and λ_2. Recall, the conditional representation for the BPZIP model proposed in this chapter does not introduce constraints that make the model identifiable. However, with the conditional representation, the λ_k's are updated somewhat through μ_1 and μ_2, both of which are directly informed by the data (see Figure 3.4). As a result, the conditional representation of the BPZIP model potentially allows for more posterior updating of the λ_k's.

Posterior distributions for the Poisson parameters and zero-inflation parameter at the elicited ages $x_i = 25$, and $x_i = 65$ are shown in Figure 3.15.

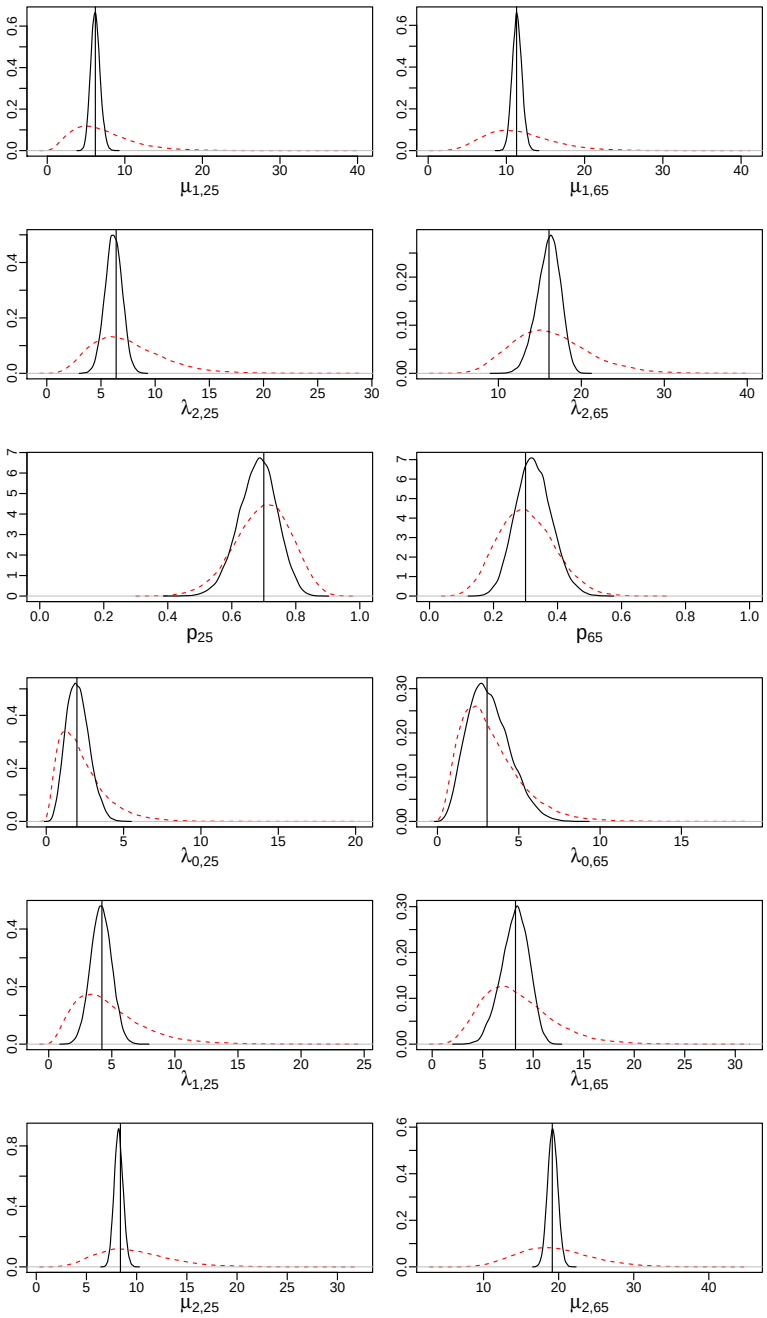

Figure 3.15: Prior (red, dashed) and posterior (black, solid) densities for model parameters at $x_i = 25$ and $x_i = 65$. Priors shown are based on expectations of a well-informed expert across all ages.

Table 3.8: Posterior inference for Poisson parameters and the zero-inflation parameter based on priors obtained from a well-informed expert across all ages.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
$\lambda_{0,25}$	1.998	2.145	0.611	2.112	1.067	3.441	2.374
$\lambda_{0,45}$	2.445	2.472	0.534	2.445	1.504	3.602	2.097
$\lambda_{0,65}$	2.993	2.923	0.932	2.852	1.305	4.925	3.621
$\lambda_{1,25}$	4.197	4.360	0.725	4.342	2.976	5.849	2.873
$\lambda_{1,45}$	5.896	5.838	0.612	5.838	4.636	7.048	2.412
$\lambda_{1,65}$	8.284	7.732	1.006	7.772	5.664	9.599	3.935
$\lambda_{2,25}$	6.400	6.459	0.688	6.461	5.099	7.793	2.693
$\lambda_{2,45}$	10.151	10.269	0.657	10.288	8.911	11.489	2.578
$\lambda_{2,65}$	16.100	16.391	1.127	16.433	14.072	18.490	4.418
p_{25}	0.700	0.722	0.051	0.724	0.615	0.817	0.201
p_{45}	0.500	0.493	0.041	0.492	0.412	0.574	0.162
p_{65}	0.300	0.286	0.048	0.264	0.177	0.365	0.188

The smaller posterior credible interval widths observed at age 45 compared to the credible interval widths at age 25 and 65 reflects the fact that there are more subjects close in this age in the hypothetical study. Thus, there is more data to update the posterior quantities at age 45 compared to age 25 and 65.

The conditional BPZIP model representation adapted in this chapter does not involve inference for covariates corresponding to the Poisson parameters, λ_0, and λ_1.[10] Nevertheless, the Bayesian paradigm readily allows for posterior inference of λ_0 and λ_1 at certain ages. Figure 3.16 shows the posterior median (red) and 95% credible interval (black) at certain ages, as well as the corresponding posterior densities for λ_0 and λ_1. The top right plot in Figure 3.16 suggests that the posterior mode for λ_0 slightly shifts from about 2 to 3 with increasing age. This suggests a slight difference in the rate of simultaneously experiencing both migraines and nausea as age increases. This is further seen in the top left graph which suggests a slight increase in the posterior median. The bottom two plots in Figure 3.16 suggest

[10] This is different from the BPZIP regression model in Chapter Two which allows for inference on regression coefficients for covariates related to λ_0 and λ_1.

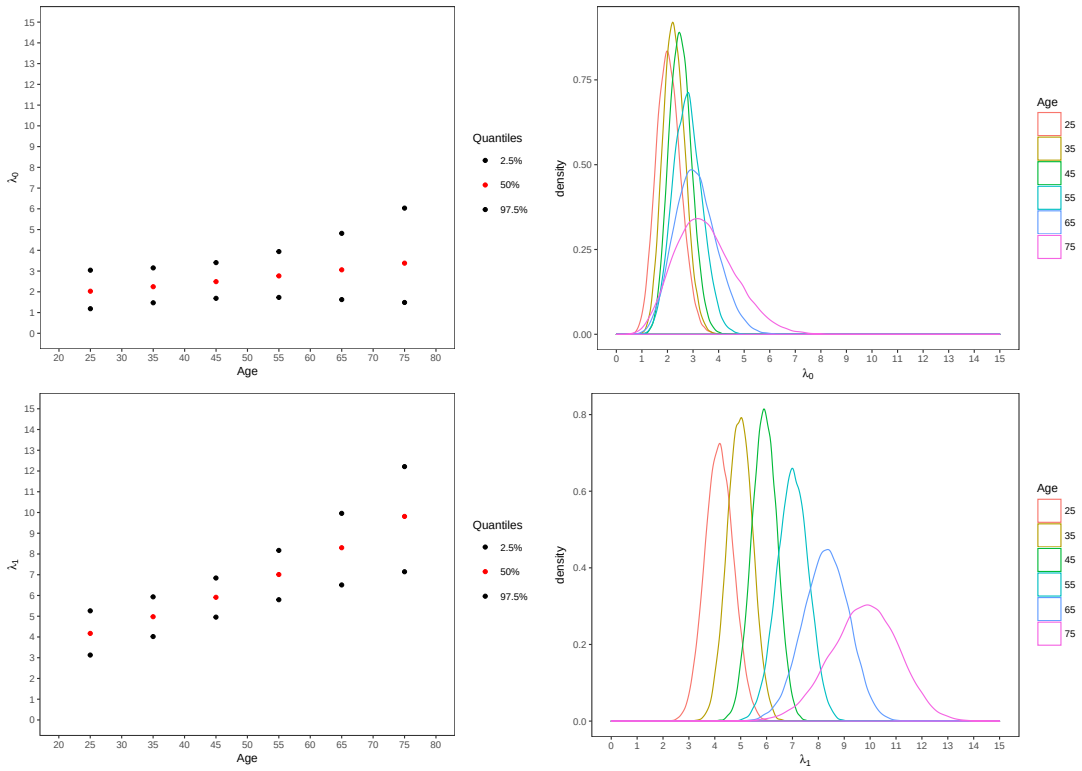

Figure 3.16: Posteriors for λ_0 and λ_1 at certain ages. The left plots show the posterior median (black dot) and 95% credible intervals (red dots) for certain ages. The right plots are the posterior densities for certain ages.

an increase in the number of exclusive migraines experienced by subjects as age increases.

Note that we do not include posterior inference for θ and corresponding regression coefficients $\boldsymbol{\alpha}$. For our purposes, the conditional representation allows us a way in which we can plausibly specify a prior for λ_0. In particular, we propose the prior assessment tasks in Section 3.4.1 as an alternative to the assessment tasks proposed in Section 2.6.4 for data that follow the BPZIP model in (2.2). Thus, for our purposes we are not interested in posterior inference for parameters θ and α. That is, our use of the conditional probability, θ, is in prior construction, not posterior inference. Determining if the conditional representation of the BPZIP model

reduces computation time as AlMuhayfith et al. (2015) found with the standard BP and bivariate zero-inflated Poisson distribution is an area for further research.

3.4.5 Posterior Predictive Distribution as a Tool for Clinician

In this section, we make use of the posterior predictive distribution as a tool for a clinician to assess what a future patient that is x_i years old taking this medication might expect to experience given everything we know (i.e. given the current data, prior information, and posterior). Figures 3.17 and 3.18 represent the posterior predictive distributions for the adverse events of interest.

Figures 3.17 and 3.18 could be used as a tool by a clinician to assess what a patient age x_i that is at-risk for both migraines and nausea episodes might expect to experience while taking this new drug. For example, suppose a clinician has a 55-year old patient that he/she has just prescribed this new drug. Based on Figure 3.17, the probability that the patient is at-risk to experience migraines is between about 63% and 70%. Furthermore, based on the conditional posterior distribution that given the patient is at-risk to experience migraines, the clinician can inform the patient that they might expect to experience about 9 migraines (and no more than 16 migraines) over a 6-month period. Similarly, Figure 3.18 suggests this patient can expect to experience about 15 nausea episodes and no more than 22, over a 6 month period.

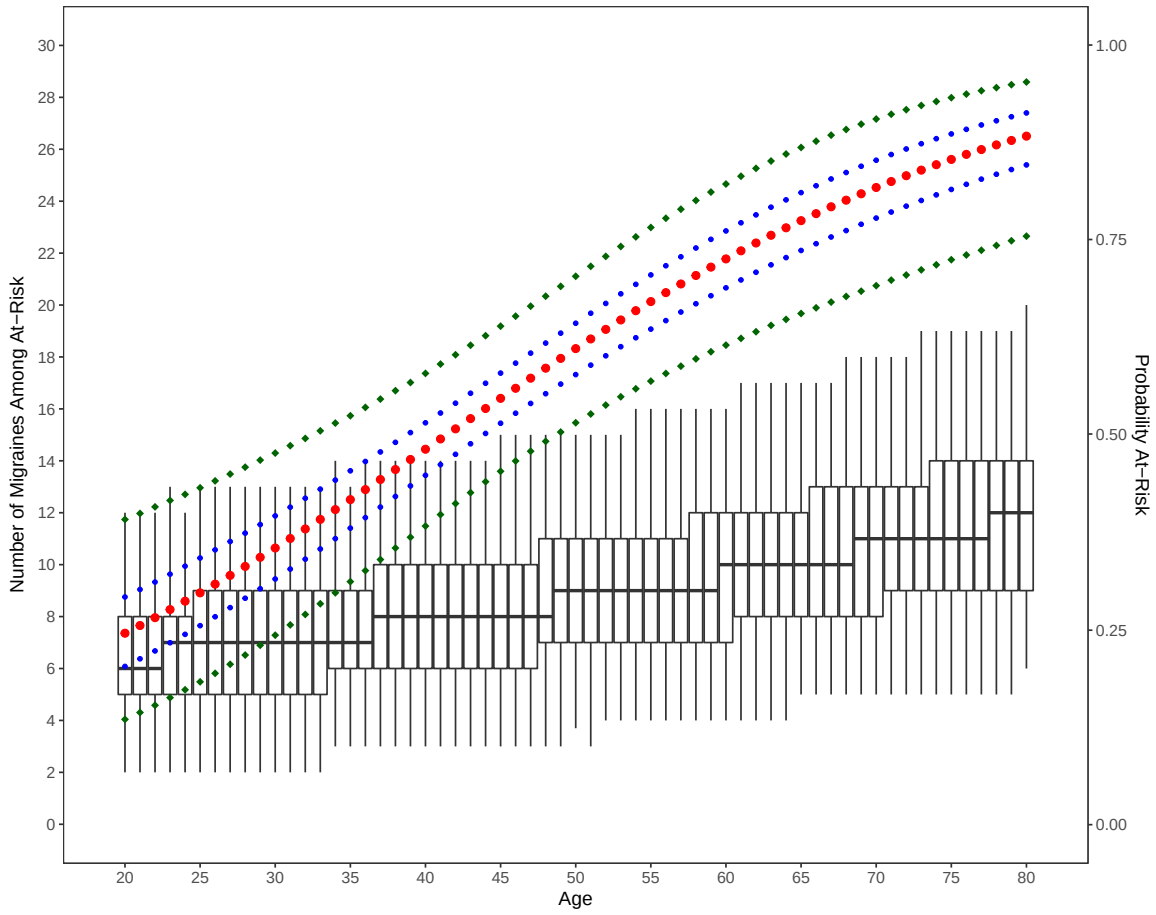

Figure 3.17: Posterior predictive distribution of number of migraines (represented by box plots) *given* a subject age x_i is at-risk for migraines, and the marginal posterior distribution for the probability that a subject age x_i is at-risk for migraines (the red dots represent the posterior median, the blue dots represent the 25th and 75th percentile, and the green dots represent the 2.5th and 97.5th percentile).

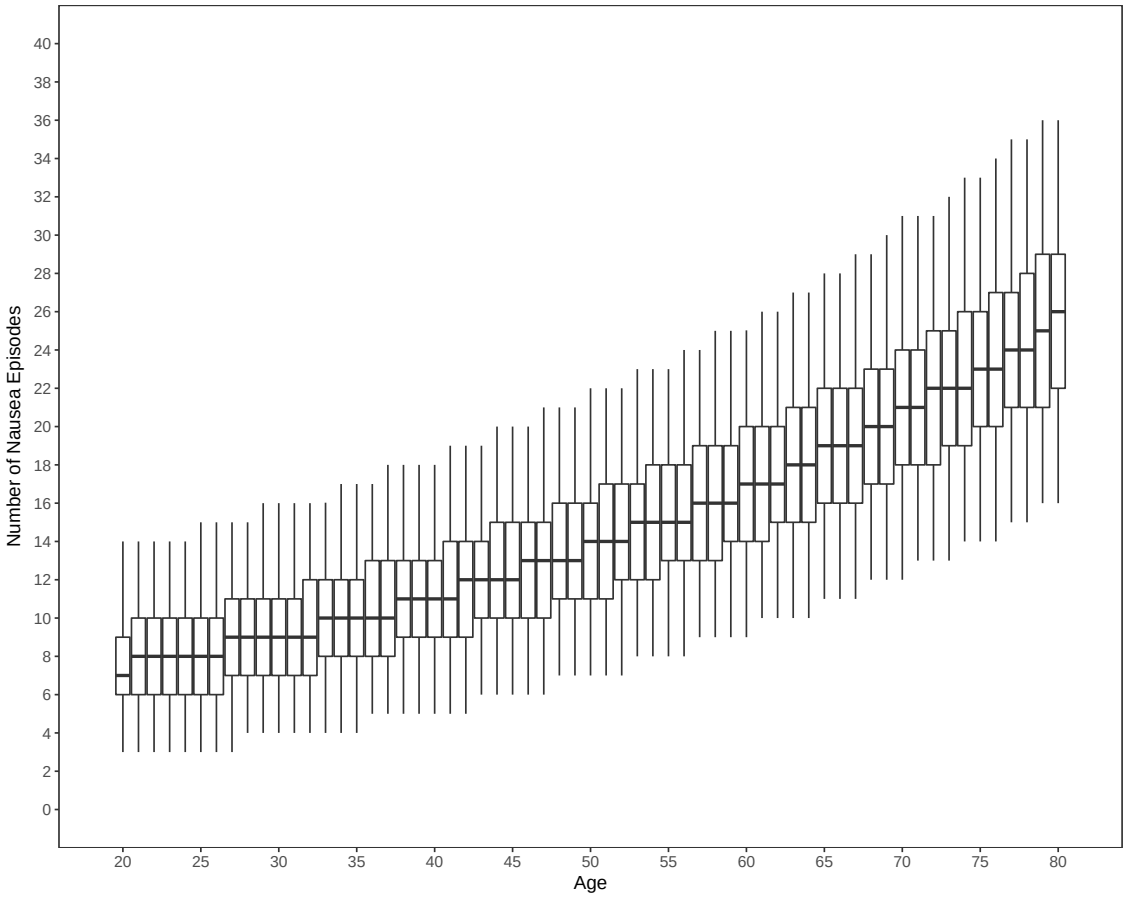

Figure 3.18: Posterior predictive distribution for number of nausea episodes (represented by box plots) for a subject age x_i.

3.4.6 Another Illustration: Expert Less Confident with Older Ages.

As another illustration, we now suppose that the expert is not equally confident across all ages. In particular, suppose that we have an expert that is comfortable making assessments about subjects age 20-30 years old, but becomes increasingly uncertain in their assessments as age increases from age 30. We demonstrate the implications of this increasing uncertainty on posterior inference. Inference was carried out using the same specifications as in Section 3.4.4. Using the same assessment tasks outlined in Section 3.4.1 we collect the information provided in Table 3.9.

Table 3.9: Prior information collected from expert that is more uncertain with increasing age.

Parameter	Mode	80th percentile	
$\mu_{1,25}$	5	10	
$\mu_{1,65}$	10	20	
$\lambda_{2,25}$	6	10	
$\lambda_{2,25}$	15	25	
Parameter	**25th percentile**	**Median**	**75th percentile**
$\theta_{\tilde{x}_i=25}$	0.25	0.32	0.39
$\theta_{\tilde{x}_i=65}$	0.17	0.27	0.37
$p_{\tilde{x}_i=25}$	0.65	0.70	0.75
$p_{\tilde{x}_i=65}$	0.18	0.30	0.42

Note that the elicited percentile information in Table 3.9 is widened to account for the increased uncertainty at older ages. This is further shown in Figures 3.19 and 3.20

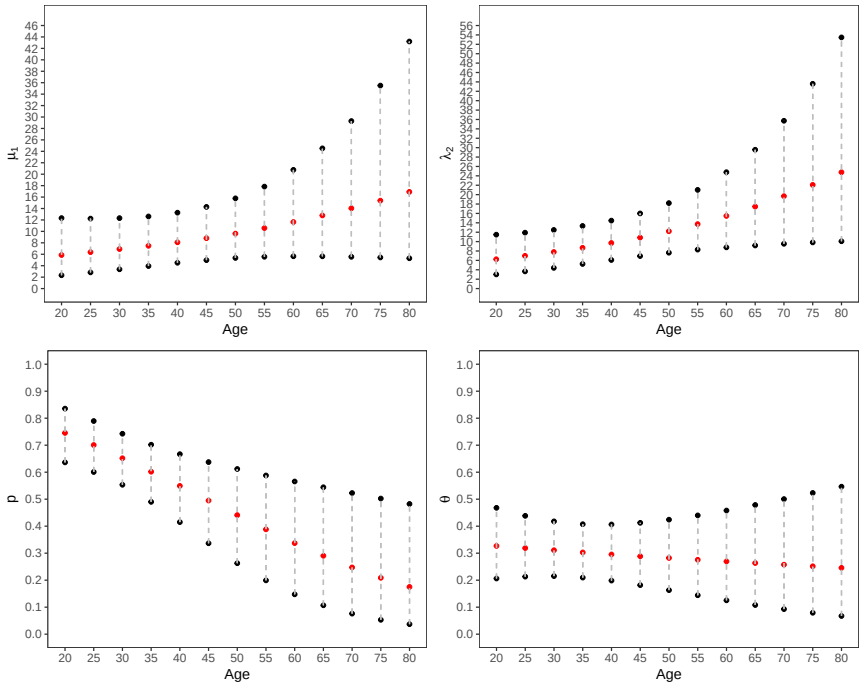

Figure 3.19: Induced prior median (red dot), 10th percentile and 90th percentile (black dots) based on expert's expectations.

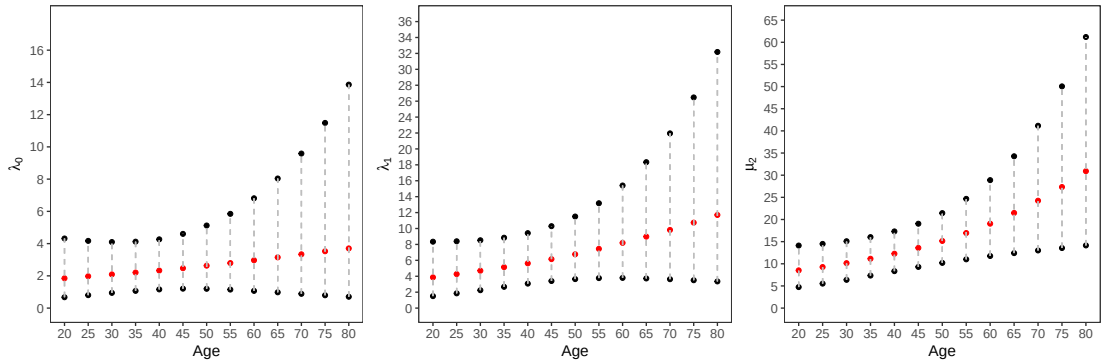

Figure 3.20: Induced prior median (red dot), 10th percentile and 90th percentile (black dots) based on expert's expectations.

The resulting posterior densities for regression coefficients are shown in Figure 3.21.

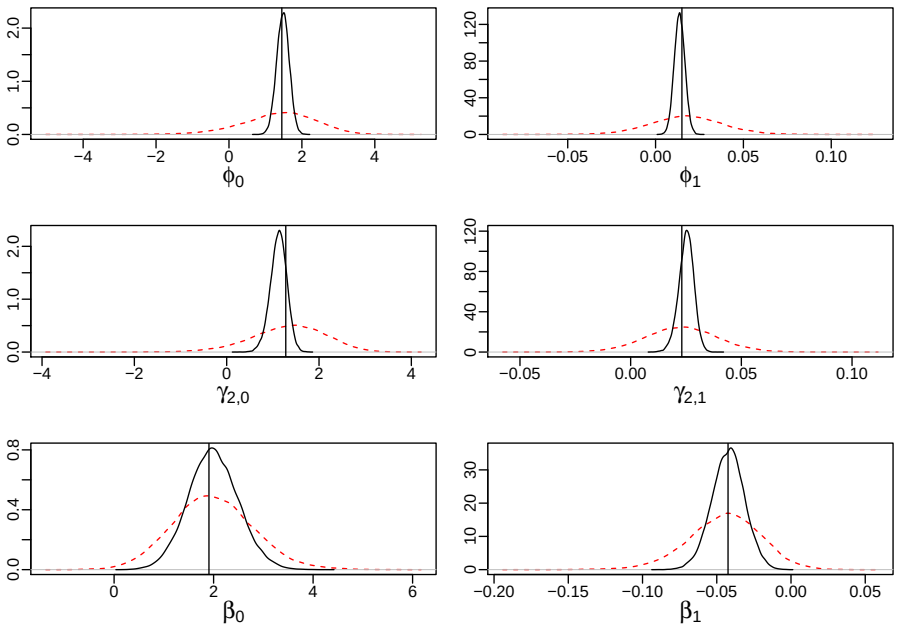

Figure 3.21: Prior (red, dashed) and posterior (black, solid) for regression coefficients. Priors reflect beliefs of an expert that is less confident making assessments about older ages.

Table 3.10: Posterior inference for regression coefficients.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
ϕ_0	1.449	1.478	0.172	1.481	1.139	1.813	0.673
ϕ_1	0.015	0.014	0.003	0.014	0.008	0.019	0.012
$\gamma_{2,0}$	1.280	1.122	0.178	1.129	0.752	1.454	0.702
$\gamma_{2,1}$	0.023	0.025	0.003	0.025	0.018	0.031	0.014
β_0	1.906	1.994	0.501	1.985	1.028	2.999	1.971
β_1	−0.042	−0.042	0.011	−0.042	−0.065	−0.021	0.044

Posterior inference for model parameters at age $x_i = 25$ and 65 is shown in Figure 3.22. The larger width of the prior densities at age 65 compared to priors at age 25 reflects the increasing uncertainty of the expert with older ages. The increased uncertainty is also reflected in the prior ESS for the priors that result from the expert's assessments. As an example, the expert's assessments about the proportion of subjects age $x_i = 25$ and 65 not at-risk for migraines in Table 3.9 can be translated into the parameters of a beta distribution using numerical methods. Namely,

$$\tilde{p}_{25} \sim \text{Beta}(26.71, 11.59)$$

and

$$\tilde{p}_{25} \sim \text{Beta}(2.98, 6.63).$$

From these priors we have that the expert's assessments for subjects age 25 is equivalent to about 38 observations, whereas the expert's assessments for subjects age 65 is equivalent to about 9 observations. Posterior inference for model parameters is provided in Table 3.11. Note that the truth for all parameters is contained within their respective 95% credible intervals. Furthermore, the 95% credible interval widths at age 65 are are comparable to those in Table 3.8 reflecting the increased prior uncertainty at older ages. We constructed a similar example with $n = 200$, in which we compared the posterior results for priors obtained from a well informed expert,

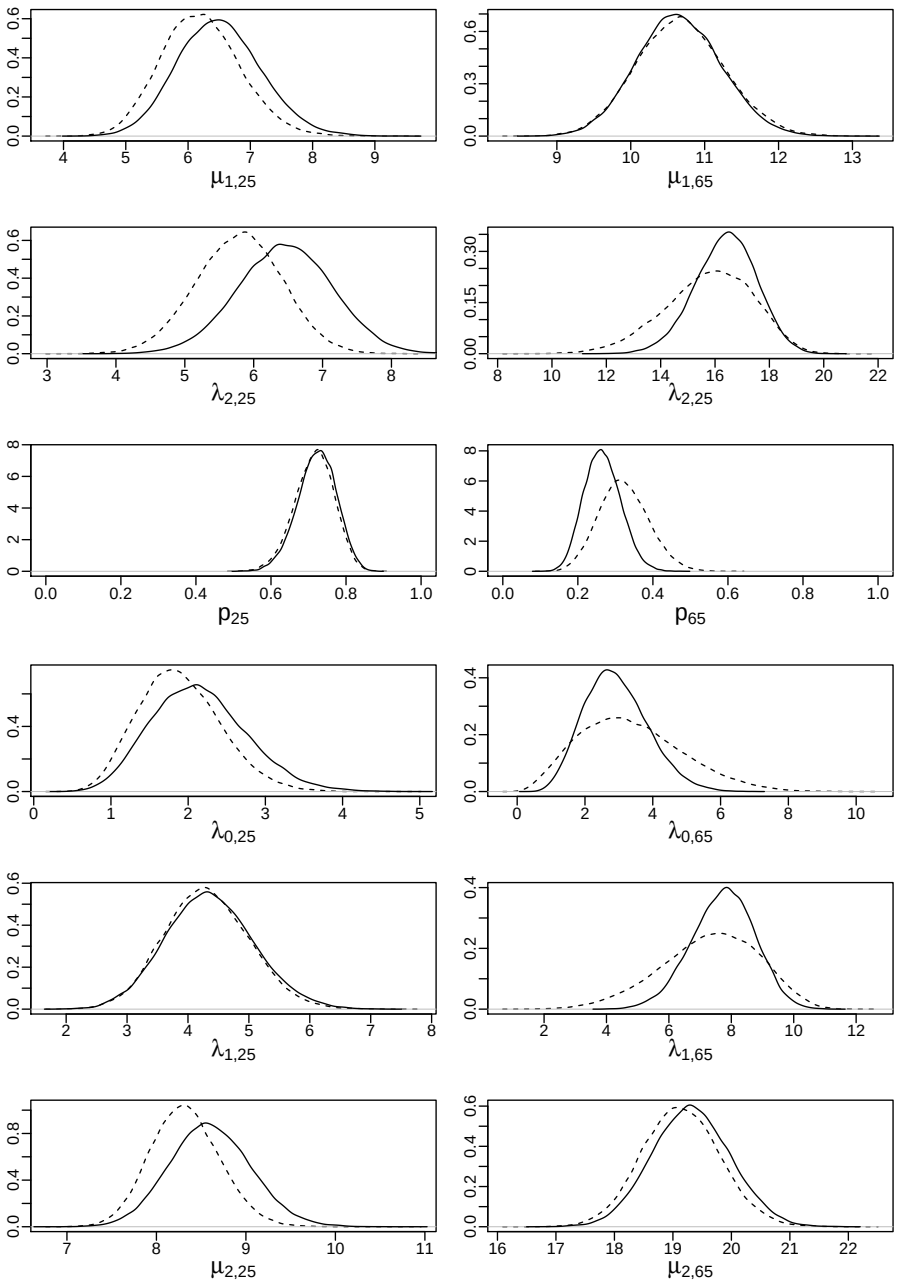

Figure 3.22: Comparison of posterior densities based on prior assessments of a well-informed expert (solid, black) and an expert that is less confident with assessments at older ages (dashed, black).

and posterior results for priors obtained from an expert that is less confident making assessments for older ages. We observed that for $n = 200$, the increased prior uncertainty was not reflected in the posterior as is the case for $n = 100$.

Table 3.11: Posterior inference for parameters based on priors obtained from an expert less confident with increasing age.

Parameter	Truth	Mean	SD	Median	2.5%	97.5%	Width
$\lambda_{0,25}$	1.9981	1.8925	0.5311	1.8588	0.9642	3.0217	2.0575
$\lambda_{0,45}$	2.4454	2.4787	0.7271	2.4415	1.1717	4.0015	2.8298
$\lambda_{0,65}$	2.9928	3.3654	1.5100	3.2184	0.8751	6.6343	5.7592
$\lambda_{1,25}$	4.1971	4.3124	0.6966	4.2910	3.0036	5.7216	2.7180
$\lambda_{1,45}$	5.8967	5.6461	0.7810	5.6553	4.0718	7.1233	3.0515
$\lambda_{1,65}$	8.2846	7.3119	1.5627	7.4143	4.0119	10.0529	6.0410
$\lambda_{2,25}$	6.4000	5.7863	0.6229	5.7978	4.5491	6.9862	2.4371
$\lambda_{2,45}$	10.1509	9.5318	0.7793	9.5634	7.9350	10.9817	3.0467
$\lambda_{2,65}$	16.1000	15.7737	1.6173	15.8800	12.3459	18.5981	6.2522
p_{25}	0.7000	0.7152	0.0525	0.7176	0.6060	0.8116	0.2056
p_{45}	0.5000	0.5218	0.0447	0.5218	0.4338	0.6092	0.1754
p_{65}	0.3000	0.3225	0.0658	0.3197	0.2009	0.4585	0.2576

Finally, we provide posterior inference for λ_0 and λ_1 at certain ages in Figure 3.23. These plots again suggest that as age increases there is a slight increase in the number of occurrence of both adverse events, an increase in the number of exclusive migraines, and the number of nausea episodes. The posterior densities and corresponding intervals for λ_0 and λ_1 suggest increased posterior variability (compared to that shown in Figure 3.16). This increased posterior variability for older ages is not observed when comparing the posterior plots for μ_2 in Figure 3.16 to those in Figure 3.23. This this not surprising as the prior dictates the posterior for λ_0 and λ_1 (nonidentifiable parameters) more so than for μ_2 (identifiable parameter).

3.5 Summary

We have developed a conditional representation of the Bayesian BPZIP model. We demonstrated that the conditional representation provides a method for prior

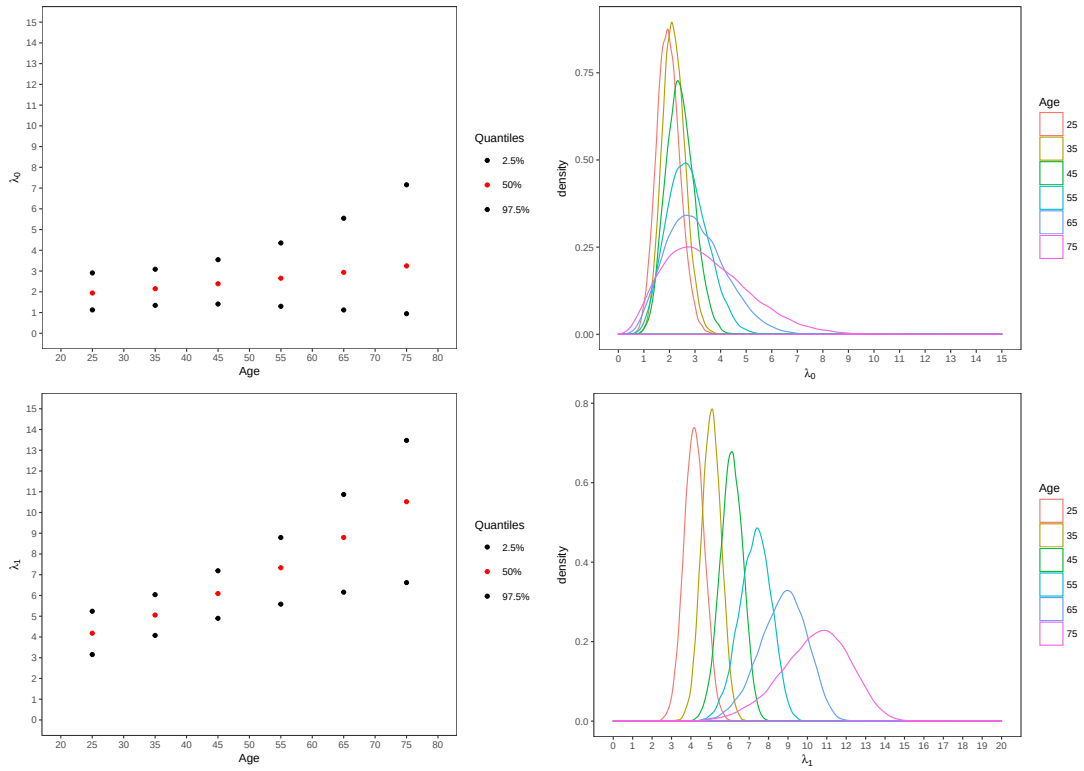

Figure 3.23: Posteriors for λ_0 and λ_1 across several ages when assuming increased uncertainty with increasing age. The left plots show the posterior median (black dot) and 95% credible intervals (red dots) for certain ages. The right plots are the posterior densities for certain ages.

construction of the association parameter, λ_0, via direct elicitation of a conditional expectation. We also demonstrated that the conditional representation potentially allows for more posterior updating of the unidentified Poisson parameters λ_0, λ_1 and λ_2 compared to the joint representation of the BPZIP model in Chapter Two. We proposed and described methods of prior elicitation for a CBPZP model within the context of a hypothetical adverse event drug safety study. We illustrated how we might use the prior predictive distribution to visualize the unobserved counts underlying the BP distribution in our BPZIP model. Finally, we propose the posterior predictive distribution as a tool for a clinician within the context of the drug safety study.

CHAPTER FOUR

A Bayesian Bivariate Zero-Inflated Poisson Model

Thus far we have discussed a BPZIP model where one outcome is zero-inflated and one outcome is not zero-inflated. In particular, we presented a Bayesian analysis and proposed methods of prior construction for model parameters. In this chapter, we extend the ideas proposed in Chapter Two and Chapter Three to a Bayesian bivariate zero-inflated Poisson (BZIP) model, where both outcomes are assumed zero-inflated. In medical, healthcare, environmental, and ecological studies, existence of excessive zeros in count data is common. If the zeros are ignored for the sake of simplifying the analysis, valuable information will be lost and can result in biased estimates of the parameters and potentially misleading findings. In this chapter, we develop Bayesian methods for fitting the general BZIP model, ultimately in a regression setting.

The organization of this chapter is as follows: In Sections 4.1 and 4.1.1 we introduce a bivariate zero-inflated Poisson model. In Section 4.2 we develop a Bayesian BZIP model, which includes diffuse prior structures and methods for posterior inference. In Section 4.3 we demonstrate potential problems that can arise in posterior inference for BZIP models, particularly in a diffuse prior setting. In Section 4.4 we discuss common approaches in the literature for mitigating the effects of non-identifiability for inference of Poisson parameters. In Section 4.5 we intoduce a clinical example to investigate the safety and efficacy of a new drug. This example serves as a vehicle for which we illustrate our proposed methods for BZIP inference throughout this chapter. In Section 4.6 we offer methods for informative prior construction for a Bayesian BZIP model. We discuss, in detail, prior elicitation from a subject-matter expert. This includes prior elicitation of a Dirichlet distribution

for zero-inflation parameters using conditional beta distributions, and a conditional means priors approach for Poisson parameters. Concluding comments are given in Section 4.7.

4.1 Bivariate Zero-Inflated Poisson Model

Suppose we have two Poisson counts denoted Y_1 and Y_2 such that (Y_1, Y_2) denotes a bivariate count vector. In Chapter Two we discussed the case where one outcome is zero-inflated and one outcome is not zero-inflated. In this chapter we consider the case where both Poisson counts, Y_1 and Y_2, are zero-inflated. The bivariate zero-inflated Poisson model is used to model count data with abundance of zero observations that come from two sources: sampling zeros and structural zeros. The most common representation of the BZIP data model found in the literature is that introduced in Li et al. (1999), and is the representation that we use for our joint count distribution. More recently, variations of the BZIP distribution proposed by Li et al. (1999), have appeared. See, for example, Walhin (2001), Wang et al. (2003), Dong et al. (2014), and Mohammadi et al. (2016).

A BZIP model can be constructed from a mixture of a point mass at (0,0), two univariate Poisson distributions with parameters μ_1 and μ_2, and a bivariate Poisson (BP) distribution with parameters $(\lambda_0, \lambda_1, \lambda_2)$ as follows:

$$(Y_1, Y_2) \sim \begin{cases} (0,0), & \text{with probability } p_0 \\ (\text{Poisson}(\mu_1), 0), & \text{with probability } p_1 \\ (0, \text{Poisson}(\mu_2)), & \text{with probability } p_2 \\ \text{BP}(\lambda_0, \lambda_1, \lambda_2), & \text{with probability } p_3, \end{cases} \tag{4.1}$$

where $p_3 = 1 - p_0 - p_1 - p_2$, $\mu_1 = \lambda_1 + \lambda_0$, and $\mu_2 = \lambda_2 + \lambda_0$. The BZIP model in (4.1) assumes that *both* counts, Y_1 and Y_2, contribute to the observed excess zeros. This is in contrast to the BPZIP model in (2.2) which assumes that the zero-inflation is attributed to only Y_1. In addition, the BZIP model assumes both structural and

sampling zeros for both Y_1 and Y_2. Again, this is in contrast to the BPZIP model which assumes structural and sampling zeros for Y_1, and only sampling zeros for Y_2.[1] As a result, the BZIP model introduces more parameters compared to the BPZIP model to account for the zero-inflation. As with the BPZIP model, we use the trivariate reduction representation of the BP distribution (see Appendix A). Thus, the BZIP model in (4.1) is appropriate for modeling bivariate zero-inflated Poisson counts that exhibit positive association.

Let $(Y_1, Y_2) \sim \text{BZIP}(\mathbf{p}, \boldsymbol{\lambda})$ where $\mathbf{p} = (p_0, p_1, p_2, p_3)$, $\sum_{j=0}^{3} p_j = 1$, and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$. Then the bivariate joint probability mass function, $f_{Y_1,Y_2}(y_1, y_2|\mathbf{p}, \boldsymbol{\lambda}) \equiv f_{Y_1,Y_2}(y_1, y_2)$, is given by, for $y_1, y_2 \in \{0, 1, 2, \ldots\}$,

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} p_0 + p_1\, e^{-(\lambda_1+\lambda_0)} + p_2\, e^{-(\lambda_2+\lambda_0)} + p_3\, e^{-(\lambda_1+\lambda_2+\lambda_0)}, & y_1 = 0, y_2 = 0, \\ \dfrac{1}{y_1} p_1\, (\lambda_1 + \lambda_0)^{y_1}\, e^{-(\lambda_1+\lambda_0)} + p_3\, \lambda_1^{y_1}\, e^{-(\lambda_1+\lambda_2+\lambda_0)}, & y_1 \neq 0, y_2 = 0, \\ \dfrac{1}{y_2!} p_2\, (\lambda_2 + \lambda_0)^{y_2}\, e^{-(\lambda_2+\lambda_0)} + p_3\, \lambda_2^{y_2}\, e^{-(\lambda_1+\lambda_2+\lambda_0)}, & y_1 = 0, y_2 \neq 0, \\ p_3\, e^{-(\lambda_1+\lambda_2+\lambda_0)} \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_2^{y_2}}{y_2!} \sum_{m=0}^{\min(y_1,y_2)} \binom{y_1}{m} \binom{y_2}{m} m! \left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^m, & y_1 \neq 0, y_2 \neq 0. \end{cases} \tag{4.2}$$

In the applications we consider in this dissertation, we typically take $f_{(Y_1,Y_2)}(0, 0)$ and $f_{(Y_1,Y_2)}(y_1, 0)$, $y_1 \neq 0$, to be large (compared to the BPZIP model). It can be shown that

$$\text{E}(Y_1, Y_2) = p_3\, [(\lambda_1 + \lambda_0)\, (\lambda_2 + \lambda_0) + \lambda_0]\,,$$

and

$$\text{Cov}(Y_1, Y_2) = p_3\, \lambda_0 + [p_3\, p_0 - p_1\, p_2]\, (\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0).$$

[1] For a comparison of the BPZIP and BZIP model see Appendix E.

Moreover, the correlation is given by

$$\rho_{\text{BZIP}} \equiv \text{Corr}(Y_1, Y_2) = \frac{p_3\,\lambda_0 + [p_3\,p_0 - p_1\,p_2]\,(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)}{\sqrt{\sigma^2_{Y_1}\,\sigma^2_{Y_2}}},$$

where $\sigma^2_{Y_1}$ and $\sigma^2_{Y_2}$ denote the variance of Y_1 and Y_2, respectively.

Let $\mathbf{y}_i = (y_{1i}, y_{2i})$, $i = 1, \ldots, n$, denote the bivariate outcomes. The corresponding likelihood function is

$$\begin{aligned}\ell(\mathbf{p}_i, \boldsymbol{\lambda}_i \,|\, \mathbf{y}_i) = &\prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = 0 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i1}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = 0 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i2}} \\ &\times \prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = y_2 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i3}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = y_2 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i4}},\end{aligned} \tag{4.3}$$

where $\mathbf{p}_i = (p_{0i}, p_{1i}, p_{2i}, p_{3i})$, $\boldsymbol{\lambda}_i = (\lambda_{0,i}, \lambda_{1,i}, \lambda_{2,i})$ and I_{ik} is an indicator function defined as

$$\begin{aligned} I_{i1} = 1: &\text{ if } y_{1i} = 0, y_{2i} = 0, \\ I_{i2} = 1: &\text{ if } y_{1i} \neq 0, y_{2i} = 0, \\ I_{i3} = 1: &\text{ if } y_{1i} = 0, y_{2i} \neq 0, \end{aligned} \tag{4.4}$$

and

$$I_{i4} = 1: \text{ if } y_{1i} \neq 0, y_{2i} \neq 0.$$

It can be shown that the marginal distribution of Y_1 is a univariate ZIP that is a mixture of point mass at 0 with probability $p_0 + p_2$ and a Poisson distribution with parameter μ_1 with probability $1 - (p_0 + p_2)$. This is denoted $Y_1 \sim \text{ZIP}(p_0 + p_2, \mu_1)$. That is,

$$Y_1 \sim \begin{cases} 0, & \text{with probability } p_0 + p_2, \\ \text{Poisson}(\mu_1), & \text{with probability } p_1 + p_3, \end{cases}$$

with corresponding marginal probability mass function

$$f_{Y_1}(y_1 \,|\, p_0, p_2, \mu_1) \equiv \Pr(Y_1 = y_1) = \begin{cases} (p_0 + p_2) + (1 - p_0 - p_2)\ e^{-\mu_1}, & y_1 = 0 \\ (1 - p_0 - p_2)\frac{\mu_1^{y_1}\, e^{-\mu_1}}{y_1!}, & y_1 = 1, 2, 3 \ldots, \end{cases} \tag{4.5}$$

where $\mu_1 = \lambda_1 + \lambda_0$ and $\sum_{j=0}^{3} p_j = 1$. Thus, Y_1 can be a structural zero with probability $p_0 + p_2$ and a sampling zero with probability $p_1 + p_3$. It follows that the mean and variance of Y_1 are given by

$$\mathrm{E}(Y_1) = (p_1 + p_3)\mu_1, \tag{4.6}$$

and

$$\mathrm{Var}(Y_1) = (p_1 + p_3)\mu_1 \left[1 + (p_0 + p_2)\mu_1\right]. \tag{4.7}$$

Similarly, it can be shown that the marginal distribution of Y_2 is a univariate ZIP, denoted $Y_2 \sim \mathrm{ZIP}(p_0 + p_1, \mu_2)$. That is,

$$Y_2 \sim \begin{cases} 0, & \text{with probability } p_0 + p_1, \\ \mathrm{Poisson}(\mu_2), & \text{with probability } p_2 + p_3, \end{cases}$$

with corresponding marginal probability mass function

$$f_{Y_2}(y_2 \,|\, p_0, p_1, \mu_2) \equiv \Pr(Y_2 = y_2) = \begin{cases} (p_0 + p_1) + (1 - p_0 - p_1)\ e^{-\mu_2}, & y_2 = 0 \\ (1 - p_0 - p_1)\frac{\mu_2^{y_2}\, e^{-\mu_2}}{y_2!}, & y_2 = 1, 2, 3 \ldots, \end{cases}$$

where $\mu_2 = \lambda_2 + \lambda_0$ and $\sum_{j=0}^{3} p_j = 1$. Thus, Y_2 can be a structural zero with probability $p_0 + p_1$ and a sampling zero with probability $p_2 + p_3$. It follows that the mean and variance of Y_2 are

$$\mathrm{E}(Y_2) = (p_2 + p_3)\mu_2, \tag{4.8}$$

and

$$\mathrm{Var}(Y_2) = (p_2 + p_3)\mu_2 \left[1 + (p_0 + p_1)\mu_2\right]. \tag{4.9}$$

4.1.1 Bivariate Zero-Inflated Poisson Regression Model

Let $(y_{1i}, y_{2i}) \sim \text{BZIP}(\mathbf{p}_i, \boldsymbol{\lambda}_i)$, for $i = 1, \ldots, n$, where n denotes the number of observations, $\mathbf{p}_i = (p_{0,i}, p_{1,i}, p_{2,i}, p_{3,i})$ denotes the vector of zero-inflation parameters for the ith observation and $\boldsymbol{\lambda}_i = (\lambda_{0,i}, \lambda_{1,i}, \lambda_{2,i})$ denotes the Poisson parameters for the ith observation. We can represent the BZIP model in such a way that the Poisson parameters, λ_k, and mixture probabilities, p_j, depend on covariates through canonical link, generalized linear models. In particular, we express the Poisson parameters λ_k, $k = 0, 1, 2$, as a function of covariates via the logarithmic link and the zero-inflation parameters, p_j, $j = 0, 1, 2$, is related to covariates via the logit link. That is, for the ith observation, we have that

$$\log(\lambda_{k,i}) = \mathbf{x}_{k,i}^T \boldsymbol{\gamma}_k, \tag{4.10}$$

for $k = 0, 1, 2$, and

$$\log\left(\frac{p_{j,i}}{1 - \sum_{j=0}^{2} p_{j,i}}\right) = \mathbf{w}_{j,i}^T \boldsymbol{\beta}_j, \tag{4.11}$$

for $j = 0, 1, 2$, where $\mathbf{x}_{k,i}$ and $\boldsymbol{\gamma}_k$ are vectors of covariates and corresponding regression coefficients, respectively, associated with the kth λ, and $\mathbf{w}_{j,i}$ and $\boldsymbol{\beta}_j$ are vectors of covariates and corresponding regression coefficients, respectively, associated with jth zero-inflation parameter (or mixture probability). Let q and r denote the number of covariates corresponding to λ_k and p_j, respectively. Specifically, $\boldsymbol{\gamma}_k$ is a $(q+1)$ vector of regression coefficients and $\boldsymbol{\beta}_j$ is a $(r+1)$ vector of regression coefficients. This parameterization allows the same or different covariates to affect the Possion rates and zero-inflation probabilities. For example, it is common to assume that the Poisson parameters depend on covariates and that the zero-inflation probabilities do not depend on covariates (e.g. Mohammadi et al. (2016)). In addition, this representation allows for the Poisson rates (and similarly the zero-inflation parameters) to depend on different covariates, which extends the use of this model to a wide

range of applications. For example, Wang et al. (2003) assume that λ_1 and λ_2 depend on covariates and that λ_0 does not depend on covariates.[2]

Let $\mathbf{y}_i = (y_{1i}, y_{2i})$, $i = 1, \ldots, n$. The BZIP regression likelihood is given by

$$\begin{aligned}\ell(\boldsymbol{\beta}_j, \boldsymbol{\gamma}_k \,|\, \mathbf{y}_i) = &\prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = 0 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i1}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = 0 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i2}} \\ &\times \prod_{i=1}^{n} \left[f(y_{1i} = 0, y_{2i} = y_2 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i3}} \prod_{i=1}^{n} \left[f(y_{1i} = y_1, y_{2i} = y_2 \,|\, \mathbf{p}, \boldsymbol{\lambda})\right]^{I_{i4}},\end{aligned} \tag{4.12}$$

where, for $i = 1, 2, \ldots, n$, $j = 0, 1, 2$, and $k = 0, 1, 2$, we have

$$p_{j,i} = \frac{\exp(\mathbf{w}_{j,i}^T \boldsymbol{\beta}_j)}{1 + \sum_{j=0}^{2} \exp(\mathbf{w}_{j,i}^T \boldsymbol{\beta}_j)},$$

$$p_{3,i} = \frac{1}{1 + \sum_{j=0}^{2} \exp(\mathbf{w}_{j,i}^T \boldsymbol{\beta}_j)},$$

and

$$\lambda_{k,i} = \exp(\mathbf{x}_{k,i}^T \boldsymbol{\gamma}_k),$$

and I_{it}, $t = 1, \ldots, 4$, is an indicator function defined as in (4.4).

Frequentist inference for bivariate zero-inflated Poisson model parameters in the literature include maximum likelihood estimation (MLE) and method of moments estimation (MME) (e.g., Li et al. (1999) and Yuen et al. (2015)). Due to the complexity of fitting multivariate zero-inflated Poisson models, the procedure of maximum likelihood estimation is difficult to implement and there are no closed form expressions for the MLEs. Li et al. (1999) suggest that confidence intervals constructed based on the MLEs may be wider than needed (particularly when Poisson means are large), but that intervals constructed based on MMEs may have less

[2] It is not uncommon in the literature to assume that λ_0 does not depend on covariates (e.g. Wang et al. (2003), Mohammadi et al. (2016)).

than desired coverage probabilty. Maximum likelihood estimation for BZIP model parameters can be implemented via an expectation-maximization (EM) algorithm (see Wang et al. (2003) and Li et al. (1999)). However, convergence of the EM algorithm for the BPZIP, is often dependent on initial values. Some propose the MMEs as initial values of parameters. Neverthless, Liu and Tian (2015) demonstrate that this is not always appropriate as the MMEs for the BZIP model may be outside the feasible region in which case they cannot be directly treated as the initial values of parameters.

4.2 The Bayesian Bivariate Zero-inflated Poisson Model

As a result of hierarchical Bayesian modeling and the availability of software to implement methods such as MCMC, Bayesian inference for multivariate zero-inflated models has become more prevalent in the literature. Applications of Bayesian multivariate zero-inflated Poisson models in the literature include occupational health data to assess significance of intervention on the reduction in the number of musculoskeletal and non-mukculosketal injuries (Wang et al. (2003)), outpatient psychiatric use data (Neelon et al. (2010)), plant population count data (Majumdar and Gries (2010)), automobile insurance claims for three different types of claims (Bermdez and Karlis (2011)), analysis of safety crash data at intersections Dong et al. (2014), and the joint modeling of the number of blood donation and the number of blood deferral (Mohammadi et al. (2016)). Bayesian inference in the aforementioned sources rely on diffuse priors for model parameters. Hence, despite the increased prevalence of Bayesian inference for bivariate zero-inflated Poisson models, use of informative prior structures applied to these models has not been explored in detail.

4.2.1 Diffuse Prior Structure: Bayesian BZIP Non-regression Model

We first consider the BZIP model in the absence of covariates. As with the BPZIP model, we assume that prior distributions for the zero-inflation parameters,

$\mathbf{p} = (p_0, p_1, p_2)$, and Poisson parameters, $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$, are independent. We will use the following conditional conjugate priors:

$$\mathbf{p} \sim \text{Dirichlet}(h_1, h_2, h_3, h_4),$$

and for $k = 0, 1, 2$,

$$\lambda_k \sim \text{Gamma}(c_k, d_k),$$

where h_k, $k = 1, \ldots, 4$, c_k and d_k are considered hyperparameters. In the absence of prior information regarding the zero-inflation parameters, uniform ($h_1 = h_2 = h_3 = h_4 = 1$) or Jeffreys prior ($h_1 = h_2 = h_3 = h_4 = 0.5$) are commonly used. In the absence of prior information regarding the Poisson parameters, small values of c_k and d_k for the prior on the λ_k's result in a diffuse prior for λ_k with large variance. For example, $c_k = 1$ and $d_k = 0.01$ for $k = 0, 1, 2$, where c_k and d_k are the shape and rate parameter, respectively. In Section 2.4, we demonstrated that such "diffuse" priors can be problematic.

4.2.2 Diffuse Prior Structure: Bayesian BZIP Regression Model

Consider the BZIP regression model outlined in Section 4.1.1, where $\mathbf{p}$ and $\boldsymbol{\lambda}$ both depend on covariates. As in the non-regression case, we assume that the prior distributions for all parameters are independent. Let $\boldsymbol{\gamma}_k = (\gamma_{k,0}, \ldots, \gamma_{k,q})$ be a $1 \times (q+1)$ vector of regression parameters consisting of an intercept and q covariates corresponding to λ_k. Further, let $\boldsymbol{\beta}_j = (\beta_{j,0}, \ldots, \beta_{j,r})$ be a $1 \times (r+1)$ vector of regression parameters consisting of an intercept and r covariates corresponding to p_j. We assume the elements of $\boldsymbol{\gamma}_k$ are mutually independent and independent of the elements of $\boldsymbol{\beta}_j$. It follows that the joint prior distribution is given by

$$\pi(\boldsymbol{\beta}_j, \boldsymbol{\gamma}_k) = \prod_{k=0}^{2} \prod_{v=0}^{q} \pi(\gamma_{k,v}) \prod_{j=0}^{3} \prod_{v=0}^{r} \pi(\beta_{j,v})$$

A typical prior structure places independent diffuse normal priors on both sets of regression coefficients. That is,

$$\boldsymbol{\gamma}_k \sim \mathrm{N}_q(\mathbf{0}, \boldsymbol{\sigma}^2_{\boldsymbol{\gamma}}\mathbf{I}_q),$$

and

$$\boldsymbol{\beta}_j \sim \mathrm{N}_r(\mathbf{0}, \boldsymbol{\sigma}^2_{\boldsymbol{\beta}}\mathbf{I}_r),$$

where $k = 0, 1, 2$ and $\sigma^2_{\boldsymbol{\gamma}}$ and $\sigma^2_{\boldsymbol{\beta}}$ are chosen to be large (e.g. 10^3) to express absence of prior information. Here $N_d(\boldsymbol{\varphi}, \boldsymbol{\Lambda})$ denotes a d-variate normal distribution with mean vector $\boldsymbol{\varphi}$ and covariance matrix, $\boldsymbol{\Lambda}$. Alternatively, we can place inverse-gamma priors on $\sigma^2_{\boldsymbol{\gamma}}$ and $\sigma^2_{\boldsymbol{\beta}}$, or proper uniform prior distributions on $\sigma_{\boldsymbol{\gamma}}$ and $\sigma_{\boldsymbol{\beta}}$.

The prior structure outlined above is common in the literature. Neverthless, we have found that, in practice, using diffuse priors for the regression coefficients in this model leads to convergence issues in MCMC implementations. For example, Majumdar and Gries (2010) presents a Bayesian analysis for bivariate plant counts in which they are primarily interested in estimation of the expected value of the two counts, $\mathrm{E}(Y_1)$ and $\mathrm{E}(Y_2)$, and the probability of zero counts, $\Pr(Y_1 = 0, Y_2 = 0)$, $\Pr(Y_1 = 0)$, and $\Pr(Y_2 = 0)$. Note that these are all functions of μ_1 and μ_2, which are updated by the data per the marginal distributions in (4.5) and (4.1). In discussion of convergence of the Gibbs sampler, Majumdar and Gries (2010) note that the corresponding summary diagnostics for these quantities indicate no potential problem with the sampler and yield consistent posterior estimates. In particular, they remark that "These diagnostics should not be taken as a proof of convergence of the chains, however if there were any problems, usually the diagnostic factors point to some potential problems." Nevertheless, we have found that this is often not the case (see Appendix 4.0.1). Via simulation we observed that convergence of the expectations and probabilities of interest does not necessarily indicate convergence of λ_0, λ_1, and λ_2, making posterior inference for such parameters problematic. It is not

surprising that diagnostics indicate convergence of these expectations and probabilities. For example, note that (4.6), (4.8) and (4.2) indicate that $\mathrm{E}(Y_1)$, $\mathrm{E}(Y_2)$ and $\Pr(Y_1 = 0, Y_2 = 0)$ are functions of identifiable quantities. Thus, these quantities are also identifiable.[3]

4.2.3 Posterior Inference for BZIP Model

Suppose we have n bivariate outcomes where the ith observation is represented by $\mathbf{y}_i = (y_{1i}, y_{2i})$ for $i = 1, \ldots, n$ and $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ represents the vector of the observed bivariate responses. Consider the non-regression case where $\ell(\mathbf{p}, \boldsymbol{\lambda} \,|\, \mathbf{y})$ denotes the likelihood given by (4.3). We assume that the prior distributions for all parameters are independent. Let $\pi(p_j)$ denote the prior distribution for p_j, $j = 0, 1, 2, 3$, and let $\pi(\lambda_k)$ denote the prior distribution λ_k, $k = 0, 1, 2$. Then, the joint posterior distribution of $(\mathbf{p}, \boldsymbol{\lambda})$ is given by

$$\pi(\mathbf{p}, \boldsymbol{\lambda} | \mathbf{y}) \propto \ell(\mathbf{p}, \boldsymbol{\lambda} \,|\, \mathbf{y}) \left\{ \prod_{j=0}^{3} \pi(p_j) \prod_{k=0}^{2} \pi(\lambda_k) \right\},$$

which has no closed form-expression. Thus, posterior computation proceeds using MCMC methods, such as Gibbs sampling, to sample from the posterior distribution.

Similar to posterior inference for the BPZIP regression model (discussed in Section 2.2.5), inference for the BZIP model relies on representing Y_1 and Y_2 in terms of latent variables. From the model in (4.1) we have that (Y_1, Y_2) consists of four underlying sub-populations (or mixtures);

(1) Subjects not at-risk for Y_1 or Y_2

(2) Subjects at-risk for Y_1 and not at-risk for Y_2

(3) Subjects not at-risk for Y_1 and at-risk for Y_2

[3] Note that functions of identifiable quantities are identifiable. However, sums that are identifiable with a common summand (e.g., λ_0) does not imply that the common summand is identifiable. Similarly, identifiable sums does not imply identifiable summands (Gustafson (2015)).

(4) Subjects at-risk for Y_1 and Y_2,

where "not at-risk" represents the case where a structural zero is always observed, and "at-risk" represents the case where either a non-zero count or sampling zero is observed. We do not actually observe which sub-population (or mixture) each subject is from. We can represent the BZIP random variables in terms of latent (unobserved) variables. In particular, we define the random variables Y_1 and Y_2 in terms of latent variables $\mathbf{U}$ and $\mathbf{X}$, where $\mathbf{U} = (U_0, U_1, U_2, U_3)$, $U_3 = 1 - U_0 - U_1 - U_2$ and $\mathbf{X} = (X_1, X_2, X_0)$ such that

$$Y_1 = (1 - U_0)(1 - U_2)(X_1 + X_0), \tag{4.13}$$

and

$$Y_2 = (1 - U_0)(1 - U_1)(X_2 + X_0), \tag{4.14}$$

where $\mathbf{U} \sim \text{Multinomial}(1, \mathbf{p})$. That is, $\mathbf{U} = \mathbf{u}$, where

$$\mathbf{u} = \begin{cases} (1,0,0,0), & \text{if } y_{1i} = 0 \text{ and } y_{2i} = 0 \\ (0,1,0,0), & \text{if } y_{1i} \neq 0 \text{ and } y_{2i} = 0 \\ (0,0,1,0), & \text{if } y_{1i} = 0 \text{ and } y_{2i} \neq 0 \\ (0,0,0,1), & \text{if } y_{1i} \neq 0 \text{ and } y_{2i} \neq 0. \end{cases}$$

Further, the latent variables, X_1, X_2, and X_0 are the underlying independent Poisson random variables with means λ_1, λ_2, and λ_0, respectively that form the standard BP (see Appendix A). That is, $Y_1 = X_1 + X_0$ and $Y_2 = X_2 + X_0$, such that the observed data are the counts Y_1 and Y_2 and the unobserved data are the counts, X_1, X_2 and X_0.[4] Denote the model parameters by $\boldsymbol{\theta} = (\mathbf{p}, \boldsymbol{\lambda})$, the observed data by $\mathbf{Y} = (Y_1, Y_2)$ and the latent (unobserved) data by $\mathbf{Z} = (\mathbf{U}, \mathbf{X})$. We are interested in the posterior distribution, $\pi(\boldsymbol{\theta}|\mathbf{Y})$, however, this is difficult to compute directly. To ease computation, we obtain the posterior, $\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Z})$ (often referred to as the

[4] Per (4.13) and (4.14), Y_1 and Y_2 both depend on four latent variables. This is in contrast to the BPZIP model, where Y_1 and Y_2 depend on differing numbers of latent variables (i.e. Y_1 depends on three, and Y_2 depends on two).

augmented data posterior, Tanner and Wong (1987)), which is more straightforward to compute. Instead of sampling directly from the posterior, $\pi(\boldsymbol{\theta} \,|\, \mathbf{Y})$, we sample from the posterior, $\pi(\boldsymbol{\theta}, \mathbf{Z} \,|\, \mathbf{Y})$. In order to implement this method within a Gibbs sampling framework, we must be able to sample from two conditional distributions, namely the posterior distribution of the augmented data, $\pi(\boldsymbol{\theta} \,|\, \mathbf{Y}, \mathbf{Z})$, and $\pi(\mathbf{Z}|\boldsymbol{\theta}, \mathbf{Y})$.

This can be extended to the regression model. The joint posterior distribution is given by

$$\pi(\mathbf{U}, \mathbf{X}, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_k, \boldsymbol{\sigma}^2_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}^2_{\boldsymbol{\beta}} \,|\, \mathbf{Y}) \propto \ell(\mathbf{U}, \mathbf{X}, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_k|\mathbf{Y}) \prod_{k=0}^{2} \left\{ \prod_{v=0}^{q} \pi(\boldsymbol{\gamma}_{k,v}) \right\} \prod_{j=0}^{2} \left\{ \prod_{r=0}^{r} \pi(\boldsymbol{\beta}_{j,r}) \right\},$$

which again has no closed form. Thus, for posterior inference we use MCMC methods, such as Gibbs sampling, to sample from the full conditionals, namely, $\pi(\boldsymbol{\theta}_{jk} \,|\, \mathbf{Y}, \mathbf{Z})$ and $\pi(\mathbf{Z}|\boldsymbol{\theta}_{jk}, \mathbf{Y})$, where here $\boldsymbol{\theta}_{jk} = (\boldsymbol{\beta}_j, \boldsymbol{\gamma}_k)$. This algorithm can be readily implemented in software such as WinBUGS or OpenBUGS (for both the non-regression case and regression case). For this dissertation, inference for the BZIP model was carried out using WinBUGS through the R package `R2WinBUGS`.[5]

4.2.4 BZIP Prior and Posterior Predictive Distributions

The prior and posterior predictive distributions are commonly used in the implementation of Bayesian analysis for prediction. By using both of these joint distributions, we can model our uncertainty completely. The prior predictive distribution is the expected value of the likelihood with respect to the prior. Often we are interested in predicting the "next" observation or observations (e.g. gauging the prospects for a future sample). That is, suppose we want to predict future bivariate observation(s) denoted $\tilde{\mathbf{y}}$ (assumed independent from our data $\mathbf{y}$). We can make use of the posterior predictive distribution, which is defined as the expected value of

[5] We explored implementation of the BZIP model in JAGS, however, specification of a multinomial distribution for $\mathbf{u}$ proved problematic (JAGS prompted that actual data be provided for $\mathbf{u}$, instead of sampled values from a Dirichlet prior). In addition, implementation of this model was explored in STAN. However, at the time STAN was not recommended for multivariate analysis.

the BZIP likelihood, evaluated at $\tilde{\mathbf{y}}$, with respect to the BZIP posterior distribution given the data. The posterior predictive distribution for the BZIP model has the form

$$\tilde{\pi}(\tilde{\mathbf{y}}|\mathbf{y},\tilde{\mathbf{x}}) = \int_{\Theta} \ell(\tilde{\mathbf{y}} \,|\, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_k, \mathbf{y}, \mathbf{x}) \pi(\boldsymbol{\beta}_j, \boldsymbol{\gamma}_k \,|\, \mathbf{y}, \mathbf{x}) d\boldsymbol{\theta}, \tag{4.15}$$

where $\mathbf{x}$ is the vector of covariates for the current sample, and $\tilde{\mathbf{x}}$ is the vector of covariates corresponding to the future observations, and $\boldsymbol{\theta}$ is the vector of parameters defined on Θ.

4.3 BZIP Examples: Problems with Posterior Inference

Similar to Section 2.4, we constructed a variety examples for posterior inference of our BZIP model. As with the BPZIP models in Section 2.4, we found convergence issues in MCMC implementations of our BZIP model, particularly in a diffuse prior setting. Namely, there is a lack of convergence for the Poisson parameters, λ_k. As before, the reason for these difficulties is a lack of identifiability. This is hardly surprising given the additional parameters needed for the BZIP model. Despite the addition of zero-inflation parameters, the mixture probabilities remain identifiable due to the constraint that $\sum_{j=0}^{3} p_j = 1$. Nevertheless, we observed that, in some cases, the model struggles to identify which of the four mixtures an observation is from. In such cases, diffuse priors for the zero-inflation parameters presented in Section 4.2.1 and Section 4.2.2 are not appropriate. This is particularly the case for small n ($n = 50, 100$) and small values of λ_k, $k = 0, 1, 2$. In this section, we demonstrate a situation in which we have poor estimation of the mixture probabilities. We refer the reader to Appendix 4.0.1 for an example demonstrating the lack of convergence for the BZIP Poisson parameters.

4.3.1 R2WinBUGS Specifications for Posterior Inference

Posterior inference was carried out with WinBUGS using the R package `R2WinBUGS`. Inference was done with two chains, and initial values were specified for each chain.

We ran 300,000 iterations and used the first 100,000 iterations as a burn-in. For the remaining 200,000 iterations we sampled every 10th value to reduce autocorrelation. Accordingly, 20,000 parameter values were retained for each chain.

4.3.2 Posterior Inference for Mixture Probabilities

In this example, we illustrate an example where diffuse priors on $\mathbf{p}$ prove problematic and results in poor posterior estimates of the p_j's. We generated $n = 50$ observations from a BZIP model with true values of parameters, $\mathbf{p} = (0.65, 0.15, 0.15, 0.05)$, $\boldsymbol{\lambda} = (1, 1, 0.5)$. The data are depicted in Figure 4.1.

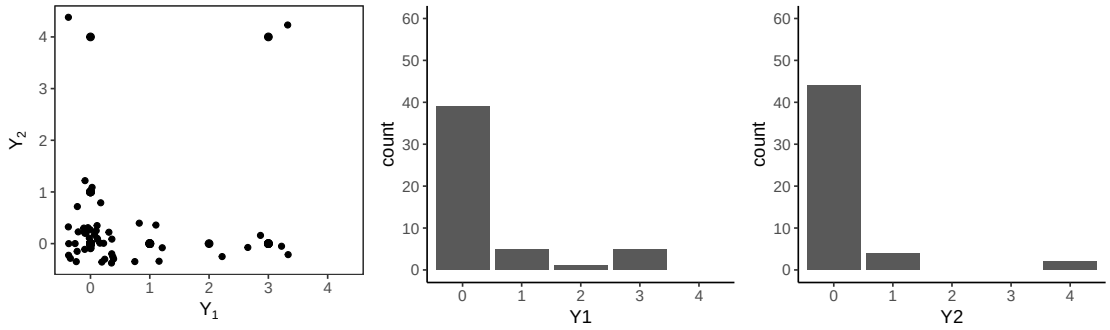

Figure 4.1: Data for $n = 50$ observations generated from a BZIP model with true values $\mathbf{p} = (0.65, 0.15, 0.15, 0.05)$, $\boldsymbol{\lambda} = (1, 1, 0.5)$, and $n = 50$.

We consider diffuse normal priors for regression coefficients, namely, for $j = 0, 1, 2$,

$$\beta_j \sim \mathrm{N}(0, \sigma^2 = 1,000).$$

The resulting induced priors on the p_j's are shown in Figure 4.2. Posterior inference was carried out with WinBUGS using the same specifications as described in Section 4.3.1. The resulting posterior densities and summary results are shown in Figure 4.3 and Table 4.1.

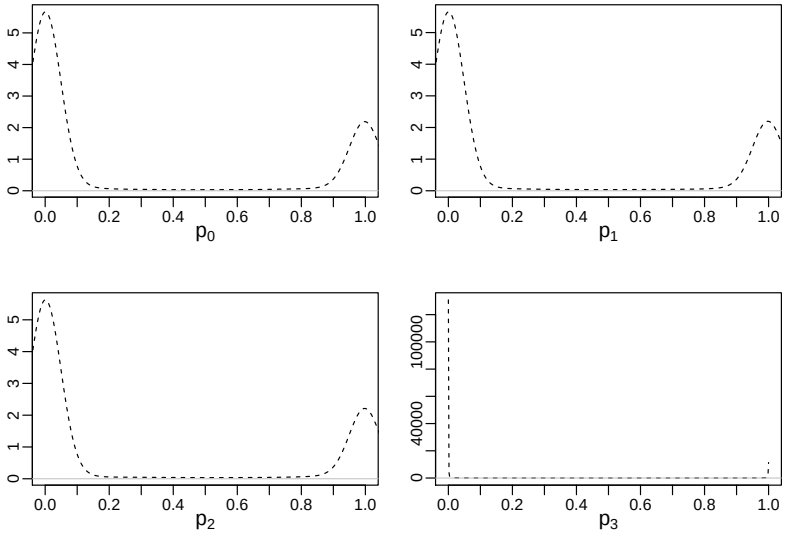

Figure 4.2: Induced priors on p_j for diffuse normal priors on regression coefficients.

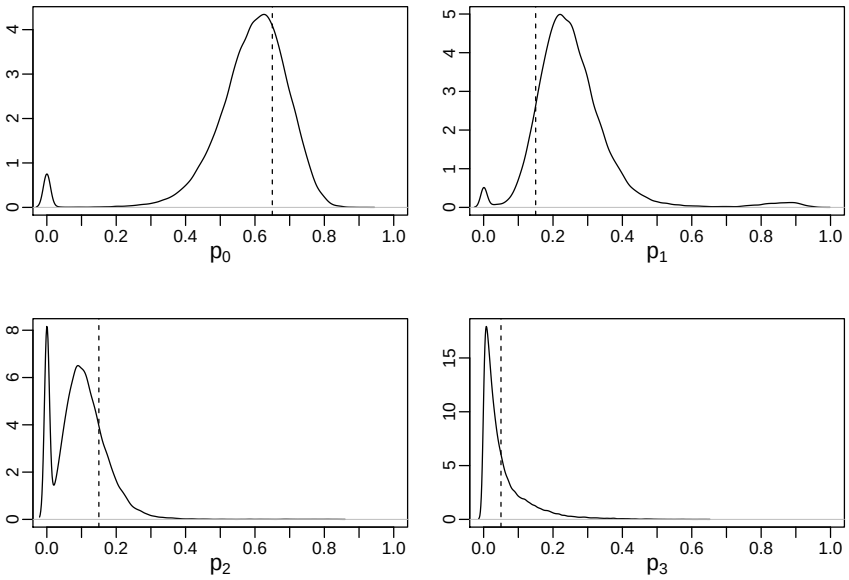

Figure 4.3: Posterior densities (black, solid line) and true value (black, dashed vertical line) for mixture probabilities with diffuse priors.

Table 4.1: Posterior summary results for mixture probabilities and $n = 50$.

Parameter	Truth	Mean	SD	2.5%	97.5%	Width
p_0	0.65	0.5852	0.1254	0.2893	0.7575	0.4682
p_1	0.15	0.2578	0.1198	0.0941	0.5310	0.4369
p_2	0.15	0.1006	0.0713	0.0000	0.2513	0.2513
p_3	0.05	0.0564	0.0696	0.0008	0.2556	0.2548

The irregular, bimodal posteriors reflect that there is little updating *a posteriori*. These examples demonstrate that, for small n and diffuse priors, we have poor estimation of the p_j's due to the small amount of observed data to update the priors. Furthermore, in this example, the Poisson rates were chosen to be small (close to zero) to illustrate that, for small λ_k and small n, it appears that the model struggles to identify which of the four mixtures in (4.1) the bivariate observations (Y_1, Y_2) are from. Increasing the initial burn-in length, chain length and thinning rate did not improve posterior results.

Now suppose we increase the sample size to $n = 100$. The data generated from a a BZIP model with true values $\mathbf{p} = (0.65, 0.15, 0.15, 0.05)$ and $\boldsymbol{\lambda} = (1, 1, 0.5)$ is shown in Figure 4.4.

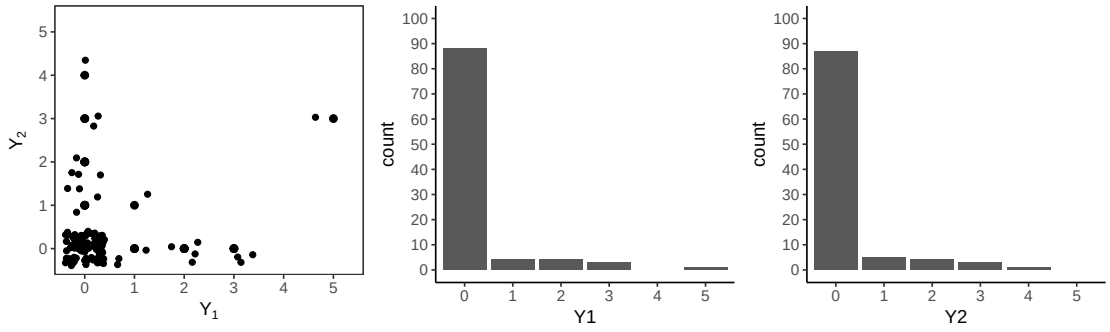

Figure 4.4: Data for $n = 100$ observations generated from a BZIP model with true values $\mathbf{p} = (0.65, 0.15, 0.15, 0.05)$, $\boldsymbol{\lambda} = (1, 1, 0.5)$.

We again consider diffuse priors on regression coefficients as shown in Figure 4.2. We repeat inference in WinBUGS using the same specifications as in Section 4.3.1. The resulting posterior densities and summary results are shown in Figure 4.5 and Table 4.2.

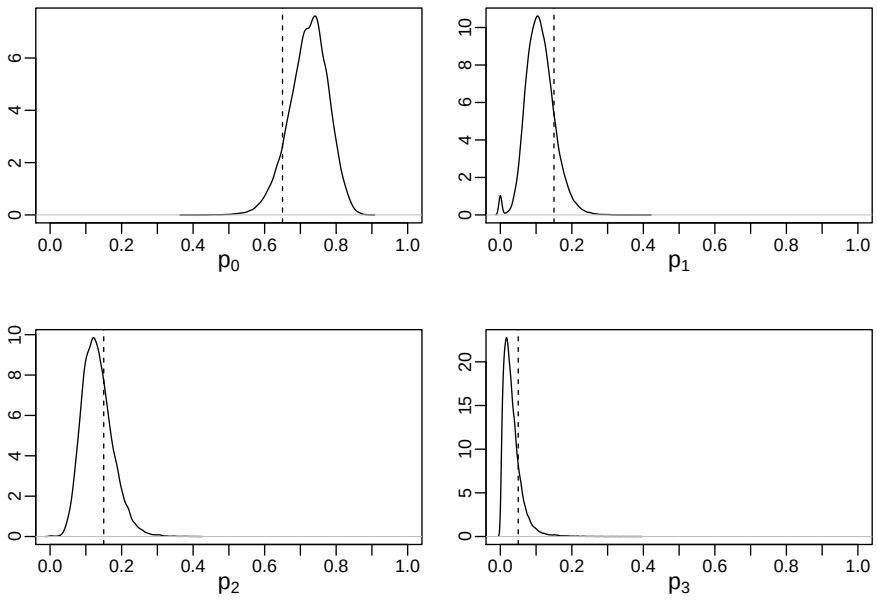

Figure 4.5: Posterior densities (black, solid line) and true value (black, dashed vertical line) for diffuse priors $\boldsymbol{\beta}_j$ for mixture probabilities.

Table 4.2: Posterior summary results for mixture probabilities and $n = 100$.

Parameter	Truth	Mean	SD	2.5%	97.5%	Width
p_0	0.65	0.7225	0.0545	0.6062	0.8197	0.2135
p_1	0.15	0.1117	0.0401	0.0422	0.1984	0.1562
p_2	0.15	0.1324	0.0429	0.0619	0.2279	0.1660
p_3	0.05	0.0334	0.0269	0.0038	0.1005	0.0967

The densities in Figure 4.5 show more posterior updating compared to the posterior densities in Figure 4.3. Moreover, Figure 4.5 and Table 4.2 indicate that, for this example, increasing the sample size to $n = 100$, improves estimation of the

mixture probabilities. In general, however, we observed that increasing sample size does not improve estimation of the p_j's.

We constructed a variety of other examples leading to the following tentative conclusions with respect to inference for BZIP mixture probabilities. Specifically, we identified certain situations that suggest value in informative priors. For example, as p_0 increases to one, the potential value of informative priors for posterior inference strengthens. This is because the effective sample size (the number of non-zero data) becomes smaller as the probability that both Y_1 and Y_2, mainly estimated by p_0, increases to one. In addition, we observed that, when the true value of p_3 is small relative to the other mixture probabilities, informative priors are worth exploring. Small values of p_3 indicate that the number of non-zero counts is small relative to the number of zero counts, which in turn can potentially impact the estimation of the Poisson rates.

4.4 Additional Remarks: The Allure of Nonidentifiability and Possible Approaches

In Section 2.5, we discussed common approaches in the literature for handling the issue of nonidentifiability. The approaches discussed therein can also be applied to the BZIP model. For example, identifiability can be obtained by setting $\lambda_1 = \lambda_2$ (i.e. assuming Y_1 and Y_2 are identically distributed), which implies that the counts, Y_1 and Y_2, have equal rates (see Yuen et al. (2015)). We could also obtain identifiability by assuming a fixed value as the prior for λ_0. Again, this is extremely informative and not reasonable in practice. Mohammadi et al. (2016) alleviate the effects of nonidentifiability by assuming different covariates to effect the individual Poisson rates. Such an approach may not be realistic in practice.

An alternative, as used in other types of models with unmeasured confounding components or measurement error models, is to add constraints. For example, we could assume a strict ordering of the Poisson rates: $0 < \lambda_0 < \lambda_1 < \lambda_2$. Another

option would be to apply constraints to the sums $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$. However, determining conceivable constraints on these sums would be a difficult task, almost unrealistically so. A potentially less extreme approach to mitigate problems nonidentifiability can cause is using moderately informative priors. In Chapters Two and Three, we discuss in detail methods for informative prior specification for the Poisson parameters. These methods can be readily implemented in the BZIP model. In contrast, the additional zero-inflation parameters requires more complex and more involved methods for prior elicitation from a subject-matter expert than those described for the BPZIP model. This will be a focus of what follows. Some applications in the literature (Wang et al. (2003), Mohammadi et al. (2016)) simplify (4.1) and instead represent the BZIP model as a mixture of a point mass at $(0, 0)$ and a bivariate Poisson. That is,

$$(Y_1, Y_2) \sim \begin{cases} (0,0), & \text{with probability } p_0, \\ \text{BP}(\lambda_0, \lambda_1, \lambda_2), & \text{with probability } 1 - p_0. \end{cases} \tag{4.16}$$

The corresponding joint probability mass function is

$$f_{(Y_1,Y_2)}(y_1, y_2 \,|\, p, \boldsymbol{\lambda}) = \begin{cases} p_0 + (1-p_0)\exp(-\lambda_0 - \lambda_1 - \lambda_2), & y_1 = 0, y_2 = 0 \\ (1-p_0) f_{BP}(y_1, y_2 \,|\, \lambda_0, \lambda_1, \lambda_2), & y_1 \neq 0 \text{ or } y_2 \neq 0, \end{cases}$$

where $f_{BP}(y_1, y_2 \,|\, \lambda_0, \lambda_1, \lambda_2)$ denotes the probability mass function of the standard BP distribution provided in Appendix A. Note that this representation reduces the number of zero-inflation parameters to one, accounting for the case that both Y_1 and Y_2 are structural zeros. For this simplified representation, the methods discussed for prior construction on the zero-inflation parameter in Chapters Two and Three are appropriate. If the researcher is only interested in inference for the probability that both outcomes are structural zeros and the probability that both outcomes are not zero-inflated then the simplified representation of the BZIP model in (4.16) is appropriate. However, if the researcher is interested in identifying observations

from the four subpopulations represented in (4.1), then the simplified model is not appropriate. For our purposes, we suppose that inference for the four zero-inflation parameters is of interest and consider the full BZIP model in (4.1).

4.5 Application: Adverse Event Study

We consider a hypothetical study to compare a new treatment with a standard of care with respect to safety. A study to evaluate the safety of this new drug has already been completed and the results suggest that the new treatment is safe. Now, we want to investigate whether the new treatment is superior to the current standard of care with respect to safety. Namely, we want to assess whether the new treatment reduces the number of two side effects, common to similar medications currently on the market. The study consists of $n = 100$ subjects that suffer from a variety of neurological disorders. Subjects are randomized to a treatment group (receive new treatment) or control group (receive current standard of care). We assume that the subjects are similar in characteristics such as age, gender, etc. Throughout the course of the six-month study, the subjects were asked to recall the number of occurrence of two adverse events, which are known to be related. We assume that the observed person-time is the same for each subject. Let (Y_{1i}, Y_{2i}) be a bivariate response count for the i^{th} subject such that

$$Y_{1i} = \text{Number of migraines experienced during study}$$
$$Y_{2i} = \text{Number of seizures experienced during study,}$$

for $i = 1, \ldots, n$, where n is the number of subjects. We assume $(Y_1, Y_2) \sim \text{BZIP}(\mathbf{p}, \boldsymbol{\lambda})$ is appropriate here. In contrast to the adverse event drug safety application considered in Chapters Two and Three, in this application we assume that both adverse events attribute to the zero-inflation.

From (4.1) we have that the data can arise from one of four mixtures:

(1) $(Y_{1i}, Y_{2i}) \sim (0, 0)$: the subject is not at-risk for migraines or seizures (with probability p_0).

(2) $(Y_{1i}, Y_{2i}) \sim (\text{Poisson}(\mu_1), 0)$: The subject is at-risk for migraines, but not at-risk for seizures (with probability p_1).

(3) $(Y_{1i}, Y_{2i}) \sim (0, \text{Poisson}(\mu_2))$: The subject is not at-risk for migraines, but is at risk for seizures (with probability p_2).

(4) $(Y_{1i}, Y_{2i}) \sim \text{BP}(\lambda_0, \lambda_1, \lambda_2)$: The subject is at-risk for both the migraines and seizures (with probability p_3).

The data we generated to simulate this hypothetical study are depicted in Figures 4.6 and 4.7.

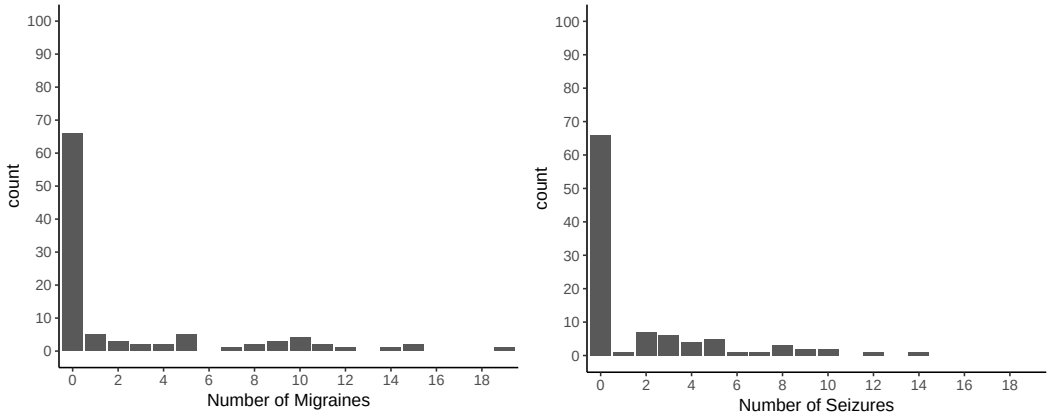

Figure 4.6: Marginal data for the number of migraines and the number of seizures for $n = 100$ subjects.

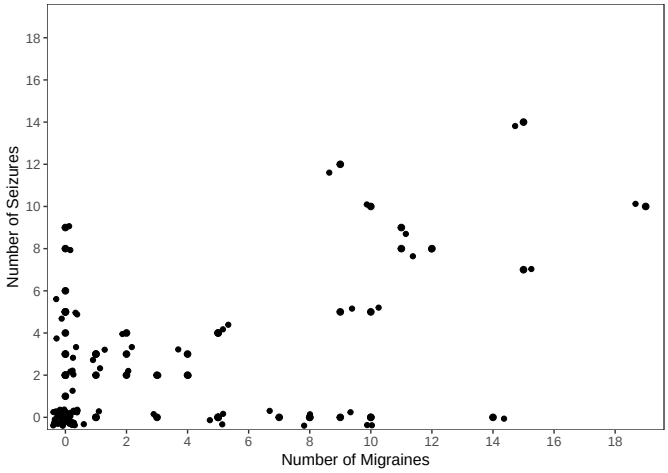

Figure 4.7: Scatterplot illustrating the joint association between the number of migraines and the number of seizures for $n = 100$ subjects.

For our purposes we assume that the Poisson parameters depend on a binary covariate that indicates treatment group for the ith subject. That is, we assume the kth Poisson mean, $k = 0, 1, 2$ and ith subject

$$\log(\lambda_{k,i}) = \gamma_{k,0} + \gamma_{k,1} x_i, \tag{4.17}$$

where

$$x_i \equiv \begin{cases} 1, & \text{new treatment,} \\ 0, & \text{current standard of care.} \end{cases} \tag{4.18}$$

We are interested in the effect, if any, this new treatment has on the reduction of the occurrence of these two adverse events. Further, we want to compare the new treatment to the current standard of care. Table 4.3 contains the model parameter summary for the ith subject.

Table 4.3: Model parameter summary for subject i.

Parameter	Interpretation
$\lambda_{1,i}$	The mean rate of exclusive migraines for subject i.
$\lambda_{2,i}$	The mean rate of exclusive seizures for subject i.
$\lambda_{0,i}$	The mean rate of both migraines and seizures for subject i.
$\gamma_{k,0}$	Average log adverse event rate for a unit for current standard of care.
$\gamma_{k,1}$	Average change in log rate of migraines for the ith subject in the new treatment group; $e^{\gamma_{k,1}}$ represents the increased (or decreased) odds of adverse event rate for a subject receiving the new treatment.
p_{0i}	The proportion of subjects not at-risk for migraines or seizures.
p_{1i}	The proportion of subjects that are at-risk for migraines, but are not at-risk for seizures.
p_{2i}	The proportion of subjects that are not at-risk for migraines, but are at risk for seizures.
p_{3i}	The proportion of subjects that are at-risk for migraines and seizures.

To evaluate whether the new treatment is superior to the current standard of care with respect to safety, suppose we are interested in the posterior predictive probability that among subjects that experience both adverse events, the number of migraines, the number of seizures, and the number of both migraines and seizures, is less for the new treatment compared to the current standard of care. That is, for some probability, δ, we are interested in the posterior predictive probability that

$$\Pr(Y_{1,1} < Y_{1,0} \,|\, \text{data}) \geq \delta, \tag{4.19}$$

$$\Pr(Y_{2,1} < Y_{2,0} \,|\, \text{data}) \geq \delta, \tag{4.20}$$

and

$$\Pr(Y_{1,1} + Y_{2,1} < Y_{1,0} + Y_{2,0} \,|\, \text{data}) \geq \delta, \tag{4.21}$$

where $Y_{1,0}$ and $Y_{2,0}$ denote the number of migraines and the number of seizures, respectively, for the current standard of care, and $Y_{1,1}$ and $Y_{2,1}$ denote the number of migraines and the number of seizures, respectively, for the new treatment. Here, we assume $\delta = 0.90$.

4.6 Informative Prior Structure for Bayesian BZIP Model

In Chapter Two we suggest the use of informative priors as a method to alleviate the effects that nonidentifiability can cause. We specifically consider prior elicitation from a subject-matter expert to construct informative priors for model parameters. The informative priors proposed for the Poisson rates in Chapters Two and Three can easily be extended to the BZIP model. In contrast, due to the increased number of zero-inflation parameters, the methods for prior construction used for the BPZIP zero-inflation parameter are not sufficient for the BZIP model. Thus, here we focus on methods of prior construction for $\mathbf{p}$ and refer the reader to Chapters Two and Three for prior construction for Poisson parameters.

The BZIP model requires prior specifications for three zero-inflation parameters (the distribution for the fourth zero-inflation parameter can be determined using the constraint that $\sum_{j=0}^{3} p_j = 1$). This is in contrast to the BPZIP model, which requires prior specification for a single zero-inflation parameter. In Section 4.3 we showed that a diffuse Dirichlet prior distribution on $\mathbf{p}$ can be problematic in some cases. In these cases, providing information on the p_j's is essential for estimation of the p_j's (and model convergence in some cases). We adapt the methods proposed in Elfadaly and Garthwaite (2013a) and propose plausible methods of prior elicitation for the BZIP zero-inflation parameters. This involves eliciting a Dirichlet distribution for a multinomial model using conditional beta variates (recall that in

a BZIP model we assume the zero-inflation parameters are multinomial and that each observation can come from only one mixture). To our knowledge, methods for eliciting an informative Dirichlet prior distribution has not been applied to mixture probabilities in a bivariate or multivariate zero-inflated Poisson model setting.

4.6.1 Dirichlet Prior: Zero-inflation Parameters

A BZIP model assumes a multinomial distribution for the zero-inflation parameters, p_j. Hence, we assume that the p_j's are independent and each bivariate observation (Y_1, Y_2) belongs to one and only one of the four mixtures that make up the BZIP model in (4.1). In Bayesian statistics, it is well known that the Dirichlet distribution is a conjugate prior for the probability parameters of multinomial models. This distribution preserves the unit-sum constraint for the probability parameters of multinomial probabilities so that a subject matter expert's assessment must satisfy a number of requirements for statistical coherence (Elfadaly and Garthwaite (2013a)). For example, if there are only two categories, the lower probability quartile of one category and the upper probability quartile of the other category must sum to one. As the number of categories increases, the number of requirements also increases. These requirements are complex and have no simple closed form. As statisticians, a critical objective of the elicitation process is to choose assessment tasks that lead to a coherent set of assessments, preferably without the expert having to be conscious of the statistical requirements.

In general, eliciting parameters of multivariate distributions is not an easy task. Elfadaly and Garthweite (2013a and 2013b) propose methods for eliciting an informative prior distribution for multinomial models. In particular, they propose methods for quantifying expert opinion about the hyperparameters of both the standard Dirichlet distribution and a generalized Dirichlet distribution, referred to as the Connor-Mosimann, both of which are conjugate prior distributions for a multino-

mial model. Elfadaly and Garthweite (2013a and 2013b) apply their methods within the context of multinomial proportions that do not depend on covariates. This is done based on conditional quartile assessments of beta conditional distributions. Decomposition of the Dirichlet elicitation process into the assessment of several beta distributions helps reduce the complexity of eliciting a multivariate distribution. The median and quartiles of univariate beta distributions are assessed and translated in to the parameters of a Dirichlet distribution. We discuss construction of a Dirichlet distribution and suggest plausible assessment tasks to prompt an expert about their expectations of the zero-inflation parameters.

4.6.2 Prior Assessment Tasks for Elicitation of a Dirichlet Distribution

Elicitiation of multivariate distributions often require difficult assessment tasks by an expert. Accordingly, Elfadaly and Garthwaite (2013a) propose elicitation methods for the Dirichlet distribution that involve asking the expert to think about univariate distributions. The assessment tasks for the elicitation of a Dirichlet involve asking the expert to assess conditional quartiles. We adapt the assessment tasks outlined in Elfadaly and Garthwaite (2013a) to apply to the BZIP zero-inflation parameters, which has not done previously. For $j = 4$ mixture probabilities, the conjugate Dirichlet prior has the form

$$\pi(p_1, \ldots, p_4) = \frac{\Gamma(N)}{\Gamma(h_1) \cdots \Gamma(h_4)} p_1^{h_1 - 1} \cdots, p_4^{h_4 - 1}, \tag{4.22}$$

where $N = \sum_{j=1}^{4} h_j$ and $h_j > 0$. It can be shown that N is the prior effective sample size for a Dirichlet prior and multinomial likelihood. To elicit the vector of hyperparameters, $\mathbf{h} = (h_1, \ldots, h_4)$, we can make use of the fact that the conditional distributions of the Dirichlet variates are scaled beta distributions.

Consider our adverse event example in Section 4.5, where we assume that the zero-inflation parameters do not depend on covariates. In this example, the zero-inflation parameters, $\mathbf{p}$, form the multinomial categories that we want to quantify

expert opinion about. For reference, the interpretations of the zero-inflation parameters in the context of the adverse event example are presented in Table 4.3. This information will be essential in the assessment tasks for the prior elicitation process. In this example, we assume that some of the $n = 100$ subjects will not be at-risk for either adverse event, whereas some subjects will be at-risk for one of both adverse event. Furthermore, we assume we have a well informed expert that is familiar with medications similar to that under study, and familiar with the subject population.

To elicit a Dirichlet prior distribution for $\mathbf{p}$ based on univariate beta distributions, the expert is first asked to order the mixture proportions from the most likely (probable) to the least likely (probable). Elfadaly and Garthwaite (2013a) suggest prompting the expert to order in this way because the conditional distributions to be assessed are less skewed which may lead to easier assessment tasks. That is the distribution that we expect to be the most skewed (corresponding to the smallest mixture proportion) will be determined automatically based on the expert's assessment of the other three mixtures.

Suppose, for our example, the subject-matter expert gives the following order:

$$p_0 > p_3 > p_1 > p_2. \tag{4.23}$$

That is, the expert believes that the majority of the 100 subjects will not be at-risk for migraines or seizures. Based on the expert's ordering in (4.23) we begin by eliciting information about the percentage of subjects not at-risk for migraines or seizures. The expert is asked a series of three of questions (one for assessment of a median value, one for assessment of the lower quartile and one for assessment of a upper quartile) in order to represent the expert's knowledge and uncertainty.

(1) To assess the median, the expert is asked "Suppose we have 100 subjects. What, do you think, is the percentage of these subjects that are not at-risk for either adverse event?" Their response will be the median value.

(2) Next, to assess a lower quartile,"Suppose the percentage of subjects that are not at-risk for either adverse event is actually *less than* your initial assessment (that is, their median assessment is too high). Given this information, what would you now estimate as the percentage?" Their response is taken to be the 25th percentile.

(3) Finally, to obtain an upper quartile the expert is asked "Suppose the percentage of subjects that are not at-risk for either adverse event is actually *greater than* your initial assessment (that is, their median assessment is too low). Given this information, what would you now estimate as the percentage?" Their response is taken to be the 75th percentile.

Suppose the expert's assessment of the lower, middle (median) and upper quartile are $p_{0,0.25} = 0.48$, $p_{0,0.50} = 0.55$, and $p_{0,0.75} = 0.62$, respectively. This information is then translated into the parameters of a beta distribution using numerical methods.

The elicitation of the remaining zero-inflation parameters relies on assessment of conditional quartiles. Per the expert's initial ordering, we next ask the expert about their expectations for the percentage of the 100 subjects that are at-risk for both adverse events. When making assessments about this percentage, the expert is told to assume that the percentage of subjects that are not at-risk to experience either adverse event *is* 55% (i.e. their median assessment). In this way, the expert is actually making assessments about $p_3|p_0$, instead of just p_3. The expert is again asked a series of three questions:

(1) To assess the median of $p_3|p_0$, the expert is asked "Given that the percentage of the 100 subjects not at-risk for either adverse event is 55%, what, do you think, is the percentage of the remaining subjects that are at-risk for *both* adverse events?" Their response is taken to be the median value.

(2) Next, to assess a lower quartile for $p_3|p_0$, the expert is asked,"Suppose the percentage of subjects that are at-risk for *both* adverse events is actually *less than* your initial assessment (that is, their median assessment is too high). Given this information, what would you now estimate as the percentage?" Their response is taken to be the 25th percentile.

(3) Finally, to obtain an upper quartile for $p_3|p_0$, the expert is asked "Suppose the percentage of subjects at-risk for *both* adverse events is actually *greater than* your initial assessment (that is, their median assessment is too low). Given this information, what would you now estimate as the percentage?" Their response is taken to be the 75th percentile.

Suppose the experts assessments for the lower, middle, and upper quartile are $p_{3,0.25}|(p_0 = 0.55) = 0.13$, $p_{3,0.50}|(p_0 = 0.55) = 0.20$, and $p_{3,0.75}|(p_0 = 0.55) = 0.27$, respectively. These conditional quartile assessments can be translated into the parameters of a conditional scaled beta distribution.

Similarly, we elicit information about the distribution of $p_1|p_0, p_3$. In their assessments about the percentage of subjects just at risk migraines, the expert is told to assume their initial (median) assessments for p_0 and p_3 are exactly $p_{0,0.50} = 0.55$ and $p_{3,0.50} = 0.20$. That is, the expert is asked to assume that 55% of the 100 subjects are not at-risk for both adverse events and 20% are at risk for both adverse events. The expert is then prompted to relay a median, lower and upper quartile for the percentage of the remaining subjects that are at-risk for just migraines. Suppose the expert's assessments for the lower, middle, and upper quartile are $p_{1,0.25}|(p_0 = 0.55, p_3 = 0.20) = 0.10$, $p_{1,0.50}|(p_0 = 0.55, p_1 = 0.20) = 0.15$, and $p_{1,0.75}|(p_0 = 0.55, p_3 = 0.20) = 0.20$, respectively.

Finally, the three conditional quartiles for the percentage of the 100 subjects that are at-risk for just seizures, $p_3|(p_0 = 0.55, p_3 = 0.20, p_1 = 0.15)$ is computed

automatically using the unit-sum constraint. The information collected from the expert is summarized in Table 4.4.

Table 4.4: Expert elicited conditional quartiles. Here the median is the expert's initial median assessment, and the 25th and 75th percentiles are the conditional lower and conditional upper quartiles, respectively.

Parameter	25th Percentile	Median	75th Percentile
p_0	0.48	0.55	0.62
$p_3 \,\vert\, p_0 = 0.55$	0.13	0.20	0.27
$p_1 \,\vert\, (p_0 = 0.55, p_3 = 0.20)$	0.10	0.15	0.20
$p_2 \,\vert\, (p_0 = 0.55, p_3 = 0.20, p_1 = 0.15)$	0.05	0.10	0.20

Often the prior elicitation process outlined above requires reconciliation of the assessed parameters for the beta distributions to ensure that the expert's assessment of the conditional quartiles satisfy mathematical properties of the Dirichlet distribution (see Elfadaly and Garthwaite (2013a) for more detail). Implementation of the above assessment tasks can be done using the free, readily available interactive PEGS-Dirichlet software (Elfadaly and Garthwaite (2013c)) developed by Elfadaly and Garthwaite (2013a). The software consists of several features that assist the subject-matter expert throughout the elicitation process. Figure 4.8 emulates some of the graphics involved of this software. To begin, the expert assesses the conditional medians. The top left plot in Figure 4.8 is a visual after the expert has assessed the conditional medians. The top of each bar represents the assessed medians by the expert. The red bars denote values that should be treated as the truth. Here, the expert specifies the median value of $p_0 = 0.55$, indicated by the red bar in the top left plot. Next, the expert specifies a value for $p_3 \,|\, p_0 = 0.55$. The red dotted lines above the bar for p_3 represents the maximum value for the assessed median of p_3 conditional on their assessment that $p_0 = 0.55$. Similarly, the red dotted line above p_1 indicates that given the experts assessments of $p_0 = 0.55$ and $p_3 = 0.20$, their as-

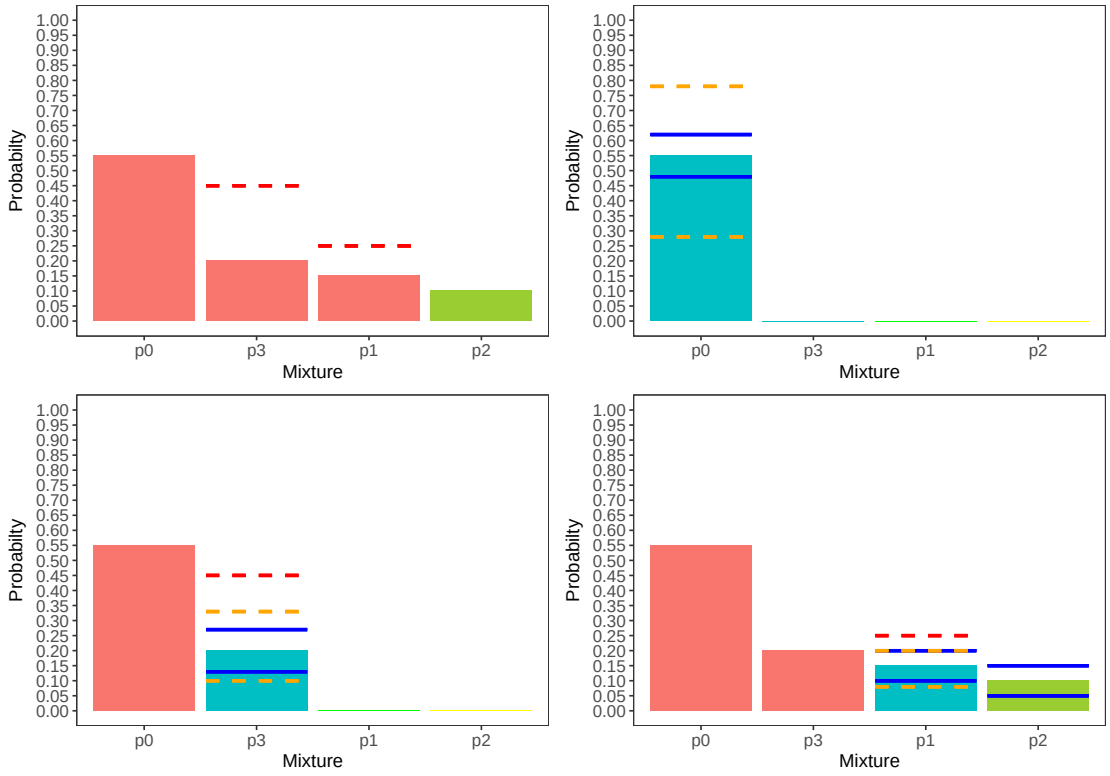

Figure 4.8: Prior elicitation of a Dirichlet distribution based on expert's assessments of conditional quartiles.

sessed median value for p_1 cannot exceed 0.25. In this way, the interactive software guides the expert to make assessments consistent with mathematical constraints, namely the unit-sum constraint. The median value for $p_2 \,|\, p_0, p_3, p_1$ (indicated by the green bar in the top left plot) is automatically computed by the software based on the expert's assessments for the previous three categories.

Next, the expert is asked to assess the upper and lower conditional quartiles. The expert begins by assessing the quartiles for the most likely category p_0, as shown by the top right plot in Figure 4.8. The orange dotted lines are suggested boundaries for the expert. Namely, the expert is advised to assess his/her conditional quartiles between these dotted lines to obtain a marginal distribution for p_0 that is unimodal with a mode that is neither near zero or one, which often is not representative of the

expert's opinion.[6] The blue solid lines represent the expert's assessments for the lower and upper quartile of p_0. Next, the expert assesses the conditional quartiles for $p_3 \,|\, p_0 = 0.55$. This is shown in the bottom left plot in Figure 4.8. The bar for p_0 becomes red to indicate that, when making assessments about p_3, the expert should assume that the true value for p_0 is 0.55. The red dotted line above the bar for p_3 in the bottom left graph represents an upper boundary for which p_3 cannot exceed based on the expert's assessment of p_0. The expert's assessed lower and upper quartile for $p_3 \,|\, p_0 = 0.55$ are indicated by the blue solid lines. Next, the expert is asked to asses the quartiles for p_1 assuming that the true values for p_0 and p_3 are 0.55 and 0.20, respectively. The expert's assessed conditional quartiles for p_1 are represented by the blue solid lines above p_1 in the bottom right plot of Figure 4.8. Finally, the upper and lower quartiles for the last category, p_2, are automatically computed based on the expert's assessments of the first three categories. This is represented by the blue solid lines over the yellow bar in the bottom left plot.

As the expert makes their assessments of the upper and lower quartiles, the software also provides a visual of the assessed distribution based on their assessments. In addition, to help the expert during the task, the software shows the resulting assessed marginal (for p_0) or conditional (for the remaining mixtures) beta distribution based on the experts assessments. This allows the expert to modify his/her assessments if they feel the resulting density does not accurately represent their beliefs. The marginal beta distribution for p_0 and conditional scaled beta distributions are shown in Figure 4.9. The mixture probabilities must sum to one and the quartile assessments for the different categories must also meet certain requirements. The PEGS-Dirichlet software computes the hyperparameters of the Dirichlet distribution by using the parameters of the beta distributions based on the expert's assessments

[6] Elfadaly and Garthwaite (2013a) developed this software for data that is not zero-inflated. If the expert believes that, for example, the vast majority of observations are not at-risk for either adverse event, the expert might make assessments near these boundaries.

as shown in Figure 4.9. This requires reconciliation of the expert's assessments in order to satisfy mathematical constraints (we refer the reader to Elfadaly and Garthwaite (2013a) for more details on this).

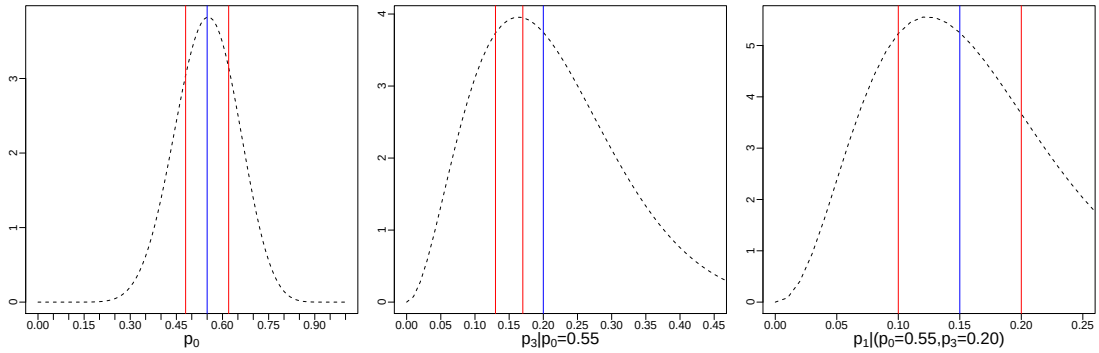

Figure 4.9: Marginal and conditional scaled distributions for mixture probabilities based on expert's assessments. The blue vertical line represents the expert's assessment of the median and the red horizontal lines represent the expert's assessment for the lower and upper quartiles.

Since it is easier for an expert to think in terms of univariate distributions, we often want to provide feedback about their assessments in terms of the marginal distribution, if possible. Even though the assessed probability quartiles are based on conditional assessments, with the exception of the most likely category (here p_0), it is easier for the expert to think in terms of the univariate distributions.

Table 4.5: Marginal (unconditional) medians and quartiles implied by the expert's conditional assessments.

Parameter	25th Percentile	Median	75th Percentile
p_0	0.48	0.55	0.63
p_3	0.14	0.19	0.26
p_1	0.09	0.13	0.19
p_2	0.05	0.09	0.14

Thus, if possible, we want to show the expert the marginal quartiles for each p_j based on their conditional assessments. Accordingly, it is the marginal quartiles that

are presented to the expert for feedback on whether or not the resulting distributions accurately reflect their beliefs. If the expert feels that the suggested marginal percentiles do not accurately reflect their belief, modifications can be made until they are satisfactory. For our purposes, we assume the expert is satisfied with the marginal quartile values provided in Table 4.5. The resulting elicited Dirichlet prior distribution is

$$\mathbf{p} \sim \text{Dirichlet}(h_1 = 10.61, h_2 = 2.79, h_3 = 2.67, h_4 = 1.90).$$

This is equivalent to a prior ESS of about $n_e = 19$ observations. In addition, based on the marginal and conditional quartiles in Table 4.5, we can compute the prior ESS for the individual mixtures. The estimated prior ESS from the marginal distribution for p_0 is about $n_e = 22$. Similarly, the estimated prior ESS from the conditional distribution for $p_3|p_0 = 0.55$ yields $n_e = 17$ and the conditional distribution of $p_1|p_0 = 0.55, p_3 = 0.20$ yields $n_e = 17$.

4.6.3 Remarks about Choice of Dirichlet as Prior Distribution

The standard Dirichlet distribution is the most widely used prior for multinomial models due to its tractability, simplicity and conjugacy. Nevertheless, there are several limitations of this distribution that have been well-documented in the literature. For example, the standard Dirichlet distribution has a limited number of parameters (i.e. a k-variate Dirichlet is specified by k parameters) and thus lacks flexibility to represent prior belief (Aitchison (1986) and O'Hagan and Forster (2005)). In addition, the Dirichlet distibution imposes a negative correlation between the mixture probabilties, which might not be appropriate or accurately represent prior belief. Lastly, Dirichlet variates that have the same mean necessarily have equal variances (O'Hagan and Forster (2005)).

Elfadaly and Garthwaite (2013a) propose the Connor-Mosimann distribution (one generalization of the Dirichlet distribution) as a more flexible option for a prior

to quantify expert opinion on proportions. The Connor-Mosimann distribution requires a larger number of parameters than the standard Dirichlet distribution. In particular, it has $2(j-1)$ hyperparameters to represent opinion about j parameters compared to the Dirichlet which has just j hyperparameters to represent j parameters. Moreover, the Connor-Mosimann distribution has desirable properties such as conjugacy and reasonable tractability, and the $2(j-1)$ hyperparameters can be determined from the same assessments as the Dirichlet distribution. The Connor-Mosimann distribution has a more flexible dependence structure. This more general covariance structure allows for positive correlation among mixture probabilities with the exception that the first mixture component (or most likely category), is always negatively correlated with the other mixture probabilities. We do not consider the Connor-Mosimann distribution here, but this is a topic for future research as in some cases it more accurately reflects prior belief of an expert.

4.6.4 Conditional Means Prior Approach for Poisson Parameters

For our hypothetical healthcare adverse event example, we assume that the Poisson parameters depend on covariates. Thus, for the Poisson parameters, we use the conditional means prior approach discussed in Chapter Three to elicit a prior on the Poisson parameters, μ_1 and λ_2, and induce priors on λ_1 and λ_0 by eliciting information about the conditional probability, θ. Thus, we combine a conditional means prior approach for the Poisson parameters with the elicitation of a Dirichlet distribution for the zero-inflation parameters as outlined above. It can be shown that we can represent the BZIP model as a product of the conditional and marginal distributions as in Chapter Three. As in Chapter Three, we use the conditional representation of the BZIP joint distribution given by (3.3). The conditional representation of the BZIP model in (4.1), $f_{\text{CBZIP}}(y_1, y_2 \,|\, \mathbf{p}, \theta, \mu_1, \lambda_2) \equiv f_{\text{CBZIP}}(y_1, y_2)$, is

given by

$$f_{\text{CBZIP}}(y_1, y_2) = \begin{cases} p_0 + (1-p_0) f_{\text{CBP}}(y_1, y_2 \,|\, \theta, \mu_1, \lambda_2), & y_1 = 0, y_2 = 0, \\ (p_1 + p_3) f_{\text{CBP}}(y, y_2 \,|\, \theta, \mu_1, \lambda_2), & y_1 \neq 0, y_2 = 0, \\ (p_2 + p_3) f_{\text{CBP}}(y, y_2 \,|\, \theta, \mu_1, \lambda_2), & y_1 = 0, y_2 \neq 0, \\ (1 - p_0 - p_1 - p_2) f_{\text{CBP}}(y_1, y_2 \,|\, \theta, \mu_1, \lambda_2), & y_1 \neq 0, y_2 \neq 0, \end{cases}$$

where $f_{\text{CBP}}(y_1, y_2 \,|\, \theta, \mu_1, \lambda_2)$ is as defined in (3.8). As before, we can represent the conditional BZIP representation in such a way that the parameters depend on covariates in a generalized linear model. Within the context of the adverse event drug safety study, for the ith subject in treatment group x_i, we have that

$$\log(\mu_{1,i}) = \phi_0 + \phi_1 x_i,$$
$$\log(\lambda_{2,i}) = \gamma_{2,0} + \gamma_{2,1} x_i,$$

and

$$\text{logit}(\theta_i) = \alpha_0 + \alpha_1 x_i,$$

where x_i is as defined in 4.18. As in Chapter Three, it follows that,

$$\lambda_{0,i} = \theta_i \times \mu_{1,i},$$
$$\lambda_{1,i} = \mu_{1,i} - \lambda_{0,i},$$

and

$$\mu_{2,i} = \lambda_{2,i} + \lambda_{0,i}.$$

Table 4.6 provides the model parameter summary for subject i and adverse event $j = 1, 2$.

Table 4.6: Model parameter summary for the ith subject.

Parameter	Interpretation
$\mu_{1,i}$	The mean rate of experiencing migraines (among those at risk to experience migraines) for a subject receiving treatment x_i.
$\mu_{2,i}$	The mean rate of experiencing a seizure for a subject receiving treatment x_i.
θ_i	The conditional probability that among subjects at-risk for both adverse events, given subject i has a migraine, subject i also has a seizure.
α_0	The log odds that an at-risk subject receiving the current standard of care has a migraine and a seizure (simultaneously experiences both adverse events).
α_1	The log odds that an at-risk subject receiving treatment x_i has a migraine and a seizure; e^{α_1} represents the increased (or decreased) odds that an at-risk subject receiving the new treatment has a migraine and has a seizure (compared to the current standard of care).
ϕ_1	Average change in log adverse event rate for experiencing migraines (among those at-risk to experience migraines) for a subject receiving the new treatment compared to current standard of care; For a subject receiving the new treatment, the expected number of migraines increases (decreases) by a factor of e^{ϕ_1} compared to current standard of care.
$\gamma_{2,0}$	Average log adverse event rate for exclusive seizures for a subject receiving the current standard of care.
$\gamma_{2,1}$	Average change in log adverse event rate for experiencing exclusive seizures for a subject receiving new treatment compared to standard of care; For a subject receiving the new treatment, the expected number of exclusive seizures increases (decreases) by a factor of $e^{\gamma_{2,1}}$ compared to current standard of care.

We refer the reader to Table 4.3 for interpretations of $\lambda_{0,i}$, $\lambda_{1,i}$, and $\lambda_{2,i}$. To begin, we elicit information about μ_1 by assessing the expert's judgment about the

number of migraines experienced by a subject. For both treatments, $x_i = 0$ and $x_i = 1$, the expert is asked a series of questions:

(1) "Consider subjects that are receiving the new treatment, what, do you think, is the most likely value for the number of migraines?" Their response is taken to be the mode.

(2) Next, we prompt the expert to suggest an upper bound to represent their uncertainty. The expert is asked "What, do you think, is the largest the number of migraines can be among subjects with the new treatment." Their response taken to be the 80th percentile.

Similarly, we prompt the expert about their expectations for the number of seizures for subjects that are in the new treatment group. The information collected from the expert is summarized in Table 4.7.

Next, we elicit information about the number of (exclusive) seizures. To do this the expert is asked:

(1) "Consider subjects that are in the treatment group, what do you think, is the most likely value for the number of just seizures?" Their response is taken to be the mode.

(2) Next, we prompt the expert to suggest an upper bound to represent their uncertainty. The expert is asked "What, do you think, is the largest the number of just seizures can be among subjects with the new treatment." Their response is taken to be the 80th percentile.

Similarly, we ask the expert about their expectations for subjects that are taking the current standard of care. Suppose the information collected from the expert is that summarized in Table 4.7.

Table 4.7: Expert elicited information on Poisson parameters.

Parameter	Mode	80th percentile
$\mu_{1,0}$	6	12
$\mu_{1,1}$	4	10
$\lambda_{2,0}$	5	10
$\lambda_{2,1}$	4	8

This information obtained in Table 4.7 can be translated into parameters of a gamma distribution. In particular,

$$\mu_{1,0} \equiv \exp\left(\phi_0 + \phi_1(0)\right) \sim \text{Gamma}(3.37, 2.53),$$

$$\mu_{1,1} \equiv \exp\left(\phi_0 + \phi_1(1)\right) \sim \text{Gamma}(2.41, 2.83),$$

and

$$\lambda_{2,0} \equiv \exp\left(\gamma_{2,0} + \gamma_{2,1}(0)\right) \sim \text{Gamma}(3.37, 2.11),$$

$$\lambda_{2,1} \equiv \exp\left(\gamma_{2,0} + \gamma_{2,1}(1)\right) \sim \text{Gamma}(3.37, 1.69),$$

These resulting gamma densities are shown in Figure 4.10. Using the CMP approach as in Section 3.4 we obtain the induced priors for $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}_2$. These induced priors have no closed form but can easily be simulated as shown in Figure 4.11.

Next, we consider elicitation of a conditional means prior for the logistic regression parameters used to model the conditional probability, θ. In the context of the drug safety study, θ represents the conditional probability that given an at-risk subject has a migraine, the subject also has migraine.[7] To elicit information the conditional probability the expert is asked to assume that the subject is at-risk for both adverse events. For $x_i = 0$ and $x_i = 1$:

(1) To assess the median, the expert is asked"Suppose we have 100 at-risk subjects from the study population that in treatment group, x_i and have a

[7] In this context,"at-risk" subjects refer to subjects at-risk for both adverse events.

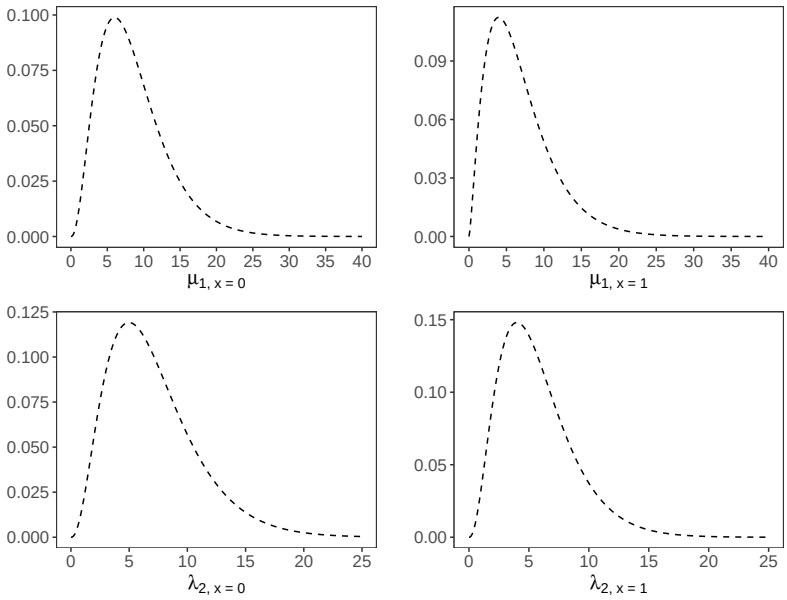

Figure 4.10: Priors based on information collected from expert about μ_1 and λ_2 for current standard of care and new treatment.

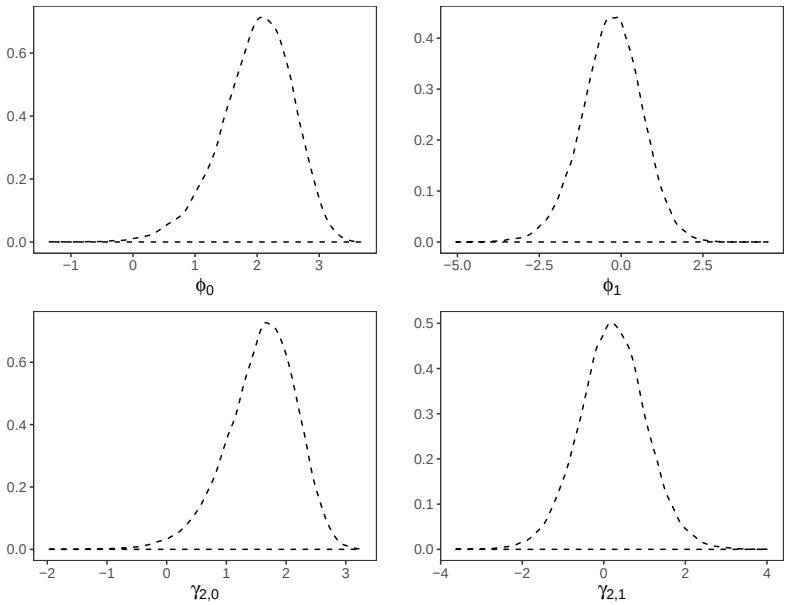

Figure 4.11: Simulated density plots for the induced priors on the regression coefficients corresponding to μ_1 and λ_2.

migraine. What, do you think, is the percentage of these subjects that also have a seizure" Their response is taken to be the median.

(2) Next, to assess the lower quartile for the percentage of subjects that have a seizure given the subject has reported a migraine, the expert is asked "Suppose the percentage of subjects that are receiving the treatment x_i that have a seizure given they have a migraine is *less than* your initial assessment. Given this information, what would you now estimate as the percentage?" Their response is taken to be the 25th percentile.

(3) Finally, to obtain an upper quartile, the expert is asked "Suppose the percentage of subjects with intervention x_i that have a seizure given they have a migraine is actually *greater than* your initial assessment. Given this information, what would you now estimate as the percentage?" Their response is taken to be the 75th percentile.

Suppose the information collected from the expert is that summarized in Table 4.8.

Table 4.8: Expert elicited information on conditional probability.

Parameter	25th Percentile	Median	75th Percentile
θ_0	0.22	0.30	0.38
θ_1	0.22	0.30	0.38

This information obtained about θ in Table 4.8 can be translated into parameters of a beta distribution using numerical methods. In particular, the resulting priors for standard of care, $x_i = 0$, and the new treatment,$x_i = 1$, are

$$\theta_0 \equiv \text{logit}^{-1}\left(\alpha_0 + \alpha_1(0)\right) \sim \text{Beta}(4.6, 10.4)$$

and

$$\theta_1 \equiv \text{logit}^{-1}\left(\alpha_0 + \alpha_1(1)\right) \sim \text{Beta}(4.6, 10.4).$$

The beta distributions that result from the information collected about the expert's expectations are shown in Figure 4.12.

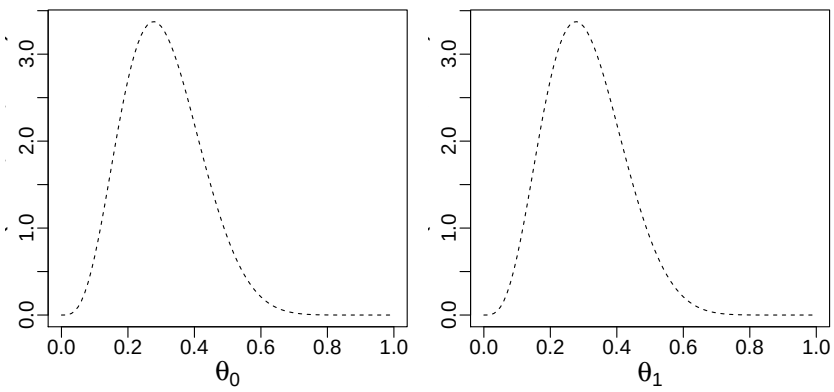

Figure 4.12: Priors based on information collected from expert about conditional probability that given a subject has a migraine, they also have a seizure.

Using the CMP approach as in Section 3.4 we obtain the induced priors for α_0 and α_1. These induced priors have no closed form but can easily be simulated as shown in Figure 4.13.

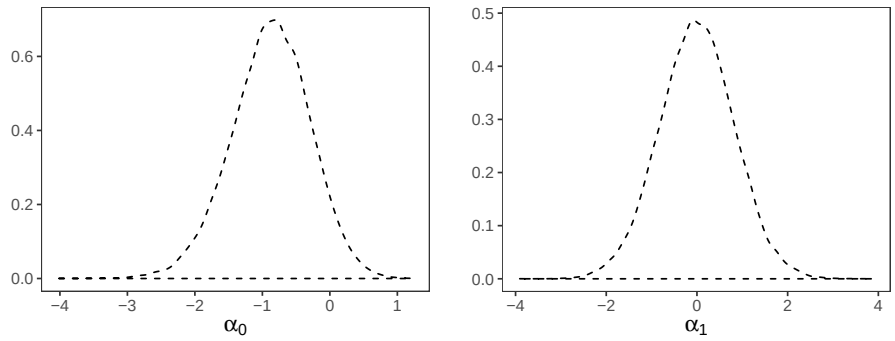

Figure 4.13: Simulated density plots for the induced priors on the regression coefficients corresponding to θ.

Finally, the induced priors on λ_0, λ_1 and μ_2 for the current standard of care and the new treatment are shown in Figure 4.14.

4.6.5 Using Prior Predictive Distribution as Tool to Provide Feedback to Expert

Here we provide feedback to the expert regarding the implications of their assessments for the mixture probabilities in Section 4.6.2 and their assessments about the Poisson parameters. Providing feedback to the expert about the implications of

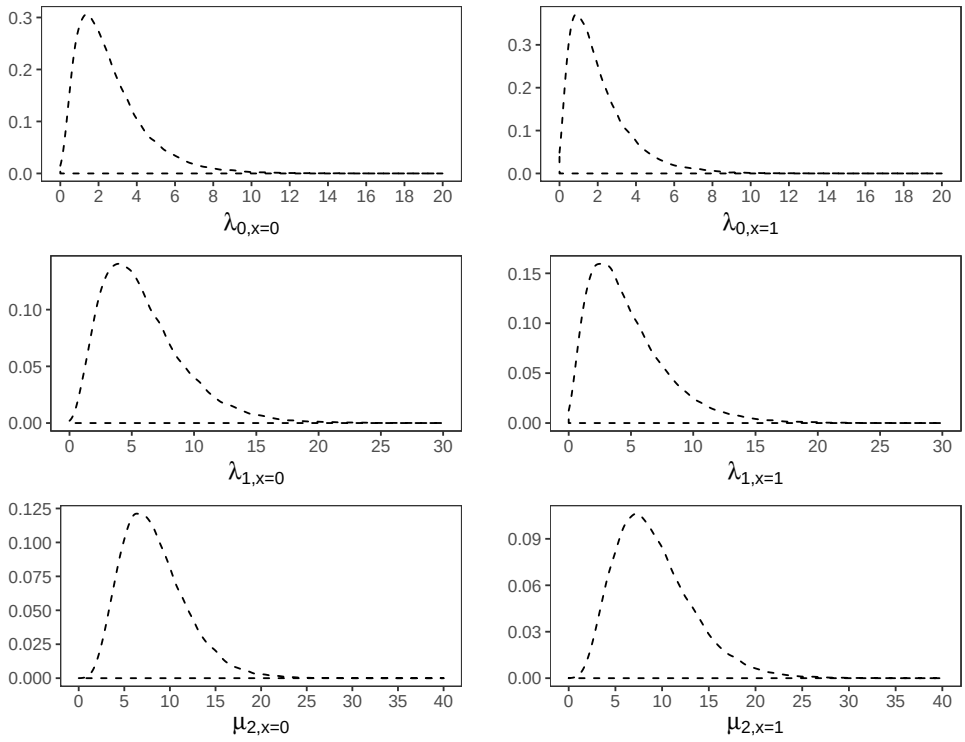

Figure 4.14: Simulated density plots for the induced priors on the λ_0, λ_1 and μ_2 for the current standard of care (left densities) and the new treatment (right densities) based on information collected from the expert.

their prior assessments is an essential part of the prior elicitation process. This allows the expert to confirm whether or not the resulting prior distributions accurately reflect their prior beliefs and if necessary allows the opportunity for modifications of the prior structure. As in Section 2.7.1, we can use the prior predictive distribution as a tool to provide the feedback in a scale that is meaningful to the expert. Specifically, we can use the prior predictive distribution to generate hypothetical data that might result given their prior assessments. This can easily be implemented in WinBUGS. We might, for example, generate data from the prior predictive distribution corresponding to the current standard of care ($x_i = 0$) and the new treatment ($x_i = 1$), for a sample size of $n = 100$. Table 4.9 summarizes possible data that might result.

Table 4.9: Summary of hypothetical dataset generated from the prior predictive distribution for current standard of care ($x = 0$) and for new treatment ($x = 1$) based on the expert's prior assessments.

Variable	Trt. Group	Mean	SD	2.5%	97.5%
Y_1	Standard	3.05	5.59	0	17
Y_1	New	1.48	3.62	0	12
Y_2	Standard	2.37	4.56	0	15
Y_2	New	1.27	2.88	0	9

Figure 4.15 provides a scatterplot and histogram for plausible data that could result based on the expert's prior judgment.

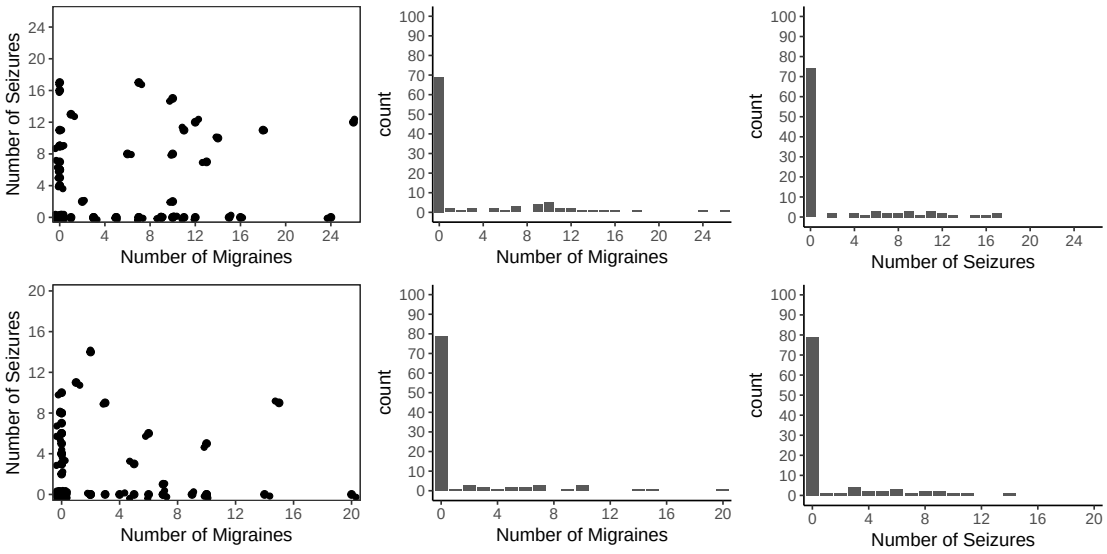

Figure 4.15: Simulated data based on prior predictive distribution for number of migraines and number of seizures for $n = 100$ subjects taking the current standard of care (top three plots) and $n = 100$ subjects taking the new drug (bottom three plots) based on information collected from expert.

Both scatterplots in Figure 4.15 suggest positive association between the number of migraines and the number of seizures. Moreover, the histograms for the number of migraines and the number of seizures indicate that the expert's assessments suggest a slight decrease in the number of these adverse events events for the new treatment compared to the current standard of care. These hypothetical data allows the expert to evaluate and modify their prior assessments in a meaningful scale. In practice we would show the expert several realizations of these prior predictive distributions.

4.6.6 Posterior Inference for Adverse Event Study

We assume that based on the feedback provided to the expert about the implications of their prior structure, they are satisfied that the elicited priors accurately reflect their beliefs. Posterior inference was carried out in WinBUGS using the same specifications as outlined in Section 4.3.1. Standard diagnostics based on trace plots and the Gelman-Rubin statistic indicated no problems with convergence. Table 4.11 provides the posterior results for the BZIP model parameters based on the expert's assessments.

Table 4.10: Posterior inference for regression coefficients.

Parameter	Mean	SD	50%	2.5%	97.5%	Width
$\gamma_{2,0}$	1.578	0.281	1.621	0.907	2.003	1.096
$\gamma_{2,1}$	-0.839	0.365	-0.855	-1.523	-0.075	1.448
ϕ_0	2.336	0.070	2.337	2.195	2.471	0.276
ϕ_1	-1.449	0.188	-1.444	-1.831	-1.097	0.734

Posterior densities for the regression coefficients are shown in Figure 4.17. Posterior inference for ϕ_1 suggests that the number of migraines experienced by subjects increases by a factor between 3.4 and 6.1 for those taking the current standard of care compared to those subjects receiving the new treatment. Posterior inference for $\gamma_{2,1}$ suggests that the number of seizures increases by a factor between

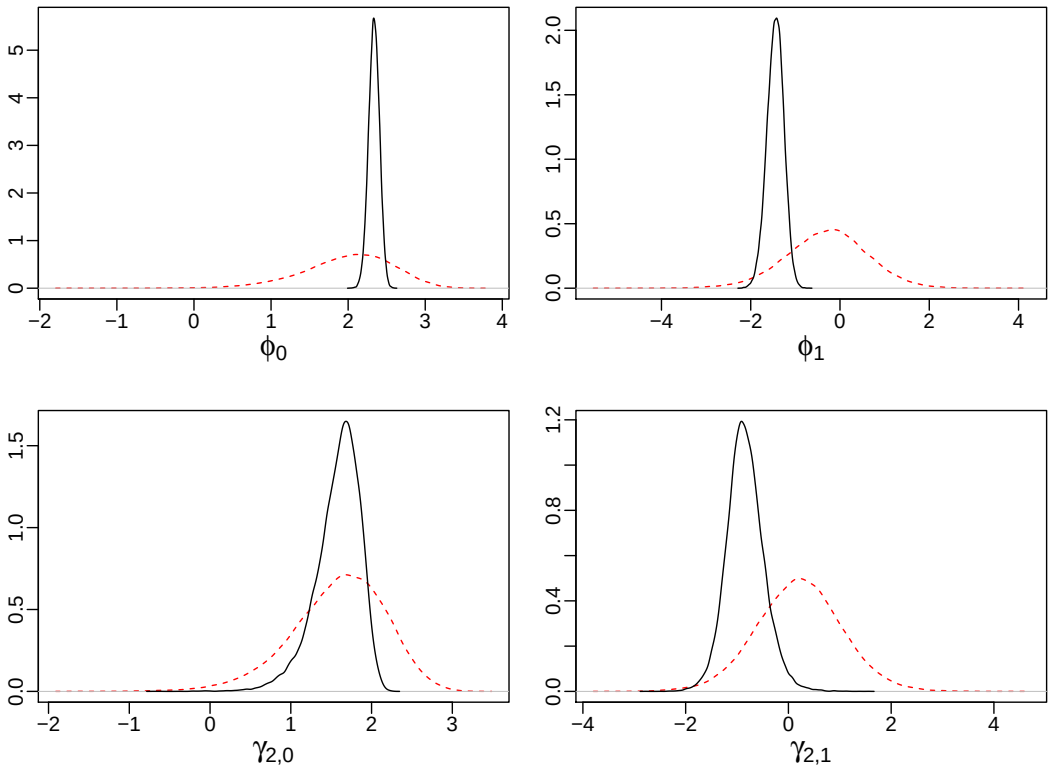

Figure 4.16: Prior (red, dashed) and posteriors (black, solid) densities for BZIP regression coefficients based on expert's assessments.

1.05 and 4.88 for subjects receiving the current standard of care compared to the treatment group.

Table 4.11 provides the posterior results for the BZIP model parameters based on the expert's assessments. The posterior densities in Figure 4.17 and the posterior results in Table 4.11 for $\lambda_{0,0}$ and $\lambda_{0,1}$ suggest that the simultaneous rate of occurrence of migraines and seizures is less for the new treatment compared to the current standard of care. Similarly, the posterior densities for λ_1 and λ_2 suggest that the rate of exclusive migraines and the rate of exclusive seizures is less for the new treatment compared to the current standard of care.

We now evaluate whether the new treatment is superior to the current standard of care with respect to safety using the criteria (4.19), (4.20) and (4.21). That is, given the data from the current study, we require that the posterior predictive

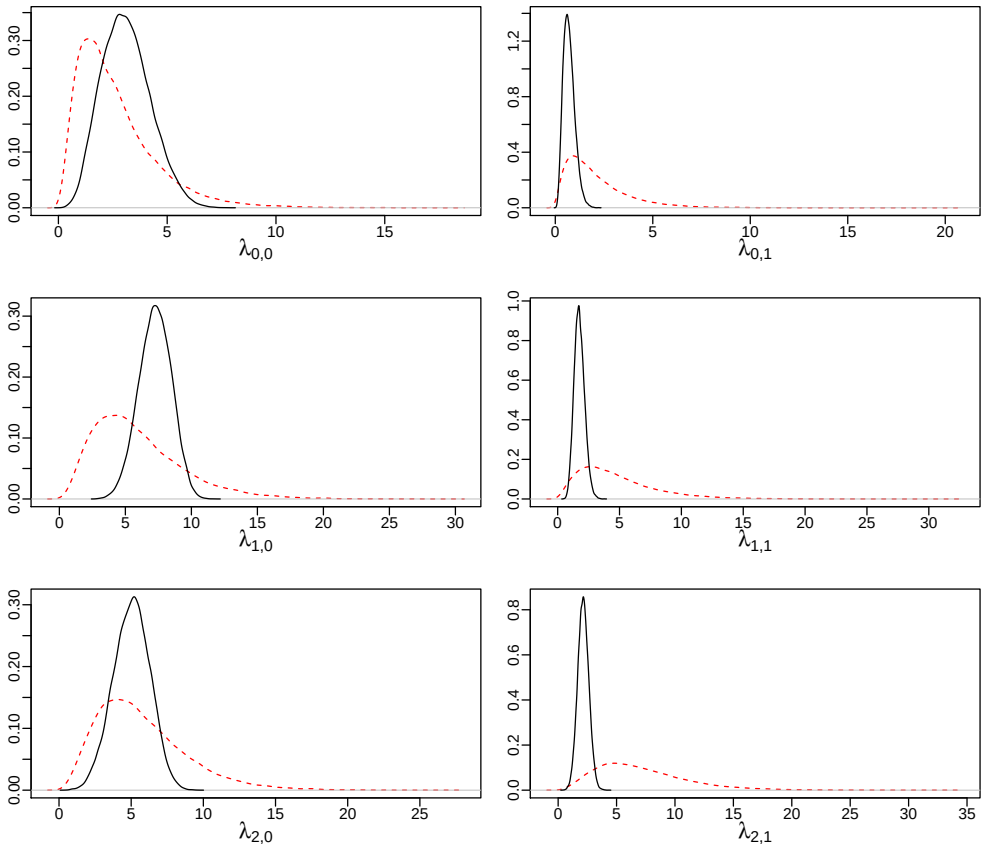

Figure 4.17: Prior (red, dashed) and posteriors (black, solid) densities for BZIP parameters based on expert's assessments.

Table 4.11: Posterior inference for Poisson parameters and mixture probabilities for both treatment groups in drug safety study.

Parameter	Mean	SD	50%	2.5%	97.5%	Width
$\lambda_{0,0}$	3.116	1.111	3.050	1.181	5.460	4.279
$\lambda_{0,1}$	0.707	0.293	0.674	0.238	1.366	1.128
$\lambda_{1,0}$	7.248	1.229	7.272	4.778	9.572	4.794
$\lambda_{1,1}$	1.758	0.419	1.732	1.004	2.642	1.638
$\lambda_{2,0}$	5.024	1.265	5.057	2.477	7.412	4.935
$\lambda_{2,1}$	2.150	0.471	2.145	1.239	3.090	1.851
p_0	0.530	0.047	0.530	0.438	0.621	0.182
p_1	0.136	0.033	0.134	0.078	0.206	0.128
p_2	0.127	0.032	0.125	0.069	0.195	0.126
p_3	0.207	0.039	0.205	0.136	0.287	0.151

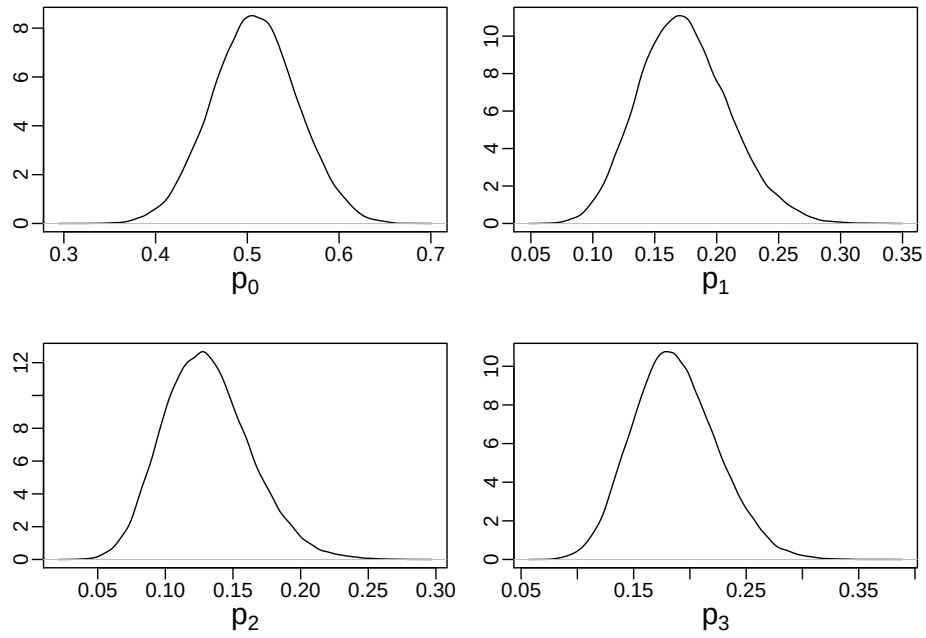

Figure 4.18: Posterior densities for mixture probabilities.

probability that, among subjects that experienced both adverse events, the number of migraines, the number of seizures, and the sum of both events, is less for the new treatment compared to the current standard of care. In particular, using the posterior predictive distribution, we have that

$$\Pr(Y_{1,1} < Y_{1,0} \,|\, \text{data}) = 0.98,$$

$$\Pr(Y_{2,1} < Y_{2,0} \,|\, \text{data}) = 0.92,$$

and

$$\Pr(Y_{1,1} + Y_{2,1} < Y_{1,0} + Y_{2,0} \,|\, \text{data}) = 0.98,$$

where, again, $Y_{1,0}$ and $Y_{2,0}$ denote the number of migraines and the number of seizures, respectively, for the current standard of care, and $Y_{1,1}$ and $Y_{2,1}$ denote the number of migraines and the number of seizures, respectively, for the new treatment. Hence, given the data from the current study, it appears that the new treatment is superior to the current standard of care in reducing the number of adverse events, by this criteria. The FDA guidance on Bayesian methods requires that prior probabili-

ties of such success criteria be considerably less than the requisite success probability, in this case δ. These prior predictive probability of success is

$$\Pr(Y_{1,1} < Y_{1,0} \,|\, \text{prior}) = 0.55,$$
$$\Pr(Y_{2,1} < Y_{2,0} \,|\, \text{prior}) = 0.57,$$

and

$$\Pr(Y_{1,1} + Y_{2,1} < Y_{1,0} + Y_{2,0} \,|\, \text{prior}) = 0.60,$$

which are indeed much less than $\delta = 0.90$. The prior structure does not unreasonably favor the desired criteria. Furthermore, Figure 4.19 shows the prior predictive distribution and the posterior predictive distribution for the number of migraines, the number of seizures, and the number of both migraines and seizures among subjects that experienced both adverse events. The top box plots reflect these distributions with respect to the expert's prior assessments. The bottom box plots reflect these distributions given the data in Figure 4.6.

The top three prior predictive box plots in Figure 4.19 reflect that the expert believes that the number of adverse events experienced by subjects receiving the new treatment will be less compared to subjects receiving the current standard of care. Nevertheless, the overlap in the prior predictive distributions for the new treatment and the current standard of care further indicates that the prior structure does not unreasonably favor the new treatment. The posterior predictive distributions in Figure 4.19 suggest that among those that experience both adverse events, the number of migraines, the number of seizures, and the number of both adverse events is less for subjects receiving the new treatment compared to those receiving the current standard of care.

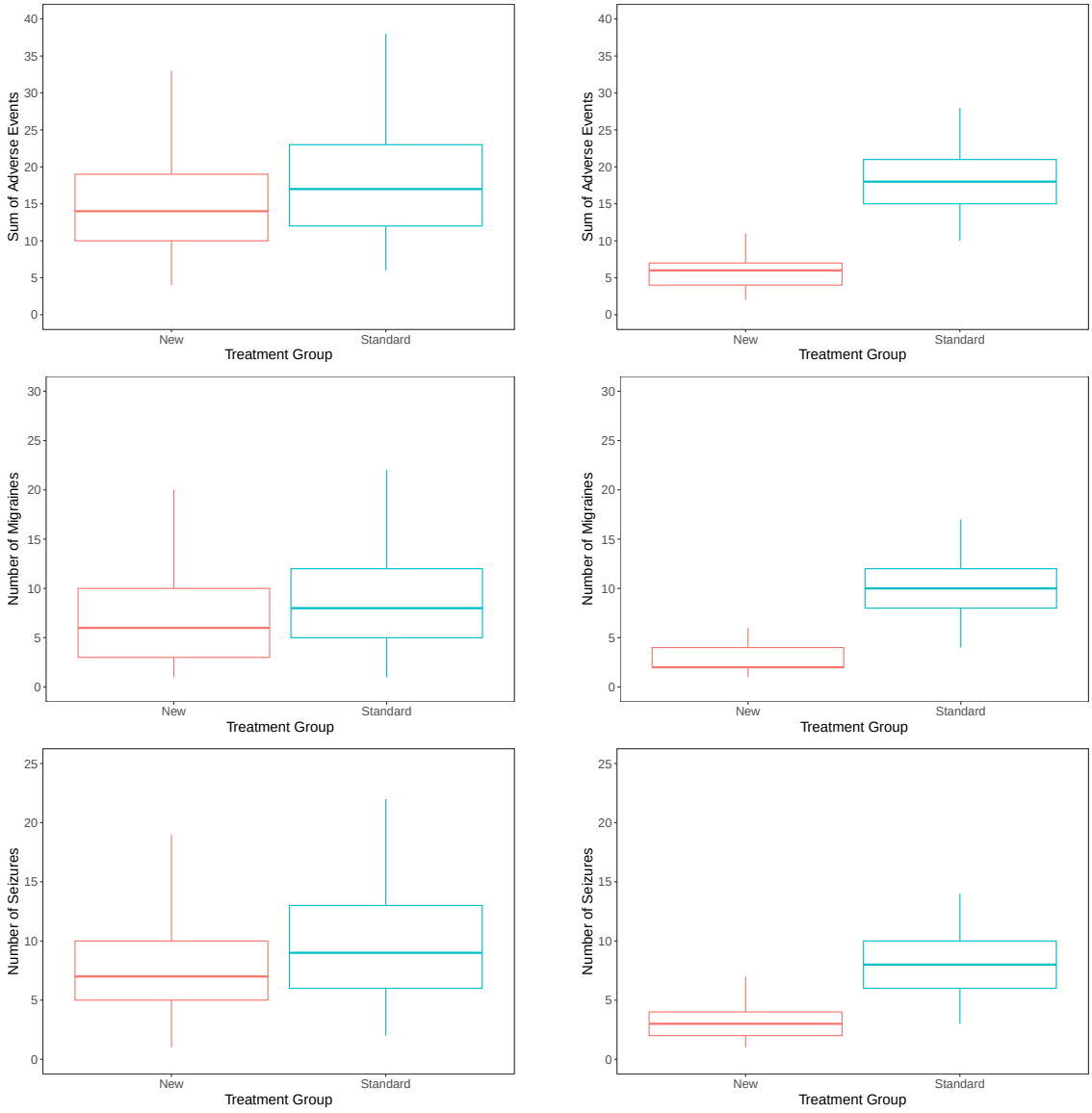

Figure 4.19: Prior predictive (left) and posterior predictive distributions (right) for the number of adverse events among those that have experienced both adverse events for the new treatment (red) and current standard of care (blue). The right boxplots reflect the distribution of the number of both adverse events (top), number of migraines (middle), and number of seizures (bottom) among those that experienced both adverse events, given the expert's prior assessments. The right boxplots reflect the distribution of the number of both adverse events (top), number of migraines (middle), and number of seizures (bottom) among those that experienced both adverse events, given the prior, data and the posterior.

4.6.7 Posterior Predictive as Tool for Clinician

In Section 3.4.5, we used the prior predictive distribution as a tool for a clinician to assess what a future patient might expect given the current data. Similarly, in this section, we make use of the posterior predictive distribution as a tool for a clinician to assess what a future patient taking treatment x_i might expect to experience given the current data. Figure 4.20 represents the posterior predictive distribution for the number of migraines and/or the number of seizures for an at-risk subject for the four subpopulations in (4.1).

Figure 4.20 could be used as a tool by a clinician to assess what a patient receiving treatment x_i, that is at-risk for either just migraines (top right boxplots), just seizures (second row, right boxplots), or both (bottom two right box plots) might expect to experience. For example, based on the conditional posterior predictive distribution that given a patient is at-risk for both adverse events, the clinician can inform a patient receiving the current standard of care that they might expect to experience about 8 (and no more than 14 seizures) over a 6-month period. Similarly, the clinician can inform a patient that is at-risk for both events receiving the new treatment that they might expect to experience about 3 seizures (and no more than 7).

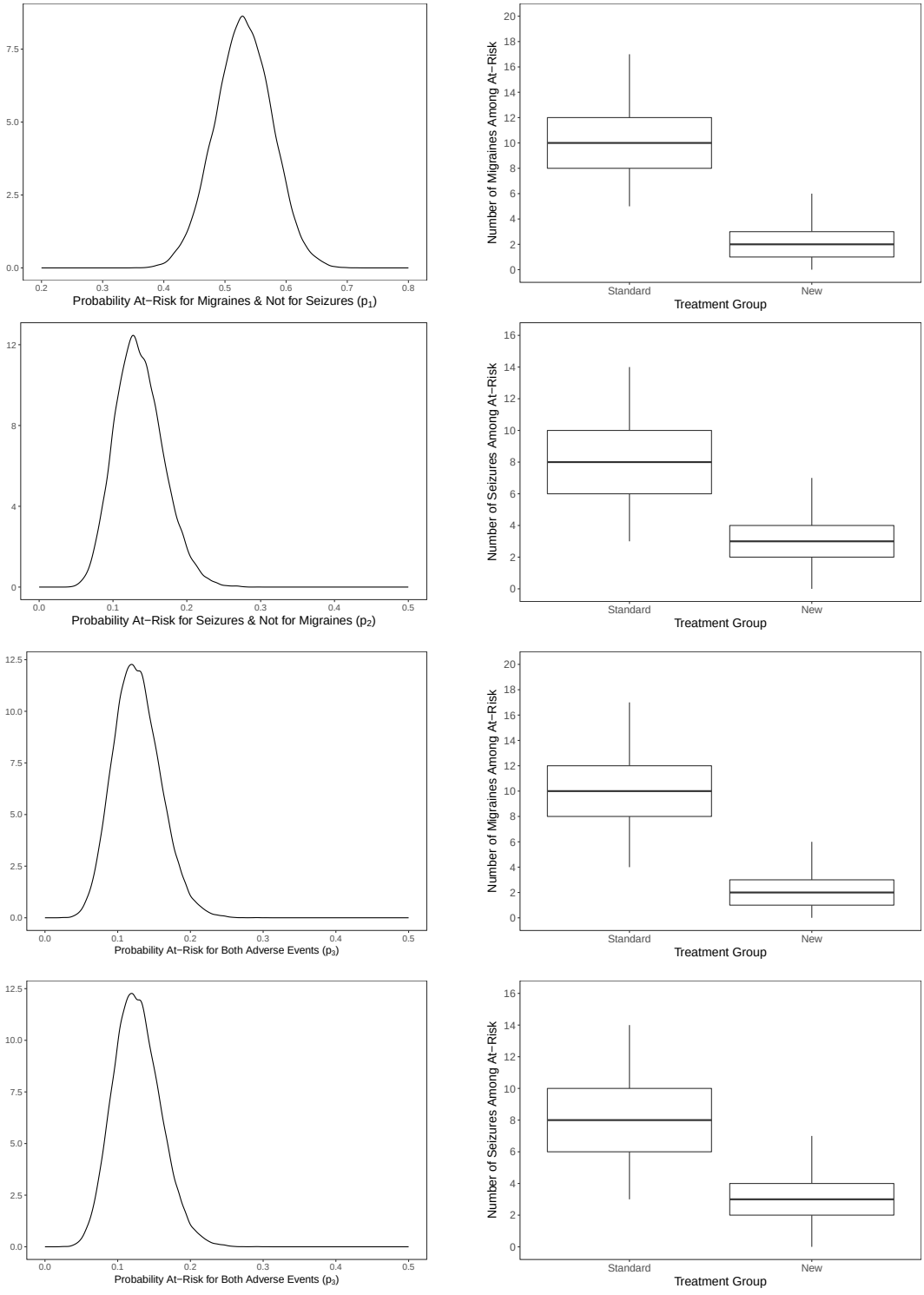

Figure 4.20: Conditional posterior predictive distribution for number of adverse events (represented by the boxplots) given a subject is at-risk and the corresponding marginal posterior density for the probability that a subject is at-risk.

4.7 Summary

We have discussed a Bayesian approach to a bivariate zero-inflated Poisson regression model. Despite the increased prevalence of Bayesian inference for zero-inflated models in the literature, informative priors for these models have not yet been explored in detail. We offer methods for prior construction and show how relatively informative priors can mitigate the effects nonidentifiability can cause. We have demonstrated how in nonidentifiable settings the prior distribution can be used as a tool to identify parts of the parameter space that are not covered by the likelihood, namely for the unobserved Poisson parameters λ_0, λ_1 and λ_2. We also demonstrated that use of such informative priors alleviates convergence issues often present when using a diffuse prior structure for model parameters. We proposed and described methods of prior elicitation for a BPZIP model within the context of a clinical example. We offer methods for prior elicitation of a multivariate distribution, namely a Dirichlet distribution for the zero-inflation parameters. Finally, we extend the methods of informative prior construction via the CMP approach developed in Chapters Two and Three to the BZIP model. We demonstrate the use of the prior predictive distribution as tool by which to provide feedback to the expert to illustrate the the implications of their prior structure.

APPENDICES

APPENDIX A

The Bivariate Poisson Distribution

Let (Y_1, Y_2) denote a bivariate Poisson (BP) such that $(Y_1, Y_2) \sim \text{BP}(\lambda_0, \lambda_1, \lambda_2)$. Several representations of a BP distribution have been proposed in the literature (Kocherlakota and Kocherlakota (1992)). For the BP distribution in our BPZIP model in Chapters Two and Three and the BZIP model in Chapter Four, we use the BP representation derived via the trivariate reduction method (Johnson et al. (1997)). This is the representation most commonly adopted in the literature. This representation is based on the joint distribution the sums of independent, latent random variables. In particular, consider the random variables joint random variables Y_1 and Y_2. Using the trivariate reduction representation we have that

$$Y_1 = X_1 + X_0, \quad \text{and} \quad Y_2 = X_2 + X_0, \tag{A.1}$$

where X_1, X_2, and X_0 are independent Poisson random variables with means λ_1, λ_2, and λ_0, respectively, such that $\lambda_i > 0$, $i = 0, 1, 2$. This representation of the BP can be used to model positively correlated count data. Alternate parameterizations of the BP distribution make use of convolutions and conditional distributions. We refer the reader to Johnson et al. (1997) for further detail. Further, extensions based on mixtures have been proposed to allow for a more flexible correlation structure and overdispersed marginal distributions, however these models are not without limitations (Nikoloulopoulos and Karlis (2009)).

Let $(Y_1, Y_2) \sim \text{BP}(\lambda_0, \lambda_1, \lambda_2)$. The joint probability mass function is given by

$$\begin{aligned} f_{BP}(y_1, y_2) &= \Pr(Y_1 = y_1, Y_2 = y_2) \\ &= e^{-(\lambda_1+\lambda_2+\lambda_0)} \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_2^{y_2}}{y_2!} \sum_{m=0}^{\min(y_1,y_2)} \binom{y_1}{m} \binom{y_2}{m} m! \left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^m, \end{aligned} \tag{A.2}$$

where $\lambda_0, \lambda_1, \lambda_2 > 0$, and $y_1, y_2 \in \{0, 1, 2, 3, \ldots\}$. It follows that the marginal distribution of Y_1 and Y_2 are Poisson with mean $\lambda_1 + \lambda_0$ and $\lambda_2 + \lambda_0$, respectively. Since the sum of independent Poisson random variables is also a Poisson random variable, it follows that the marginal distribution of Y_1 is Poisson with rate $\lambda_1 + \lambda_0$. Similarly, the marginal distribution of Y_2 is Poisson with rate $\lambda_2 + \lambda_0$.

A.1 Inference for the BP Distribution

Let $y_i = (y_{1i}, y_{2i})$, $i = 1, \ldots, n$ denote the observed bivariate outcomes. The corresponding likelihood function is

$$L(\boldsymbol{\lambda} \,|\, \mathbf{y}) = \prod_{i=1}^{n} e^{-(\lambda_1+\lambda_2+\lambda_0)} \frac{\lambda_1^{y_{1i}}}{y_{1i}!} \frac{\lambda_2^{y_{2i}}}{y_{2i}!} \sum_{m=0}^{\min(y_{1i}, y_{2i})} \binom{y_{1i}}{m} \binom{y_{2i}}{m} m! \left(\frac{\lambda_0}{\lambda_1 \lambda_2}\right)^m, \qquad \text{(A.3)}$$

where $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$. Inference for the BP model is not an easy task, due to the complicated and intractable form of the likelihood, which involves the product of n summations, as shown in expression (A.3). Historically, the complicated form of the likelihood has been an obstacle for both Frequentist inference and Bayesian inference for the BP model (Karlis and Tsiamyrtzis (2008)).

The trivariate reduction representation of the BP in (**??** allows for estimation of parameters, $\boldsymbol{\lambda}$, via the EM algorithm in the frequentist paradigm. However, convergence of the MLEs obtained via the EM algorithm can be sensitive to initial values.

A.2 Potential Limitations of Trivariate Reduction Representation

Although the random variables Y_1 and Y_2 are generated from sums of additive variables, Y_1 and Y_2 are *not* independent. The covariance between Y_1 and Y_2 is

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(X_1 + X_0, X_2 + X_0) = \mathrm{Var}(X_0) = \lambda_0$$

and the correlation coefficient between Y_1 and Y_2 is given by

$$\rho_{\mathrm{BP}} = \mathrm{Corr}(Y_1, Y_2) = \frac{\lambda_0}{\sqrt{(\lambda_1 + \lambda_0)(\lambda_2 + \lambda_0)}}.$$

Holgate (1964) showed that the correlation coefficient cannot exceed the square root of the ratio of the smaller to the larger of the means of the two marginal distributions. This is a potential limitiation of this representation of the BP distribution. Moreover, this representation assumes $\lambda_0 > 0$. Hence this representation only allows for positive correlation between Y_1 and Y_2. We refer the reader to Johnson et al. (1997) for a representation of the BP distribution that allows for negative correlation between Y_1 and Y_2.

Finally, Berkhout and Plug (2004) proposes a general bivariate count model using conditional probabilities that can be applied to Poisson counts and can be used to estimate two correlated count data processes, allowing for negative as well as positive correlation. The model is referred to as conditional Poisson model (CPM). Nevertheless, the model proposed by Johnson et al. (1997) that allows for positive correlation is the most prevalent in the literature and the representation that we adopt in this dissertation. Accordingly, we consider applications where the assumption of positive correlation is appropriate.

APPENDIX B

Data Generation of Bivariate Partial Zero-Inflated Poisson Model

The process used to generate data for the BPZIP model introduced in Chapter Two is described here. We discuss the data generation within the context of the adverse event drug safety study in Section2.3. Recall that a BPZIP model can be constructed as follows:

$$(Y_1, Y_2) \sim \begin{cases} (0, \text{Poisson}(\mu_2)), & \text{with probability } p \\ \text{BP}(\lambda_0, \lambda_1, \lambda_2), & \text{with probability } 1-p, \end{cases} \tag{B.1}$$

We begin by discussing how the zero-inflation parameter, p is generated. In the context of the adverse event drug safety study, an observation is not at-risk for Y_1 with probability p and is at-risk for Y_1 with probability $1-p$. The `rmultinom` function in R generates a specified number of random values from a specified multinomial distribution. In this case, we can use the `rmultinom` function with specified parameters `n` (sample size), `size` (in this case 1) and `prob` (in this case a vector of length 2 with values for p and $1-p$. Specifically, we generate the zero-inflation parameters with `rmultinom(n,1,probs=c(p,1-p))`. This produces a $2 \times n$ matrix, where each column is a random vector consisting of a 0 or 1, which determines if the ith subject is at-risk for Y_1. If the first row is a 1, then the subject is not at-risk for Y_1 (and Y_1 is a structural 0), and if the second row is a 1, then the subject is at-risk for Y_1. Once we determine whether the ith observation is at-risk for Y_1 (e.g. which mixture in (B.1) each observation is from), we need to generate the corresponding bivariate counts. To do this we use the `rpois` function in R. In particular,

- For all subjects for which row one is assigned a value of 1 by `rmultinom`, Y_1 is a structural zero and Y_2 is a Poisson random variable. Thus, Y_1 is assigned

a value of zero (structural zero) and the corresponding positive count for Y_2 can be generated using the `rpois` function with specified parameters n (in this case 1) and λ (in this case $\lambda_2 + \lambda_0$, where values of λ_2 and λ_0 are provided by the user.

- For all subjects for which row two is assigned a value of 1 by `rmultinom`, Y_1 and Y_2 are Poisson random variables with parameters $\mu_1 = \lambda_1 + \lambda_0$ and $\mu_2 = \lambda_2 + \lambda_0$, respectively. This is our bivariate Poisson distribution. To do this we use the `rpois` function in R as follows:

```
X0 <- rpois(1, lam0)
X1 <- rpois(1, lam1)
X2 <- rpois(1, lam2)

Y1[i] <- X0 + X1
Y2[i] <- X0 + X2
```

  where `lam0`, `lam1` and `lam2` are specified by the user. The above process is then repeated n times, where n is the desired number of bivariate observations.

In the adverse event drug safety study in Section 2.3, we assume that both the probability that a subject is not at-risk for migraines, p, and the Poisson parameters, $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$, depend on a single covariate, age. Assume that subjects in the study are all between the ages of 20 and 80 years old, with slightly more subjects closer to age 20 than age 80. Further, assume the mean age for subjects in the study is around 48 years. To reflect this, we generate the subjects' ages from a shifted four-parameter beta distribution. Specifically, $\text{Beta}_{[20,80]}(1.4, 1.6)$, rounded to the nearest whole number.

For examples in the dissertation that do not assume the model parameters depend on covariates (e.g. age), we generate the bivariate counts as described above with constant values of p, λ_0, λ_1, and λ_2. In adverse event drug safety study, where we assume parameters depend on age, slight modifications are required for data

generation. Instead of constant values for p, λ_0, λ_1, and λ_2, these values change depending on subject's age. In particular, for the ith subject, age x_i, we have

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, \tag{B.2}$$

and for $k = 0, 1, 2$,

$$\lambda_{k,i} = \exp(\gamma_{k,0} + \gamma_{k,1} x_i). \tag{B.3}$$

To best assess the proposed methods of prior construction, we want the generated data to accurately reflect the expert opinion in Section2.7. That is, we want the data to reflect the expert's belief that the probability a subject is at-risk for migraines increases with increasing age, and that the number of adverse events increases with increasing age. Specifically, we want the data to reflect the expert's expectations at the ages for which we elicit information (i.e. $x_i = 25$ and $x_i = 65$). We considered two methods to generate the Poisson counts based on the assumed relationship that $\log(\lambda_i)$ is linearly related to age. Both involve working backwards and yield comparable results. One method we considered involved using a scaled beta distribution to transform λ_i on a desired interval (where the desired value for λ_i at 25 and 65 is the minimum and maximum, respectively). We subsequently took the logarithm of these values, and employed maximum likelihood methods in R to obtain estimates for $\gamma_{k,0}$ and $\gamma_{k,1}$ that reflect the desired relationship. Finally, using the maximum likelihood estimates for $\gamma_{k,0}$ and $\gamma_{k,1}$, we can generate values of λ_i which subsequently can be used to generate the bivariate adverse event counts.

An alternative method is to work backwards and set up a system of two equations using the logarithm of (B.3). Namely, set λ_i at desired values for $x_i = 25$ and $x_i = 65$. This yields a system of two equations and two unknowns to solve for estimates of $\gamma_{k,0}$ and $\gamma_{k,1}$. These estimates can be used to generate the λ_i which can subsequently be used to generate the Poisson counts. Similarly, we can work backwards using (B.2) to set up a system of two equations and two unknowns to obtain

estimates for β_0 and β_1, which can subsequently be used to generate the probability that a subject is not at-risk to experience migraines. For our adverse event drug safety study, this yields $\beta_0 = 1.9064$, $\beta_1 = -0.0424$, $\gamma_{0,0} = 0.4397$, $\gamma_{0,1} = 0.0101$, $\gamma_{1,0} = 1.0094$, $\gamma_{1,1} = 0.017$, $\gamma_{2,0} = 1.2797$, and $\gamma_{2,1} = 0.0231$.

APPENDIX C

Data Generation for BZIP Model

The process used to generate data for the BZIP model discussed in Chapter Four is described here. To do this we extend the data generation process is described in Section B, too account for the additional zero-inflation parameters, or mixtures in our BZIP model. A BZIP model can be constructed from a mixture of a point mass at (0,0), two univariate Poisson distributions with parameters μ_1 and μ_2, and a bivariate Poisson (BP) distribution with parameters $(\lambda_0, \lambda_1, \lambda_2)$ as follows:

$$(Y_1, Y_2) \sim \begin{cases} (0,0), & \text{with probability } p_0 \\ (\text{Poisson}(\mu_1), 0), & \text{with probability } p_1 \\ (0, \text{Poisson}(\mu_2)), & \text{with probability } p_2 \\ \text{BP}(\lambda_0, \lambda_1, \lambda_2), & \text{with probability } p_3, \end{cases} \tag{C.1}$$

To generate which of the four possible mixtures an observation is from, we again use the `rmultinom` function in R. In this case we use the `rmultinom` function with specified parameters `n` (sample size), `size` (in this case 1) and `prob` (in this case a vector of length 4, representing the four zero-inflation parameters). This produces a $4 \times n$ matrix, where each column is a random vector consisting of one row with a value of 1 and three rows of 0's, which determines which mixture the ith observation is from. In particular,

- If the first row is assigned a value of, 1 then the ith observation is not at-risk for either adverse event and both Y_1 and Y_2 are assigned a value of zero (i.e. both Y_1 and Y_2 are structural zeros).

- If the second row is assigned a value of 1, then the ith observation is at-risk for Y_1, but not at-risk for Y_2. We generate the Poisson count for Y_1 using the

`rpois` function in R with specified parameters `n` (in this case 1) and `lambda` (in this case $\lambda_1 + \lambda_0$), and assign Y_2 a value of zero.

- If the third row is assigned a value of 1, then the ith observation is not at-risk for Y_1, but at-risk for Y_2. We generate the Poisson count for Y_2 using the `rpois` function in R with specified parameters `n` (in this case 1) and `lambda` (in this case $\lambda_2 + \lambda_0$), and assign Y_1 a value of zero.

- Finally, if the fourth row is assigned a value of 1, then the ith observation is at-risk for both Y_1 and Y_2. We generate the bivarate Poisson counts as described in Appendix B.

To best assess the proposed methods of prior construction, we want to generate data for the adverse event study in Section 4.5 that reflects the expert's beliefs. To reflect the expert's belief that the at-risk probabilities is not dependent on treatment group, we assume constant values for the zero-inflation parameters, $\mathbf{p}$. To reflect the expert's belief the new medication reduces the number of adverse events experienced among at-risk subjects, we can work backwards as described in Section B to obtain the estimates: $\gamma_{0,0} = 1.2528$, $\gamma_{0,1} = -1.2528$, $\gamma_{1,0} = 1.9459$, $\gamma_{1,1} = -1.2528$, $\gamma_{2,0} = 1.3863$, and $\gamma_{2,1} = -0.6931$.

APPENDIX D

Bivariate Zero-Inflated Poisson Model

4.0.1 Example: Demonstrating Nonidentifiability of λ_k's for BZIP Model

Consider an set of outcomes from the hypothetical drug efficacy study. We use the method described in Appendix C to generate data for a sample of $n = 250$ subjects with true values of parameters $\boldsymbol{\lambda} = (4, 7, 1)$. Histograms of the marginal distributions of the data are provided in Figure D.1.

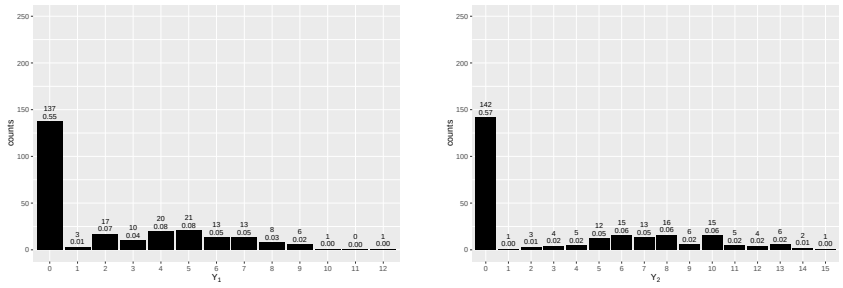

Figure D.1: Marginal distributions of Y_1 (left) and Y_2 (right) for $n = 250$ subjects.

Consistent with what is done in the literature, we place diffuse normal priors on regression coefficients. For example, in the case where the Poisson rates $\boldsymbol{\lambda}$ do not depend on covariates we have for the ith individual that

$$\log(\lambda_{k,i}) = \beta_k, \quad k = 0, 1, 2,$$

where $\beta_k \sim \mathrm{N}(0, \sigma^2 = 100)$. For example, the prior on β_1 (interpret) and the resulting induced prior on the rate of experiencing just adverse event A (λ_1) are shown in Figure D.2.

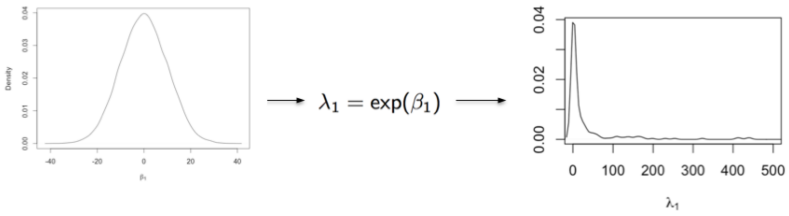

Figure D.2: Induced prior on λ_1 with diffuse prior on β_1: $\beta_1 \sim \mathrm{N}(0, \sigma^2 = 100)$.

Similarly, we induce priors on the rate of experiencing just adverse event B (λ_2) and the rate of experiencing both adverse event A and adverse event B (simultaneously). Figure D.3 shows the resulting posterior densities for λ_0, λ_1 and λ_2.

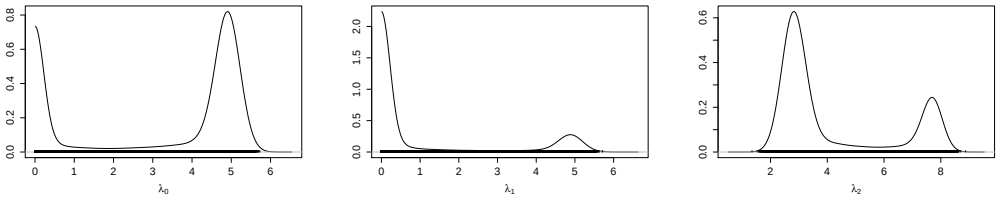

Figure D.3: Posterior densities for λ_0, λ_1 and λ_2 with diffuse priors.

The bimodal posterior densities suggest a lack of convergence. Convergence diagnostics such as the autocorrelation plots, trace plots (see Figure D.6), and Gelman-Rubin statistic further suggest a lack of convergence.

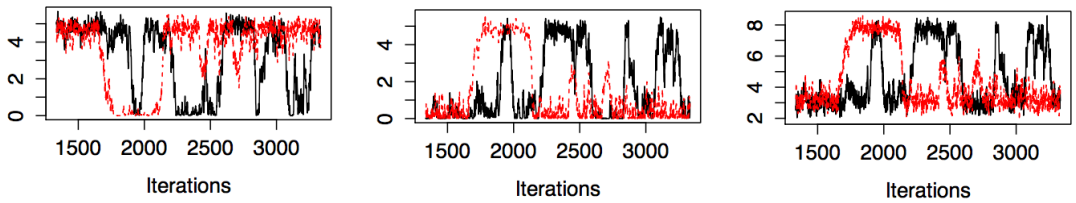

Figure D.4: Trace plots for two MCMC chains with diffuse priors for λ_0 (left plot), λ_1 (middle plot), and λ_2 (right plot).

Thus, posterior results are meaningless. Furthermore, the lack of updating *a posteriori* is characteristic of what is observed for nonidentifiable models. Thus, we propose the culprit for the bimodal posteriors as nonidentifiability. The posterior densities for μ_1 and μ_2 are shown in Figure D.5.

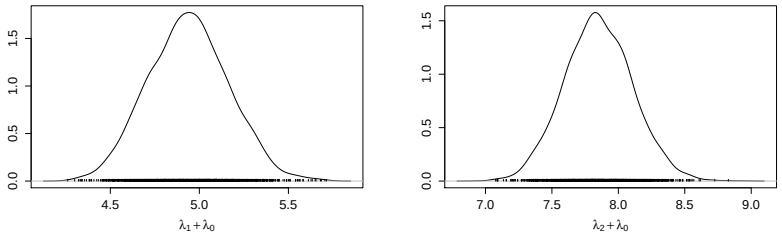

Figure D.5: Posterior densities for $\mu_1 = \lambda_1 + \lambda_0$ and $\mu_2 = \lambda_1 + \lambda_0$.

The unimodal, smooth posterior densities for μ_1 and μ_2 suggest convergence (standard diagnostic tests such as trace plots and the Gelman-Rubin statistic indicate convergence). This is characteristic of identifiable parameters and thus suggest that although the individual summands λ_1, λ_2 and λ_0 are nonidentifiable, the sums, μ_1 and μ_2 are identifiable.

APPENDIX E

Comparison of BPZIP Model and BZIP Model

Table E.1: Comparison of bivariate partial zero-inflated Poisson model and bivariate zero-inflated Poisson model.

Property	BPZIP	BZIP
Y_1	Assumes not all at-risk for Y_1	Assumes not all at-risk for Y_1
Y_2	Assumes all at-risk for Y_2	Assumes not all at-risk for Y_2
Marginal Distributions	$Y_1 \sim \text{ZIP}(p, \lambda_1 + \lambda_0)$ $Y_2 \sim \text{Poisson}(\lambda_2 + \lambda_0)$	$Y_1 \sim \text{ZIP}(p_0 + p_2, \lambda_1 + \lambda_0)$ $Y_2 \sim \text{ZIP}(p_0 + p_1, \lambda_2 + \lambda_0)$
Sources of Zeros	Y_1: structural and sampling Y_2: sampling	Y_1: structural and sampling Y_2: structural and sampling
Number of Mixtures	2	4
Zero-Inflation Parameters	p	$\mathbf{p} = (p_0, p_1, p_2, p_3)$
Poisson parameters	$\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$	$\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2)$
Latent variable representation	$Y_1 : (U, \mathbf{X})$ $Y_2 : \mathbf{X}$	$Y_1 : (\mathbf{U}, \mathbf{X})$ $Y_2 : (\mathbf{U}, \mathbf{X})$

APPENDIX F

Simulation Results Comparing BPZIP and CBPZIP Representation

In Chapter Three we discuss a conditional representation for the BPZIP model and illustrate how this representation provides a route of prior construction for the association parameter, λ_0. We suggest that the conditional representation potentially allows for more posterior updating of the nonidentified parameters, λ_0, λ_1, and λ_2, namely through μ_1. Here we consider a small simulation study with 100 replications to compare the credible interval widths for λ_0, λ_1 and λ_2 between the joint representation of the BPZIP model and the conditional representation of the BPZIP model. We generate 100 data sets for sample sizes $n = 100$ and 200. For each data set, appropriate priors were specified for the joint and conditional representation such that the prior variability for all model parameters is comparable. Simulations were conducted in JAGS using the specification in Section2.3.1. A random sample of the $M = 100$ simulations was selected to check for convergence. Standard diagnostics based on trace plots and the Gelman-Rubin statistic indicate no problems with convergence.

Results for these simulations are summarized below. Tables F.1 and F.2, show the true value for each parameter, as wells as the mean and median of the 100 posterior means for $n = 100$. These tables also include the median width of the 95% credible intervals, and their coverage. Tables F.4 and F.5 include this information for simulations with $n = 200$. As expected, the coverage for λ_0, λ_1 and λ_2 is one for all simulations. These results further illustrated as box plots in Figures F.1 and F.2 for $n = 100$ and Figures F.3 and F.4 for $n = 200$. In each box plot, the horizontal line represents the true value for the corresponding parameter. The center point of each box represents the median of the 100 posterior means. The average

of 100 posterior 2.5th and 97.5th percentiles. Finally, the grey boxes represent ± 1 simulation standard deviation.

Simulation comparing the interval widths for λ_1, λ_2 and λ_0 using the conditional representation versus the joint representation of the BPZIP model.

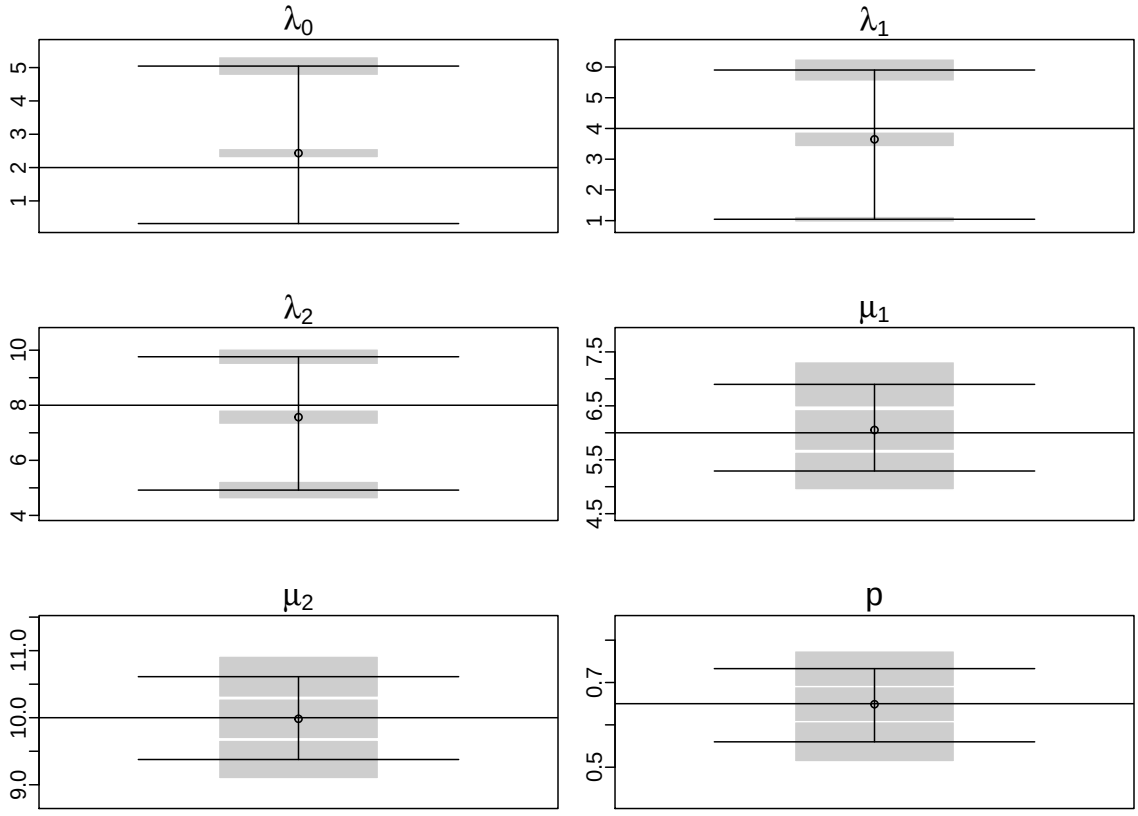

Figure F.1: Simulation results for $n = 100$ and $M = 100$ for the joint representation of BPZIP model.

Table F.3 provides comparison of the average posterior credible interval widths for the Poisson parameters obtained with the joint representation and the conditional representation. Note that the widths for the conditional representation are smaller compared to that obtained with the the joint representation. This further seems to suggest that the conditional representation allows for more posterior updating. Furthermore, the standard deviation of the 100 posterior credible interval widths is

Table F.1: Simulation results for $n = 100$ and $M = 100$ for the joint representation of BPZIP model.

Parameter	Truth	Mean	Median	2.5%	97.5%	Avg. Posterior Width	Coverage
λ_0	2	2.4325	2.3419	0.3200	5.0452	4.7264	1
λ_1	4	3.6457	3.7201	1.0447	5.9017	4.8494	1
λ_2	8	7.5676	7.6540	4.9180	9.7619	4.8841	1
μ_1	6	6.0518	6.0427	5.2913	6.8977	1.6300	0.98
μ_2	10	9.9837	9.9802	9.3767	10.6120	1.2355	0.98
p	0.65	0.6486	0.6494	0.5599	0.7328	0.1728	0.95

Table F.2: Simulation results for $n = 100$ and $M = 100$ for the conditional representation of BPZIP model.

Parameter	Truth	Mean	Median	2.5%	97.5%	Avg. Posterior Width	Coverage
λ_0	2	2.0149	1.9670	0.8014	3.5116	2.7127	1
λ_1	4	4.0397	4.0571	2.5319	5.4554	2.9070	1
λ_2	8	7.9586	7.9961	6.3981	9.4042	2.9779	1
μ_1	6	6.0657	6.0555	5.2508	6.9157	1.6376	0.95
μ_2	10	10.0707	10.0678	9.4636	10.7019	1.2405	0.95
p	0.65	0.6489	0.6500	0.5600	0.7330	0.1726	0.96

smaller for the conditional representation compared to the joint representation for all parameters.

Table F.3: Simulation results for $n = 100$ and $M = 100$ for the joint representation and conditional representation of the BPZIP model.

Parameter	Representation	Prior Width	Mean Posterior Width
λ_0	Joint	10.3256	4.7264
λ_0	Conditional	10.5851	2.7127
λ_1	Joint	12.4168	4.8494
λ_1	Conditional	12.3533	2.9070
λ_2	Joint	13.3676	4.8841
λ_2	Conditional	13.3676	2.9779
μ_1	Joint	19.6538	1.6300
μ_1	Conditional	20.0135	1.6376
μ_2	Joint	20.4389	1.2355
μ_2	Conditional	20.7644	1.2405

The smaller 95% credible interval widths for λ_0, λ_1 and λ_2 with the conditional representation compared to the joint representation suggests more posterior updating of these parameters for $n = 100$. This is also seen below with $n = 200$. Furthermore, the 95% credible interval widths for λ_0, λ_1 and λ_2 for the joint representation are comparable for $n = 100$ and $n = 200$. On the other hand, the 95% credible interval for the conditional representation for $n = 200$ are slightly less wide than those for $n = 100$. Additional simulations are needed to further investigate this. Furthermore, the standard deviation of the 100 posterior credible interval widths is smaller for the conditional representation compared to the joint representation for all model parameters.

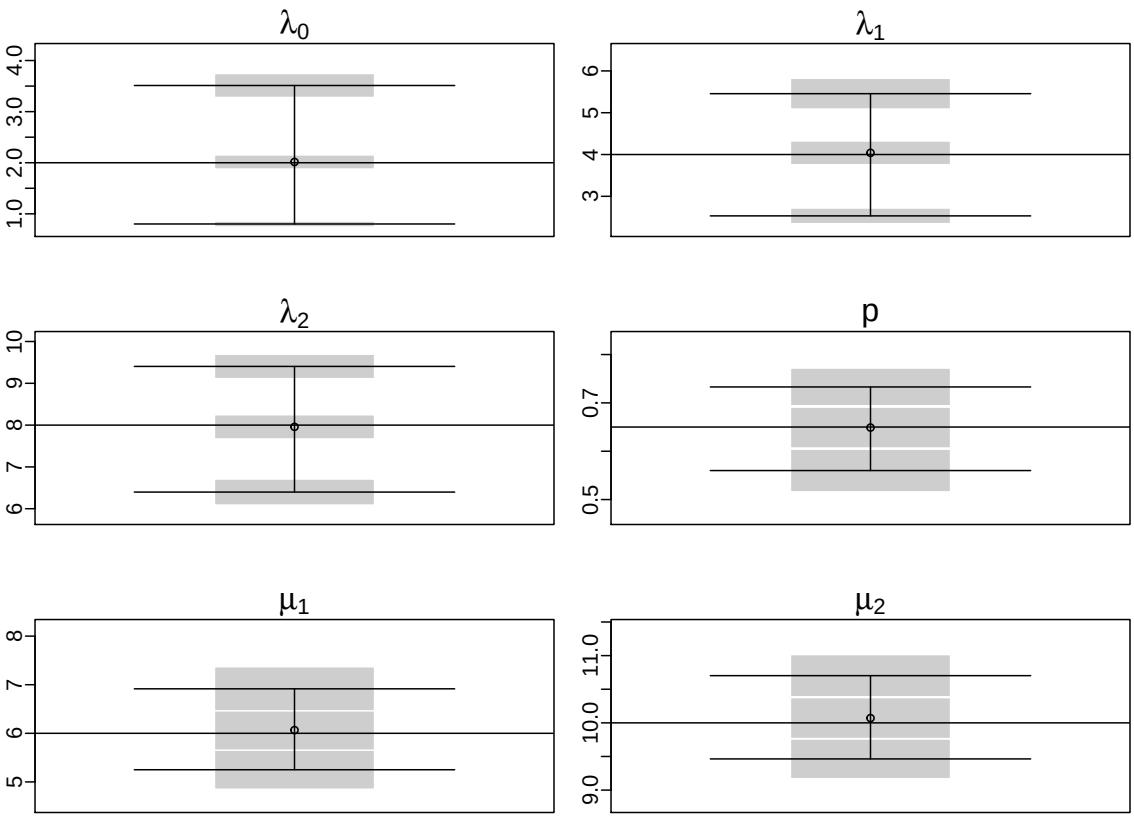

Figure F.2: Simulation results for $n = 100$ and $M = 100$ for conditional representation of the BPZIP.

Table F.4: Simulation results for $n = 200$ and $M = 100$ for the joint representation of the BPZIP model.

Parameter	Truth	Mean	Median	2.5%	97.5%	Avg. Posterior Width	Coverage
λ_0	2	2.4190	2.3338	0.3206	5.0234	4.7010	1
λ_1	4	3.6359	3.7212	1.0614	5.8249	4.7662	1
λ_2	8	7.5599	7.6420	4.9455	9.6984	4.7711	1
μ_1	6	6.0639	6.0589	5.4845	6.6774	1.1770	0.95
μ_2	10	9.9639	9.9624	9.5332	10.4058	0.8743	0.95
p	0.65	0.6537	0.6542	0.5888	0.7156	0.1269	0.95

Table F.5: Simulation results for $n = 200$ and $M = 100$ for the conditional representation of BPZIP model.

Parameter	Truth	Mean	Median	2.5%	97.5%	Avg. Posterior Width	Coverage
λ_0	2	2.0102	1.9640	0.8022	3.4631	2.6618	1
λ_1	4	4.0061	4.0341	2.5496	5.3052	2.7578	1
λ_2	8	7.9544	7.9939	6.4537	9.2518	2.8008	1
μ_1	6	6.0160	6.0111	5.4579	6.6039	1.1532	0.96
μ_2	10	9.9851	9.9832	9.5526	10.4278	0.8744	0.97
p	0.65	0.6530	0.6536	0.5882	0.7151	0.1269	0.93

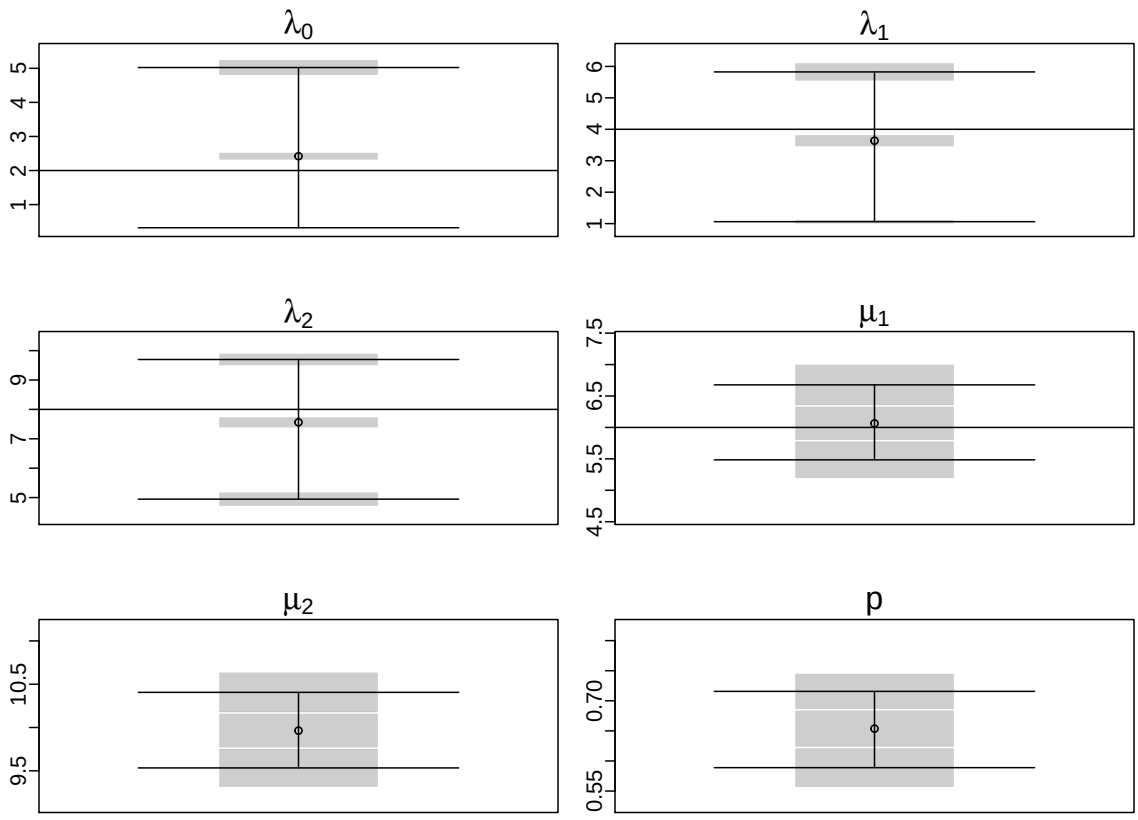

Figure F.3: Simulation results for $n = 200$ and $M = 100$ for joint representation of the BPZIP model.

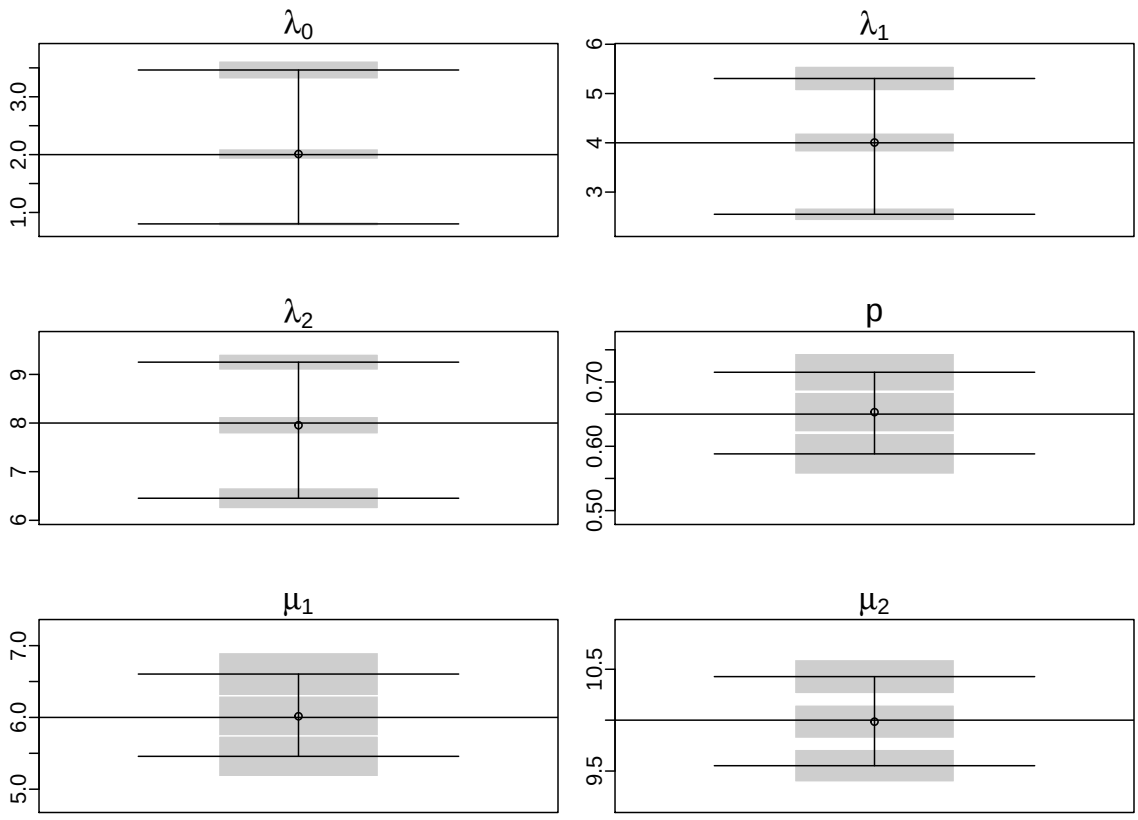

Figure F.4: Boxplots for $n = 200$ and $M = 100$ for conditional representation of the BPZIP.

BIBLIOGRAPHY

Aitchison, J. (1986), *The statistical analysis of compositional data*, England;United Kingdom.

AlMuhayfith, F. E., Alzaid, A. A., and Omair, M. A. (2015), "On bivariate Poisson regression models," *Journal of King Saud University - Science*, 28, 178–189.

Angers, J.-F. and Biswas, A. (2003), "A Bayesian analysis of zero-inflated generalized Poisson model," *Computational Statistics and Data Analysis*, 42, 37–46.

Arnold, B., Castillo, E., and Sarabia, J. (2004), "Distributions with Beta Conditionals," in *Handbook of Beta Distribution and Its Applications*, eds. Gupta, A. and Nadarajah, S., New York: Marcel Dekker, chap. 10, pp. 255–282.

Baughman, A. L. (2007), "Mixture Model Framework Facilitates Understanding of Zero-Inflated and Hurdle Models for Count Data," *Journal of Biopharmaceutical Statistics*, 17, 943–946.

Bedrick, E. J., Christensen, R., and Johnson, W. (1996), "A new perspective on priors for generalized linear models," *Journal of the American Statistical Association*, 91, 1450–1460.

Berkhout, P. and Plug, E. (2004), "A bivariate Poisson count data model using conditional probabilities," *Statistica Neerlandica*, 58, 349–364.

Bermdez, L. and Karlis, D. (2011), "Bayesian multivariate Poisson models for insurance ratemaking," *Insurance Mathematics and Economics*, 48, 226–236.

Bermdez, L. and Morata, L. (2009), "A priori ratemaking using bivariate Poisson regression models," *Insurance Mathematics and Economics*, 44, 135–141.

Brooks, S. P. and Gelman, A. (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Chen, M.-H. and Ibrahim, J. G. (2003), "Conjugate Priors For Generalized Linear Models," *Statistica Sinica*, 13, 461–476.

Cohen, A. (1963), "Estimation in Mixtures of Discrete Distributions," *Proceedings of the International Symposium on Discrete Distribuions*, 373–378.

Congdon, P. (2005), *Bayesian models for categorical data*, Chichester;New York: Wiley.

Dagne, G. A. (2004), "Hierarchical Bayesian Analysis of Correlated Zero-inflated Count Data," *Biometrical Journal*, 46, 653–663.

Dong, C., Richards, S. H., Clarke, D. B., Zhou, X., and Ma, Z. (2014), "Examining signalized intersection crash frequency using multivariate zero-inflated Poisson regression," *Safety Science*, 70, 63–69.

Dwivedi, A. K., Dwivedi, S. N., Deo, S., and Shukla, R. (2010), "Statistical models for predicting number of involved nodes in breast cancer patients," *Health*, 2, 641–651.

Elfadaly, F. G. and Garthwaite, P. H. (2013a), "Eliciting Dirichlet and ConnorMosimann prior distributions for multinomial models," *TEST*, 22, 628–646.

— (2013b), "Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models," *Statistics and Computing*, 27, 449–467.

— (2013c), "PEGS-Dirichlet," .

— (2015), "Eliciting prior distributions for extra parameters in some generalized linear models," *Statistical Modelling*, 15, 345–365.

FDA (2010), "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," *Guidance for Industry and Food and Drug Administration*.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005), "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*, 100, 680–700.

Gelman, A. (2006), "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, 1, 515–534.

Gelman, A. and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.

Ghosh, S. K., Mukhopadhyay, P., and Lu, J.-C. (2006), "Bayesian analysis of zero-inflated regression models," *Journal of Statistical Planning and Inference*, 136, 1360 – 1375.

Gustafson, P. (2004), *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*, Boca Raton: Chapman & Hall/CRC.

— (2015), *Bayesian inference for partially identified models: exploring the limits of limited data*, vol. 141, Boca Raton, Florida: CRC Press.

Gustafson, P., Hossain, S., and MacNab, Y. C. (2006), "Conservative Prior Distributions for Variance Parameters in Hierarchical Models," *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 34, 377–390.

He, H., Tang, W., Wang, W., and Crits-Christoph, P. (2014), "Structural zeroes and zero-inflated models," *Shanghai archives of psychiatry*, 26, 236.

Holgate, P. (1964), "Estimation for the Bivariate Poisson Distribution," *Biometrika*, 51, 241–245.

Hu, M.-C., Pavlicova, M., and Nunes, E. V. (2011), "Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an HIV-Risk Reduction Intervention Trial," *The American Journal of Drug and Alcohol Abuse*, 37, 367–375.

Hughes, G. and Madden, L. V. (2002), "Some methods for eliciting expert knowledge of plant disease epidemics and their application in cluster sampling for disease incidence," *Crop Protection*, 21, 203–215.

Jang, H., Lee, S., and Kim, S. W. (2010), "Bayesian analysis for zero-inflated regression models with the power prior: Applications to road safety countermeasures," *Accident Analysis and Prevention*, 42, 540–547.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997), *Discrete multivariate distributions*, New York: Wiley.

Karlis, D. and Ntzoufras, I. (2003), "Analysis of Sports Data by Using Bivariate Poisson Models," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52, 381–393.

Karlis, D. and Tsiamyrtzis, P. (2008), "Exact Bayesian modeling for bivariate Poisson data and extensions," *Statistics and Computing*, 18, 27–40.

Kassahun, W., Neyens, T., Faes, C., Molenberghs, G., and Verbeke, G. (2014), "A zero-inflated overdispersed hierarchical Poisson model," *Statistical Modelling*, 14, 439–456.

King, S. Y., Kung, M. S., and Fung, H. L. (1984), "Statistical prediction of drug stability based on nonlinear parameter estimation," *Journal of pharmaceutical sciences*, 73, 657.

Klar, B. (2015), "A note on gamma difference distributions," *Journal of Statistical Computation and Simulation*, 85, 3708–3715.

Kocherlakota, S. and Kocherlakota, K. (1992), *Bivariate discrete distributions*, vol. 132, New York: M. Dekker.

Lambert, D. (1992), "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing," *Technometrics*, 34, 1–14.

Li, C.-S., Lu, J.-C., Park, J., Kim, K., Brinkley, P. A., and Peterson, J. P. (1999), "Multivariate Zero-Inflated Poisson Models and Their Applications," *Technometrics*, 41, 29–38.

Liu, H. and Powers, D. A. (2007), "Growth Curve Models for Zero-Inflated Count Data: An Application to Smoking Behavior," *Structural Equation Modeling*, 14, 247.

Liu, Y. and Tian, G.-L. (2015), "Type I multivariate zero-inflated Poisson distribution with applications," *Computational Statistics & Data Analysis*, 83, 200–222.

Majumdar, A. and Gries, C. (2010), "Bivariate zero-inflated regression for count data: A bayesian approach with application to plant counts," *International Journal of Biostatistics*, 6, 1–24.

McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models*, London: Chapman & Hall, 2nd ed.

Mohammadi, T., Kheiri, S., and Sedehi, M. (2016), "Analysis of Blood Transfusion Data Using Bivariate Zero-Inflated Poisson Model: A Bayesian Approach," *Computational and Mathematical Methods in Medicine*, 2016, 1–7.

Morita, S., Thall, P. F., and Muller, P. (2008), "Determining the Effective Sample Size of a Parametric Prior," *Biometrics*, 64, 595–602.

Nadarajah, S. and Kotz, S. (2005), "On the Product and Ratio of Gamma and Beta Random Variables," *Allgemeines Statistisches Archiv*, 89, 435–449.

Neelon, B. H., OMalley, A. J., and Normand, S.-L. T. (2010), "A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use," *Statistical Modelling*, 10, 421–439.

Nikoloulopoulos, A. K. and Karlis, D. (2009), "Modeling Multivariate Count Data Using Copulas," *Communications in Statistics - Simulation and Computation*, 39, 172–187.

Ntzoufras, I. and Karlis, D. (2005), "Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R," *Journal of Statistical Software*, 14.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., D.J., J., J.E., O., and T., R. (2006), *Uncertain Judgements: eliciting expert probabilities*, Chichester: Wiley.

O'Hagan, A. and Forster, J. (2005), "Kendall's Advanced Theory of Statistics, Vol. 2B: Bayesian Inference (2nd Ed.)," *Journal of the American Statistical Association*, 100, 1465–1466.

Plummer, M. (2003), "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," .

Robert, C. P. (1994), *The Bayesian choice: a decision-theoretic motivation*, New York: Springer-Verlag.

Rose, C. E., Martin, S. W., Wannemuehler, K. A., and Plikaytis, B. D. (2006), "On the Use of Zero-Inflated and Hurdle Models for Modeling Vaccine Adverse Event Count Data," *Journal of Biopharmaceutical Statistics*, 16, 463–481.

Sturtz, S., Ligges, U., and Gelman, A. (2005), "R2WinBUGS: A Package for Running WinBUGS from R," *Journal of Statistical Software*, 12, 1–16.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Walhin, J. F. (2001), "Bivariate ZIP Models," *Biometrical Journal*, 43, 147–160.

Wang, K., Lee, A. H., Yau, K. K. W., and Carrivick, P. J. W. (2003), "A bivariate zero-inflated Poisson regression model to analyze occupational injuries," *Accident Analysis and Prevention*, 35, 625–629.

Wilks, S. S. (1962), *Mathematical statistics*, New York: Wiley.

Yuen, H.-K., Chow, S.-C., and Tse, S.-K. (2015), "On Statistical Tests for Homogeneity of Two Bivariate Zero-Inflated Poisson Populations," *Journal of Biopharmaceutical Statistics*, 25, 44–53.